

# **DATA SCIENCE FOR MARKETING**



# PORTUGUESE TOURISM VISITORS: A DATA ANALYSIS REPORT

Beatriz Fragoso | 20231531 Madalena Cunha | r2015484 Sofia Franco | 20230097 João Dias | 20230072

# **EXECUTIVE SUMMARY**

# **Objective**

The objective of this report is to provide comprehensive insights and recommendations to the Portuguese National Tourism Board Organization (NTBO) on positioning Portugal as a premier tourism destination and how COVID-19 affected tourist patterns and behaviours. Through user-generated content, particularly TripAdvisor reviews, our goal is to characterize and describe the patterns of visitors to Portuguese attractions simultaneously comparing it to the main Portuguese competitors. The study focuses on understanding shifts in the tourism dynamic due to the pandemic, identifying visitor preferences and behaviours by using data analysis techniques and empowering the Portuguese NTBO with data-driven strategies, optimising marketing efforts, and strengthening Portugal's competitiveness in the global tourism landscape.

# **Key Findings**

- **Tourist Behaviour and Preferences:** A shift in visitor patterns post-COVID-19, with a notable increase in lower ratings, reflecting changes in visitor satisfaction.
- **Visitor Demographics and Seasonality:** Predominance of visitors from the UK and the USA, with a high preference for summer travel, but a more even seasonal distribution for Portugal.
- **Attraction Analysis:** Historical and religious sites are the most popular, reflecting Europe's rich cultural heritage.
- **Sentiment Analysis:** Generally positive sentiments in reviews, with moderate subjectivity, indicating balanced visitor feedback.
- **Temporal Analysis:** A decrease in average ratings during the pandemic, followed by a gradual recovery.
- **Comparative Country Analysis:** Insights into how Portugal compares with competitors regarding visitor origins and trip types.
- **Developed Models:** Using data-driven tools and algorithms to understand tourist movements and preferences.
- **Marketing Recommendations:** Tailored strategies based on attraction types and visitor segmentation, including special campaigns for key visitor segments like 'Gold Users'.
- **Future Strategies:** Emphasis on data-driven decisions and adaptive strategies for NTBO to sustain and grow tourism.

#### Recommendations

- **Attraction Types Analysis:** Establish partnerships with cultural institutions, museums, and historical preservation organizations. Collaborate on joint initiatives to promote heritage site preservation. Utilize partnerships to create unique and exclusive visitor experiences.
- Temporal and Seasonal Analysis: Position Portugal as a year-round destination. Develop marketing
  materials showcasing diverse experiences in every season. Highlight unique attractions, festivals,
  and events during spring and winter.
- Holiday Travel Patterns: Optimize marketing strategies for peak holiday travel, with targeted campaigns and exclusive offers for Labour Day and high seasons. Launch promotions in January, May, Spring, and Winter, and create special packages for couples around Labour Day and busy travel times.
- **Sentiment Analysis:** Conduct sentiment analysis to identify Portugal's strengths as a tourist destination. Pinpoint specific attractions, services, or experiences highly praised by visitors. Investigate how holidays coincide with positive or negative sentiments in reviews for further analysis.

# **TABLE OF CONTENTS**

EXECUTIVE SUMMARY	2
TABLE OF CONTENTS	3
INTRODUCTION	4
TOURISM OVERVIEW PRE- AND POST-PANDEMIC	4
PORTUGAL TOURISM BACKGROUND	4
METHODOLOGY	5
SCOPE OF THE PROJECT:	5
DATA COLLECTION	5
INSIGHTS FROM USER-GENERATED CONTENT  VISITOR BEHAVIOR AND PREFERENCES PRE- AND POST-PANDEMIC	7 8 8 8
REFERENCES	20

#### INTRODUCTION

# TOURISM OVERVIEW PRE- AND POST-PANDEMIC

The global tourism landscape has undergone a paradigm shift with the COVID-19 pandemic, affecting the sector in a deep and unprecedented way. Before the pandemic, international travel was characterized by an intricate and well-connected network of destinations – due to the emergence of low-cost airlines, making travel accessible to more people – and a flow of passengers crossing borders daily to explore different geographies, new cultures and, thus, contributing to the economic vitality of numerous countries.

The outbreak of the pandemic at the end of 2019, however, brought this era of rampant global mobility to an abrupt halt and with it a plethora of challenges that have had an impact on the whole tourism industry, such as health and safety concerns, travel restrictions, economic repercussions, and supply chain disruptions due to global lockdowns.

The post-COVID scenario of tourism has been marked by a fundamental revaluation of established norms. The pandemic exposed vulnerabilities in the industry, prompting destinations to rethink their approach. We have witnessed the rise of domestic tourism, a heightened emphasis on health and safety measures, and a shift in traveller priorities. Digital transformation also accelerated, with technology becoming a vital tool for contactless experiences, online bookings, and destination marketing.

In this new dynamic, the tourism industry faces the dual challenge of recovery and reinvention. In Portugal, tourism has historically played a leading role in the national economy, as the country's main export activity. Being one of the most affected sectors during COVID-19, it has been identified as a priority key sector for the country's development strategy, recovery, and modernization.

In May 2021, NTBO Turismo de Portugal initiated the 'Reactivate Tourism | Build the Future Action Plan' to guide the recovery of the nation's tourism from the pandemic. The plan includes analysing social media reviews to understand changing visitor patterns and the pandemic's long-term effects on travel habits, crucial for competing in the rejuvenating global tourism market.

In essence, this report aims to investigate the complex visitor behaviour and preferences at Portuguese tourist sites through a comprehensive data-driven approach stepping into the role of a strategic advisor for the Portuguese NTBO and to find if and how these patterns were affected by the pandemic, and how to differentiate Portugal from its competitors on the global tourism stage. Our proposition is through the main findings suggest how to create effective marketing strategies that truly connect with tourists, thereby enhancing Portugal's charm as a top travel choice.

#### PORTUGAL TOURISM BACKGROUND

The Tourism sector, recognised as a priority one for the country's development strategy, was also one that suffered with the pandemic impact the most. Analysing two of the most common indicators - Tourism Consumption in the Economic Territory (TCET) and Gross Value Added generated by Tourism (GVAT) -, we can see that in 2020 the sector's downturn largely affected the national economy (Figure 1), but that in 2021 the recovery was already clearly noticeable and in 2022 the contribution of tourism to Portugal's GDP exceeded pre-pandemic levels, showing a surprising recovery.

Also, according to INE (National Statistics Bureau), we found that during COVID, in 2020, domestic tourism slightly increased by 8.4% when compared to 2019 values, although this increase may only have been due to the fall in foreign tourism (from 64.4 per cent in 2019 to 51 per cent in 2020). These data show us tendencies for country visits before and after the pandemic that we intend to confirm in our further analysis.

#### **METHODOLOGY**

# SCOPE OF THE PROJECT:

# **DATA COLLECTION**

The starting point for our analysis was by addressing two datasets:

- 1. **EuropeTop100Attractions:** considered the primary dataset for the analysis, was obtained from TripAdvisor and represents reviews from the top 100 tourist attractions in Europe between January 1, 2019, to August 21, 2021. Besides the reviews for the attractions, it also provides information from the user who has written the review, such as their locations, the type of trip and when the customer visited the attraction.
- 2. **Holidays:** provides information on worldwide public holidays, which is fundamental to understanding peak tourist times.

# DATA ANALYSIS TOOLS AND TECHNIQUES

Through a combination of data-driven tools and to ensure comparability across datasets we employed data techniques such as normalising ratings and rankings to a common scale and standardize values to ensure consistency in the analysis.

The use of Python and Excel as our primary tools for data analysis was strategic. Python's advanced data handling capabilities enabled us to manage large volumes of data with efficiency and perform complex analyses and Excel's widespread use and accessibility make our analysis transparent and easier to read.

#### **RATIONALE**

To ensure that our methods align with the NTBO's objectives and effectively proceed with the study, we aimed to understand which approaches should be utilized, which we identified the following:

#### 1. IMPACT OF THE COVID-19 PANDEMIC

**Objective:** Evaluate how visitor patterns have been affected by the COVID-19 pandemic.

**Rationale:** Perform a thorough analysis of visitor reviews and ratings before and after the pandemic to detect any changes in the volume and behaviour of visitor feedback.

#### 2. IDENTIFY POPULAR ATTRACTIONS AND ACTIVITIES:

**Objective:** Determine the most popular attractions and activities based on user-generated content.

**Rationale:** Knowing which attractions and activities are highly rated and frequently mentioned by visitors helps in promoting key experiences and optimizing resource allocation.

# 3. UNDERSTAND VISITORS' DEMOGRAPHICS

**Objective:** Identify and understand the demographics of visitors to Portuguese attractions.

**Rationale:** Knowing the nationality and other demographic factors of visitors helps tailor marketing strategies and services to better meet their preferences.

#### 4. SEASONAL AND TEMPORAL PATTERNS

**Objective:** Analyse patterns in visitation based on temporal factors such as seasons and holidays.

**Rationale:** By correlating the information on holidays with the season it is possible to identify peak visitation and make informed strategic planning and promotional efforts.

#### 5. EVALUATE VISITORS' SENTIMENTS:

**Objective:** Assess sentiment analysis of reviews to gauge visitor satisfaction and identify areas for improvement. **Rationale:** Understanding the sentiment of reviews helps in identifying aspects of the visitor experience that are positively received or need attention, providing insights for enhancing tourism offerings.

#### 6. BENCHMARK AGAINST COMPETITORS

**Objective:** Compare user-generated content and visitor patterns for Portuguese attractions with those of key competitors in the tourism industry.

**Rationale:** Benchmarking against competitors provides insights into Portugal's competitive position, allowing the NTBO to identify strengths, weaknesses, opportunities, and threats.

#### **ANALYSIS AND FINDINGS**

Considered the core of the study – the complex patterns and insights that were able to be retrieved from the vast amount of tourist reviews, by examining the behaviours and preferences of visitors as they interact with the top 100 attractions in Europe with the power of data-driven techniques, we gathered valuable insights that serve to enlighten future NTBO strategic decisions. The analysis is divided into sub-sections, each addressing a distinct dimension of the visitor experience, behaviour, and perception.

# **INSIGHTS FROM USER-GENERATED CONTENT**

The original dataset presented an extensive collection of 92,210 values, each entry potentially holding considerable information across 15 distinct columns, ranging from general ratings to user-specific contributions. Throughout the meticulous process of data cleaning and standardization, we employed a series of techniques aimed at refining the dataset's quality and usability. Our final dataset comprises 78,409 entries, each of which offers a snapshot of individual tourist experiences, ratings, and reviews and on a more generalized analysis, we observed the following:

Rating Distribution: The analysis indicates a significant skew towards higher ratings. A predominant 71% of reviews have a 5-star rating, followed by 20% of reviews with a 4-star rating. Lower ratings, ranging from 1-3 stars, collectively contribute to less than 10% of the total reviews, indicating a generally positive reception from users (See Figure 2).

#### **Visitor Demographics & Behaviours:**

<u>Visitor Origins:</u> A significant portion of the visitors are from the United Kingdom (approximately 38%) and the United States (about 26%). Other notable countries include Australia, Canada, and Ireland. It's important to note the potential bias in the data towards these nations, likely influenced by the predominant user base of the source platform<sup>1</sup> (See Figure 3).

<u>Visited Countries:</u> The analysis identifies Spain, England, Scotland, Italy, Portugal, Poland, and France as the top visited countries. This information is vital for understanding tourism trends and the global appeal of certain destinations (See Figure 4).

<u>Seasonality:</u> The seasonal distribution of travel visits among users indicates a preference for summer travel which sees the highest number of visits. This peak is likely due to favourable weather conditions, school vacations (considering that 'Family' is the second most common type of travel among users), and the cultural inclination to travel during this period. The moderate visitation numbers in spring and autumn suggest these seasons are also favourable for travel, potentially due to the pleasant weather and fewer crowds, making them attractive for those seeking a more relaxed travel experience. Also, both together represent almost 50% of the dataset, which could be a good opportunity for a boost in travel promotions during those months. Lastly, Winter is the least favoured season for travel, as indicated by the lowest visitation numbers, which could be a result of the colder climate and a tendency for individuals to remain close to home during the holiday season.

On the other hand, the seasonal visitation for Portugal surprisingly displays a balanced distribution throughout the year, with a slight preference for Spring (28.6%) and Autumn (25.84%), indicating Portugal's

-

<sup>&</sup>lt;sup>1</sup> https://6sense.com/tech/reservation-and-online-booking/tripadvisor-market-share

appeal during milder weather and potentially fewer crowds during peak season. Summer and Autumn visitations are nearly equal, and even Winter has 19.82% of visits, suggesting that Portugal attracts tourists beyond the typical high season. This contrasts, with the global trend where Summer is the clear peak season, provides an advantage for the NTBO to sustain a more consistent level of tourism throughout the year and shows that Portugal's favourable climate and cultural offerings likely contribute to this more even spread, benefiting the tourism industry by providing steady visitor numbers year-round (See Figure 5).

<u>Trip Types:</u> The data reveals that the most prevalent trip type is 'Couples', accounting for approximately 50% of the reviews. This is followed by 'Family' and 'Friends' trip types, suggesting a trend towards travel for leisure and personal relationships (See Figure 7). While looking specifically at Portugal, the scenario does not change that much either. Couples are still top of the list with a bit higher percentage, followed by Friends and Family, however, are evenly placed for second place (See Figure 6).

Holidays: Cross-referencing holiday data with the attractions and reviews allowed us to identify periods of peak and off-peak visitation. Our data shows that most people travel abroad during holidays time and: visit mostly Spain, England, Scotland, Italy, and Portugal; specially during January and May, and Spring and Winter seasons; go visit the Basilica of the Sagrada Familia (Spain), Tower of London (London), Edinburgh Castle (Scotland), Parc Guell (Spain) and Schönbrunn Palace (Austria); and the predominant trip types are 'Couples' and 'Family'. We found that the most popular holiday to travel is Labour Day. Portugal got 35070 visitors during this holiday for the period 2019-2021 (see Figures 7-11).

<u>Tourism Origins</u>: Interestingly, our data confirms what we saw in the tourism overview: most of the tourism origin in our dataset is not domestic, so NTBO should focus solely on targeting foreign tourists in its future marketing activities (see Figure 12).

#### VISITOR BEHAVIOR AND PREFERENCES PRE- AND POST-PANDEMIC

The dataset provides intriguing insights into the distribution of ratings and the evolution of visitor behaviour in response to the global pandemic.

Before the pandemic, the landscape of visitor ratings was characterized by generous approvals by the tourists, as is evident from the fact that 71% of the ratings were a full 5 stars, meaning exceptional satisfaction.

With the appearance of the pandemic, however, there has been an impressive shift in this trend. The data indicates a notable growth in the lowest ratings, with 1-star reviews becoming increasingly prevalent and the dominant 5-star ratings experienced a decline. This shift could be attributed to different factors, including potentially diminished service quality or visitor experiences as a direct consequence of stringent health and safety measures. Other pandemic-related impacts, such as restricted access to facilities and amenities or modifications to the visitor experience to comply with social distancing norms, could also have played a role in shaping these perceptions.

Despite the escalation in lower ratings, most reviews continue to skew positively, with 5-star ratings still representing the most common feedback post-pandemic.

re	viewRating	1	2	3	4	5
	Covid					
	After	0.040833	0.036472	0.068385	0.198811	0.655500
	Before	0.013140	0.015880	0.055831	0.203751	0.711398

Figure 13 – Review Ratings pre- and post-pandemic

It was also expected the volume of visits would decrease drastically considering the dataset has very few values, being the last input from August 2021. To have a better insight into how the pandemic affected visitor

behaviour, it would be ideal to have more user-generated content from different social media platforms such as Booking, Expedia, etc. (See Figure 14)

#### ATTRACTIONS ANALYSIS

Our analysis also tried to combine the different attraction types of attractions, such as Historical, Religious and Cultural Sites to identify the most common types of attractions by visitors, provides insights into visitor values and preferences, and it's a crucial factor in the strategic planning for tourism and culture development.

There is a wide range of attractions available, with a strong emphasis on historical and religious sites with a percentage of 42.37% and 22.03%, respectively. This emphasis aligns with Europe's renowned legacy, showcasing a continent deeply rooted in rich historical and cultural heritage (See Figure 15).

# **SENTIMENT ANALYSIS**

We analysed the sentiment of TripAdvisor reviews to find out what the predominant sentiment was among users and discovered that many reviews are positive, i.e., they contain positive textual aspects. We also looked at the degree of subjectivity of the reviews, which measures how emotional and personal the expressed sentiments are in reviews. Our dataset subjectivity results suggest that most reviews tend to be moderately subjective, followed by a high degree of subjectivity. This means that even our positive reviews are not especially opinion-based or emotionally charged. Looking into Portugal reviews we found that they tend, on average, to have a positive sentiment, and the content of these reviews is moderately subjective, indicating a mix of objective and subjective expressions (Figure 16).

#### **TEMPORAL ANALYSIS**

Drawing a time series analysis, our focus was to try to find trends over periods by understanding how the number of reviews and average ratings for locations have varied, specifically before and after the onset of the COVID-19 pandemic. We found that the monthly average reviews rating suffered an abrupt fall during the last months of 2020, when the COVID-19 pandemic hit, slightly increased and fell again during January 2021, when another pandemic outbreak and lockdown occurred (Figure 17). From January 2021 onwards the review ratings start slowly increasing. Considering seasonality for the most visits – in which Summer was predominant – and comparing it to holiday trips seasonality, where we found that Spring and Winter were the most chosen by visitors that travel on holiday days, we found it interesting that doesn't match with previous results.

# COMPARATIVE ANALYSIS WITH OTHER COUNTRIES

We looked at how Portugal compares with other countries in a dataset in terms of visitors' origin countries and trip types. We defined Portugal's main competitors by comparing them to the countries analysed in the last INE Tourism Satellite Account for Portugal 2022, cited before, that are present in our dataset, being: Spain, Hungary, Czech Republic, France, and Italy. The results show that the top visitors for Portugal come from the US, UK, Canada, and Australia, and for competitors, the top 3 are slightly inverted (UK, US, Australia, and Canada). Regarding trip types, we can see that for both Portugal and the main competitors 'Couples' is the main. For Portugal, it's followed by 'Friends' and 'Family', for competitors, the other way around (Figures 18-19).

#### MARKETING RECOMMENDATIONS

Based on the analysis that combines different attraction types, particularly highlighting the prevalence of Historical, Religious, and Cultural Sites, a good marketing strategy would be to establish partnerships with cultural institutions, museums, and historical preservation organizations. Collaborate on joint initiatives to promote the preservation of heritage sites and artefacts. Use these partnerships as a foundation for unique and exclusive visitor experiences.

By examining temporal analysis and seasonality for Portugal, we recommend NTBO to position the country as a year-round destination by developing marketing materials showcasing the diverse experiences available in every season and highlighting unique attractions, festivals, and events during spring and winter to attract holiday travellers during these periods.

By examining holidays travels patterns, marketing strategies can be fine-tuned to align with the expectations and experiences of visitors during specific holiday seasons, like designing a special marketing campaign for Labour Day, promoting exclusive deals, events, or packages for travel to Portugal during this period; offer seasonal promotions and discounts, especially during the months of January and May, as well as the Spring and Winter seasons; and maybe craft romantic getaways for couples during Labour Day or other popular travel periods. A potentially interesting further analysis to investigate could be to examine how holiday coincide with positive or negative sentiments in reviews.

Regarding positive sentiment analysis, it can allow NTBO to identify Portugal strengths as a tourist destination by helping to pinpoint specific attractions, services, or experiences that are highly praised by visitors.

# DEVELOPED MODELS AND FURTHER POSSIBLE ACTIONS

We used different methods to perform an analysis that could answer our business problems and questions. 

1st. To understand how tourists move between attractions, we started by using a method to find patterns in the places they visit. Imagine it like looking for common paths in a maze. We found that tourists are more likely to visit places that are near each other. So, attractions close in distance often ended up being visited together. For example, tourists who visit Casa Batló will visit the Sagrada Familia. Continuing with the same approach, we did another analysis. This time, instead of looking at individual attractions, we focused on countries to see which ones are most popular or frequently visited. An interesting finding was that Spain and Portugal are the top two countries that tourists tend to visit together, meaning that those who visit Spain, sooner or later, tend to go to Portugal as well. This insight is what we plan to leverage in developing our strategy.

2<sup>nd</sup>. The next approach was focusing on tourists who've been to Spain but not to Portugal and identifying similarities between them. First, we matched Spanish attractions with similar ones in Portugal using a similarity check and identified 'La Lonja de la Sela' as a similar attraction to 'Torre de Belém'.

	Spain_LocalID	Portugal_LocalID	Similarity
69	La Lonja de la Seda	Torre de Belém	0.133673
82	Palace of Catalan Music	Quinta da Regaleira	0.106347
18	Basilica of the Sagrada Familia	Ponte de Dom Luís I	0.105551
39	Catedral De Burgos	Ponte de Dom Luís I	0.097683
61	Cathedral De Santiago de Compostela	Quinta da Regaleira	0.088970
71	Mezquita Cathedral de Cordoba	Cais da Ribeira	0.081096
4	Acueduct of Segovia	Ponte de Dom Luís I	0.069518
96	Real Alcazar de Sevilla	Quinta da Regaleira	0.058904
50	Catedral de Sevilla	Cais da Ribeira	0.049833
9	Alcázar of Segovia	Mosteiro dos Jeronimos	0.049298
108	The Walls of Avila	Park and National Palace of Pena	0.040626

Figure 20 – Example of similar attractions between Spain and Portugal

To make our list perform even better, we made sure it only included people who enjoyed their time in Spain (those who gave a rating of 3 or more) – We cleaned up the list to show just the usernames and the Portuguese attractions we think they'll like based on the similarities we just referred and with this. Deliver a list of valuable usernames and the attractions NTBO should promote to these users.

	userName	Name	localID	Country	reviewRating	Recommended_Attraction
0	Maurice M@MauriceM171	Basilica of the Sagrada Familia	MAG001	Spain	5	Ponte de Dom Luís I
1	nickchamb@nickchamb	Basilica of the Sagrada Familia	MAG001	Spain	5	Ponte de Dom Luís I
2	Arthur Villon@arthur4588	Basilica of the Sagrada Familia	MAG001	Spain	5	Ponte de Dom Luís I
3	James U@696jamesu	Basilica of the Sagrada Familia	MAG001	Spain	5	Ponte de Dom Luís I
4	Kamal K Sirra@Sirra63	Basilica of the Sagrada Familia	MAG001	Spain	4	Ponte de Dom Luís I

Figure 21 – Example of the list with the users that might visit Portuguese attractions.

As we mentioned earlier, our initial analysis assessed the support and confidence levels between various attractions. We concluded that attractions with higher confidence levels are typically those located geographically closer to each other. To enhance our marketing campaign, we conducted a cluster analysis based on geographical coordinates, using latitude and longitude. We plan to provide these cluster analysis results to the marketing team as well. This will enable them to compare the 'Recommended Attraction' with the cluster file and suggest additional nearby attractions to the users. The necessity of creating these clusters arises from scenarios where a user might have visited multiple attractions but only reviewed one. Given the high confidence metrics for geographically proximate locations, we can utilize these clusters to suggest more potentially interesting attractions to the users.

3<sup>rd</sup>. We conducted a comprehensive analysis utilizing the RFM tool to categorize customers based on their behaviour.

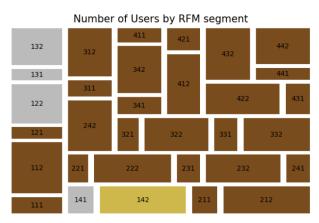


Figure 22 – RFM Segment Graphic

A "gold user" refers to individuals displaying substantial value and active engagement. These users have not only recently shared their feedback, signifying ongoing interaction with the platform, but also consistently contribute reviews, showcasing a loyal and robust relationship with TripAdvisor. Furthermore, their meaningful engagement adds value to the platform, underscoring their significance in contributing to its overall success. In essence, the term "gold user" characterizes those who, through their recent and frequent contributions, play a vital role in enhancing the quality and appeal of the platform, thereby contributing to its sustained growth and higher user satisfaction.

Among the identified segments, we have an exclusive group of 3013 **Gold Users**. For this distinguished segment, we are implementing a unique marketing strategy, offering them a special promotion valid for any attraction across Portugal.

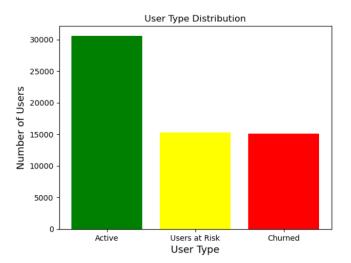


Figure 23 – User segmentation

While we acknowledge the presence of other user segments, such as Active users, Users at risk, and Churned users, our current focus is on the gold user segment. However, we emphasize the importance of considering these other segments in future campaigns.

We recommend tailoring distinct marketing strategies for each user segment. Active users will be the target of either a Cross-sell or an Up-sell campaign. Users identified as at risk will be the focal point of a Retention campaign, and Churned users will be the primary audience for a Reactivation campaign. This segmentation approach allows us to optimize our marketing efforts by aligning them with the specific needs and behaviours of each user group.

#### **CONCLUSION**

In conclusion, this report, through its meticulous application of advanced data science techniques like RFM segmentation, sentiment analysis, and the strategic use of similarity and Apriori algorithms, offers a deep dive into the evolving patterns of tourism in Portugal and its main competitors. By carefully analysing visitor behaviours, preferences, and demographic shifts, especially in the context of the COVID-19 pandemic, we have crafted a detailed picture of the current tourism landscape. This analysis not only provides the Portuguese NTBO with actionable insights for targeted marketing and improving visitor experiences but also underscores the importance of adapting to dynamic global trends. The implementation of data-driven strategies, as demonstrated in our report, is key to revitalizing and sustaining the allure of Portugal as a prime tourist destination.

Following that idea, key takeaways include the need for targeted marketing strategies tailored to visitor segments, a focus on enhancing visitor satisfaction through sentiment analysis, and adapting to changing patterns post-pandemic. The prevalence of historical and cultural attractions suggests a unique selling point, offering opportunities for themed experiences and educational initiatives. NTBO can leverage these insights to craft strategic campaigns that resonate with diverse traveller interests, foster recovery, and position Portugal as a resilient and enriching destination.

Overall, we believe that the project magnitude could benefit from a delivery calendar extension given the different specific potentialities this analysis can target. Time was a limiting factor for a deeper understanding of how all the results relate to each other. That said, we believe that we target all business objectives proposed prior. We believe that these objectives can be further developed by NTBO and converted into marketing strategies.

#### **APPENDICES**

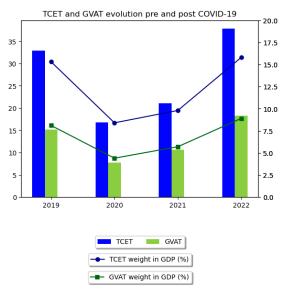


Figure 1 - Tourism Consumption in the Economic Territory (TCET) and Gross Value Added generated by Tourism (GVAT) evolution in million years and weight in GDP (%) for 2019-2022. (source: INE, Turismo de Portugal)

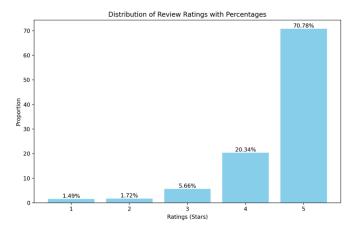


Figure 2 - Rating distribution with Percentage.

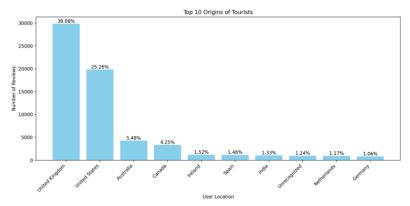


Figure 3 – Top 10 Countries visited by tourists.

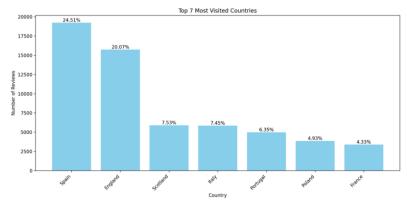


Figure 4 – Top 7 Most Visited Countries.

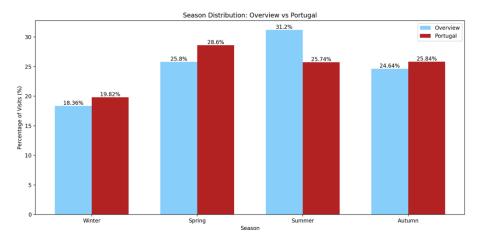


Figure 5 – Season distribution comparison between the overview dataset vs Portugal.

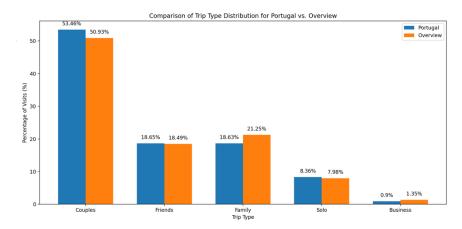


Figure 6 – Comparison of trip type distribution for Portugal vs overview

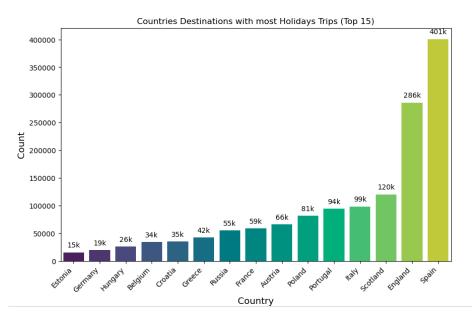


Figure 7 - Top 15 destination countries for holidays trips.

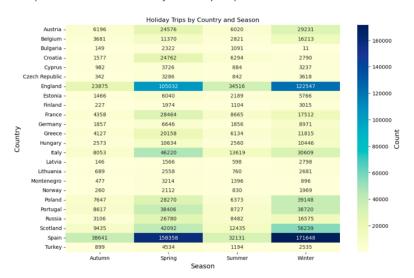


Figure 8 – Holidays trips by country and season.

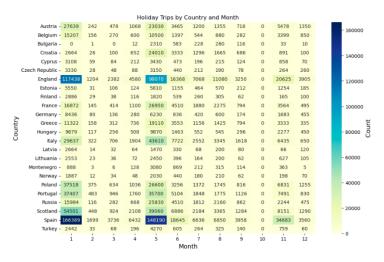


Figure 9 - Holidays trips by country and month.

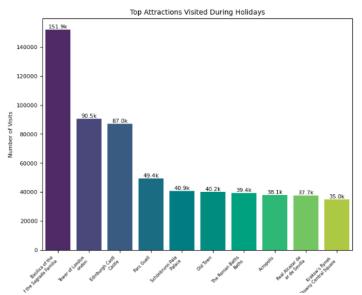


Figure 10 – Top Attractions for holiday trips.

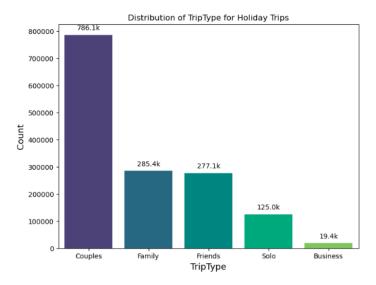


Figure 11 – Trip type distribution for holidays trips.

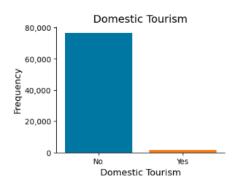


Figure 12 – Comparison frequency of Domestic Tourism in overall dataset.

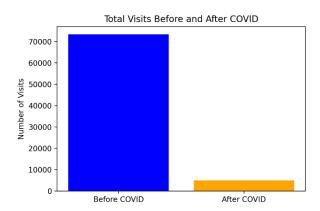


Figure 14 – Total visits before and after COVID.

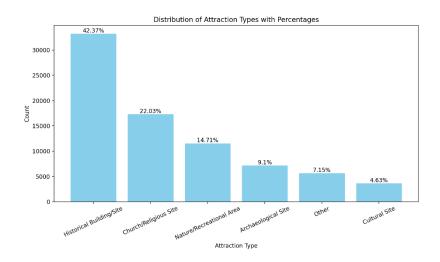


Figure 15 – Distribution of attraction types.

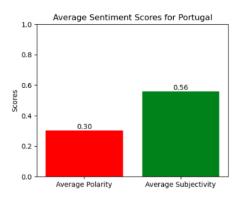


Figure 16 - Average sentiment scores for Portugal from TripAdvisor users' reviews.

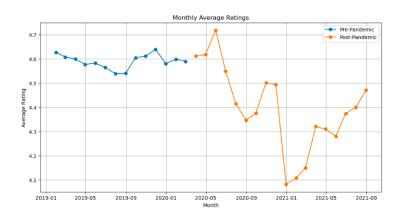


Figure 17 – Monthly average reviews rating pre (2019) and post pandemic years (2020 and 2021).

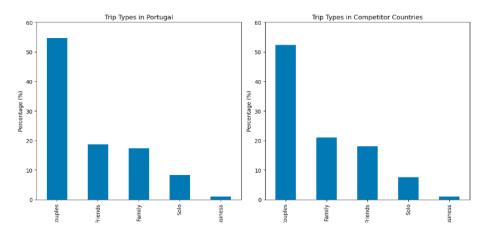


Figure 18-Comparison of trip types and visitor origins between Portugal and competitor countries.

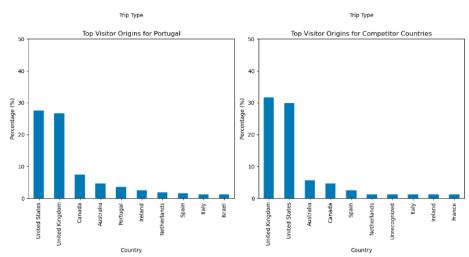


Figure 19 – Comparison of trip types and visitor origins between Portugal and competitor countries.

# **REFERENCES**

Conta Satélite do Turismo para Portugal 2022 (INE, Junho 2023):

https://www.gee.gov.pt/pt/indicadores-diarios/ultimos-indicadores/32941-conta-satelite-do-turismo-para-portugal-ine-2 (Relatório completo:

https://www.ine.pt/ngt\_server/attachfileu.jsp?look\_parentBoui=618144221&att\_display=n&att\_download=y)

TripAdvisor (6sense Technographic Insights): <a href="https://6sense.com/tech/reservation-and-online-booking/tripadvisor-market-share">https://6sense.com/tech/reservation-and-online-booking/tripadvisor-market-share</a>