TECNOLOGIA SETÚBAL
POLITÉCNICO SETÚBAL
SCHOOL • POLYTECHNIC UNIVERSITY

Edeilson Ferreira da Silva

Developing a Data Analysis and Visualization Educational Decision Support System (EDSS)

Dissertation submitted as a partial requirement to obtain the degree of **Master in Software Engineering**

Júri
Presidente (Prof. Dr., Prof. Claúdio Miguel Garcia Loureiro dos Santos Sapateiro, Polytechnic Institute of Setubal)
Advisor (Prof. Dr., Miguel A. Guevara Lopez, Polytechnic Institute of Setubal)
Vogal (Prof. Dr., Osvaldo Rocha Pacheco, University of Aveiro)

December 2024

Abstract

Institutional language providers are constantly seeking to deliver better experience for its students, a movement that fosters implementation of different technologies with focus on this industry. Following literature, to accomplish this objective any designed strategy must involve a comprehensive way to deal with its historical data. The quality of its generated data will allow institutions to see what happened in the past, understand the current situation and make informed decisions for future improvements. Conjecture that encourages the development of an Educational Decision Support System (EDSS). To create it, a Systematic Literature Review was carried out to understand the trends in digital solutions for the educational management field, followed by consistent methodology. Methodology Design Science Research (DSR) was the path to respond to this class of problem by building a satisfactory solution with pragmatic validity. The result shows that EDSS can lead educational institutions to embrace data culture, adopt data governance and consequently get more quality. It also reveals the system was easy to operate but with room for improvement regarding data update rate.

Keywords: academic management; data-informed decision making; educational decision support system (EDSS); english language provider; institutional administration.

Resumo

Instituições de ensino de idiomas estão constantemente buscando oferecer a melhor experiência para seus estudantes, movimento que promove a implementação de diferentes tecnologias com foco neste setor. Na literatura, para atingir este objetivo, qualquer estratégia projetada deve envolver uma maneira abrangente de lidar com seus dados históricos e na qualidade deste ativo, isso permitirá uma visão consistente do passado, compreensão do presente e fundamento para decisões futuras. Conjuntura que motivou o desenvolvimento de um Sistema de Suporte à Decisão Educacional (EDSS). Para criá-lo, foi realizada uma Revisão Sistemática da Literatura para entender as tendências em soluções digitais para o campo da gestão educacional, seguida de metodologia consistente. A metodologia Design Science Research (DSR) foi o caminho para responder a esta classe de problema construindo uma solução satisfatória com validade pragmática. Os resultados mostram que o EDSS pode levar as instituições educacionais a abraçar a cultura de dados, adotar a governança de dados e, consequentemente, garantir a qualidade na aquisição dos dados, que é base para tomada de decisões. Além disso, revela que o sistema era fácil de operar, mas com espaço para melhorias em relação à taxa de atualização de dados.

Palavras-chave: administração institucional; gestão acadêmica; provedor de língua inglesa; sistema de suporte à decisão educacional (EDSS); tomada de decisão baseada em dados.

Developing a Data Analysis and Visualization Educational Decision Support System (EDSS)

Contents

List of Figures

List of Tables

List of Acronyms

CA	*Correspondence Analysis*
CMS	*Content Management Systems*
CRM	*Customer Relationship Management*
DR	*Design Research*
DS	*Design Science*
DSR	*Design Science Research*
DSS	*Decision Support System*
DW	*Data Warehouse*
ERD	*Diagram Entity Relationship*
EDSS	*Education Decision Support System*
ETL	*Extraction, Transformation and Load process*
GDPR	*General Data Protection Regulation*
IELTS	*International English Language Testing System*
ILEP	*Following the Interim List of Eligible Programmes*
ISSM	*Information System Success Model*
KPI	*Key Performance Indicators*
NPS	*Net Promoter Score*
OLAP	*Online Analytical Processing*
PRISMA	*Reporting Items for SystematicReviews and Meta-Analyses*
RDMS	*Relational Database Management System*
SQL	*Structure Query Language*
TAM	*Technology Acceptance Model*

1. Introduction

As any other field, educational providers are constantly looking to be up to date with the most recent technological tools to be aligned with market needs and deliver better services and, for private institutes, make profits. To adapt themselves, remain competitive and relevant, entities start developing their own solutions to monitor and manage their own operation. Presently, we can easily find examples with the use of technologies to assist institutions to better understand its own scenario using predictive learning analytics techniques (Sghir *et al.* 2023), improve its management support through concepts like mobile administration (Terence *et al.* 2021) and implementation of blockchain approaches to boost data governance (Lianny *et al*. 2023).

Considering the nature of the business, it is very common to find technologies that analyse students' performance over learning behaviour and support them with information to achieve better results on exams, however, it is not the focus of this study. This research target follows another vertent of this ambience, the educational management. Every educational institution has, or should have, the premise of providing an environment with great conditions for students. Situation that can have its potential enhanced by implementing digital solutions that are relevant for both scenarios academic performance and educational management.

To Berges *et al*. (2021), understanding that updating, adopting or replacing technologies in a near future is necessary to avoid obsolescence of processes and services that could consequently generate loss of competitiveness and consequently students. Implementing new resources involves management reform and change several aspects performed under a systematised approach (Zhao, 2023), but it is also pretty quite clear that most of institutions already realise it and are moving towards the implementation of digital solutions that prioritises reliable and available information to support its operation model from basic tasks to long-term strategy. Phenomenon that certainly leads to data dependency with collection, storage and processing (Gaftandzhieva et al. 2023).

In this context, we have invited as the main use case, an institution that provides English Language Programmes located in Dublin, Ireland. The institute delivers training like English for general, academic and professional purposes to non-English speakers. Following the Interim List of Eligible Programmes[1] (ILEP), governmental guidelines with regulation for the sector, 1136 language programs, 74% located in Dublin, provided by 52 different institutes are available in 2024. Observing the amplitude of this specific market, it is easy to understand how high the level of competitiveness and how strategic is to be supported with cutting edge technology that helps the management with future decisions.

Following an agreement where the institution's name should be kept confidential mainly because the basis for this study look over its operational process and scrutinise its private datasources. In general, the institution can be considered an educational provider consolidated at the market for its almost 10 years of activities, operating in different cities with branches located at the capital, demonstrating resilience and adaptiveness demonstrated during the pandemia. Therefore, although it manifests a strong position between other players, the management is aware that to keep performing well the entire process must be reviewed from time to time to ensure its compliance and high performance on daily operation. More than that, it is conscious about adapting new solutions that support its growth strategy.

Currently, to manage its activities, the institution uses a Content Management Systems (CMS) connected with a Relational Database Management System (RDMS), a third party Customer Relationship Management (CRM) with its own datasource and several linked spreadsheets that deliver some parallel solutions. This set of tools support operations through the registration and manipulation of data regarding agents, professors, classrooms, students, exams, assessments, etc. The cluster of tools acts as an interface between branches allowing the interchange of information but has very limited analytical resources. Despite its utility to perform

[1] https://www.irishimmigration.ie/coming-to-study-in-ireland/what-are-my-study-options/interim-list-of-eligible-programmes-ilep/

ad-hoc routinary tasks the system was not designed to provide information that focus on support management to make formed decisions.

Figure 1.1 presents the current model for students to engage the institution under an eight months language program. Generally the program includes a General English course and a mandatory exam, the type of exam can change according to student level. Although it can suffer some variation between institutions, each course is composed of 25 weeks classes and 4 weeks holidays and here the nuances create some complexity for the scheme. Each student has the opportunity to enrol in classes at the beginning and opt to take holidays at the end of their course, while others can take it in the middle.

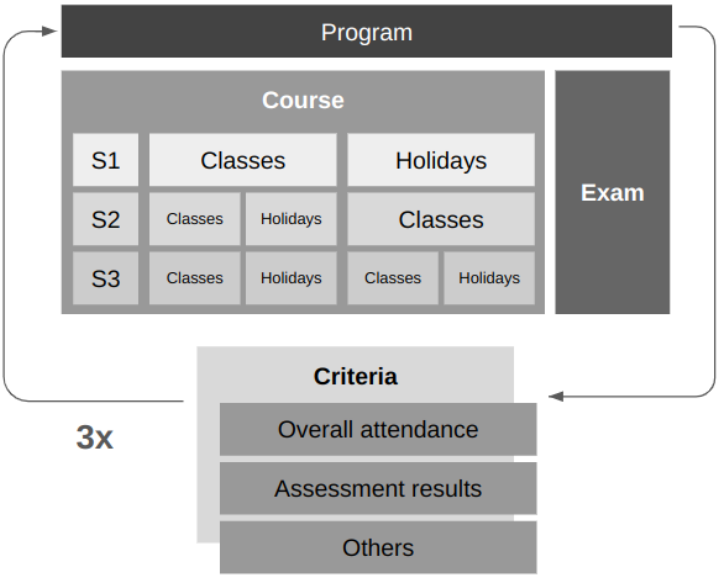

Figure 1.1 - 8 months program model.

A third scenario still can come up with classes-holidays-classes-holiday letting the comprehension and decision more susceptible to mistakes. Notwithstanding, approved leaves regarding different reasons and emergency situations faced by students can create a turbulent scenario. With all that, every student can renew course two times based on some criteria such as attendance, assessment results between others. Conjecture that evidence the need for a tool to conduct analysis

initially based on its own compiled data and presents real time information. In this way, as an educational provider looking for understanding its own behaviour using data it is indispensable to raise questions such as:

Q1 - How many students are starting classes at different levels? Are the teachers enough or should the institution hire?
Q2 - What about the classrooms, is the space enough? Should new classrooms in a new building be booked or can we cancel some contract?
Q3 - Is the distribution of students per class reasonable, when a replacement level test should be applied?
Q4 - What is the average time of a student by class level, are they happy with the provided learning environment?
Q5 - Is there any relevant factor that helps to predict what students will keep studying at the institution?
Q6 - Who are the students with the greatest potential to renew their course? Do they follow any pattern?

To answer these questions we propose the development of a **custom-made Educational Decision Support System (EDSS) as an effective way to assist the institution to have a clear vision of its management situation and start to make informed decisions based on its own generated data.** Following Sharda et al. (2019), EDSS fits in the category of Online Analytical Processing (OLAP) system and relies on data integration to present resumes with the intention of providing information, generating insights and supporting the decisions. To Bimonte et al. (2021) using this data structure can potentialise performance because it is supported by a mental model, which makes the perception and interpretation easy for humans.

Contributing to the discussion Berges *et al.* (2021) defend that indicators derived from OLAP analysis should be associated with a dashboard that displays a set of relevant standards to inform the stakeholders. Moreover, characteristics of drill-down and roll-up the management can perform a faster analysis and support the decisions (Barros *et al.* 2023). Uvalieva *et al*. (2015) punctuate improvement of education management and the removal of social tensions as most relevant

administrative objectives in the educational field. Once framed, those elements can contribute to reducing uncertainty and improving decision's quality by administration.

In brief, EDSS is a digital solution that focuses on gathering those elements to provide educational decision support for school administration presenting real time information to understand, monitor and predict situations from historical data. In other words, our intention is to design an innovative way for educational management bodies to get answers for specific questions, gain actionable insights and find founded information to make decisions towards positive changes that helps stakeholders to meet their goals. Fortunately, the advance of technology allows, independent of its size, institutions to design such structures with considerable low cost and few resources.

1.1. Motivation and objectives

Given the need to improve the decision-making on education management, we intend to take advantage of knowledge acquired during the Master Degree in Software Engineering, specifically from subjects like Information Systems, Data Analysis and Information Visualization. Together these topics nourish an overview of data, its importance and potential usage. Secondly, the decision to use this specific context comes from the researcher's experience with the institution, knowing the current process aligned with acquired expertise created an opportunity of improvement.

Most decisions are being taken by the institution's board naturally based on its operation but it is also known the current system was not developed to support those decisions using data. We comprehend that those decisive moments can be better sustained with a solution equipped with transparency, precision, accuracy through indicators that explains a given scenario information. Having a tool that complements the current administrative system will require some individual objectives though.

- Integrate data consolidating it from existing datasource;

- Develop descriptive statistic elements to provide current information;
- Implement real-time analytics to give information that support decision-making;
- Create a storyline to answer why students has chosen the institution;
- Perform exploratory analysis from a storyline to generate insights and contribute with decision-making;
- Develop an intuitive, user-friendly interface that allows non-technical staff to easily navigate, utilise and interpret EDSS.

To achieve that many tasks will be performed such as identifying the users, elucidation of requirements using different techniques such as interview and survey, define the most valuable visualisations based on prioritisation techniques, processing data points, analyse algorithms results and EDSS as a whole. Those objectives and tasks will be described and discussed throughout this document.

1.2. Expected Results

Achieving a platform with a technology readiness level (TRL) of 5-6 as final deliverable, we aim to provide nuanced insights that contribute to strategic decision-making in the academic management field, specifically, English learning providers. It was expected that the produced framework can offer a comprehensive range of functional capabilities to aid decision processes through an intuitive and user-friendly analytic system that meets a minimal performance considering the provided infrastructure and lead the management to gain information based on its own historical data. We believe that it could positively impact the way the settlements are made and consequently save time, financial resources and human capital to solve empirical problems.

1.3. Contributions from dissertation

Many aspects from this study can be considered beneficial for the institution, beginning with the sense of awareness when it allows the investigation to scrutinise its processes looking for possible gaps and improvement opportunities. By setting the lights over its operational process the management reinforce its knowledge by consolidating what is already known, figure out unknown divergences that are currently happening or even find opportunities to improve its operation. Scenarios that obviously need some solid information and that is another point this dissertation can contribute.

Looking for patterns in institutional data can spawn evidence to support decisions for the institution as a whole. This probe can lead to the creation of several performance indicators that can be used to measure the operational efficiency and work as a fundamental mechanism for accurate decision-making. The advantage here can be seen beyond the management level, for example, when the department starts to understand the impact of its input for other areas.

To be efficient, the decision-making has to rely on outcomes derived from this investigation and to achieve that the data was compiled using the basic concept of statistics with descriptive and exploratory methods. To improve the final user experience the dissertation has performed different machine learning algorithms to support the decision-making through an intuitive digital tool available at computers running under the institution's networking, which makes it fast to access, easy to maintain and gives a safety layer.

1.4. Dissertation structure

For a better comprehension of this dissertation the document was structured to communicate the taken path in a comprehensive way. The introduction naturally brought the context and circumstances that it was developed, the motivation and objectives followed by expected results gave a framework on how the problematic

was understood, faced and a solution proposed. To complete it, we have discussed the main contributions of this dissertation to the invited institution.

The next chapter brings a Systematic Literature Review to ground the theory and understand what literature has already investigated, what are the common problems and main proposed solutions. With that the research was able to summarise the findings to get the directions from the studied subject; The chapter followed up with the applied Method Design. At this point a convenient research method has been explored with the intention of supporting the investigation as a whole, how it was conducted and outcomes evaluated.

The result of proposed Methodology is presented at subsection 2.2. Based on Design Science Research the Educational Decision Support System (EDSS) presents its own version of the method with some areas and attributes that give robustness to the investigation. Once the investigation path is presented, characteristics of EDSS are discussed with details, debating available features, necessary technologies to release EDSS, data analysis and applied machine learning algorithms.

Next we discuss results with performance of visual elements and how the most efficient element has been chosen. A process that involved statistical tests such as A/B test, heuristic evaluation and inferences, between different analyses from applied algorithms. Then we complete with a conclusion and ideas for future work.

2. Systematic Literature Review

From the objective of implementing EDSS for educational management we first proceed with the intention of understanding where this type of system is usually applied. This section will describe the broad picture that has been captured in order to acquire knowledge about the theme, how they are related, what are the trending topics and its major contributions. To achieve it, we have followed a systematic review protocol based on Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) to give transparency during the subject identification, selection and critical appraising of related paperworks have been done (Page et al. 2021).

To collect those papers the investigation has followed few established criteria on PRISMA guidelines[2] such as information identification and restrictions applied. The search was delineated based on descriptors that represent the technology and the context, which has resulted in queries presented at Figure 2.1. To resume, the repository Web of Science[3] has returned 33 registers while Scopus[4] giving us a total of 1,463 results. To get more relevant results at the second query we have narrowed the search by delimiting it with the search on fields Title, Abstract and Keywords and a limitation regarding the language, as a result 62 documents were found.

Repository	Query	Records
Web of Science	((ALL=("Decision Support System" OR DSS)) OR ALL=("Online Analytical Process" OR OLAP)) AND (ALL=("Education* Management") OR ALL=("Academic administration"))	33
Scopus	(TITLE-ABS-KEY ("decision support system" OR DSS) OR TITLE-ABS-KEY ("online analytical processing" OR OLAP) AND TITLE-ABS-KEY ("education* management") OR TITLE-ABS-KEY ("academic administration")) AND (LIMIT-TO (LANGUAGE , "English"))	62

Figure 2.1 - Applied queries.

[2] http://prisma-statement.org/PRISMAStatement/FlowDiagram
[3] https://www.webofscience.com
[4] https://www.scopus.com

Once the registers were combined, 15 duplicate records were removed considering the criteria of inclusion and exclusion. Those characteristics carefully decide whether a study was eligible or not based on characteristics such as:

- Relevance - The paper must explicitly inform that is working with the DSS or OLAP;
- Context - As the proposed solution have been addressed to a education scenario, the paper should present some implementation, analyse or reflection about the it;
- Published year - As the intention is to get what is most recent paperwork in that area a period of 10 years have been set;
- Availability - The document must be completely available for the analysis; and,
- Language - Only documents in English where kept; were considered at the inclusion.

In parallel, the exclusion criteria were defined by: Irrelevance - Investigations that mention systems created to support decision but are not effectively working with that into educational context; and Out of the context - When the subject is education is very common to find studies and purposed that aims to understand the environment phenomenon from a learning perspective and education quality, as the scope of this work covers education management or academic administration these papers must consequently been removed, resulting on 46 selected for the next step. Details available at Appendix A - Screening from selected studies.

The procedure has continued by screening the title and abstract of the records. A critical and exploratory reading were done under inclusion and exclusion criteria to select relevant records. At this stage two registers were not available and consequently removed, as a result, a corpus with 14 registers have been selected with eligibility to a full reading, here it is important to reinforce that elements set for inclusion or exclusion were, again, considered. As PRISMA allows addition of new items at this stage of the process, 4 new items from different information sources were added, the inclusion criteria follows the previous standard and the result

detailed at Appendix B - Notes from PRISMA eligible readings. Figure 2.2 gives an overview of the selection process.

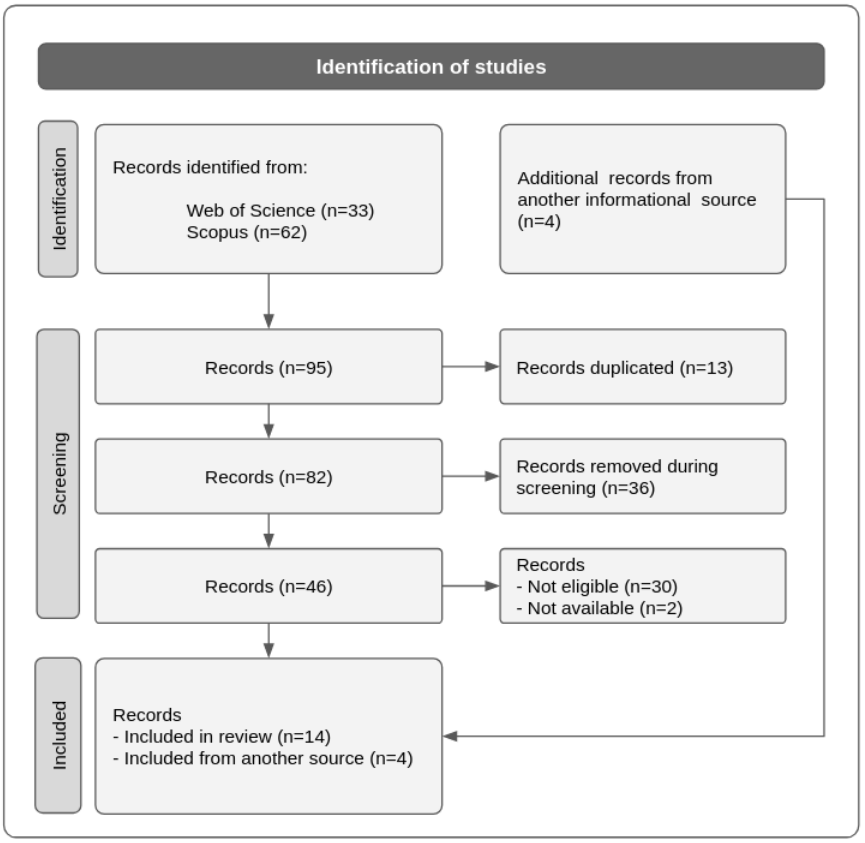

Figure 2.2 - State-of-the-Art selection process.

The next section presents a summary of the analysis from selected records concerning the idea of EDSS as the main challenge for the educational management context.

2.1. **Summarising the finds**

The current section describes how the selected papers have been scrutinised and discussed under the light of EDSS development. A complete reading and a rigorous analysis of those documents are established as requirements from PRISMA, the framework however, does not establish a limitation for the investigated characteristics. With DSR artefact as directional axis of this investigation, it was defined that attributes like type of problem faced, proposed solution, applied

algorithms, methods and the results would give a dimension of how digital solution evolving decision-making support tools are being used by educational institutions.

Education management can face several issues in its operation and our inference from the research corpus concludes that several problems are related with technology management, most of them regarding the lack of ability to deal with the massive volume of data. Sometimes fragmented in different data sources resulting in manual processing with inefficient communication and lack of awareness about the strategic use of data. To Xie and Chu (2022) investing in information technology and making efforts to adapt itself to the most recent technological advance is inevitable.

Becoming data-driven and taking advantage of the academic administration seems to be the answer but several aspects of the business can become a challenge for this shift. To Skittou *et al.* (2022) a very significant objection in this context is data collection because of its nature as structured, semi-structured or unstructured, thus it can be formatted or unformatted. Gaftandzhieva *et al.* (2023) addresses the problem by developing an autonomous data warehouse where the data in an appropriate format could be stored and then, with a proper data mining process, extract meaningful information. In the same way J. Chen (2021); Y. Chen (2017); Skittou *et al*. (2022) advocate data warehouses as components for their solution.

Y. Chen (2017) delineates that to build a decision support tool is essential to combining data warehouse technology and data mining techniques are criterias to building a decision-making tool to support educational institutions. The author understands the development of a should include servers to perform Extraction, Transformation and Load process (ETL) process and store raw data, provide OLAP tools to work on the selected data, a server to store the already analysed data and provide the service, and be accessible through a web browser. Driving our understanding that it enhances the process and takes more responsibility over the non-structured data.

In parallel, the corpus presents a trend on use of Data Mining techniques to generate a possible solution. Between the proposals Y. Liu (2020) warns of the fact

that problems like data recognition and analysis can sometimes take longer and for that reason it is interesting to ponder on how to use the technique. The author sees different approaches using data source → internet/business data/Internet of things; or data processing layer → data acquisition/storage/processing and visualisation; Furthermore, reinforces that it can suffer variation accordingly with the environment which is applied.

To encapsulate this process, Dik *et al.* (2014) state the need to be prepared to store a vast amount of characteristics, open for integrations regarding different technologies and user-friendly. Skittou *et al*. (2022) complements by mentioning how the web-based management system has been facilitating the interconnection for the stakeholders which is fundamental. At this point our synthesis from records, except Xiao *et al.* (2023) that focuses on face recognition, other efforts seek to address common issues like data fragmentation, inefficient decision-making, and the need for objective analysis in education management, which reinforce the need for a robust application.

Records also present different approaches to develop systems that supports decisions, for instance the application of Analytic Hierarchy Process devoted for hierarchical decision-making, to work with situations where quantitative data are not available, with that the system could provide ways to measure attributes (Uvalieva *et al.* 2015) and Spatial Decision Support System that has decreased the students' total travelling (Batsaris *et al.* 2021)**.** The authors also suggest a module with maps that inform students' location as improvement for a system that supports decisions.

The applied algorithms also gave us a good understanding of what is being used to solve problems in the academic context. Zhao (2023) suggests different perspectives on how to use statistical algorithms like association algorithms to mine the data in exams, quantitative analysis for evaluation, time series model to analyse teaching effects and academic performance, and cluster analysis to manage students by group. In a broad sense, grouping techniques seems to be a common practice for educational management solutions designed by C. Liu and Song (2021); Peng and Pei (2022); Skittou *et al.* (2022); Zhao (2023); Zheng and Zhou (2021).

Another interesting insight was the observation of the use of decision tree for classification by J. Chen (2021); Peng and Pei (2022); Wang (2023); Zhao (2023).

Results from applied algorithms usually appeared in visual format like charts, tables, measurement graphs. Presenting those outcomes seems to be more intuitive as a decision support tool because it presents hidden patterns, trends and anomalies come to the light. From that, the decision-making body can obtain historic information, monitor progress over time aligning it to regulatory compliance and get actionable insights from ongoing processes. Gaftandzhieva *et al.* (2023); Skittou *et al.* (2022) suggests three dimensions to evaluate the resource, first is to consider the system as a whole to check if the solution meets the requirements, what is the level of defects and its usability. The second point is regarding its functionality to ensure that the integration with the current system and at least check the system performance indicator, should support a large data query and concurrent operation.

To sum up, it is notable that a movement towards the improvement of the education field goes beyond the students' performance, the scope embraces management as a strategic point for educational providers. It is also noticeable that cutting edge technologies are being applied for different challenges in the industry. The initial steps to adopt a system to support formed decisions seems to require a reasonable effort that involves management willing to try new digital tools, equipment to store data and host applications. As researchers can manipulate technologies such as data mining, predictive analytics, and recommendation algorithms; the institution might need a skilled professional who can interpret the generated outcomes to take actions that improve the segment. It is also noticeable that many of the utilised tools are open source so financial investment would not be a strong impediment.

2.2. **Proposed Method**

This section was accomplished considering the wide options of research methods and, under the light of M.Sc. program, the preoccupation with a theoretical

approach that could not just address theoretical questions but also contribute with a relevant and pragmatic production. In this way this investigation has applied a method recognized as Design Science Research (DSR), an approach that helps the researcher to demonstrate adopted procedures during the investigation and aims to get the academic production closer to real world problems (Dresch *et al.* 2015).

To comprehend DSR possibilities we need to take a step back to point out where this method is situated. Johannesson and Perjons (2021) explains that Design Research (DR) cannot be considered an empirical science where the focus of research is made to describe, explain, and predict the natural world. The idea behind DR is the changing, improvement, and creating new worlds through artificial solutions to attend peoples demand by solving problems and finding new opportunities. Design Science (DS) is a strand of DR with the intention of creating models, methods, and systems called artefacts that could assist the development, use, and maintain IT solutions. DSR is a way to operationalize DS projects that aims to produce both an artefact and knowledge with its effects on the environment. Figure 2.3 shows a visual explanation.

The objective of DSR is a recommendation or intentional materialisation of an artefact that helps to solve a real-world problem. Dresch *et al.* (2015) convey that it should be developed looking at the fulfilment of a domain problem and, in parallel, contributes with science. Declaration that can be complemented with the few ideas to be followed during the process, first the research method must be rigorous in other to contribute with a new knowledge of general interest, than it is important to make used of the know-how already generated to ensure that the theory have a consolidated basis and give a special attention to the communication because it should be directed to practitioners and researchers (Johannesson & Perjons, 2021).

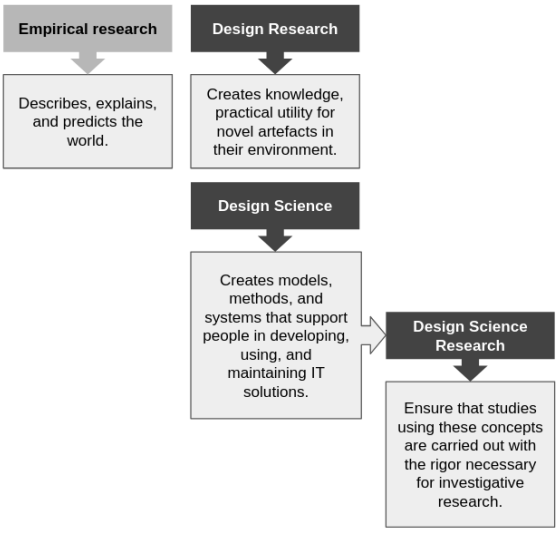

Figure 2.3 - DSR Method; Adapted from (Johannesson & Perjons, 2021, p.1)

It is also important to highlight that, the fact it's being built to sort a specific problem does not mean the solution must be optimal, the main idea is to decrease the gap found between theoretical studies and the effective practice by drawing a satisfactory solution that is pragmatic validity and belongs to a general class of problem (Dresch *et al.* 2015). What generates the following conclusion: It must be sufficiently appropriate and feasible in the context, even if that is a partial solution; Must be subject to generalisation, for instance, as a singular feature from a whole system; and must be based on a designed solution requirements and it should work well. Figure 2.4 clarifies the idea showing the mechanism that ensures the relevance of the deliverable allowing the identification of challenges and opportunities while nourishing the rigour of deliverable.

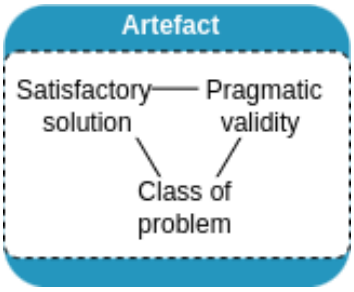

Figure 2.4 - Artefact structure. Adapted from (Dresch et al. 2015, p.59)

To achieve a reasonable level of relevance and rigour Dimov *et al*. (2023) amend with the fact that an artefact cannot be drawn isolated and arguments that create and evaluate are characteristics of design while theorising and justifying are fundamental components for scientific validation. With that we assume that relevance and rigour operate in parallel dimensions covering the project workflow from the beginning. For example, the rigour starts with theorisation and justification and goes until the ending, where the communication takes the stage. Relevance in other hand consider business process, technical and personal skills to define a domain problem.

On top of that, to deliver a robust artefact, researchers also present variations of DSR method that includes different perspectives and scenarios, the circumstance reinforce the evidence of the use of a mainstream with identification of the problem, designing of a potential solution, and communicating the results Ermolaev *et al.* (2023); Lefebvre *et al*.(2023); Müller and Reuter-Oppermann (2023). With that we conclude that the flexibility of the DSR allows the researchers to highlight what is most relevant for their research and as a result many adaptations can be found. A condition that also foments a design for this specific research to better understand the concept and also to develop something relevant guaranteeing the rigour of the produced artefact.

Summarising, to start the development of a useful resource is essential to ask the stakeholders what are their expectation regarding the tool when talking about insights, what is the resume they would like to see in order to assist decisions, if is there any critical metrics or Key Performance Indicator (KPI) that should be tracked, what is the granularity or required level of details, is there any specific trends that is essential to monitor, any colour schemes or design aesthetics that should be followed and how often should it be refreshed. Based on that a compiled survey with eight questions have been designed:

1. How do you envision yourself using a visualisation tool, would it help you to make analysis or support decisions?
2. What information or actionable insight would you like to get from visual elements (e.g. predictive analytics, forecasting, data comparisons)?

3. Which departments will use the system?
4. Do departments have distinctive demands, should the information be restricted by role?
5. Is there any sensitive information that must be considered?
6. Is there any preference for the type of chart, graphs or visualisations?
7. Should the dashboard be interactive and allow the drill down or filter data?
8. What criteria (metrics) should be used to evaluate the effectiveness of the proposed solution?

The answers resulting from the interview can certainly nourish a backlog with a lot of desired features, however, it is reasonable that a screening to establish something relevant and possible to be accomplished be performed into the proposed scope. For this matter techniques of prioritisation like MosCow[5], and critical thinking can be helpful to define not just what is a real problem but also what is feasible to be built at the given conditions. In parallel with the refining, grounding theory and support the justification, a State-of-the-Art study is imperative because it can bring inferences on what has been done and what is attainable from other experiences. A clear example comes from Schoormann *et al*. (2023) that formulates questions regarding principles of design and the concept of reuse fomenting the use of what is already known.

The design phase is where the suggested solution is developed and the artefact comes to life. At this point it is worth highlighting that the artefact is a result of a complex interaction that aims to solve a real-world problem, which has already been discussed, and is involved in a cyclical dynamic of development and evaluation. To start this phase the development can borrow the principle of reuse for the chosen technology (Johannesson & Perjons, 2021; Schoormann *et al.* 2023).

It is well known that many software libraries maintained by companies, governments and communities are available out there and this is essential to achieve pragmatic validity from the monetary costs perspective. It is equally important to emphasise that the evaluation can bring new questions to the board and

[5] https://en.wikipedia.org/wiki/MoSCoW_method

consequently leads to modifications at the design to match with the initial requirements. Assuming the relevance of artefact evaluation it will be better explored in the next topic.

Once the artefact is pragmatically valid, it is time to communicate and, although it is not a preponderant task, the communication performs a fundamental role into DSR because of its main characteristic that is getting the academic projects and transform into a real-world solution. As a result, the outcome informs the results and also contributes with different stakeholders, be they scientist, practitioners or singular person that have general interest. Note that this research carefully considers the communication through this dissertation and final presentation as the outcome of the process but also an outcome required from the chosen methodology.

2.3. Artefact Evaluation

Evaluating the outcome and understanding the practical impact it has is certainly not an easy task even though that DSR requires that develop of a new artefact to solve a specific problem also requires some way adequately evaluated the utility of the solution (Dresch *et al.* 2015; Hevner & vom Brocke, 2023). To be recognized the evaluation must be rigorously conducted and different techniques, tools and formats can be combined to get it done, starting by checking if the requirements have been fulfilled, what improvements were made in the context operation, usability, and overall performance, etc.

Alturki *et al.* (2011) explains the objective is not to explain ‘why’ or even ‘how’ the delivered solution works but ‘how well’ the artefact performs into the proposed context. The authors defend that the verification should be done in two stages: Undergo the artefact to simulation or experiments, also known as artificial evaluation; and in case of failure and, perform a natural evaluation involving people, processes, and other variables. Presuming the artefact is a digital solution it makes complete sense because it covers technical aspects such as measuring the time to complete specific tasks, get the frequency and types of errors, and the percentage of successfully completed tasks under controlled conditions. And a holistic approach by collecting feedback from users regarding their overall satisfaction, assess how well

the solution aligns with users' workflow and tasks in their day-to-day work provided by the naturalistic evaluation.

Nevertheless, it is important to take a step back and remember that the delivered resource does not have to be an optimal solution, the idea is to develop a desirable artefact from the institution perspective, technically feasible to be built and economically viable. The cyclical characteristic of the DSR model proposed will allow future increments if there is a need. With that, considering that the presented artefact is an EDSS for an educational institution, it is reasonable to first consider dimensions like data accuracy, interactivity and visualisation effectiveness to be evaluated and then define the method.

Those evaluation topics could help to validate the accuracy of presented data on dashboard, assess the possibility of users to drill down it into details, make sure the filters are working, that the interaction with visual elements are potentialise the gain of deeper insights and check if the chosen charts, visualisations and other elements are appropriate for the type of data being displayed. From that perspective to define the success of the artefact the feature must fulfil the required dimensions. To illustrate the validation, consider the following question: Q1 - How many students are starting classes at different levels? Are the teachers enough or should the institution hire? Table 2.1 presents our proposed validation method.

Dimension	Evaluation Method	Description
Data Accuracy	Data Validation	Collect a sample from dataset used in the EDSS and manually verify the accuracy against the available data presented at the current system
	Cross-Verification	Perform a cross-verification between the data presented at the dashboard and an independent instance of the database to identify discrepancies
Interactivity	User performance in different scenarios	Simulates a specific task that requires users to interact with different features of the dashboard
	Task completion time	Check the time it takes to perform the task. E.g.Finding the number of students enrolled in a particular course level

Data Visualisation Effectiveness	Heuristic Evaluation	Collect user feedback to ensure that charts provide clear insights into enrollment trends at different levels
	A/B Testing	Conduct A/B testing with alternative visualisations to determine which representations are most effective for conveying the information needed to answer the question

Table 2.1 - Evaluating artefact method.

In fact a digital solution can be evaluated in several dimensions such as accuracy, performance and general reliability to take a screenshot of the current moment; therefore, it can always be complemented with different tools. Gartner's maturity model for enterprise information management[6] is a reasonable example to illustrate the current situation regarding decisions based on data, with six levels of maturity the mapping starts from companies that are not aware of the movement without ownership of its own data; Those, classified as ‘aware’, who understand it but the silos do not have full trust on its data; Reactive companies that shares information without proper data management department; Proactive companies that have adopted information systems and move towards data governance; Managed companies that have standards and policies well as process matured.

The report completes the levels with what is considered an effective company. At this ‘optimised’ stage the companies have a data strategy already established, the quality of data is constantly monitored in cyclical programs relying on a rigorous evaluation process of data. Considering Gartner’s report, this study can safely categorise the current institution as a ‘reactive’ model because of the fact it fully relies on the IT department for data storage and governance at the same time it lacks specific roles related with data management.

Placing the institution at this milestone, the successful adoption of EDSS as an supportive apparatus that guides institutional decisions can be evaluated by checking if the institution has taken a step forward on Gartner's model. For example, providing information based on its own data asset the system generates clear evidence for the need of data governance but the ‘proactive’ approach nevertheless requires more than an adoption data integration tool, it demands data management

[6] https://www.gartner.com/en/documents/3236418

policies. An empirical inference of this movement can be measured by checking a professional strictly assigned for role in data.

With settled methodology, outcomes delineated and evaluation defined, the investigation moves towards discussing the development customised method that paves the development of a digital solution called Educational Decision Support System. The next section presents the methodology adapted for this research, explores features that make up the system, applied technologies to release the artefact, how machine learning algorithms applied and data analysed to create a comprehensive tool to support informed decisions.

3. Educational Decision Support System (EDSS)

Since the phrase ‘data is the new oil’ has been widely broadcast and computational power has become accessible, small businesses around the world also started to look for ways to explore data and extract useful information from it. In the educational field it isn't different, general management, administrative department and coordinators are daily looking at its own generated data to enhance its productivity and performance to improve the way it delivers education as a product. From a student perspective, we can easily think about academic performance as a main result from its own efforts during the program but behind the scenes many other variables such as presences in class, absences, studying time, travelled distance to arrive at school, among other demographic data could impact its achievement.

To an educational management body, creating a broader picture is a key element to lead the departments on offering better services with customised responses instead of just answering demands. In this context we can safely consider data as raw-material, even more valuable than oil for a few reasons, starting with the main characteristic of finitude, the more humankind uses crude oil the less we have it, this logic is reversed when considering data. Petroleum is difficult and expensive to extract while data is generated every second through several connected devices. With that even small institutions can better understand its behaviour by extracting useful historical data, convert it into information to support informed decisions.

In the education field many decisions are predictable such as pursuing a new technology, adopting a course material or even hiring a new professor while some are unexpected. Each situation conveys many possible options, sometimes considerable like process change but in fact most of the decisions are smaller and taken on a daily basis over the current process, with this idea a set of right decisions is essential to design and succeed in long-term plans. Assuming a set of formed decisions can enable the education providers to thrive for long years or lead it into

failure when it is taken without support, for that reason this research has grounded the development of EDSS creating an adapted method presented at Figure 3.1.

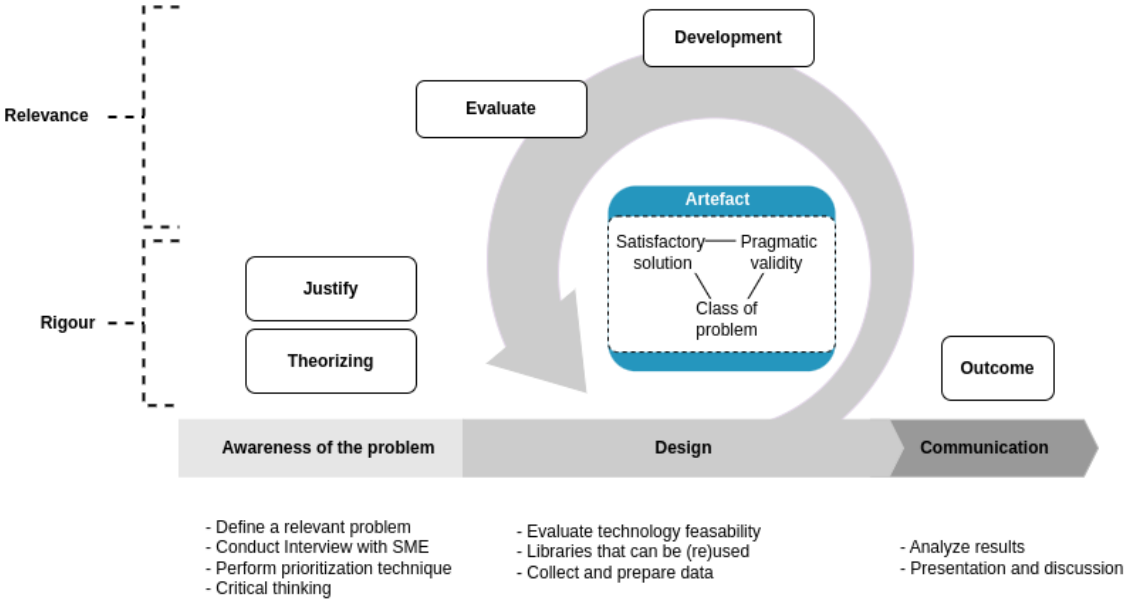

Figure 3.1 - Adapted DSR methodology.

The figure presents a framework that considers three areas: Awareness of the problem, Design and Communication; those areas gather phases such as justification, theory, design, evaluation and outcomes and work as guidelines for this research, they are intersected with Relevance and Rigour which establishes boundaries to make sure we are solving a relevant problem and that it can be clearly communicated through generated outcomes. In addition to supporting different phases, each area accommodates a set of tasks, techniques and tools set with the objective of facilitating the process.

At the first phase we gain consciousness regarding the problem to be tackled, justify the relevance and need for proposed solution. An example of an applied tool at this phase was interviewing with Subject Matter Experts (SME) which can provide broad and sensitive information, different from questionnaires that would generate straightforward information (Johannesson & Perjons, 2021). Next we tighten it theoretically by scrutinising different research to get the State-of-the-Art as a starting point to designing EDSS, all this process has been presented and discussed early.

The design phase, better discussed in the next section, accommodates the development, analysis and evaluation of EDSS. This phase helped us to define the features that compound the proposed digital artefact, the essential technologies to make the system available, aspects of data governance, preprocessing and the way the data storytelling was created to make sense for decision takers. At this stage studies were performed to understand why some visualisation types were more appropriated for this supportive context. Then different machine learning algorithms were applied to explore relevant factors for students to choose the institution as a learning partner, a proposed tool set with the intention of monitoring general feeling, investigating the renewal possibilities and probabilities.

Moving to the final phase, it represents the outcomes that drive the communication and disclosure of investigation. The final document and presentation are examples of instruments where the results were analysed, presented and discussed. Here is where final discussion and new ideas for future investigation are stated, that will consequently be discussed at further.

3.1. EDSS features

The purpose of building an Educational Decision Support System (EDSS) goes beyond regular Data-Driven Decision Making (DDDM), which plays a more deterministic role while minimizing human factors. A great example could be a system where students' previous assessments are compiled to build a new learning path on their course.

This situation minimizes human actions and decisions once the algorithm has predefined rules and is consequently responsible for the process as a whole. Here we converge with Data-Informed Decision Making (DIDM) first because this model assumes data as valuable input and then because it depends on human expertise to make decisions (Webber & Zheng, 2020).

To define what features would be implemented on EDSS, an initial brainstorming session has brought many ideas that should increase transparency over the daily activities and raise the knowledge surrounding the institution's operation. The ideas were scrutinised considering viability of available technology, research knowledge and impact for decisions regarding requested features. Even though the management has final word, a reasonable way to list potential elements part of EDSS was applying a prioritisation technique where each feature was categorised into ‘Must Have’, ‘Should have’, ‘Could have’ and ‘Won’t have’ categories. The result of this dynamic is presented at the following table.

Must Have:	Should Have:	Could Have:	Won't Have:
Total number of enrolled students by class, campus and period.	Option to select classes considering options by course ID, start date, or duration to facilitate detailed analysis.	Inform how far are the students from school, it would be helpful to analyse cases where a student has been late for classes.	Financial information related to students (e.g., total revenue, outstanding payments).
An overview of student attendance rates together, checking one by one is too exhaustive.	Detailed analysis of attendance patterns for each class or course.	A dashboard component illustrating trends in student enrolment and course completion rates over time.	Data collection, communication or integration with students social media.
The distribution of students across different classes or courses.	Showing the percentage of students who successfully complete each course.	This Visualisation could include line charts displaying enrolment trends by semester or academic year, as well as completion rates for different courses.	Demographics data involving personal information such as health records or medical history.
See all students starting next couple weeks	Displaying key performance metrics such as average grades or completion rates.	Additionally, it could incorporate filtering options by program ID, course ID, or admission status for deeper insights.	
Visualisation showcasing the distribution of students based on enrolment status (e.g., active, pending, graduated).	Trend analysis of course enrollment over time.	Advanced analytics such as predictive modelling for individual student academic performance.	
Visualisation presents key demographic information such as gender, nationality, and age.	Compile feedback forms to gauge overall satisfaction		
Real-time availability of courses or classes for enrolment.			

Table 3.1 - MoSCoW table for applied prioritisation technique.

Given desired features, to evaluate the viability of building a customised artefact under a relatively short time frame it was necessary to take small steps towards available technologies. In spite of the amount of open source resources out there, each task must be taken with parsimony considering conditions such as minimum requirements regarding power of processing, complexity to integrate and make it available.

3.2. EDSS architecture

Aside from the objectives, building an Educational Decisions Support System requires many decisions and concerns regarding the quality of collected data, the way it can be resumed to become useful information, visual components that truly represent the studied phenomenon and necessary technology to release it. Gaftandzhieva *et al.* (2023) point out that implementing analytical tools is a long process and it usually runs with problems because of challenges related to technology, privacy, ethical and responsible use of data.

With those concerns on the table, this research has proceeded with what is inherent to every data project, it is worth to say. The EDSS was initially designed to use a minimal logical architecture to make it available for final users, Figure 3.2 represents this proposal. Although it demonstrates the applied mechanism to ensure that data is accessible, relevant, and up-to-date it will be better discussed at the topic 3.3 Data Governance and Preprocessing. Therefore, at this moment it is important to highlight that all data used for this project comes from internal databases, CRM tools and CSV files provided by the institution.

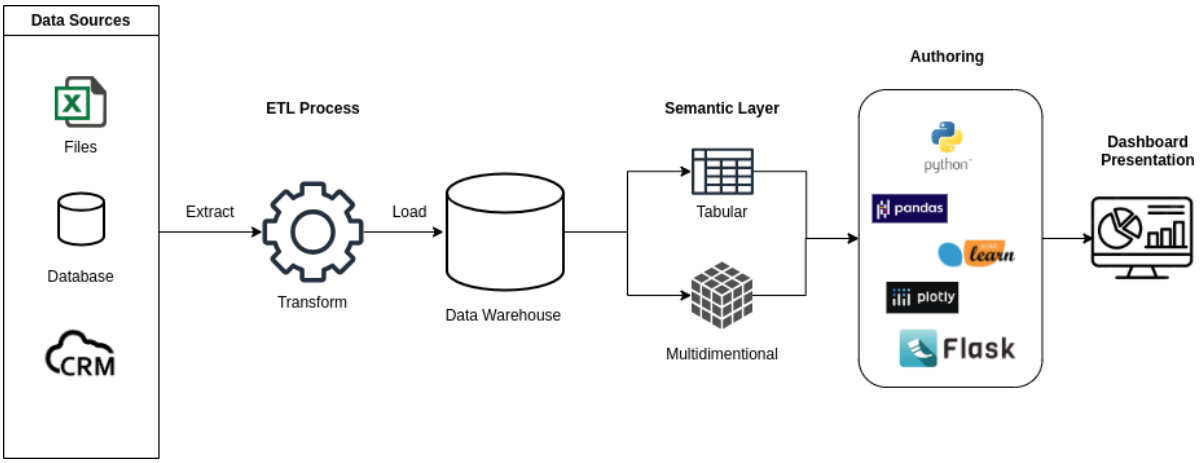

Figure 3.2 - EDSS Architecture.

After gathering data from different sources it is essential to cleaning data, handling missing values and normalising data to store into Data Warehouse (DW). Known as ETL, the process of extract, transform and load could be done using tools like Apache NiFi[7], Talend[8] or Microsoft SSIS[9]. However, the research has chosen a less sophisticated path. That decision was made under circumstances like time span to develop and produce the final dissertation, expertise on mentioned resources and complexity to create a minimum viable product. For that reason, the Structured Query Language (SQL) was the only technology used to access, manipulate in some cases and retrieve data from objects in the database while Python transformed and processed it. Combining both technologies was enough to gather, transform, filter, and merge datasets into a Data Warehouse (DW).

Building a DW requires a semantic definition of how data should be structured, this process that includes tabular and multidimensional data. The use of multidimensional data is more complex than tabular structures, but one model does not eliminate the need of another. While tabular data, structured in rows and columns, is suitable to apply a Correspondence Analysis, to be considered a business intelligence tool it must make use of OLAP cubes.

[7] https://nifi.apache.org/
[8] https://www.talend.com/
[9] https://learn.microsoft.com/en-us/sql/integration-services/sql-server-integration-services

The authoring layer addresses the development of the backend, the stage where communication with DW occurs, transformations are performed and statistical algorithms processed. For presentation layer namely front-end, the research has use Flask[10], a web application framework to structure and the artefact available over the internet, libraries like Pandas[11] for data exploration, analysis and understanding, Plotly[12] to build visual and interactive resources and Scikit-learn[13] library well known by its machine learning algorithms.

To test the features and also paving the way for the future use in a real scenario, the EDSS will be hosted in a local server provided by the institution and released for final users through a web browser under the internal networking, with that no monetary values were involved in order to run trials or make EDSS available. Combined this set of technologies would present historical behaviour, insightful visualisations, predictions for the decision makers.

3.3. Data Governance and Preprocessing

Despite not having a Data Governance department, the institution deals with its own data as a serious matter under the IT department. As it operates beneath the General Data Protection Regulation (GDPR)[14], any action involving students' personal data must be submitted to an internal committee that evaluates the impact of the data usage. Examples of that intervention could be seen at Table 3.1, specifically at the ‘Won’t have’ category, where data generated related to financial debts or health information and consequently must be avoided.

In summary, the scope of the proposed EDSS was limited to use academic data which were already available at the main database and external data sources where the institution does not manipulate sensitive information. In practice, the research could not generate or collect data through instruments like feedback form,

[10] https://flask.palletsprojects.com/en/3.0.x/
[11] https://pandas.pydata.org/
[12] https://plotly.com/python/
[13] https://scikit-learn.org/stable/index.html
[14] https://gdpr-info.eu/

special cases where visualisation requires student identification ‘Student ID’ should be used. Even though this attribute would not reveal students characteristics the IT department had to approve the use for any presented study. The system also had to care about users, only authorised staff with username and password could access it to guarantee data privacy.

Once desired features were selected and risk compliance evaluated by the IT department, the next step was to design multidimensional schemes that allows EDSS to perform efficient querying. Figure 3.3 presents part of this process. To have a clear idea of what multidimensional tables would be, Diagram Entity Relationship (ERD) was used as a guideline to generate OLAP cubes. The diagrams has been designed using a visual database design tool called MySQL Workbench[15], by presenting the dynamic between tables and attributes it becomes easier to understand design necessary queries at the same time we about data compliance.

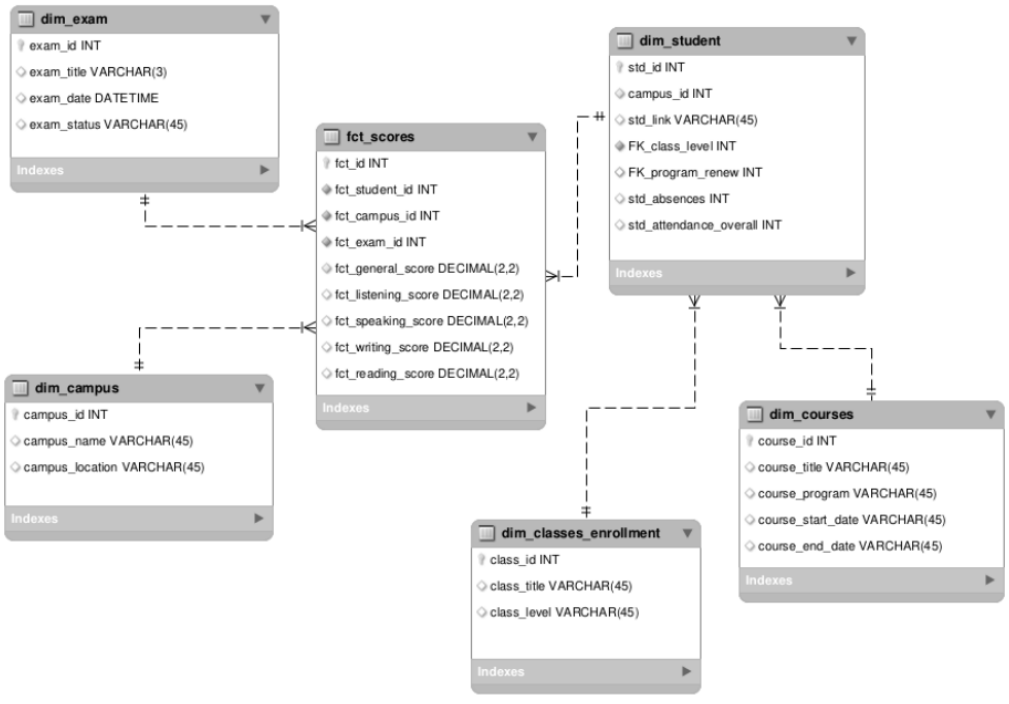

Figure 3.3 - Star scheme for business analysis

[15] https://www.mysql.com/products/workbench/

The process continues by designing queries that represent each ERD, those queries were fundamental elements to create a source for descriptive statistics, to perform algorithms and produce visual elements. Figure 3.4 presents a SQL script that extracts the most recent exams taken, mind that to keep the student privacy the students were identified by its ID (a). Following the premise of allowing specific users to access student profiles the query generates a link (b) for each student which gives more flexibility to EDSS.

```
WITH RankedExams AS (
   SELECT ts.student_code AS 'id',                                                                    (a)
          CONCAT('https://…student/details?id=', ts.student_key, '#exams') AS 'Student',              (b)
          DATE_FORMAT(tse.result_at, '%Y-%m-%d') AS 'Date',
          te.title AS 'Exam',
          Tse.status AS 'Status',
          tse.`result` AS 'Result',
          Tar.total_percentage  AS 'Overall',
          ROW_NUMBER() OVER (PARTITION BY ts.student_code ORDER BY tse.result_at DESC) AS row_num
   FROM tbl_student_exams tse
       LEFT JOIN tbl_students ts ON ts.student_id = tse.student_id
       LEFT JOIN tbl_exams te ON te.id = tse.exam_id
       LEFT JOIN tbl_attendance_reports tar ON tar.student_id = tse.student_id
   WHERE tse.student_id IN (
          SELECT ta.student_id
          FROM tbl_admissions ta
          WHERE (ta.course_duration = 25 AND ta.duration_for LIKE 'week'
                 OR ta.course_duration = 8 AND ta.duration_for LIKE 'month')
          ORDER BY ta.course_start_date ASC
       )
       AND tse.result_at BETWEEN '2024-01-01' AND '2024-12-31'
)
     SELECT * FROM RankedExams
          WHERE row_num = 1
     ORDER BY id ASC;
```

Figure 3.4 - Query performed to obtain a fact table.

For institutions that are not familiar with its own data, the solution could pass through data classification for easier management, security, and compliance, which would help to categorise sensitive information; And a ontology mapping to create a comprehensive understanding of how different data elements relate to each other into its context. At the wake of this process, it is also imperative to know the available data type and measure its quality from different perspectives such as accuracy, completeness, consistency, compliance, duplication and timing in access and update.

A good example of data consistency can be found in systems that operate in different continents, a situation faced in this investigation. It is quite common to find documents with issues regarding date consistency like ‘15/04/1986’ and ‘1986-15-04’ which influences on date calculations and obviously must be treated accordingly. Having a clear vision about what should be collected combined with a critical analysis regarding data quality would avoid inconsistent cases and consequently decrease difficulties on its manipulation and mistakes in case the attribute needs to be calculated.

Data preprocessing is one of most important tasks to develop a custom-made EDSS, it is the basis for next phases and mistakes on that stage impact the analysis and consequently decisions based on provided information. The care on data preprocessing has started by retrieving only necessary data from desired tables and then performing necessary data transformations which included cleaning, filtering and aggregating data when it was necessary. To resume, when dealing with a complex and inconsistent scenario several non-exclusive must be taken to ensure data quality. Figure 3.5 shows a representation of a problematic scenario faced during the EDSS development.

	date	assessed_by	student_id	assessment
1	2024-03-23	4	510	80% progression
2	2024-03-23	4	510	80%
3	2024-03-23	4	488	80% progression
4	2024-03-23	18	573	82 Complete IELTS Reading and Listening
5	2024-03-23	18	573	67 IELTS Grammar, Vocabulary, Reading
6	2024-03-23	4	2	Progression wk 43 -80%
7	2024-03-23	4	2	p 44 -75%
8	2024-03-23	10	189	intermediate test 69%
9	2024-03-23	8	274	55%
10	2024-03-23	8	274	80%
11	2024-03-23	8	274	79%
12	2024-03-23	8	274	75%
13	2024-03-23	8	353	50%
14	2024-03-23	8	353	41 IELTS Reading & Listening
15	2024-03-23	8	353	55 Grammar Review
16	2024-03-23	8	353	70 Unit progression
17	2024-03-23	8	353	37 Unit Progression
18	2024-03-23	8	353	67 IELTS Reading, Unit 3
19	2024-03-23	8	353	60 Unit 1, 2
20	2024-03-23	8	353	74 IELTS Reading and Listening
21	2024-03-23	8	250	90 Unit Progression
22	2024-03-23	8	250	80 Unit Progression
23	2024-03-23	8	250	87 Unit Progression
24	2024-03-23	8	250	50 Unit Progression
25	2024-03-23	8	250	95 Unit Progression
26	2024-03-23	8	250	74%
27	2024-03-23	8	553	47%

Figure 3.5 - Sample of fact table.

It is well known that data is often obtained in a format that needs processing before answering questions and that is exactly what the previous figure presents. The column named ‘assessment’ that stores International English Language Testing System (IELTS) results for a small set of students. This exam usually provides an overall band score complemented by individual evaluations regarding listening, reading, writing and speaking. The founded scenario represents a regular situation in datasets where standards are poorly communicated, and systems somehow were not designed to support new models resulting in low quality historical data regarding students' achievement.

In defence of the current system, the institution started to handle different kinds of exams after a few years and the database was not modelled to receive this specific assessment results. Yet it is equally important to understand that this circumstance reveals fragility in the change management process, which may cause a direct impact on the quality of analysis based on that piece of information. ERD diagrams can be helpful to mapping databases but the results sometimes can bring unpredictable values. The trade-off is that instead of getting a resultset that could immediately foster a visual element, the column ‘assessment’ presents a non-standard information demanding data treatment.

To overcome this situation we have analysed the distribution of concerned field, identifying patterns and frequencies of each format (numeric, percentage, text, etc.), determined the percentage of null values to decide how to handle it, pondered whether to remove, impute, or flag null values based on each specific analysis scenario. Another measure was to convert all numeric representations to a single scale format (e.g. 0-10), applied regular expressions to extract numeric values from text entries and performed a Text-to-Numeric Conversion process.

In conclusion, ensuring data quality and consistency is a basic premise not just to produce a valuable information source for future analysis, it is also essential to create storylines that demonstrate the full picture with past present and future. In the following section we explore the idea of storytelling.

3.4. Storytelling and Dashboard

In fact, there is no exclusive way to build an analytical panel that summarises data through graphical components, nonetheless it is still compulsory that provided information generates actionable insights. Distinct dimensions at the decision process create a need to deeply understand the general user's proposals, its target and goals; and the variability of data aligned with management proposals can provide different ways to support the board decisions, even if that is in the same industry.

It is indeed not a trivial task and, considering the available options to present a single informational dimension, the consonance between elements can create more doubt than insights. To avoid misunderstandings and produce a valuable comprehensive storyline Bach *et al.* (2022) argues that the amount of streaming data allows designers to process, abstract and simplify visual representation combining different elements to produce a clear and interactive interface. Although we have a clear understanding that providing a ‘what-if’ scenario could bring more value for the solution, the applied technology, type of available data and researchers' knowledge could create limitations on its development.

The found solution was to design elements individually, considering data and graphical components that care initially about reliability of presented information, simplicity on users access and interaction and, after that, gather created elements. To International, D. (2017) visual patterns allow users to discern periodic changes in data, gain insights from its trends and anticipate the effects of related data before they occur. Together those elements could explain a given situation, in this case, a student profile and possible scenario where students could renew their course.

Based on our systematic review, Bach *et al.*(2022) have mapped common dashboard solutions and proposed a wide comprehension by splitting it into content patterns which describe information through data and meta information, visual representation and composition patterns that describe how components are distributed and organised. Between recommendations such as page layout, screen space, interactions and colour scheme at the composition patterns, the authors

describe a possibility of using layout with multiple pages since it follows a regular structure keeping a visual representation, which is useful to build storytelling and provide personalised information.

With that no threshold regarding elements or number of pages were set at the beginning, the idea was to design elements and validate those elements individually to set only well evaluated elements on the EDSS. Each page was designed to answer a broader question through one or more visual representations, the full story should be taken from the elements combined. A good explanation of this situation can be seen in the next figure.

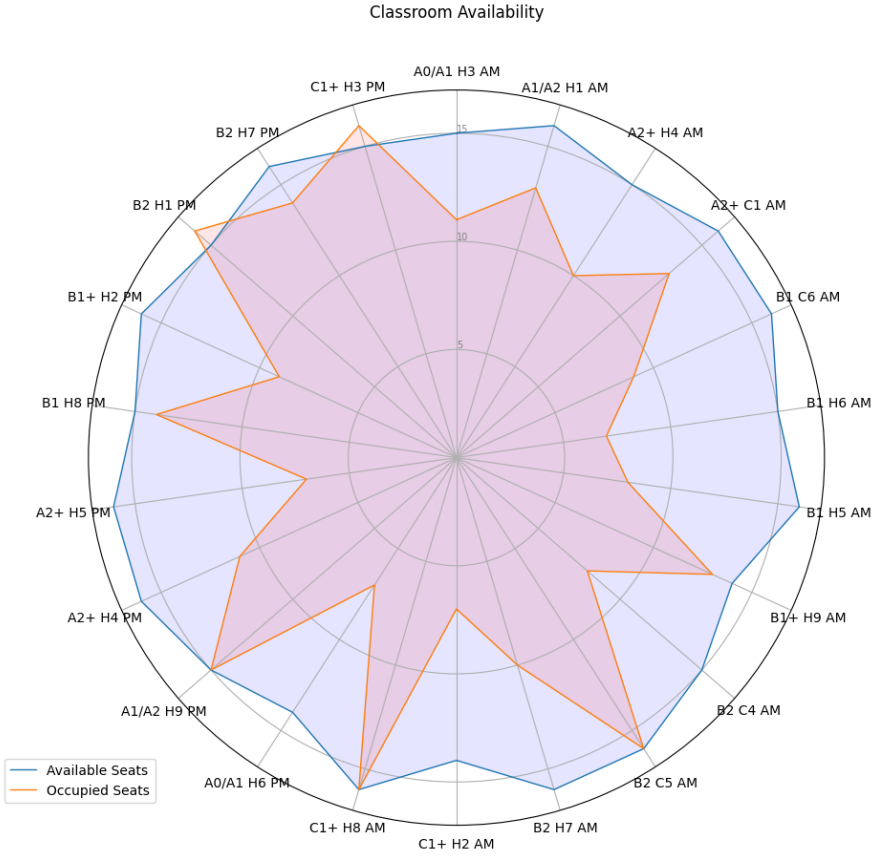

Figure 3.6 - Radar Chart with available seats by classroom.

A Radar chart was designed to present the amount of available seats and students by classroom. In spite of possible interaction, the information comes with some difficulty at first glance, at this stage the use of tables has been discussed and

pondered. Circumstance that has shown that the best strategy of providing informed information is not to create fancy charts but something useful and clear for the end user and tables can provide values that confer significance for what is presented instead. Figure 3.7 presents an alternative way to present previous information, it evidences that dynamic table empowers the users to drill down into specific data points without losing information.

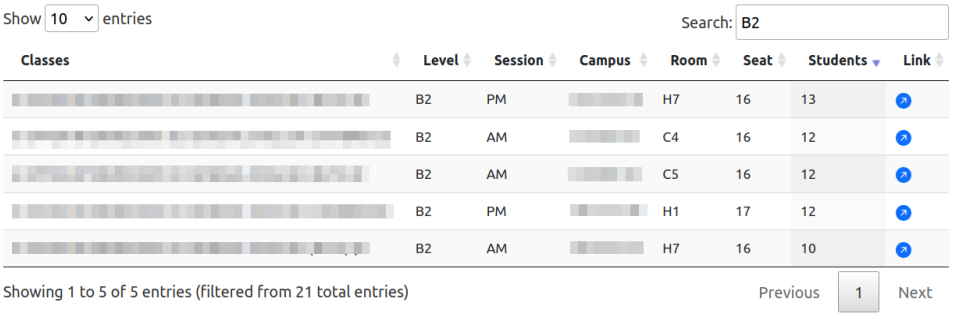

Figure 3.7 - Table presenting available seats by classroom.

Adopting an interactive table leads EDSS to not just gain information with more details available, it also amplifies the utility of the dashboard allowing actions beyond the system through specific links that send the straight to the classes edition. For instance, by offering a special link from EDSS to CMS, the users can access straight away specific sections allowing them to make punctual changes like moving students between classes, adding a new teacher, changing classroom number, etc. As the link forwards the user to a specific interface, the CMS only allows editions after a login process, in other words, it creates a layer of security. Those possibilities undoubtedly strengthen the dashboard and contribute to storytelling.

3.5. **Data analysis and visualisation types**

Interviewing stakeholders is essential to obtain the right direction since the beginning of development, at this phase, brainstorm sessions provide valuable information to create the relevant deliverables. After all, the EDSS is only worthwhile if it highlights trends based on available data and generates actionable insights to

support future decisions. The artefact, however, is not a panacea and it will not be supportive enough if the data doesn’t have quality and consistency.

To build a solid decision support tool it is imperative to rely on data, neither from an internal or external source and it never hurts to remember that this investigation has focused primarily on internal available data to produce an analytical system that supports management with routinary decisions. It is also worth highlighting that since its conception the system has been thought to be open for new data integration with the objective to enhance the EDSS.

From the backlog of suggested and developed features, a screening of items has been done to perform data analysis, a task that requires enough technical effort to develop a solution in a reasonable time and to display comprehensive outcomes. Descriptive statistics techniques were firstly applied to describe and summarise a set of data. Although measures like minimal and maximum number of students per class were already available with the ongoing system, information such as amount of students per nationality, overall attendance per classroom, even number of students beginning classes in a specific level were not available.

At the current institution system, students that were completing classes should be indicated and processed manually by the academic team before delivering it to the administration in order to issue certificates and other important course letters. Task that sounds simple but here nuances of the operation comes to the light and gives some complexity for the analysis, see Figure 1.1. From the business model a student completes a course period while the class level has no ending date. To solve this problem Figure 3.8 shows a proposed interactive barchart with the students completing courses by date, it allows staff members to identify trends with the amount of students completing classes at the same period.

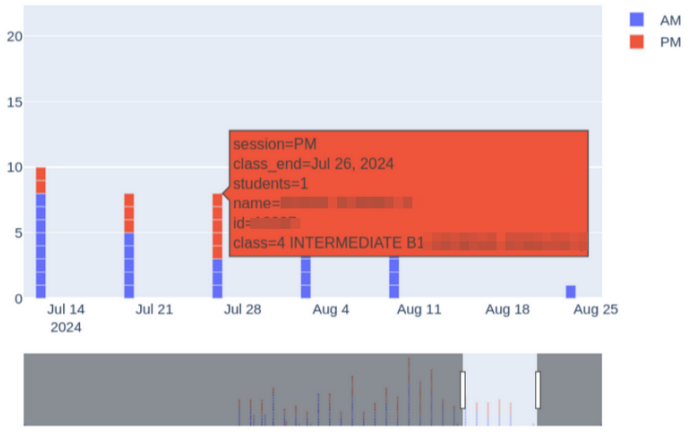

Figure 3.8 - Students completing course v.1

Using an interactive component brings several advantages because it creates the possibility to see beyond the amount of students completing classes in a determined period through the bars. The presented illustration shows an artefact that supplies six different dimensions: counting, trends, data range, outliers, period and individual details at once. Beginning with the amount by date which is trivial but with a closer look shows that each square loads student details. With ‘details’ we mean individual necessary information that was improved with a link which drives the user to the student profile on CMS. In practical terms, it saves a considerable time when compared to checking a physical or even digital static report because it requires browsing to CMS, manually type student ID and searching the student register.

The same visualisation still brings more information, following business rule students must complete classes period once a week, though provided visualisation students out of this standard can be easily identified. Figure 3.9 which is a second cut of the same visual element, it presents the option of (a) filtering by period different term, (b) segment students in a data range highlighting missing actions or wrong imputed data and, this example demonstrates a student who is supposed to have classes period gone and consequently out of classroom and (c) students completing classes on weekend, which is against the institution policies.

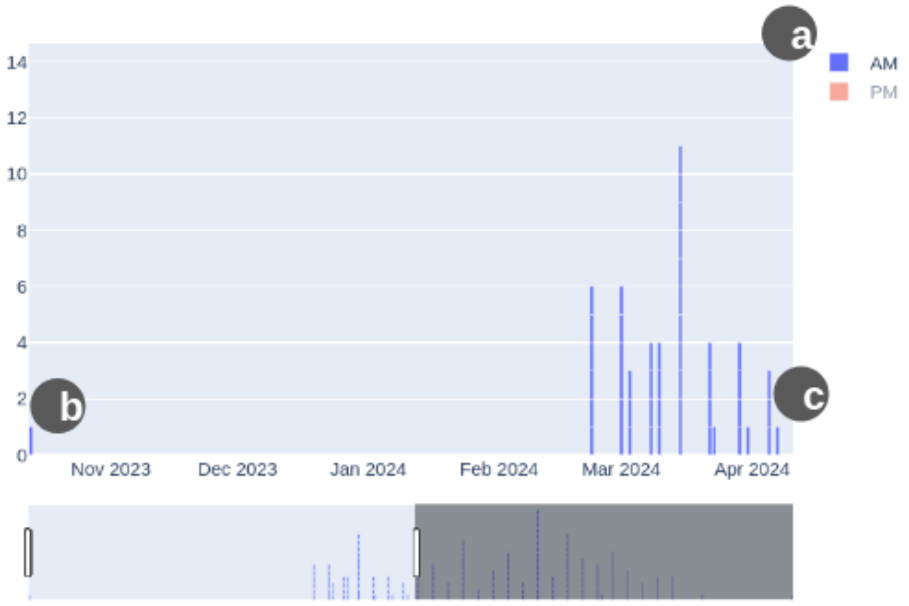

Figure 3.9 - Students completing course v.2

Drawn initially to observe students that were completing classes period, the visual element presents itself as a powerful resource not just for academic purposes, it also enhances administrative and marketing views regarding this topic. The visual component reveals tendencies that facilitate routinary actions such as issuing letters and certificates, furthermore contribute with subsidies to promote marketing campaigns related with students which are completing classes in a specific period to be approached with marketing campaigns. In fact, descriptive statistics is powerful and it can be boosted with a proper visual plot, data and clear goals.

Another widely requested feature deals with attendance, this is one of the most sensitive information regarding the nature of the business, this indicator is the basis for many decisions that can directly impact the students and consequently the institution. The compliance with local regulation requires high attention, constant monitorization and communication with the government about this matter, students with overall under 85% cannot renew their course resulting in a loss for the institution. This gauge is also behind decisions like warning letters to inform low attendance and, in extreme cases, suspension which impacts students' VISA.

This information is already available in two ways, the staff interested in that information can browse on CMS searching for the specific person, which is obviously counterproductive if there is a need to search for more than one student. The second way is selecting a specific classroom, with that the ongoing system generates a list of placed students. It, however, is a static list and even with the amount of desired students grouped it is not useful for attendance comparison. The profile of each student must be accessed by typing results in a manual process, it creates unnecessary workload and does not work valuable for fast decisions.

To present this information in a way that management, academic team or any other stakeholders could make precise decisions, EDSS has come with different plots and main objective of displaying the current attendance behaviour graphically and then understand the impact of each visualisation for end users. Early on we wrapped up that more dimensions could provide more consolidated information and consequently facilitate analysis which promoted the gain of valuable insights, as we do not have this possibility the research have followed a different principle where the same information can be presented in different format resulting in (a) Histogram and (b) Box-Plot, examples are presented in the next figure, note that both provides the option of segmentation by term.

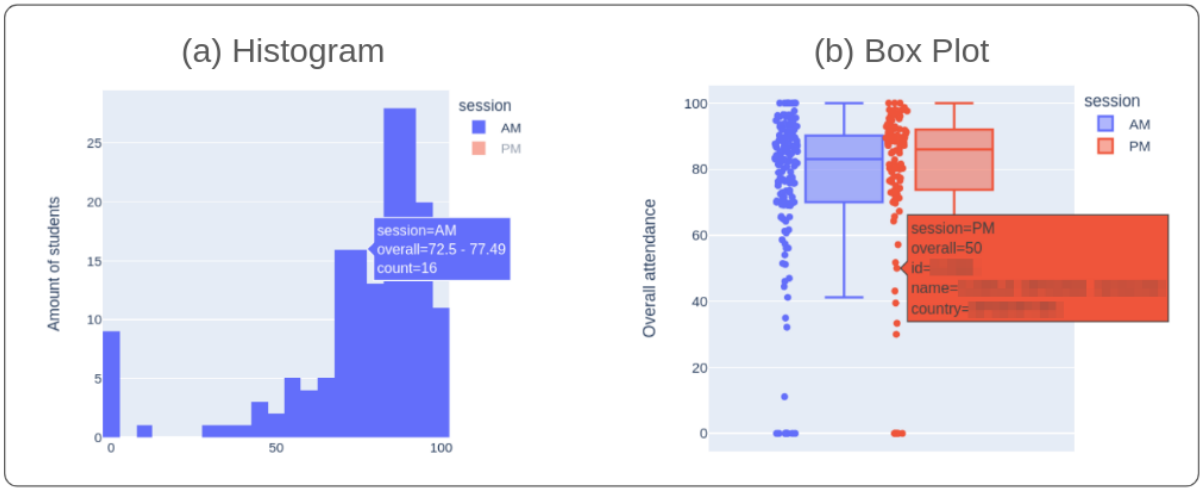

Figure 3.10 - Attendance distribution.

To understand the impact of each proposed visualisations to make efficient decisions the research has conducted and an hypothesis test. Ensuring that the chosen visual component is capable of improving the team performance with safer decision-making requires some rigorous method to measure it. The hypothesis test A/B is a statistical procedure that checks the veracity of a statement which represents a determined population. The applied method and its results have been properly explored at the Results and Discussion section.

To be in compliance with the market regulations, the Irish language provider must monitor students' attendance regularly and take several actions based on that indicator; it is imperative. Notwithstanding, before any action the institution must deeply understand each individual case and its conditions. Following ILEP guidelines the students can receive approved leave regarding adverse circumstances and a common attribute appraised by the academic body is the student location.

The students' requests are usually regarding medical conditions, bank appointments between other bureaucracies, to approve it the responsible staff has to analyse the request considering the commuting time for students to be in class. Brainstorm meeting reveals that this task is not that simple, to make a final decision the student address must be up to date with all necessary information and, although the official spoken language is English, many addresses are in Irish, which creates mistyping when the information is provided, with that ZIP code becoming the most reliable information. It however requires too much effort to research what was the distance and time.

In fact, being based solely on empirical knowledge to define the distance is not adequate and searching on Google Maps[16] demonstrates that it might not be the most efficient process. It was comprehended that to make this decision a resource that offers more precise and fast information was indispensable. During the effort to better understand the situation it becomes evident that such a tool would not just answer specific demand but also increase the awareness regarding students that live near campus, as the institution has more than one campus this is useful

[16] https://www.google.com/maps

information to avoid long commuting. Ireland is a country that faces many adverse moments regarding climate, knowing where students are located can also improve communication in emergency situations.

Still about emergency situations, one of the arguments used to request a visual component equipped with a global position comes from past emergencies regarding health. By plotting students' location and campuses, the resource could complement the emergence plan with nearest hospitals highlighted on the map. Equally, the information would help on marketing proposes by considering special outside classes at parks, museums, etc. To resume, it was evident that the institution can certainly gain many insights by simply mapping students' locale.

Given the scenario, the situation was tackled using the Python library GeoPy[17] and OpenStreetMap[18] available under the Open Database License. The resource was proposed initially to show the students location with some details accessed during the interaction. It allowed the school management to literally see the distribution within a specific radius, a scheme definitely easy to read and in interpreting to take formed decisions in a reasonable time. The next figure gives an overview of the result.

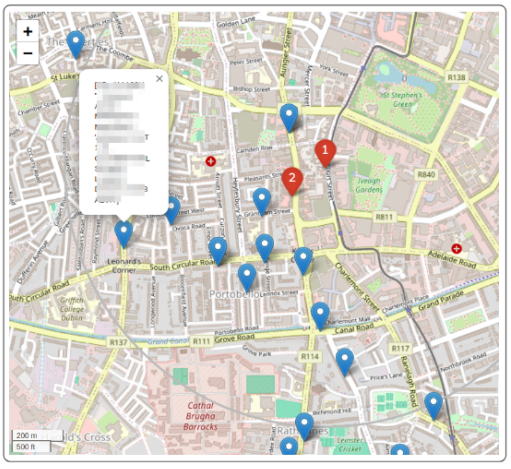

Figure 3.11 - Campuses and hypothetical location of students.

[17] https://geopy.readthedocs.io
[18] https://www.openstreetmap.org/

Merging this interactive solution into EDSS can certainly provide many useful insights and aid the users to get valuable information but it is also important to let clear that the presented information relies on the quality of raw material. Processing it using a full address for each individual register requires high computational power and it is not viable to be performed on the current server, even with libraries that make it automatically, gaps on data entry have drastically impacted the performance. The conclusion here is that the element can definitively contribute to make informed decisions; however the institution policies regarding data input must be refined, an alternative way is to split address from ZIP code. By getting only this last piece of information the process moves a step towards conversion to global coordinates avoiding unnecessary preprocessing.

To summarise, descriptive analysis can support decisions when there are good information sources and it can furnish EDSS with valuable information, it however is not a threshold, as a decision support tool it can always be enriched with implementation of exploratory analysis involving different algorithms that increase its relevance for future decisions. Again, it is important to reinforce that advanced algorithms also demand appropriate data available, consistency and reasonable refresh rate. With that in mind we take a step forward to improve EDSS.

Starting from the central objective that is to understand the main reasons for the students to choose the institution as a partner in the language learning process, we have looked at different instruments to create a comprehensive panorama. From documents provided by the department of marketing the research has initially tried to map the main factors for that choice. Believing it can contribute with valuable clues and, assuming it can give significant direction for the previous finds, it is necessary to monitor students' feelings. The reasons and the sentiment of students can be an interesting method to measure the tendency of the amount of students over time, for that reason a linear regression can enhance with some predictability. To complete the exploratory analysis a comprehensive analytical element that presents the probability of renewal faced several dimensions would provide valuable information for decision makers.

3.6. **Identifying retention factors**

This phase begins with an exploratory analysis based on data provided by the department of marketing. The sector runs seasonal surveys to leverage the students' feeling regarding the institution, the instrument is quite dense and gathers useful information for their strategic decisions. For instance, the survey checks the opinion of the last promoted social event to decide what could be the next format to engage more students, aggregate the possibility of a friend's recommendation which results in a scale called Net Promoter Score (NPS)[19] and some academic aspects to promote new marketing campaigns. The inference comes from a fast review of the results and insights from students notes, and apart from reading answers and calculating NPS, no deeper analysis is made.

Poderating the available dataset this investigation has considered a way to provide more valuable information, not in a real time considering the characteristic of seasonality of the survey, but in a way that the gain of information could be real and provide the effective relationship between the analysed attributes with the student's conditions at the institution. For example, answering if the enrolled students are happier with the social activities or staff helping is more relevant for them. To evaluate the relationship between the students' status and the most relevant attributes with an association level between them, the researchers have first compiled the number of answers in a contingency table. It measures the discrepancy between observed and expected values (Fávero & Belfiore, 2017).

Status	Facilities	Friend's Recommendation	Staff Support	Learning Resources	Social Events	Total
Completed	12	4	8	15	0	39
Enrolled	27	12	42	35	17	133
Renewed	8	0	25	12	22	67
Total	**47**	**16**	**75**	**62**	**39**	**239**

Table 3.2 - Contingency table with the most relevant attributes by students status.

[19] https://en.wikipedia.org/wiki/Net_promoter_score

The table represents a segmentation of a quality survey sent by the marketing team to different students in March, 2024. Following the purpose of this investigation only two characteristics were selected, ‘Status’ and ‘Reasons to study with us’. The first characteristic gathered students who were ‘enrolled’, representing students registered for the first time in a language course; students who had already ‘completed’ the course and possibly were not in Ireland anymore and those who were enrolled for the second or even third time, considered ‘renewed’. The second characteristic allows students to choose one or more reasons to choose the institution. With those qualitative variables the next step was to represent this correlation graphically through a simple Correspondence Analysis (CA).

To Fávero and Belfiore (2017), that technique is very efficient to explore association between two categorical variables because of its nature, which avoids the concept of ‘arbitrary weighting’ and produces a visual outcome knowledge as perceptual map that can represent similarities and behaviour between selected variables. To proceed with the analysis the authors convey about the need of a cross-tabulation generated from the frequency of categorical variables observed, this is also a source to perform the validation by using Chi-Square (χ^2) test. Based on it the technique has been applied in a dataset with 239 observations and the results presented in the following figure.

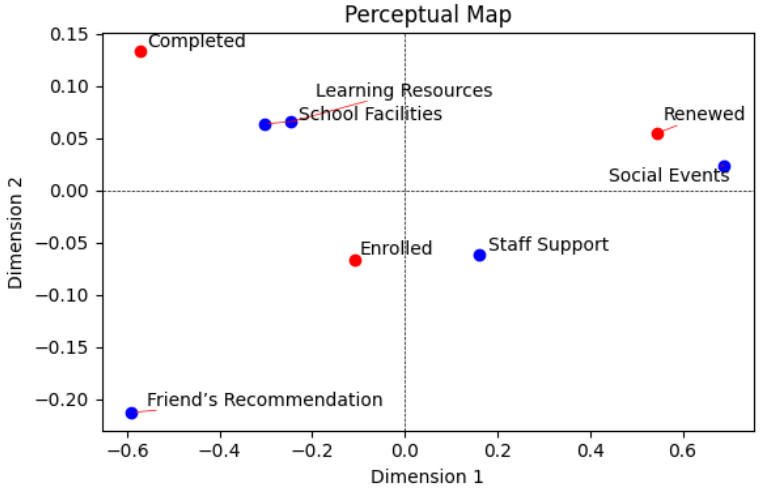

Figure 3.12 - Perceptual map with reasons for students to choose the institution.

The perceptual map presents association strength between status and categories. The distance between them indicates the strongness of associations and, consequently, nearest points have strong associations. Beginning with status ‘Enrolled’, which presents students that are currently registered and have not renewed course yet, the closeness with the ‘Staff Support’ suggests a higher similarity and provides valuable information for the marketing team. As the information comes from their attempt of comprehending students' thought with the objective of creating new campaigns it certainly provides some information gain for the sector but also gives a wide comprehension for the management of students' motivations.

When focusing on students who have renewed courses once or more we have gotten the confirmation and effectiveness of marketing strategies. During meetings and brainstorm sessions it was explained that the main way to promote school was through Social Events. They are diverse and are split between celebrations, with focus on studies, exam preparations and job interviews, the events can take place in institution premises and outside. Following the team responsible for those promotions, the main objective is to create better experiences for students and consequently lead students to keep studying at the same institution. Intention that Perceptual Map has demonstrated by showing the closeness between the status and category.

Another interesting fact on the studied dataset resulted from trying to understand why the majority of presented registers shows status ‘completed’. At this point it was declared that the subset of students invited to answer the survey were made up of those who completed the program in three months, in that specific case, students who had completed classes until January 2024. As the institution is constantly taking actions to improve processes, from a marketing team perspective, it is important to keep this data range in order to avoid answers from students that left before many changes. Possible reason behind only 39 status ‘completed’ when the database presents 4135 registers with same status. To complete, having this attribute close to ‘school facilities’ and ‘learning resources’ contributes to creating a full picture revealing how students that left sees the institution.

The map also shows what the marketing team considers a key question in their survey, ‘friend’s recommendation’. As explained early in this chapter, this attribute is the basis to calculate NPS, an instrument considers scores from 0 to 6 as detractors, between 7 and 8 are neutrals and 9 to 10 the promoters, the calculation gives a ratio from a subtraction of detractors in promoters, the result gives the indication of where the students are positioned. Although it's been used as a guide, no other information is provided and consequently the measure becomes isolated information. The Factor map in this scenario provides contextualised information by presenting a low similarity between different status and this category.

In spite of the applied instrument having no focus on course renewal, with data it becomes evident that students who have decided to renew courses with the institution consider social activities as the most valuable attribute for their experiences. It has explicitly demonstrated that the marketing strategy on promoting those kinds of interaction has some impact on students' decisions to renew their course with the institution. To evaluate those results and deliver a more precise information the study has performed a X^2 test better described at the Results and Discussion section.

3.7. Students feeling

Classifying the interlocutor's speech is an intrinsic need for human beings and it is fundamental for communication, understanding and decision-making. Currently, different artefacts are being used to establish and optimise this process with a transversal application in several fields. Those digital solutions are usually based on reviewing opinion from text, specifically known as text mining, a process that involves analysing the textual data to extract information about the writer's opinions, attitudes, and emotions towards a particular topic or product. (Alshamsi *et al*. 2020; Mouthami *et al*. 2013; Zahoor *et al*. 2020).

With that understanding every industry that cares about the customer's impression regarding its products and services must have a tool that evidences

current customers' feeling as an indicator to support decisions. (Carvalho, 2022) argues that the technique called Sentiment Analysis aims to detect the feeling expressed in sentences while Mata *et al.*(2020) highlights the growing importance of this technique due to the increase of social network users' interactions and online criticism. According to the author, the development of services based on Natural Language Processing (NLP) presents an automated way to collect, focused on categorising, understanding and interpreting the emotional tone present in texts.

For an educational institution this type of analysis can generate several benefits starting from a clear visualisation of a KPI where the management can monitor the students feeling over the time. This instrument obviously can be applied in different circumstances with different purposes, for instance, students who already have completed their studies period can provide useful information for the market team while a middle term feedback could benefit academics with the ongoing situation regarding teachers, course material, environment, etc.

From that perspective, implementing such analysis on EDSS can initially help in monitoring and over the time support decisions. To Redhu (2018), Mehta and Mehta (2020) and Cyril *et al*. (2021) one of the most common approaches to sentiment analysis is the use of machine learning techniques. It involves training algorithms on labelled datasets to classify the student perception expressed as positive, negative, or neutral. However, many other techniques such as VADER (Valence Aware Dictionary and sEntiment Reasoner)[20], Stanford CoreNLP[21], NLTK (Natural Language Toolkit)[22], Spacy[23], among others.

Between this variety of options, the researchers have considered a few points to implement a sentiment analysis tool on EDSS beginning with its intuitiveness. It is important to have something useful, simple to be integrated in Python and lightweight to be performed in a regular computer. Time to implement was also an important variable to consider, so the choice was made with a pre-trained model with

[20] https://vadersentiment.readthedocs.io/en/latest/
[21] https://stanfordnlp.github.io/CoreNLP/
[22] https://www.nltk.org/
[23] https://spacy.io/

scalability in mind to provide real-time analysis of student feedback and possible improvement.

Once more it is important to reinforce that no instrument to collect information was built, all the information provided derived from current feedback instruments already in use by the institution. As the amount of data could not be considered Big Data this situation was also relevant for the choice. So, the found solution was TextBlob[24], a powerful Python library for processing textual data. To analyse the sentiment of input text TextBlob calculates the polarity of the text with a float value between -1 and 1 classifying the text as 'Positive', 'Negative', or 'Neutral'. The result can be seen below.

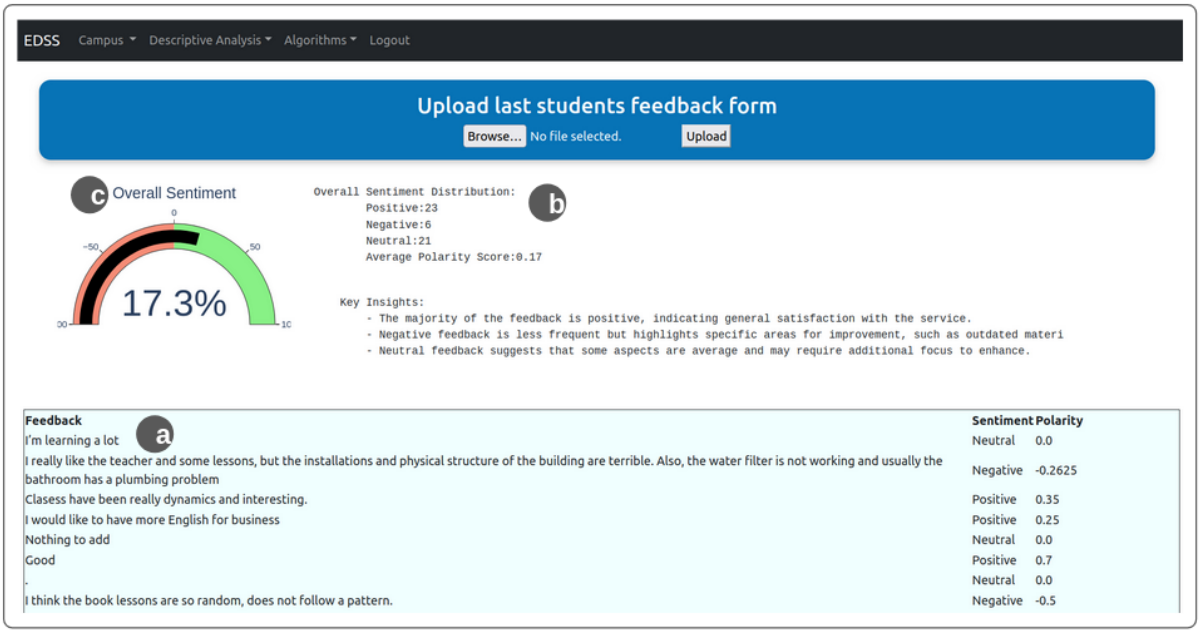

Figure 3.13 - Students feeling

Although individual classification could bring value for the institution when analysed one by one (a), a comprehensive sentiment analysis must provide visual and textual reports to help understand the overall sentiment of students (b) from different events. This flexibility allows not just management to measure it but also departments who are involved in social events that last a couple hours or academic team checking periodically the evolution when a new teacher assumes a class. Moreover EDSS can enhance the analysis with a gauge chart that provides a quick,

[24] https://pypi.org/project/textblob/

visual representation of the overall perception (c), making it easy for users to interpret the results at a glance.

Understanding the general perception by monitoring it monthly can give some indicator of the overall situation but is obviously not enough. To make a formed decision this research had to go deeper looking for data that best explain students' motivation to keep the institution as a partner during its learning process. In the next section we start to look at the attendance, considering it as one of the main criteria to remain in that country as a student, it makes sense to investigate it. More details regarding sentiment analysis will be discussed at the section Results and Conclusion.

3.8. Trends in enrollment based on students attendance

In fact, educational English learning providers are constantly looking to deliver great experiences for students by improving their environment with new facilities, course material, and specialised human resources, etc. Assuming it will be achieved at some point, it requires continuous monitoring to identify main trends over the time, that is a basic premise to comprehend, evaluate and tune current scenarios. With that in mind another purpose of EDSS is to improve mechanisms already used by the institution and offer a new way to predict future enrollments.

The current algorithm utilised by the institution is actually running considering variables such as number of the week, term, number of available seats, together with students ‘applied’, ‘enrolled’ and ‘completing’ status. For obvious reasons the model cannot be shared in this dissertation. Although the solution presents accurate tendencies, it demands that staff members gather those numbers manually, which becomes extremely susceptible to wrong data input. The problem was understood as a linear because it works with one or more explanatory variables that present itself in a linear way to answer the dependent variable.

To Fávero and Belfiore (2017), evaluating the influence of multiple independent variables on a phenomenon results in a regression problem where

several knowledge fields study the relationship between dependent and independent variables to make predictions. They acrescent that the main focus of a regression analysis is to estimate the relationship between this answer with one, simple linear model of regression defined by: $Y=\beta_0+\beta_1 x+u$. By estimating the coefficients of the regression equation we can understand that the relationship between independent variables are organised and calculated in a way to predictions about independent variables.

Once the technique to tackle this specific this problem was defined the implementation of a Linear Regression[25] Model was implemented on EDSS with the intention to predict the amount of enrolled students over time. On the process traditional libraries such as Numpy to deal with numbers, Pandas[26] pandas for data manipulation, Sklearn to perform machine learning algorithms and Plotly to present results. After loading the dataset the data has been pre-processed split between feature and target, specifically attendance was set as X (feature) while a dichotomous variable called 'renew' was set as dependent y (target). Next Figure presents the outcome of the applied algorithm.

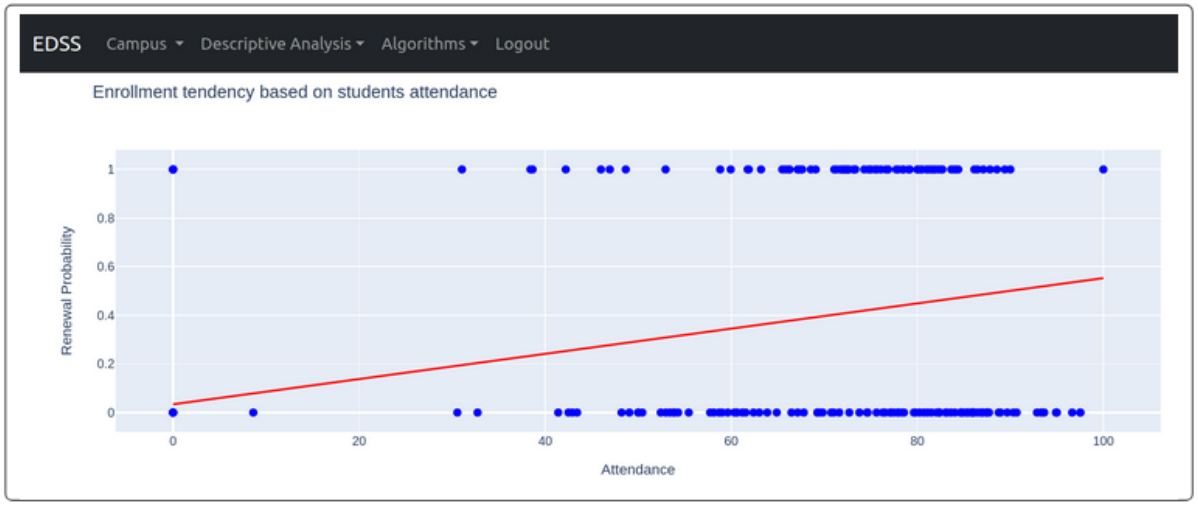

Figure 3.14 - Results from Linear Regression Model.

[25] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html
[26] https://pandas.pydata.org/

Yet the visual element represents the real intention of split students with potential to renew, the algorithm evaluation through different metrics was necessary to understand how well the model's predictions aligned with the observed values. The research looked at evaluating model's assumptions to identify potential issues through residual statistics such as mean and standard deviation and capture systematic errors using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE). Those information were compiled into a complementary panel representing the main achieved result (visualisation) in a more comprehensive way. Statistical details like results of the strength and direction of the linear relationship between the predicted and actual values will be presented at the Results and Discussion chapter.

3.9. Probability of renewal

To create robustness in this research has implemented a Machine Learning algorithm with the intention of indicating the probability of students to keep studying at the same institution. Undoubtedly this information is strategic because it can lead the management to canalise efforts on students that are truly interested in renewing the course program, undertake necessary efforts to approach those students and evaluate relevant information in that context. The identified problem for this situation is a classification, it is delimited between students that have studied once and decided to renew their VISA and those who decided to not repeat experience.

Predict the probability of students renewing their course period is a task that could be performed by different machine learning algorithms, for instance Decision Trees are useful for capturing non-linear relationships and interactions between features, Support Vector Machines (SVM) that can be effective for both linear and non-linear classification problems or even Neural Networks that is mostly applied for complex patterns, it however requires may require large dataset and more computational resources. Considering complexity, amount of available data and facility to interpretable once the final users are not a technical person, this research

has considered Logistic Regression a starting point because it approaches binary classification problems providing probability estimates for class membership.

Fernandes *et al*. 2020, suggest a five stages model to apply Logistic Regression beginning with the identification of a categorical and dichotomous variable that will serve as the target of our study, in our case the number of times that a student has completed a study period. The second stage is dedicated to note technical requirements by giving attention to multicollinearity problems, generated when a model finds a strong correlation between two or more independent variables and impacts the individual effect from independent variables over the target, and an excessive number of outliers.

Once the model was established it is time to estimate, fit the model and share resources that could facilitate the replicability and give transparency for the results. At this point it is important to reinforce that this research is using data from a private institution. It is evident that it cannot be shared, however, this research believes that by demonstrating step by step on how it was preceded other investigators could easily reproduce the applied model. As explained before, the Logistic Regression model derived from Scikit-learn[27] was performed to create binary classification and estimate the probability of students to renew service. Figure 3.14 shows the results with a dataset with 221 new registers representing current students.

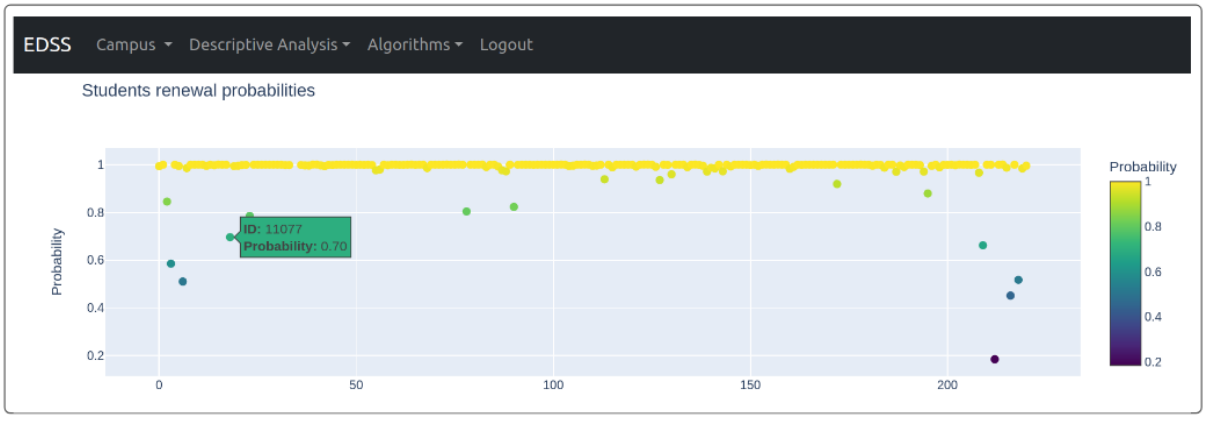

Figure 3.15 - Logistic Regression to get probability to renew course.

[27] https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

This phase was implemented with many others steps included, to not mention loading dataset, clean and prepare necessary features to run exploratory analysis, the research applied encoding of categorical variables using LabelEncoder()[28]. The objective here was to encode target labels with values between 0 and n-1and it was followed by the need of normalisation executed by StandardScaler[29] library, a crucial step Machine Learning algorithms because it ensures convergence, equal contribution between features and interpretability.

With all procedures adopted and model fit, Fernandes et al. (2020) suggests the stages to interpret and validate results, fulfilled by measuring a performance evaluation with 0.9706 of accuracy obtained from 2892 registers split between 2313 trained and 579 tested. Even with an excellent performance the result also presented significant imbalance in the dataset (530 renew vs. 49 not renewed in the test set), the reason why the model performs exceptionally well on the 'renew' class (Class 1) with high precision, recall, and F1-score, results that will be better explored at next session.

[28] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.LabelEncoder.html
[29] https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html

4. Results and discussion

Evaluating a digital system is indeed not trivial and creating an ideal scale that measures the impact of the proposed solution is undoubtedly challenging, to mitigate it and get minor errors during the proceeding this research cannot rely on a singular method. In this section we present some of the mechanisms used to understand the impact of EDSS for educational institutions but before starting the answer it is important to position the particular institution's vision regarding data using and its potential for decision-making.

Between interviews, meetings and feedback it was clear that the institution has not achieved data maturity for a few reasons, lack of a data strategy is one example. It is fundamental that, in a movement that takes informed decisions based on data, a specific department to care about data governance be implemented. Regardless it is not an instantaneous solution, at least a professional focused on data management available to examine the existence and effectiveness of data governance policies and procedures should be part of the strategy. That would be helpful to create a data culture and give focus on that matter by introducing levels of data literacy across the institution.

It is also important to remember that data is already being used at daily operation and a data professional would evaluate how effectively data is being used transforming it into a data asset. Infrastructure and available technology complement this scenario, it is providential to have someone that understand the current state of data storage and appraise the use of dedicated servers considering a minimal computational power to process analytics tools. With that we move on to evaluate the digital artefact.

The evaluation begins with individual visual elements individually, explains how it was used to perform analysis and the final plot presented. Together, the outcomes contribute to creating storytelling. Once the researcher tries to understand the system as a whole through an heuristic evaluation, checking if the solution meets requirements under the light of management decisions.

4.1. A/B testing on visual elements

Deciding which is the best visualisation to compose EDSS asks for different strategies to evaluate the efficiency of each visual element. Beginning with the objectives, each component must be built considering the quality of available data to summarise and plot it in a way the stakeholders can clearly interpret those data points. On top of what to present, how we present it is also liable for analysis and discussion with the objective of enhancing interpretation. For instance, attendance is a bottom-line information for the students and the institution, but when presented together with the ending classes date causes many doubts.

The classes run weekly with two classes per day, a hypothetical student that started classes on April 15th and had one abstention will complete that week with an overall of 80%. Another student that started two weeks before and missed the same day would complete the same weekly achieving 93%. Although it is just Math, when students and staff look at the attendance overall without considering starting day the result is a lack of communication and misinterpretation.

To avoid cases where visualisation might not deliver full information this research has evaluated a specific context to understand how attendance should be presented in order to optimise the decision-making. Here the common sense of presenting categorical and numeric variables with disambiguous visualisation was purposeful because, in fact, some elements can clearly, represent and better explain a set of data than others.

For this situation two different visualisations were suggested, previously presented at Figures 4.8 and 4.9. The intention is to gain maximum information at the first glance and support deeper analysis through the manipulation of the component. To support this decision a statistical procedure capable of rejecting or not the hypothesis of non-significative difference between the Box-Plot and Histogram on present relevant information has been performed by following the guideline (Fávero & Belfiore, 2017. p.195).

1. Selecting an appropriate statistical test;
2. Present the null hypothesis (H_0) and alternative hypothesis (H_a);
3. Set the value of alpha (α);
4. Perform the calculation and determine the critical region considering α;
5. Reject or not reject H_0.

With that 19 participants, members of staff body, were stimulated with both visualisations, after the manipulation they were invited to anonymously answer, through Google Forms[30]. What is the component which makes more sense for their daily routine and then rating the iteration using a scale that goes from 0 to 10. While 0 represents lower effectiveness, 10 was the maximum considering their routinary decision process, Table 4.1 presents the results used to decide whether to use Box-Plot or Histogram for representing a better distribution of students' attendance.

Looking at sample size, the chosen method to measure effectiveness was the paired *t-test*. To Fávero and Belfiore (2017, p. 196) one of the utilities of *t-test* is the possibility of comparing the means between two independent groups and determine whether they are statistically significantly different from each other, in other words, comparing the rating of visualisations that has presented two different methods to provide the same information. Next step is define H_0 for the desired test and H_a for its contrast, described as:

- H_0 - There is no significant difference between present attendance using Box-Plot or Histogram;
- H_a - There is a significant difference between provided visualisations.

Practically speaking, by rejecting H_0 the visual resource should be submitted to a different test that provides enough information to establish the statistical significance and efficiency of proposed visualisation and, in this case, the research would simply use the most chosen. Following Fávero and Belfiore (2017, p. 194) by making a decision based on the sampling two errors could come out, reject the null hypothesis when it is true or do not reject the null hypothesis when it is false.

[30] https://www.google.com/forms/about/

Assuming the significance level of α=0.05 the idea is to get a balance between errors.

Participant	Preference	Rating
1	Histogram	8
2	Histogram	7
3	Box-Plot	6
4	Box-Plot	9
5	Histogram	8
6	Histogram	5
7	Box-Plot	7
8	Histogram	9
9	Box-Plot	6
10	Histogram	8
11	Box-Plot	7
12	Histogram	9
13	Histogram	6
14	Box-Plot	8
15	Box-Plot	5
16	Histogram	7
17	Box-Plot	9
18	Histogram	8
19	Box-Plot	6

Table 4.1 - Visualisation rating.

With t-statistic:0.81243 and α:0.42777 the test has failed to reject the H_0 resulting in no significant difference between the visualisations in users preference. It means that both models would, from a statistical perspective, provide an effective representation of the distribution and help the team with decisions regarding this matter.

It is evident that not every visual element that makes up EDSS has a second option to present the same information and consequently does not provide subsidies to be evaluated on this matter. However, this method can be useful to demonstrate different indicators and layouts in order to build analytical panels for decision support that deliver more value for a specific team. Measuring the potentiality of an individual

element is certainly useful but only looking at isolated items might not be the best way to measure how good is the presented solution, for that reason understanding the impact of EDSS as a whole with an heuristic evaluation is necessary.

4.2. Performance and heuristic evaluation

Measuring the impact of individual elements on EDSS is challenging but analysing it as a whole also has its nuances, the reason why we understand that empirical experience should also be considered between the results. Each visualisation can deliver valuable information by itself while others only have its full potential achieved when combined into a storyline. Even with that understanding the impact of a custom-made digital solution is necessary to establish initial parameters, after that it can be pivoted, adapted and improved in future.

In this way, one of the most relevant attributes to measure EDSS effectiveness is the performance, characteristic of being easy to manipulate, clear to be interpreted and provide reasonable insights for immediate decisions could be lost if the system presents low performance. As the system was built using a web application framework, the most influential element in this indicator was the speed visualisation is loaded for users, it demands a bundle of tasks such as performing query, dataset transformations and the plot to run well.

Once the EDSS was in production the the visual elements have being observed and the performance measured using Lighthouse Report Viewer[31], an open source tool built to audit best practices, accessibility, performance, between other elements for web pages. Here it is important to consider the size of the dataset and understand that this process was not performed over a Big Data. The test was accomplished by loading pages with class numbers, elements that show students completing courses on that week and pages that showed students with better probability to renew courses. Next figure presents results from accessibility and

[31] https://developer.chrome.com/docs/lighthouse

performance, it is clear that, in spite of good results on accessibility, there is plenty of room for improvement.

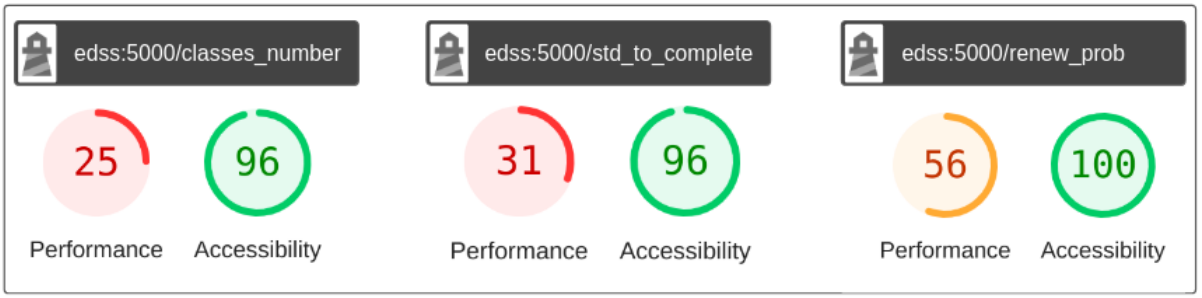

Figure 4.14 - Lighthouse performance and availability scores.

Speed Index indicator, another Lighthouse attribute that checks how quickly the contents of a page are visibly populated, resulted in on average 11.9 seconds to load pages except for the visualisation that presents students' location. Added as a ‘could have’, the feature was implemented considering its relevance and feasibility. Between the pre-processing where students post code have been extracted from its addresses, transform it into coordinates and plot the map for current students only, the load under production conditions took around 6 minutes.

Waiting too long to obtain answers can jeopardise a fast decision-making process. For this specific feature some of potential causes to have a slow load were at the pre-processing with missing or wrong information resulting in non-compatible Irish addresses and consequently problems to retrieve post code. Of course the code can always be refactored and the applied algorithm improved, it could be helpful to minimise the problem. Provided infrastructure is also a concern in this situation because the server was not dedicated for this task.

With that we conclude that the fact of being a desired feature doesn't mean it should be implemented on EDSS without a proper infrastructure and careful analysis on how addresses are being imputed on the database. The current example can impact general performance resulting in bad experience for final users, after all the solution must be truthful and offer good conditions for those who are manipulating it. Knowing the importance of users experience about the proposed solution the next step was to capture the amplitude of feelings and leverage EDSS as:

1) Very Poor - The system's performance is severely lacking, with critical failures across all attributes;
2) Poor - The system's performance is unsatisfactory, failing to meet most requirements effectively;
3) Fair - The system performs below expectations in several areas and requires significant improvements;
4) Satisfactory - The system meets basic requirements and functions adequately but has room for improvement;
5) Good - The system performs above average, with strong performance in most areas;
6) Excellent - When it performs exceptionally well across all attributes, meeting or exceeding all expectations.

To support it, this dissertation understands the a proposed framework that combines multiple theories such as Technology Acceptance Model (TAM), which leverages how users come to accept and use a technology, and Information System Success Model (ISSM) that seeks to provide a comprehensive understanding of information systems success through different dimensions, is a reasonable way to create a customized model. An example comes from Ashahril *et al*. (2023, p. 133) that proposed a framework to evaluate Citizen Engagement in using Open Government Data.

The authors model also establishes elements from the Theory of Human Behavior which are cognitive, affective, and conative. Through those dimensions the objective was to capture the quality of information and system; levels of satisfactions, perceived usefulness, perceived ease of use, trust; and level of engagement in using the tool. That said, we move forward to design our own model with five different criterias and weights to evaluate the user perception on using EDSS.

Beginning with the simplicity to operate, which is justifiable in many contexts, user-friendliness and ease of operation often determine whether a solution will be adopted and used effectively. The intention with this is to understand the quality over flexibility of the system (Cognitive) that directly impacts the perceived ease of use

(Affective). It was followed by the relevance of the solution, as mentioned before, the relevance is inherent to an artifact produced when the DSR methodology is followed, it also relates the aspect of information quality (Cognitive).

Then we mapped the clarity and intelligibility of provided information, this point tends to measure the accuracy of information quality on end users point of view but also can influence user satisfaction which is an affective attribute. The next considered point was regarding stakeholders expectations on EDSS, by capturing the reliability of users actions we once more try to get the quality of the system and the trust. To complete the scenario we have used data freshness as part of information quality completeness to understand if it somehow can influence the user's engagement with the tool.

The outcome is presented at Table 4.2, it has been defined assuming some weight for each element, it was set looking to prioritize user experience and practical usability (simplicity and relevance) over technical aspects like data freshness. With that, the users were invited to evaluate EDSS under presented criterias in a scale from 1 to 5.

N	Criteria	Question	Weight
1	Relevance of the solution	How well does EDSS address specific needs and requirements for your routinary job?	0.3
2	Simplicity to operate	How ease of use and user-friendliness of the EDSS interface and operations.	0.2
3	Clarity and Intelligibility of provided information	How clear and understandable is the information presented by the EDSS?	0.2
4	Reliability to my actions	The degree of confidence in the EDSS's ability to produce accurate and consistent results.	0.2
5	Data freshness	In your opinion, was the information retrieved using EDSS?	0.1

Table 4.2 - Heuristic Evaluation.

With 18 participants from institutional staff, the results were far from the expected. With an average of 3.2 derived from the somatory of each vote multiplied by the question weight, the inquiry has demonstrated that EDSS somehow does not attend the general purpose. Results also highlighted useful information with best

somatory score (1.1) to ‘EDSS relevance’ and the lowest (0.2) to ‘data freshness’. It shows that a decision support tool is really necessary to support decisions in that context and artefact is relevant for the institution but is also highly eligible for improvements.

4.3. Machine Learning algorithm results

Although performance, accessibility and heuristic evaluation can be gathered to appraise the impact of EDSS to answer business questions and organise it in a way that allows management to take action requires another perspective of analysis from its results. During the development of a support decision tool the use of different data sources and sophisticated algorithms are imminent therefore attention on statistical details is necessary. With more dimensions to explore the management can get different points of view under the same process, finding out answers for its demand and validate actions.

The first example of this behaviour was presented at Figure 3.12, generated from the association between attributes and categories. It shows the strength of the relationship between different variables in function of statuses to present the preference of students who have chosen the institution as a partner in the learning process. At first glance the model provided relevant information positioning students who have renewed courses and social events near each other but a deep analysis to understand this scenario is obviously necessary. In this direction the results from the applied CA model have been scrutinised and χ^2 test applied. The validation has started by checking the column masses and the total marginal frequencies of the categories for each column in the contingency table, see Table 3.2.

The most expressive mass was computed to ‘Staff Support’ with 31.38% indicating that this attribute has a significant aspect in the dataset. On the other hand, the total observations related to ‘Friend’s Recommendation’ has a smaller column mass of 6.69%. ‘Social Events’, ‘School Facilities’ and ‘Learning Resources',

that have respectively scored 16.31%, 19.66% and 25.94%, complete the pattern and relationships between the row and column categories.

Another essential measure in CA models is the Eigenvalues, it quantifies the amount of variability explained by each dimension, aiding in dimension reduction and interpretation of the results. In other words, it indicates the most influential dimension in the relationships between the row and columns from the contingency table. Figure 4.1 presents the result of the applied model with a reduction in two dimensions to explain the variability, note that with an Eigenvalue of 0.143, the first component captures 95.83%, most of the variability in the data. While the second component is significantly smaller, indicating that it explains only 4.17% of the total variance.

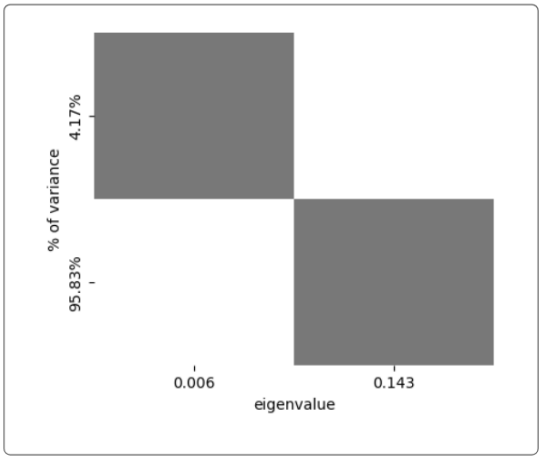

Figure 4.1 - Eigenvalues from applied Correspondence Analysis.

The next step was to check how statistically significant the presented results are by applying a hypothesis test to capture how much the observed frequencies deviate from the expected frequencies if the variables were independent. Assuming a Null Hypothesis (H_0) when no relevance at the association strength between attributes and categories is considered and Alternative Hypothesis (H_a) consequently fails to reject the null hypothesis.

The research also considered that a p-value smaller than 0.05 suggests that the association is statistically significant. With 8 degrees of freedom the returned χ^2 of 35.6430 and p-value 2.0408 we reject the null hypothesis of independence and

conclude that there is a statistically significant association between status and the other variables. In other words, social events have some relevant contribution for students to remain at the same institution.

This motivation can obviously change along the time, with that in mind and in the wake of creating a full picture to demonstrate students' motivation to renew course at the same institution, reason why Sentiment Analysis technique was applied to monitor students feeling in different periods, results from an inquiry applied in April, see figure 3.13. The result reveals that from 50 answers classified as positive (23), negative (6) and neutral (21) an latent polarity score rises with an average of 0.17 in a scale that goes from -1 to 1.

The majority of positive feedback indicates in general a good satisfaction with the provided service for that period and set of students. Result that demands careful look to avoid misunderstandings and turns into a vanity metric. First because those are punctual outcomes and must be analysed from a full perspective involving events it was applied, students integration and everyday situations that the model cannot notify. At the same time, neutral feedback suggests some aspects that may require additional focus to enhance while negative ones claim for attention and possible improvement in some areas, such as outdated course materials and facility issues.

Compiling the main reason for students to choose the institution aligned with the general feeling can be a clue to understand the scenario but in practical terms, the students decision to remain in Ireland is also subject to the regulatory laws and the main one is the attendance. Without a good average students just cannot renew the course program. The minimum required by the government is 85% so the staff is constantly monitoring this indicator looking to have improved. Next figure is a current representation of attendance distribution.

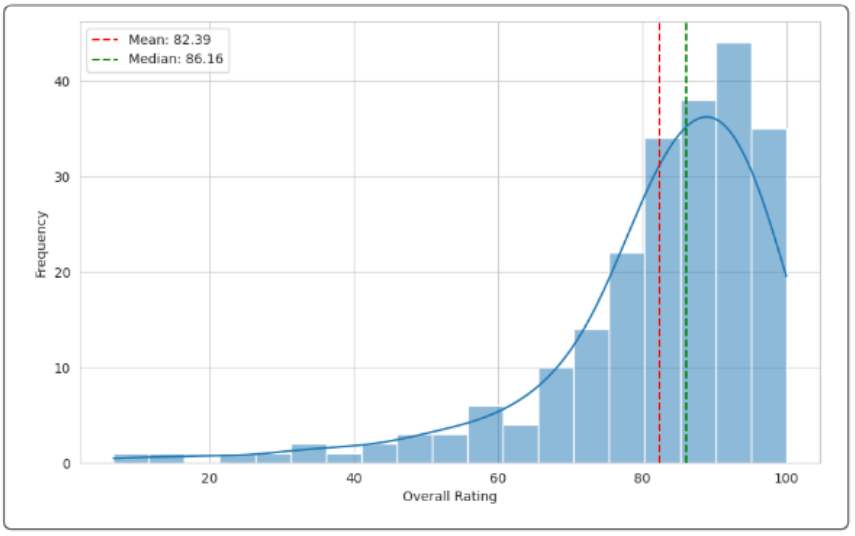

Figure 4.2 - Distribution of Overall Ratings.

The plot with overall attendance also presents some statistical measures starting with Central Tendency Measures of mean (82.39), median (86.16), and mode (100.00) which gives a good idea of general engagement and become an actionable asset for decision regarding attendance. The variability shows a standard deviation of 15.87 indicates a moderate spread of values around the mean, negative skew (-1.89) and kurtosis value of 4.66 indicating a heavy tail and a higher, sharper peak compared to a normal distribution. Results that can be enhanced with granularization such as filtering by data range, nationalities, term, between others.

Advancing at the analysis but still based on students' attendance a Linear Model was applied (presented early at Figure 3.14) generating interesting results. Beginning the analysis with actual and predicted values the presented scenario describes a correlation of 0.5002, it indicates a moderate positive relationship reflecting and it is reflected on mean of Residuals: -0.0099, which is good because it is close to zero but contrast with higher standard deviation of residuals 0.3022 in a scale from 0 to 1.

Following up on the facts, EDSS presents a coefficient: 0.0052 indicating that input variables have a minimal effect on the predicted output and intercept: 0.0342. While target variance is explained by a R^2 score of 0.2491 suggesting that the model's predictive power is also relatively weak. Based on those results EDSS

presents 0.0914 on average squared for difference between predicted and actual values (MSE), on top of that, model's predictions deviate from the current values in about 0.3024 units (RMSE) while the absolute difference between predicted and current numbers takes about 0.1876 (MAE) units.

Conjecture that evidences a weak predictive power for this dataset. The low R^2, moderated correlation, and high error rates considering the scale of the target variable suggest that a linear model may not be the most appropriate choice for this dataset. Reason why we have taken a step forward in order to analyse different dimensions and algorithms with the intention of capturing the probability of students renewing courses.

Looking to overcome this problem we have first selected categorical variables such as ‘nationality’, ‘agent’, ‘payment_method’ and ‘gender’, to understand the significance level between them and the target variable called ‘renew’, it was done using Chi-Square(χ^2) that resulted in χ^2: 1348.0261 with p-value: 6.95732 under 1106 degrees of freedom. The p-value (>0.05) indicates that the relationship between the categorical and the target variable is highly statistically significant, in other words, those categorical variables are important predictors for the 'renew' classification.

In addition, correlation between numerical variables was explored, results are presented at Figure 4.3. The matrix shows a moderate positive moderate correlation between dependent variable and assessment (0.2699), payment (0.2603), weak with attendance (0.1882) and negative with age (-0.090899). It is also notable that between independent variables (without variable ‘renew’) there is no correlation above 0.7 or less than -0.7. In summary, not high enough (0.8 or 0.9) to immediately indicate severe multicollinearity.

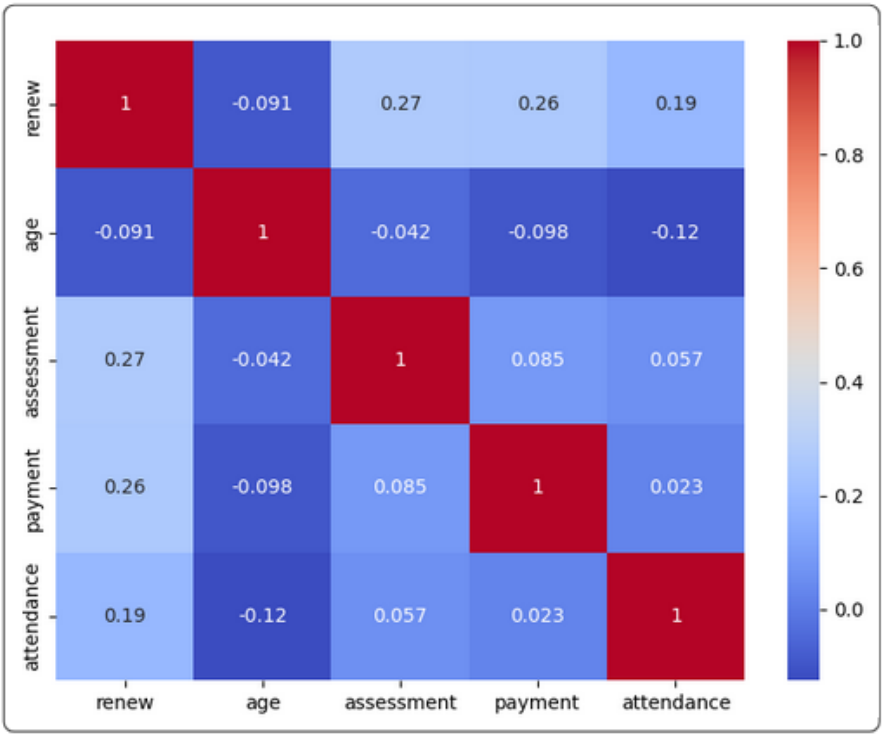

Figure 4.3 - Heatmap with correlation between numerical variables.

At the sequence of analysis a cross-validation was performed to capture how good the model was through a generalisation using K-Fold[32] technique with n_splits = 5 which gave us approximately 570 registers per fold. Our understanding is that this threshold provides a good balance between validation set size and number of iterations considering n_splits = 10 will generate less fold but more iterations while n_splits = 5 will create bigger fold with little interactions. With an average accuracy of 95.816% and standard deviation: 0.767% the described scenario suggests a consistent model considering the nature of the tackle problem presenting a reliable performance across different subsets of the data.

Applying a Logistic Regression over 2892 registers with the intention of getting the probability of students to keep studying at the same institution. The binary classification approach have considered following variables: 'nationality', 'date_of_birth', 'gender', 'agent', 'renew', 'assessment', 'payment_method', 'payment', and 'attendance', after treatment, training and application the model resulted on values that can be seeing at Table 4.3.

[32] https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.KFold.html

Classification Report

Class	Precision	Recall	F1-Score	Support
0	0.92	0.71	0.8	49
1	0.97	0.99	0.98	530

Table 4.3 - Classification results.

The presented table shows positive class, representing students that have renewed course period at least once, with majority of support and it describes an imbalanced dataset. Precision and recall have an inherent trade-off whereby improving one often decreases the other, more specifically, from all the samples the model labelled as positive precision gets the percentage of those who are really positive, it is calculated as: *True Positives / (True Positives* + *False Positives).* While recall checks the percentage of actual positive samples that were correctly identified through: *True Positives / (True Positives* + *False Negatives)*, as a result and despite imbalance, students that had no intention to renew performed reasonably which present potential for Improvement as lower recall (0.71).

The F1-score provides a balanced measure between precision and recall indicating a good performance when scores above >=80%. In this way the result of formula: 2 * (Precision * Recall) / (Precision + Recall) demonstrates strong predictive power with potential for fine-tuning on non-renewals. To complete this stage the research tried to understand the coefficient magnitude and the importance of each feature in predicting the target variable 'renew'. The table presents the payment method (0.3080) as the most influential feature as consequence it strongly affects the likelihood of renewal. Followed by assessment (0.0634) and attendance (0.0464) as the most influential feature, indicating that attendance records play a role in renewals.

Coefficient magnitude

Feature	Magnitude
payment method	0.308011
assessment	0.063415
attendance	0.046491
nationality	0.036767
gender	0.027942
age	0.008361
agent	0.000961
payment	0.000122

Table 4.4 - Coefficient results.

Results that work as initial answers for Chen (2021) initial purpose; following the authors' proposal for future investigations with a decision support system of education management based on data mining. On the other hand, EDSS diverge regarding applied technologies endorsing the architecture proposed for Chen (2017) that makes a decision support system available through a web browser.

To get a comprehensive result the ROC (Receiver Operating Characteristic) curve, a visual tool used to evaluate the performance of a binary classification model, was applied. It represents the relationship between the False Positive Rate (FPR) and the True Positive Rate (TPR) for different decision threshold values. The AUC (Area Under the Curve) presented at Figure 4.5 quantifies the model's performance, ranging from 0 to 1, where higher values indicate better performance.

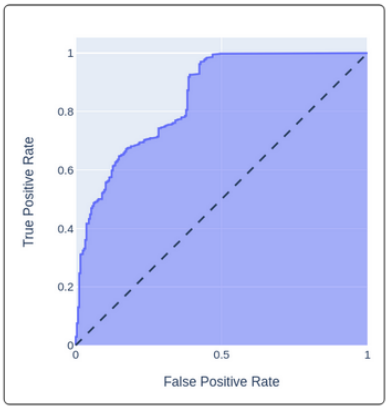

Figure 4.5 - Area under the curve.

Analysis of the ROC curve shows a model that performs well in classifying positive and negative examples, with an AUC of 0.8488. The sensitivity (TPR) increases as the threshold decreases, but this leads to a higher false positive rate (FPR), which is a common trade-off in classification models. The objective is to adjust the threshold to balance these two factors, depending on the needs of the problem.

The challenge of using machine learning algorithms to give more capacity to EDSS was first to provide robustness on what was presented, then create some tool that delivers information without doubts. Following the institution management, there is no point in providing a fancy tool. Assumption that highlights the difference between the Fuzzy Multi-Criteria Decision Making approached by Gaftandzhieva *et al.* (2023). In this way EDSS has been built looking more to what Skittou *et al*. (2022) call a recommendation system. A tool based on predictions that handle analysis results and present measures for decision makers.

Another distinguished point on EDSS is the way it has been designed, given the final objective on finding hypothetical students willing to renew their course program, the data exploration makes use of different algorithms to create a full picture. Meanwhile, authors such as Wang (2021) make use of the ID3 algorithm to work on issues on decision-making due to reliance on intuition rather than data-driven analysis and Zhao (2023) that performs Decision trees to evaluate and manage the data information to take advantage as an educational resource.

After thoroughly examining and presenting the research results, we move on to close this study. The findings discussed in the previous sections have provided valuable insights from the implementation of EDSS into an educational institution. In the following section, we will synthesise these results drawing meaningful conclusions that address our initial research questions and hypotheses. To round out this dissertation, we also outline potential avenues for future research.

5. Conclusion and future work

Researching a computer-based tool to aid educational institutions on taking informed decisions is undoubtedly a challenge. Integrate data, design models and build analytical tools to provide timely and useful information requires a huge effort on synchronising the available technology, research expertise, institutions' needs and available resources. Given such complexity the intention of this dissertation was not to exhaust the subject by delivering a singular solution, instead we have explored the possibilities and potentialities in a scenario where a digital decision support system based on data visualisation could serve as a monitoring tool enhancing a decision-making providing actionable insights.

This investigation makes use of current institution data sources, and for a data project the amount of data is obviously extremely relevant. It is well known that results with a small set can lead to unbalanced scenarios and biassed information. However, it does not reduce the importance of this study, it highlights the importance of initiating a data culture and the importance of implementing data governance. The first example can be explained with the disjunction of the term and concept. From numbers presented on EDSS the management figured out that a second type of holiday, called emergence break, would be necessary. The reason is quite simple, considering that the emergence break should not be counted on attendance the impact would indisputably be huge.

Inevitably the numbers also raise more doubts on management that consequently ask more questions, which is a positive change made through EDSS. The resource designed to answer initial questions reveals that the decision-making body makes use of outcomes as conducting wire to different questions, what is natural and desired. Assuming possible answers for those new questions, the decision process tends to rely more on provided solutions resulting in a crescent demand for data governance and quality.

Another clear example comes from vocabulary, during the meetings the institution figured out that students' status ‘completed’ were considered as a hypothetical student has completed their program period by marketing team. Meanwhile the Academic team uses that status to set students who classes but still have a holiday period. In practical terms this situation resulted in a misunderstanding with frequent need for double checking increasing workload and invisible failure on marketing campaigns because many students were taking holidays in the middle of the program.

To make informed decisions based on EDSS indicators it is imperative that presented information foments a strategic layer and that is just possible through wide extensive communication, development of valuable resources and feedback loops to measure if outcomes make sense for final users. The investigation have wrapped up that it is preferable that the system compile it and provide a small number with useful indicators through user-friendly interfaces is most valuable to the stakeholders than creating fancy visualisations for users to interpret just because it is possible, the temptation of using complex statistical methods with advanced visualisations can be discouraging for end user.

Here the example comes from higher management, to them since presented information is faithful, the preference for compiled numbers in a report. Situation that reveals missing characteristics on EDSS but also presents an interesting point regarding its update. Most of the time the data emerges from routinary tasks but some analysis will only present immediate circumstances such as the feelings of students after some action, context where periodical reports can provide reasonable answers. The cycle that involves data collecting, processing and results presentation is essential to generate this valuable resource but explains that using data in real time sometimes will just not work, therefore it still relies on a strong pipeline to treat data and make it available for analysis at the moment a decision process occurs.

For that reason we propose an adoption of implementation of different ETL tools to enhance EDSS power on data collecting, transformation and processing to advance this study. The applied way to use .CSV files derived from CRM, using tools like Apache NiFi could consume APIs automatizing the loading step. The amount of

generated data from analysis made on EDSS would also allow future research based on reuse of those results providing subsidies widely planned. For instance, a perceptual map is currently analysing factors originated from monthly research, gathering several months might lead the management to another understanding.

Future research will also demand a more robust dataset with many other dimensions to be analysed but we strongly believe that adopting EDSS since it has a small amount of data gives the institution the power to take an important step towards accepting a culture based on data. Recognizing a decision support tool as part of the operation can lead the management to make informed decisions in parallel with high concern of collecting, storing and protecting the generated data, in other words, it transforms data into a valuable asset.

References

Alshamsi, A., Bayari, R., & Salloum, S. (2020). Sentiment analysis in English texts. Advances in Science, Technology and Engineering Systems Journal, 5(6), 1683-1689. https://doi.org/10.25046/aj0506200

Alturki, A., Gable, G. G., & Bandara, W. (2011). *A design science research roadmap*. 107–123. https://doi.org/10.1007/978-3-642-20633-7_8

Ashahril, S. M., Isa, A. M., & Anwar, N. (2023). Developing a Framework of Citizen's Engagement in Open Government Data's Website. Environment-Behaviour Proceedings Journal, 8(SI12), 129-136.

Bach, B., Freeman, E., Abdul-Rahman, A., Turkay, C., Khan, S., Fan, Y., & Chen, M. (2022). *Dashboard design patterns*. *29*(1), 342–352. https://doi.org/10.1109/TVCG.2021.3058914

Barros, F., Rodrigues, B., Vieira, J., & Portela, F. (2023). *Pervasive Real-Time Analytical Framework—A Case Study on Car Parking Monitoring*. *14*(11), 584. https://doi.org/10.3390/info14110584

Barry, E. S., Merkebu, J., & Varpio, L. (2022). *State-of-the-art literature review methodology: A six-step approach for knowledge synthesis*. *11*(5), 281–288.

Batsaris, M., Kavroudakis, D., Hatjiparaskevas, E., & Agourogiannis, P. (2021). Spatial Decision Support System for Efficient School LocationAllocation. *European Journal of Geography*, *12*(4), 31–044. Scopus. https://doi.org/10.48088/ejg.m.bat.12.4.031.044

Berges, A., Ramirez, P., Pau, I., Tejero, A., & Crespo, A. G. (2021). A Framework for Strategic Intelligence Systems Applied to Education Management: A Pilot Study in the Community of Madrid. *IEEE Access*, *9*, 75313–75323. Scopus. https://doi.org/10.1109/ACCESS.2021.3081734

Bimonte, S., Edoh-Alove, E., & Coulibaly, F. A. (2021). *Map4OLAP: A web-based tool for interactive map visualization of OLAP queries*. https://doi.org/10.1109/BigData52589.2021.9671574

Chen, J. (2021). *Horizontal Model of Higher Education Management Policy Support System Based on Data Mining*. *1283*, 836–840. Scopus. https://doi.org/10.1007/978-3-030-62746-1_132

Chen, Y. (2017). *Application Research of Big Data Mining and Decision Analysis System in Colleges and Universities* (W. Jing, X. Ning, & Z. Huiyu, Eds.; WOS:000426710100244; Vol. 73, pp. 1195–1200).

Cyril, C. P. D., Beulah, J. R., Subramani, N., Mohan, P., Harshavardhan, A., & Sivabalaselvamani, D. (2021). An automated learning model for sentiment analysis and data classification of Twitter data using balanced CA-SVM. Concurrent Engineering, 29(4), 386-393. https://doi.org/10.1177/1063293X211031485

Dik, V. V., Urintsov, A. I., Dneprovskaya, N. V., & Pavlekovskaya, I. V. (2014). Prospective of e-learning toolkit enhanced by ICT development. *Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu*, *4*, 152–156. Scopus.

Dimov, D., Maula, M., & Romme, A. G. L. (2023). Crafting and assessing design science research for entrepreneurship. *Entrepreneurship Theory and Practice*, *47*(5), 1543–1567.

Dresch, A., Lacerda, D. P., & Antunes Jr, J. A. V. (2015). *Design science research* (pp. 67–102).

Ermolaev, E., Abellán Álvarez, I., Sedlmeir, J., & Fridgen, G. (2023). *z-Commerce: Designing a Data-Minimizing One-Click Checkout Solution* (pp. 3–17).

Fávero, L. P., Belfiore, P. (2017). Manual de análise de dados: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. Elsevier Brasil.

Fernandes, A. A. T., Figueiredo Filho, D. B., Rocha, E. C. D., & Nascimento, W. D. S. (2020). Read this paper if you want to learn logistic regression. Revista de Sociologia e Política, 28(74), 006. https://doi.org/10.1590/1678-987320287406en

Gaftandzhieva, S., Hussain, S., Hilčenko, S., Doneva, R., & Boykova, K. (2023). Data-driven Decision Making in Higher Education Institutions: State-of-play. *International Journal of Advanced Computer Science and Applications*, *14*(6), 397–405. Scopus. https://doi.org/10.14569/IJACSA.2023.0140642

Gerber, A., & Baskerville, R. (Eds.). (2023). *Design Science Research for a New Society: Society 5.0: 18th International Conference on Design Science Research in Information Systems and Technology, DESRIST 2023, Pretoria, South Africa, May 31–June 2, 2023, Proceedings* (Vol. 13873). Springer Nature.

Hevner, A. R., & vom Brocke, J. (2023). A Proficiency Model for Design Science Research Education. *Journal of Information Systems Education*, *34*(3), 264–278.

Houhamdi, Z., Athamena, B., Abuzaineddin, R., & Muhairat, M. (2019). A Multi-Agent System for Course Timetable Generation. *TEM JOURNAL-TECHNOLOGY EDUCATION MANAGEMENT INFORMATICS*, *8*(1), 211–221. https://doi.org/10.18421/TEM81-30

International, D. (2017). DAMA-DMBOK: data management body of knowledge. Technics Publications, LLC.

Johannesson, P., Perjons, E., Johannesson, P., & Perjons, E. (2021). Demonstrate Artefact. An Introduction to Design Science, 137-139.

Lefebvre, H., Flourac, G., Krasikov, P., & Legner, C. (2023). *Toward Cross-Company Value Generation from Data: Design Principles for Developing and Operating Data Sharing Communities* (pp. 33–49).

Lianny, R., Rizki, M., Harpito, N., & Umam, M. I. H. (2023). *Blockchain implementation in the academic administration service system*. 195–203. https://doi.org/10.1049/icp.2023.1781

Liu, C., & Song, B. (2021). Impact Assessment of Big Data on Higher Education Management Based on Time-Varying Clustering Sampling Algorithm. *Computational Intelligence and Neuroscience*.

Liu, Y. (2020). *Research on the construction of school education management decision system based on data mining framework*. 2215–2218. Scopus. https://doi.org/10.1109/ICMCCE51767.2020.00480

Mehta, P., & Pandya, S. (2020). A review on sentiment analysis methodologies, practices and applications. International Journal of Scientific and Technology Research, 9(2), 601-609.

McCoy, C., & Rosenbaum, H. (2019). Uncovering unintended and shadow practices of users of decision support system dashboards in higher education institutions. *Journal of the Association for Information Science and Technology*, *70*(4), 370–384.

Mouthami, K., Devi, K. N., & Bhaskaran, V. M. (2013). Sentiment analysis and classification based on textual reviews. 2013 International Conference on Information Communication and Embedded Systems (ICICES), 271–276. https://doi.org/10.1109/ICICES.2013.6508366

Müller, H. M., & Reuter-Oppermann, M. (2023). *Persuasive Blood Donation App Design for Individualist and Collectivist Cultures* (pp. 157–172).

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery*, *88*, 105906.

Peng, B., & Pei, X. (2022). A Decision Support System Model for Middle School Education Management Based on Sparse Clustering Algorithm. *Mobile Information Systems*, *2022*. Scopus. https://doi.org/10.1155/2022/4807395

Rahman, A. A., Adamu, Y. B., & Harun, P. (2017). *Review on dashboard application from managerial perspective*. 1–5.

Schoormann, T., Möller, F., Di Maria, M., & Große, N. (2023). Guiding Design Principle Projects: A Canvas for Young Design Science Researchers. *Journal of Information*

Systems Education, *34*(3), 307–325.

Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in Predictive Learning Analytics: A decade systematic review (2012–2022). *Education and Information Technologies*, *28*(7), 8299–8333.

Sharda, R., Delen, D., & Turban, E. (2019). *Business Intelligence e Análise de Dados para Gestão do Negócio-4*. Bookman Editora.

Skittou, M., Merrouchi, M., & Gadi, T. (2022). *A Model of an Integrated Educational Management Information System to Support Educational Planning and Decision Making: A Moroccan Case*. *745*, 167–177. Scopus. https://doi.org/10.1007/978-981-33-6893-4_16

Terence, T., GROOTBOOM, N., ZHOU, M., GUVHU, R., & Shukla, V. K. (2021). *Addressing Academic Administration Challenges in Higher Educational Institutions through Mobile Technologies and Cloud Computing*. 1–6. https://doi.org/10.1109/ICRITO51393.2021.9596144

Thuan, N. H., Tate, M., Drechsler, A., & Antunes, P. (2023). Special Issue Editorial: Introduction to Design Science Education. *Journal of Information Systems Education*, *34*(3), 256–263.

Uvalieva, I., Garifullina, Z., Utegenova, A., Toibayeva, S., & Issin, B. (2015). *Development of intelligent system to support management decision-making in education*. 6th International Conference on Modeling, Simulation, and Applied Optimization, ICMSAO 2015 - Dedicated to the Memory of Late Ibrahim El-Sadek. Scopus. https://doi.org/10.1109/ICMSAO.2015.7152249

Wang, J. (2023). *Decision Support System Model of Education Management Based on Cloud Storage Technology*. *465 LNICST*, 385–393. Scopus. https://doi.org/10.1007/978-3-031-23950-2_41

Wang, W. (2021). Model Construction and Research on Decision Support System for Education Management Based on Data Mining. *Computational Intelligence and Neuroscience*, *2021*. Scopus. https://doi.org/10.1155/2021/9056947

Xiao, X., Su, Z., Ye, Q., Qin, Z., & Wu, L. (2023). Intelligent education management system design for universities based on MTCNN face recognition algorithm. *Applied Mathematics and Nonlinear Sciences*, *9*(1).

Xie, X., & Chu, J. (2022). Data collection and visualization application of VMware workstation virtualization technology in college teaching management. *Mathematical Problems in Engineering*.

Yücel, E., & Erol, S. (2020). *The Gender Analysis of Enrolled Students: A Comparison Study of Austrian and Turkish Higher Education*. 36–47. Scopus. https://doi.org/10.1007/978-3-030-31343-2_4

Zahoor, K., Bawany, N. Z., & Hamid, S. (2020). Sentiment Analysis and Classification of Restaurant Reviews using Machine Learning. 2020 21st International Arab Conference on Information Technology (ACIT), 1–6. https://doi.org/10.1109/ACIT50332.2020.9300098

Zhao, H. (2023). Research On Construction Of Educational Management Model Based On Data Mining Technology. *Journal of Applied Science and Engineering (Taiwan)*, *26*(5), 613–621. Scopus. https://doi.org/10.6180/jase.202305_26(5).0004

Zheng, C., & Zhou, W. (2021). *Research on Information Construction and Management of Education Management Based on Data Mining*. *1881*(4). Scopus. https://doi.org/10.1088/1742-6596/1881/4/042073

Zhu, Z., & Sun, Y. (2023). Personalized information push system for education management based on big data mode and collaborative filtering algorithm. *Soft Computing*, 1–11.

Appendix A - Screening from selected studies

#	Author	Title	Year	Screening note	Status
1	Liu, S., Huang, B., Zhang, L., Li, S., Zhu, W., & Zhu, Z.	Empirical Research on the Application of OLAP to Book Publishing Decision Optimization	2014	OLAP applied on book publishing field	Excluded
2	Dik, V. V., Urintsov, A. I., Dneprovskaya, N. V., & Pavlekovskaya, I. V.	Prospective of e-learning toolkit enhanced by ICT development	2014	Exploring ICT's role in enhancing e-learning with personalized learning paths using Decision Support Systems.	Selected
3	Bhayat, I., Manuguerra, M., & Baldock, C.	A decision support model and tool to assist financial decision-making in universities	2015	Proposing a model and tool for systematic financial decision-making in universities, optimizing project portfolios and risk-return analysis.	Excluded
4	Uvalieva, I., Garifullina, Z., Utegenova, A., Toibayeva, S., & Issin, B.	Development of intelligent system to support management decision-making in education	2015	Creating an intelligent system for educational management decisions using the Analytic Hierarchy Process.	Selected
5	Feng, S.	Physical Training Model Design and Evaluation under Informatization	2015	Developing a Decision Support System (DSS) for physical training management, enhancing planning and assessment.	Excluded
6	Gan, S., & Du, H.	Research and Application of Multi-dimensional Analysis and Graphical Technology in Business Intelligence.	2015	Implementing metadata-based emergency information resource organization enhances data integration and management, facilitating decision support systems.	Excluded
7	Li, Z., Teng, X., & Wang, J.	Decision Support System on Government Emergency Management for Urban Emergency	2016	Developing a decision support system for urban emergency management to enhance rapid response and risk mitigation.	Excluded
8	Li, Q., Jiang, W., Lin, S., Gao, X., Sun, A., Luo, G., & Xu, Z.	Examination data analysis and evaluation platform based on cloud computing	2016	Developing an exam data analysis platform using cloud computing to enhance efficiency and accuracy in educational management.	Excluded
9	Yuan, Z.	Intelligent Decision Support System Development Technology of Automotive Mechanical System	2016	Developing an intelligent decision support system for automotive mechanical systems using CBR and RBR.	Excluded

10	Lu, Y., Wu, D., & Zhou, L.	Research of Emergency Information Resourse Organization Based on Metadata	2016	Implementing metadata-based emergency information resource organization enhances data integration and management, facilitating decision support systems.	Excluded
11	Yi, Z.	The Application of Data Warehouse and Data Mining Technology in Power System.	2016	Utilizing data warehouse and mining for efficient power system management: addressing integration and information extraction challenges.	Excluded
12	Li, D.	The Exploration of Enterprise E-Commerce Intelligence System	2016	Exploring enterprise e-commerce intelligence system for enhanced competitiveness: functions, components, and key techniques.	Excluded
13	Yu, J., Shu, X., Shang, C., & Liu, F.	Analysis of Power Marketing Data Mining Based on the Big Data Technology	2017	Utilizing big data and data mining in power marketing decision support systems, with focus on neural network algorithms.	Excluded
14	Chen, Y.	Application Research of Big Data Mining and Decision Analysis System in Colleges and Universities	2017	Exploring big data mining for enhanced subject management in colleges, aiming at objective decision-making.	Selected
15	Redjeki, D., & Rochyanto, H.	EFFECTIVENESS OF EDUCATION FOR KNOWLEDGE USE OF GENITALIA ANTISEPTICS FOR ADOLESCENTS	2017	Evaluating education effectiveness on genital antiseptic use among adolescents: Pretest-posttest analysis.	Excluded
16	Houhamdi, Z., Athamena, B., Abuzaineddin, R., & Muhairat, M.	A Multi-Agent System for Course Timetable Generation	2019	Developing a multi-agent system for efficient course timetable generation in universities, considering student and faculty preferences.	Selected
17	Girsang, A. S., Sunarna, D. A., Syaikhoni, A., & Ariyadi, A.	Business Intelligence for Education Management System	2019	Implementing business intelligence for education management to analyze student progress, using data warehouse and OLAP	Excluded
18	Ho, M., & Lu, J.	School competition in Hong Kong: A battle of lifting school academic performance	2019	Examining school competition in Hong Kong: Investigating the impact of academic performance and marketing practices on student intake. Results suggest a competitive focus on elevating academic standards.	Excluded
19	Masethe, M. A., Ojo, S. O., & Odunaike, S. A. (2019)	Taxonomy of recommender systems for educational data mining (EDM) techniques: A systematic review	2019	Creating a taxonomy for educational data mining techniques using recommender systems, aiding decision-makers in effective EDM method selection.	Excluded
20	Dmitriev, O.	Conceptual and Instrumental Universalization and Implementation of Scheduling Algorithm of Extreme	2020	Developing scheduling algorithms for extreme experiments on complex objects, focusing on	Excluded

		Experiment for Complex Objects		universalization and implementation.	
21	Silva, R., de Pontes Bernardo, C., Watanabe, C. Y. V., da Silva, R. M. P., & da Silva Neto, J. M.	Contributions of the internet of things in education as support tool in the educational management decision-making process	2020	Exploring IoT's role in education for informed decision-making, addressing system challenges.	Excluded
22	Suryati, L.	Decision Support System for Academic Administration Staff Achievement in STMIK IBBI Using TOPSIS-HFLTS Method	2020	Developing a DSS for academic staff achievement using TOPSIS-HFLTS method, accommodating diverse assessments.	Excluded
23	Liu, Y.	Research on the construction of school education management decision system based on data mining framework	2020	Exploring data mining for efficient school education management decision systems, addressing challenges posed by overwhelming information data.	Selected
24	Yücel, E., & Erol, S.	The Gender Analysis of Enrolled Students: A Comparison Study of Austrian and Turkish Higher Education	2020	Comparative gender analysis of enrolled students in Austrian and Turkish higher education using a data-driven decision support system.	Selected
25	Mokhtari, Y., Abra, O. E., Serrar, O., & Qbadou, M.	Towards a Decision Support System Based on a Multi-Agent Systems for Educational Management by School Project	2020	Creating a multi-agent system for decision support in Moroccan educational management through school projects.	Excluded
26	Berges, A., Ramirez, P., Pau, I., Tejero, A., & Crespo, A. G.	A Framework for Strategic Intelligence Systems Applied to Education Management: A Pilot Study in the Community of Madrid	2021	Developing a strategic intelligence system for education management, enhancing decision-making in complex educational policies.	Selected
27	Masethe, M. A., Ojo, S. O., Odunaike, S. A., & Masethe, H. D.	Framework of Recommendation Systems for Educational Data Mining (EDM) Methods: CBR-RS with KNN Implementation. In Transactions on Engineering Technologies	2021	Proposing a recommendation system framework to aid decision-making in educational data mining methods, enhancing accuracy.	Excluded
28	Chen, J.	Horizontal Model of Higher Education Management Policy Support System Based on Data Mining	2021	Developing a data mining-based decision support system for higher education management policies.	Selected
29	Wang, W.	Model Construction and Research on Decision Support System for Education Management Based on Data Mining	2021	Building a decision support system for education management using data mining, aiming to enhance efficiency and effectiveness.	Selected
30	Zheng, C., & Zhou, W.	Research on Information Construction and Management of Education Management Based on Data Mining	2021	Utilizing data mining to enhance education management through improved decision-making and information utilization.	Selected

31	Batsaris, M., Kavroudakis, D., Hatjiparaskevas, E., & Agourogiannis, P.	Spatial Decision Support System for Efficient School LocationAllocation	2021	Developing a Spatial Decision Support System for optimal school location-allocation, addressing capacity and proximity challenges in Greece.	Selected
32	Peng, B., & Pei, X.	A Decision Support System Model for Middle School Education Management Based on Sparse Clustering Algorithm	2022	Developing a decision support system for middle school education management using sparse clustering algorithm, enhancing efficiency and data utilization.	Selected
33	Skittou, M., Merrouchi, M., & Gadi, T.	A Model of an Integrated Educational Management Information System to Support Educational Planning and Decision Making: A Moroccan Case	2022	Creating an integrated educational management system for enhanced planning and decision-making: insights from Morocco.	Selected
34	Zuo, J., & Kummer, M. G. C.	A New Student Behavior Analysis Method Based on K-Means Algorithm and Consumption Data of Campus Smart Card	2022	Proposing a student behavior analysis method using K-means and campus smart card data, enhancing education management.	Excluded
35	Xie, J.	Bayesian Networks in the English Language Proficiency Test	2022	Utilizing Bayesian networks for predictive modeling in college English proficiency tests, aiding education management.	Excluded
36	Mohamed Hashim, M. A., Tlemsani, I., & Matthews, R.	Higher education strategy in digital transformation. Education and Information Technologies	2022	Developing a qualitative model for leveraging digital transformation to build competitive advantages in higher education.	Excluded
37	Alameen, A.	Improving the Accuracy of Multi-Valued Datasets in Agriculture Using Logistic Regression and LSTM-RNN Method	2022	Enhancing plant disease detection accuracy in agriculture via Logistic Regression and LSTM-RNN.	Excluded
38	Wang, J.	SOA-based Information Integration Platform for Educational Management Decision Support System	2022	Platform integration to enhances educational management: Integrates systems for efficient, timely, and secure decision support. Unifies resources and adapts to changing demands.	Excluded
39	Miftakul Amin, M., & Dwitayanti, Y.	Additive Ratio Assessment Model for Lecturer Performance Evaluation	2023	Developing an Additive Ratio Assessment model for evaluating lecturer performance in higher education.	Excluded
40	Zhang, J.	Application of Big Data in Comprehensive Management and Service of Sports Training System Under the Background of Informatization	2023	Utilizing big data for enhanced sports management and service, aiming at scientific decisions and humanized approaches	Excluded

41	Gaftandzhieva, S., Hussain, S., Hilčenko, S., Doneva, R., & Boykova, K.	Data-driven Decision Making in Higher Education Institutions: State-of-play	2023	Advancing data-driven decision-making in higher education for improved performance and sustainability.	Selected
42	Teixeira, J., Alves, S., Mariz, P., & Almeida, F.	Decision support system for the selection of students for Erasmus+ short-term mobility	2023	Developing a DSS for Erasmus+ mobility student selection, promoting inclusivity and team homogenization.	Excluded
43	Wang, J.	Decision Support System Model of Education Management Based on Cloud Storage Technology	2023	Utilizing cloud storage for an effective decision support system in educational management.	Selected
44	Amin, M. M., & Dwitayanti, Y.	Evaluation based on Distance from Average Solution and Copeland Score for The Selection of Practical Lecturers	2023	Developing a decision model for selecting practical lecturers using DSS: Evaluation via Distance from Average Solution and Copeland Score.	Excluded
45	Zhao, H.	Research On Construction Of Educational Management Model Based On Data Mining Technology	2023	Utilizing data mining for educational management model construction enhances decision support in student training.	Selected
46	Liu, Y.	The Application of OLAP and Web Technology in the Evaluation of Higher Educational Quality and the Design of Management System	2023	Exploring OLAP and web technology for effective higher education quality evaluation and management system design.	Excluded

Appendix B - Notes from PRISMA eligible readings

#	Author	Problem	Proposed Solution (Artefact)	Algorithm	Method	Results
1	Batsaris, M., Kavroudakis, D., Hatjiparaskevas, E., & Agourogiannis, P. (2021)	Regarding overlocation of students	Spatial Decision Support System using rshiny	Distances are calculated by using Dijkstra's shortest path algorithm.	Authors have used Open Street Map to present students location	The outcomes of the implementation of the SDSS are improved compared with the current allocation practices with a decrease of 8 percent of the total traveling .
2	Berges, A., Ramirez, P., Pau, I., Tejero, A., & Crespo, A. G. (2021)	Combine fragmented data sources that is leading to a lack of communication	Strategic Intelligence System to streamline data access and integration on Education Management	CRISP-DM methodology for predictive analysis; OLAP cubes for dynamic data analysis.	Systematic approach involving analysis of current model, needs detection, technology analysis, platform implementation, and functional validation.	The tool wasn't implemented
3	Chen, J. (2021)	Lack of efficient decision-making support due to the inability utilize massive data	DSS with Data Mining to extract valuable insights from educational data	Decision tree algorithm for data analysis and prediction, programmed using Visual Basic	Extract data from different information source, use SQL to build a data warehouse and create an ad-hoc application to present information	-
4	Chen, Y. (2017)	Inefficiency of decision-making under the volume of data generated by management information systems.	DSS to support higher education management and decision-making	Data mining algorithms for qualitative and quantitative analysis, including numerical analysis and knowledge processing.	ETL server for data cleaning and integration; Data warehouse server for storage; OLAP and data mining tools for analysis.	A prototype that combines data warehouse and data mining technology has been suggested to be built but no results were presented
5	Dik, V. V., Urintsov, A. I., Dneprovskaya, N. V., & Pavlekovskaya, I. V. (2014)	The study focus on individual learning trajectories	DSS as complement of Learn Management System	Markov Chain process and Electronic Performance Support System (EPSS) integrated within the DSS framework	-	-

6	Gaftandzhieva, S., Hussain, S., Hilčenko, S., Doneva, R., & Boykova, K. (2023)	Challenges in making informed decisions due to the complexity of data sources, manual data processing, and lack of awareness about the strategic significance of data	Implement data-driven decision-making tools like educational data mining, learning analytics, and business intelligence	Not specified but the mention of the need to analyze datasets and produce predictions.	Data integration; Data extraction; Software modularization	Authors call attention to the lack of quality data like inconsistent, incomplete or unavailable.
7	Houhamdi, Z., Athamena, B., Abuzaineddin, R., & Muhairat, M. (2019)	**Not available**	**Not available**	**Not available**	**Not available**	**Not available**
8	Liu, Y. (2020)	Challenges due to the overwhelming volume of data, making manual analysis and decision-making difficult and inefficient	School Education Management Decision System using Data Mining architecture	Web crawling, Batch Acquisition, and Neural Networks	Data collection, analyzes and suggests decisions based on trends	Experimental results showed improved academic performance, validating the system's efficacy in enhancing school management.
9	Skittou, M., Merrouchi, M., & Gadi, T. (2022)	Development of an integrated decision support system	Integrated Information System with DSS, Early Warning System (EWS) to predict problems, and a Recommen-dation System (RS) to propose realistic and effective measures. Makes use of Client/Server structure.	Algorithms of Classification, Clustering and Natural Language Text	ETL process with a data warehouse created from several datasources and basis to build a Visualization System (VS). The VS integrates an Early Warning System (EWS) and Recommendation System (RS). RS measures to remedy the problems raised by the EWS.	The model is a suggestion, it wasn't applied.
10	Uvalieva, I., Garifullina, Z., Utegenova, A., Toibayeva, S., & Issin, B. (2015)	Lack of efficient tools for objective analysis	DSS using the Analytic Hierarchy Process	Analytic Hierarchy Process (AHP) is utilized for hierarchical decision-making, pairwise comparisons, priority assessments, and alternative evaluations	Problem description, hierarchy construction, pairwise comparisons, priority evaluations, and consistency checks at each stage	The applied model has considered improvement of education and removal of social tensions as most influential management objectives.

11	Wang, J. (2023)	**Not available**	**Not available**	**Not available**	**Not available**	**Not available**
12	Wang, W. (2021)	Issues on decision-making due to reliance on intuition rather than data-driven analysis	Decision Support System (DSS) for education management based on data mining techniques to provide scientific and efficient decision-making support.	The ID3 algorithm, a decision tree algorithm, was utilized and improved upon to enhance efficiency and reduce time complexity.	DSS was built using a B/S model, incorporating data mining techniques to preprocess data, build models, test models, and apply them to analyze factors affecting teaching quality.	-
13	Yücel, E., & Erol, S. (2020)	The gender of enrolled students higher education	DSS using Eurostat data with focus on gender equity in higher education.	No specific algorithms are mentioned, but statistical methods were used to analyze gender disparities.	Data visualization techniques are applied to compare enrolled student numbers and percentages by education level and field.	The outcomes demonstrate a trend towards increased gender enrollment.
14	Zhao, H. (2023)	Increasing volume of data and how to take advantage as an educational resource	WEB-based education management systems	Decision tree to evaluate and manage the data information. Clustering to implement the information classification analysis. Association rule to extract the rules in the data.	The application has been split in different perspectives evaluation, examination, and student management, curriculum setting, and teaching methods.	Simulation results show that proposed algorithms have accurate results and consider the time of execution a valuable point.
15	Zheng, C., & Zhou, W. (2021)	Challenges on Education Management regarding growthing.	Data mining application to enhance teaching management	Behavior-related algorithms, cluster analysis, attribute selection, and sampling are employed to analyze student data, predict trends, understand reading preferences, and optimize course offerings.	Extracting student data from various sources and analysis aids in setting up courses, identifying students' strengths and weaknesses, and providing targeted teaching methods.	-
16	Peng, B., & Pei, X. (2022)	Issues on management due to manual methods	DSS for middle school education management	Sparse Clustering Algorithm, Alternating Direction Multiplication (ADMM) for sparse coefficients, spectral clustering	-	The paper purposes a system to manage educational information and there is no practical results

17	Zhu, Z., & Sun, Y. (2023)	Deal with overload information in education due to the vast amount of online learning resources.	Push System based on recommendation algorithms	Entropy, standard deviation and clustering algorithm	ETL with data cleaning and preprocessing, analysis and integration of clustering algorithms.	The study shows a low Mean Absolute Error (MAE) value which indicates a good recommendation accuracy
18	Xiao, X., Su, Z., Ye, Q., Qin, Z., & Wu, L. (2023)	Improve the education management system with a face recognition	Built a tailored face recognition system for student monitoring and system operation	Multi-task convolutional neural network (MTCNN), using TensorFlow and OpenCV for computational task	Encompasses logical and technical layers, including infrastructure, data storage, technical support, and user interfaces	Performance tests demonstrate the effectiveness of the system.
19	Xie, X., & Chu, J. (2022)	Lack of educational management	Implement VMware Workstation virtualization technology to streamline data collection and visualization	-	SQLServer for database management and VMware Workstation for virtualization technology, facilitating data processing, analysis, and visualization	-
21	Liu, C., & Song, B. (2021)	Assessment lacks clarity due to undefined metrics	Application that uses Time-Varying Clustering Sampling Algorithm	Time-varying clustering	Simulation to identify key factors influencing teaching effectiveness.	-