Instituto Superior
de Contabilidade
e Administração

Politécnico de Coimbra

Instituto Superior
de Contabilidade
e Administração

Politécnico de Coimbra

Descriptive and predictive analyses for the Portuguese tourism sector

Bianca Sales da Costa

ISCAC | 2021

Bianca Sales da Costa

Descriptive and predictive analyses for the Portuguese tourism sector

Coimbra, Outubro de 2021

Instituto Superior
de Contabilidade
e Administração
Politécnico de Coimbra
COIMBRA BUSINESS SCHOOL
ISCAC.pt

Bianca Sales da Costa

Descriptive and predictive analyses for the Portuguese tourism sector

Dissertação submetida ao Instituto Superior de Contabilidade e Administração de Coimbra para cumprimento dos requisitos necessários à obtenção do grau de **Mestre em Análise de Dados e Sistemas de Apoio à Decisão**, realizada sob a orientação da Professora Joana Jorge de Queiroz Leite.

Coimbra, Outubro de 2021

TERMO DE RESPONSABILIDADE

Declaro ser a autora desta dissertação, que constitui um trabalho original e inédito, que nunca foi submetido a outra Instituição de ensino superior para obtenção de um grau académico ou outra habilitação. Atesto ainda que todas as citações estão devidamente identificadas e que tenho consciência de que o plágio constitui uma grave falta de ética, que poderá resultar na anulação da presente dissertação.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

AGRADECIMENTOS

Agradeço aos docentes do Mestrado em Análise de Dados e Sistemas de Apoio à Decisão pela vasta transmissão de conhecimento.

Agradeço especialmente a Professora Joana Jorge de Queiroz Leite, que me orientou durante todo o desenvolvimento desta dissertação com muita dedicação, sempre me incentivando a melhorar.

Agradeço a minha família e meu esposo, que estiveram ao meu lado durante esta longa jornada.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

RESUMO

Considerando a importância do setor do turismo em Portugal, e o cenário global atual, onde cada vez mais pessoas estão conectadas e buscam informações na internet, propõe-se análises descritivas relacionadas aos dados de hóspedes em alojamentos turísticos em Portugal e aos dados de busca sobre os principais destinos turísticos de Portugal disponibilizados pelo Google Trends ao longo dos últimos. Também propõe-se análises preditivas do número de hóspedes em Portugal. As análises foram realizadas com base nos dados dos oito principais países de origem dos hóspedes (residentes e não residentes), nomeadamente: Portugal, Grã-Bretanha, Irlanda, Espanha, França, Alemanha, Brasil e Itália. Procurou-se avaliar se os dados de buscas, em diferentes defasagens, são relacionados com os dados dos hóspedes e se a inclusão destes dados aprimoram as previsões do número de hóspedes de Portugal. As análises preditivas foram realizadas através da execução de modelos de previsão das classes ETS, ARIMA e regressão dinâmica em diferentes procedimentos de avaliação da capacidade preditiva. Os resultados indicam a correlação entre os dados para cada país em diferentes defasagens. Eles também indicam qual é o melhor método para produzir previsões ao horizonte 1 e vários horizontes e em quais dos países analisados a inclusão dos dados de pesquisa melhora as previsões. Com base nos resultados de cada país, foram realizadas previsões para o ano de 2020 com o objetivo de comparar a quantidade expectável de hóspedes com os dados concretos e estimar as perdas do sector em decorrência da pandemia do Coronavírus (COVID-2019).

Palavras-chave: Turismo; Análises descritivas; Análises preditivas; Google Trends.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

ABSTRACT

Considering the relevance of the tourism sector in Portugal, and the current global stage where people are increasingly more connected and seek information on the internet, it is proposed to perform a descriptive analysis related to the guest data on Portuguese tourism accommodations and on the search results of the main Portuguese tourism destinations made available by Google Trends of the last few years. It is also proposed a predictive analysis of the number of guests in Portugal. The analyses were performed based on data from the eight main countries of origin of the guests (resident and non-resident), which are: Great Britain, Ireland, Spain, France, Germany, Brazil, and Italy. It was assessed if the data, considering different lag values, is related and if the inclusion of the search results improves the forecasts made by the predictive analyses of the number of guests in Portugal. The predictive analyses were performed using ETS, ARIMA and dynamic regression forecasting models considering different methods of training and test data. The results show a correlation between the data for each country for different lag values. They also indicate which is the best forecasting model for 1-month and several months horizons and for which of the analysed countries the search data inclusion was beneficial for the forecasts. Based on the results for each country, forecasts were performed for the year of 2020 with the aim to compare the expected number of guests with the actual number in order to estimate the losses on the sector due to the Coronavirus (COVID-2019) pandemic.

Keywords: Tourism; Descriptive analysis; Predictive analysis; Google Trends.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

GENERAL INDEX

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

FIGURE INDEX

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

TABLE INDEX

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

List of abbreviations and acronyms

ACF - Autocorrelation Function

AR - Autoregressive model

ARIMA - Autoregressive Integrated Moving Average

ARMA - Autoregressive‑moving-average model

ETS - Exponential Smoothing

GDP - Gross Domestic Product

GT - Google Trends

INE – Instituto Nacional de Estatística

MA - Moving-average model

PACF – Partial Autocorrelation Function

UNWTO - World Tourism Organization

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

INTRODUCTION

The tourism industry contributes directly and indirectly to the economy of almost every country. Tourist activities demand goods and services that need to be produced and supplied, resulting in the generation of jobs and increased demand for transportation, accommodation and restaurant services. As a result, in many countries, including Portugal, tourism is an important part of the national product and contributes significantly to the balance of payments and regional development.

The World Tourism Organization ([UNWTO], 2008) describes tourism as a cultural, social, and economic phenomenon that involves the movement of people to countries or places outside their usual environment.

According to the latest World Tourism Barometer report (2020), international tourist arrivals grew globally in 2019 for the tenth consecutive year, with Europe leading in terms of the number of international arrivals with 51% of the global market.

The tourist arrivals in Portugal grew between 2013 and 2019. According to the Tourism Statistics (2020) report by *Instituto Nacional de Estatística* (INE), the number of non-resident tourists arriving in Portugal reached 24.6 million in 2019, corresponding to a growth of 7.9% compared to the previous year (+7.5% in 2018). Moreover, there were 77.8 million overnight stays, corresponding to an increase of 4.3% (+3.3% in 2018). The domestic market generated 26.1 million overnight stays (33.6% of the total) and the external market generated 51.7 million overnight stays (66.4% of the total). Another report by this entity, the *Conta Satélite do Turismo* (2020), shows that tourism activities represented 15.4% of the national Gross Domestic Product (GDP) in the same year.

This scenario of continuous growth changed drastically with the coronavirus pandemic in 2020. In order to contain the spread of the virus, several countries closed their borders and introduced self-isolation measures. This resulted in a shrinkage in the global tourism sector. The UNWTO Tourism Dashboard (2021) shows that the number of international tourist arrivals shrank by 74% globally in 2020 when compared with the previous year. In Portugal, it decreased by 75%. However, the second semester of 2021 already displays signs of recovery.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Considering the importance of tourism, official entities, tourism associations and researchers have been paying close attention to the production of descriptive and predictive analyses in this sector for the last few years (Song, Qiu, & Park, 2019). These analyses can help companies and governments to formulate their future strategies by understanding the behavior of the sector over the previous years and by forecasting tourism demand. Usually, these analyses are performed in terms of tourist arrivals, tourism expenses and length of stay from data made available by government institutions through statistical reports.

Although these reports are periodically published, they do not always satisfy the needs of users, because they are incomplete or published too late. In the face of that, some researchers have used other sources available on the Internet, especially data from search engines, associated with traditional data to make theirs analyses (Li et al., 2017). The rational that supports this option is that, over the last decades, the Internet has been changing the way people have access to both tourism information and the choice of destination. Due to an increasingly connected world, it is expected that even more people will use the Internet to search for this information.

In this context, Google holds a prominent place, since it is the most used search engine (StatCounter, 2021), and Google Trends (GT) is the associated data source, as it provides search frequency results for a given term in relation to the total volume of searches performed on the Google platform, given a certain period and region.

Regarding the literature on the use of GT data in tourism demand forecasting, several studies have been published over the last decade such as Park et al. (2016), Dinis et al. (2017) and Antolini and Grassini (2019).

Considering the importance of tourism in Portugal, the aim of this study is to carry out a descriptive analysis of tourism demand of the main countries of origin of tourists who arrived in Portugal between 2013 and 2019. These countries are: Portugal, Great Britain, Ireland, Spain, France, Germany, Brazil, and Italy.

Another objective is to perform a descriptive analysis of the GT data. Additionally, to analyze if the volume of web searches of the main tourist destinations in Portugal, provided by GT, are correlated with the conventional statistics provided by INE and to

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

verify whether the use of GT data, incorporated in different lag values, improves the tourist demand forecasting models. Moreover, another objective of this study is to make tourist demand forecasts for the year of 2020 and to compare the results with real data to measure the losses in the sector as a result of the COVID-19 pandemic.

This study was split in three main chapters. Chapter 1 describes a literature review about tourism data, GT data and forecasting. Chapter 2 details the used methodology and Chapter 3 presents the final results and discussion. It ends with the conclusions.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

1 LITERATURE REVIEW

The objective of this chapter is to review the literature regarding the relevant concepts and methods for tourism forecast. Section 1 is dedicated to tourism data, namely its definition, sources, and issues. Section 2 presents GT's data and its potential. Section 3 is a short overview on time series and the main concepts related to exploratory analysis. Section 4 introduces forecasting, specifically its steps and relevant aspects for evaluating the predictive ability of the methods. Finally, Section 5 reviews the most used forecasting methods in tourism.

1.1 Tourism data

According to UNWTO (2008), "Tourism is a social, cultural and economic phenomenon which entails the movement of people to countries or places outside their usual environment for personal or business/professional purposes.". Camilleri (2018) explains that individuals become tourists when they voluntarily travel to another region, other than its residence, and get involved in different activities. Tourists can be classified as national or foreign. The former refers to individuals who travel exclusively within the national borders of their country of residence, while the latter refers to those who travel to other countries.

Tourism data is generally made available by government institutions that are responsible for producing and disseminating official statistical data in their countries. The main indicator is the number of tourist arrivals. Additionally, there are other variables used to measure the consumption of goods and services activities in the tourism sector, such as the number of guests, overnight stays, and total revenue in tourist accommodation establishments.

According to the Travel & Tourism Competitiveness Report (2018), Portugal ranked 75th out of 136 countries, in terms of quality and coverage of statistical information for the tourism sector in 2017. This indicator measures the quality and frequency of the provision of travel and tourism data by government institutions.

Portugal's tourism data is made available by INE, which operates as an official and independent Portuguese entity. There are two main reports on tourism: 'Survey on Guests

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Stays on Hotels and Other Accommodation Establishments' and 'Tourism Statistics'. The former is published monthly, and it contains data about the number of tourist arrivals, number of guests, overnight stays, and total revenue in tourist accommodation establishments in the previous months. The latter report is usually published annually in the middle of the year with data related to the previous year. It contains additional relevant data not available in the other report, such as reason for the trip, gender and age of tourists, country of origin of guests by region of Portugal.

Dinis (2016) assumed that, due to financial constraints in countries like Portugal, it has been increasingly difficult to collect and disseminate statistical information on a regular basis, since statistical data, mainly about tourism demand, is gathered through surveys.

1.2 Google Trends data

According to Dinis et al. (2017), people increasingly use the Internet in all phases of the travel cycle, and they usually start the decision-making process using a search engine. Therefore, search engine data can reflect the interests and desires of the tourist consumer with the advantage of being always available. In addition, Hu et al. (2020) highlight that the search traffic, provided by search engines, is useful for quickly detecting a particular phenomenon and is consequently an excellent monitoring tool. Therefore, search engine data can be seen as an explanatory variable for tourist demand.

The Search Engine Market Share Worldwide report, produced by StatCounter (2021), recognizes Google as the most used search engine, representing 92% of the total market. The Google ecosystem provides several tools, namely GT, which is a public website that makes available its search data and shows how frequently a given search term is entered into Google's search engine relative to the site's total search volume over a given period of time (Google, 2020).

GT is an excellent platform for observing people's information-seeking activities. It offers real-time data on the needs, desires, demands and interests of its users. In addition, it also offers a variety of options for comparing search terms (Jun et al., 2018). However, it is important to note that GT data does not report the current search volume, since it is normalized to the maximum value over the entire period, since 2004. Therefore, that the maximum value of the returned series is 100 (Antolini & Grassini, 2019).

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Choi and Varian (2012) explain that the GT data is a query index that starts with the query share: the total query volume for search terms in a given geographic region divided by the total number of queries in that region at a point in time. However, although it is possible to assume that the GT data contains valuable information about travel intentions, there is no knowledge about the sample picture, and it is not probabilistic (Carrière-Swallow and Labbé, 2013).

Antolini and Grassini (2019) suggest that, even though data from search engines has gained space in recent studies, there is a challenge in choosing the data to incorporate it with official statistical data to improve the forecasting performance of the models. Choi and Varian (2012) state that, due to the immense number of queries, the big challenge is to "determine exactly which queries are the most predictive for a particular purpose". Moreover, the query results are limited to the language and expressions that people use as search terms.

Data from search engine are structured time series data that reflect user attention on certain topics through keywords. Furthermore, the data is collected on a daily, weekly, and monthly basis (Li et al., 2017).

Currently, GT classifies the search terms in 26 categories and 1106 sub-categories. The category related with tourism activities is referred as Travel, which has 11 sub-categories. The full list of categories and the list of travel sub-categories are shown in Table 1.1 and Table **1.2**, respectively.

Table 1.1 Google Trend Categories: name and identification number

Name	ID	Name	ID	Name	ID
All Categories	0	Games	8	People & Society	14
Arts & Entertainment	3	Health	45	Pets & Animals	66
Autos & Vehicles	47	Hobbies & Leisure	65	Real Estate	29
Beauty & Fitness	44	Home & Garden	11	Reference	533
Books & Literature	22	Internet & Telecom	13	Science	174
Business & Industrial	12	Jobs & Education	958	Shopping	18
Computers & Electronics	5	Law & Government	19	Sports	20
Finance	7	News	16	Travel	67
Food & Drink	71	Online Communities	299		

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Table 1.2 Google Trend Travel Subcategories

Name	ID
Air Travel	203
Bus & Rail	708
Car Rental & Taxi Services	205
Carpooling & Ridesharing	1339
Cruises & Charters	206
Hotels & Accommodations	179
Luggage & Travel Accessories	1003
Specialty Travel	1004
Tourist Destinations	208
Travel Agencies & Services	1010
Travel Guides & Travelogues	1011

Filtering categories and sub-categories is a useful resource that allows restricting query results with generic terms. It is also useful for restricting words that have more than one meaning. Furthermore, it is possible to filter the results by region/country and to use punctuation in filtering search results as mentioned in Table 1.3.

Table 1.3 Search term and types of results

Search term	Type of results
Portugal Lisbon	Results include searches containing both Portugal and Lisbon in any order. No misspellings, spelling variations, synonyms, plural, or singular versions of your terms are included.
“Portugal Lisbon”	Results include the exact phrase inside double quotation marks, possibly with words before or after.
Portugal + Lisbon	Results include searches containing the words "Portugal" OR "Lisbon".
Portugal - Lisbon	Results include searches containing the word "Portugal," but exclude searches with the word "Lisbon".
Center + Centre + Centere	Results include alternative spellings like “Centre” or “Centere,” and common misspellings like "Centere." Trends considers each version of a word a different search, including misspellings.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

1.3 Time Series

Statistically, periodical tourism and search engine data forms time series. Latorre and Cardoso (2001) define a time series, also called a historical series, as a sequence of data obtained at regular intervals of time during a specific period. This set can be obtained through periodic observations of the event of interest.

Hyndman et al. (2008) explain that, when viewing data as a time series (i.e., respecting its sequentiality), it is possible to produce specific descriptive analysis, which summarizes and explores the behavior of the data and its components, such as trend (T), seasonal (S), cycle (C), and irregular or error (E). Each of these concepts can be explained as follows:

- Trend: There is a trend when there is a long-term increase or decrease in data.
- Cyclic: A cycle occurs when the data shows ups and downs that do not have a fixed frequency. These fluctuations are generally due to economic conditions.
- Seasonal: A seasonal pattern occurs when a time series is affected by seasonal factors, such as the time of year or the day of the week. Seasonality is always fixed and known.
- Irregular or error: The unpredictable component of the series.

According to Athanasopoulos and Hyndman (2021), through the decomposition process it is possible to separate the time series into its components. Also, when a time series is decomposed, the trend and cycle components are often combined into a single component. This happens when the length of the time series does not allow to capture the economic cycles. Thus, a time series can be seen as composed by the following components: a trend-cycle, a seasonal, and a remaining component (containing everything else in the time series).

Some of the most prominent methods in the literature are X-11 and STL decomposition. The X-11 method assumes that the main components of a time series follow both the additive and multiplicative decomposition. The process is automatic and tends to be highly robust to outliers and level changes in the time series (Dagum & Bianconcini, 2016). Athanasopoulos and Hyndman (2021) explain that STL is an acronym for "Seasonal and Trend decomposition using Loess," where loess is a method for estimating

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

nonlinear relationships. Also, STL provides several advantages over the X-11 method, since it handles any type of seasonality, and it is robust to outliers.

According to Shmueli and Lichtendahl Jr. (2016) outliers are extreme data values and their presence can affect or distort descriptive and predictive analyses. Therefore, it is important to carefully evaluate and, if necessary, replace these values. It is possible to detect outliers in a time series by using decomposition, since it allows removing the trend-cycle and seasonal components. After that, anomaly (i.e., outlier) detection is performed on the "remainder" (i.e., the irregular component) using, for example, the standard method based on the interquartile range (IQR).

The dependence of consecutive observations in a time series can be exhibited through the autocorrelation function (ACF), as autocorrelation measures the linear relationship between lagged values of a time series (Athanasopoulos & Hyndman, 2021). If a time series shows no autocorrelation, it is called white noise.

Another important aspect that requires attention during the time series analysis process is the concept of stationarity. A time series is stationary when its statistical properties, such as mean, variance, and autocorrelation, are constant over time. On the other hand, it is non-stationary when its statistical properties change over time. There are some ways to analyze if a time series is stationary or non-stationary, such as the time plot of the data and the ACF plot (Athanasopoulos & Hyndman, 2021).

1.4 Forecasting

According to the authors Athanasopoulos and Hyndman (2021, p. 2), “Forecasting has fascinated people for thousands of years, sometimes being considered a sign of divine inspiration, and sometimes being seen as a criminal activity.” Furthermore, the term Forecasting can be understood as the process of predicting the future as accurately as possible, given all available information, including historical data and knowledge of any future events that may impact the forecasts.

Shmueli and Lichtendahl Jr. (2016) point out that there are some basic steps that must be followed during the forecasting process:

1. Goal definition;

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

2. Gathering data;
3. Exploratory analysis;
4. Selection of a set of potential forecasting methods;
5. Application of the set of forecasting methods;
6. Comparison in terms of forecast accuracy;
7. Forecast generation using the best method.

Step six of the forecasting process draws attention to the need of evaluating forecast accuracy. This is done by comparing the predictions with the actual observed values. However, the proper assessent of a model, which results of the application of a method, is done by comparing its preditions with observed values that were not used to fit it. Therefore, there are two aspects to consider: the experimental setup and the accuracy measures.

Regarding the experimental setup, Athanasopoulos and Hyndman (2021) clarify that one common method consists of separating the data series into two parts: training and test data. The training data is used to estimate any parameters of a forecasting model and the test data is used to evaluate its accuracy. In time series, this split respects the timeline, meaning that the training set is composed by the earliest observations and the most recent observations are reserved for the test set. A sophisticated extention of this method is the time series cross-validation, in which a series of training and test sets are produced on subsets of data.

After the application of the set of forecasting methods, it is important to evaluate the forecast accuracy. According to Lawrence and Klimberg (2010), there are several metrics used to evaluate the size of the error, which is the difference between the observed and the estimated values. The most common ones are: Mean Error (ME), Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Scaled Error (MASE), and Mean Absolute Percentage Error (MAPE).

Hyndman and Koehler (2006) drew a comparison between different forecast precision metrics while highlighting their advantages and disadvantages. They suggested that MAE, MSE and RMSE may be preferred if all series are on the same scale, the MASE is

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

the best available measure in data which is close to zero or negative, and the MAPE may be preferred if all data is positive and much greater than zero.

1.5 Forecasting methods in Tourism

In their study, Song et al. (2019) reviewed general trends and the evolution of tourism demand forecasting methods from an historical perspective. They summarized the methodological development of prediction models from 211 key studies published between 1968 and 2018, and showed that these prediction methods continue to evolve. According to these authors, the methodological approaches to forecasting tourism demand fall into four categories: time series models, econometric models, artificial intelligence based models, and judgment methods. However, time series models were the most used among the analyzed studies.

The time series models identify, through the collection of repeated measurements over time, the behavior of a variable and extrapolate the data to the future based on this behavior (Dinis, 2016). It is suggested that first it is necessary to identify the presence of these characteristics in the data in order to, subsequently, choose the forecasting method that is capable of capturing and dealing with the patterns in an appropriate manner.

According to Song et al. (2019), the time series models include the Naïve, autoregressive (AR), simple exponential smoothing (ES), moving average (MA), while the advanced ones are autoregressive integrated moving average (ARIMA) and seasonal autoregressive integrated moving average (SARIMA).

In view of the variety of forecasting methods available, researchers have attempted to summarize the time series methods that are most used and that show the best results in the tourism sector. In one of these studies, Song and Li (2008) concluded that ARIMA models were the most used to forecast tourism demand among the articles analyzed in their study.

Nonetheless, Athanasopoulos et al. (2011) believe that, despite the best efforts of the authors, the diversity of the studies did not lead to a consensus on the relative forecasting performance of the methods commonly used when applied to tourism data.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

In view of that, a literature review was carried out on the main time series forecasting methods in tourism, in order to choose those that best apply to the data set in this study. The methods were organized in three classes: exponential smoothing, ARIMA and dynamic regression.

1.5.1 Exponential smoothing

Hyndman et al. (2008) affirm that "Historically, exponential smoothing describes a class of forecasting methods…" and that each method has its own properties but have in common that predictions are weighted combinations of previous observations, so that weights decrease exponentially as the observations get older.

According to Athanasopoulos and Hyndman (2021) the idea appears to have originated with Robert G. Brown around 1944. However, in the 1950s, he extended this method from continuous time series to discrete time series and included terms to deal with trend and seasonality. Charles Holt was also working independently on an exponential smoothing method. This method is known as the Holt's Linear Method and it differed from Brown's with respect to the smoothing of the trend and seasonal components (Hyndman et al., 2008). In 1960 Peter Winter published a paper that provided empirical tests for Holt's methods on additive and multiplicative seasonal exponential smoothing. This method is similar to Holt's linear method, with one additional equation for dealing with seasonality. As a result, seasonal versions of Holt methods are often referred to as Holt-Winters methods. These methods continued to evolve, culminating with systematization and extension proposed by Hyndman et al. (2008), where the exponential smoothing methods were expanded from algorithms that could only generate point forecasts to fully statistical models, which were also able to provide prediction intervals.

Due to the wide variety of different configurations present on the exponential smoothing methods, they must be chosen according to the characteristics of the used time series. Collectively, the exponential smoothing class methods are referred to as ETS models, from Error, Trend and Seasonality. Hyndman et al. (2008) explain that ETS can also be considered an abbreviation of ExponenTial Smoothing. Furthermore, the notation ETS(·,·,·) helps in remembering the order in which the components errors, trend and seasonality are specified, as below:

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

$$\text{Error} = \{\text{A}, \text{M}\}$$

$$\text{Trend} = \{\text{N}, \text{A}, \text{A}_d, \text{M}, \text{M}_d\}$$

$$\text{Seasonal} = \{\text{N}, \text{A}, \text{M}\}$$

where: N = None, A = Additive, M = Multiplicative, d = Damped. The combinations of all versions led to a total of 30 variations of ETS, even though some of them can present numerical stability issues. In view of this, an automated procedure for model selection becomes useful.

1.5.2 ARIMA

According to Kirchgässner and Wolters (2007), the autoregressive processes emerged in traditional econometrics in 1949, through Donald Cochrane and Guy H. Orcutt who used the first order autoregressive process to model the residuals of a regression equation.

An autoregressive model of order q is referred to as AR(p) and it assumes that the current value of the series is a linear combination of the past p values of the series and a white noise (Santos, 2012). In other words, in an autoregressive model the variable of interest is predicted through the linear combination of its previous values (Athanasopoulos & Hyndman, 2021). A moving average model of order q is referred to as MA(q) and it uses previous forecast errors in a regression model instead of using previous values of the forecast variable in a regression (Adhikari & Agrawal, 2013). The combination of the autoregressive processes of order p and moving averages of order q originates the autoregressive and moving average model, referred to as ARMA(p, q).

According to Fava (2000), the combination of the Autoregressive (AR), Integration (I) and Moving Average (MA) components result in the Autoregressive Integrated Moving Average (ARIMA). It is possible to write it as ARIMA(p, d, q), where:

p = order of the autoregressive part;

d = degree of first differencing involved;

q = order of the moving average part.

According to Song et al. (2019), the seasonal ARIMA model was first proposed by Box and Jenkins in 1970 and it is known as SARIMA$(\text{p}, \text{d}, \text{q})(\text{P}, \text{D}, \text{Q})\text{m}$, where:

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

P = the order of the seasonal autoregressive model;

D = the number of the seasonal differences;

Q = the order of the seasonal moving average model;

m = the seasonal period.

A SARIMA model is formed by including additional seasonal terms in the ARIMA. The seasonal part of the model consists of terms that are very similar to the non-seasonal components of the model, but they involve backshifts of the seasonal period (Athanasopoulos & Hyndman, 2021).

1.5.3 Dynamic regression

Athanasopoulos and Hyndman (2021) describe the basic concept of regression models as the projection of the time series of interest y assuming there is a relationship to another time series x.

To measure the relationships between variables and what they represent, it is necessary to use the correlation coefficients (r). Schober et al. (2018) describe that the correlation is a measure of monotonic association between two variables. A monotonic relationship between two variables is one in which, as the value of one variable increases, so does the value of the other variable; or as the value of one variable increases, the value of the other variable decreases. The relationship can be classified as:

- Positive correlation: the other variable also tends to increase;
- Negative correlation: the other variable tends to decrease;
- No correlation: the other variable does not tend to increase or decrease.

Moreover, according to the authors, several approaches suggest that the correlation coefficient can describe a "weak", "moderate" or "strong" relationship. Generally, researchers agree that a correlation coefficient lower than 0.1 indicates an insignificant relationship and that a coefficient higher than 0.9 suggests a very strong relationship, however the intermediate values are debatable.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Some of the common correlation coefficients are Pearson, Spearman and Kendall. Depending on the form and how the variables behave, one correlation coefficient may be more appropriate than another.

There are different regression models and the choice of the best one to be used depends on the number of independent variables, type of dependent variables and the shape of the regression line. In this study, the literature review was focused on the dynamic regression model.

According to Peter and Silvia (2012), when an ARIMA model includes other exogenous variables, it is sometimes referred to as an ARIMAX model and this model is referred to as a dynamic regression.

Athanasopoulos and Hyndman (2021) explain that in a dynamic regression, the error term is an ARIMA process, whereas in an ordinary regression, the error term is white noise. This ARIMA process is where the historical information about the time series is incorporated.

Sometimes the impact of a x-reg argument that is included in a regression model will not be simple and immediate. During the travel planning process, for example, people are expected to initiate the decision-making process through online searches some time before the departure date. In situations like this, it can be helpful to allow for lagged effects of the predictor.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

2 METHODOLOGY

The aim of this chapter is to present the adopted methodology during the empirical development of this study. Since the approaches considered are, to some extent, related to the possibilities the software offered, in Section 1 it is presented the software used to perform the data gathering, processing and analyses. Section 2 consists of a detailed description about the data gathering and description processes. Session 3 describes which descriptive analyses were carried out. Section 4 details the data pre-processing that was performed after the descriptive analyses. Section 5 describes the forecasting methods

2.1 Software

The Rstudio software was chosen to perform the tasks related to data gathering, data processing, descriptive and predictive analyses. It is an integrated development environment (IDE) for the R programming language, which is open source and widely used for data analysis. Some of the functions used on this study were made available through packages, which are collections of functions and data sets developed by the community. In the following sections, along with a description of the procedures, the R functions and packages used are also listed.

2.2 Data gathering

Two variables were used on this study: **guests** and **hits**. The former is a time series that represents the total number of guests in tourism accommodations in Portugal on monthly intervals between Jan-2013 and Dez-2020. The latter is also a time series that corresponds to the GT results considering a combination of search terms of the main tourist regions in Portugal. It is described on monthly intervals between Jan-2012 and Dez-2020.

Both variables were gathered taking into account the 8 countries that were considered in this study. Therefore, it consisted of a total of 16 time series. The countries considered correspond the main countries of origin guests in Portugal, namely: Portugal, Great Britain, Ireland, Spain, France, Germany, Brazil, and Italy.

Due to its more frequently publication, the ‘Survey on Guests Stays on Hotels and Other Accommodation Establishment’ (2021) was chosen as the source of the guests variable. As mentioned previously, the report contains the main statistics on the tourism

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

accommodation sector in Portugal. It is available in Excel format, and it contains 19 tables, as shown in Table 2.1.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Table 2.1 Survey on Guests Stays on Hotels and Other Accommodation Establishment

Table Name	Table Description
1.Sintese	Summary Information
2.H_N	Guests in tourist accommodation establishments, by NUTS II
3.HNR_P	Non-resident guests in tourist accommodation establishments, by country of residence
4.D_N	Overnight stays in tourist accommodation establishments, by NUTS II
5.D_Tipo	Overnight stays in tourist accommodation establishments, by type of establishment
6.DR_N	Overnight stays in tourist accommodation establishments by residents in Portugal, by NUTS II
7.DNR_N	Overnight stays in tourist accommodation establishments by non-residents, by NUTS II
8.DNR_P	Overnight stays in tourist accommodation establishments by non-residents, by country of residence
9.EM_N	Average stay in tourist accommodation establishments, by NUTS II
10.PT	Total revenue in tourist accommodation establishments, by NUTS II
11.PT_Tipo	Total revenue in tourist accommodation establishments, by type of establishment
12.PA	Revenue from accommodation in tourist accommodation establishments, by NUTS II
13.PA_Tipo	Revenue from accommodation in tourist accommodation establishments, by type of establishment
14.RevPAR_N	Revenue per available room (RevPAR) in tourist accommodation establishments, by NUTS II
15.ADR_N	Average Daily Rate (ADR) in tourist accommodation establishments, by NUTS II
16.TO_N	Net bed occupancy rate in tourist accommodation establishments, by NUTS II
17.D_MN	Overnight stays in tourist accommodation establishments, by municipality
18.DR_MN	Overnight stays in tourist accommodation establishments by residents in Portugal, by municipality
19.DNR_MN	Overnight stays in tourist accommodation establishments by non-residents, by municipality

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The tables related to guests in tourist accommodation establishments, 2.H_N and 3.HNR_P, were chosen to create an Excel file called ine_2020.xlsx. On the same file, an additional table called 'dt_guests_country_residence' was included, which contains 4 attributes, as shown in Table 2.2, with 768 observations, 96 for each country of origin.

Table 2.2 Guests attributes

Atribute Name	Type	Description
country	chr	Name of the country
guests	num	Total number of guests
geo	chr	Country name acronym
date	mth	Date (monthly)

The variable GT Hits was obtained through queries on the GT platform using the gtrends() function from the gtrends package (Massicotte & Eddelbuettel, 2021). The combinations of search terms used are the main tourist regions of Portugal in the native language of each country, as shown in Table 2.3 and Table 2.4. The category Travels (67) and the subcategory Tourist destinations (208) were selected.

Table 2.3 Country Query Information

Country	First Language	GEO Location	ISSO Code
United Kingdom	English	GB	en-GB
Ireland	English	IE	en-IE
Spain	Spanish	ES	es-ES
France	French	FR	fr-FR
Germany	German	DE	de-DE
Brazil	Portuguese	BR	pt-BR
Portugal	Portuguese	PT	pt-PT
Italy	Italian	IT	it-IT

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Table 2.4 Keywords Query

Country	Keyword
United Kingdom	"Porto Portugal+Lisbon+Algarve+Azores+Madeira Islands"
Ireland	"Porto Portugal+Lisbon+Algarve+Azores+Madeira Islands"
Spain	"Porto Portugal+Lisboa+Algarve+Azores+Isla de Madeira"
France	"Porto Portugal+Lisbonne+Algarve+Açores+Île de Madère"
Germany	"Porto Portugal+Lissabon+Algarve+Azoren+Auf Madeira"
Brazil	"Porto Portugal+Lisboa+Algarve+Açores+Ilha da Madeira"
Portugal	"Porto Portugal+Lisboa+Algarve+Açores+Ilha da Madeira"
Italy	"Porto Portogallo+Lisbona+Algarve+Azzorre+Isola di Madeira"

An example of the query can be seen in Figure 2.1.

Figure 2.1 Google Trends Query

```
geo_query <- "GB"
country <- "United Kingdom"

# Google Trends Query
google_trends <- gtrends(keyword = keywords,
                  geo = geo_query,
                  time = "2012-01-01 2019-12-31",
                  gprop = c("web"),
                  category = 208,
                  #hl = hl_query,
                  low_search_volume = FALSE,
                  cookie_url = "http://trends.google.com/Cookies/NID",
                  tz = 0,
                  onlyInterest = TRUE)

query_gt <- data.frame(google_trends$interest_over_time)

# Save in .RData
file_google_trends <- paste("google_trends_", geo_query, ".RData", sep = "")
setwd(here("3_rdata/1_google_trends/"))
save(query_gt, file = file_google_trends)
```

The query results data was obtained for each country, and it contains 7 attributes, as described in Table 2.5, with 108 observations per country.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Table 2.5 Hits attributes

Atribute Name	Type	Description
keyword	chr	Keywords used
hits	num	Total number of hits
geo	chr	Country name acronym
date	mth	Date (monthly)
time	chr	Query date range
gprop	chr	Product type (GT)
category	int	Category used

2.3 Exploratory data analysis

The guests and hits variables were selected from both datasets, and an exploratory data analysis was made for each one by performing a time series plot, a seasonality plot, an ACF plot, a PACF plot and a features table.

The time series and seasonality plots were made using the plot_ly() function from the plotly package (Sievert, 2020).

The ACF and PACF plots were made using the ts_cor () function from the plotly TSstudio package (Krispin, 2020). These plots were made to analyze if the time serie is stationary or not. In a stationary time series, the ACF graph will drop to zero relatively quickly, while in a non-stationary time series the ACF decreases slowly (Athanasopoulos & Hyndman, 2021). Furthermore, a Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was performed using the features() function from the feasts package (O'Hara-Wild et al., 2021).

The features table was made using the features() function from the feasts package. Some of the features are: Seasonal Strength Year which indicates the timing of the peaks, Seasonal Peak Year which indicates the timing of the troughs, Spikiness which measures the prevalence of spikes in the remainder component R_t of the STL decomposition, and Linearity which measures the linearity of the trend component of the STL decomposition.

Since GT results do not provide the exact number of searches and the data is normalized to the maximum value over the entire period, an additional analysis was performed in the

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

'hits' attribute to identify outliers by using the plot_anomaly_diagnostics() function from the timetk package (Dancho & Vaughan, 2021).

2.4 Pre-processing

Considering that all analyzed countries presented at least one value indicated as a possible outlier in the time series, two additional attributes were created: anomaly and hits_out_rmv. The former indicates which observations from the 'hits' variable were detected as outliers. The latter is a copy of the original hits time series but with replaced values for the outlier. These values were calculated by using the Kalman Smoothing method implemented in the na_kalman() function from the imputeTS package (Moritz & Bartz-Beielstein, 2017).

The guests and hits with outliers removed variables were decomposed through the X11 and STL methods. The decomposition process was performed using the model() and components() functions from the forecast package (Hyndman et al., 2021).

Considering the possibility of the hits variable presenting irregular variations that could mask the trend and seasonality and affect the forecasting process, a new variable called hits_irr_rmv was created by removing the irregular values calculated during the X11 decomposition process from the hits_out_rmv values. Therefore, there are three variations of the hits variable:

- Original;
- With outlier removed;
- With outliers and irregular components removed.

Considering that the datasets were extracted from different sources and have different formats, a new and unique dataset called query was created using the full_join() function from the dplyr (Wickham et al., 2021) package with data from different datasets.

The query data contains 14 attributes, as described in Table 2.6, with 108 observations per country.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Table 2.6 Query attributes

Atribute Name	Type	Description
date	mth	Date (monthly)
keyword	chr	Keywords used
geo	chr	Country name acronym
time	chr	Query date range
gprop	chr	Product type (GT)
.model	chr	Decomposition type used
irregular	num	Irregular component
category	int	Category used
guests	num	Total number of guests
hits	num	Total number of hits
hits_out_rmv	num	Hits with outliers removed
anomaly	chr	Indicative of anomaly
hits_irr_rmv	num	Hits with outliers and irregular component removed

The correlation coefficients between the guests and the three variations of the hits variables were calculated considering different lag values between 0 and 12, as shown in the Appendices 9, 10 and 11. This made possible to identify in which lag value the correlation is stronger, that is, how long before traveling tourists from each country searched for Portuguese tourist destinations.

The lag process has been carried out by using the lag() function from the dplyr package and the correlation coefficient was calculated using the cor() function from the stats package (R Core Team, 2021).

2.5 Forecasting methods

Given the growing importance and complexity of forecasts in the most diverse contexts, choosing the forecasting methods that best fit the given dataset can become a major challenge.

Considering all the aspects that were analyzed during the literature review of forecasting methods, three model classes were selected to be used during the forecasting process: exponential smoothing, ARIMA and dynamic regression. Regarding the experimental setup, two procedures were used:

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

- Time series split into training data (80%) and test data (20%);
- Time series cross-validation with initial value in 67 and step value of 1.

A group of 5 models were applied, as shown in Table 2.11.

Table 2.7 Models

Model	Class	Data Splitting	Description
Model 1	ETS	Split Cross-Validation	ETS model using only the guests variable for each country
Model 2	ARIMA	Split Cross-Validation	Auto ARIMA model using only the guests variable for each country
Model 3	Dynamic Regression	Cross-Validation	Auto ARIMA model using guests and lagged original hits variables where the correlation is the highest for each country
Model 4	Dynamic Regression	Cross-Validation	Auto ARIMA model using guests and lagged hits with outliers removed variables where the correlation is the highest for each country
Model 5	Dynamic Regression	Cross-Validation	Auto ARIMA model using guests and lagged hits with outliers and irregular component removed variables where the correlation is the highest for each country

The time series split was implemented using the window() function from stats package. The time series cross-validation was performed through the stretch_tsibble() function from tsibble package (Wang et al., 2020).

The parameters for the ETS class models were estimated using the ets() function. The parameters for the ARIMA class models were estimated using the auto.arima() function. The dynamic regression class models were estimated using the auto.arima() function with one additional argument called the x-reg argument. The point forecast was obtained using the forecast() function and the accuracy measures were calculated using accuracy() function. All these functions are included in the forecast package.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

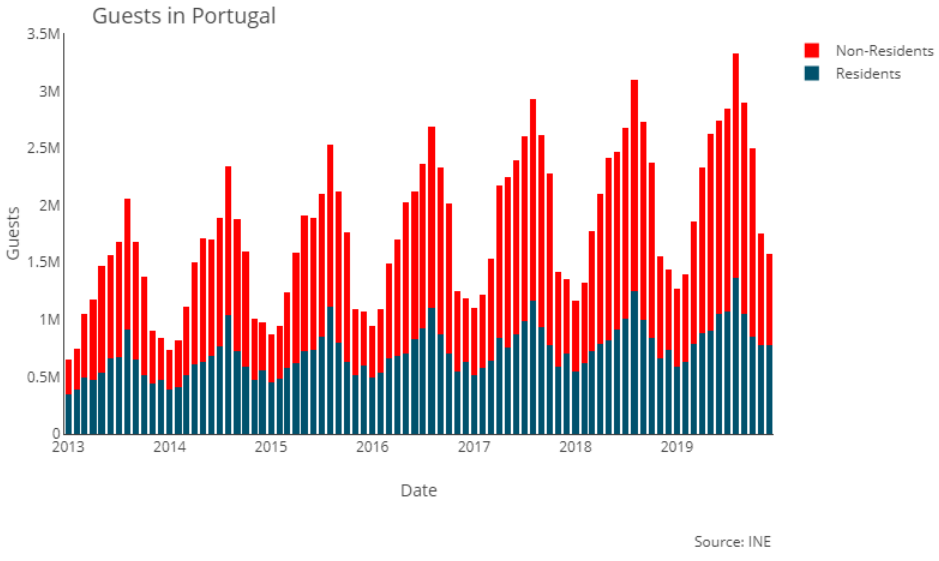
3 RESULTS AND DISCUSSION

This chapter consists of an overview of the guests in Portugal, the presentation of results for some of the analysed countries and discussions. Section 1 contains an overview of the total number of guests in tourism accommodations in Portugal. Section 2 and 3 present the main results obtained through descriptive and predictive analyses of guest data from Portugal and Brazil, respectively. Section 4 contains the main discussion points.

3.1 Overview of Guests in Portugal

Through an overview of the total number of guests in tourism accommodations in Portugal, it was possible to observe that, although the number of resident guests has been expressive, the number of non-resident guests has shown a greater growth rate over the last few years. In some years it even exceeded the number of resident guests during the summer periods.

Figure 3.1 Overview residents and non-residents guests

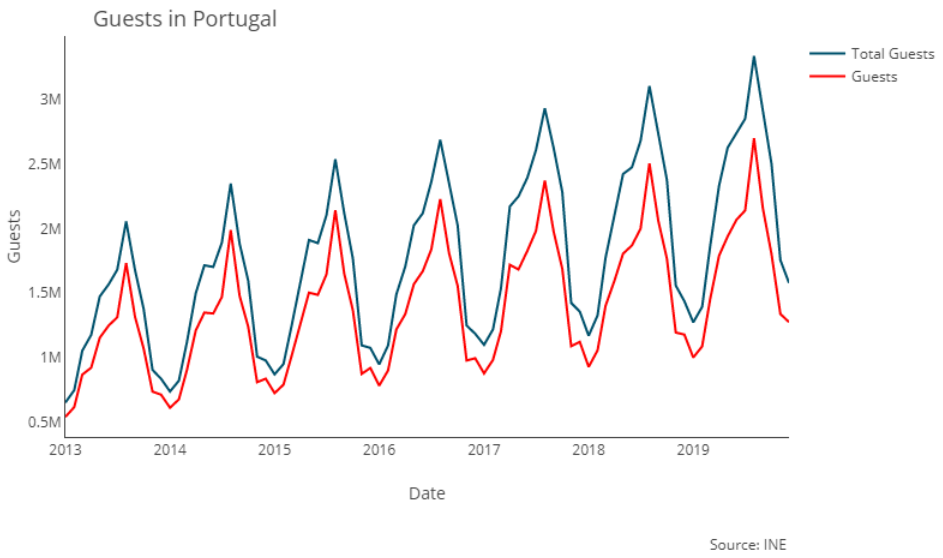


In view of this, in addition to Portuguese guests, the guests from the 7 main countries of origin of tourists in Portugal were also analyzed in this study, namely: Great Britain, Ireland, Spain, France, Germany, Brazil, and Italy.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The number of guests from Portugal and from these countries corresponds, on average, to 77% of the total number of guests in Portugal over the last few years. Figure 3.2 shows the total number of guests compared to the number of guests from the countries selected for this study.

Figure 3.2 Overview residents and non-residents guests



3.2 Portugal

Through the analysis of the Portuguese guest and hits variables, shown in Figure 3.3 and 3.4, respectively, it is possible to observe a growing trend up until 2019.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.3 Portugal Guests Time Series

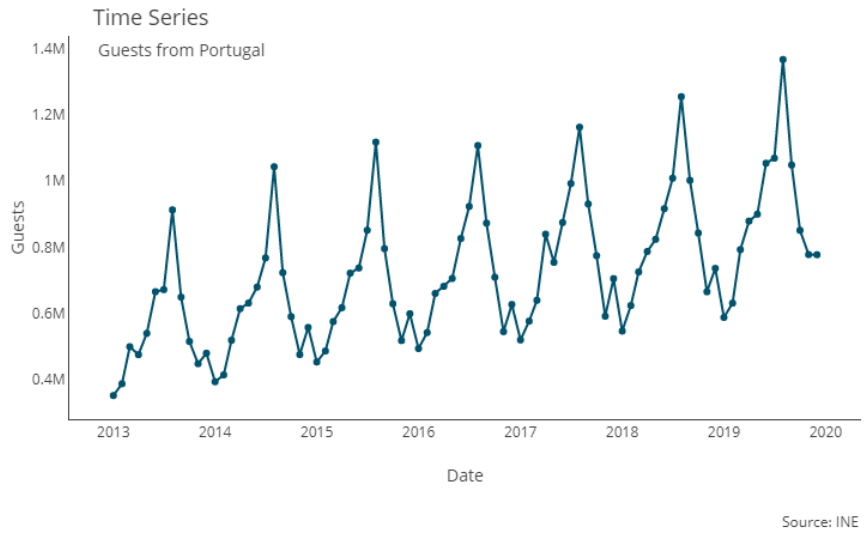
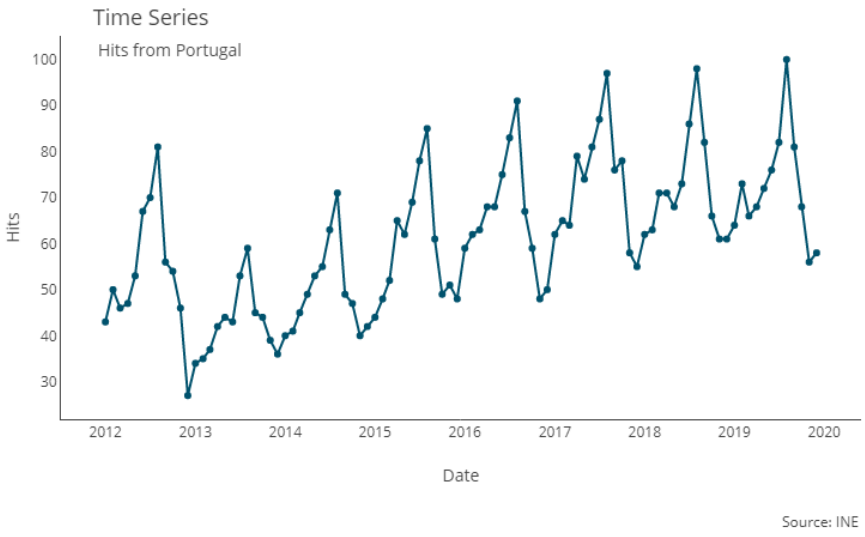


Figure 3.4 Portugal Hits Time Series



It is possible to confirm a high seasonality in both variables, with its peak in August, by performing the seasonality plot, as shown in Figure 3.5 and 3.6, respectively.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.5 Portugal Guests Seasonality

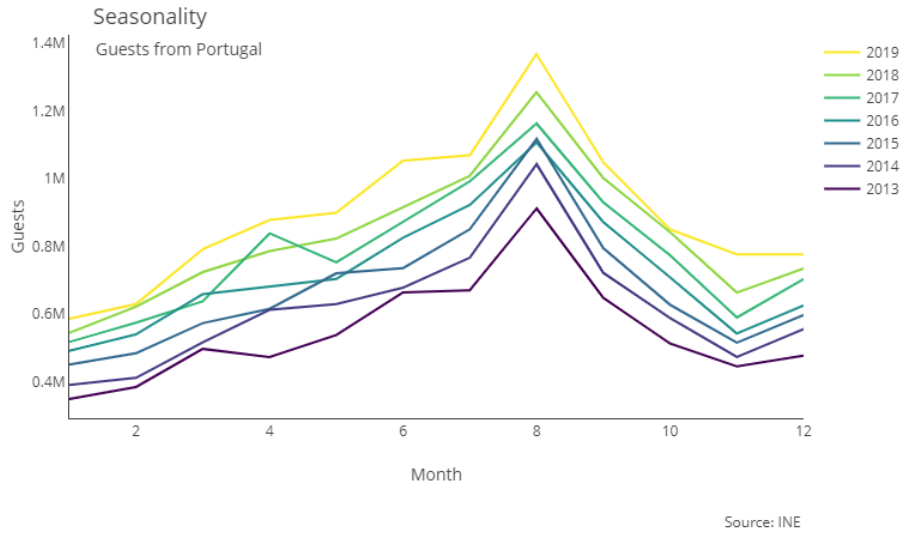
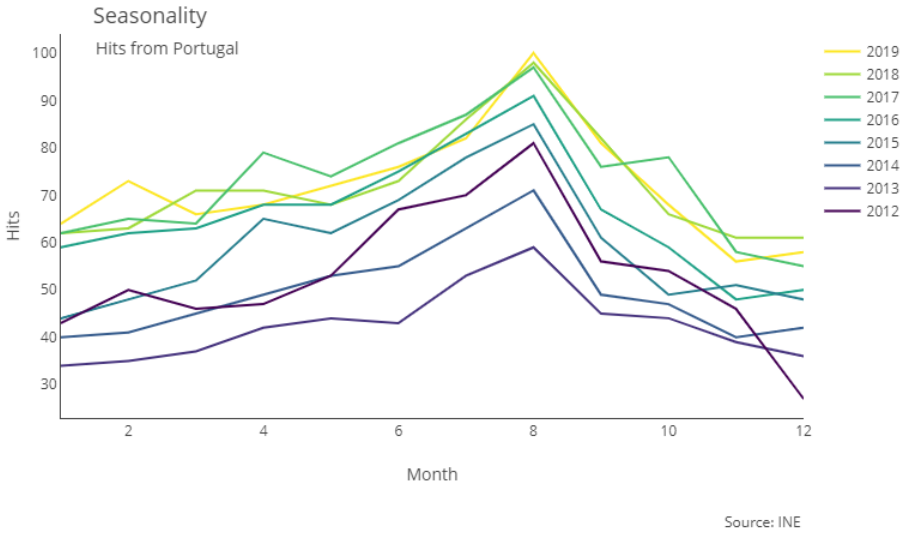


Figure 3.6 Portugal Hits Seasonality

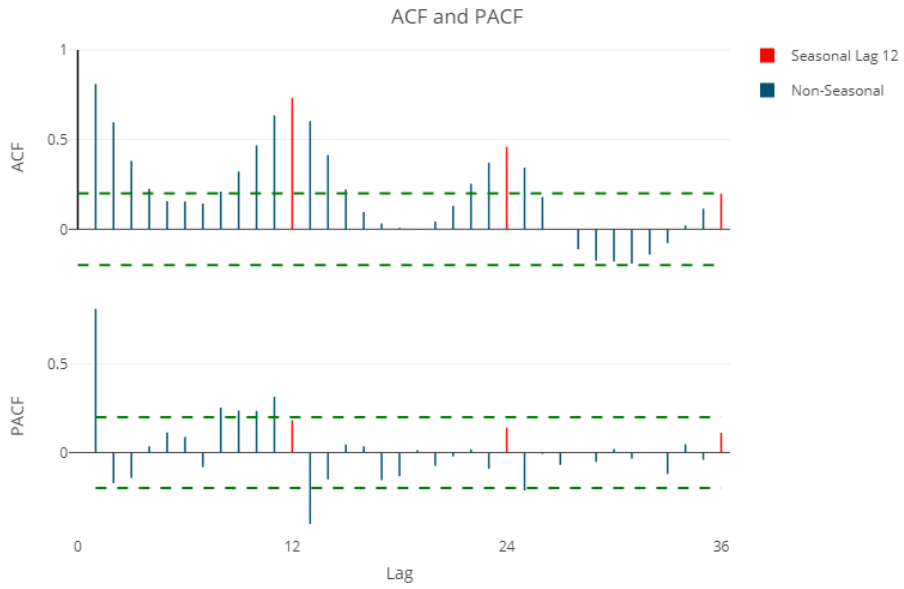


By analyzing the ACF and PACF plots, as shown in Figure 3.7 and 3.8, it is possible to identify that both time series present observations that are considered statistically correlated, thus they are autocorrelated.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.7 Portugal Guests ACF and PACF

Figure 3.8 Portugal Hits ACF and PACF



Considering the KPSS Stat (0.97) and KPSS P (0.01) values, the null hypothesis, which implies that the guest time series is stationary, was rejected.

Considering the KPSS Stat (0.97) and KPSS P (0.01) values, the null hypothesis, which implies that the hits time series is stationary, was rejected.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

After performing an outlier detection analysis, no outlier was detected for the hits time series. The guests and hits time series decompositions can be seen in Figure 3.9 and 3.10, respectively.

Figure 3.9 Portugal Guests Decomposition

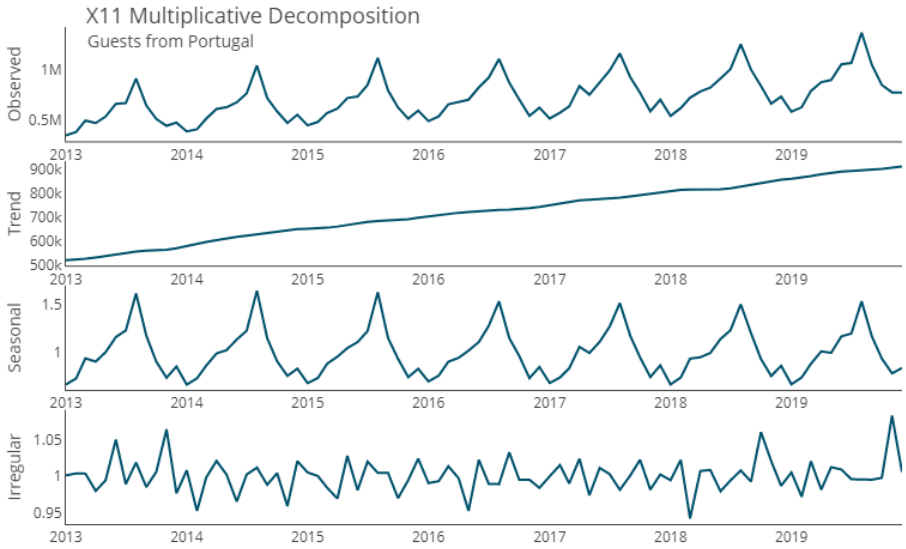
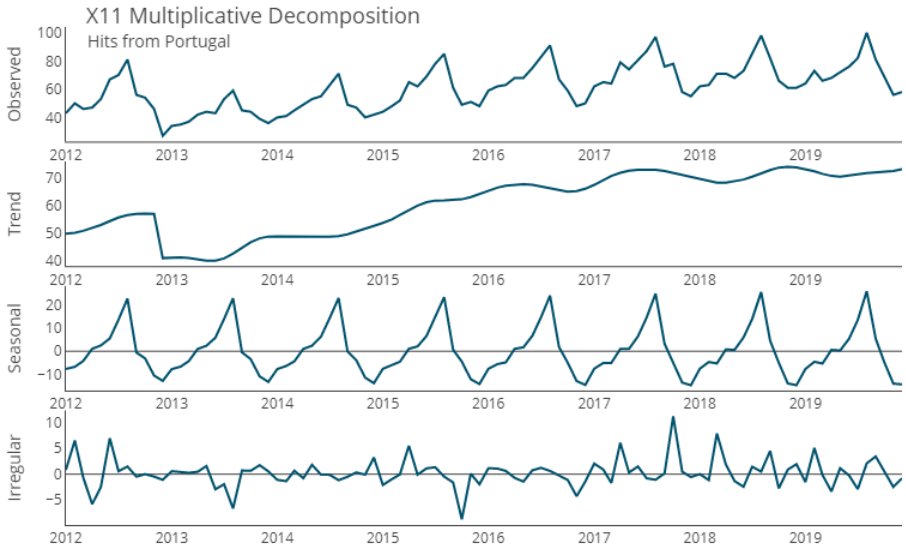


Figure 3.10 Portugal Hits Decomposition



Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The correlation coefficient between the guest and hits variables from Portugal suggest that the stronger correlation occurs when the lag considered is 0.

As mentioned in the methodology chapter, the predictive analysis consists of comparing 5 different models. The models 1 and 2 used the guest time series, which was split in training (80%) and test (20%) data. The models 3, 4 and 5 were made using guests and hits through using the cross-validation method. The accuracy models are shown in Table 3.1.

Table 3.1 Portugal Accuracy Models

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Portugal	Model 1	Test	Split	17440.9	44941.23	34908.68	3.84
Portugal	Model 1	Test	CV	11382.56	41803.78	30938.11	3.38
Portugal	Model 2	Test	Split	12943.23	40226.25	32340.35	3.68
Portugal	Model 2	Test	CV	9197.21	37893.06	30951.88	3.58
Portugal	Model 3	Test	CV	19073	46065.78	42159.46	4.87
Portugal	Model 4	Test	CV	19073	46065.78	42159.46	4.87
Portugal	Model 5	Test	CV	7429.03	40759.03	36479.7	4.2

In the Split setup, Model 2 is the one that presents the lowest RMSE and therefore was the one chosen to be used to perform forecasts for 2020. Its parameters are shown in Figure 3.11.

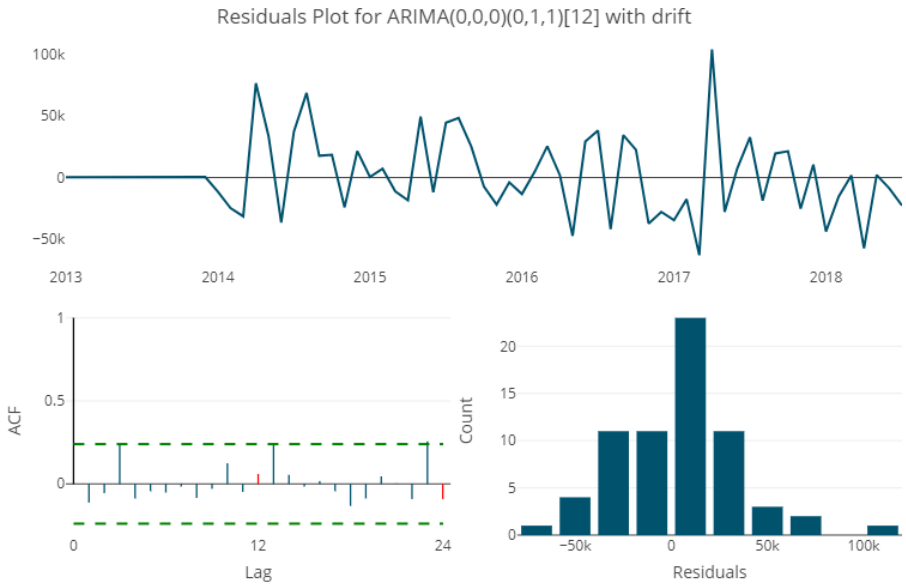
Figure 3.11 Portugal Best Model Summary

```
## Series: train_guests
## ARIMA(0,0,0)(0,1,1)[12] with drift
##
## Coefficients:
##          sma1      drift
##       -0.4726  4540.3866
## s.e.   0.1659   255.1511
##
## sigma^2 estimated as 1.209e+09:  log likelihood=-653.64
## AIC=1313.28   AICc=1313.75   BIC=1319.3
##
## Training set error measures:
##                    ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 1569.572 30922.52 22709.44 -0.1294311 3.249074 0.3893519
##                    ACF1
## Training set -0.1117644
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The residuals plot is shown in Figure 3.12.

Figure 3.12 Portugal Best Model Residuals



3.3 Brazil

Through the analysis of the Brazilian guests and the hits variables, shown in Figure 3.13 and 3.14, respectively, it is possible to observe a growing trend up until 2019.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.13 Brazil Guests Time Series

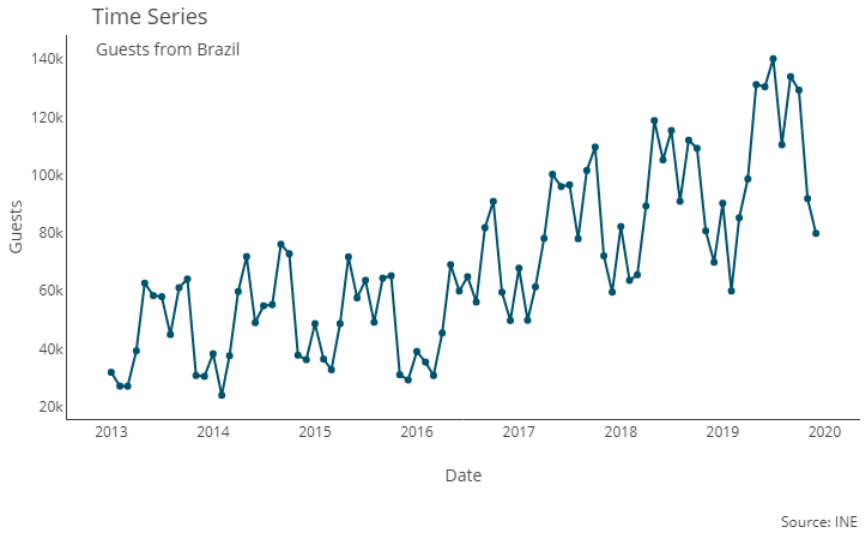
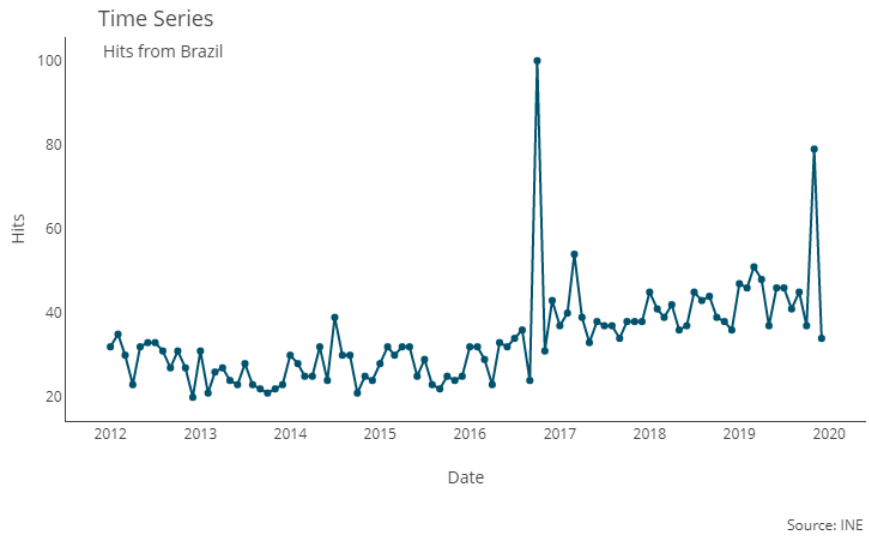


Figure 3.14 Brazil Hits Time Series



Regarding the guests variable, it is possible to confirm a strong seasonality in the months of May, July and October by performing the seasonality plot, as shown in Figure 3.15. Unlike the guests variable, hits variable does not present such strong seasonality, as shown in Figure 3.16.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.15 Brazil Guests Seasonality

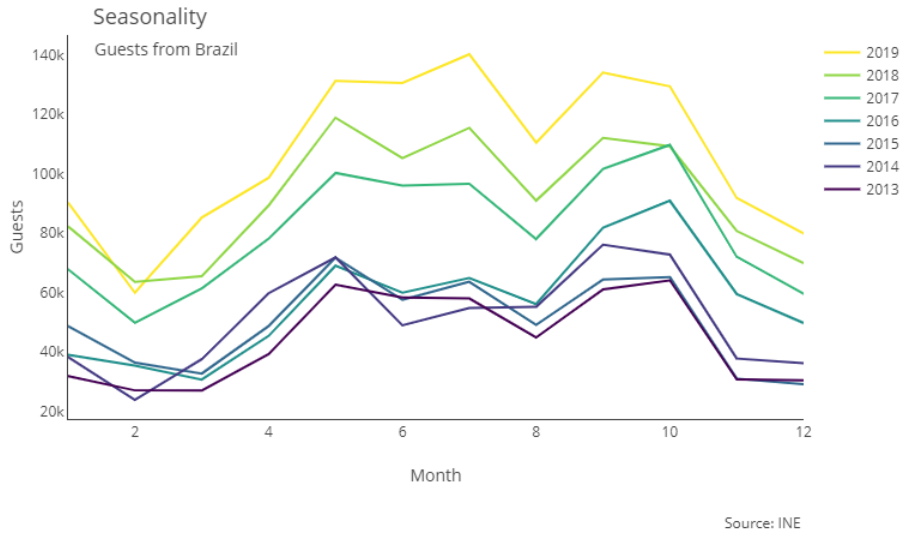
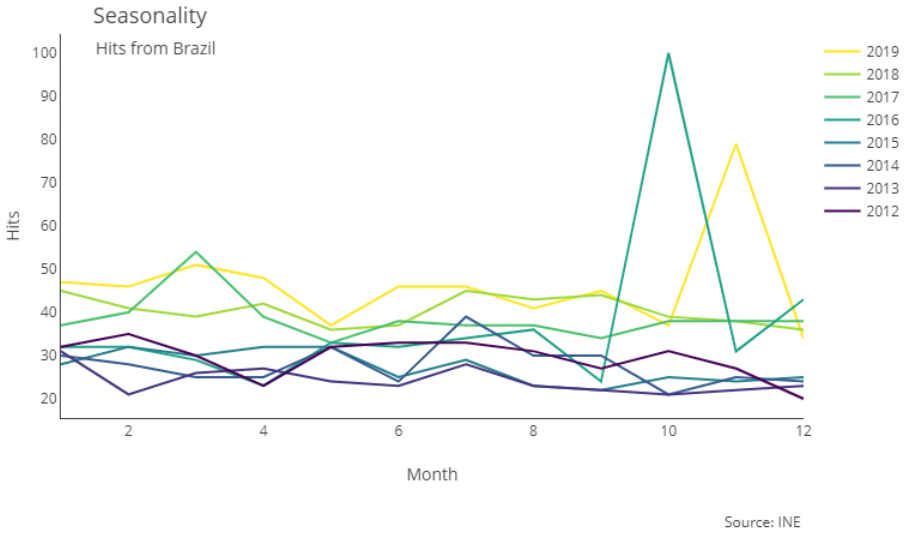


Figure 3.16 Brazil Hits Seasonality



By analyzing the ACF and PACF plots, as shown in Figure 3.17 and 3.18, it is possible to identify that both time series present observations that are considered statistically correlated.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.17 Brazil Guests ACF and PACF

Figure 3.18 Brazil Hits ACF and PACF

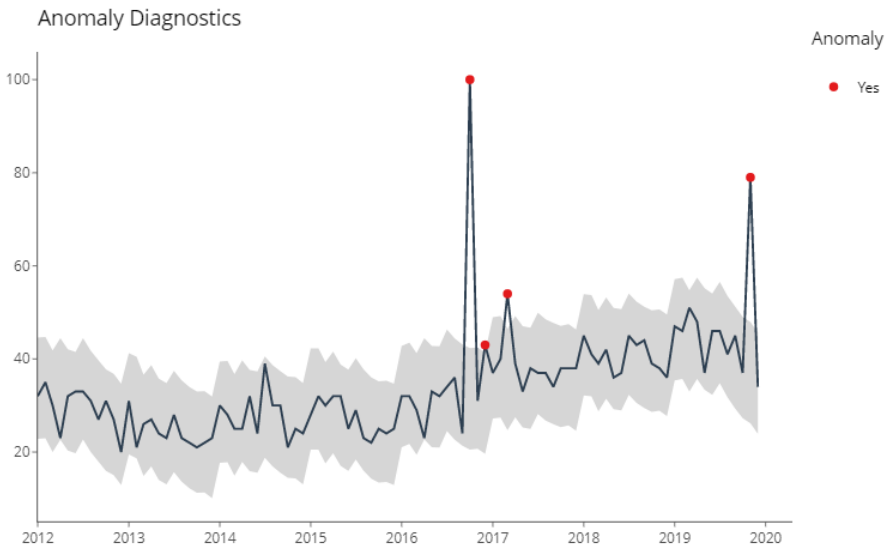
Considering the KPSS Stat (1.60) and KPSS P (0.01) values, the null hypothesis, which implies that the guest time series is stationary, was rejected.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Considering the KPSS Stat (1.70) and KPSS P (0.01) values, the null hypothesis, which implies that the hits time series is stationary, was rejected.

After performing an outlier detection analysis, four outlier observations were detected for the hits time series, as shown in Figure 3.19. These values were replaced through method mentioned in the methodology.

Figure 3.19 Brazil Hits Outlier



The guests and hits time series decompositions can be seen in Figure 3.20 and 3.21, respectively.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.20 Brazil Guests Decomposition

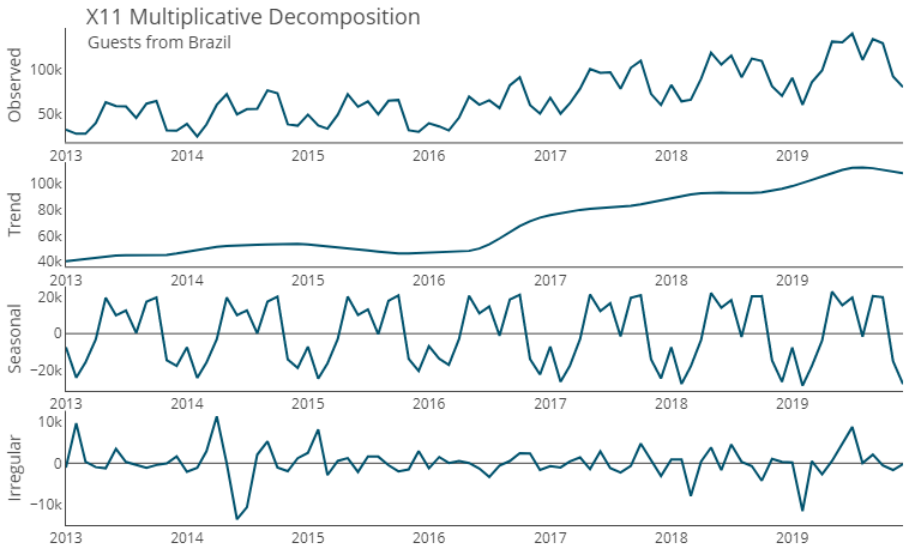
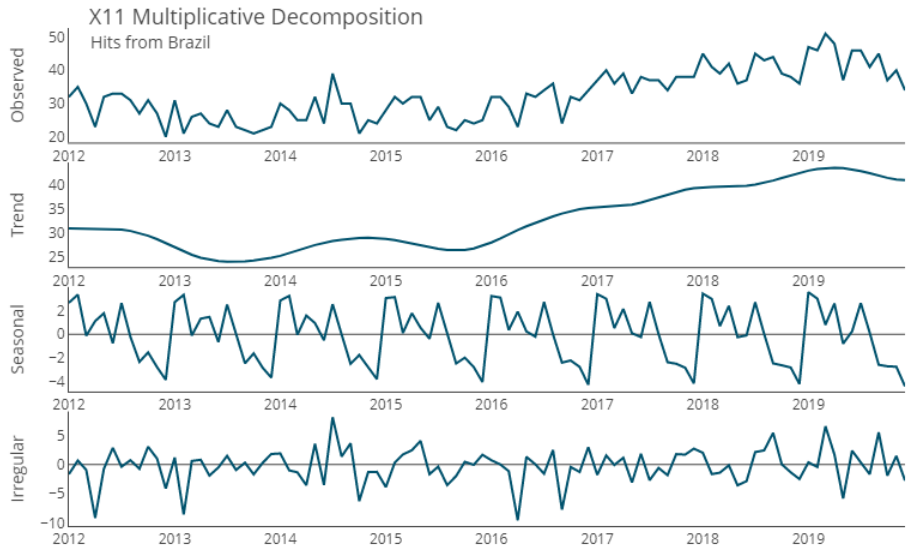


Figure 3.21 Brazil Hits Decomposition



The correlation coefficient between the guest and hits variables from Brazil suggest that the stronger correlation occurs when a lag value of 3 is considered.

The accuracy models are shown in Table 3.2.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Table 3.2 Brazil Accuracy Models

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Brazil	Model 1	Test	Split	-5123.05	12133.79	10329.87	11.45
Brazil	Model 1	Test	CV	-374.33	9972.38	8692.04	8.96
Brazil	Model 2	Test	Split	-11342.5	14895.66	13374.18	15.11
Brazil	Model 2	Test	CV	-167.93	9121.12	7090.06	7.65
Brazil	Model 3	Test	CV	-2131.4	12981.88	10484.87	11.61
Brazil	Model 4	Test	CV	699.76	8823.88	7078.27	7.72
Brazil	Model 5	Test	CV	1037.28	7869.15	6161.52	6.84

In the Split setup, model 1 is the one that presents the lowest RMSE and is therefore the one chosen to perform forecasts for 2020. Its parameters are shown in Figure 3.22.

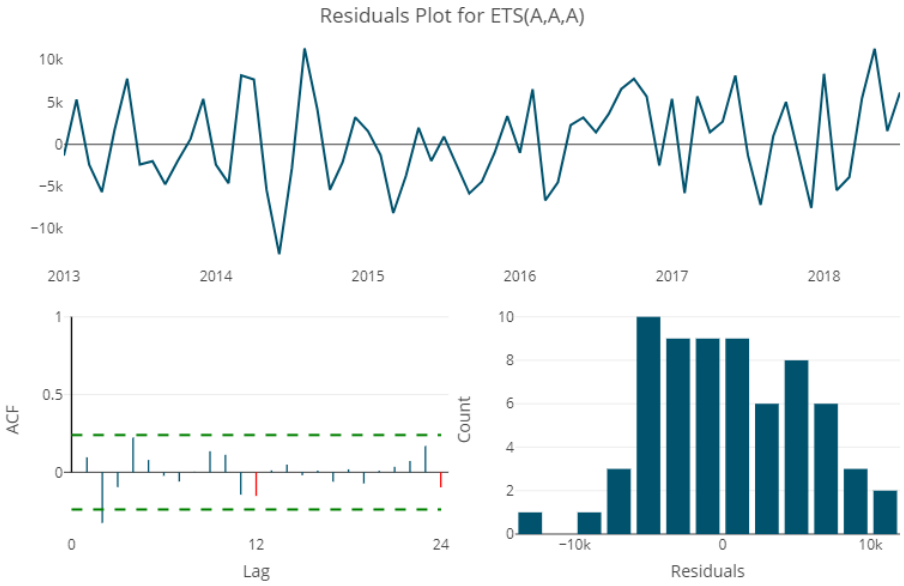
Figure 3.22 Brazil Best Model Summary

```
## ETS(M,A,M)
##
## Call:
##  ets(y = train_guests)
##
##   Smoothing parameters:
##     alpha = 0.7839
##     beta  = 0.0064
##     gamma = 1e-04
##
##   Initial states:
##     l = 111217.0223
##     b = 1239.5353
##     s = 0.3617 0.5201 1.3143 1.5377 1.3226 1.3926
##           1.4627 1.4003 1.0469 0.7353 0.5121 0.3938
##
##   sigma:  0.032
##
##      AIC     AICc      BIC
## 1415.826 1428.316 1453.306
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -742.0545 5181.376 3576.552 -0.367059 2.265939 0.2670718
##                    ACF1
## Training set 0.03954793
```

The residuals plot is shown in Figure 3.23.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.23 Brazil Best Model Residuals



Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

3.4 Discussion

The guests time series from all countries show an increasing trend and seasonality until 2019. It is possible to observe that the high seasonality occurs just before, during, or soon after the summer period, however the countries present seasonality in different months. The United Kingdom and Ireland, for example, have similar behavior with high seasonality in the months of May, June, and September. Spain, Portugal, and Italy have extremely strong and accentuated seasonality in August. France presents seasonality in the months of May and August while Germany's occurs in May and September. Brazil is the one that presents the most different seasonal behavior, corresponding to the months of May, July, and October.

The hits time series also show an increasing trend until 2019, however their seasonality is not as clear as observed in the guest series. Through the complementary analysis to detect outliers, it was possible to observe that practically all countries presented at least one value considered as outlier. Furthermore, through the analysis of the decomposition, it was possible to observe the presence of a significant irregular component in the time series of some countries. After removing the irregular component from the hits variable and replacing the outliers, it was possible to observe that the correlation between the guests and hits variables improved for practically all the analyzed countries.

Regarding value for the lag that corresponds to the strongest correlations, it varied between 0 to 3 among the analyzed countries. The correlation between guests and hits is the strongest for Portugal at lag 0, while for United Kingdom, France, Spain, and Italy it is at lag 1. Ireland and Germany present the strongest correlation at lag 2 and Brazil is the only one country where it happened with a lag value of 3.

When considering forecasts for a short horizon, all the model groups were tested through the cross-validation method with a forecast horizon of 1 month. On the other hand, when considering longer horizons, groups 1 and 2 were tested through the split method.

After performing a comparative analysis between the models that were cross-validated, it was possible to observe that the ARIMA class models were the ones that presented the lowest RMSE for seven of the analyzed countries, and consequently were the ones

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

considered as the best short horizon forecasting models. The countries were: United Kingdom, Ireland, Spain, France, Brazil, Portugal, and Italy.

Regarding the dynamic regression class models, only Brazil, Ireland and United Kingdom presented superior results compared to the models without the hits variable. Additionally, the models for these three countries also presented better results when the outliers and irregular components were removed.

Germany was the only country that presented the best results when using an ETS class model.

Regarding the comparative analysis for the models that were tested through the split validation, the ARIMA class models also presented the best results for the majority of the countries: Ireland, Spain, France, Portugal and Italy. The ETS class models presented the bests results for United Kingdom, Germany, and Brazil.

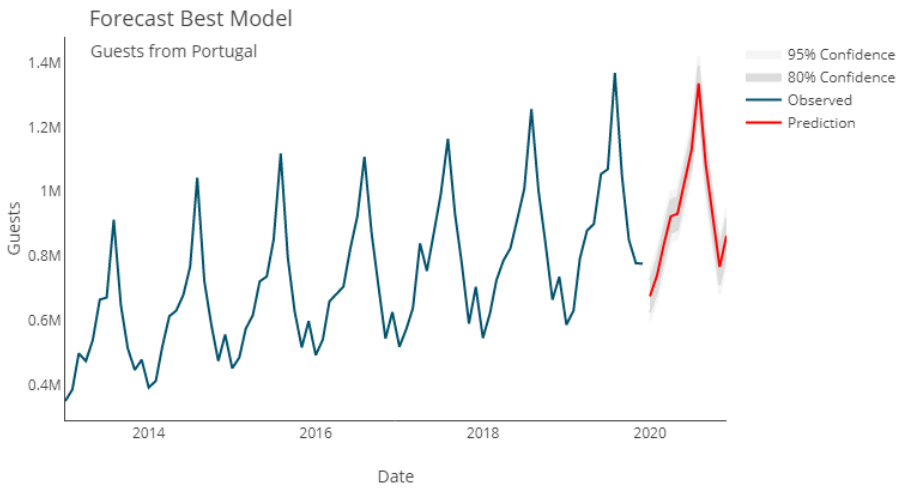
After identifying the best model that was tested with the split validation for each individual country, an additional guests forecast was performed for a 12 month horizon for the year of 2020. The aim was to estimate the losses on the tourism sector by comparing the actual number of guests with the expected values if the pandemic had not occurred.

Regarding the forecasts for guests for each country, it was possible to observe that the guests amount dropped in all the analyzed countries, with an average drop higher than 60%. An exception was Portugal, where the average drop was of 42%.

The forecast plot for Portugal can be seen in Figure 3.24.

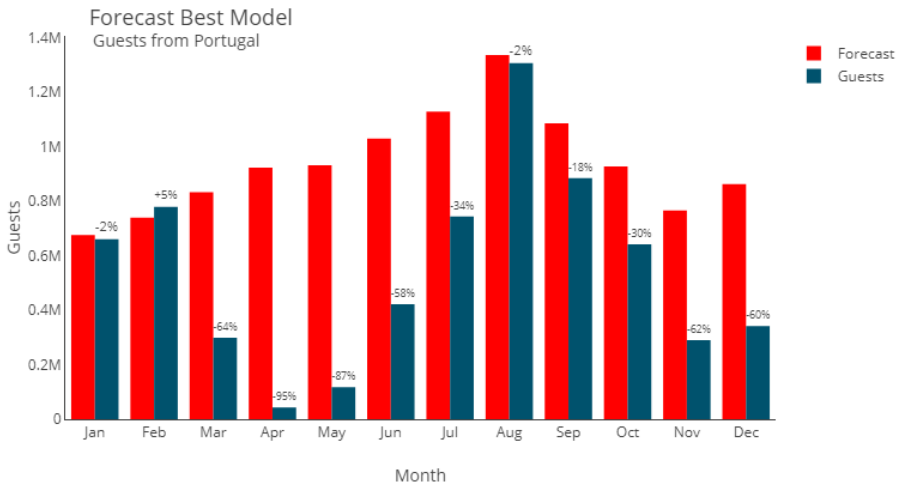
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Figure 3.24 Portugal Forecast



The comparative results between guest data and their forecasts are shown in Figure 3.25. It is possible to observe that the number of guests began to fall in March, however as of June this number grows again.

Figure 3.25 Portugal Forecast vs actual guests



Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The guests from Brazil were the most affected ones, with an average drop of 76%. The amount of guests dropped significantly throughout the year with few signs of recovery. The forecasted data and the comparison between it and the actual data are shown in Figure 3.26 and 3.27, respectively.

Figure 3.26 Brazil Forecast

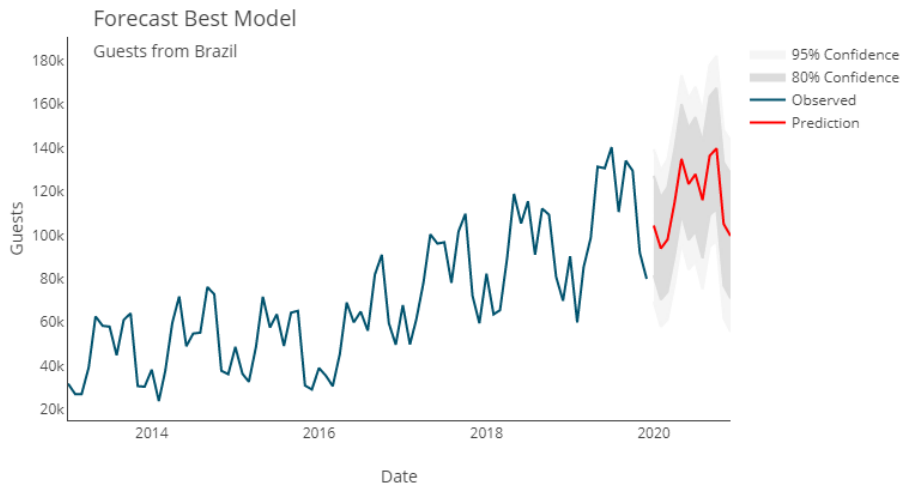
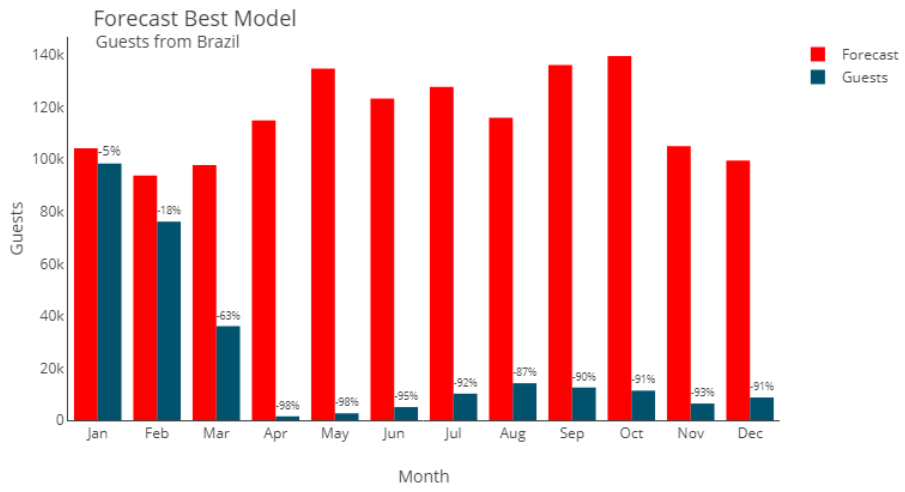


Figure 3.27 Brazil Forecast vs actual guests



Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

CONCLUSION

By analyzing the descriptive analyses results it was possible to conclude that the guests series is much more regular than the hit series. It was possible to observe an increasing trend on the guests series over the last few years until 2019 with high seasonality. Although the hits time series also showed an increasing trend, it was not possible to observe the seasonality as clearly as observed in the guests time series.

Regarding the correlation coefficients, it was observed that the correlation became stronger depending on the lag value applied to the hits variable and it varied between 0 to 3 among the analyzed countries. The correlation also became stronger for most countries after processing the hits data in order to remove the irregular component and to replace the outlier values. It is possible to conclude that the best lag value increases with the distance between the analyzed country and Portugal .

In respect of the predictive analyses and the setup method, the model groups that were cross-validated with a 1-month forecast horizon presented superior performance when compared to the model groups that were tested through the split method with longer horizons.

Regarding the model classes, in both split and cross-validation setup, the most promising models with the best results were the ones from the ARIMA class. On these ARIMA class models, the insertion of the hits variable presented superior results in only 3 countries when compared to the models without the hits variable. Additionally, the models for these three countries also presented better results when the irregular component was removed and the outlier values were replaced.

Therefore, it is possible to state that, although there is a correlation between the data, the inclusion of the hits variable did not improve the results of forecasts for all countries.

Even though the short-term forecast achieved better results, additional guests forecasts were performed for a 12 month horizon for the year of 2020. though the comparison with the actual number of guests in the period, it was possible to conclude that even though the tourism sector had significant losses mainly regarding the low number of international tourists due to travel restrictions, it presented significant signs of recovery when considering the number of resident tourists during peak season.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

The main contributions of this study are the descriptive analyzes of the guests and hits time series over the last few years performed individually considering the guests country of origin.

Through the analyses of the guests time series, it was possible to observe in which months there were more guests from each country staying in tourist establishments in Portugal. Regarding the hits data, it was possible to observe that although it is not so regular, it is correlated with the guests data.

The correlation analysis between the data considering different lag values for the hits data made it possible to understand how long before the trip the guests of each country usually search on the internet about the main tourist destinations in Portugal.

Regarding the descriptive analyses, it was possible to identify the most promising forecasting models and experimental setup for each country. In addition, through the forecasts for the year 2020, it was possible to measure the number of guests for each country and assess the losses in the sector due to the COVID-19 pandemic.

Due to time constrains, it was not possible to test the split method for the dynamic regression model class for longer horizons. Additionally, it was not possible to perform the cross-validation method for longer horizons, which is a suggestion for future studies. Furthermore, another suggestion would be to perform descriptive and predictive analyses considering the individual Portuguese regions using more specific search terms related to tourism. Moreover, it would be possible to incorporate machine learning techniques to test other models and assess if it would improve their accuracy.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

REFERENCES

Adhikari, R., & Agrawal, R. (2013). *An Introductory Study on Time Series Modeling and Forecasting.*

Antolini, F., & Grassini, L. (2019). Foreign arrivals nowcasting in Italy with Google Trends data. *Springer Netherlands*, pp. 2385-2401.

Athanasopoulos, G., & Hyndman, R. (2021). *Forecasting: Principles and Practice.* O Texts.

Athanasopoulos, G., Hyndman, R., Song, H., & Wu, D. (2011). The tourism forecasting competition. *International Journal of Forecasting*, pp. 822-844.

Camilleri, M. (2018). *The Tourism Industry: An Overview. In Travel Marketing, Tourism Economics and the Airline Product.* Cham, Switzerland: Springer Nature.

Carrière-Swallow, Y., & Labbé, F. (2013). Nowcasting with Google trends in an emerging market. *Journal of Forecasting*, pp. 289-298.

Choi, H., & Varian, H. (2012). Predicting the Present with Google Trends. *The Economic Record*, pp. 2–9.

Dancho, M., & Vaughan , D. (2021). timetk: A Tool Kit for Working with Time Series in R. Retrieved from https://CRAN.R-project.org/package=timetk

Dinis, G., Costa, C., & Pacheco, O. (2017). Similarities and correlation between resident tourist overnights and Google Trends information in Portugal and its tourism regions. *Tourism & Management Studies*, pp. 15-22.

Dinis, M. (2016). *Indicadores do Comportamento online e tendências da procura turística.* Aveiro.

Fava, V. (2000). Análise de séries de tempo. Atlas.

Google. (2020). *Trends Help*. Retrieved Set 07, 2020, from Google: https://support.google.com/trends/?%20hl=ko#topic=6248052

Hu, M., Li, G., & Li, H. (2020). Forecasting tourism demand with multisource big data. *Annals of Tourism Research journal*.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Hyndman, R., & Koehler, A. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*.

Hyndman, R., Athanasopoulos , G., Bergmeir , C., Caceres , G., Chhay , L., O'Hara-Wild, M., . . . Yasmeen , F. (2021). Forecasting functions for time series and linear models. Retrieved from https://www.jstatsoft.org/article/view/v027i03

Hyndman, R., Koehler, A., Ord, J., & Snyder, R. (2008). *Forecasting with Exponential Smoothing.* Australia: Springer.

Instituto Nacional de Estatística. (2020). *Conta Satélite do Turismo.*

Instituto Nacional de Estatística. (2020). *Estatísticas do Turismo - 2019.*

Jun, S., Yoo, H., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting & Social Change*, pp. 69-87.

Kirchgässner, G., & Wolters, J. (2007). *Introduction to Modern Time Series Analysis.* Springer.

Latorre, M., & Cardoso, M. (2001). Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. *Revista Brasileira de Epidemiologia*.

Lawrence, K., & Klimberg, R. (2010). *Advances in Business and Management Forecasting: Forecasting Sales.* Emerald Group Publishing Limited.

Li, X., Pan, B., Law, R., & Huang, X. (2017). Forecasting tourism demand with composite search index. *Tourism Management, 59*, pp. 57-66.

Massicotte , P., & Eddelbuettel, D. (2021). gtrendsR: Perform and Display Google Trends Queries. Retrieved https://CRAN.R-project.org/package=gtrendsR

Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time Series Missing Value Imputation in R. Retrieved from https://doi.org/10.32614/RJ-2017-009

O'Hara-Wild, M., Hyndman, R., & Wang, E. (2021). feasts: Feature Extraction and Statistics for Time Series. Retrieved from https://CRAN.R-project.org/package=feasts

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Organization, W. T. (2008). *Glossary of Tourism Terms*. Retrieved 2021, from UNWTO: https://www.unwto.org/glossary-tourism-terms

Park, S., Lee, J., & Song, W. (2016, May 04). Short-term forecasting of Japanese tourist inflow to South Korea using Google trends data. *Journal of Travel & Tourism Marketing, 34*(3), pp. 357-368.

Peter, Ď., & Silvia, P. (2012). ARIMA vs. ARIMAX – which approach is better to analyze and forecast macroeconomic time series. *Academia*.

R Core Team. (2013). R: A Language and Environment for Statistical Computing.

R Core Team. (2021). R: A Language and Environment for Statistical Computing. Retrieved from https://www.R-project.org/

Santos, L. (2012). *Uso de modelos autoregressivos e gráficos de controle para monitorar volatilidade de ativos financeiros.* São Paulo.

Shmueli, G., & Lichtendahl Jr., K. (2016). *Practical Time Series Forecasting with R: A Hands-On Guide.* Axelrod Schnall Publishers.

Sievert, C. (2020). Interactive Web-Based Data Visualization with R, plotly, and shiny. Retrieved from https://plotly-r.com

Song, H., & Li, G. (2008). Tourism demand modelling and forecasting-A review of recent research. *Tourism Management, 29*(2), pp. 203-220.

Song, H., Qiu, R. T., & Park, J. (2019). A review of research on tourism demand forecasting. *Annals of Tourism Research, 75*, 338-362.

StatCounter. (2021). *Search Engine Market Share Worldwide - August 2021*. Retrieved 09 29, 2021, from StatCounter - GlobalStats: https://gs.statcounter.com/search-engine-market-share

UNWTO. (2008). *Handbook on Tourism Forecasting Methodologies.* Madrid: The World Tourism Organization.

Wang, E., Cook, D., & Hyndman, R. (2020). A new tidy data structure to support exploration and modeling of temporal data. *Journal of Computational and Graphical Statistics*.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Wang, E., Cook, D., & Hyndman, R. (2020). A new tidy data structure to support exploration and modeling of temporal data. Retrieved from https://doi.org/10.1080/10618600.2019.1695624

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A Grammar of Data Manipulation. Retrieved from https://CRAN.R-project.org/package=dplyr

World Economic Forum. (2018). *Travel & Tourism Competitiveness Report 2017*. World Economic Forum.

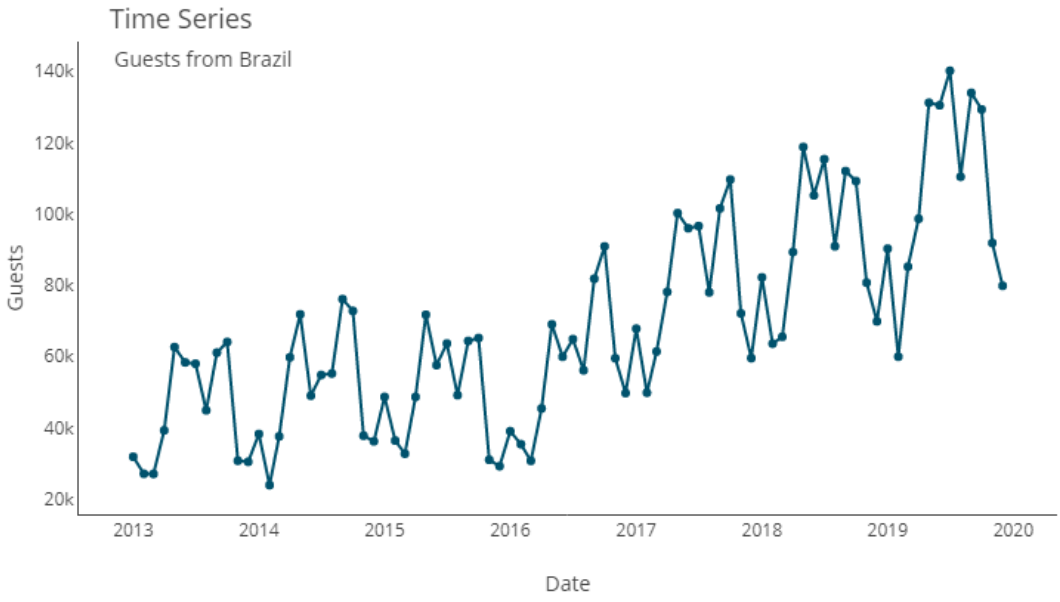
World Tourism Organization. (2020, Jan 20). *International Tourism Growth Continues to Outpace the Global Economy*. Retrieved Mar 03, 2021, from UNWTO: https://www.unwto.org/international-tourism-growth-continues-to-outpace-the-economy#:~:text=1.5%20billion%20international%20tourist%20arrivals,in%20view%20of%20current%20uncertainties.

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

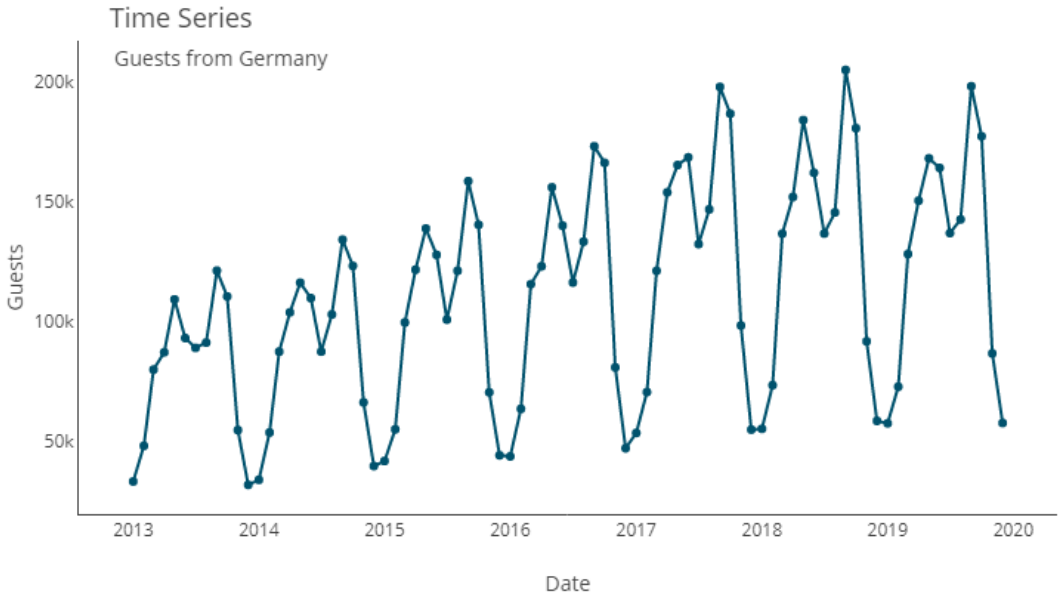
APPENDICES

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 1. GUESTS TIME SERIES PLOTS

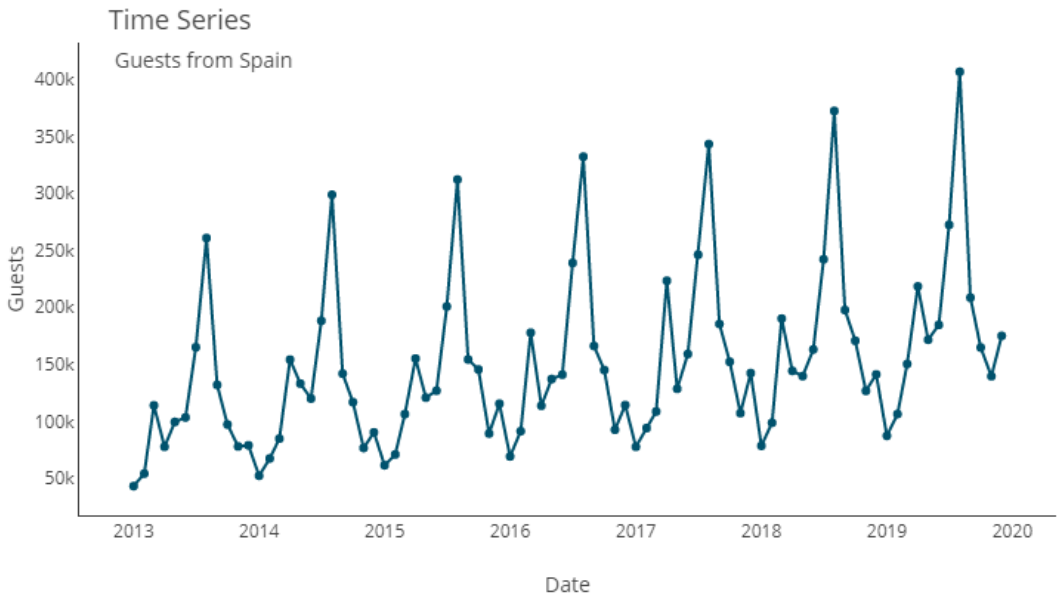


Source: INE

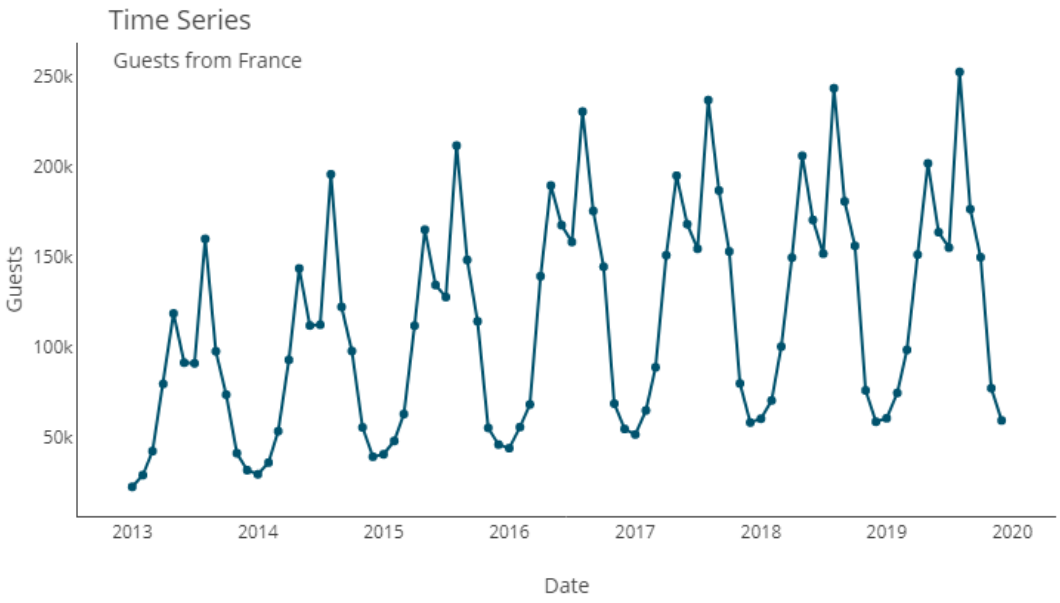


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

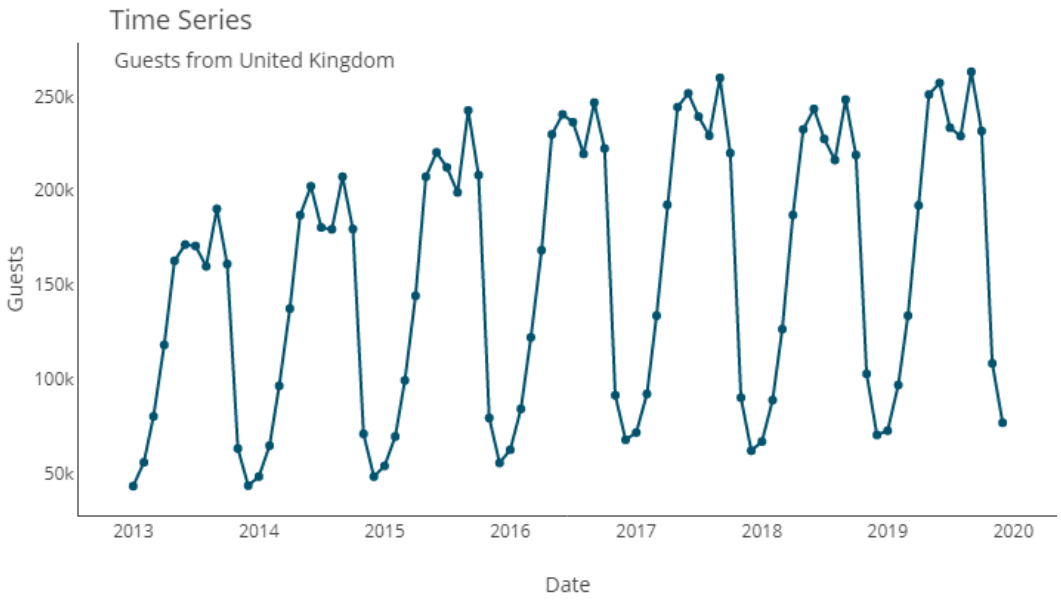


Source: INE

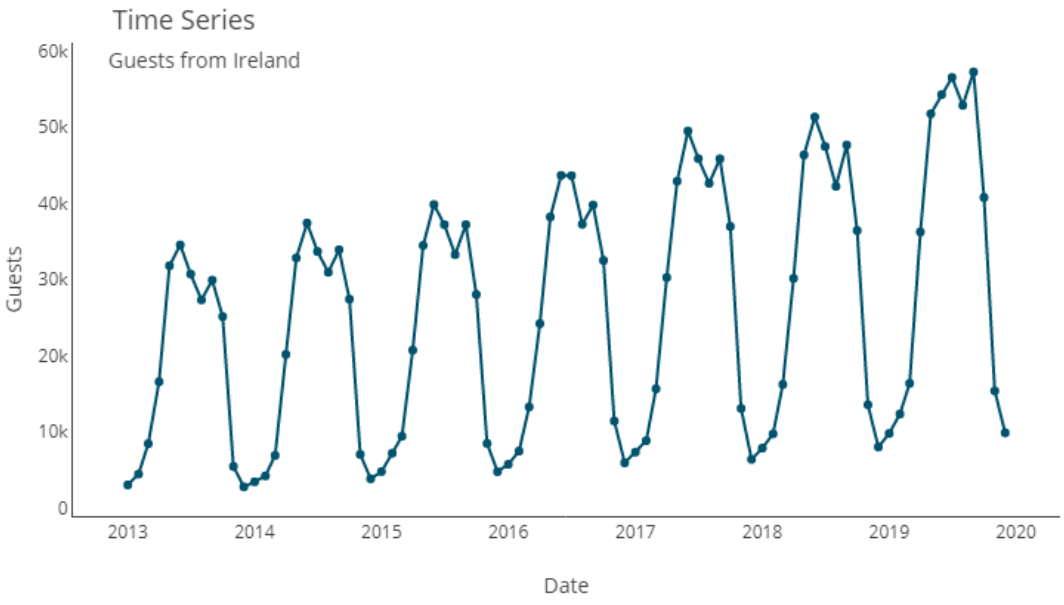


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

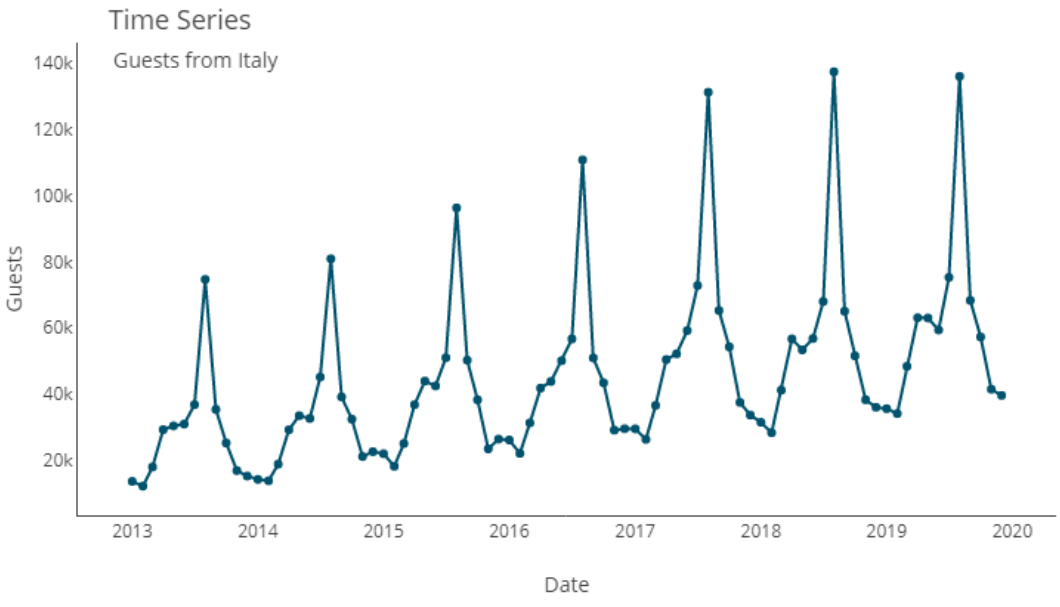


Source: INE

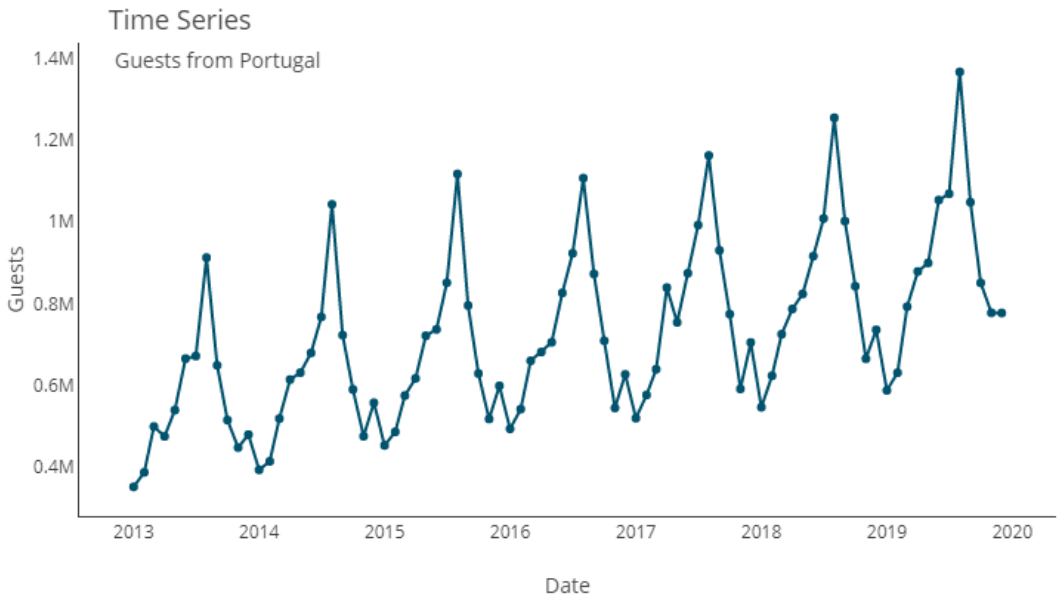


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



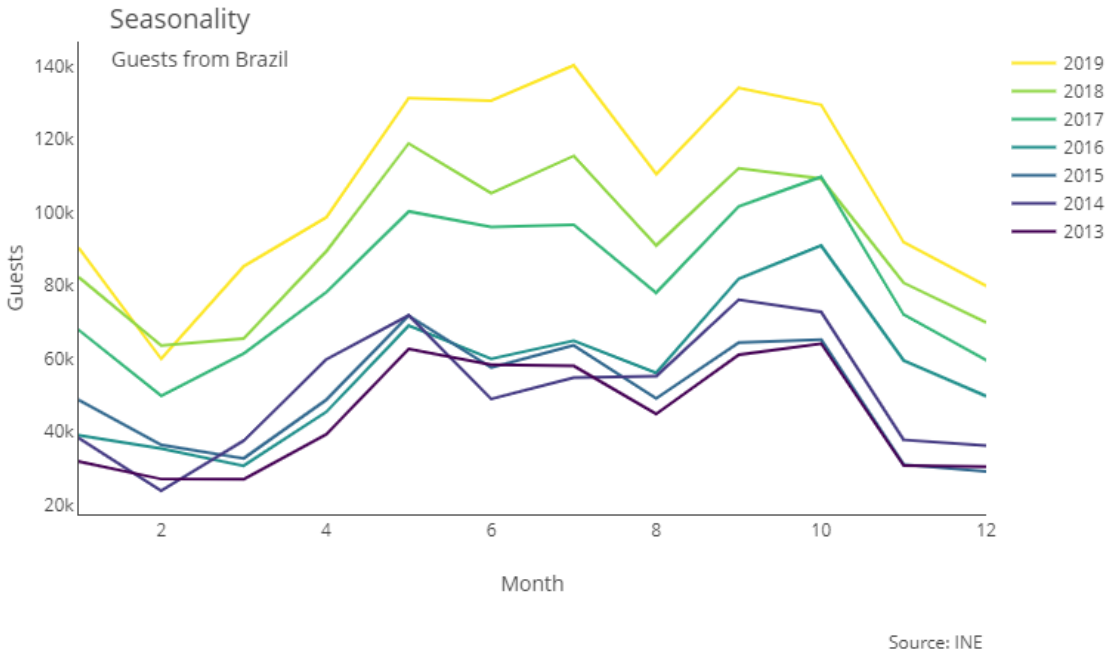
Source: INE



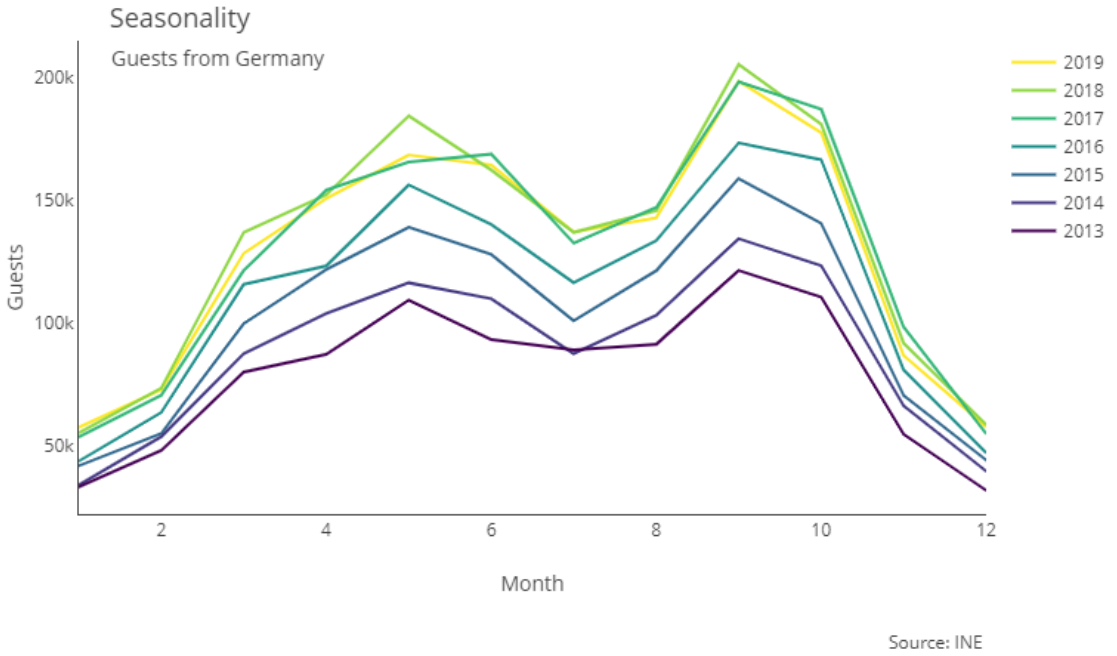
Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 2. GUESTS SEASONALITY PLOTS

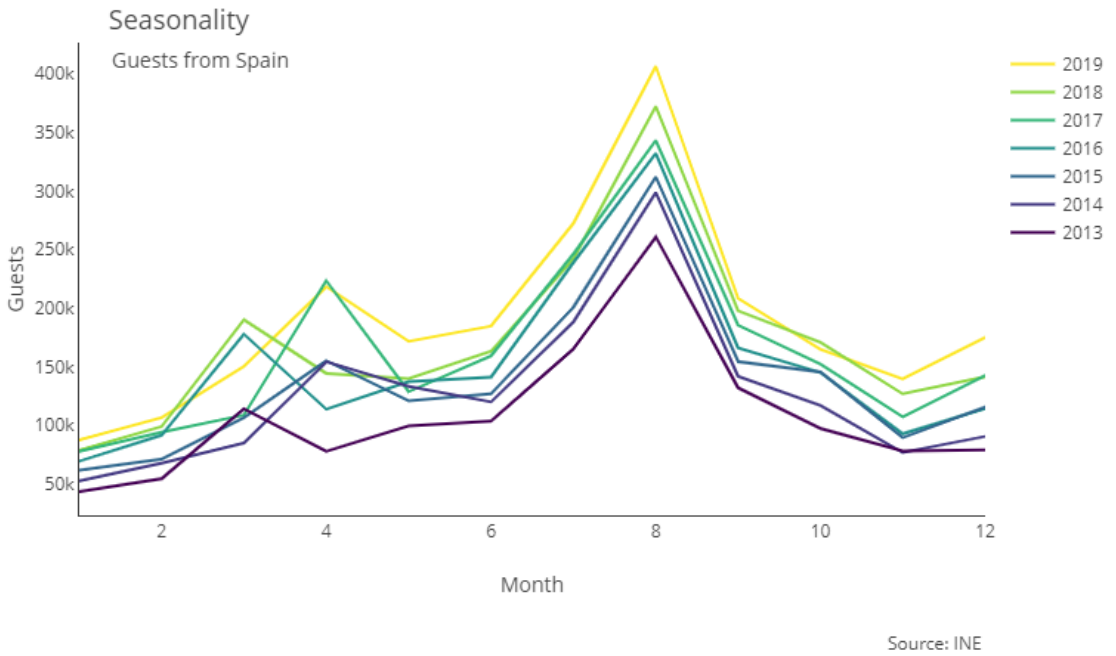


Source: INE

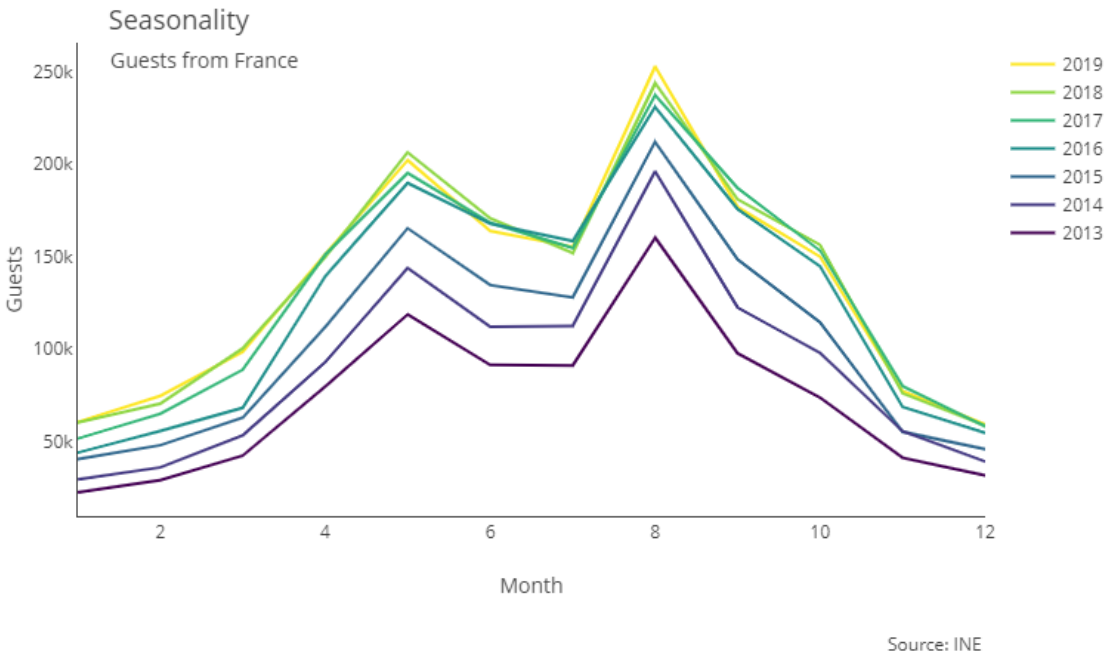


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

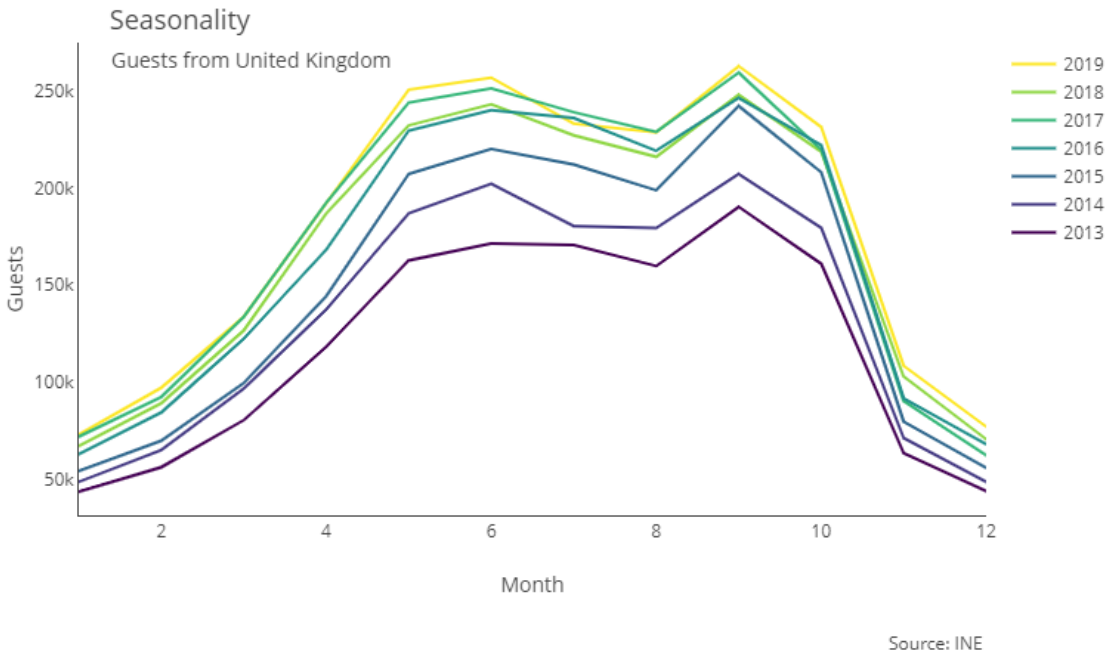


Source: INE

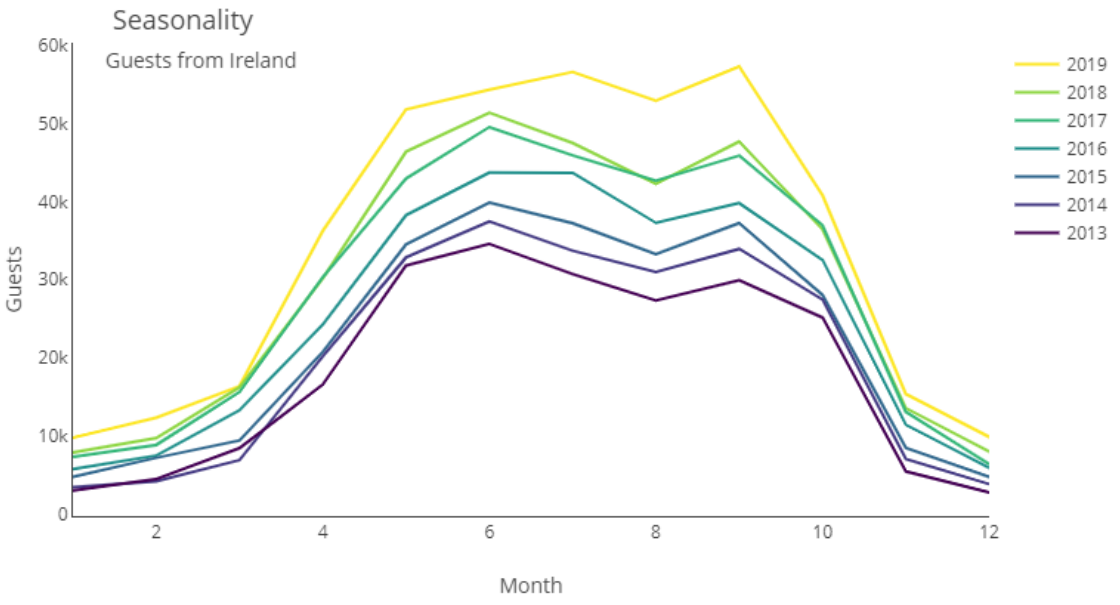


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

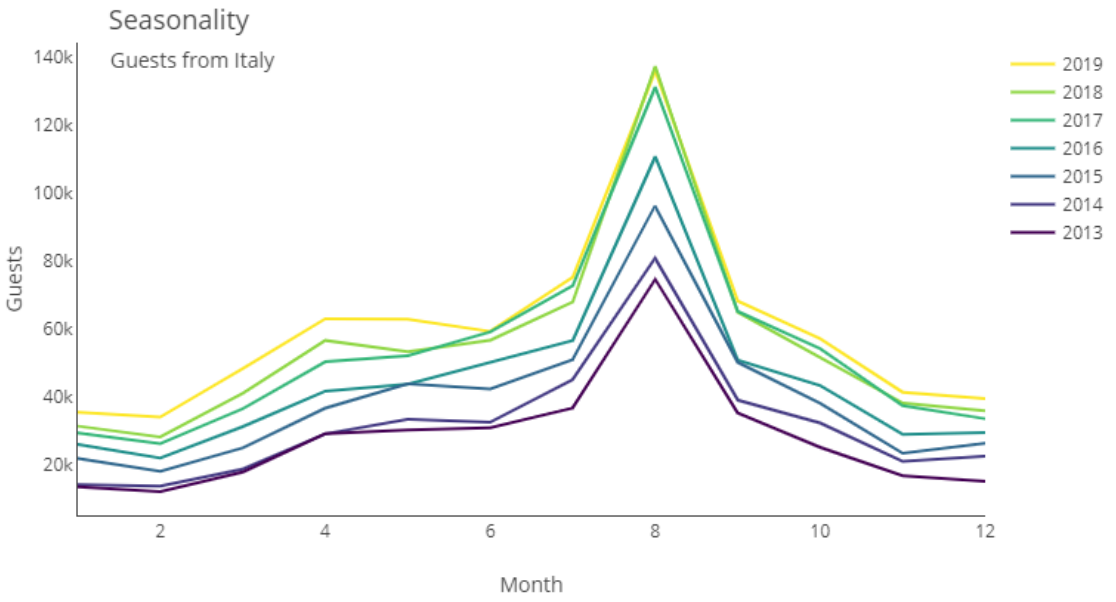


Source: INE

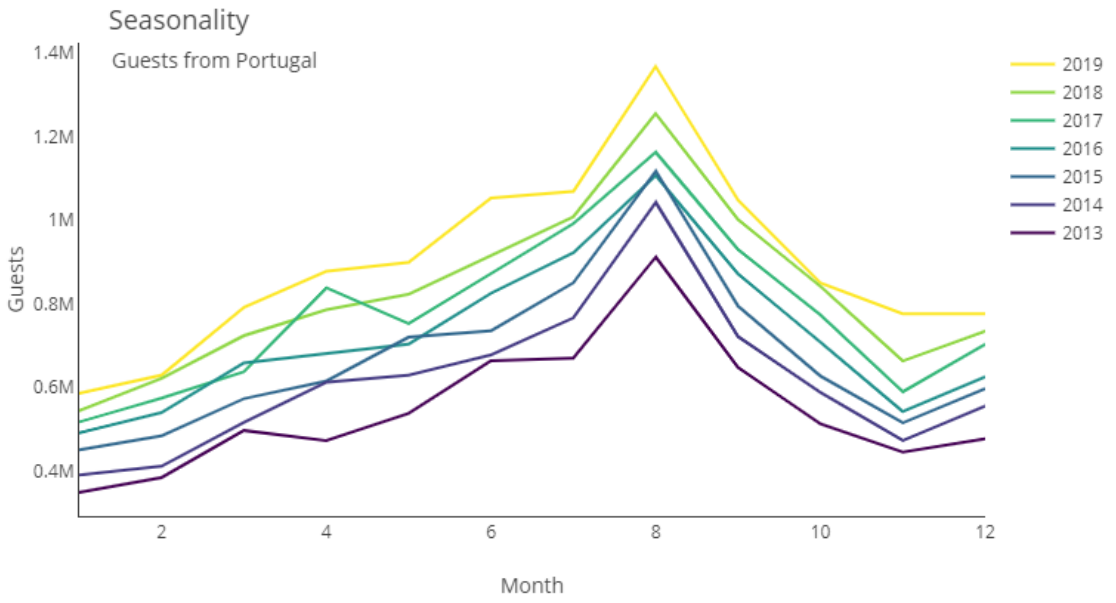


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



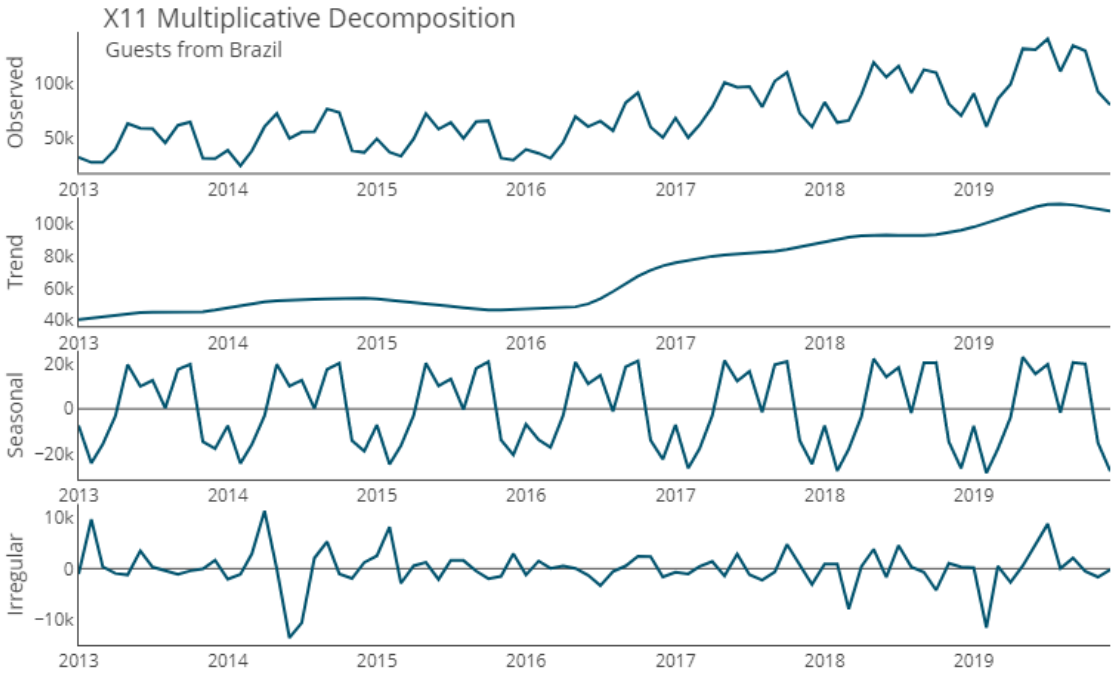
Source: INE

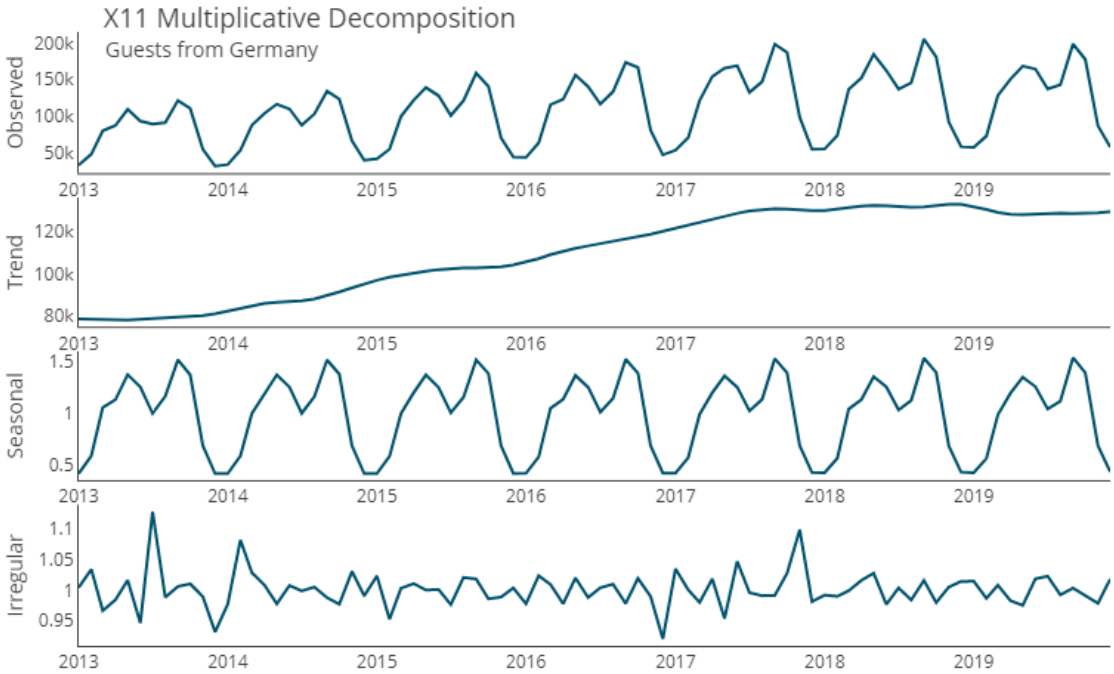


Source: INE

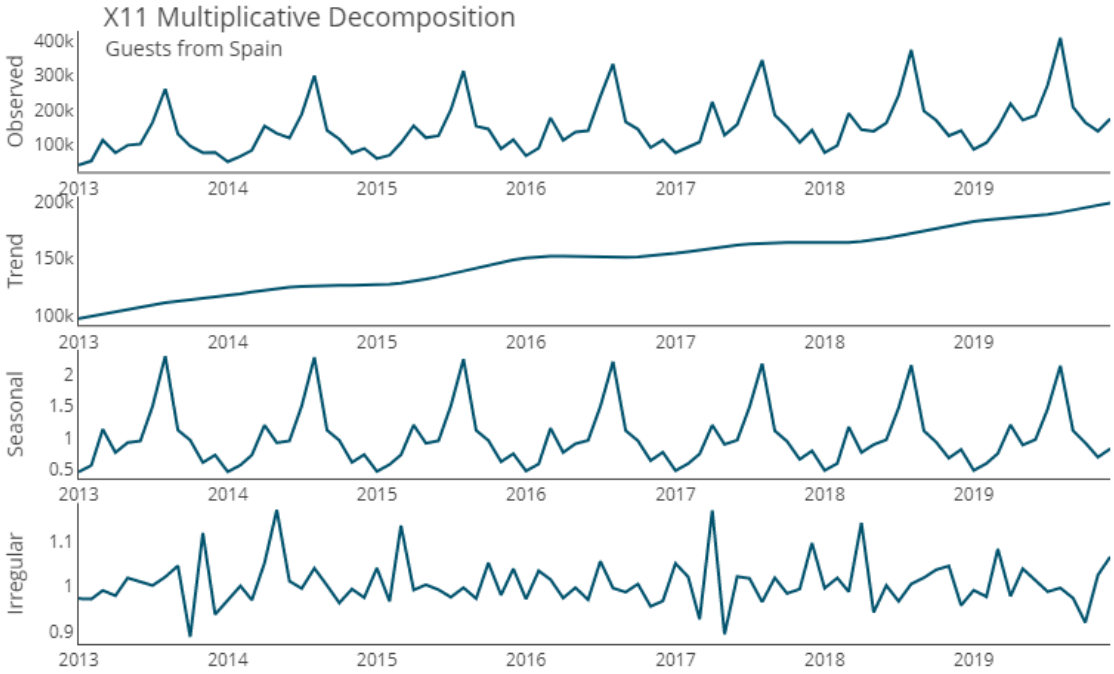
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

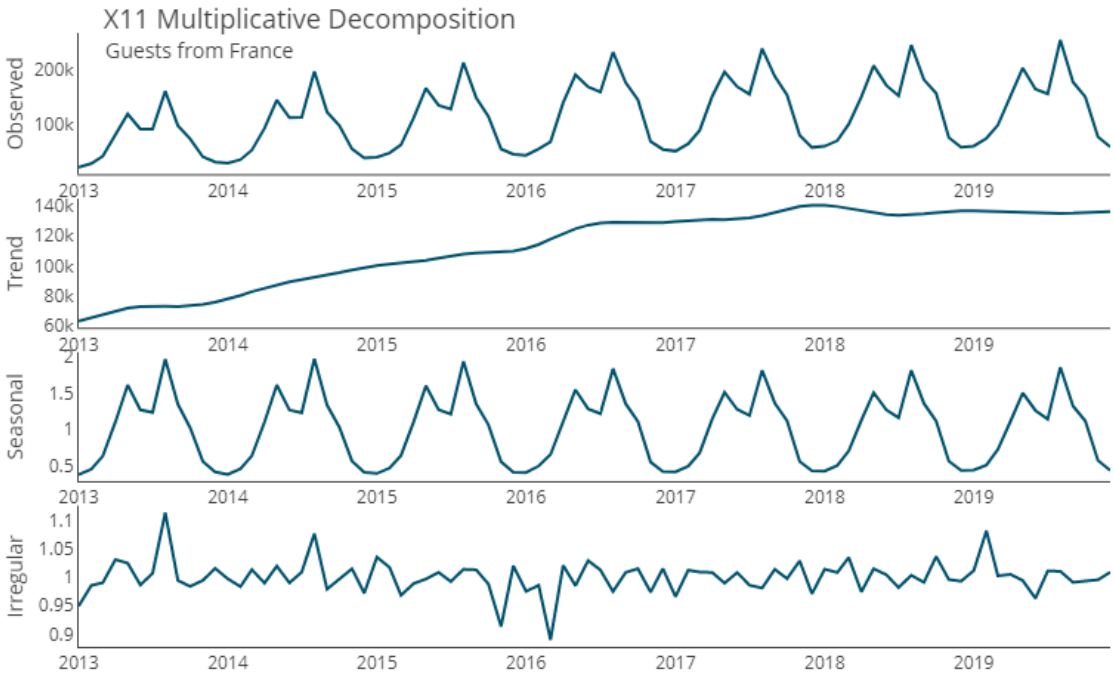
APPENDIX 3. GUESTS X-11 DECOMPOSITION PLOTS



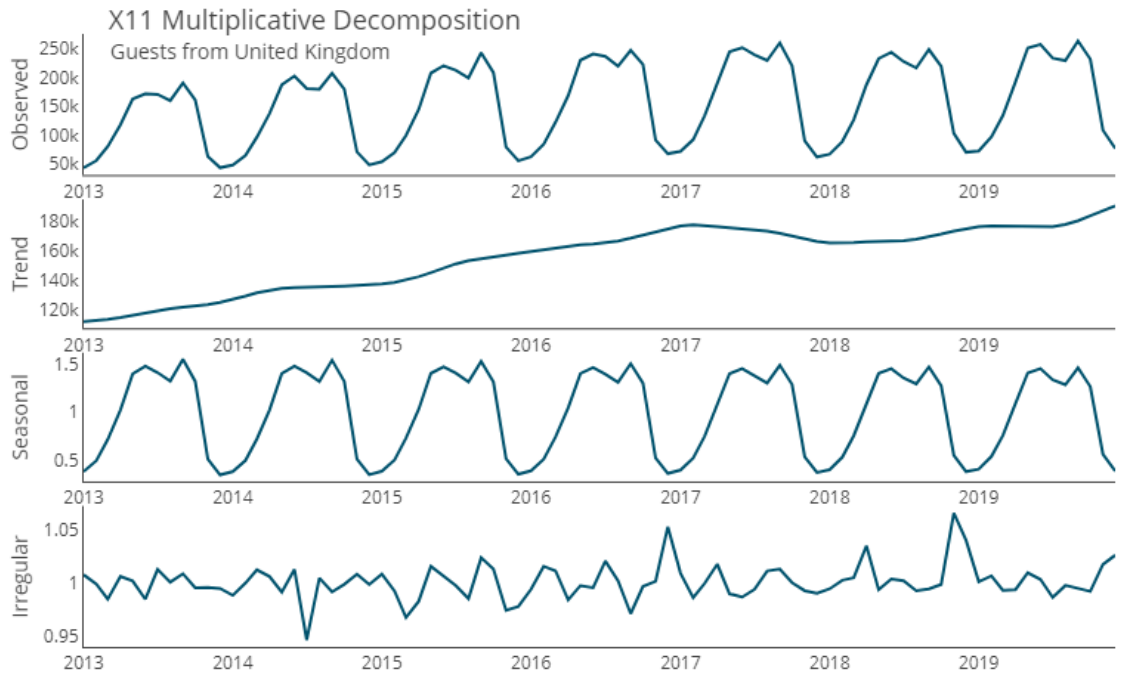


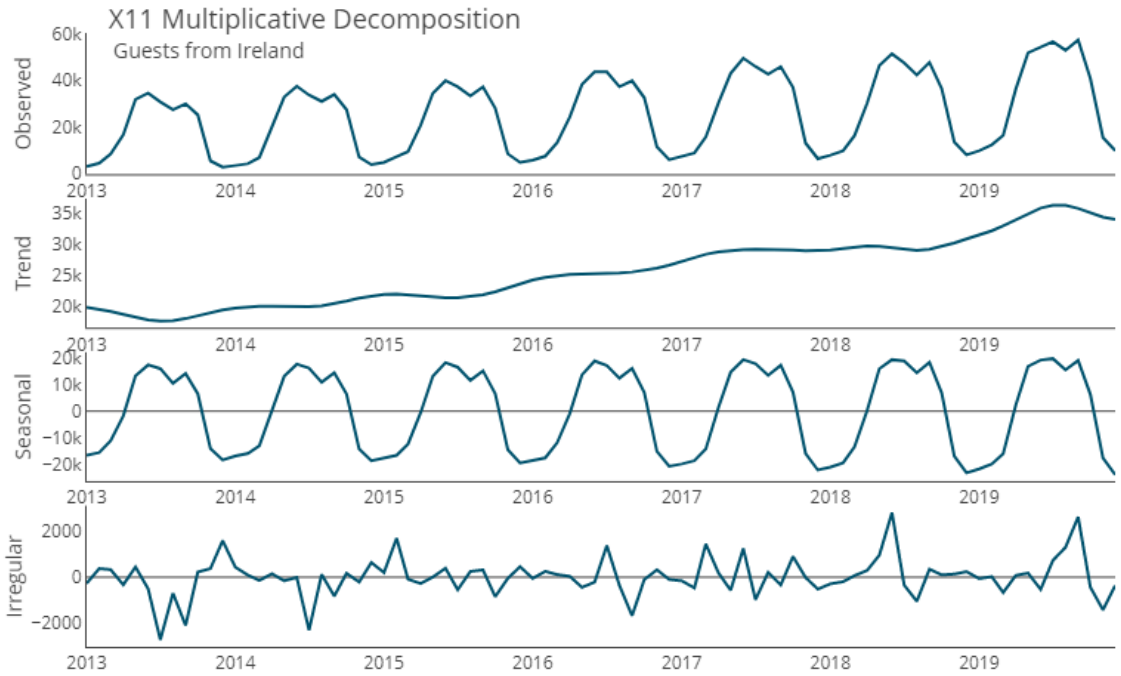
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



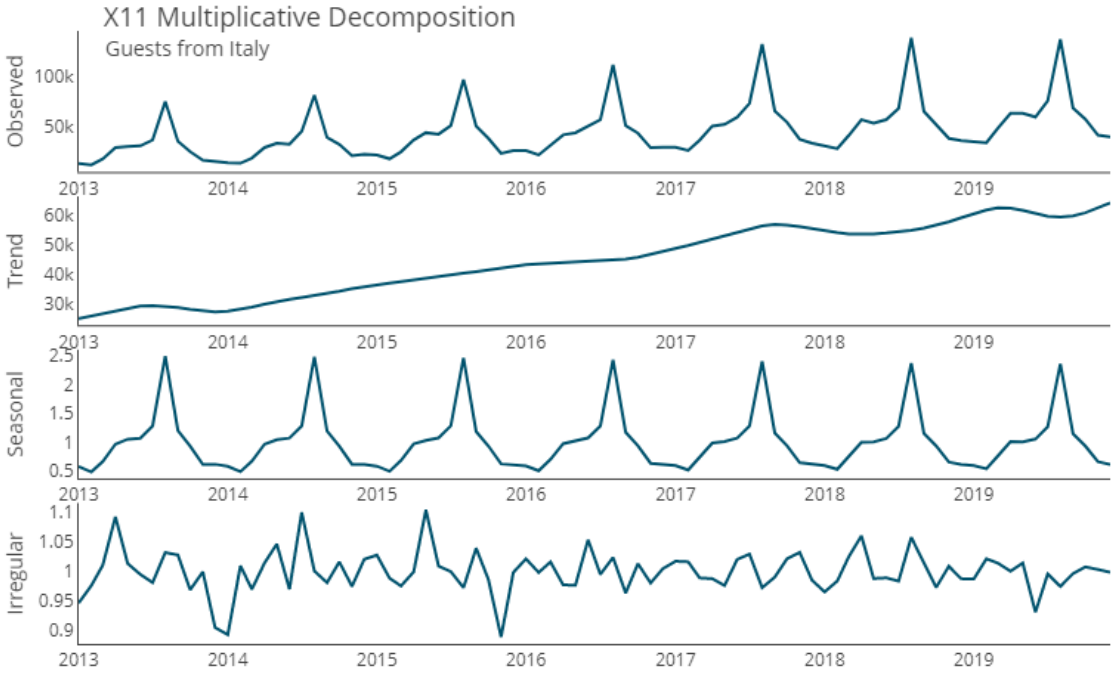


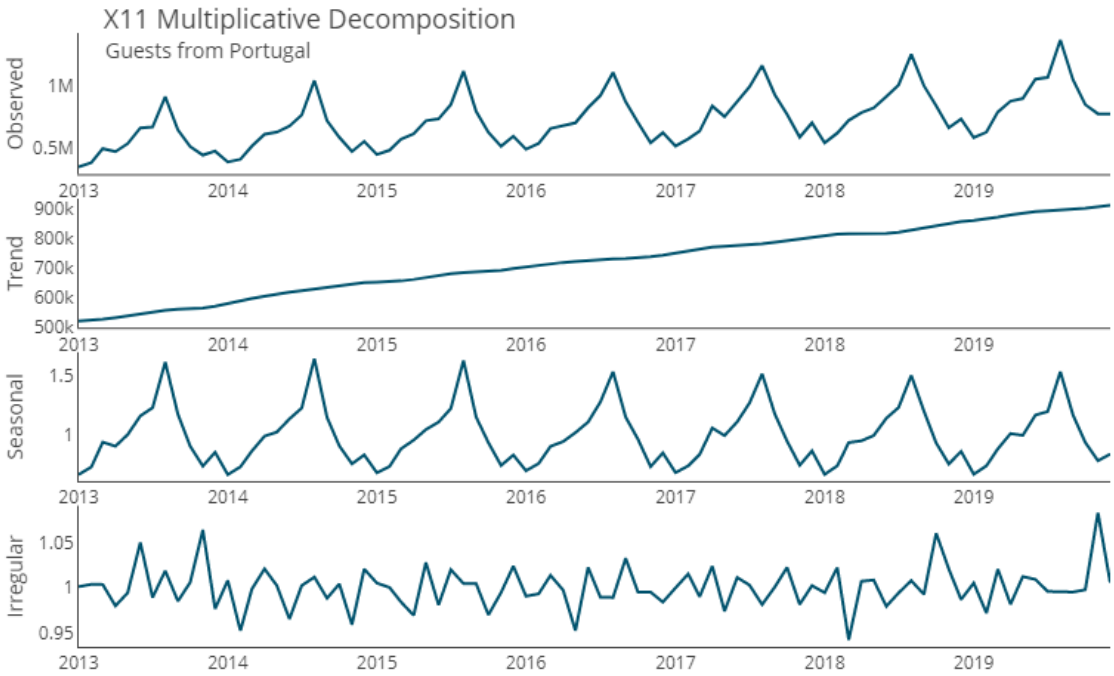
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt





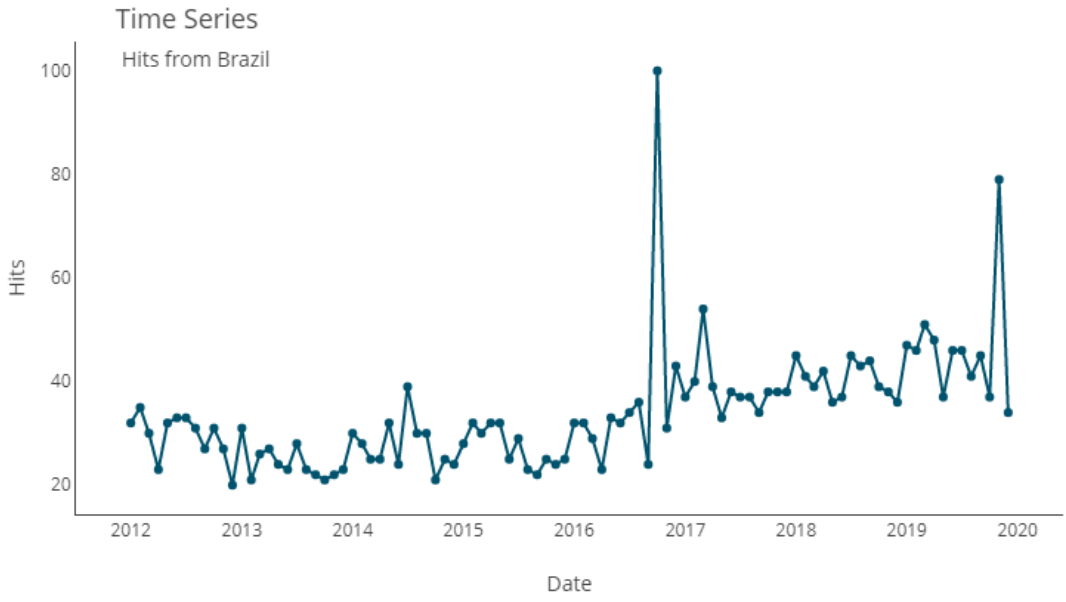
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



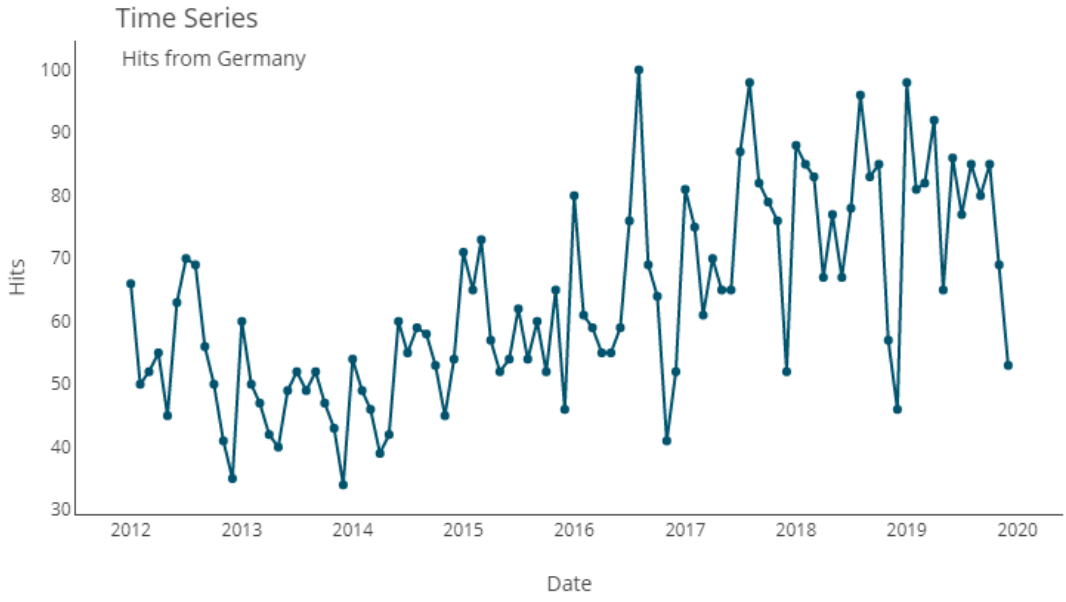


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 4. HITS TIME SERIES PLOTS

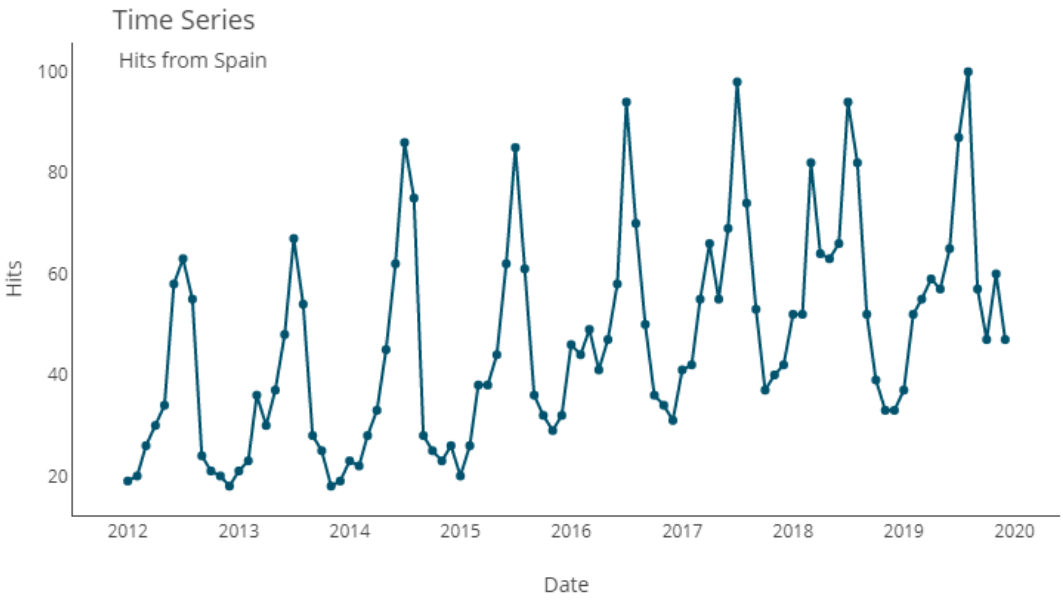


Source: INE

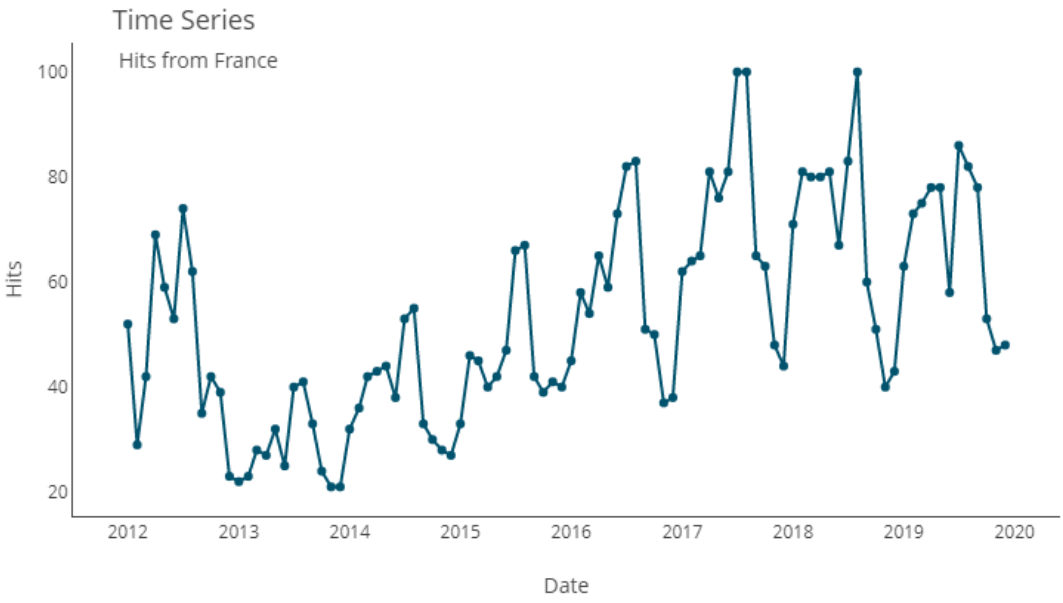


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

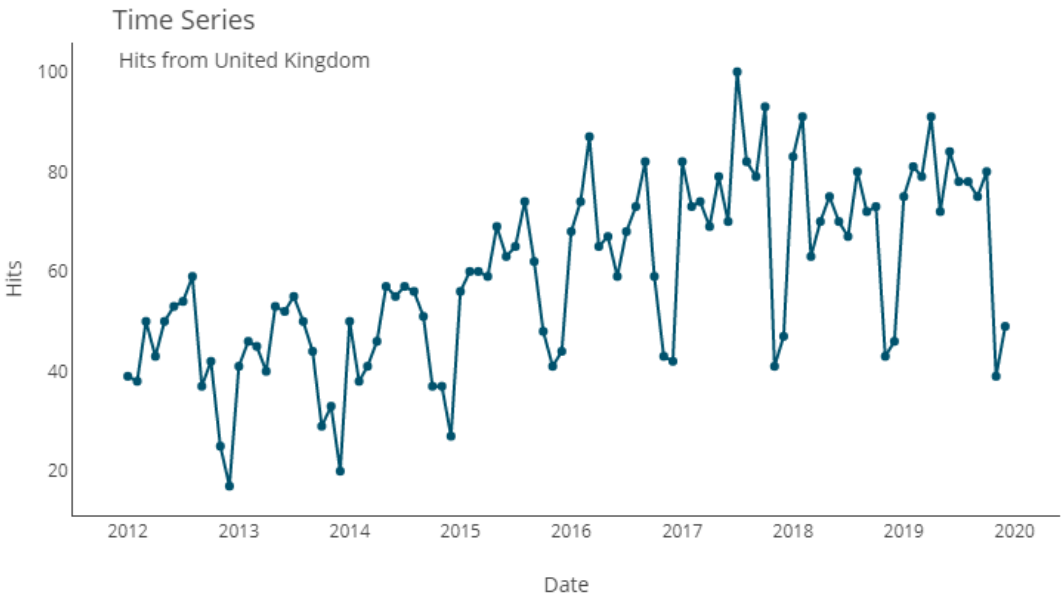


Source: INE

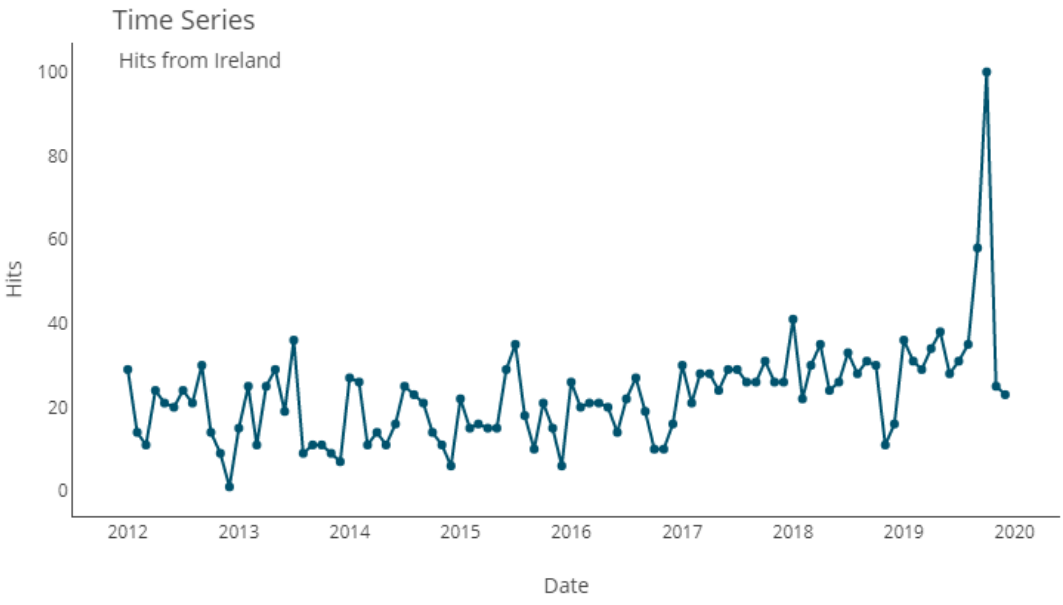


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

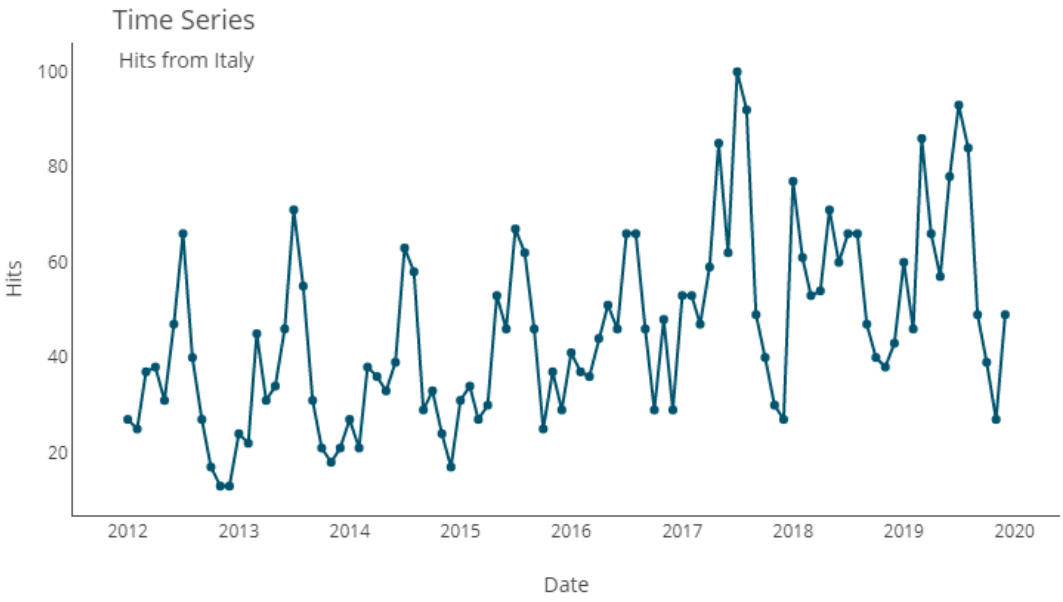


Source: INE

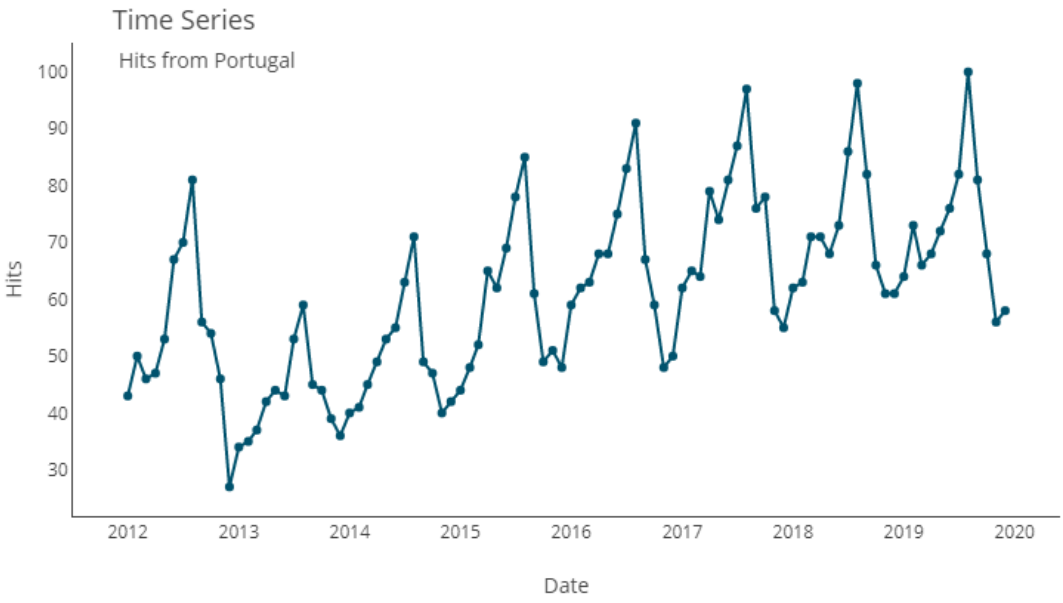


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



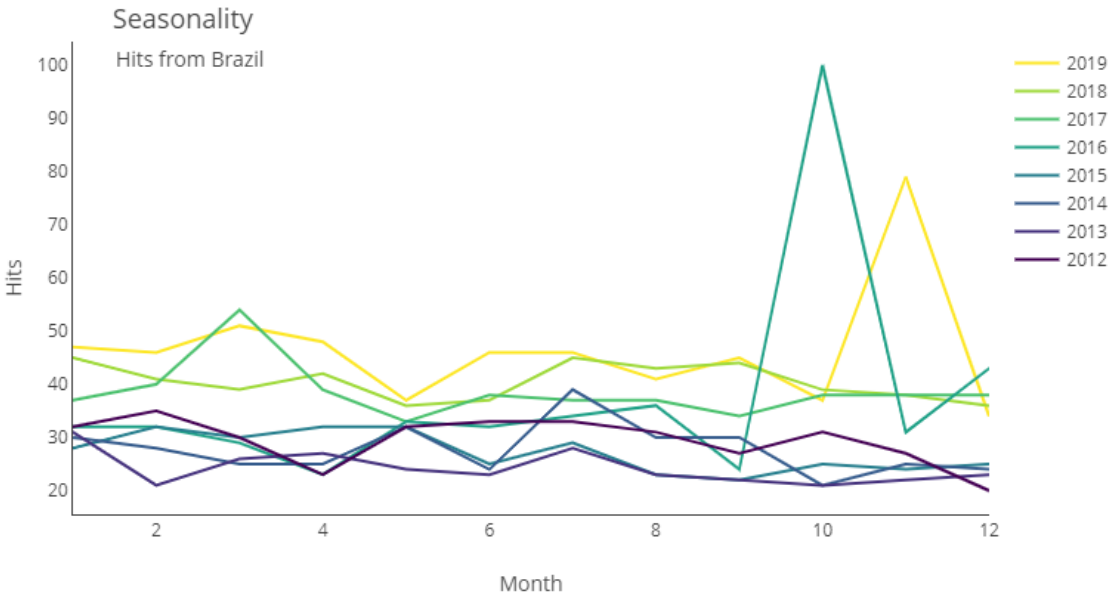
Source: INE



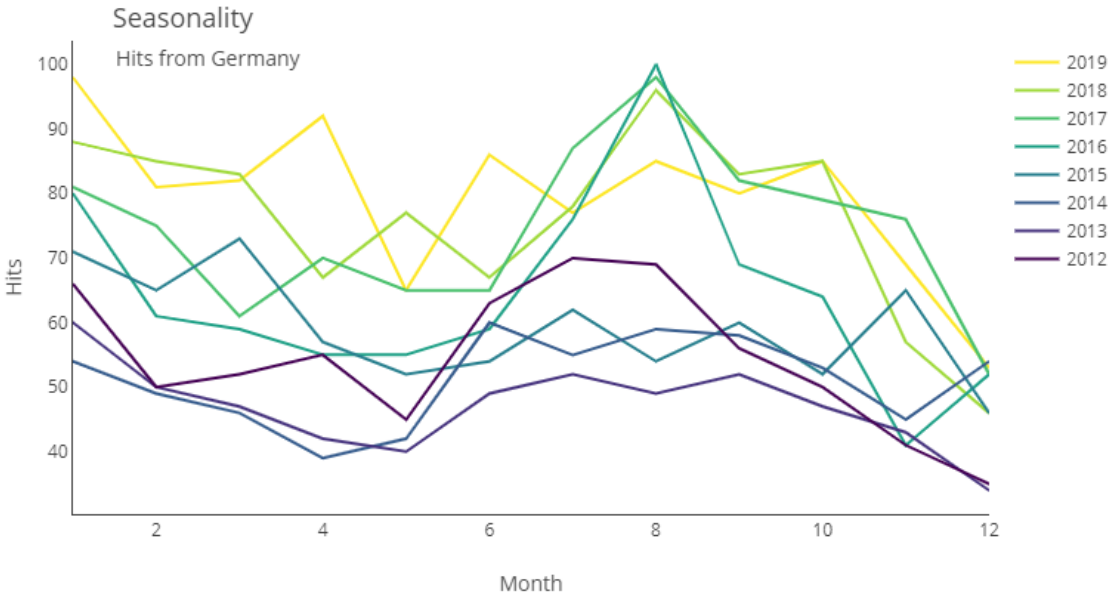
Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 5. HITS SEASONALITY PLOTS

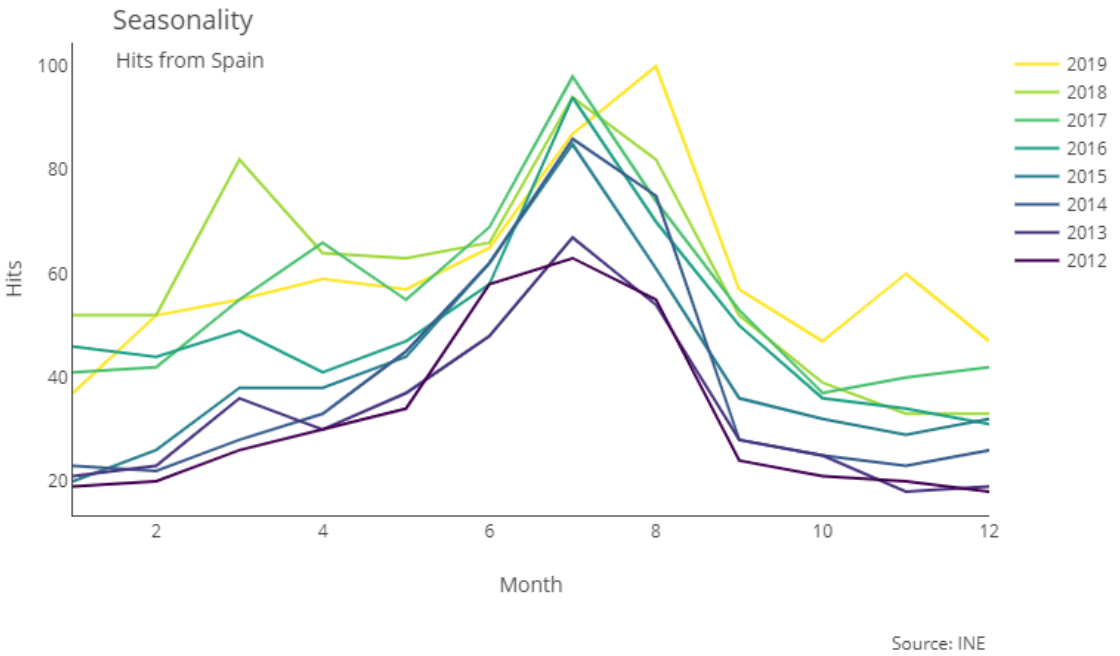


Source: INE

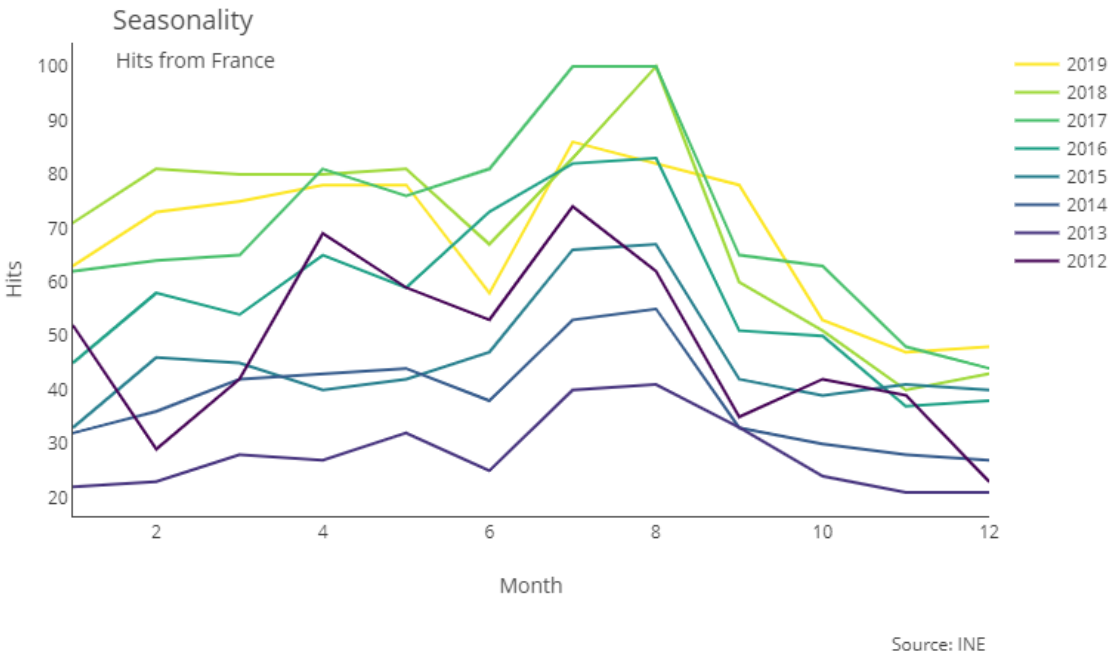


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

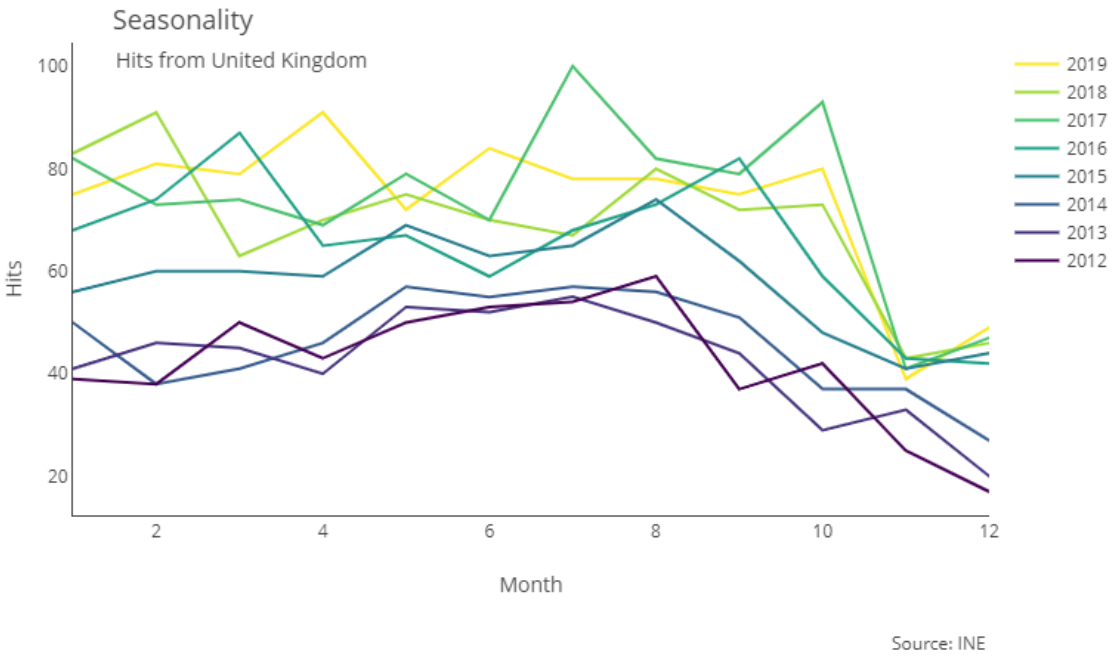


Source: INE

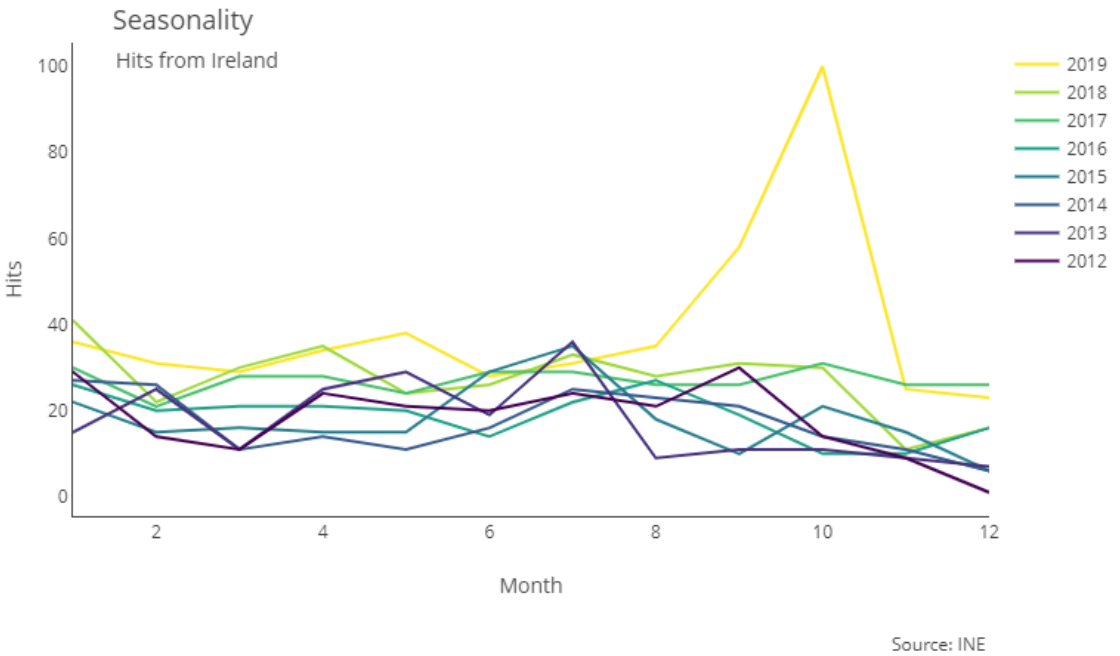


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

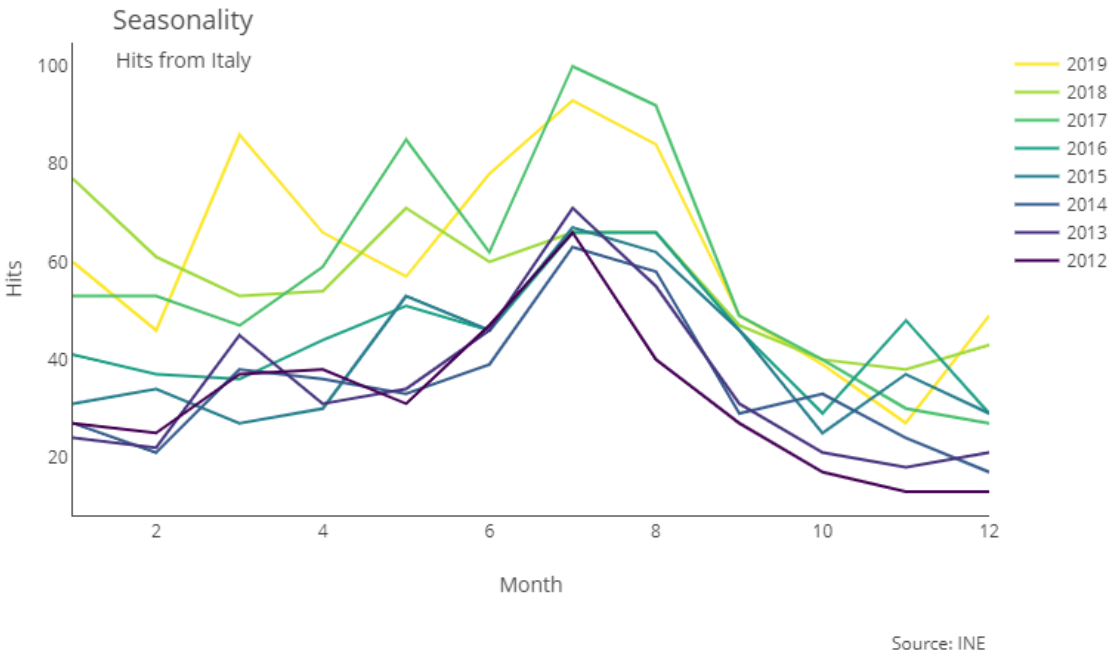


Source: INE

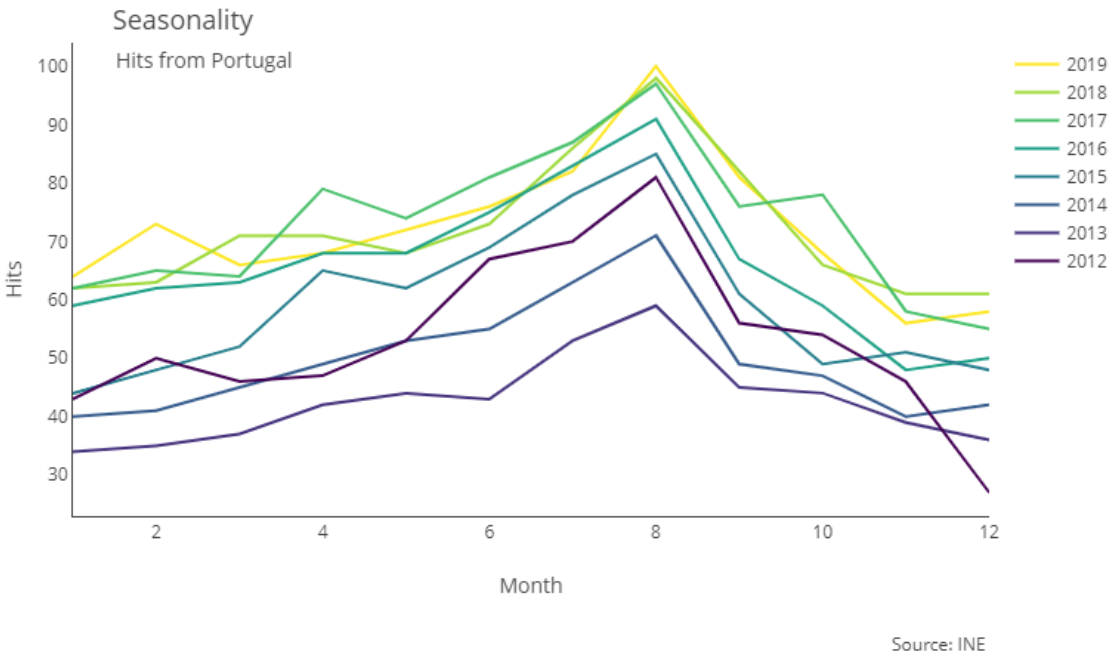


Source: INE

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



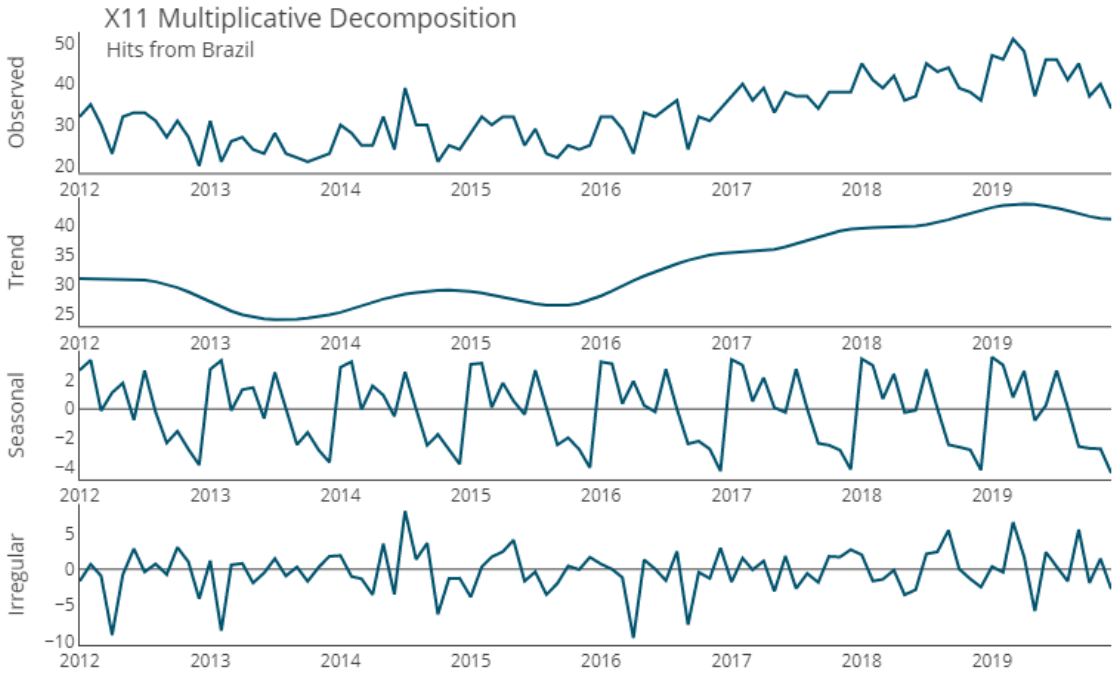
Source: INE

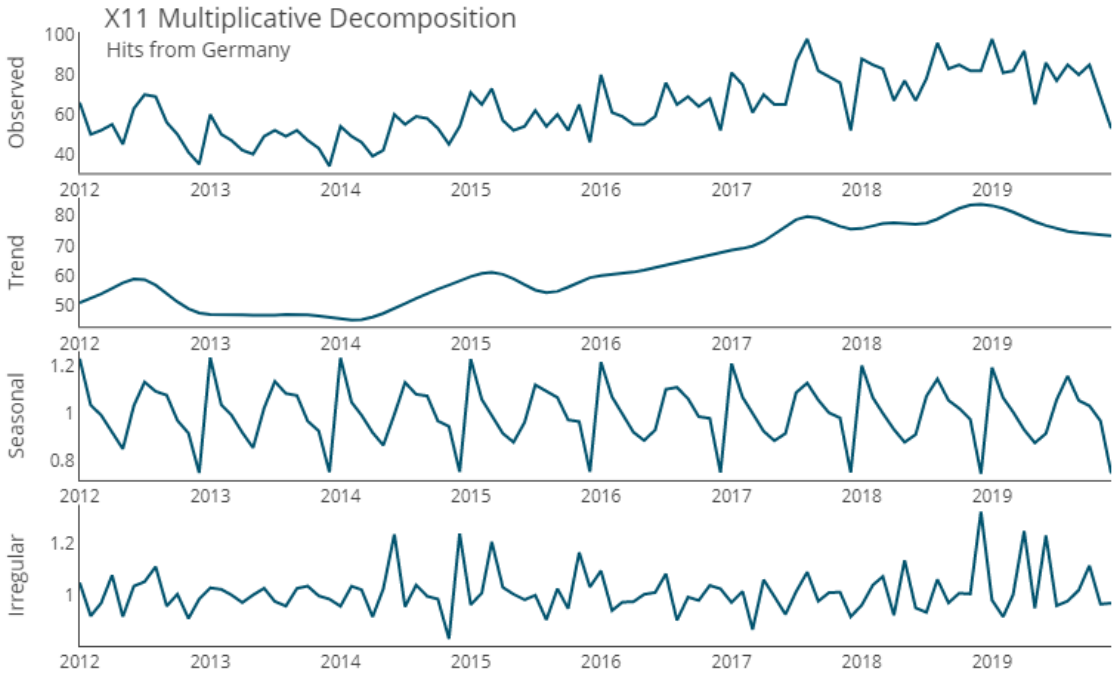


Source: INE

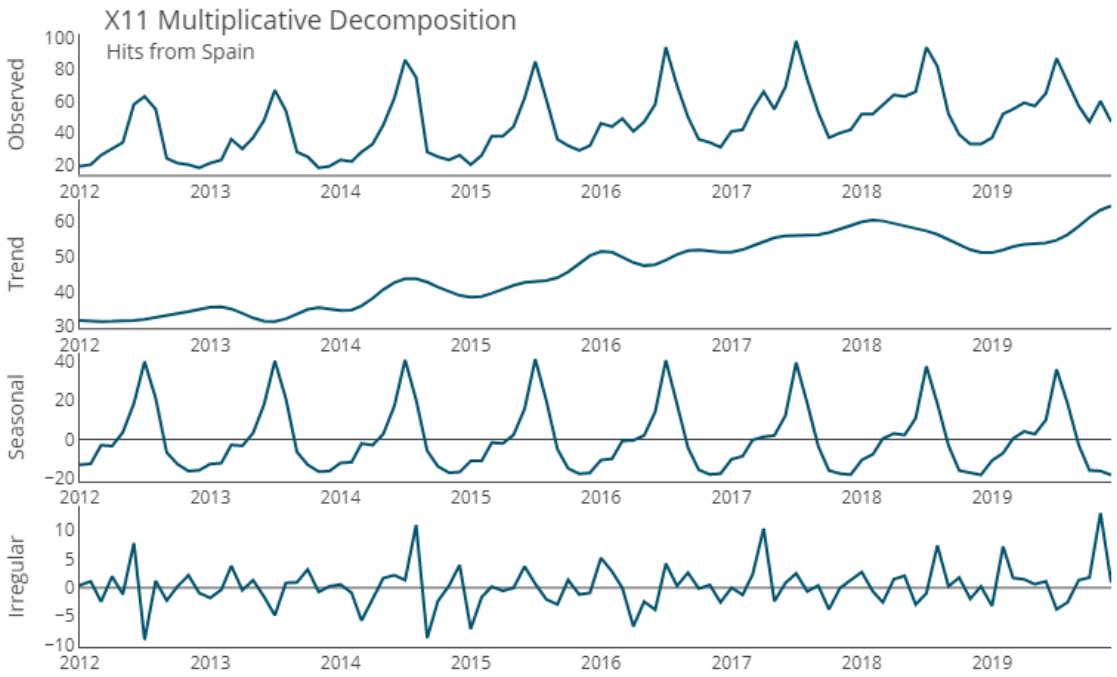
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

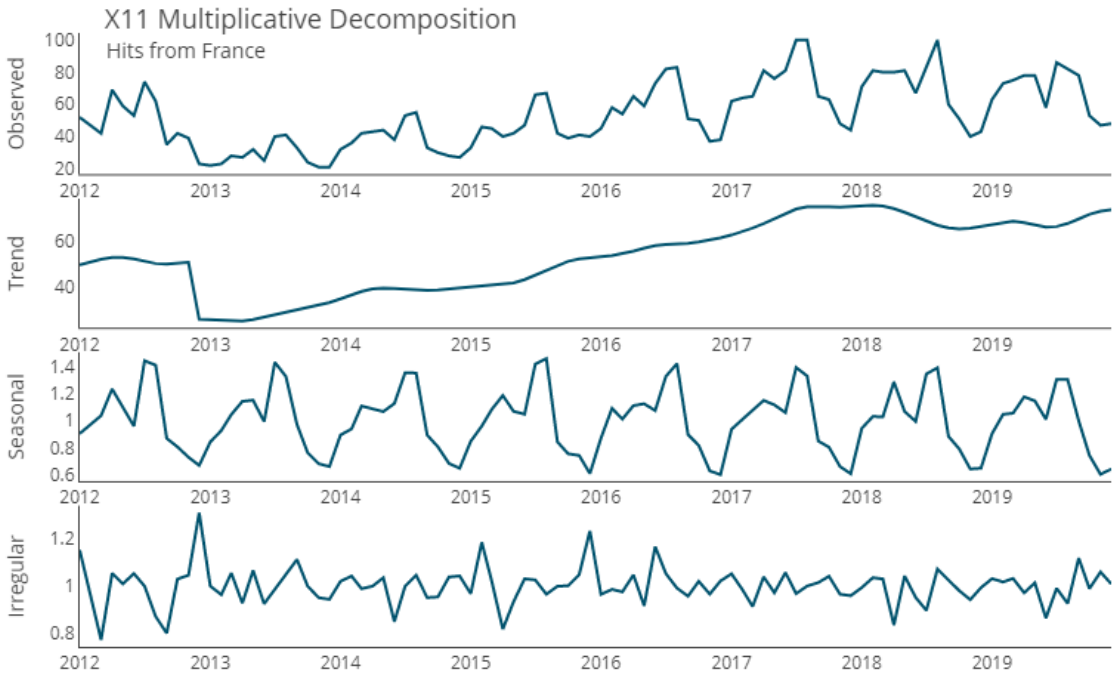
APPENDIX 6. HITS X-11 DECOMPOSITION PLOTS



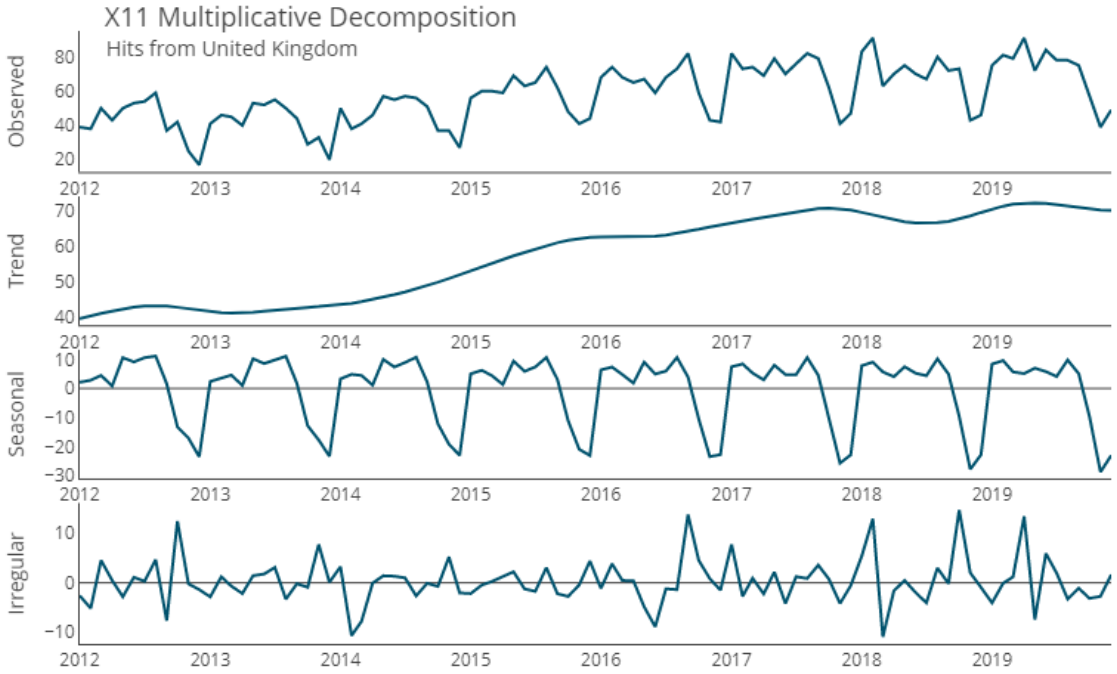


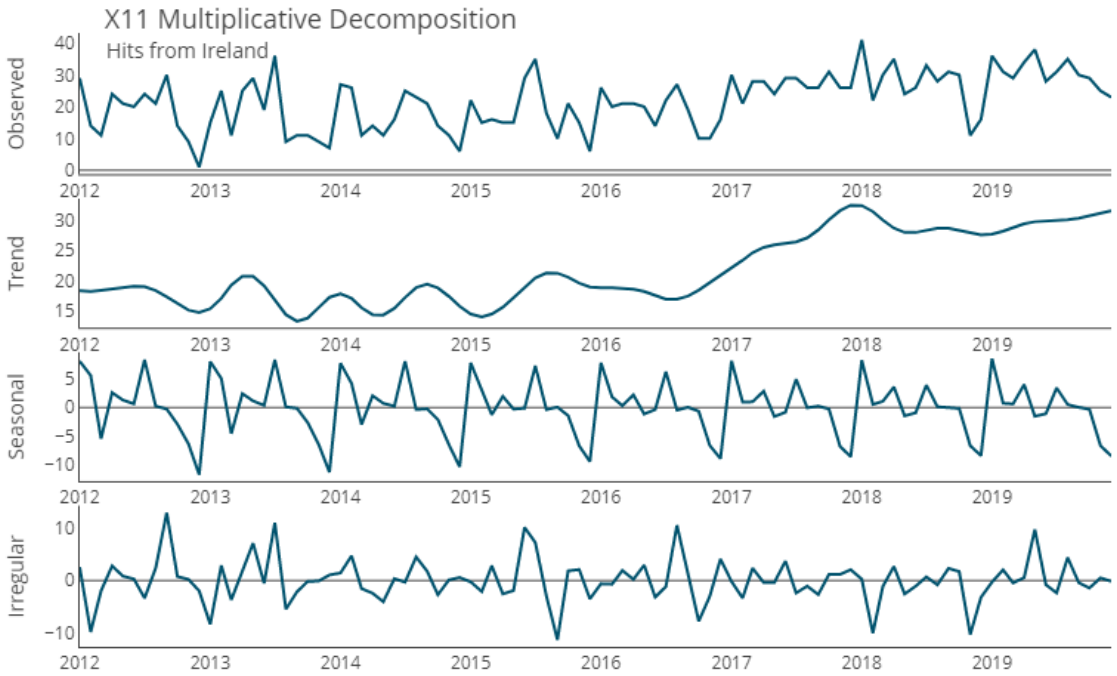
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



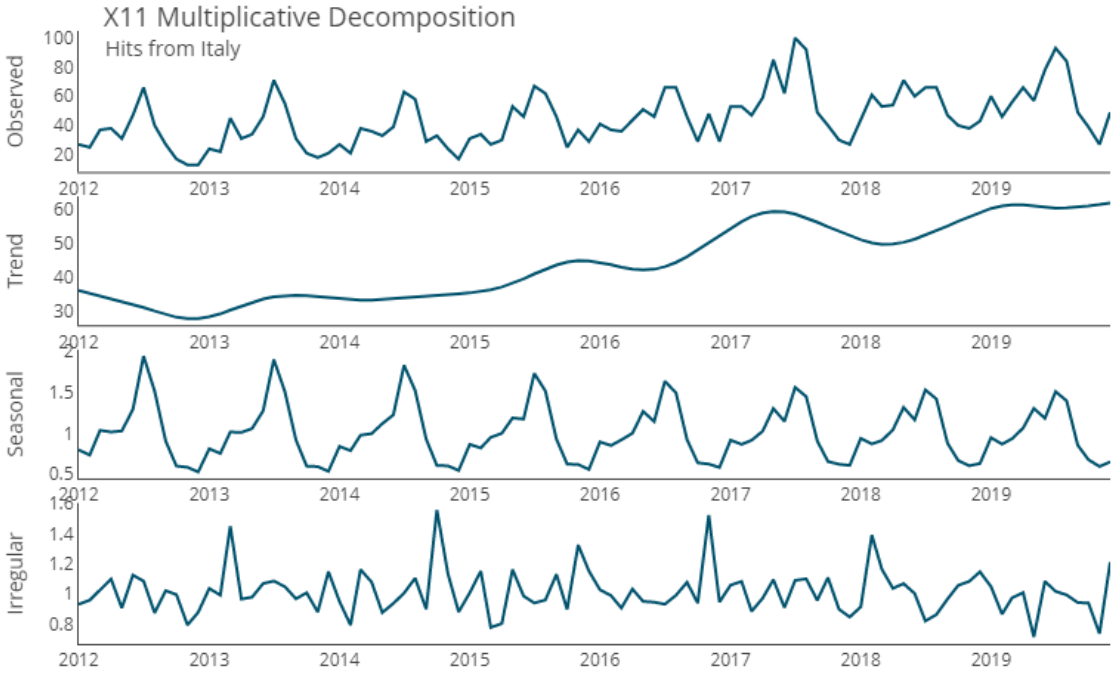


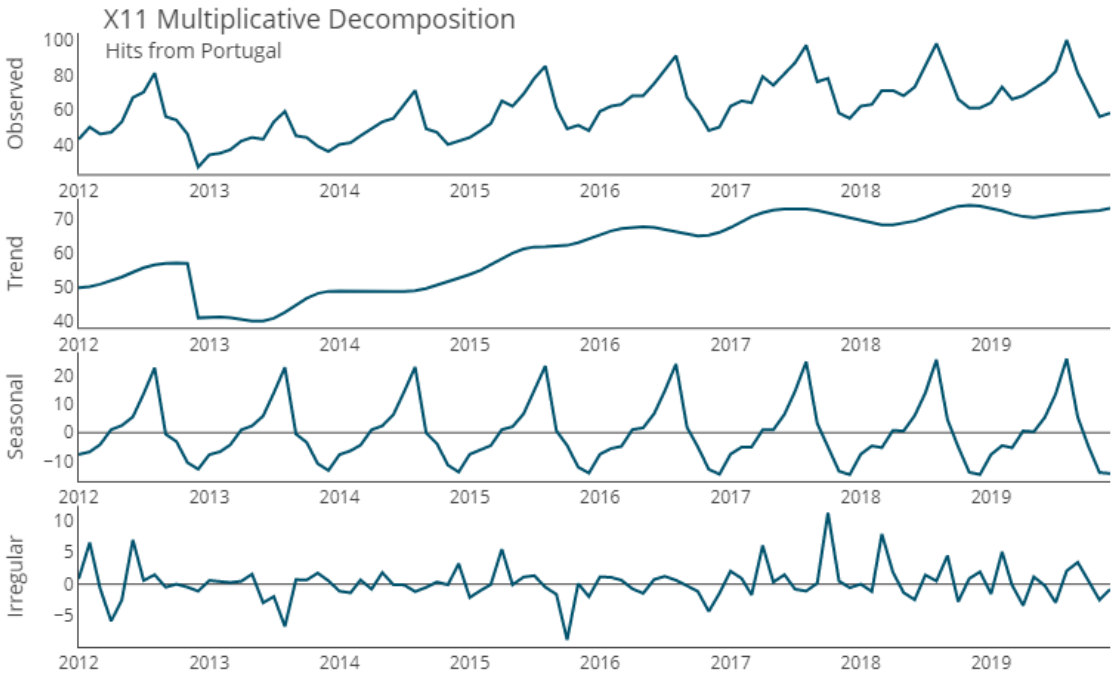
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt





Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt





Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 7. STATIONARITY TEST

Country	KPSS Stat	P Value	Hypothesis	Guests Time Serie
Brazil	1,6007856	0,01	Rejected	Non-stationary
Germany	0,7321259	0,0106249	Accepted	Stationary
Spain	0,6247601	0,0203854	Accepted	Stationary
France	0,5270429	0,0355759	Accepted	Stationary
United Kingdom	0,3164211	0,1	Accepted	Stationary
Ireland	0,3821455	0,0848511	Accepted	Stationary
Italy	0,8408935	0,01	Rejected	Non-stationary
Portugal	0,9741541	0,01	Rejected	Non-stationary

Country	KPSS Stat	P Value	Hypothesis	Hits Time Serie
Brazil	1,7036091	0,01	Rejected	Non-stationary
Germany	1,8156206	0,01	Rejected	Non-stationary
Spain	1,0032457	0,01	Rejected	Non-stationary
France	1,3740963	0,01	Rejected	Non-stationary
United Kingdom	1,642976	0,01	Rejected	Non-stationary
Ireland	1,2364706	0,01	Rejected	Non-stationary
Italy	1,1734395	0,01	Rejected	Non-stationary
Portugal	1,3616524	0,01	Rejected	Non-stationary

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 8. CORRELATION FIRST DIFFERENCE

LAG	GB	IE	ES	FR	DE	BR	PT	IT
Lag 0	0.335	0.336	0.538	0.45	0.022	0.11	0.669	0.335
Lag 1	0.224	-0.124	0.523	0.469	0.181	-0.171	0.296	0.386
Lag 2	0.214	0.212	-0.109	0.075	0.318	-0.04	0.041	-0.115
Lag 3	0.093	0.051	-0.147	0.136	-0.059	0.237	-0.229	0.132
Lag 5	0.032	-0.025	-0.061	0.025	-0.118	-0.114	-0.012	-0.156
Lag 5	-0.135	-0.031	0.137	-0.207	-0.252	-0.072	-0.255	0.04
Lag 6	-0.291	-0.065	-0.298	-0.137	-0.122	0.051	-0.125	-0.178
Lag 7	-0.085	-0.207	-0.164	-0.185	0.078	0.057	-0.119	0.189
Lag 8	0.144	0.152	0.008	-0.223	0.417	-0.029	-0.049	-0.265
Lag 9	-0.042	-0.064	0.083	0.107	0.049	0.087	-0.34	-0.034
Lag 10	-0.401	-0.106	-0.289	-0.017	-0.303	-0.14	0.033	-0.055
Lag 11	-0.056	-0.003	-0.129	-0.379	-0.119	-0.013	-0.013	-0.084
Lag 12	0.278	-0.041	0.324	0.389	-0.059	0.188	0.632	0.247

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 9. CORRELATION ORIGINAL DATA

LAG	GB	IE	ES	FR	DE	BR	PT	IT
Lag 0	0.519	0.368	0.764	0.663	0.395	0.497	0.884	0.669
Lag 1	0.617	0.281	0.836	0.767	0.524	0.438	0.809	0.773
Lag 2	0.584	0.246	0.517	0.651	0.563	0.519	0.628	0.572
Lag 3	0.46	0.381	0.27	0.477	0.339	0.597	0.404	0.466
Lag 5	0.284	0.282	0.133	0.266	0.144	0.508	0.277	0.246
Lag 5	0.064	0.117	-0.002	0.077	0.042	0.488	0.068	0.151
Lag 6	-0.082	-0.043	-0.244	-0.025	0.119	0.524	-0.032	-0.016
Lag 7	-0.077	-0.122	-0.269	-0.087	0.344	0.525	-0.042	0.002
Lag 8	-0.029	-0.054	-0.186	-0.091	0.489	0.491	0.02	-0.139
Lag 9	-0.041	-0.042	-0.091	0.044	0.352	0.483	0.162	-0.033
Lag 10	-0.036	-0.003	-0.035	0.173	0.137	0.41	0.403	0.047
Lag 11	0.237	0.206	0.271	0.3	0.174	0.443	0.633	0.308
Lag 12	0.508	0.364	0.708	0.582	0.329	0.47	0.851	0.655

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 10. CORRELATION WITH OUTLIERS REMOVED

LAG	GB	IE	ES	FR	DE	BR	PT	IT
Lag 0	0.505	0.373	0.734	0.663	0.344	0.657	0.884	0.705
Lag 1	0.671	0.433	0.864	0.767	0.426	0.675	0.809	0.804
Lag 2	0.632	0.471	0.54	0.651	0.482	0.778	0.628	0.585
Lag 3	0.485	0.452	0.283	0.477	0.346	0.823	0.404	0.465
Lag 5	0.323	0.282	0.116	0.266	0.222	0.76	0.277	0.226
Lag 5	0.088	0.117	-0.045	0.077	0.148	0.699	0.068	0.08
Lag 6	-0.093	-0.043	-0.258	-0.025	0.228	0.676	-0.032	-0.055
Lag 7	-0.103	-0.122	-0.278	-0.087	0.416	0.592	-0.042	-0.088
Lag 8	-0.039	-0.054	-0.185	-0.091	0.521	0.607	0.02	-0.161
Lag 9	-0.058	-0.042	-0.091	0.044	0.416	0.602	0.162	-0.037
Lag 10	-0.061	-0.003	-0.023	0.173	0.228	0.553	0.403	0.054
Lag 11	0.216	0.206	0.284	0.285	0.217	0.555	0.633	0.322
Lag 12	0.515	0.364	0.721	0.57	0.267	0.579	0.851	0.678

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 11. CORRELATION WITH OUTLIERS AND IRREGULAR COMPONENT REMOVED

LAG	GB	IE	ES	FR	DE	BR	PT	IT
Lag 0	0.528	0.355	0.733	0.663	0.344	0.71	0.902	0.705
Lag 1	0.725	0.402	0.891	0.767	0.426	0.723	0.824	0.805
Lag 2	0.705	0.441	0.549	0.651	0.482	0.828	0.632	0.585
Lag 3	0.559	0.464	0.288	0.477	0.346	0.885	0.387	0.465
Lag 5	0.369	0.353	0.111	0.266	0.222	0.818	0.251	0.227
Lag 5	0.098	0.214	-0.047	0.077	0.148	0.769	0.038	0.08
Lag 6	-0.12	0.067	-0.281	-0.025	0.228	0.729	-0.058	-0.055
Lag 7	-0.155	-0.009	-0.286	-0.087	0.416	0.641	-0.04	-0.088
Lag 8	-0.109	0.039	-0.2	-0.091	0.521	0.669	0.022	-0.162
Lag 9	-0.125	0.006	-0.097	0.044	0.416	0.627	0.164	-0.039
Lag 10	-0.102	-0.001	-0.015	0.173	0.228	0.541	0.418	0.052
Lag 11	0.188	0.175	0.287	0.285	0.217	0.581	0.651	0.322
Lag 12	0.532	0.331	0.723	0.57	0.267	0.629	0.875	0.678

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 12. MODELS ACCURACY

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Brazil	Model 1	Test	Split	-5123.05	12133.79	10329.87	11.45
Brazil	Model 1	Test	CV	-374.33	9972.38	8692.04	8.96
Brazil	Model 2	Test	Split	-11342.5	14895.66	13374.18	15.11
Brazil	Model 2	Test	CV	-167.93	9121.12	7090.06	7.65
Brazil	Model 3	Test	CV	-2131.4	12981.88	10484.87	11.61
Brazil	Model 4	Test	CV	699.76	8823.88	7078.27	7.72
Brazil	Model 5	Test	CV	1037.28	7869.15	6161.52	6.84

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Germany	Model 1	Test	Split	-13479.4	16392.65	13671.52	10.15
Germany	Model 1	Test	CV	-3044.27	7025.25	5905.85	5.02
Germany	Model 2	Test	Split	-17774	19761.19	17774	15.67
Germany	Model 2	Test	CV	-4812.2	9336.97	6962.6	5.85
Germany	Model 3	Test	CV	-4325.87	8869.73	6964.63	5.47
Germany	Model 4	Test	CV	-3018.63	7682.94	5811.58	4.61
Germany	Model 5	Test	CV	-3018.63	7682.94	5811.58	4.61

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Spain	Model 1	Test	Split	1646.38	15937.97	12141.26	6.64
Spain	Model 1	Test	CV	2049.9	15163.3	11208.53	6.39
Spain	Model 2	Test	Split	3025.17	14001.31	10309.27	5.82
Spain	Model 2	Test	CV	1173.92	14412.16	11838.62	7.03
Spain	Model 3	Test	CV	25433.32	46450.8	37421.69	18.68
Spain	Model 4	Test	CV	13869.87	34508.98	27154.4	14.02
Spain	Model 5	Test	CV	-271.78	16566.88	13815.37	8.67

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
France	Model 1	Test	Split	-7417.23	16151.63	12588.67	8.47
France	Model 1	Test	CV	-2550.64	11231.63	7604.85	5.04
France	Model 2	Test	Split	-10463.8	13110.42	11181.71	10.5
France	Model 2	Test	CV	-1588.18	6472.73	5301.8	4.03
France	Model 3	Test	CV	-4041.57	7694.29	6024.17	4.75
France	Model 4	Test	CV	-4041.57	7694.29	6024.17	4.75
France	Model 5	Test	CV	-4041.55	7694.29	6024.18	4.75

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
United Kingdom	Model 1	Test	Split	2719.33	9034.03	7951.95	6.24
United Kingdom	Model 1	Test	CV	-41.64	6992.28	5296.56	3.66
United Kingdom	Model 2	Test	Split	20525.92	23293.89	20809.67	15.73
United Kingdom	Model 2	Test	CV	2108.01	6793.22	4648.74	2.84
United Kingdom	Model 3	Test	CV	224	6076.15	4895.63	3.11
United Kingdom	Model 4	Test	CV	159.84	6050.89	4902.79	3.05
United Kingdom	Model 5	Test	CV	884.37	5931.64	4312.07	2.85

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Ireland	Model 1	Test	Split	950.33	5284.7	4498.35	19.84
Ireland	Model 1	Test	CV	322.76	3475.67	2650.4	9.79
Ireland	Model 2	Test	Split	672.46	3256.58	2492.15	8.24
Ireland	Model 2	Test	CV	468.87	2877.5	2157.03	6.31
Ireland	Model 3	Test	CV	458.15	2884.85	2189.46	6.55
Ireland	Model 4	Test	CV	473.62	2743	2081.94	6.19
Ireland	Model 5	Test	CV	679.58	2873.78	2218.27	7.28

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Italy	Model 1	Test	Split	5768.08	6405.74	5768.08	11.39
Italy	Model 1	Test	CV	350.16	4151.19	3271.27	5.46
Italy	Model 2	Test	Split	-917.5	3584.93	3033.9	5.5
Italy	Model 2	Test	CV	-397.2	3820.96	2883.43	4.27
Italy	Model 3	Test	CV	-351.88	3904.26	2956.49	4.32
Italy	Model 4	Test	CV	-331.18	3972.66	3017.02	4.45
Italy	Model 5	Test	CV	-330.86	3977.29	3019.47	4.45

Country	Model	Set	Setup	ME	RMSE	MAE	MAPE
Portugal	Model 1	Test	Split	17440.9	44941.23	34908.68	3.84
Portugal	Model 1	Test	CV	11382.56	41803.78	30938.11	3.38
Portugal	Model 2	Test	Split	12943.23	40226.25	32340.35	3.68
Portugal	Model 2	Test	CV	9197.21	37893.06	30951.88	3.58
Portugal	Model 3	Test	CV	19073	46065.78	42159.46	4.87
Portugal	Model 4	Test	CV	19073	46065.78	42159.46	4.87
Portugal	Model 5	Test	CV	7429.03	40759.03	36479.7	4.2

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

APPENDIX 13. BEST MODEL SUMMARY

Brazil

```
## ETS(A,A,A)
##
## Call:
##  ets(y = train_guests)
##
##   Smoothing parameters:
##     alpha = 0.6815
##     beta  = 1e-04
##     gamma = 1e-04
##
##   Initial states:
##     l = 41111.7056
##     b = 595.5728
##     s = -19817.71 -13760.8 21432.18 18629.68 -1060.835 11406.5
##           7491.518 19635 334.7761 -16159.93 -19615.9 -8514.475
##
##   sigma:  5957.534
##
##      AIC     AICc      BIC
## 1462.216 1474.705 1499.695
##
## Training set error measures:
##                    ME     RMSE      MAE         MPE     MAPE      MASE
## Training set 460.6567 5197.736 4393.131 -0.05046885 8.232642 0.3429732
##                    ACF1
## Training set 0.09608287
```

Germany

```
## ETS(M,A,M)
##
## Call:
##  ets(y = train_guests)
##
##   Smoothing parameters:
##     alpha = 0.2192
##     beta  = 1e-04
##     gamma = 1e-04
##
##   Initial states:
##     l = 75020.2317
##     b = 922.185
##     s = 0.4081 0.6915 1.3823 1.51 1.151 1.0289
##           1.2555 1.3523 1.1835 1.0247 0.5896 0.4228
##
##   sigma:  0.0466
##
##      AIC     AICc      BIC
## 1422.588 1435.078 1460.068
##
## Training set error measures:
##                   ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
## Training set -123.46 4147.884 3236.078 -0.2991878 3.333109 0.2712262 -0.1552104
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Spain

```
## Series: train_guests
## ARIMA(0,0,1)(1,1,1)[12] with drift
##
## Coefficients:
##           ma1     sar1     sma1      drift
##       -0.2783  -0.5129  -0.4817  1028.1177
## s.e.   0.1305   0.1548   0.2001    71.7339
##
## sigma^2 estimated as 336267473:  log likelihood=-621.98
## AIC=1253.97   AICc=1255.19   BIC=1264
##
## Training set error measures:
##                    ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
## Training set 1143.512 15998.9 11371.43 -1.491834 8.222633 0.5321654 -0.01329659
```

France

```
## Series: train_guests
## ARIMA(1,1,1)(0,1,1)[12]
##
## Coefficients:
##          ar1      ma1     sma1
##       0.6526  -0.9661  0.2881
## s.e.  0.1434   0.0855  0.1740
##
## sigma^2 estimated as 54970945:  log likelihood=-557.26
## AIC=1122.52   AICc=1123.33   BIC=1130.47
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -336.0234 6468.659 4750.156 -1.120251 4.706268 0.3429373
##                     ACF1
## Training set -0.07022857
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

United Kingdom

```
## ETS(M,A,M)
##
## Call:
##  ets(y = train_guests)
##
##   Smoothing parameters:
##     alpha = 0.7839
##     beta  = 0.0064
##     gamma = 1e-04
##
##   Initial states:
##     l = 111217.0223
##     b = 1239.5353
##     s = 0.3617 0.5201 1.3143 1.5377 1.3226 1.3926
##           1.4627 1.4003 1.0469 0.7353 0.5121 0.3938
##
##   sigma:  0.032
##
##      AIC     AICc      BIC
## 1415.826 1428.316 1453.306
##
## Training set error measures:
##                     ME     RMSE      MAE       MPE     MAPE      MASE
## Training set -742.0545 5181.376 3576.552 -0.367059 2.265939 0.2670718
##                    ACF1
## Training set 0.03954793
```

Ireland

```
## Series: train_guests
## ARIMA(1,0,0)(0,1,0)[12] with drift
##
## Coefficients:
##          ar1     drift
##       0.4918  198.8857
## s.e.  0.1172   33.2656
##
## sigma^2 estimated as 2424514:  log likelihood=-481.44
## AIC=968.88   AICc=969.35   BIC=974.91
##
## Training set error measures:
##                   ME     RMSE      MAE       MPE     MAPE      MASE        ACF1
## Training set 21.3656 1384.882 1028.412 -3.630086 7.391306 0.4106839 -0.05169008
```


Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

Italy

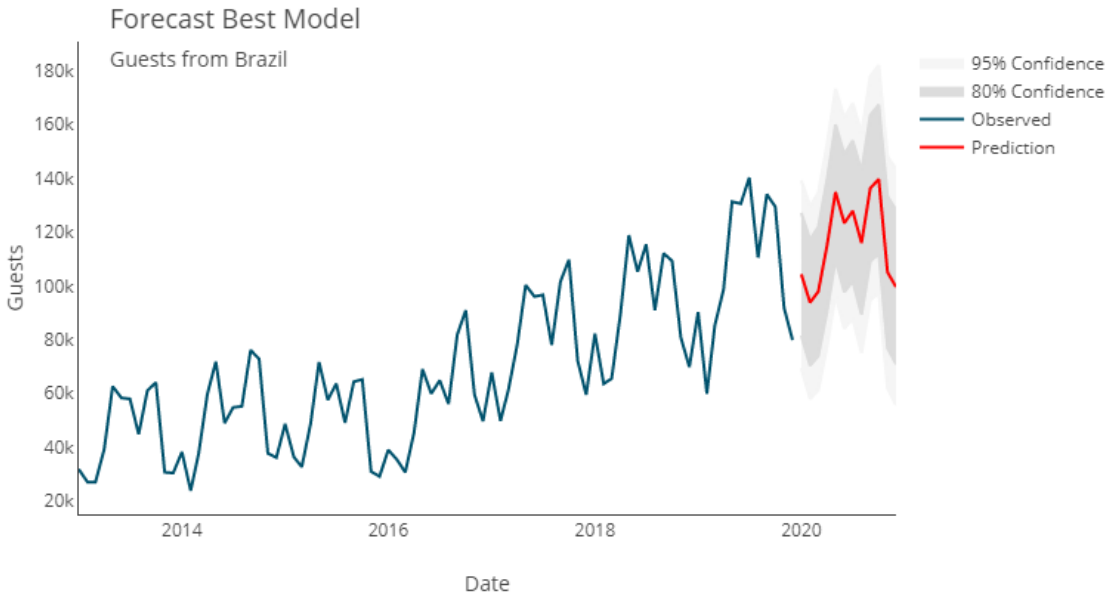
```
## Series: train_guests
## ARIMA(1,0,0)(0,1,0)[12] with drift
##
## Coefficients:
##          ar1     drift
##       0.6234  446.0635
## s.e.  0.1126  108.9742
##
## sigma^2 estimated as 14319543:  log likelihood=-530.39
## AIC=1066.78   AICc=1067.25   BIC=1072.8
##
## Training set error measures:
##                    ME    RMSE      MAE       MPE     MAPE      MASE        ACF1
## Training set 64.31732 3365.62 2251.743 -1.079464 5.420213 0.3709071 -0.07091822
```

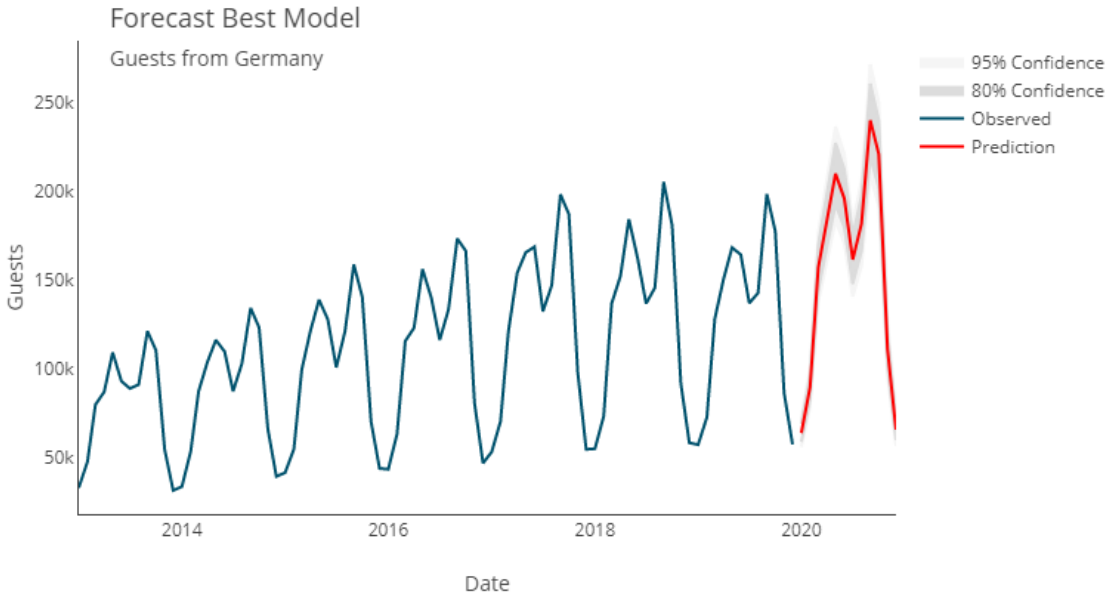
Portugal

```
## Series: train_guests
## ARIMA(0,0,0)(0,1,1)[12] with drift
##
## Coefficients:
##          sma1      drift
##       -0.4726  4540.3866
## s.e.   0.1659   255.1511
##
## sigma^2 estimated as 1.209e+09:  log likelihood=-653.64
## AIC=1313.28   AICc=1313.75   BIC=1319.3
##
## Training set error measures:
##                    ME     RMSE      MAE        MPE     MAPE      MASE
## Training set 1569.572 30922.52 22709.44 -0.1294311 3.249074 0.3893519
##                    ACF1
## Training set -0.1117644
```

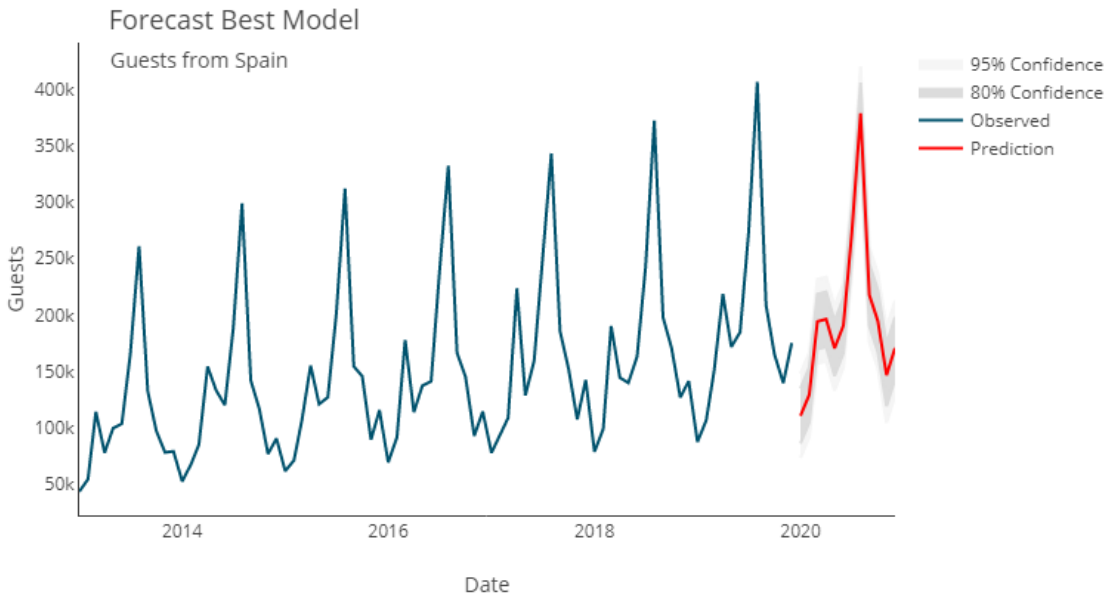

Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

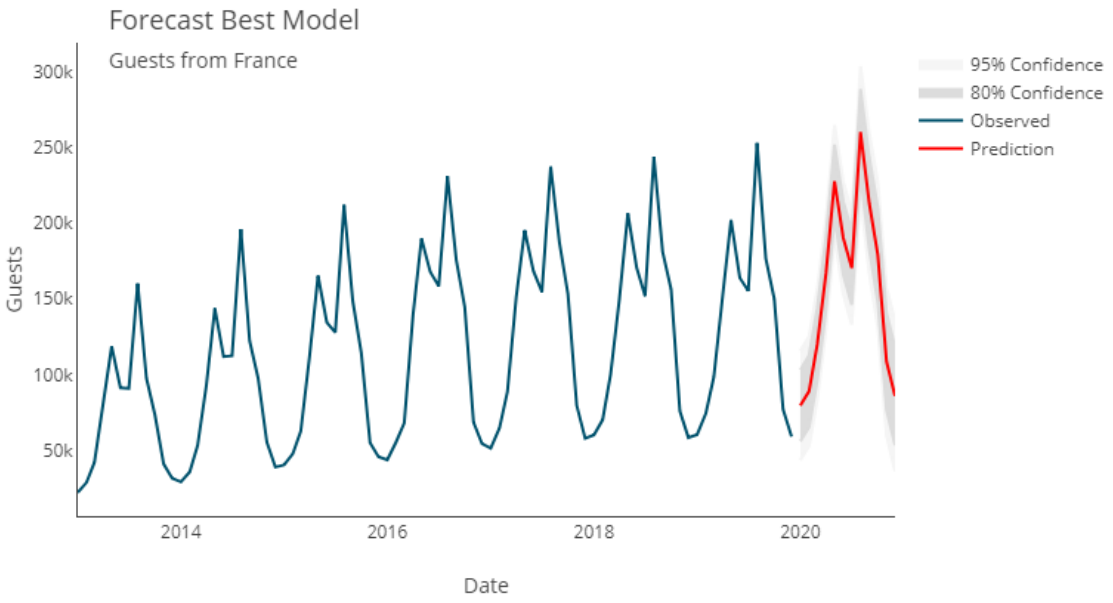
APPENDIX 14. FORECAST BEST MODEL



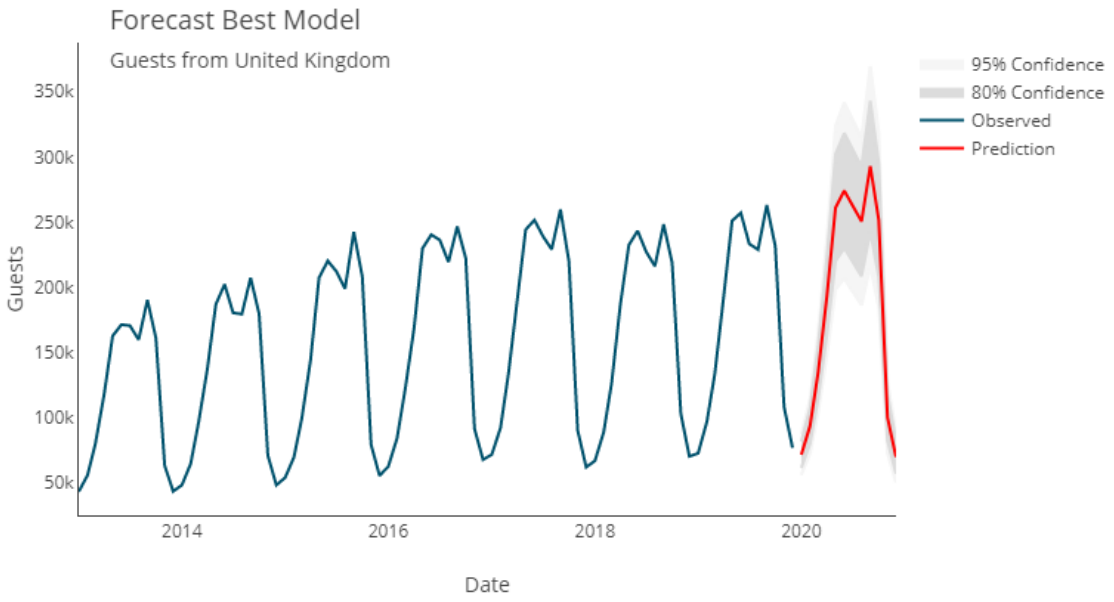


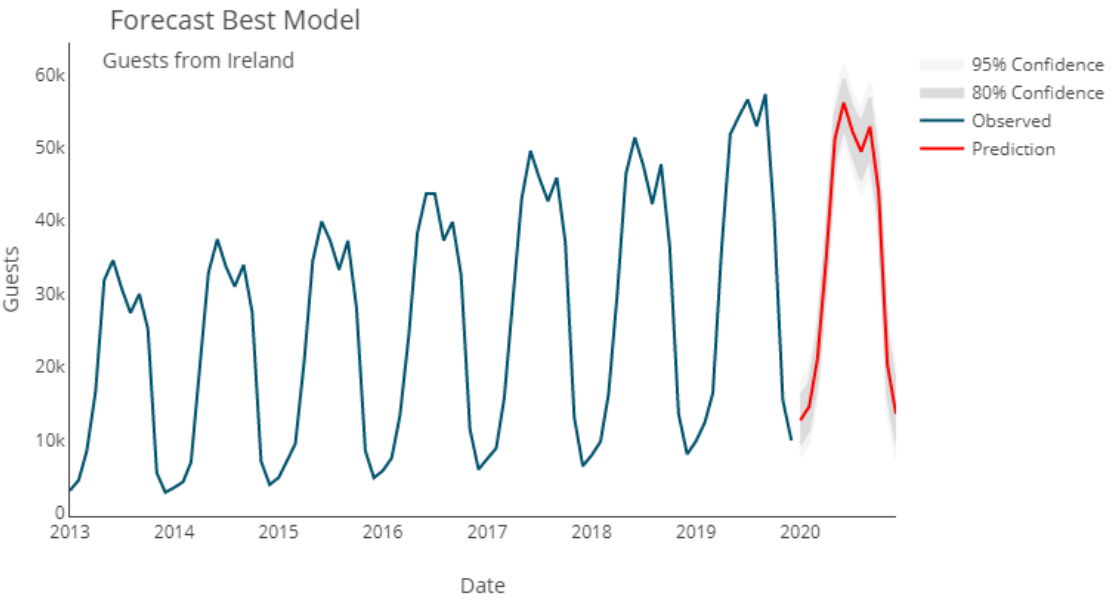
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



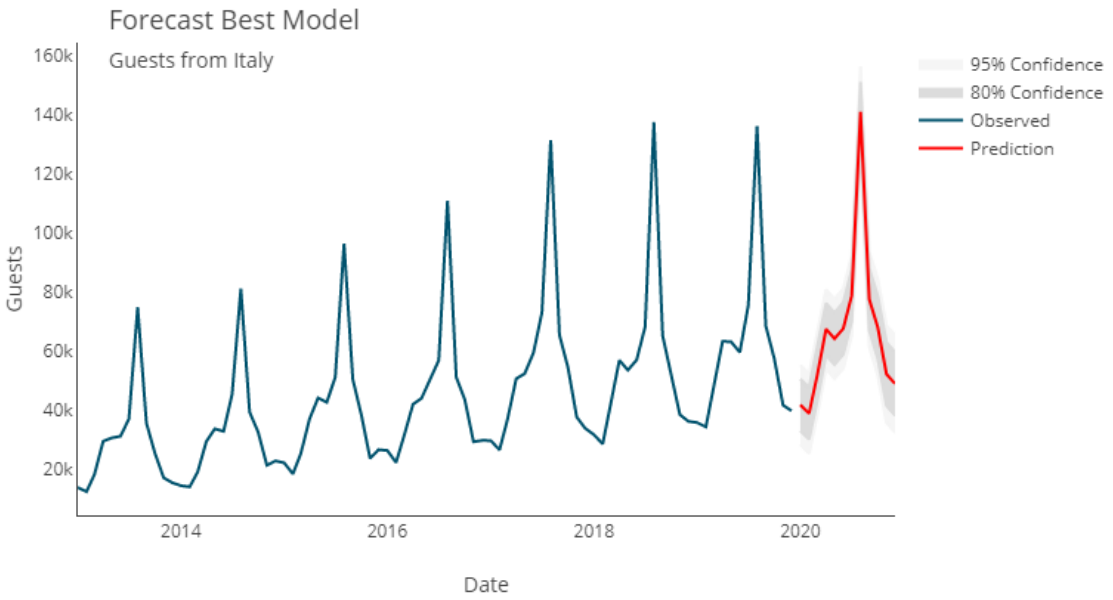


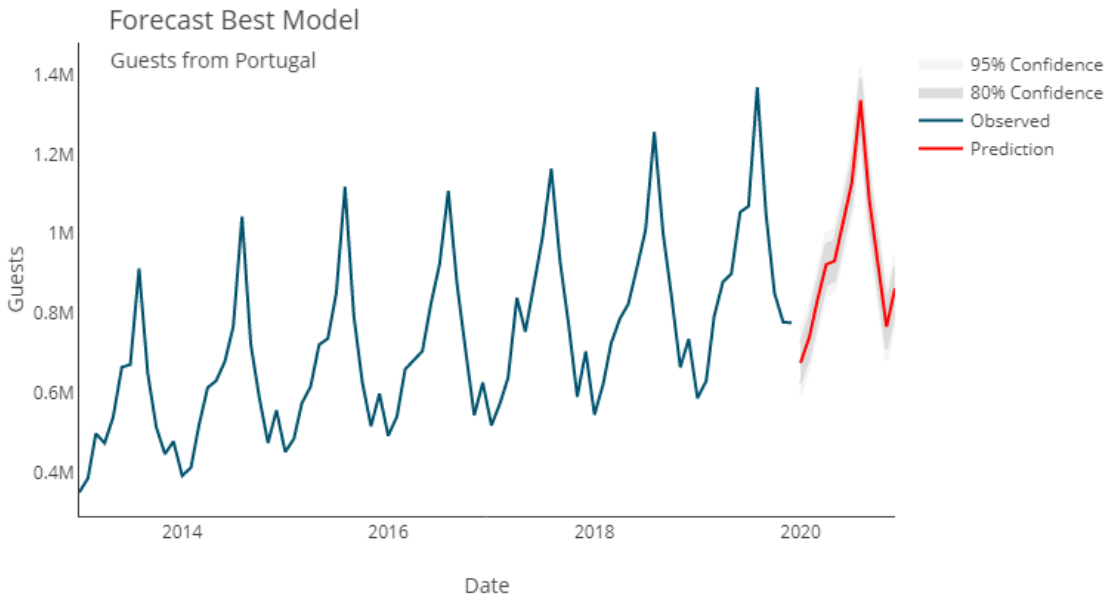
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt





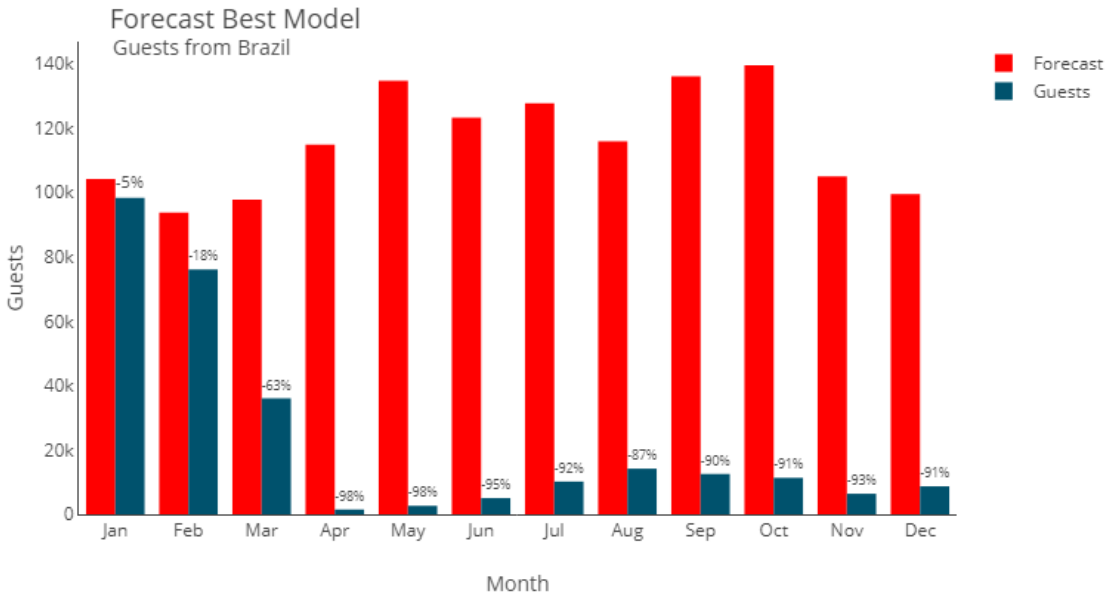
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

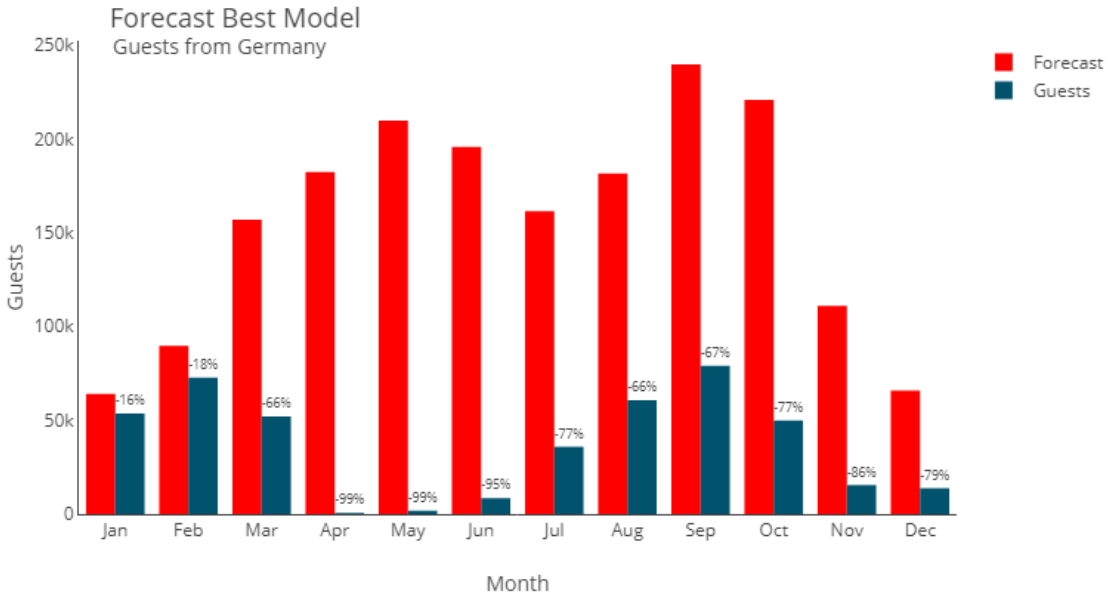




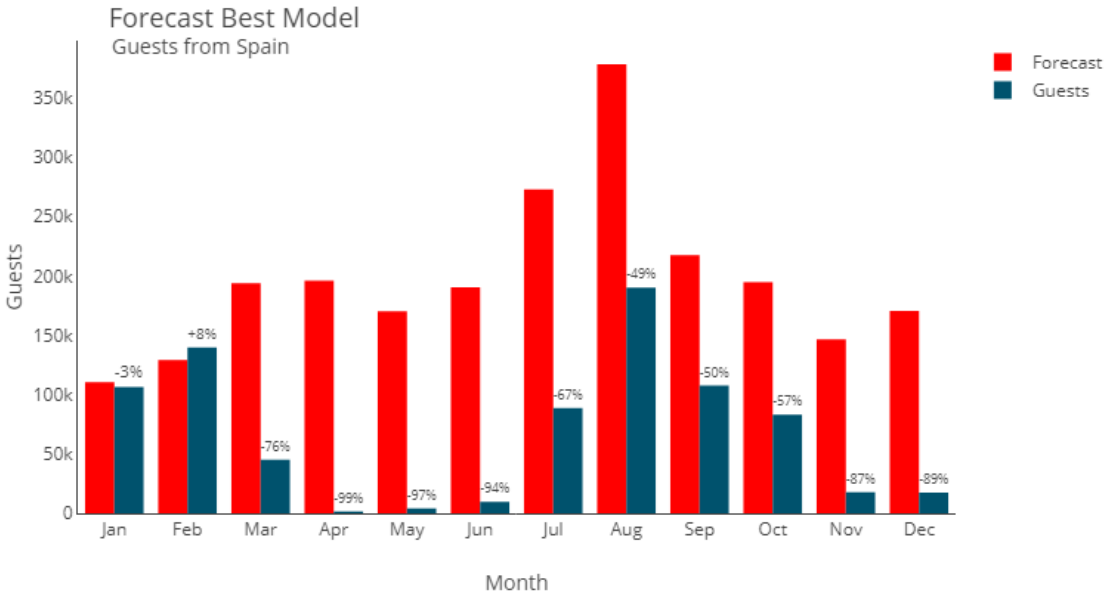
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt

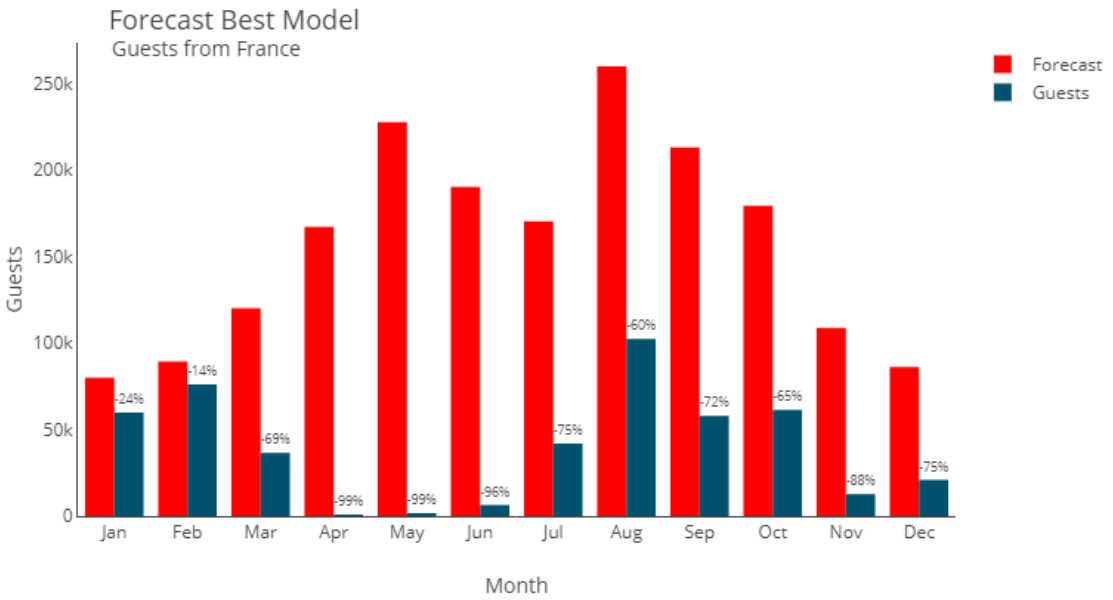
APPENDIX 15. COMPARATIVE ANALYSIS



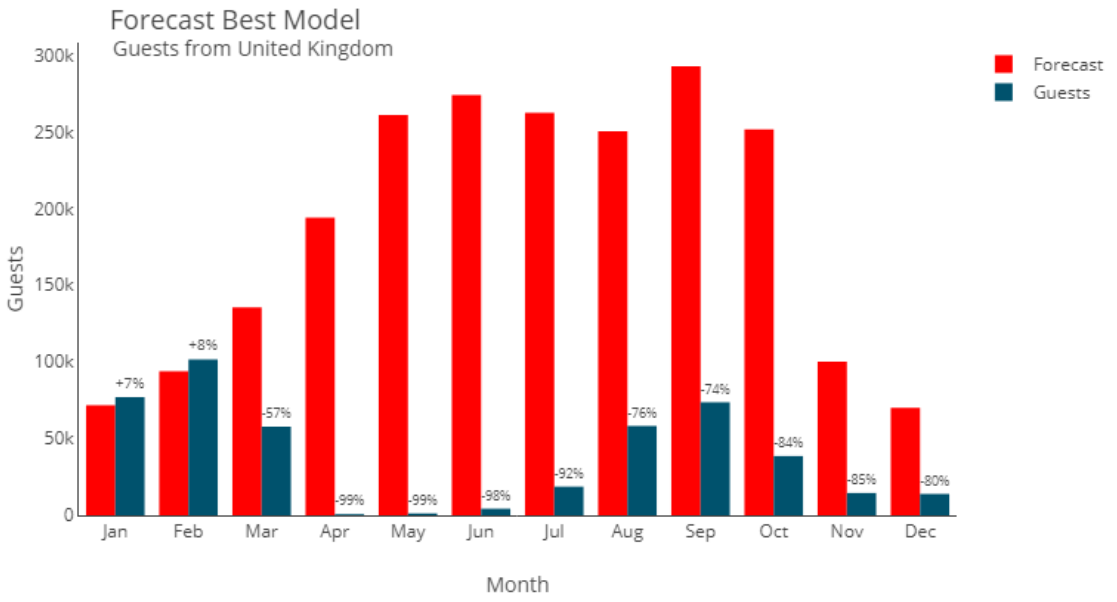


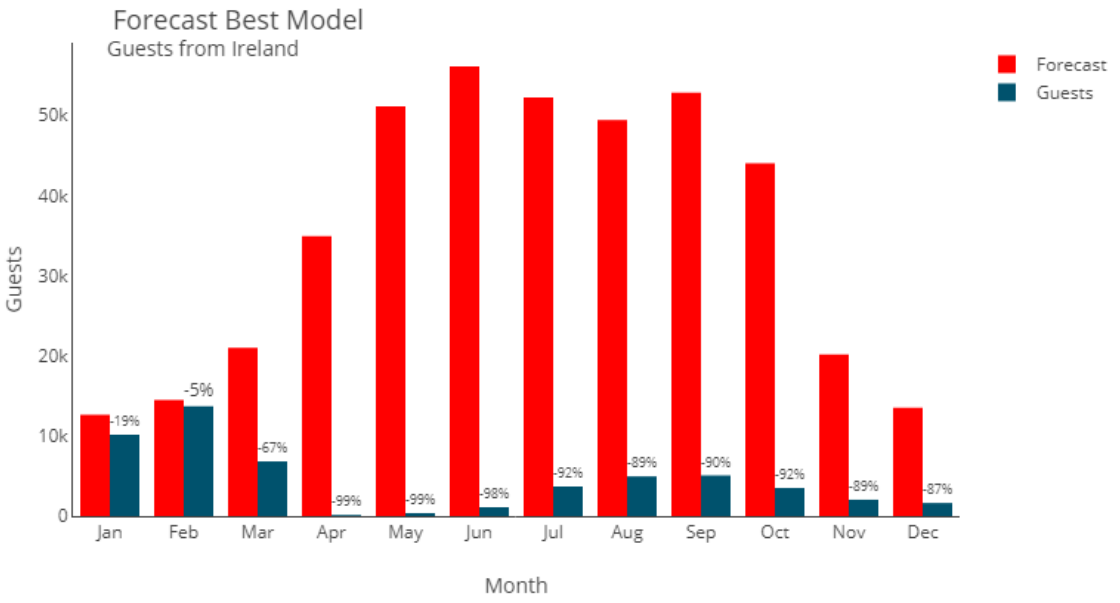
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt



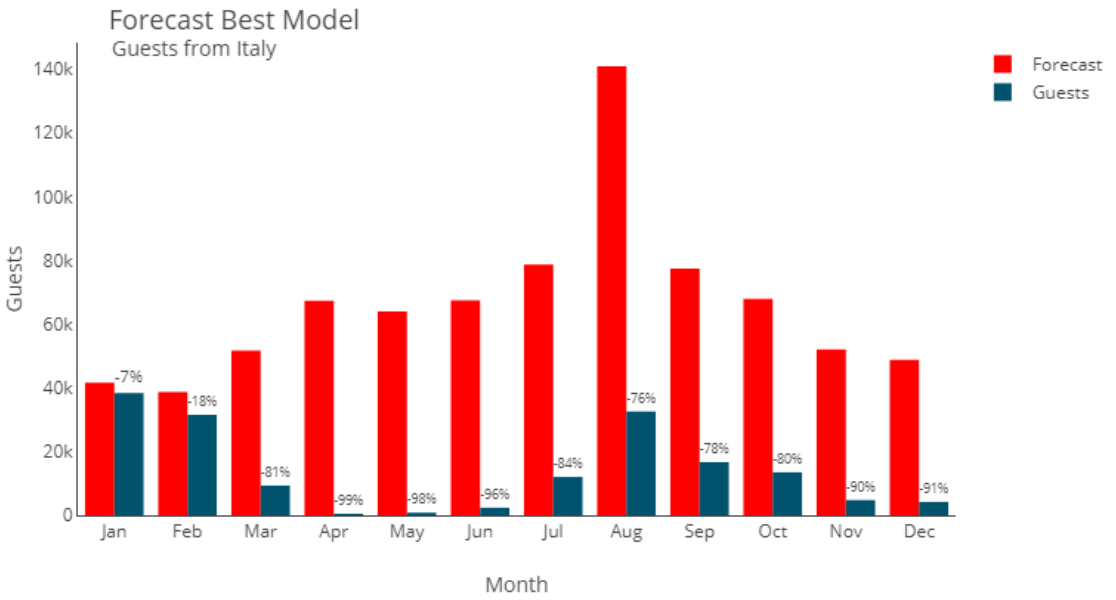


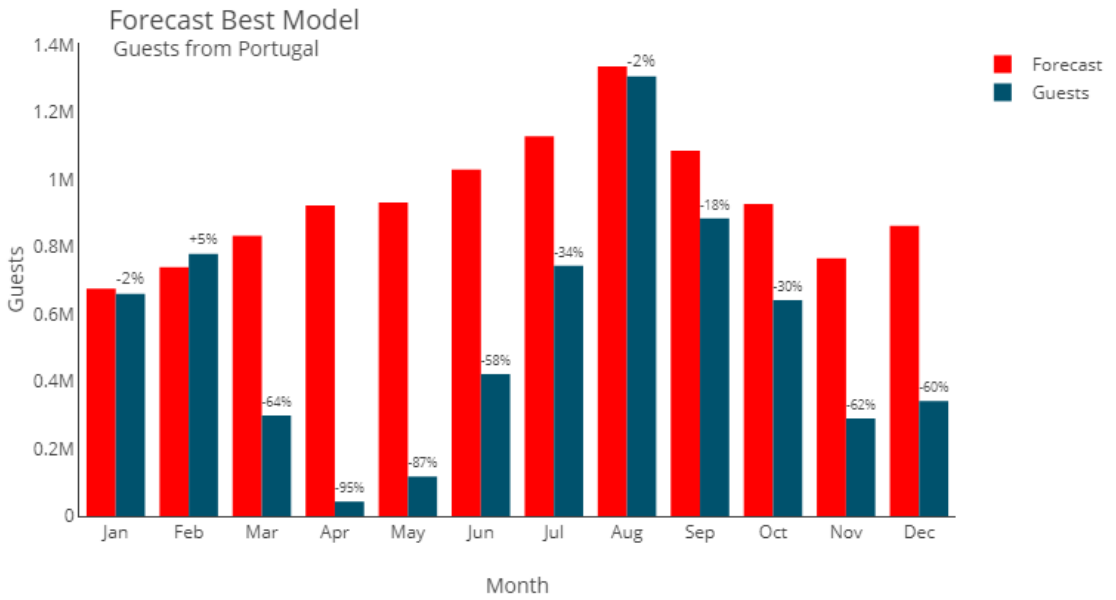
Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt





Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt





Coimbra Business School | ISCAC
Quinta Agrícola – Bencanta, 3045-601 Coimbra
Tel +351 239 802 000; E-mail: secretariado@iscac.pt; www.iscac.pt