

Data Science

An Evaluation of Portugal's Market
Position and its Competitors in
tourism based on TripAdvisor
Reviews

Emrecañ Duran | 123456789

January
2023

TABLE OF CONTENTS

03

Case Description

06

Overview of Portugal

03

Business Understanding

07

Competitors' Overview

04

Data Understanding

09

Holidays

05

Data Cleaning

11

Source of Visitors: UserLocation

05

Briefing

15

Modeling

06

Data Preparation

16

A Path forward

Data Understanding

After uploading the Reviews dataset and reviewing the general information it became apparent that some columns are not complete and they have missing values, in particular, the country of origin of the users, the time of travel, and what type of travel it was.

Furthermore, upon merging the reviews and the attractions datasets together it was noted the destination country list and attraction destinations for tourism also had missing values.

At the first glance, the user location info should still be normalized as the results may be distorted. Most visitors to Portugal are from the United Kingdom and are traveling with their partners.

The rest of the primary understanding of the dataset is presented in bullet points:

- There are 5 trip types in total
- Some same users may travel to different attractions places or countries.

Their behavior might be interesting to pay attention to. Yet it is essential to check them against any possible duplicates.

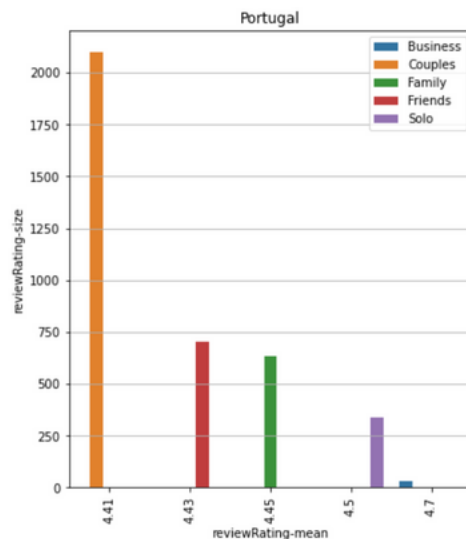
- The user Location column flags out excessive unique numbers. It is a must to investigate to comprehend what is happening there.
- The number one tourist destination is Spain and sightseeing is the Basilica of the Sagrada Familia leaning on the number of reviews written.
- All the reviews were written in English
- Most visits occurred in the year of 2019
- The timeframe of the visit data ranges from 2015 to 2021
- The average review rating on Tripadvisor was 4.58 with a minimum rating of 1 and a maximum of 5.

In the meantime, it was detected that the year 2019 had an overwhelmingly huge number of reviews in comparison to other yearly data. The reason can be rooted in sudden changes to TripAdvisor's popularity. No online evidence was found to support the statement. Yet, the number of differences could also be related to Trip Advisor's data collection.

Additionally, a question was brought up to the team's attention: what is meant by review rating metrics? Is it only the attraction itself or other factors might influence forming an opinion? It was decided that the rating would explicitly refer to the attraction itself.

In parallel, COVID-19 regulation and isolation policies are varying for different geolocations.

These were the team's initial observations.



According to the bar plot above, we can clearly see 3 distinct patterns:

- Most of the reviews are coming from couples as they provided more than 2000 reviews, however, they have the lowest average of reviews among the categories in question.
- At a lower rate, Portugal is visited by people visiting as a family or group of friends with almost 1400 reviews averaging 4.44.
- Finally, we find only a few people have visited Portugal for business or as one personal visit with a combined total of 300 reviews having the highest average among all the groups.

Outcome:

Portugal is very appealing to couples as it has a lot to offer especially to couples as it offers scenic beauty, delicious cuisine perfect for romantic meals, and an intimate dining experience.

- We recommend focusing more on improving the quality of service offered for couples and highlighting them as they are the most important category of visitors to Portugal.

It is clear that Portugal is highly admired as a business destination, as a direct result of numerous factors that makes this country a business-friendly environment.

- We recommend that we focus more on the aspect of attracting business and promoting international events focused on new technologies in order to attract businesses and business people as they will become the most profitable category of visitors and tourists if a strategy is implemented correctly. Portugal should promote itself more in order to increase awareness about its offerings to those who are coming here for a business trip. Some short tourism packages and day tours can be offered. Arranging some agreements with airline companies or Web Summit kind of big business events can help make Portugal more popular and desired by business travelers.

In our opinion, we consider "Solo" visitors as independent travelers, as Portugal has been recently a very popular choice for digital nomads and globetrotters considering the affordable cost of living, various residency options, and the expansion of co-working spaces across the country. A great solution to capture this opportunity and attract those more by developing targeted marketing campaigns to create more awareness about the working visa options such as the D7 visa for digital nomads and highlighting the benefits such as cost of living, start-up initiatives, and freelancer benefits, high-speed internet, general weather conditions, and vibrant cities.

Modeling

We will be applying below an apriori algorithm for this analysis in search of associations among the countries as well as the visitor profiles of Portugal.

	support	itemsets
18	0.267193	(United Kingdom)
16	0.245478	(Spain)
10	0.091728	(Italy)
14	0.064621	(Portugal)
13	0.059115	(Poland)
6	0.055340	(France)
0	0.049001	(Austria)
8	0.046354	(Greece)
15	0.035636	(Russia)
2	0.035276	(Croatia)
1	0.029721	(Belgium)

Associations among the countries

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
1	(Portugal)	(Spain)	0.064621	0.245478	0.005572	0.086220	0.351233	-0.010291	0.825715
0	(Spain)	(Portugal)	0.245478	0.064621	0.005572	0.022697	0.351233	-0.010291	0.957102
2	(Spain)	(United Kingdom)	0.245478	0.267193	0.005196	0.021166	0.079217	-0.060394	0.748653
3	(United Kingdom)	(Spain)	0.267193	0.245478	0.005196	0.019446	0.079217	-0.060394	0.769485

- Nearly 6% of all the reviews made is for Portugal, meaning that, Portugal's support is 6% whereas Spain's support is almost 25%.
- Nearly 0,5% percentage of all the reviews is made for both Portugal and Spain by a user, meaning that Portugal-Spain itemset's support level is nearly 0,5%.
- The confidence level of users who visited and reviewed Portugal also reviewing Spain is approximately 9% meaning that a user reviewing Portugal will also review Spain with a probability of 9%.
- On the other hand, the confidence level is nearly 2% for Spain. It means that, a user reviewing Spain will also review Portugal with a probability of 2%.
- This information can create an idea the users reviewing Portugal associates to Spain more than the users reviewing Spain associates to Portugal.
- In addition to that, users that reviewed Spain, also reviewed United Kingdom with a percentage of 2%, meaning that United Kingdom might be the competition of Portugal for users reviewing Spain.

Snapshot of recommendations

- Reach out to visitors who are planning trips to Spain and offer some discounts and exclusive offers to them as a complementary part of their trip.
- Portugal is a destination for couples. Endorse romantic activities in Portugal. We can track and target people who are searching for a couple's trip to Europe and display the ads to those.
- Increase marketing activities for user locations from our neighbour countries as they are distantly and culturally near Portugal but they are not the most frequent visitors of Portugal.
- The ratings for Portugal from the top visitor user Locations that are the UK and Spain are lower than average. It might be relevant to make a text mining analysis to figure out the cause of lower satisfaction rate.
- Data Mining analysis to have a better understanding of the way for the low ratings rather than the what. Why are people leaving lower reviews for Portugal rather than for the competitors?
- Target the Business trip travellers and monitor their rating behavior with more substantial data.
- Side suggestion: Encourage visitors to leave more ratings on TripAdvisor for two reasons: spreading popularity and gaining more data for further analysis and improvements.
- Reach more exposure in Middle and Eastern Europe which are giving higher rates to Portugal. Also, more exposure would be helpful to collect more data and test those geolocations' behaviour. The initial assumption is leaning on high ratings from Slovakia, Slovenia etc. Yet, the data is not sizeable.