National Tourism Promotion: An Exploratory Data Analysis

Catarina Nunes (20230083); Inês Castro (20220156); Pedro Silva (20230211)

NOVA IMS – Information Management School.

Abstract. Overview of the tourism industry in Portugal and the challenges faced, especially due to the COVID-19 pandemic.

Contents

Index of Figures

Index of Tables

1   Executive summary: Impact of COVID-19 on Portuguese Tourism and Strategic Recommendations

The tourism industry in Portugal has faced unprecedented challenges due to the COVID-19 pandemic. As consultants for the Portuguese National Tourism Board Organizations (NTBO), our objective was to analyze user-generated content, specifically TripAdvisor reviews, to gain insights into visitor patterns and understand the impact of the pandemic on tourism.

The onset of the COVID-19 pandemic in early 2020 triggered swift border closures across Europe, resulting in a staggering drop in visits to European attractions, including Portugal. The *"Rush to Reopen"* during the summer led to a brief recovery, but a second wave and associated travel restrictions caused tourism to plummet again, with a total **drop of 85.30% in visits from 2019 to 2020**. Amid these challenges, a noteworthy shift in review ratings post-COVID-19 has been observed, **with negative reviews gaining prominence**. Heightened concerns about **hygiene standards**, **crowd management**, and **waiting times** have contributed to a shift in customer sentiments. Businesses should proactively address these concerns to adapt to changing travel preferences and enhance customer satisfaction.

The pandemic has sparked a surge in **Domestic Tourism Motivations**, with locals becoming influential marketing ambassadors. Targeted campaigns and events fostering pride encourage residents to explore and endorse diverse attractions, creating a ripple effect of appreciation for national heritage.

In Portugal, a distinct travel trend has emerged, marked by a significant rise in **solo trips** and an increase in **trips with friends**. Accommodation providers should adapt to cater to the changing dynamics, organizing group activities for solo travelers and ensuring clean and safe environments. Despite challenges, Portugal's overall rating reputation appears to have improved during the pandemic. There has been an increase in '5' ratings, suggesting that positive experiences outweigh the relatively modest increase in '1' ratings.

On a global scale, **Spain emerges as an exceptionally attractive destination**, especially for a user demographic primarily composed of English-speaking countries. In fact, **33.33% of users who visited Portugal also explored Spain**, indicating an overlap in tourist preferences. The high confidence level of 79.17% underscores a strong likelihood that a user visiting Portugal would extend their exploration to Spain. This must present an opportunity for collaboration rather than competition, and Portugal-Spain travel packages can enhance the **overall tourism experience for visitors exploring the Iberian Peninsula**.

Certain Portuguese attractions, such as **Quinta da Regaleira**, **Mosteiro dos Jeronimos**, **and Torre de Belém**, exhibit a strong association, with nearly 90% likelihood of users visiting them together. Strategic initiatives like **connected tours**, **tourist cards/packages**, **online promotions**, and **local advertising** can play a big role in promoting these attractions and enhancing visitor experiences.

In Portugal, **Labor Day** significantly leads in tourist activity, with New Year's also drawing considerable numbers, indicating prime periods for tourism. The Torre de Belém is the top attraction, followed closely by the Mosteiro dos Jerónimos and Park and National Palace of Pena, highlighting focal points for tourism marketing efforts. Globally, **Christmas** and **New Year** are pivotal holidays for marketing campaigns, with a focus on couple-centric travel preferences. Christmas sees a majority of 64.6% preferring couple travel, emphasizing its intimate nature, while New Year maintains couple-centric dominance at 62.7%. Family travel slightly increases during New Year, and solo trips emerge at 7.1%, reflecting communal and individual celebrations, respectively.

In the realm of travel, **RFM (Recency, Frequency, Monetary)** analysis proves instrumental for tailoring personalized recommendations based on individual customer behaviors. Our exploration of diverse customer segments, including **"Best Travelers"**, **"Loyal Travelers"**, **"Spenders Travelers"**, and strategies for re-engaging **"Churned Cheap Travelers"**, **"Gold Travelers"**, and **"Silver Travelers"**, offers a nuanced understanding of travel patterns and preferences. Additionally, we leverage RFM analysis to delve into the behaviors of Portuguese travelers, providing targeted suggestions for reactivation. Furthermore, our detailed examination of attraction and **country similarities** enables strategic recommendations, enhancing thematic coherence and providing global and Portuguese travelers with interconnected journeys.

In conclusion, navigating the challenges brought by the pandemic requires a multifaceted approach, emphasizing domestic tourism, adapting to changing traveler behaviors, and leveraging strategic partnerships to foster tourism recovery in Portugal.

2 Business Understanding

2.1 Background & Context

The challenge faced by the Portuguese National Tourism Board Organizations (NTBO) is to comprehend the profound impact of the COVID-19 pandemic on the tourism sector. The pandemic, which has led to an 11-fold reduction in demand compared to the 2008/9 financial crisis, significantly disrupted the freedom of people's movement and consumer confidence (Tomé, 2022). This unprecedented disruption has resulted in substantial shifts in traveler behaviors, preferences, and overall engagement patterns with tourist attractions. The NTBO seeks to navigate these multifaceted challenges by leveraging an analysis focused on user-generated content, particularly social media reviews. Through this lens, we aim to unravel the evolving dynamics of visitor patterns influenced by the pandemic.

2.2 Business Problem

Our main objective is to analyze and describe the complex patterns of visitors who interact with Portuguese tourist attractions. We will take into account the time-related nuances that have been influenced by the COVID-19 pandemic. By studying the content generated by users, we aim to understand changes in their preferences, sentiments, and overall behavior. This analysis will not only provide valuable insights into the current state of tourism but will also serve as a foundation for strategic decision-making and targeted interventions by the NTBO.

2.3 Methods

In addressing our multifaceted business problem, we adopt the **CRISP-DM methodology**, employing various analytical techniques including descriptive statistics. Our objective is to capture the nuanced dimensions of visitor behaviors and preferences. To bolster the robustness and practicality of our analysis, we integrate insights from established approaches in the tourism sector and other best practices.

Focusing on the exploration of the tourism industry within the Data Science for Marketing course, our analysis centers on two primary datasets: Europe-Top100Attractions_ENG_20190101_20210821 (excel) and Holidays.csv. The EuropeTop100Attractions_ENG_20190101_20210821 dataset encompasses TripAdvisor reviews from January 1, 2019, to August 21, 2021, for the top 100 European tourist attractions, with two sheets containing reviews and attraction information. The Holidays.csv dataset augments our analysis by providing crucial temporal context related to worldwide public holidays. The project's overarching goal is to analyze visitor patterns, comprehend the impact of events such as the COVID-19 pandemic, and pinpoint opportunities for marketing initiatives.

3 Data Understanding

3.1 Exploratory Data Analysis (EDA)

Europe-Top100Attractions_ENG_20190101_20210821: Reviews

1. **Univariate EDA**

- **userLocation**: Featuring 12,613 different values, userLocation requires normalization to enhance consistency in the representation of user locations.
- **reviewLanguage**: The single value 'en' was found.
- **globalRating**: Featuring three values (4, 4.5, 5), globalRating exhibits minimal variation, indicating a consistent trend in user ratings.
- **tripType**: 5 tripe types with predominance in couples' trips.

- **reviewRating**: A preponderance of reviews with a rating of '5' suggests a positive sentiment among users, potentially influencing overall attraction perceptions.
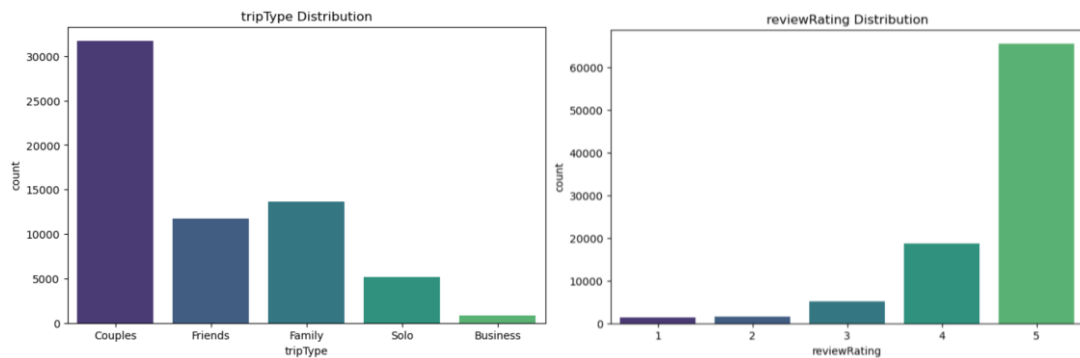
***Figure 1.** Categorical Variables distribution*

Outlier detection: In our analysis, the skewness of user contributions reveals a highly positive skew, indicating that a small percentage of users contribute a significant number of reviews, while most users contribute a limited amount. Additionally, kurtosis shows an extremely leptokurtic distribution for user contributions, suggesting a heightened likelihood of users with exceptionally high contribution counts, indicating a concentration of prolific contributors. These findings prompted further investigation into the outliers.

Fluctuations in the total number of attractions on TripAdvisor's regional ranking (sitesOnRanking) are expected at the extraction date due to extraction time, necessitating no critical analysis, and outliers require no special treatment. Despite being an outlier, the substantial interest and popularity indicated by the totalReviews for Sagrada Familia (MAG001) make it valuable for further analysis, though its significant visitation may influence overall trends. To address this, specialized models will be developed as needed to understand its specific impact. For userContributions, proposing imputation with the maximum value from valid maximum values is suggested, considering the change as insignificant. This approach allows the inclusion of the user within the frequent contributors' range without distorting the analysis.
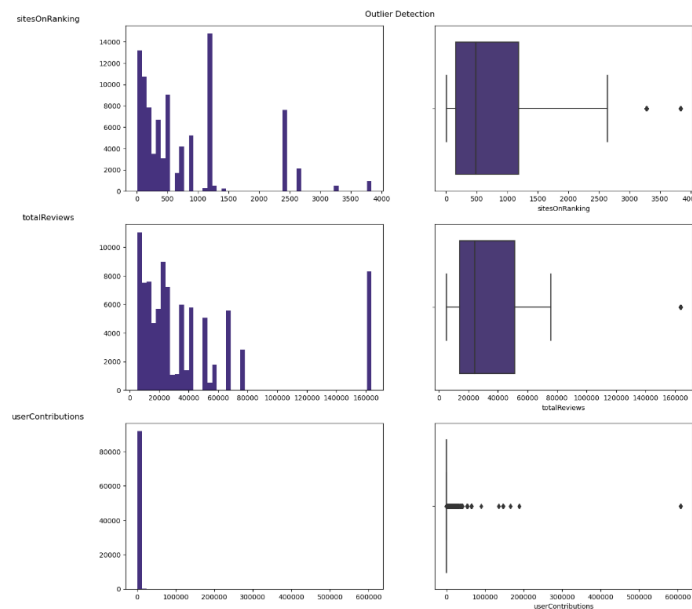
***Figure 2.** Outlier detection*

Missing values: Our main goal is to assess the correlation between attractions and reviews. Fortunately, there are no missing values in the critical columns essential for this analysis. However, for a thorough examination, we plan to compare reviews both before and after the pandemic. While the reviewWritten

column is complete, the reviewVisited column has some gaps. Regarding the userName, userLocation, and tripType columns, although it is not ideal to have missing values in these optional fields, a small number of such gaps will not significantly impact the analysis.

2. Bivariate EDA: Patterns found.

Categorical – Quantitative:

The analysis reveals a notable pattern in the distribution of review ratings associated with user contributions. Users with fewer reviews tend to give a higher percentage of poorer ratings, possibly due to sampling bias and the influence of impactful experiences. In contrast, users with higher contribution volumes provide a more balanced viewpoint, stemming from a larger and diverse sample size. This underscores the importance of considering user review history for a nuanced and representative assessment of attractions.
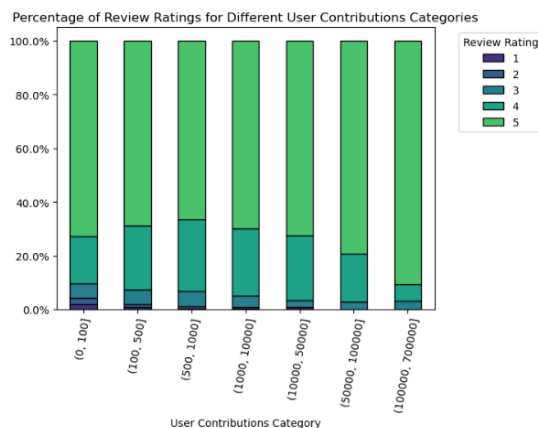
Figure 3. *Rating distribution per User Contributions Category*

Categorical-Categorical:

The analysis aimed to identify attractions with the highest visit percentages for different trip types, such as Business, Couples, Family, Friends, and Solo. To mitigate the impact of MAG001 (Sagrada Familia) as an outlier, data normalization was performed by converting values to percentages. This normalization revealed specific LocalIDs for each trip category, providing clearer insights into the proportional contributions of attractions to their respective trip types.

In reviewing attractions with the highest review ratings (total values), anomalies were identified in the LocalID data, particularly with 'u' as an unexpected identifier for Review Ratings 3 and 4. Despite irregularities, attractions MAG075 and MAG002 were identified with the most trips for Review Ratings 1 and 2, respectively, while MAG001 dominated in Review Rating 5. A normalization process was employed, demonstrating that MAG075 maintained the highest percentages for both Review Ratings 1 and 2, MAG002 led in Review Rating 3, MAG097 in Review Rating 4, and MAG038 in Review Rating 5. This investigation highlights the importance of careful data interpretation, addressing unexpected occurrences for accurate insights.
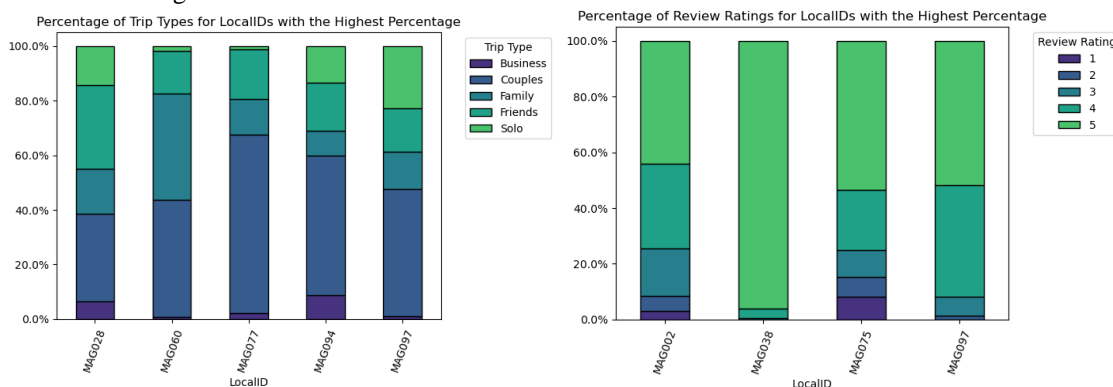
Figure 4. *Attractions with the biggest Trip type and Rating type %*

Time Series Analysis: COVID-19 effect: In our preliminary analysis comparing visit variations between 2019 and 2020, the profound impact of the COVID-19 pandemic on attraction visits is evident.

3. Multivariate EDA:

Redundancy/Duplicate rows: In this analysis, the extractionDate column is excluded from consideration. With 91,896 different values, this column reflects temporal variations in data extraction. The decision to remove it during the multivariate duplicate assessment is essential because the extraction date and time uniquely identify each data entry, posing challenges in identifying genuine duplicate rows. Additionally, given the time taken for extraction, it is plausible that some edited reviews may appear as duplicates in the dataset.

Europe-Top100Attractions_ENG_20190101_20210821: Attractions

2. Bivariate EDA: Spain, Italy, France, and the Czech Republic prominently emerge as the countries with the highest number of attractions featured in the top rankings.

4 Data Preparation

4.1 Data Cleaning and Transformation

- **Data Cleaning for the "Attractions" Sheet:**

In the "Attractions" sheet, a few data quality issues were identified and need correction. Firstly, on line 62, "Scot" should be corrected to "Scotland." Additionally, on line 44, the ISO code for Poland is incorrectly listed as "HR" and should be updated to the correct code. Furthermore, the assignment of Curaçao to Europe is incorrect, and this needs to be adjusted. Lastly, on line 77, for Italy, considering Vatican City with the Italy ISO code is suggested for better alignment.

- **Data Cleaning for the "Reviews" Sheet:**

The "Reviews" sheet has been thoroughly assessed, revealing areas that require attention. The **userLocation** column, a valuable source for exploration, presented variations in data formats, including "NY," "New York," or "United States." To normalize this column an external notebook was utilized to create an algorithm extracting ISO codes from each location. Addressing missing values, imputation with a constant value, "NULL," was employed for categorical variables. **Local IDs** were also corrected.

Outliers in **userContributions** are treated by assigning them the maximum value from valid maximum values. To handle these outliers, a deviation from the standard 1.5 multiplier in the interquartile range (IQR) was chosen, opting for a multiplier of 3 due to the highly positively skewed distribution. This choice, typically used for extreme outliers, aims to maintain a broader range for evaluating frequent contributors. The impact analysis revealed that only 4% of users were affected by this decision, making the 3-multiplier reasonable.

For duplicate identification, the focus is on the **review full text**. Duplicates with the same reviews from the same visit are retained, keeping the most recent one. Additionally, it was observed that longer text created new rows, leading to the aggregation of these rows.

- **Data Preparation for "Holidays" Sheet:**

In the "Holidays" sheet, a function has been created to normalize the **normalizeHolidayName** column by translating holiday names into English. Missing values in the ISO column have been addressed, shortening the range if necessary.

4.2   Feature Engineering: Data Integration

Merging operations have been conducted to enrich the dataset. This involves merging userLocClean with reviews to capture the user ISO code, merging reviews with attractions to collect additional attraction information, and combining this with holiday data. The result 'New Variable list' is present in the Appendix.

5   Modeling and Key Findings

5.1   Comparison Before and After Pandemic:
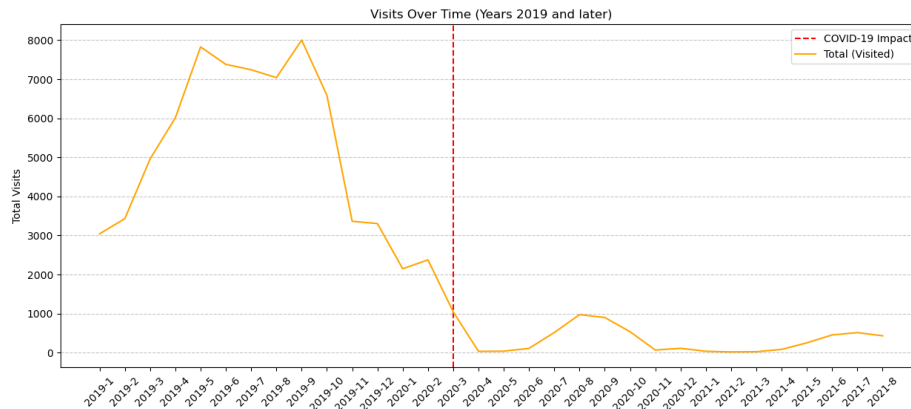
Figure 5. *Total Visits to EU Attractions, 2019 and later*

This analysis delves into the dynamic relationship between the pandemic's progression and its impact on attraction visitation patterns. The red line in March 2020 denotes the start of the lockdowns in Europe.
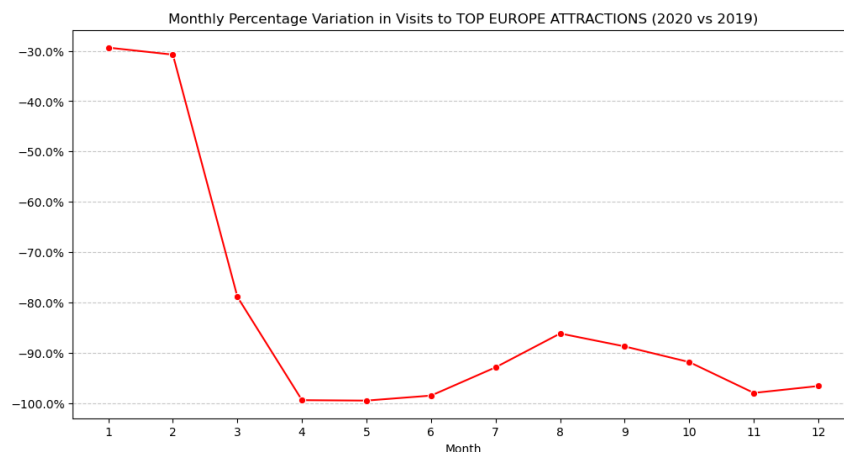
Figure 6. *Variation in Visits to EU Attractions (2020 vs 2019)*

As COVID-19 cases began surfacing in Europe early in 2020, countries swiftly implemented border closures in March 2020:

"*As a first response to the COVID-19 outbreak in the EU and to slow down the transmission of the virus, EU leaders agreed on 17 March 2020 on a coordinated temporary restriction of non-essential travel to the EU, which applied until 30 June 2020.*" - **European Council.**

The subsequent visual representation illustrates a sharp decline in visits during this period. With the onset of the summer of 2020, borders cautiously reopened, leading to a modest improvement in visitation, a phenomenon aptly termed the ***"Rush to Reopen"*** by Franco (2020), as reported by the New York Times. However, as the second wave of COVID-19 cases surged in many European countries in mid-August, albeit

with low initial case numbers, the growth rate escalated rapidly, resulting in a new decrease in attraction visits during that timeframe (Franco, 2020).

5.2   Geographic Visualization:
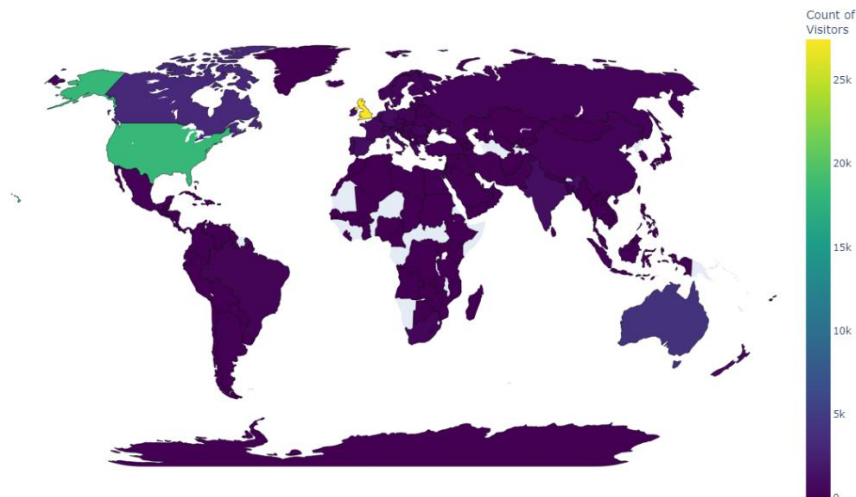
Figure 7. Visual representation of visitor origins.

In this phase, it is noticeable that our dataset is '**language biased**'. Since the reviews were all in English, most of the reviews came from people who are from English-speaking countries.

The graph below represents the percentage change in the number of visitors from the top 10 countries with the highest weight (contributing the most to the total number of visitors.), comparing the years 2019 and 2020. It visually shows the impact of the pandemic on tourism from these top-contributing countries. This helps to identify the countries that were most affected by the pandemic in terms of a decline in visitor numbers, considering their substantial impact on overall tourism. It provides a visual representation of the changes in tourism patterns for these key contributor countries during the pandemic.
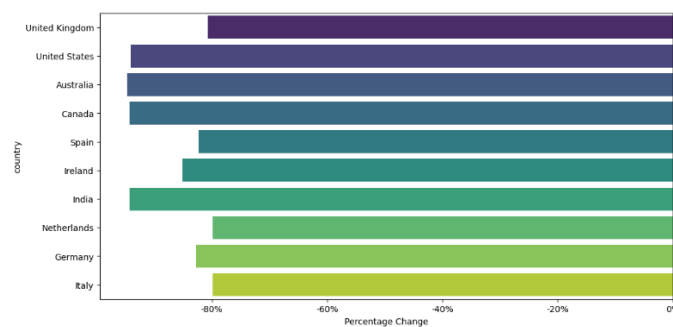
Figure 8. TOP Origin countries and their % visit drops (2019 to 2020)

- **Domestic Tourism:** Not surprisingly, we observe a significant increase in domestic tourism, suggesting that individuals are exploring more of their own countries due to border restrictions.
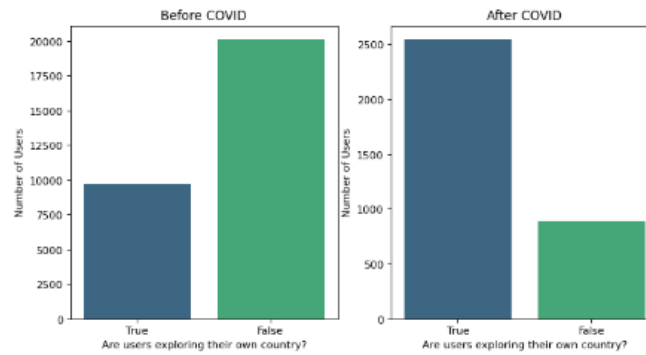
Figure 9. *Domestic Tourism Evolution*

- **Portugal Enthusiasts: Unveiling Their Origins**
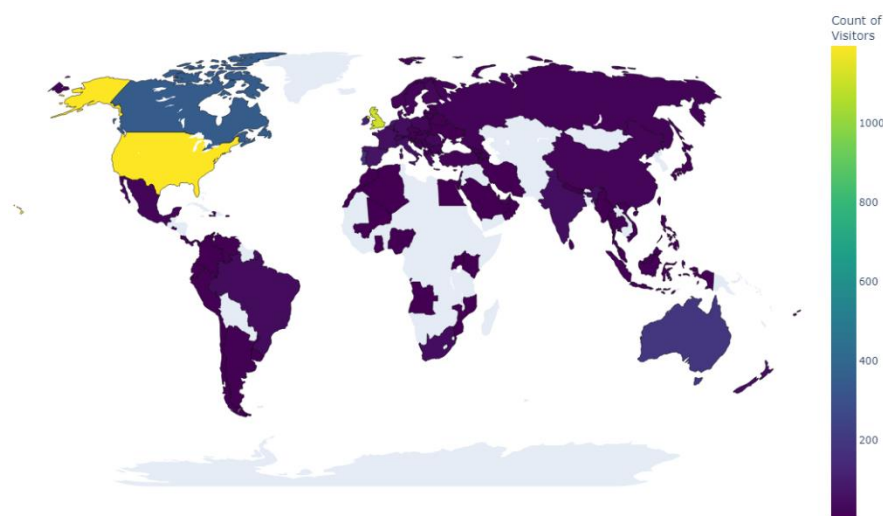
Figure 10*. Visual representation of visitors of Portugal origins.*

The top countries contributing to tourism in Portugal are led by the United States with 1194 visits, followed closely by the United Kingdom at 1109 visits. Canada takes the third spot with 347 visits, while Australia, Portugal itself, and Ireland secure positions four to six, each contributing significantly to the Portuguese tourism landscape.
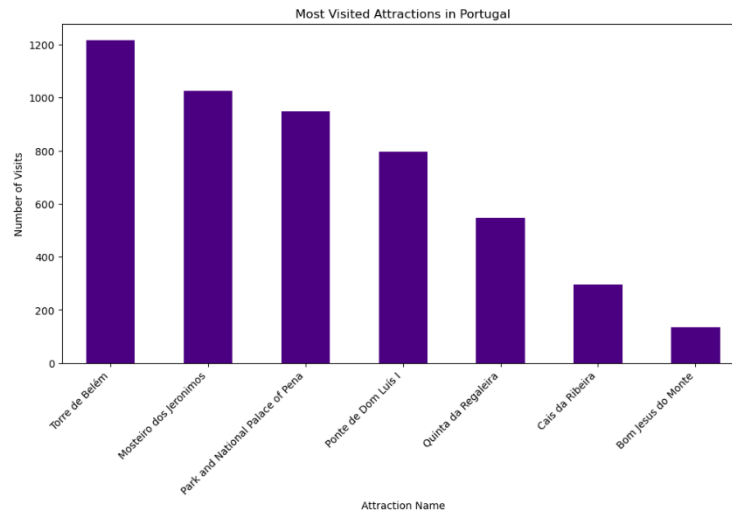
Figure 11. *Top Attractions in Portugal*

This figure represents the most popular attractions in Portugal since 2019.

Tied to the rest of Europe, Portugal did not escape unscathed from the impact of COVID-19 in early 2020. Although it initially remained slightly above, from March 2020 onward, it also had to adhere to EU policies, leading both trends to follow a similar path until the end of the year:
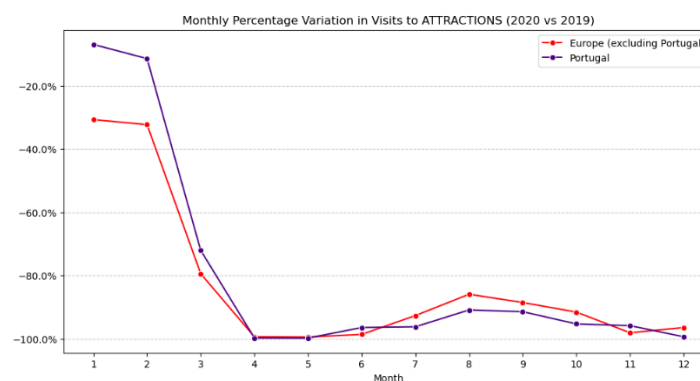
Figure 12. *Monthly Percentage Variation in Visits (2020 vs 2019) – Comparing Portugal and the rest of Europe.*

5.3   Type of Trip, Reviews, and COVID-19 Impact:

- **Transformations in Travel Profiles:**

In analyzing this graph, our goal was to pinpoint changes in travel patterns, particularly focusing on the percentage distribution of various trip types. We noted a slight drop in the percentage of trips with **friends**, linked to factors like nightclub closures and reduced social gatherings during the ongoing pandemic. Conversely, there's a noticeable rise in the percentage of **family trips**, suggesting a shift towards more intimate travel experiences. **Solo trips** also saw a slight increase, possibly reflecting a preference for individual travel due to safety concerns or personal choices. Regarding **business travel**, there's a percentage decrease, possibly driven by the widespread adoption of remote work during the COVID-19 pandemic. The acknowledgment of health risks associated with travel further contributed to the decline in business-related trips.
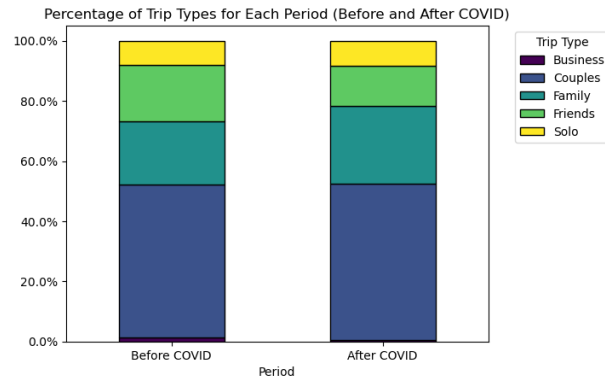
Figure 13. *Var % of Trip Types (Before Covid and After Covid)*

- **Transformations in Review Ratings:**

We've observed a shift in review ratings since the COVID-19 pandemic, particularly with **negative reviews gaining prominence**. This change is likely tied to heightened concerns and reflects a shift in customer sentiments post-COVID-19. The increase in negative reviews can be attributed to elevated concerns about various aspects of the travel experience, such as **hygiene standards**, **crowd management**, or **shorter waiting times**. It's possible that travelers now prioritize safety and efficiency, and perceived shortcomings in these areas contribute to a higher incidence of negative feedback. As expectations continue to evolve globally, businesses should proactively address these concerns to enhance customer satisfaction and adapt to the changing landscape of travel preferences.
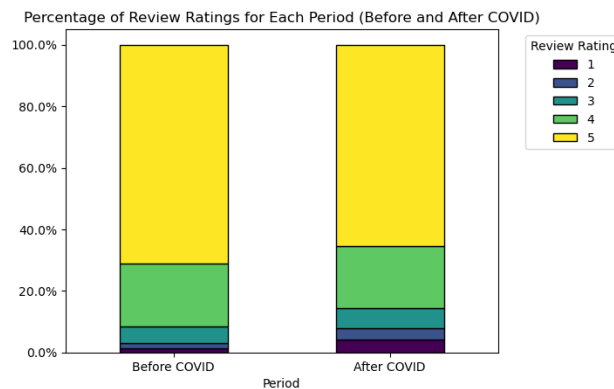
Figure 14. *Var % of Reviews (Before Covid and After Covid)*

- **COVID-19 Impact on Travel and Reputation in Portugal**
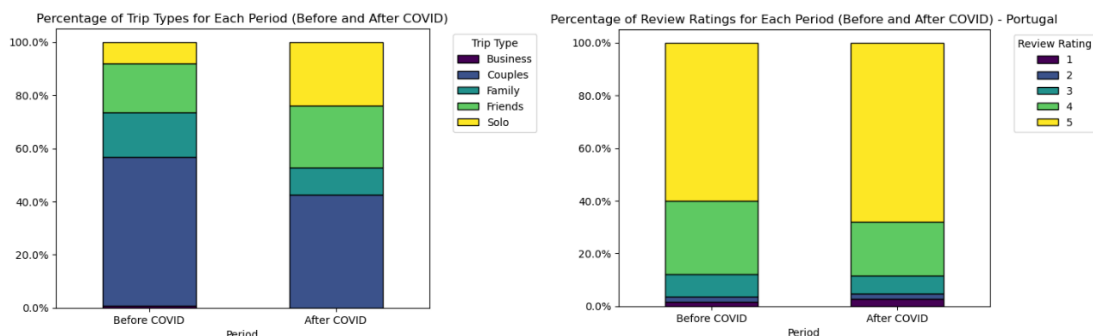
Figure 15. *COVID-19 Impact on Travel and Reputation in Portugal*

In Portugal, the analysis reveals a compelling trend marked by a significant **rise in solo trips**, aligning with the country's reputation as a top destination for solo travelers. The increased preference for solo exploration is complemented by a notable uptick in trips with **friends**, indicating a flourishing social dynamic among tourists. Concurrently, there is a reduction in other travel categories, signaling a distinct **shift in the tourist profile**. These insights are valuable indicators for tourism companies, revealing evolving traveler preferences and behavior.

Also, there has been an increase in extreme reviews during the analyzed period, particularly with a significant surge in **'5' ratings**. While there is a mild uptick in **'1' ratings**, it may not be as pronounced. On average, the pandemic appears to have played in **favor of Portugal's reputation** in terms of rating. The heightened positive experiences, reflected in increased '5' ratings, seem to outweigh the relatively modest increase in '1' ratings. This suggests that, on the whole, the impact of the pandemic has contributed positively to Portugal's overall rating reputation.

5.4   Frequent Itemset Associations (Market Basket Analysis):

- **Patterns of Countries frequently visited together (users with more than 8 reviews):**

1. **Portugal -> Spain:** About 33.33% of users who visited Portugal also visited Spain. The confidence of 79.17% indicates a strong likelihood that a user visiting Portugal also explored Spain. The lift of 1.22 suggests a positive association, making the probability of visiting Spain 1.22 times higher when a user visits Portugal compared to the overall probability. The proximity shared cultural ties, and travel convenience between Portugal and Spain may drive this strong association. It's possible to capitalize on this association by promoting travel packages that include both Portugal and Spain. Emphasize the seamless travel experience between the two countries.
2. **(Italy, Scotland) -> (England):** Users who visited both Italy and Scotland have an 87.50% likelihood of also visiting England. The lift of 2.27 indicates a strong positive association, making the probability of visiting England 2.27 times higher when users visit both Italy and Scotland compared to the overall probability.

Other important results (Values in Table Appendix):
3. **Poland -> Italy;**
4. **Croatia -> Spain;**
5. **Scotland -> England;**

- **Patterns of attractions frequently visited together (users with more than 7 reviews):**

1. **(MAG004: Casa Batllo) -> (MAG001: Basilica of the Sagrada Familia):** Users who visited Casa Batllo (MAG004) have a 95.65% likelihood of also visiting Basilica of the Sagrada Familia (MAG001). The lift of 2.49 indicates a strong positive association, making the probability of visiting the Basilica of the Sagrada Familia 2.49 times higher when users visit Casa Batllo compared to the overall probability.
2. **(MAG012: Real Alcazar de Sevilla) -> (MAG019: Catedral de Sevilla):** Users who visited (MAG012) have a 91.30% likelihood of also visiting (MAG019), and the lift value of 3.52 indicates a positive correlation.
3. **(MAG023: Casa Milà - La Pedrera) -> (MAG001: Basilica of the Sagrada Familia):** Users who visited Casa Milà - La Pedrera (MAG023) have an 87.50% chance of also visiting Basilica of the Sagrada Familia (MAG001), and the lift value of 2.28 indicates a positive correlation.

Other important results (Values in Table Appendix):

4. **(MAG002: Parc Guell) -> (MAG001: Basilica of the Sagrada Familia);**
5. **(MAG019: Catedral de Sevilla) -> (MAG012: Real Alcazar de Sevilla);**
6. **(MAG001: Basilica of the Sagrada Familia) -> (MAG002: Parc Guell).**

- **Portuguese Attractions (users with more than 7 reviews):**

While Spain dominates the Europe tourism panorama, our focus is on analyzing the patterns within Portugal. To gain insights into Portuguese attractions, we filtered the dataset specifically for reviews related to Portugal.

1. **Quinta da Regaleira (MAG047) and Mosteiro dos Jeronimos (MAG014):** About 22.86% of users who visited Quinta da Regaleira also explored Mosteiro dos Jeronimos. A high confidence of 88.89% indicates a strong likelihood of users visiting Mosteiro dos Jeronimos after Quinta da Regaleira. The lift of 1.41 suggests a positive association, making the probability of visiting Mosteiro dos Jeronimos 1.41 times higher when visiting Quinta da Regaleira compared to the overall probability.
2. **Quinta da Regaleira (MAG047) and Torre de Belém (MAG010):** Approximately 22.86% of users who visited Quinta da Regaleira also explored Torre de Belém. The high confidence of 88.89% indicates a strong likelihood of users visiting Torre de Belém after Quinta da Regaleira. The lift of 1.30 suggests a positive association, making the probability of visiting Torre de Belém 1.30 times higher when visiting Quinta da Regaleira compared to the overall probability.

Other important results (Values in Table Appendix):
3. **(MAG047: Quinta da Regaleira, MAG014: Mosteiro dos Jeronimos) -> (MAG010: Torre de Belém);**
4. **(MAG010: Torre de Belém, MAG047: Quinta da Regaleira) -> (MAG014: Mosteiro dos Jeronimos);**
5. **(MAG032: Park and National Palace of Pena, MAG014: Mosteiro dos Jeronimos) -> (MAG010: Torre de Belém);**

Tour bundles or promotional packages incorporating Quinta da Regaleira, Mosteiro dos Jerónimos, and Torre de Belém are supported by strong associations. The promotion of cross-visitation, achieved through strategically placing marketing materials or offering ticket packages at these pivotal attractions, stimulates visitors to comprehensively explore the interconnected sites. Exploring online promotional opportunities, such as joint ticket discounts or exclusive online offers, enhances appeal, further incentivizing visits to these attractions in a synergistic combination.

5.5   Text Mining for Visitor Sentiments

Portuguese Travelers

Figure 16. *Text mining for Portuguese Travelers*

In our text mining analysis focused on Portuguese travelers, a **consistent pattern** emerged in their expressions, emphasizing terms such as "beautiful," "amazing," and "nice." Concurrently, their discourse frequently featured references to "city," "palace," "history," and "architecture." These linguistic tendencies collectively convey a **positive sentiment** and an evident **appreciation for cultural and aesthetic aspects**. Moreover, the inclusion of terms like "city," "palace," "history," and "architecture" signifies a pronounced interest in cultural and historical dimensions. Portuguese travelers appear to be drawn to destinations that offer rich historical narratives, impressive architectural structures, and vibrant urban environments.

Portugal Attractions

Figure 17. *Text mining for Portugal*

The prevalence of positive sentiments in the reviews suggests a high level of satisfaction and enjoyment among visitors to Portuguese attractions. This positive language reflects the favorable experiences and impressions that travelers have during their visits, contributing to a positive overall perception of the attractions.

The inclusion of keywords like "tour" indicates a preference for **guided experiences**, the repeated mention of "city" suggests a focus on **urban experiences**, emphasizing the appeal of cityscapes like Lisbon and Porto, that are also mentioned. "palace" implies an interest in **historical and architectural landmarks**, and "river" could be linked to attractions situated along rivers (ex. Ribeira do Porto).

The use of "interesting" further underscores a desire for unique and engaging experiences. Visitors to Portuguese attractions seek not only beauty but also cultural richness and intriguing aspects that make their travels memorable.

5.6   Visitor Patterns During Holidays

The data indicates that New Year and Christmas stand out as the most significant holidays, with occurrences substantially higher than others, suggesting these times are key for launching major marketing campaigns. Labor Day, All Saints, and Workers Day exhibit moderate but significant occurrence frequencies and should be considered for campaign focuses.

Holidays with fewer occurrences, such as Halloween and Toussaint, may represent opportunities for niche marketing efforts and signals more specialized observances.
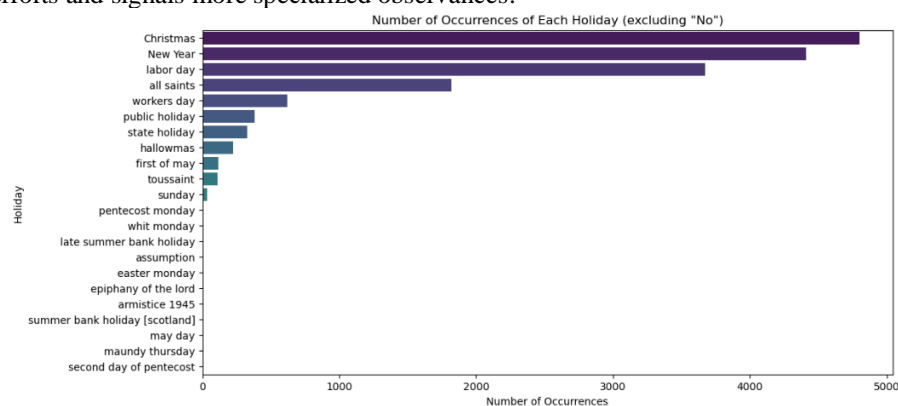
Figure 18. *Occurrences of each Holiday*

- **Type of Trip during Christmas and New Year**

During Christmas, a significant **majority of 64.6%** prefer traveling as **couples**, emphasizing the holiday's association with intimate and romantic celebrations. **Families** contribute to **18.9% of travel**, showcasing a strong tradition of familial gatherings. Travel with friends is less common at 11.2%, indicating a lesser inclination for group excursions during the festive period.

For the **New Year**, couple-centric travels slightly decline to 62.7%, **remaining the dominant mode** of celebration. Family travel sees a marginal increase to 15.5%, and travel with **friends rises to 13.6%,** suggesting a shift towards more communal celebrations for the start of the year. Solo trips, at 7.1%, emerge

uniquely in the New Year data, indicating a subset of the population seeking solitary reflections or personal adventures as they transition into a new year.

The contrast in solo travel between the two holidays provides a nuanced understanding of the holidays' social contexts—Christmas as a time for shared experiences and New Year as an occasion that also accommodates individual pursuits.
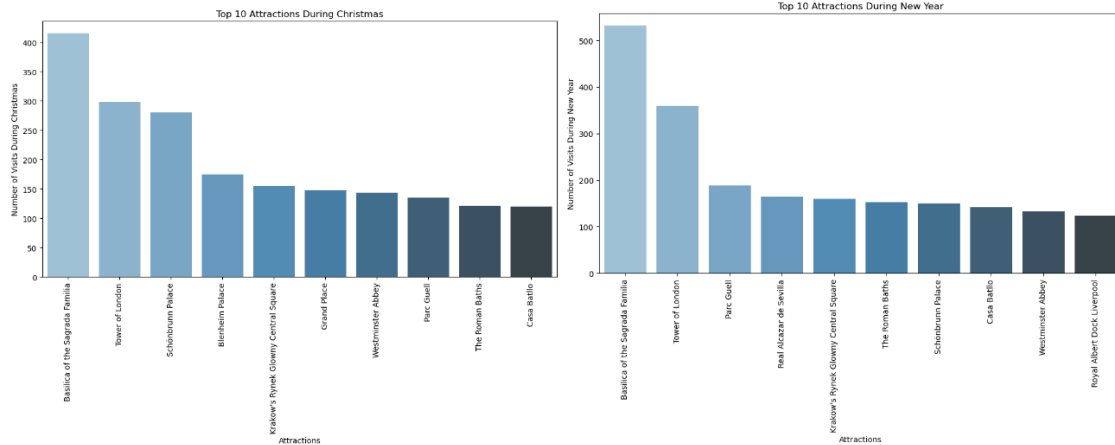
Figure 19. Top Attractions during Christmas (left) and New Year (right)

5.7   Seasonal Trends in Visitor Behavior

Recognizing that each season harbours distinct preferences and attractions, we aim to provide our clients with meticulously tailored travel plans that align with the offerings of specific times of the year. Seasonal itineraries allow us to curate journeys that optimize the weather, cultural events, and natural beauty characteristics of each season. By acknowledging that traveller's seek different experiences and activities based on the time of year, we endeavour to personalize our offerings.

Winter: Lovers of the winter trips

- **Solo, friends, couples, family or business:**

Winter beckons travellers' of various preferences, encompassing solo adventurers, friends, couples, families, and even business enthusiasts. An analysis reveals that **couples** dominate the winter journey landscape, constituting 50% of trips. The enchanting winter scenery, coupled with a romantic ambiance, creates an ideal setting for intimate getaways. **Families**, comprising 22% of winter trips, find this season opportune for festive reunions and creating cherished holiday memories. Additionally, 17% of itineraries cater to **friend groups** seeking shared experiences in the winter wonderland.

- **Where do they come from?**
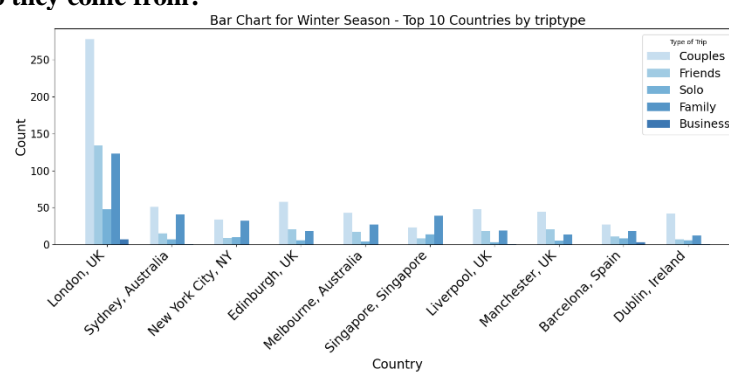
Figure 20. Winter Season characterization

- **Winter Favorites Countries to visit**

Delving into the winter travel preferences reveals that Spain, England, Scotland, and Italy reign supreme as favored destinations. Spain's mild winter climate, England's festive ambiance, Scotland's picturesque landscapes, and Italy's rich cultural tapestry make these countries stand out for winter escapades.
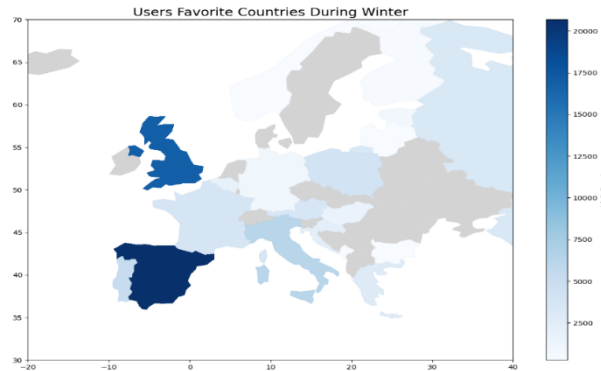
***Figure 21.** Most visited Countries in Winter*

5.8   RFM (Recency, Frequency, Monetary) & Similarities Analysis

In the travel industry, the application of RFM (Recency, Frequency, Monetary) analysis emerges as a pivotal strategy for curating personalized travel recommendations based on individual customer behaviors.

We analyze customers' recent travel activity through three key dimensions: **Recency** ('R') looks at how recently they explored new destinations, assessed by the **'reviewVisited'** variable. **Frequency** ('F') considers how often customers travel, using the **'Country'** variable to gauge diversity and regularity. **Monetary** factors ('M') traditionally involve direct spending, but lacking specific data, we cleverly use **'reviewRating'** as a satisfaction substitute. To enhance segmentation, we sum up 'reviewRating' instead of averaging, giving a more nuanced view of customers' satisfaction levels and potential contributions to the overall travel experience.
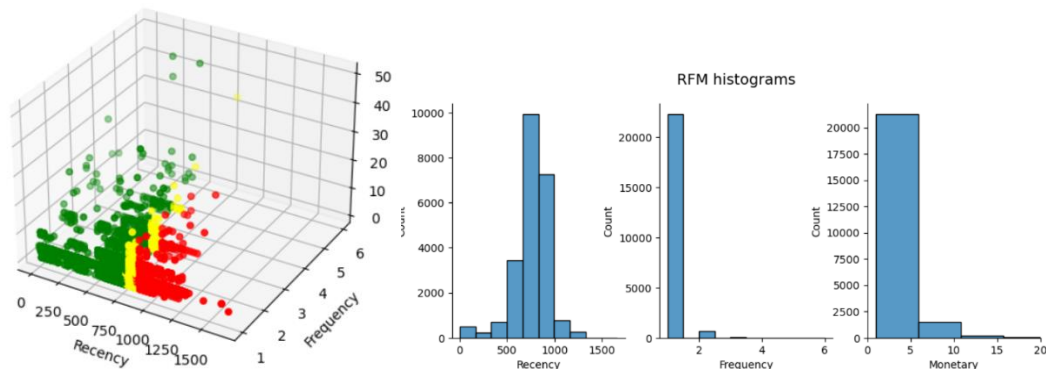
***Figure 22.** RFM vizualization Plots*

Analyzing the 3D scatter plot of RFM values reveals interesting patterns. While most customers travel infrequently, a notable group engages in rare yet recent and satisfying travel experiences. Looking at the histograms, we see that the majority of travelers had their last journey within the past 500 to 100 days (recency). **Many customers travel only once** (frequency), reflected in the histograms. In terms of spending (monetary), most fall within the 0 to 5 and 0 to 10 ranges, aligning with the fact that many customers have only taken one trip with ratings capped at 5.

In order to understand what kind of travelers we have, we created diverse segments:

1. **The "Best Travelers"**

This segment represents our best customers, characterized by an RFM score of 144, **who travelled most recently, most frequently, and gave better reviews.** Predominantly hailing from England, the United States, Australia, Canada, the Netherlands, and Israel, it is characterized by **couples and friends** with a shared love for exploration. Their favored destinations encompass Spain, England, Italy, Portugal, and Scotland. Interestingly, they often choose to travel **during non-holiday** periods but show a distinctive preference for holiday journeys, particularly during the New Year. Their top attractions include Canal Grande in Italy, the Tower of Hercules in Spain, Blenheim Palace in England, Parc Güell in Spain, the Palace of Versailles in France, and Schönbrunn Palace and Schönbrunn Gardens in Austria. This selection aligns seamlessly with their travel preferences. In order to keep them the following suggestions will be applied:

- Personalized VIP Experiences;
- Exclusive Destination Insights;
- Preferred Accommodation Packages;
- Advance Booking Privileges.

2. The" Loyal Travelers"

Our loyal travellers, identified by a x4x RFM score, underscore a **high frequency of travel.** They are primarily originated from **England and Lithuania**, with a notable preference for either **solo** trips or **couples'** escapades. Their favourite destinations predominantly include Spain, Italy, and Austria. Delving into their preferred attractions, the list comprises iconic landmarks, like Canal Grande, Tower of Hercules, Blenheim Palace, Parc Güell, Palace of Versailles, Schönbrunn Palace and Schönbrunn Gardens.
The goal is to make these group of customers continuing being frequent with their travels while making them spend more. For this the following suggestions will be applied:
- Solo Traveler Packages;
- Tiered Loyalty Rewards;
- Special Occasion Celebrations: special perks on birthdays, anniversaries, or other;
- Loyalty Points for Solo Journeys.

3. The "Spenders Travelers"
Spenders emerge through an xx4 RFM score and are the ones that give **higher reviews.** They come from diverse regions, including the United States, England, Australia, Canada, India, and New Zealand. This group primarily consists of **couples and families,** with occasional friend trips. Their favored destinations span Spain, England, Russia, and Portugal. While their typical travel patterns align with non-holiday periods, it's noteworthy that they also venture out during **Labor Day and New Year celebrations**. Delving into their preferred attractions, this group demonstrates a keen interest in iconic landmarks such as the Basilica of the Sagrada Familia, Parc Güell, Casa Batlló, Tower of London, Edinburgh Castle, Real Alcazar de Sevilla, and Casa Milà - La Pedrera.

- Exclusive Family Packages;
- Couples Retreats
- Indian Festival Travel Deals: Recognize the significance of Indian festivals;
- Labor Day Escapes

4. The "Churned Cheap Travelers"

This segment represents customers that had a **last travel a long time ago, travelled few, and gave poor reviews.** It comprises individuals from England, the United States, Australia, Canada, and Ireland, predominantly engaging in couples' trips but also including families and friends. Their preferred destinations primarily focus on Spain and England, with occasional visits to Scotland and Portugal. Notably, despite being categorized as 'churned cheap,' this segment exhibits a consistent interest in exploring renowned attractions. Among their favorite attractions are the Basilica of the Sagrada Familia, Tower of London, Edinburgh Castle, Parc Güell, Schönbrunn Palace, The Roman Baths, and Krakow's Rynek Glowny Central Square.
We want to turn these customers into better ones, that travel more and give higher reviews, therefore we will provide them with the following suggestions:

- Reactivation Promotions;
- Personalized Recommendations;
- Feedback Incentive by offering incentives such as discounts or additional amenities;
- Loyalty Program.

5. The "Gold" Travelers"

Positioned just after the best customers, they represent a group that travels with high frequency, and recentness, and consistently provides higher reviews. Primarily hailing from England, this segment consists of couples and friends. Their favorite destinations include Austria, England, Spain, and Italy. They are inclined to travel during non-holiday periods and among their preferred attractions are Schönbrunn Palace, Parc Güell, Palace of Versailles, Schönbrunn Gardens, Blenheim Palace, Duomo di Pisa, and the Old Town.

6. The" Silver Travelers"

The silver segment, primarily comprising travelers from **Lithuania**, **England**, and the **United States**, predominantly consists of **couples** and **solo** explorers. Their preferred destinations encompass **Spain**, **England**, and **Italy**, with a notable tendency to **avoid holiday periods**. Among their favored attractions, Canal Grande, Tower of Hercules (Torre de Hercules), Basilica of the Sagrada Familia, Parc Güell, Casa Batlló, Casa Milà - La Pedrera, and Catedral de Sevilla stand out. These attractions encapsulate a mix of cultural, historical, and architectural marvels, contributing to the diverse and enriching travel experiences enjoyed by this discerning silver segment.

When examining the segmented customer groups, a notable trend emerges, indicating that a substantial portion falls into the 'Spenders' category. This observation suggests that customers within this segment consistently provide **high reviews**, even if their **travel frequency is relatively low**. Within the broader analysis, two segments align with 'Loyal Customers,' emphasizing a high frequency of travel, while two segments correspond to 'Churned Cheap,' signifying a lapse in both recent travel and spending. We also have the best travelers!

7. Portuguese Travelers
Portuguese customers are in the **churned cheap segment,** which represents a valuable opportunity to reactivate and transform into active, frequent customers. Implementing personalized strategies can help rekindle their interest in travel, therefore here are targeted suggestions:

- Exclusive Country-Specific Packages: Craft exclusive travel packages focusing on their preferred countries.
- Design promotions around their favorite attractions
- Off-Peak Travel Deals
- Cultural Immersion Packages: Tailor experiences that delve into the cultural and historical richness of Portugal.

8. Portugal Attractions
The segment with **RFMScore 112** represents travelers who have recently visited Portugal but exhibit low frequency and monetary engagement. To enhance their experience and encourage repeat visits, we suggest the following:

- Personalized Welcome Back Offers;
- Feedback and Improvement;
- Tailored Loyalty Programs;
- Curated Itineraries;
- Social Proof and Testimonials: Showcase positive testimonials and experiences of other travelers who initially belonged to Segment 112 but found renewed enjoyment during subsequent visits. Social proof can instill confidence and inspire return trips.

9. Most Frequent Users

In our strategic initiative to enhance user engagement and broaden our traveler base, we meticulously examined the behaviors and preferences of our **top 10 users.** Our approach involved identifying the **most frequently visited countries** by our top users and determining countries with high similarity metrics. For example, by understanding that a user has explored Spain but not Portugal, Italy, or England, we can specifically target them with suggestions for these unvisited countries. The goal is also to provide them with personalized recommendations of trips, promotions, and deals to similar countries. **We also examined users' favorite attractions** and devised a strategic plan incorporating similar attractions for them to explore.

To re-engage churned customers, we leverage positive feedback from experienced travelers, since sharing these testimonials can act as a powerful incentive for churned customers to revisit and rediscover these captivating destinations.

- **Attractions similarity**

In our pursuit of enriching travel experiences, we embarked on a detailed analysis of attraction similarities. We meticulously selected the **top 10 attractions** based on both the strength of correlations and their frequency, ensuring a robust foundation for our insights. By closely examining the attractions with the most substantial correlations, we uncovered a curated list that truly resonates with each other. This thoughtful selection process enhances the thematic coherence of our recommendations, allowing us to provide travelers with a more interconnected and meaningful journey.

For instance, consider the Aqueduct of Segovia. By strategically choosing the ten most similar attractions, such as Alcázar of Segovia, Basilica of the Sagrada Familia, Casa Batlló, and Casa Milà - La Pedrera, **we create a narrative that aligns with the traveler's preferences.** This method goes **beyond mere correlation; it's a strategic approach** that crafts a symphony of attractions, ensuring a seamless and captivating exploration. Our methodology ensures that every recommendation is grounded in a thorough understanding of attraction relationships.

- **Portuguese Attractions Similarity**

By meticulously analyzing the **top attractions in Portugal** and their corresponding similar attractions, we unlock a treasure trove of strategic possibilities. This information serves as a powerful tool to not only entice global travelers to explore Portugal but also to **inspire Portuguese visitors** to embark on more enriching journeys.

By capitalizing on the resemblances between Portugal's top attractions and those in other regions, we can strategically deploy targeted advertising, promotions, and recommendations. **This approach not only attracts global travelers to explore Portugal but also stimulates** residents to embark on more extensive journeys. For example, suggesting similar attractions to Portuguese visitors not only enriches their travel experiences but also stimulates a desire for broader exploration, creating a positive feedback loop that benefits both domestic and international tourism.

We also delved into **identifying the most similar attractions to Portugal's top destinations,** excluding those within Portugal itself. This strategic move enables us to unravel potential travel hotspots that may appeal to Portuguese travelers. By pinpointing these similar attractions, we can tailor our recommendations and promotions, encouraging residents to explore new and exciting destinations beyond their borders. This targeted approach not only enriches the travel experiences of Portuguese individuals but also contributes to the diversification and expansion of their travel preferences.

- **Countries Similarity**

Through a methodical selection process, we identified the **top 10 countries with the highest frequency and robust correlations.** For instance, consider the scenario where Estonia is a focal point. By exploring the countries with the highest correlations, such as Latvia, Russia, Finland, and Lithuania, we uncover a network of destinations that share thematic, cultural, or experiential similarities.

6   Deployment

6.1   Suggestions for improving the visitor experience.

- **Recommendation system**

To enhance customer engagement and foster a sense of spontaneity in travel planning, we've introduced an interactive system where customers can share their favorite countries and attractions. By simply entering three preferred countries and attractions, our system generates personalized suggestions for similar destinations, providing customers with a curated list tailored to their interests.

```
Enter your 3 favorite countries (comma-separated):  France,Italy,Spain
Enter your 3 favorite attractions (comma-separated):  Torre de Belém,Piazza delle Erbe,Quinta da Regaleira

Top 2 recommended countries:
Austria
Latvia

Top 3 recommended attractions:
Abbaye du Mont-Saint-Michel
Place Stanislas
Salzburger Altstadt
```

Figure 23. Recommendation system output

How it Works:
1. Input Preferences:
   Users enter their 3 favorite countries (comma-separated).
   Users enter their 3 favorite attractions (comma-separated).
2. Recommendation Generation: The system processes user inputs to identify preferences. Top 2 recommended countries and top 3 recommended attractions are generated based on user preferences.

This innovative feature not only adds a delightful element of customization to the travel experience but also encourages customers to explore new and exciting destinations based on their individual preferences.

6.2   Recommendations for the Portuguese NTBO: Marketing strategies based on identified patterns

COVID-19 impact insights:

- **Domestic Tourism Motivations:** Residents play a crucial role as influential marketing ambassadors within their own country. Recognizing their unique insights and genuine enthusiasm for their homeland, tourism initiatives can strategically tap into this connection. Engaging residents through targeted campaigns, special offers, and events fosters pride, encouraging them to explore and endorse diverse attractions. This grassroots approach harnesses authentic local experiences, creating a ripple effect that inspires fellow citizens to appreciate the richness of their national heritage.
- **Group Focus:** Portugal Hotels, Hostels, and other types of Local Accommodations need to consider the shift in tourists' behavior. Portugal is showing signs of becoming a good choice for Solo Travelers and Friends Groups. For solo travelers, it's important to provide ways to make tourists meet other solo travelers. Solo travelers are always more open to trying new things and experimenting— organizing group activities like tours with games to help them know each other, bar rallies, and creating spaces for people to meet. Ensure that accommodation is clean and safe.

Frequent Itemset Associations (Market Basket Analysis):

- **Partnerships:** Considering the importance of Spain in the global spectrum, Portugal can view Spain as either a main competitor or a potential ally in the tourism industry. Encouraging Portugal -> Spain travel packages can strengthen collaboration and enhance the overall tourism experience for visitors exploring the Iberian Peninsula.

- **Interrail Programs:** Support initiatives like Interrail programs to promote multi-country travel experiences. Leverage tourists' natural inclination to explore nearby destinations, contributing to a more integrated travel network.
- **Connected Tours:** Develop thematic tours that connect associated attractions such as Quinta da Regaleira, Mosteiro dos Jeronimos, and Torre de Belém. This approach facilitates visits to multiple attractions in a single journey, offering a more cohesive travel experience.
- **Tourist Cards/Packages:** Introduce tourist cards or packages that provide discounts or benefits for visiting multiple associated attractions. Creating a "Lisbon Tourist Card" with discounts on attractions like Quinta da Regaleira, Mosteiro dos Jeronimos, and Torre de Belém can attract more visitors.
- **Online Promotions:** Implement online promotions to encourage early booking of tours or acquisition of tourist packages. Collaborate with booking platforms to offer exclusive discounts, making it more attractive for tourists to plan and book in advance.
- **Local Advertising:** Execute targeted advertising campaigns at specific locations, emphasizing the interconnectedness of attractions. Highlight the benefits of visiting suggested combinations like Quinta da Regaleira, Mosteiro dos Jeronimos, and Torre de Belém together, enhancing awareness among potential visitors.

7   Conclusion

In conclusion, our analysis provides a comprehensive understanding of visitor patterns at Portuguese attractions and facilitates informed decision-making for the NTBO. By harnessing the power of user-generated content, we enable the development of strategic initiatives to revive and enhance the tourism industry in Portugal post-pandemic.

References

Tomé, A. (2022, 16 de maio). *TravelBI by Turismo de Portugal - 2 anos de Pandemia - O Impacto Económico no Setor Turístico | 2022*. TravelBI by Turismo de Portugal. https://travelbi.turismode-portugal.pt/turismo-em-portugal/2-anos-de-pandemia-o-impacto-economico-no-setor-turistico-2022/

Publico.pt. (2020, 15 de março). Países encerram fronteiras para conter a pandemia de coronavírus: https://www.publico.pt/2020/03/15/mundo/noticia/paises-encerram-fronteiras-conter-pandemia-coronavirus-1907869

Expresso.pt. (2020, 31 de maio). Covid-19: Fronteiras na Europa abrem-se aos poucos ao turismo: https://expresso.pt/internacional/2020-05-31-Covid-19.-Fronteiras-na-Europa-abrem-se--aos-poucos--ao-turismo

The New York Times. (2020, 4 de dezembro). Europe's Covid Deaths: The Untold Story of Europe's Pandemic: https://www.nytimes.com/interactive/2020/12/04/world/europe/europe-covid-deaths.html

Appendix A

Table A 1. Nomenclature and Categorization of Variables

xi; *"variable name"*; "type of variable"			
localID	NC	totalReviews	DN
userName	NC	userContributions	DN
userLocation	NC	ID	NC
tripType	NC	Name	NC
reviewLanguage	NC	Country	NC
reviewFullText	NC	ISO	NC
globalRating	DN	countryOrRegion	NC
reviewRating	OC	holidayName	NC
positionOnRanking	OC	normalizeHolidayName	NC
extractionDate	OC	isPaidTimeOff	NC
reviewWritten	OC	countryRegionCode	NC
reviewVisited	OC	date	OC
sitesOnRanking	DN		

Where, DN - discrete numerical; OC – ordinal categorical, and NC – nominal categorical.

Table A 2. *Missing values*

Variable	Total Rows	Missing Values
userName	92 120	20
userLocation	92 120	13 468
tripType	92 120	29068
reviewVisited	92 120	710

Table A 3. *Skew and Kurtosis Values*

Variable	Skew	Kurtosis
globalRating	- 0.406169	4.761004
positionOnRank-ing	4.096502	28.204280
sitesOnRanking	1.566002	2.052919
totalReviews	2.034586	3.255460
userContributions	72.009683	5855.302513
reviewRating	-2.318242	5.887076

Table A 4. *Feature engingeering: New variables List*

Variable	type	Description
userISOCode	string	ISO code of the location of where the user who posted the review lives. This is not a mandatory field.
review-FullText_agg	string	full text of the review now aggregated
Name	string	Name of the attraction
Country	string	Country of the attraction
ISO	string	ISO code of the Country of the attraction
holiday	bool	1 if the day of the visit was a holiday; 0 if not
holiday_name	string	Name of the holiday
Year	numeric	Year of the visit
Month	numeric	Month of the Visit
Day	numeric	Day of the visit

Table A 5. *Frequent Countries Sets – Total Reviews (Rules by Confidence – users with more than 8 reviews):*

Antecedents	Consequents	Support	Confidence	Lift
(Italy, Scotland)	(England)	0.122807	0.875000	2.267045
(Poland)	(Italy)	0.157895	0.818182	2.027668
(Portugal)	(Spain)	0.333333	0.791667	1.219595
(Croatia)	(Spain)	0.140351	0.727273	1.120393
(Scotland)	(England)	0.210526	0.705882	1.828877
(Scotland)	(Spain)	0.192982	0.647059	0.996820
(Spain, Scotland)	(England)	0.122807	0.636364	1.648760
(England, Italy)	(Spain)	0.122807	0.636364	0.980344
(England, Italy)	(Scotland)	0.122807	0.636364	2.133690

Table A 6. *Frequent Countries Sets – Total Reviews (Rules by Confidence – users with more than 7 reviews):*

Antecedents	Consequents	Support	Confidence	Lift
(Portugal)	(Spain)	0.269231	0.800000	1.363934
(Scotland)	(England)	0.182692	0.760000	2.026667
(Croatia)	(Italy)	0.115385	0.545455	1.620779
(Italy)	(England)	0.182692	0.542857	1.447619
(England)	(Scotland)	0.182692	0.487179	2.026667
(England)	(Italy)	0.182692	0.487179	1.447619
(France)	(Italy)	0.125000	0.464286	1.379592

Table A 7. *Frequent Attractions Sets – Total Reviews (Rules by Confidence users with more than 7 reviews):*

Antecedents	Consequents	Support	Confidence	Lift
(MAG004)	(MAG001)	0.211538	0.956522	2.486957
(MAG012)	(MAG019)	0.201923	0.913043	3.516908
(MAG023)	(MAG001)	0.201923	0.875000	2.275000
(MAG002)	(MAG001)	0.230769	0.827586	2.151724
(MAG019)	(MAG012)	0.201923	0.777778	3.516908
(MAG001)	(MAG002)	0.230769	0.600000	2.151724
(MAG001)	(MAG004)	0.211538	0.550000	2.486957
(MAG001)	(MAG023)	0.201923	0.525000	2.275000

Table A 8. Frequent Attractions Sets – Reviews for Portugal (Rules by Confidence users with more than 7 reviews)

Antecedents	Consequents	Support	Confidence	Lift
(MAG047)	(MAG014)	0.228571	0.888889	1.414141
(MAG047)	(MAG010)	0.228571	0.888889	1.296296
(MAG014, MAG047)	(MAG010)	0.200000	0.875000	1.276042
(MAG010, MAG047)	(MAG014)	0.200000	0.875000	1.392045
(MAG032, MAG014)	(MAG010)	0.285714	0.833333	1.215278
(MAG014)	(MAG010)	0.514286	0.818182	1.193182
(MAG047)	(MAG010, MAG014)	0.200000	0.777778	1.512346
(MAG047)	(MAG032)	0.200000	0.777778	1.512346
(MAG032, MAG010)	(MAG014)	0.285714	0.769231	1.223776
(MAG010)	(MAG014)	0.514286	0.750000	1.193182
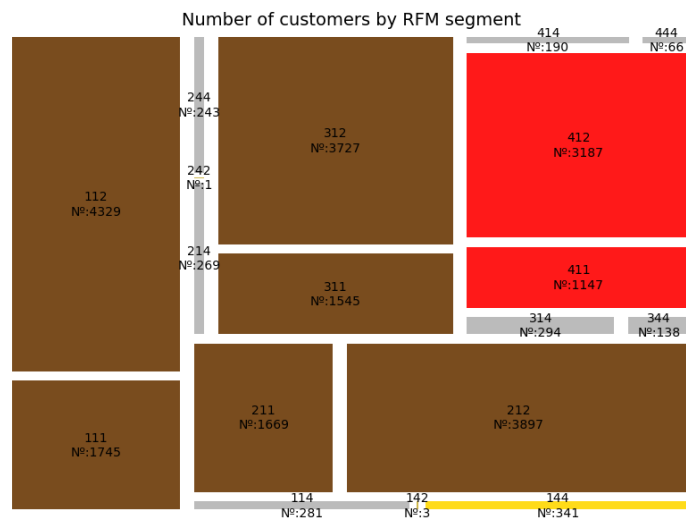
Figure 24. *RFM Best loyal spenders and churned segments*

Figure 25. *RFM Gold and Silver Segments*

Figure 26. *Average Regency (days) by Monetary and Frequency Levels*

Figure 27. Gold segment

	Country	TripType	UserISOCode	Holiday	Attraction
0	Country_Spain	tripType_nan	userISOCode_LT	holiday_name_No	Name_Canal Grande
1	Country_Italy	tripType_Couples	userISOCode_US	holiday_name_all saints	Name_Tower of Hercules (Torre de Hercules)
2	Country_England	tripType_Solo	userISOCode_GB	holiday_name_New Year	Name_Basilica of the Sagrada Familia
3	Country_Russia	tripType_Family	userISOCode_nan	holiday_name_hallowmas	Name_Parc Guell
4	Country_Portugal	tripType_Friends	userISOCode_AU	holiday_name_restoration of independence	Name_Casa Batllo
5	Country_Scotland	tripType_Business	userISOCode_CA	holiday_name_labor day	Name_Casa Milà - La Pedrera
6	Country_Poland	-	userISOCode_IN	holiday_name_toussaint	Name_Catedral de Sevilla
7	Country_Austria	-	userISOCode_MY	-	Name_Real Alcazar de Sevilla
8	Country_France	-	userISOCode_ES	-	Name_The State Hermitage Museum
9	Country_Greece	-	userISOCode_DE	-	Name_Grand Peterhof Palace
10	Country_Belgium	-	userISOCode_AE	-	Name_Tower of London
11	Country_Hungary	-	userISOCode_IE	-	Name_The Roman Baths
12	Country_Croatia	-	userISOCode_SE	-	Name_Mezquita Cathedral de Cordoba
13	Country_Germany	-	userISOCode_IL	-	Name_Edinburgh Castle

Figure 28. *Silver segment*

	Country	TripType	UserISOCode	Holiday	Attraction
0	Country_Spain	tripType_nan	userISOCode_GB	holiday_name_No	Name_Canal Grande
1	Country_England	tripType_Couples	userISOCode_US	holiday_name_New Year	Name_Tower of Hercules (Torre de Hercules)
2	Country_Italy	tripType_Friends	userISOCode_nan	holiday_name_all saints	Name_Blenheim Palace
3	Country_Portugal	tripType_Family	userISOCode_AU	holiday_name_labor day	Name_Parc Guell
4	Country_Scotland	tripType_Solo	userISOCode_CA	holiday_name_restoration of independence	Name_Palace of Versailles
5	Country_France	tripType_Business	userISOCode_NL	holiday_name_hallowmas	Name_Schönbrunn Palace
6	Country_Poland	-	userISOCode_IL	holiday_name_toussaint	Name_Schönbrunn Gardens
7	Country_Austria	-	userISOCode_RO	holiday_name_public holiday	Name_Duomo di Pisa
8	Country_Greece	-	userISOCode_ES	holiday_name_workers day	-
9	Country_Russia	-	userISOCode_BR	holiday_name_state holiday	-
10	Country_Belgium	-	userISOCode_RU	holiday_name_labour day	-
11	Country_Croatia	-	userISOCode_IT	holiday_name_day of restoration of the indepen...	-
12	Country_Germany	-	userISOCode_FI	-	-
13	Country_Hungary	-	userISOCode_DE	-	-
14	Country_Estonia	-	userISOCode_BE	-	-

Figure 29. *Best Segment*

	Country	TripType	UserISOCode	Holiday	Attraction
0	Country_Spain	tripType_Couples	userISOCode_GB	holiday_name_No	Name_Canal Grande
1	Country_Italy	tripType_Solo	userISOCode_LT	-	Name_Tower of Hercules (Torre de Hercules)
2	Country_Austria	tripType_nan	-	-	Name_Blenheim Palace
3	Country_England	tripType_Friends	-	-	Name_Parc Guell
4	Country_France	-	-	-	Name_Palace of Versailles

Figure 30. *Loyal Segment*

	Country	TripType	UserISOCode	Holiday	Attraction
0	Country_Spain	tripType_Couples	userISOCode_US	holiday_name_No	Name_Basilica of the Sagrada Familia
1	Country_England	tripType_nan	userISOCode_GB	holiday_name_labor day	Name_Parc Guell
2	Country_Russia	tripType_Family	userISOCode_nan	holiday_name_New Year	Name_Casa Batllo
3	Country_Portugal	tripType_Friends	userISOCode_AU	holiday_name_all saints	Name_Tower of London
4	Country_Scotland	tripType_Solo	userISOCode_CA	holiday_name_workers day	Name_Edinburgh Castle
5	Country_Italy	tripType_Business	userISOCode_IN	holiday_name_state holiday	Name_Real Alcazar de Sevilla
6	Country_Austria	-	userISOCode_NZ	holiday_name_public holiday	Name_Casa Milà - La Pedrera
7	Country_Poland	-	userISOCode_IT	holiday_name_hallowmas	Name_Catedral de Sevilla
8	Country_France	-	userISOCode_ZA	holiday_name_day of restoration of the indepen...	Name_The Roman Baths
9	Country_Belgium	-	userISOCode_CN	holiday_name_first of may	Name_Schönbrunn Palace
10	Country_Croatia	-	userISOCode_ES	holiday_name_labour day	Name_The State Hermitage Museum
11	Country_Hungary	-	userISOCode_SG	holiday_name_restoration of independence	Name_Edinburgh Old Town
12	Country_Greece	-	userISOCode_IL	holiday_name_sunday	Name_Westminster Abbey
13	Country_Estonia	-	userISOCode_RO	holiday_name_toussaint	Name_Torre de Belém
14	Country_Germany	-	userISOCode_GR	-	Name_Church of the Savior on Spilled Blood

Figure 31. Spenders Segment

	Country	TripType	UserISOCode	Holiday	Attraction
0	Country_Spain	tripType_Couples	userISOCode_GB	holiday_name_No	Name_Basilica of the Sagrada Familia
1	Country_England	tripType_Family	userISOCode_US	holiday_name_New Year	Name_Tower of London
2	Country_Scotland	tripType_Friends	userISOCode_nan	holiday_name_day of restoration of the indepen...	Name_Edinburgh Castle
3	Country_Portugal	tripType_nan	userISOCode_AU	-	Name_Parc Guell
4	Country_Italy	tripType_Solo	userISOCode_CA	-	Name_Schönbrunn Palace
5	Country_Poland	tripType_Business	userISOCode_IE	-	Name_The Roman Baths
6	Country_Austria	-	userISOCode_ES	-	Name_Krakow's Rynek Glowny Central Square
7	Country_France	-	userISOCode_NL	-	Name_Acropolis
8	Country_Russia	-	userISOCode_IT	-	Name_Westminster Abbey
9	Country_Greece	-	userISOCode_IN	-	Name_Torre de Belém
10	Country_Belgium	-	userISOCode_SG	-	Name_The Alhambra
11	Country_Hungary	-	userISOCode_PT	-	Name_Real Alcazar de Sevilla
12	Country_Germany	-	userISOCode_GR	-	Name_Royal Albert Dock Liverpool
13	Country_Croatia	-	userISOCode_DE	-	Name_Casa Batllo
14	Country_Cyprus	-	userISOCode_BE	-	Name_Catedral de Sevilla

Figure 32. Churned cheap segment

	Spain	Italy	England	Portugal
1	Portugal	Croatia	Scotland	Spain
2	Italy	France	France	Italy
3	France	Spain	Italy	Czech Republic
4	England	Greece	Spain	France
5	Belgium	Montenegro	Belgium	Belgium

Figure 33. Similar Countries to the favorite countries of the top users

	Basilica of the Sagrada Familia	Mosteiro dos Jeronimos	Parc Guell	Torre de Belém
1	Parc Guell	Torre de Belém	Basilica of the Sagrada Familia	Mosteiro dos Jeronimos
2	Casa Batllo	Park and National Palace of Pena	Casa Milà - La Pedrera	Park and National Palace of Pena
3	Casa Milà - La Pedrera	Quinta da Regaleira	Casa Batllo	Quinta da Regaleira
4	Palace of Catalan Music	Cais da Ribeira	Palace of Catalan Music	Ponte de Dom Luís I
5	The Alhambra	Ponte de Dom Luís I	The Alhambra	Cais da Ribeira

Figure 34. Similar attractions to the favorites of the top users

	Attraction	Top_1_Similar	Top_2_Similar	Top_3_Similar	Top_4_Similar	Top_5_Similar	Top_6_Similar	Top_7_Similar	Top_8_Similar	Top_9_Similar	Top_10_Similar
0	Torre de Belém	Mosteiro dos Jeronimos	Park and National Palace of Pena	Quinta da Regaleira	Ponte de Dom Luís I	Cais da Ribeira	Bom Jesus do Monte	The Walls of Avila	Catedral de Sevilla	Acueduct of Segovia	Real Alcazar de Sevilla
1	Park and National Palace of Pena	Quinta da Regaleira	Mosteiro dos Jeronimos	Torre de Belém	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Alcázar of Segovia	Real Alcazar de Sevilla	The Alhambra	Catedral de Sevilla
2	Mosteiro dos Jeronimos	Torre de Belém	Park and National Palace of Pena	Quinta da Regaleira	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Catedral de Sevilla	The Walls of Avila	Catedral Primada	Mezquita Cathedral de Cordoba
3	Bom Jesus do Monte	Cais da Ribeira	Ponte de Dom Luís I	Park and National Palace of Pena	Torre de Belém	Mosteiro dos Jeronimos	Cathedral De Santiago de Compostela	Quinta da Regaleira	Mezquita Cathedral de Cordoba	The State Museum St. Isaac's Cathedral	Historic Centre of San Gimignano
4	Ponte de Dom Luís I	Cais da Ribeira	Torre de Belém	Bom Jesus do Monte	Mosteiro dos Jeronimos	Park and National Palace of Pena	Quinta da Regaleira	Cathedral De Santiago de Compostela	Tower of Hercules (Torre de Hercules)	The State Museum St. Isaac's Cathedral	Staromestske namesti
5	Quinta da Regaleira	Park and National Palace of Pena	Torre de Belém	Mosteiro dos Jeronimos	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Cathedral De Santiago de Compostela	Casa Milà - La Pedrera	Casa Batllo	Basilica of the Sagrada Familia
6	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Torre de Belém	Mosteiro dos Jeronimos	Park and National Palace of Pena	Cathedral De Santiago de Compostela	Quinta da Regaleira	Tower of Hercules (Torre de Hercules)	Staromestske namesti	Schönbrunn Gardens

Figure 35. Similar attractions to the portuguese attractions

Figure 36. *Similar attractions to the portuguese favorite attractions*

	Attraction	Top_1_Similar	Top_2_Similar	Top_3_Similar	Top_4_Similar	Top_5_Similar	Top_6_Similar	Top_7_Similar	Top_8_Similar	Top_9_Similar	Top_10_Similar
0	Torre de Belém	Mosteiro dos Jeronimos	Park and National Palace of Pena	Quinta da Regaleira	Ponte de Dom Luís I	Cais da Ribeira	Bom Jesus do Monte	The Walls of Avila	Catedral de Sevilla	Acueduct of Segovia	Real Alcazar de Sevilla
1	Park and National Palace of Pena	Quinta da Regaleira	Mosteiro dos Jeronimos	Torre de Belém	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Alcázar of Segovia	Real Alcazar de Sevilla	The Alhambra	Catedral de Sevilla
2	Mosteiro dos Jeronimos	Torre de Belém	Park and National Palace of Pena	Quinta da Regaleira	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Catedral de Sevilla	The Walls of Avila	Catedral Primada	Mezquita Cathedral de Cordoba
3	Bom Jesus do Monte	Cais da Ribeira	Ponte de Dom Luís I	Park and National Palace of Pena	Torre de Belém	Mosteiro dos Jeronimos	Cathedral De Santiago de Compostela	Quinta da Regaleira	Mezquita Cathedral de Cordoba	The State Museum St. Isaac's Cathedral	Historic Centre of San Gimignano
4	Ponte de Dom Luís I	Cais da Ribeira	Torre de Belém	Bom Jesus do Monte	Mosteiro dos Jeronimos	Park and National Palace of Pena	Quinta da Regaleira	Cathedral De Santiago de Compostela	Tower of Hercules (Torre de Hercules)	The State Museum St. Isaac's Cathedral	Staromestske namesti
5	Quinta da Regaleira	Park and National Palace of Pena	Torre de Belém	Mosteiro dos Jeronimos	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Cathedral De Santiago de Compostela	Casa Milà - La Pedrera	Casa Batllo	Basilica of the Sagrada Familia
6	Cais da Ribeira	Ponte de Dom Luís I	Bom Jesus do Monte	Torre de Belém	Mosteiro dos Jeronimos	Park and National Palace of Pena	Cathedral De Santiago de Compostela	Quinta da Regaleira	Tower of Hercules (Torre de Hercules)	Staromestske namesti	Schönbrunn Gardens

Figure 37. *Similar attractions (outside of portugal) to the portuguese favorite attractions*

	Attraction	Top_1_Similar	Top_2_Similar	Top_3_Similar	Top_4_Similar	Top_5_Similar	Top_6_Similar	Top_7_Similar	Top_8_Similar	Top_9_Similar	Top_10_Similar
0	Torre de Belém	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila	The Walls of Avila
1	Park and National Palace of Pena	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia	Alcázar of Segovia
2	Mosteiro dos Jeronimos	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla	Catedral de Sevilla
3	Bom Jesus do Monte	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela
4	Ponte de Dom Luís I	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela
5	Quinta da Regaleira	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela
6	Cais da Ribeira	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela	Cathedral De Santiago de Compostela

Figure 38. *Countries with highest similarity*

	Country	Country1	Country2	Country3	Country4	Country5	Country6	Country7	Country8	Country9	Country10
0	Estonia	Latvia	Russia	Finland	Lithuania	Poland	Hungary	Belgium	Norway	Scotland	Montenegro
1	Latvia	Lithuania	Estonia	Finland	Russia	Poland	Belgium	Bulgaria	Hungary	Spain	Italy
2	Austria	Hungary	Czech Republic	Poland	Germany	Italy	France	Belgium	Croatia	England	Montenegro
3	Hungary	Austria	Czech Republic	Poland	Germany	Italy	Belgium	Estonia	England	Russia	Portugal
4	Italy	Croatia	France	Spain	Greece	Montenegro	England	Austria	Hungary	Poland	Portugal
5	Croatia	Montenegro	Italy	Austria	Greece	Poland	France	Spain	Hungary	Turkey	England
6	Lithuania	Latvia	Estonia	Finland	Poland	Russia	Hungary	Montenegro	Croatia	Germany	Spain
7	Russia	Estonia	Finland	Latvia	Lithuania	Turkey	Portugal	Spain	Hungary	England	Poland
8	France	England	Italy	Spain	Belgium	Germany	Austria	Portugal	Norway	Croatia	Poland
9	England	Scotland	France	Italy	Spain	Belgium	Germany	Austria	Poland	Hungary	Portugal
10	Spain	Portugal	Italy	France	England	Belgium	Russia	Croatia	Germany	Scotland	Austria