

National Tourism Promotion

DATA SCIENCE FOR MARKETING
FINAL PROJECT





INDEX



- 1. Introduction**
- 2. Business Understanding**
- 3. Data Visualization**
- 4. Data Understanding**
- 5. Data Preparation**
- 6. Data Similarity Analysis**
- 7. Market Basket Analysis**
- 8. RFM Analysis**
- 9. Recommendations and Insights**



INTRODUCTION

The COVID-19 pandemic had a negative impact on many industries and lives. With all restrictions and safety measures, tourism was intensely affected and now companies have to be creative in order to keep in business.

In this project, we act as a consultant for NTBO (National Tourism Board Organizations), guiding them to achieve its previously identified objective: to study the market through the UGG (User Generated Content).

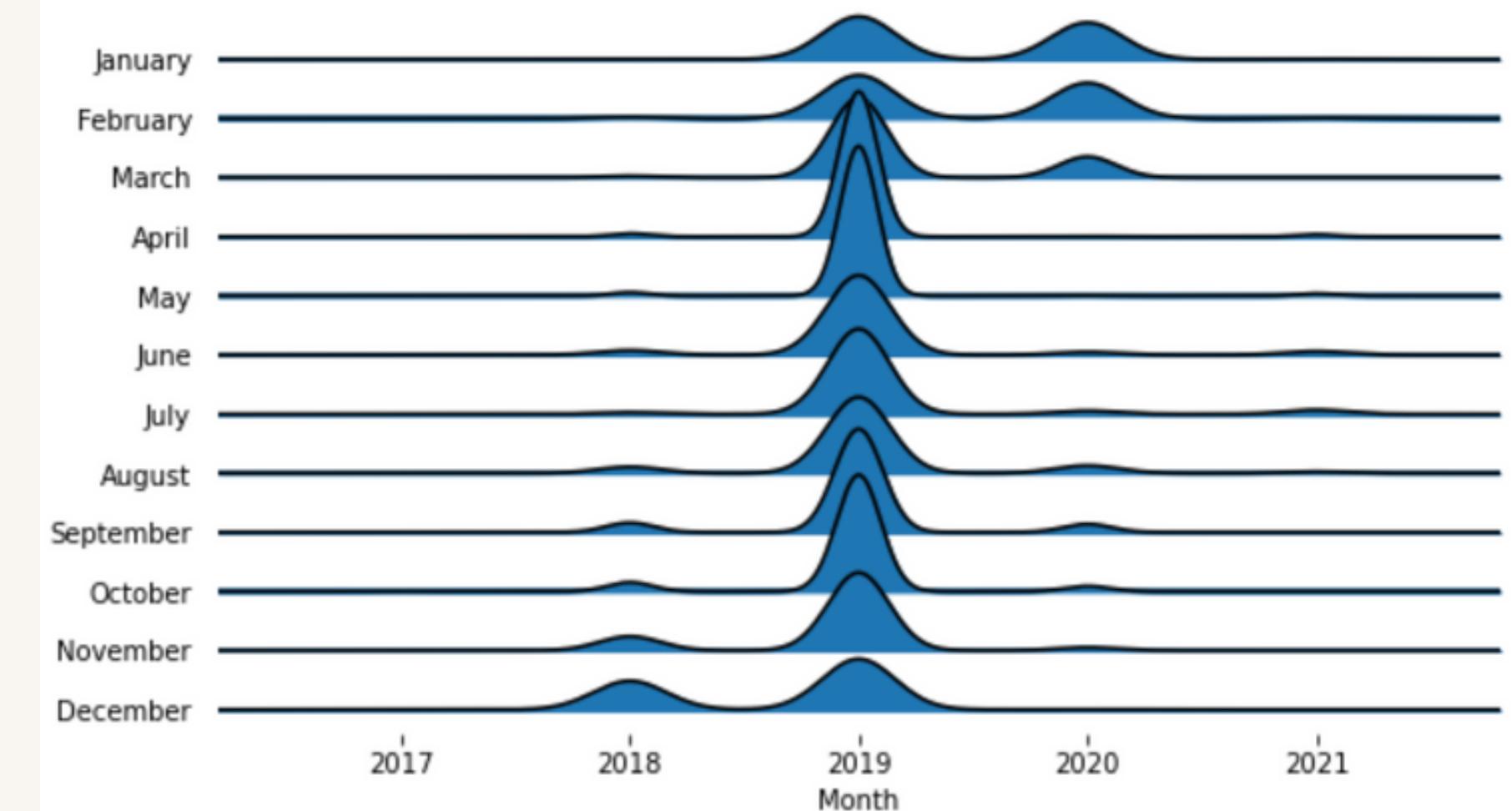
We used the CRISP-DM model to characterize and describe patterns of visitors to Portuguese attractions and compare them with other countries/attractions.



BUSINESS UNDERSTANDING

In 2021, the tourism percentage worldwide increased by 4% compared to the year before – the year the COVID-19 pandemic started.

This number is still 72% inferior to the numbers of 2019.



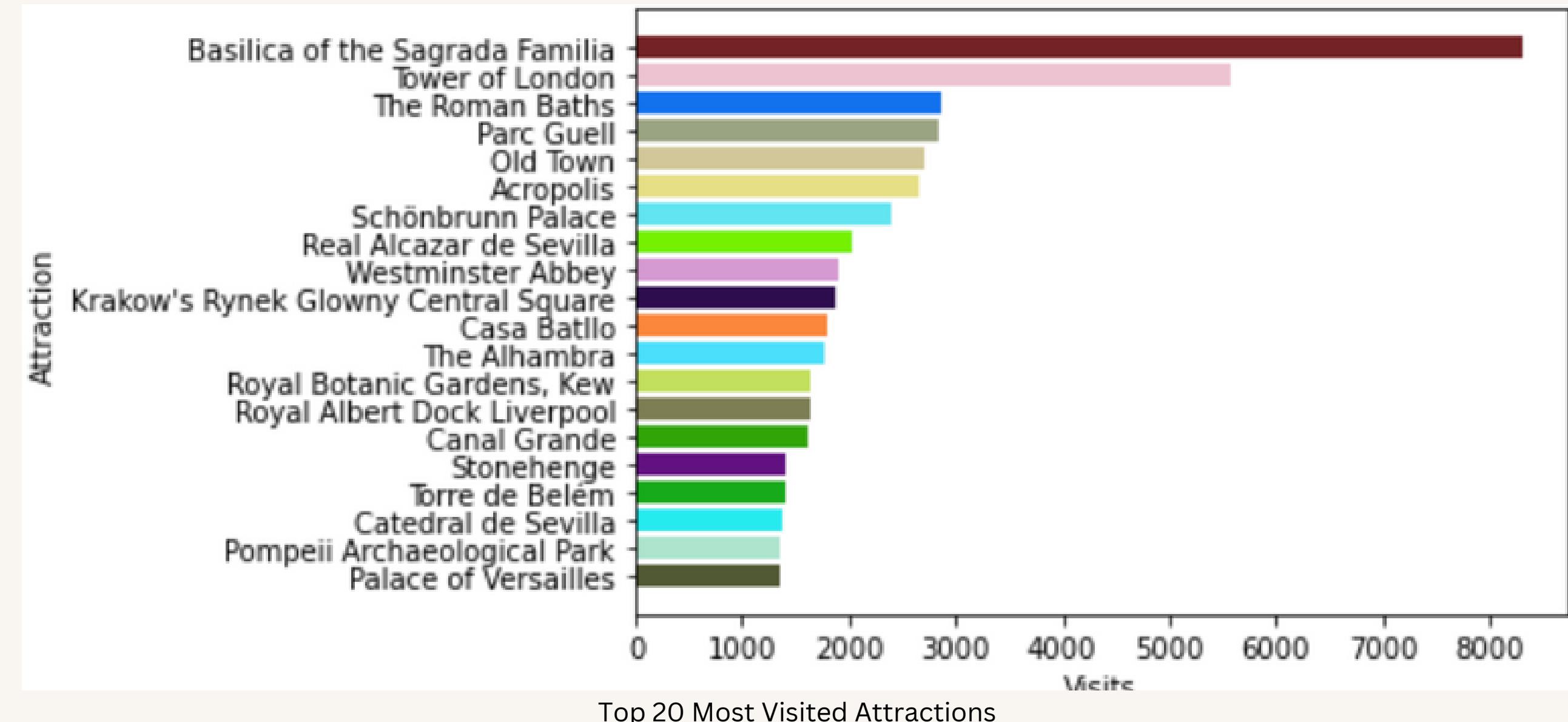
Tourism volume in Portugal by Month and Year



DATA VISUALIZATION

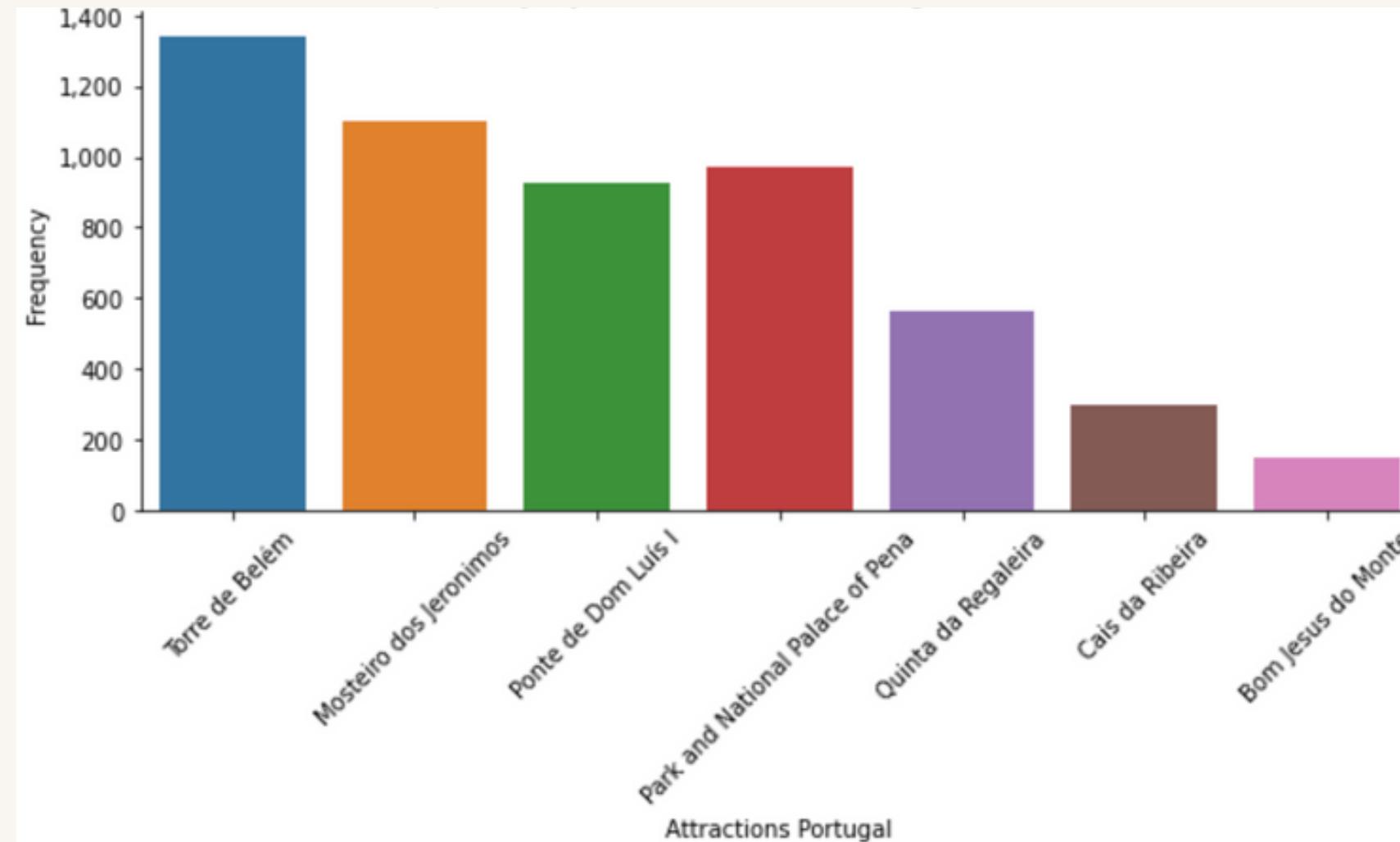
Spain has the most visited attractions.

Portugal can be seen in the top 20, with Torre de Belém as the 17th most reviewed attraction.

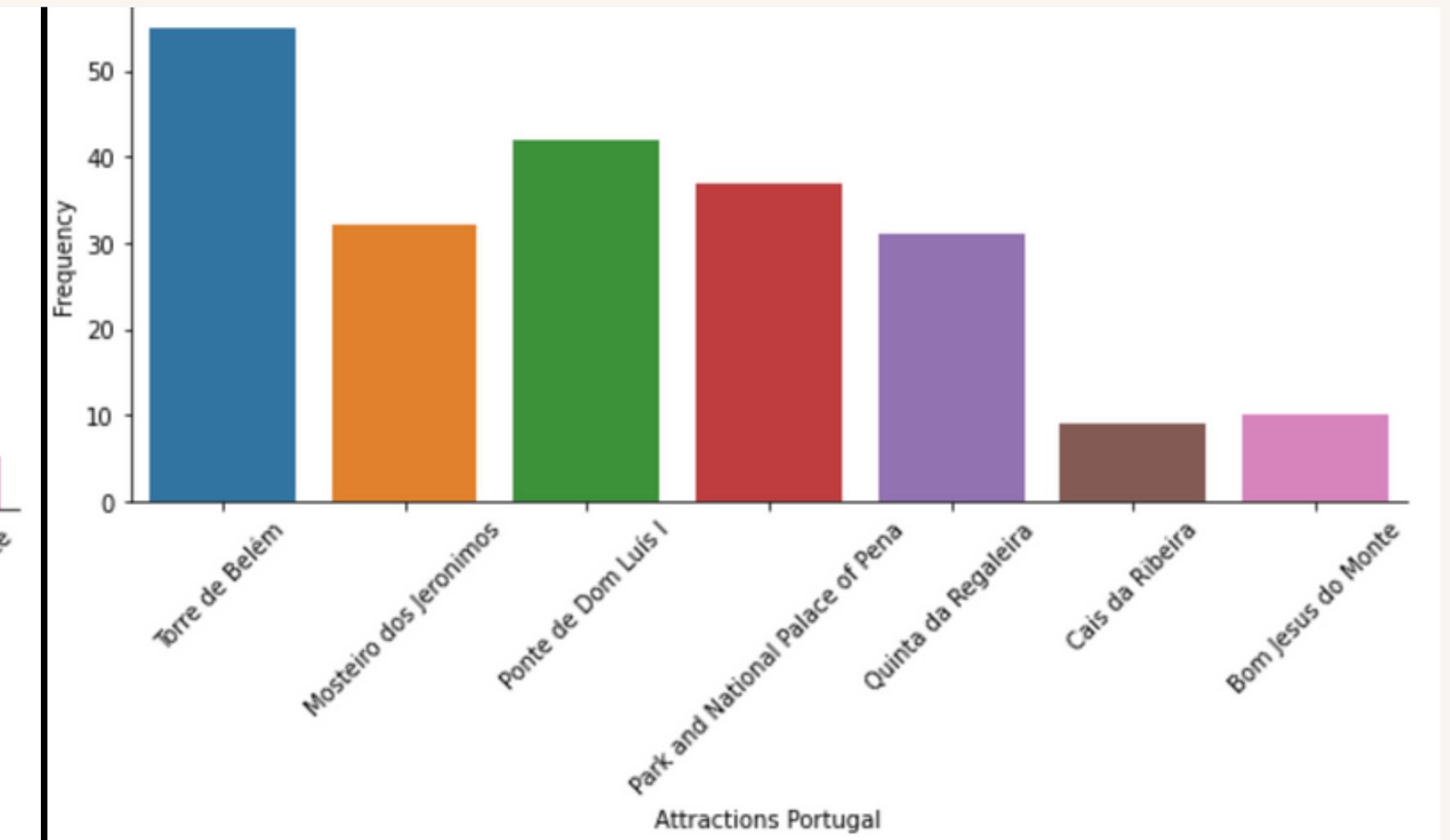




DATA VISUALIZATION



Frequency by Attractions in Portugal pre-pandemic



Frequency by Attractions in Portugal pandemic

DATA VISUALIZATION



Before merging the data using Python, we counted the percentage of Portuguese attractions on the dataset available.

The most reviewed was **Torre de Belém**.

attraction	count	part
Torre de Belém	1397	1.61%
Mosteiro dos Jeronimos	1130	1.31%
Ponte de Dom Luís I	967	1.12%
Park and National Palace of Pena	1007	1.16%
Quinta da Regaleira	595	0.69%
Cais da Ribeira	304	0.35%
Bom Jesus do Monte	159	0.18%
Total	86560	

Count and percentage of the Portuguese attractions from the dataset





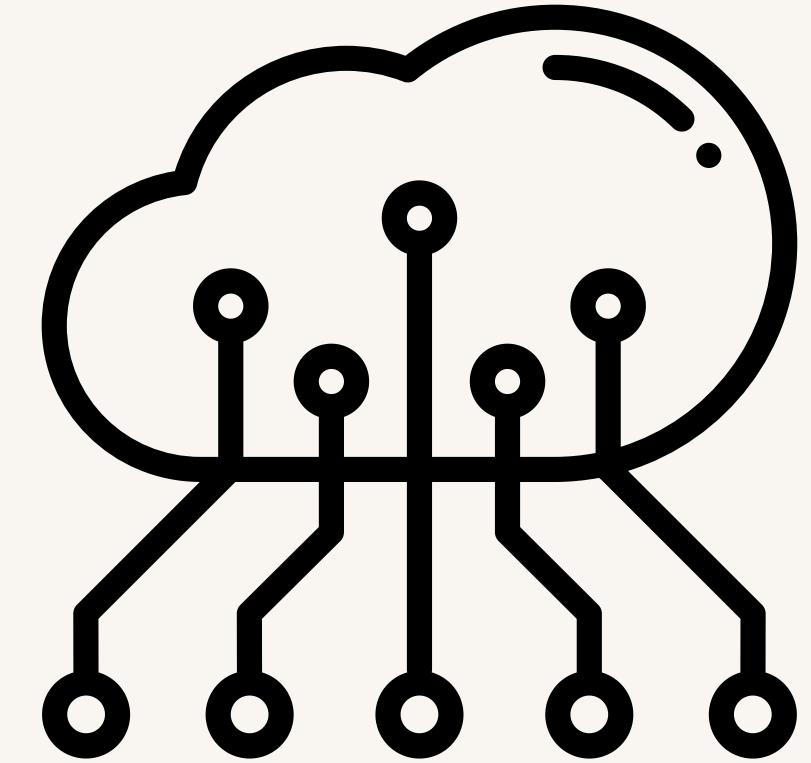
The best-rated attraction was **Quinta da Regaleira**, with more than 80% of its ratings on 5 stars. The worst was **Park and National Palace of Pena**, with the most “1 star” ratings about 5% of its total ratings.

Bom Jesus do Monte is also a well-rated attraction, with most 5 and 4 stars. The only attraction with zero “1-star” rating is **Cais da Ribeira**.

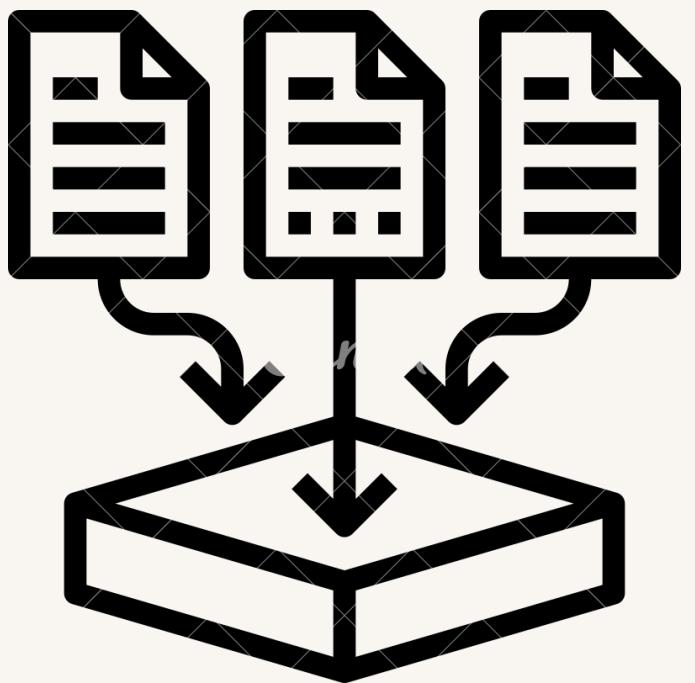
attraction	5 stars	4 stars	3 stars	2 stars	1 stars
Torre de Belém	44.45%	36.72%	15.25%	2.43%	1.15%
Mosteiro dos Jerónimos	59.82%	29.12%	7.43%	1.95%	1.68%
Ponte de Dom Luís I	73.84%	23.16%	2.69%	0.21%	0.10%
Park and National Palace of Pena	53.92%	25.82%	10.92%	4.77%	4.57%
Quinta da Regaleira	83.03%	14.12%	2.02%	0.34%	0.50%
Cais da Ribeira	64.80%	30.26%	4.28%	0.66%	0.00%
Bom Jesus do Monte	77.36%	16.98%	4.40%	0.63%	0.63%

How the Portuguese attractions are reviewed by tourist

DATA PREPARATION



- The dataset has **92.120 rows** and **16 columns**;
- There are a total of **92.120 reviews**;
- **6.03%** of the total is reviews of Portuguese attractions;
- There are **97** attractions and the most visited one was the **Basilica de La Sagrada Familia** (Barcelona, Spain)
- Most of the users came from London, UK;
- All reviews are in English;
- The main review's rating is **4.58/5**.





DATA PREPARATION

In order to generate our analysis focusing on Portugal's attraction, we needed to first **merge the data**.

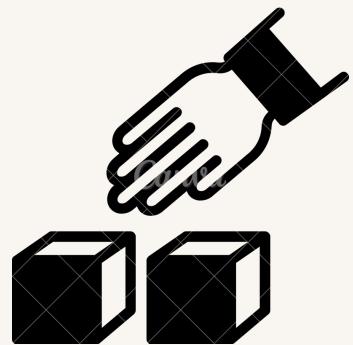
We merged the Reviews data framework with the Attractions data framework, and filtered the merged data frame to include only the rows where the "Country" column was equal to "Portugal".

reviewID	attractionDate	globalRating	positionOrRanking	sitesOrRanking	totalReviews	username	userLocation	userContributions	tripType	reviewWritten	reviewVisited	reviewRating	reviewLanguage	reviewFullText	Name	Country	ISO
0	2021-08-20 08:24:40.077	4.5	1	1109	169828	Maurice MauriceMTI	Indianapolis, IN	29	Couples	2021-08-18	2021-08-01	5	en	Absolutely breathtaking use of light; an amaz...	Basilica of the Sagrada Família	Spain	ES
1	2021-08-20 08:24:40.097	4.5	1	1109	169828	Nicholas NicholasN	London, Ireland	7	Couples	2021-08-17	2021-08-01	5	en	A really interesting and beautiful place to v...	Basilica of the Sagrada Família	Spain	ES
2	2021-08-20 08:24:40.117	4.5	1	1109	169828	Arthur Villegas ArthurVillegas101	Valencia, Spain	268	Friends	2021-08-16	2021-08-01	5	en	Beautiful place and must visit place in Barcel...	Basilica of the Sagrada Família	Spain	ES
3	2021-08-20 08:24:40.147	4.5	1	1109	169828	James U JamesU	n/a	7	Fun	2021-08-16	2021-07-01	5	en	A MUST when visiting Barcelona. It is conce...	Basilica of the Sagrada Família	Spain	ES
4	2021-08-20 08:24:40.159	4.5	1	1109	169828	Kamal C KamalCSirra@Gmail	Madrid, Spain	516	Family	2021-08-15	2021-07-01	5	en	I visited about 20 years ago and there has bee...	Basilica of the Sagrada Família	Spain	ES

DATA SIMILARITY ANALYSIS



Objective: generate a matrix that would show us the similarities between the Attraction's name, the userName (meaning the user who visited the attractions) and the reviewRating.



Similarity: refers to the numeric measure of how alike two data objects are.



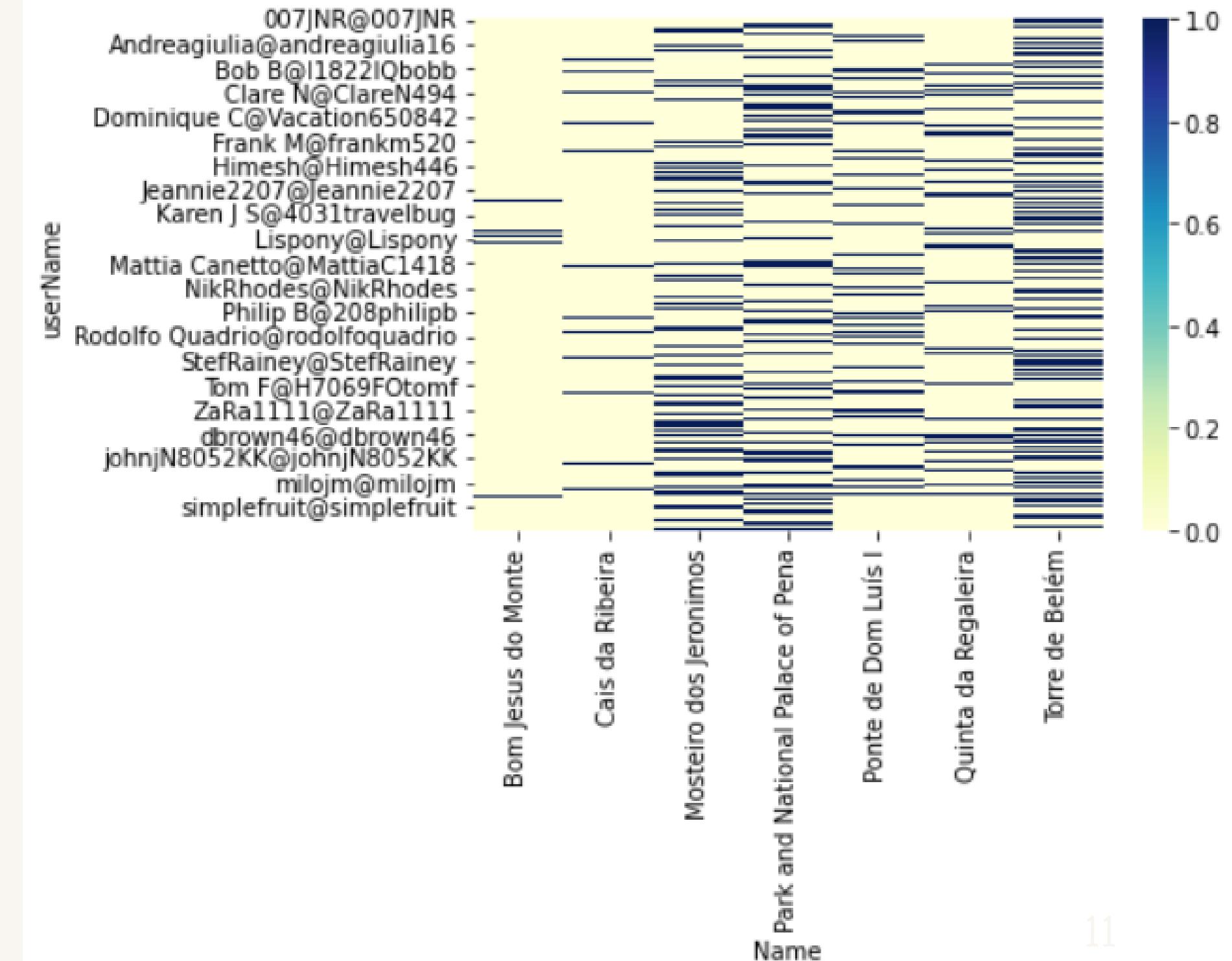
DATA SIMILARITY



This **heatmap** shows all the userNames that have gone to one or more attractions and have rated it with a review.

We identify that they are **7** most visited attractions in Portugal and **3.957** tourists that visited these attractions.

HEATMAP



DATA SIMILARITY

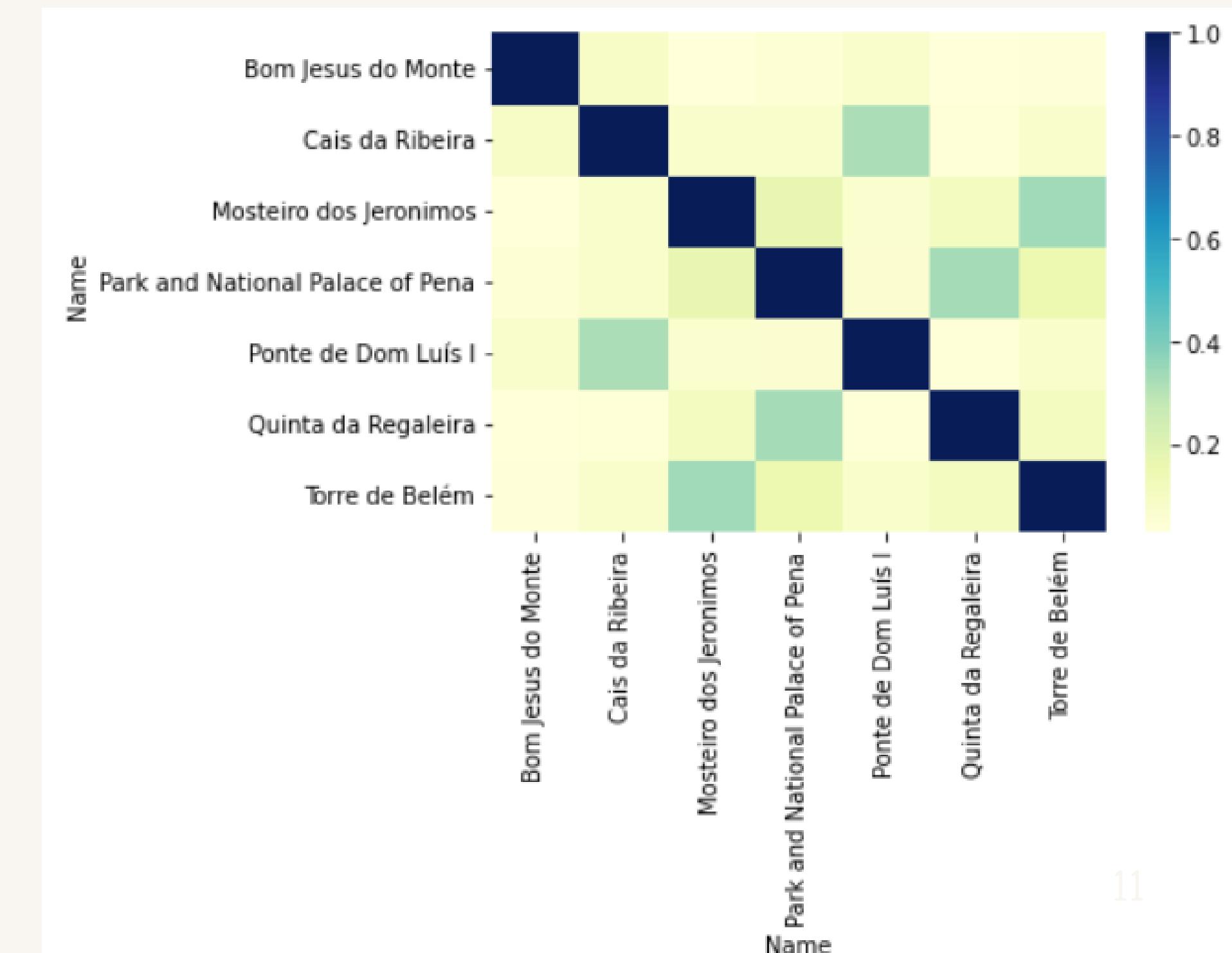


Attraction to Attraction Similarity Analysis

This Heatmap shows **which Attractions are more similar** to one another based on the rating that the users have given them. So, to userNames who have visited for example "Torre de Belém" will be most likely visiting "Mosteiro dos Jerónimos" if we offer it to them.

This help us to identify which attractions could be sell together as a package in a marketing plan, or which one we should be advertising together.

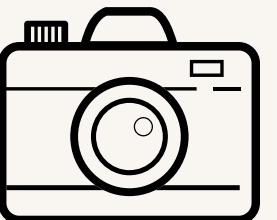
HEATMAP



MARKET BASKET ANALYSIS



Objective: Understand the consumer's behaviour, and illustrate which attractions are most visited in conjunction, and provide recommendations based on these findings.



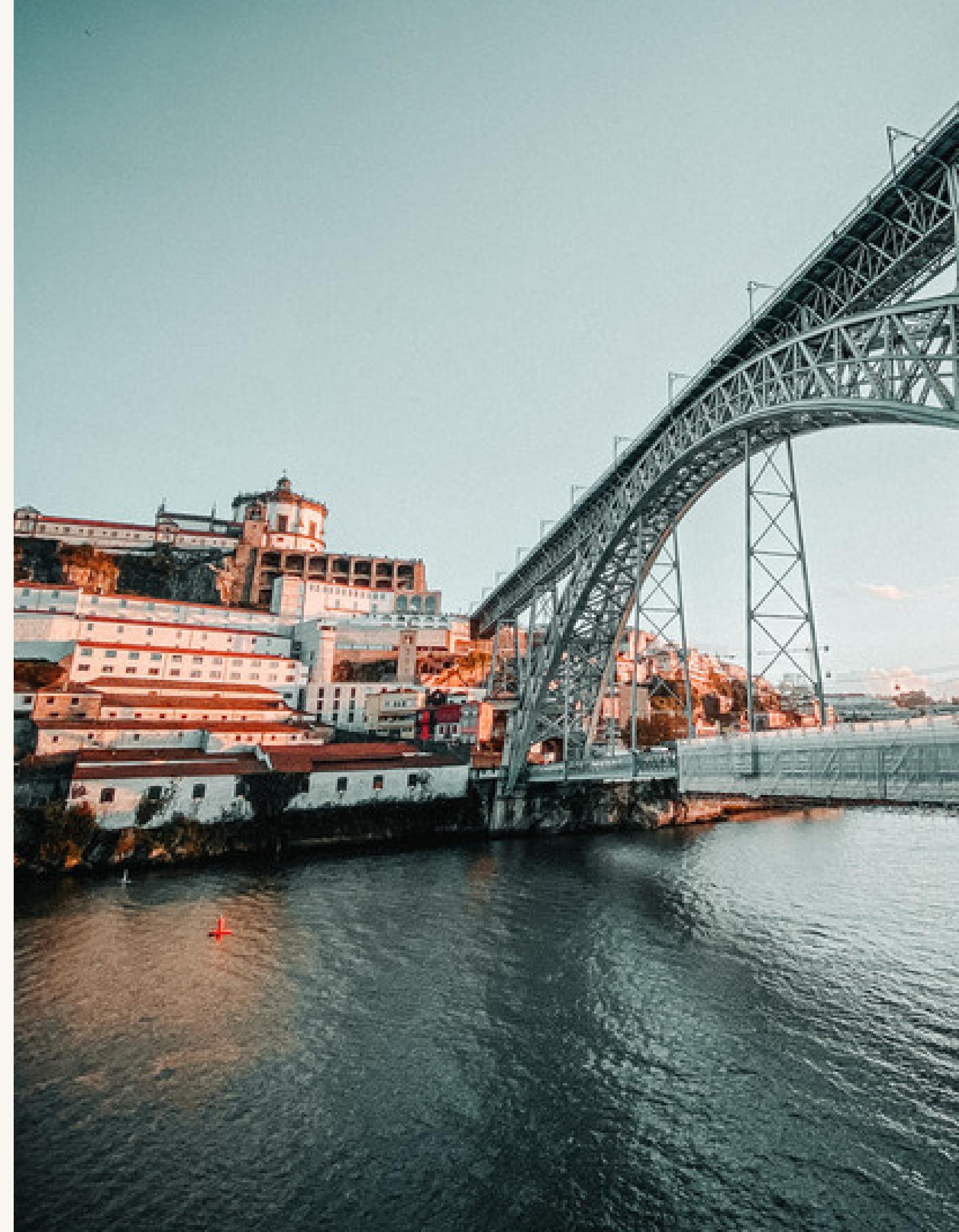
Analysis was done using the **Apriori algorithm**.



TOP 3 COMBINATIONS

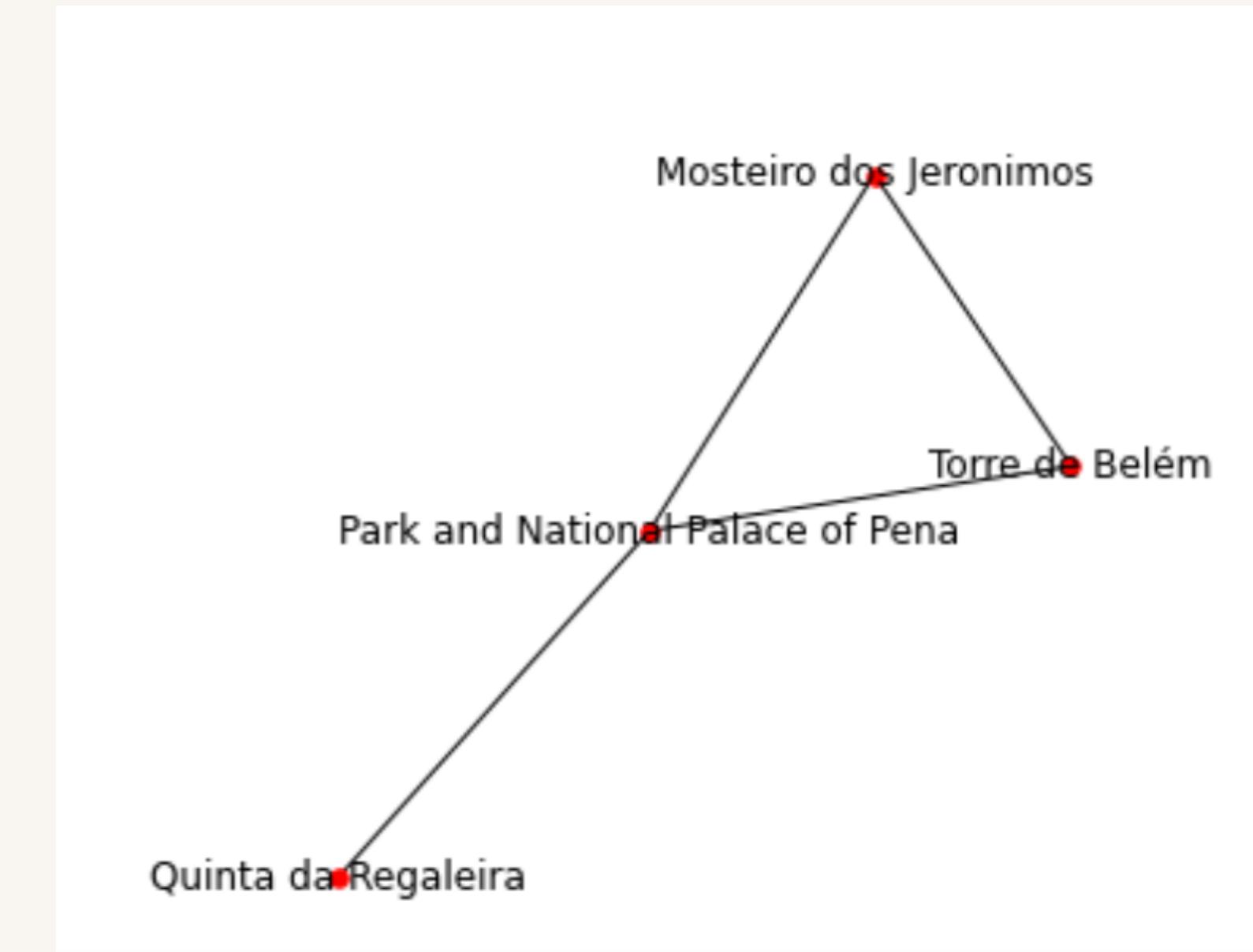
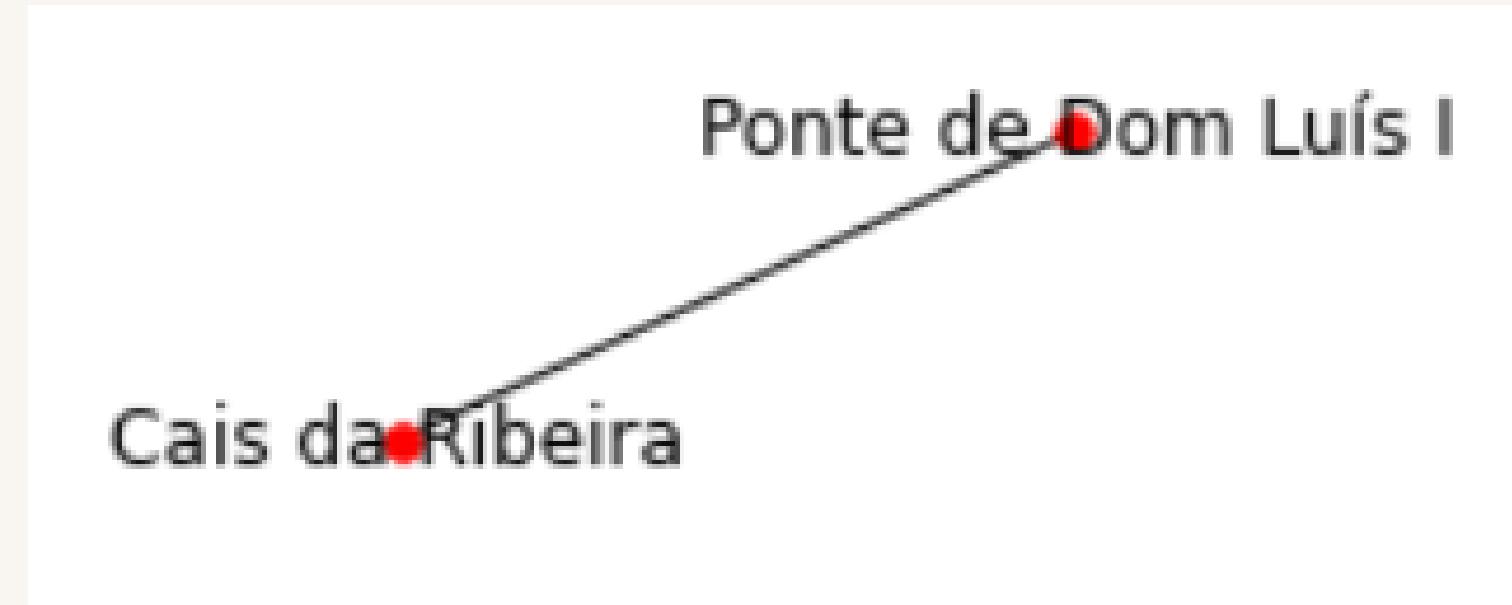


- Cais da Ribeira and Ponte de Dom Luís I,
Porto
- Quinta da Regaleira and Park and National
Palace of Pena, Sintra
- Mosteiro dos Jerónimos and Torre de
Belém, Belém





NETWORKS OF ATTRACTIONS



RFM ANALYSIS



Objective: identify different segments of travelers to align the best strategy to impact them, promoting new trips, reviews comments, and improving reviews rating.



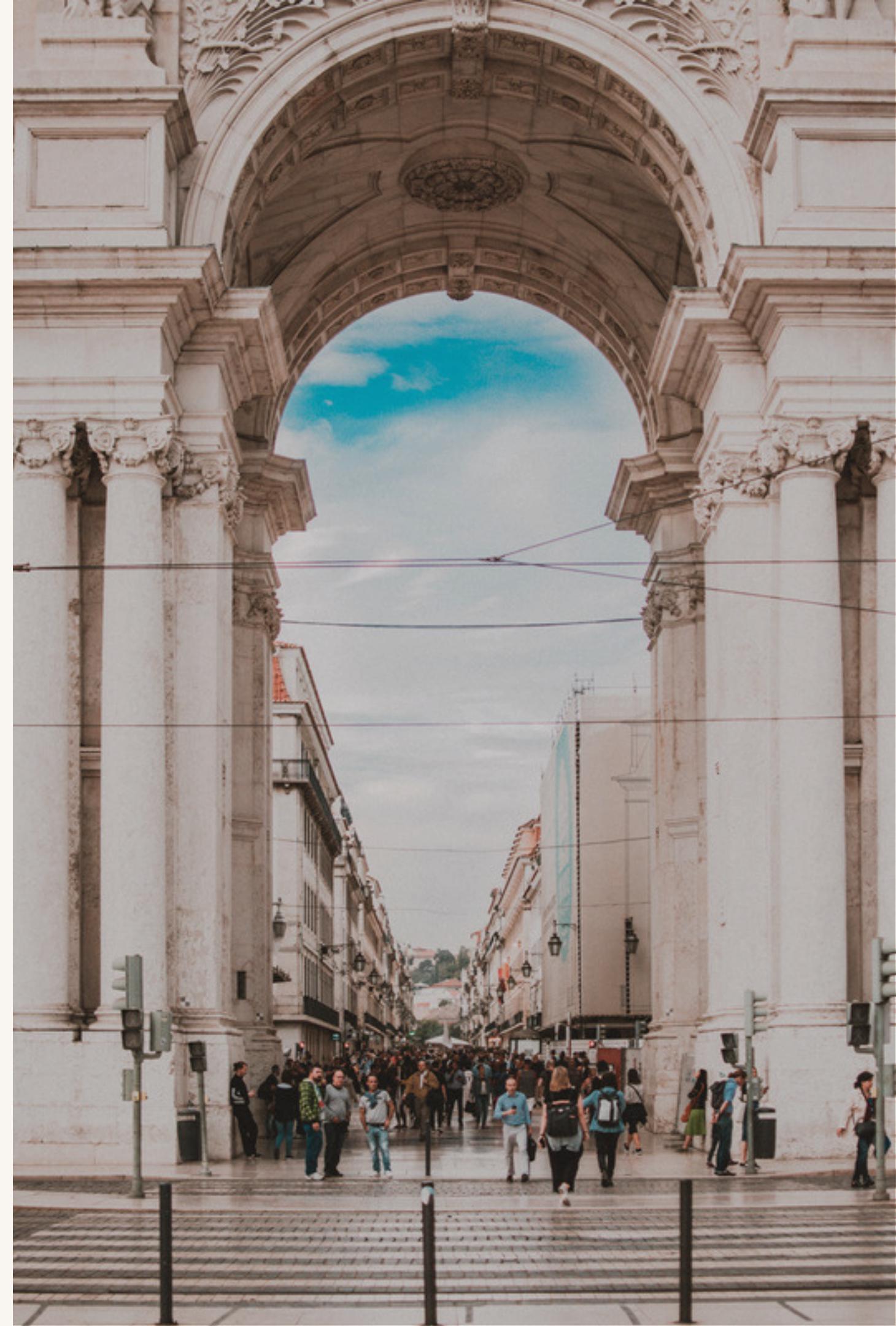
Recency = reviewVisited



Frequency = userContributions



Monetary = reviewRating



RFM ANALYSIS

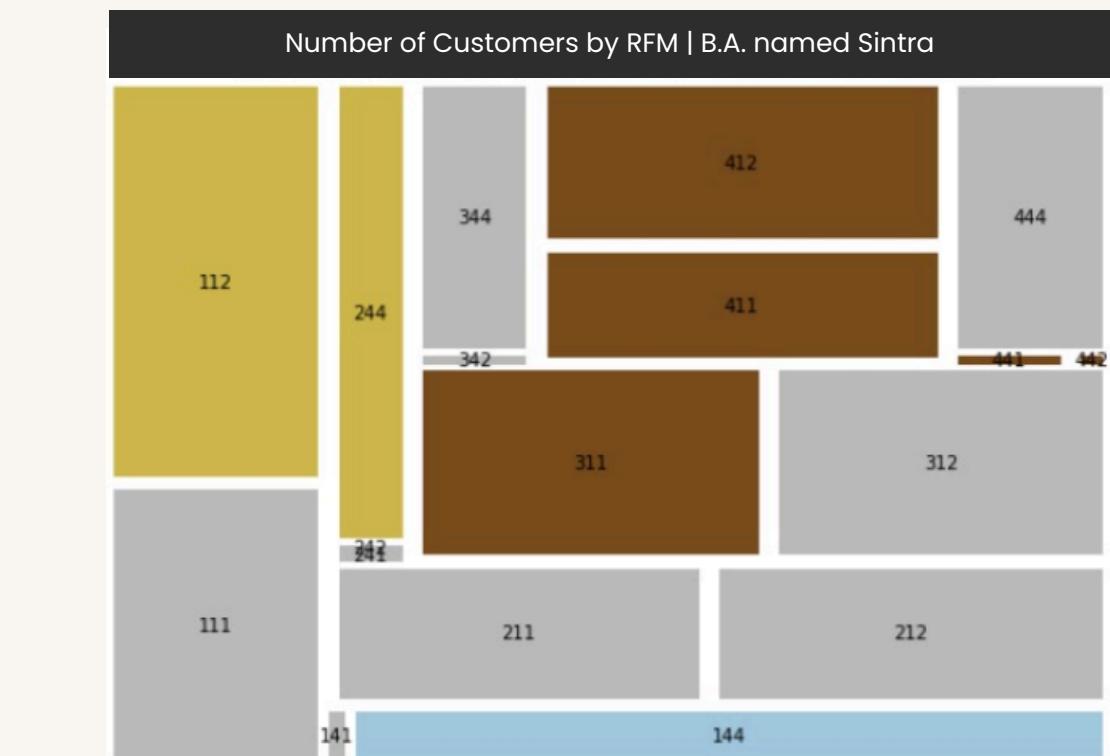
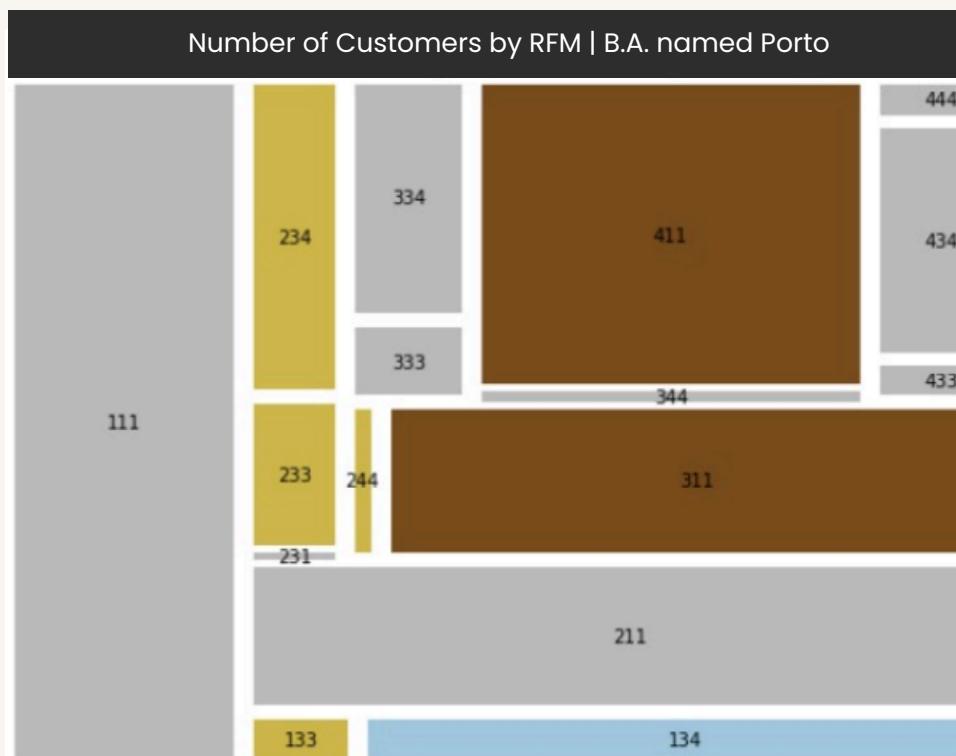


Segments:

- 1 and 2 The most loyal ones
- 3 Retention ones (at risk)
- 4 The churned ones

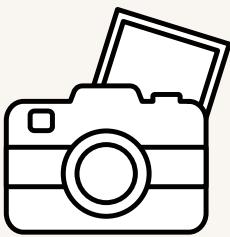


Recency Table						
R	F/M	1	2	3	4	
4	1	411	4	412	413	414
4	2	421	422	423	424	
4	3	431	432	433	434	
4	4	441	442	443	444	
3	1	311	312	313	314	
3	2	321	322	323	324	
3	3	331	332	333	334	
3	4	341	342	343	344	
2	1	211	212	213	214	
2	2	221	222	223	224	
2	3	231	232	233	234	
2	4	241	242	243	244	
1	1	111	112	113	114	
1	2	121	122	123	124	
1	3	131	132	133	134	
1	4	141	142	143	144	



RECOMMENDATIONS AND INSIGHTS

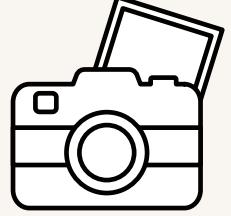


 Create a package which creates a link with Porto, therefore, a 4-day package could be created to include all 3 combinations. For instance, you could create a package which would have tours to Belém and Sintra, on two consecutive days and then a night in Porto for a day trip to the attractions in the Porto combination. The segment impacted to this strategy are the most loyal ones, their next reviews will contribute to propagating this combo as a new and great opportunity.



RECOMMENDATIONS AND INSIGHTS

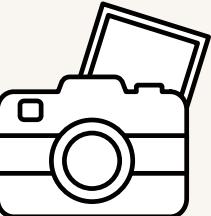


 Create discounts for tourists who go to both Sintra and Belem attractions to get an aggressive discount if they give a review in attractions on the website and have to use it until the next year. The users impacted by this promotion would be the ones we need to retain.



RECOMMENDATIONS AND INSIGHTS



 Create a package for the 4 attractions in Belém and Sintra at a discounted rate. This package would be more cost-effective, than going to the attractions individually. With this one we will work with churned segment, with a great opportunity with limited time to purchase. Later, we will send communications asking for an evaluation of the experience, so they can get more opportunities like this for the next experiences.



Thank you!

