
Destination Analysis and Targeted Marketing for Egyptian Tourism

Ramy Anwar El-Adl ramy.a.eladl@gmail.com
Mohamed Khaled m.salem250@gmail.com
Amr Essam amrmoustafa5@gmail.com
Omnia Gad omnia.gad92@gmail.com
Nesma Silman Nesma.silman@gmail.com

1. Introduction

Tourism is an essential aspect of Egypt's economy, and understanding the preferences and interests of tourists can help promote specific locations as an appealing tourist destination effectively. To achieve this goal, we will conduct comprehensive research and gather data from various sources. We will then use machine learning algorithms to analyze this data and identify patterns and trends that can help us understand tourists' preferences better.

By leveraging the power of machine learning, we will gain valuable insights into the key factors that motivate tourists to choose specific destinations. This information will enable us to develop effective marketing campaigns that target specific tourist segments and promote Egypt's unique attractions. We are excited to embark on this journey and look forward to working on this project. We are confident that our efforts will help position Egypt as a top tourist destination and boost the country's tourism industry.

2. Project Objectives

We have some questions in mind related to tourism industry patterns, and we have conducted this analysis to help us find answers to them. Our questions are:

- a. What are the preferred travel destinations for different nationalities? And how do their preferences change according to the season?
- b. Can we cluster our tourist places into specific groups with common features? What are the main themes of each group?
- c. For a given place, can we determine the distribution of visitors from different nationalities? And can we plan to increase the number of visitors from specific countries?

3. Expected Deliverables

- I. Comprehensive analysis of popular tourist destinations in selected countries.
- II. Detailed report on tourist preferences and what they find appealing in these destinations.
- III. Comparative analysis highlighting Egypt's unique selling points and competitive advantages.

4. Data and Analysis

We started the project by preparing a list of familiar tourist places in Egypt and Turkey, and then extended the data to add places from Thailand. This list was obtained from Chatgpt. Our initial data included each place and the category to which this place could be assigned. Below is a description of the features we extracted,

- **Location**: The city in which each place is located.
- **Country**: The country in which each place is located.
- **Historical Significance**: Relates to the importance of the location in terms of its historical events, heritage, and cultural value. It allows travelers to explore the rich history and traditions of the place.
- **Natural Beauty**: Encompasses the scenic landscapes, biodiversity, and natural wonders that the location offers. It appeals to travelers seeking aesthetically pleasing environments and outdoor experiences
- **Adventure Activities**: Includes various thrilling and exciting experiences such as hiking, rock climbing, ziplining, or any other adrenaline-pumping activities that cater to adventurous travelers.
- **Accessibility**: Refers to how easily a location can be reached or explored by different means of transportation. Good accessibility allows for smooth travel and exploration.
- **Shopping**: Describes the availability of shopping opportunities in the location, including markets, malls, and unique local products. It attracts travelers interested in purchasing souvenirs or experiencing the local retail scene.
- **Nightlife**: Represents the entertainment and social activities that take place during the evening and nighttime. It appeals to travelers looking for vibrant nightlife experiences, such as clubs, bars, or cultural events.
- **Water Sports**: Encompasses a variety of activities that take place in water bodies, like swimming, snorkeling, kayaking, or surfing. It attracts travelers who enjoy aquatic adventures.
- **Wildlife Viewing**: Focuses on the opportunities to observe and appreciate the local fauna and wildlife in their natural habitats. It appeals to nature enthusiasts and wildlife photographers.
- **Scuba Diving**: Specifically refers to the underwater diving activity that involves using self-contained underwater breathing apparatus (scuba). It allows travelers to explore the marine world and its diverse ecosystems.
- **Type**: This may refer to the categorization of the destination, such as whether it is a beach resort, a historic city, an eco-tourism destination, or an adventure sports hub. It helps travelers choose a location based on their preferred type of travel experience.

These columns contain categorical values that express the relevance of each feature to each place.

In our data collection phase, our main source of data was <https://tripadvisor.com/>. Which is considered the world's largest travel site. The site contains data about tourist destinations around the world and their reviews. We took the below two steps to collect our data,

Step 1: Using web scraping, we collected all the URLs of our places from the website.

Step 2: We used the URLs collected from step 1 to extract the reviews written about these places and their details (review date, title, star rating...etc.) as well as reviewers' data.

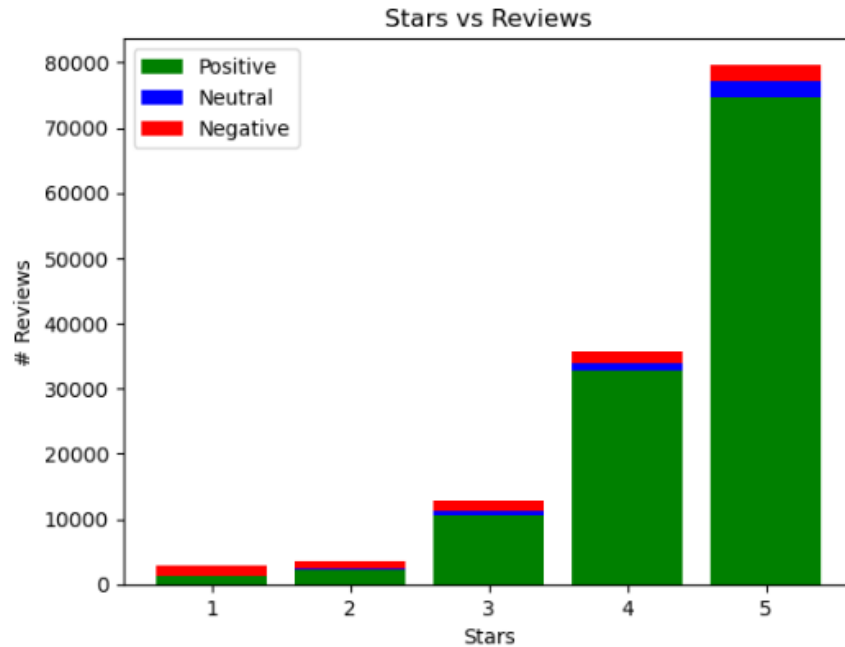
We restricted the number of collected reviews to a maximum of 2000 reviews per place to optimize the time consumption. These 2000 reviews are the most recent reviews for each place since reviews on the website are sorted from newest to oldest. We thought that retrieving such fresh data would make our analysis more relevant to the current places' reviews. However, this means we are ignoring information that was included in old reviews. For any future work, if we wanted to get the full picture of any place, we might consider using sampling to include reviews from different years.

Below are the **pre-processing** steps taken on the collected data before starting our analyses:

- We noticed that some URLs collected from scraping do not match the correct name of our place of interest, so we used Levenstein distance technique to validate the data and use accurate URLs.
- We noticed that the values in "Country" column are inconsistent. For example, some reviewers wrote their home country as "UK", some used "United Kingdom" and some used "London". So, we unified all the different names that belong to the same country by using geopy.geocoders Nominatim.
- We wanted to understand how the trend of visits changes per season. However, the site doesn't contain information about the date of visits. So, we assumed that reviews were written nearly in the same season in which the visit took place, and we mapped the reviews date to each of the four seasons and used the data in our analysis.

The analysis:

We performed sentiment analysis on reviews to identify their polarity using Vader, and we classified reviews into three categories (Positive, Neutral and Negative). We made a comparison between our classification and the stars rating given by the reviewers, and we noticed that the results of our sentiment analysis are more accurate for reviews with a greater number of stars.



It was also noticed that the total number of positive reviews is much higher than the negative ones, which might indicate that travelers don't post their negative reviews as much as they do with positive ones.

We calculated the % of positive reviews per place so that we could filter out places with a given threshold and use this information to improve tourists' experience in places with low ratings.

| | sentiment | negative | neutral | positive | total | positive% |
|-------------------------|-----------|----------|---------|----------|-------|-----------|
| Tourist place | | | | | | |
| Qasr el-Nil Bridge | | 0 | 0 | 3 | 3 | 100.00 |
| Egyptian Textile Museum | | 0 | 0 | 13 | 13 | 100.00 |
| Ras Shitan | | 0 | 0 | 1 | 1 | 100.00 |
| City Centre Maadi | | 0 | 0 | 15 | 15 | 100.00 |
| Pierre Loti Hill | | 5 | 1 | 352 | 358 | 98.32 |

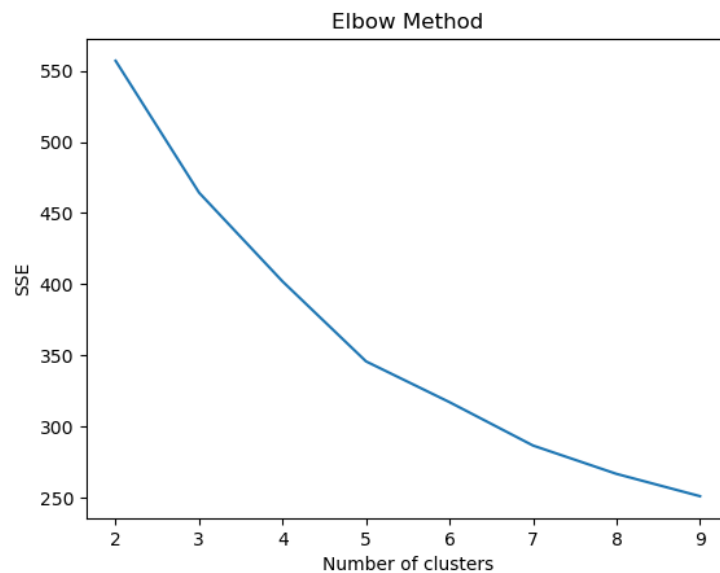
We took steps to perform topic modeling on our data using Gensim library. However, the outcome wasn't informative enough. So, we skipped this step.

Finally, we performed clustering on our places using K-means technique.

5. Methodology

We set out with the objective of grouping the list of locations by utilizing the categories gathered during the initial phase. Our goal was to generate distinct clusters, each consisting of comparable places. We

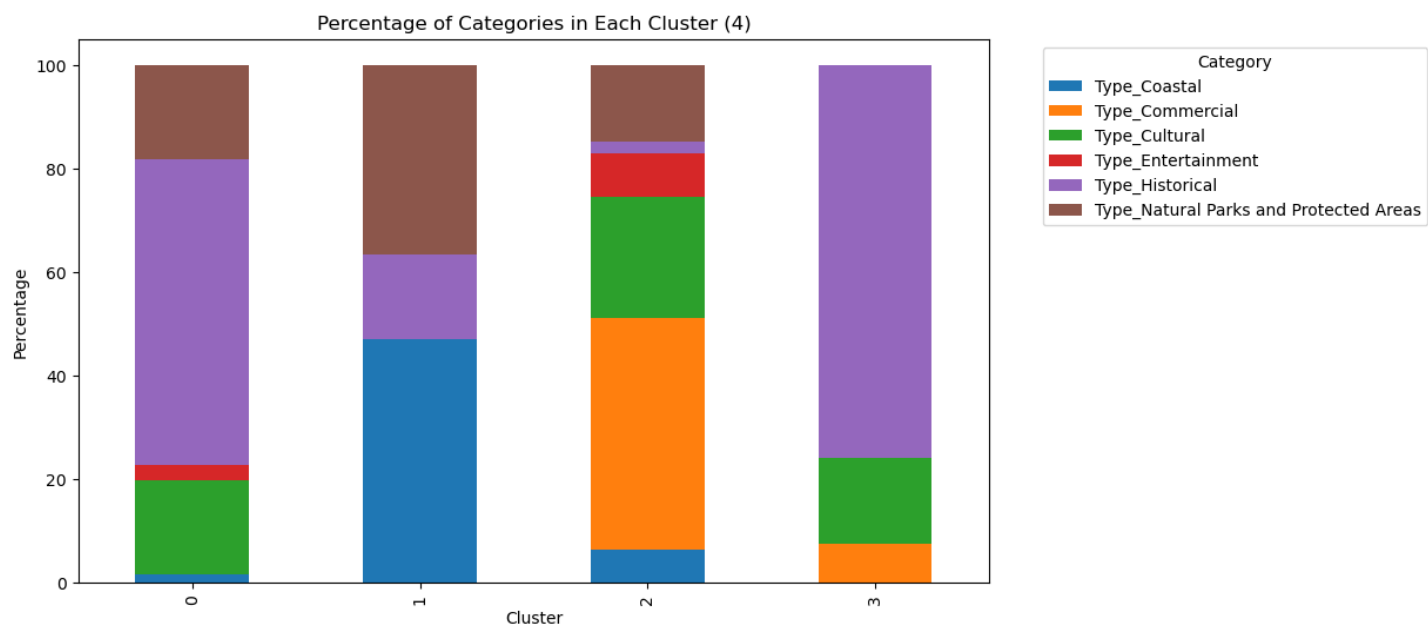
experimented with creating 7, 6, 5, and 4 clusters. And we observed the Elbow Method and silhouette below,



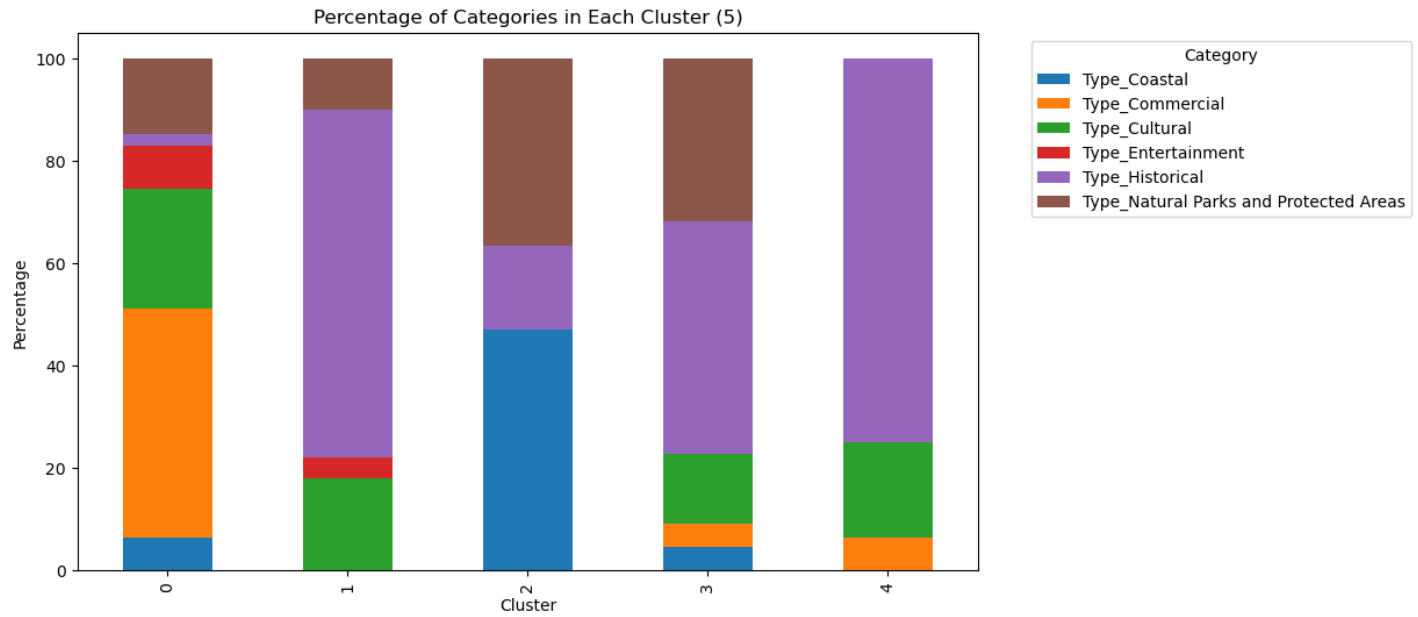
- Silhouette score for k=4: 0.27775321651688833
- Silhouette score for k=5: 0.298341387444546
- Silhouette score for k=6: 0.3213010496381913
- Silhouette score for k=7: 0.3218767204461577

Visualizing the relation between the categorizes and the number of clusters:

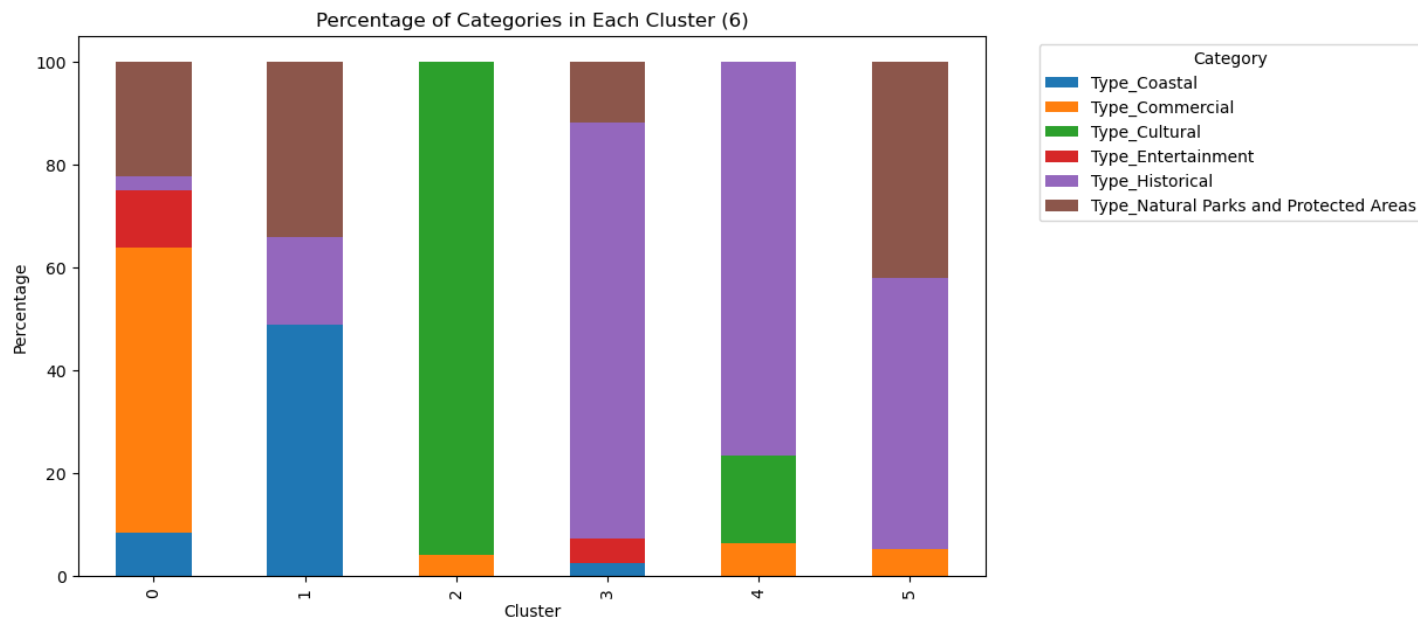
➤ **4 Clusters:**



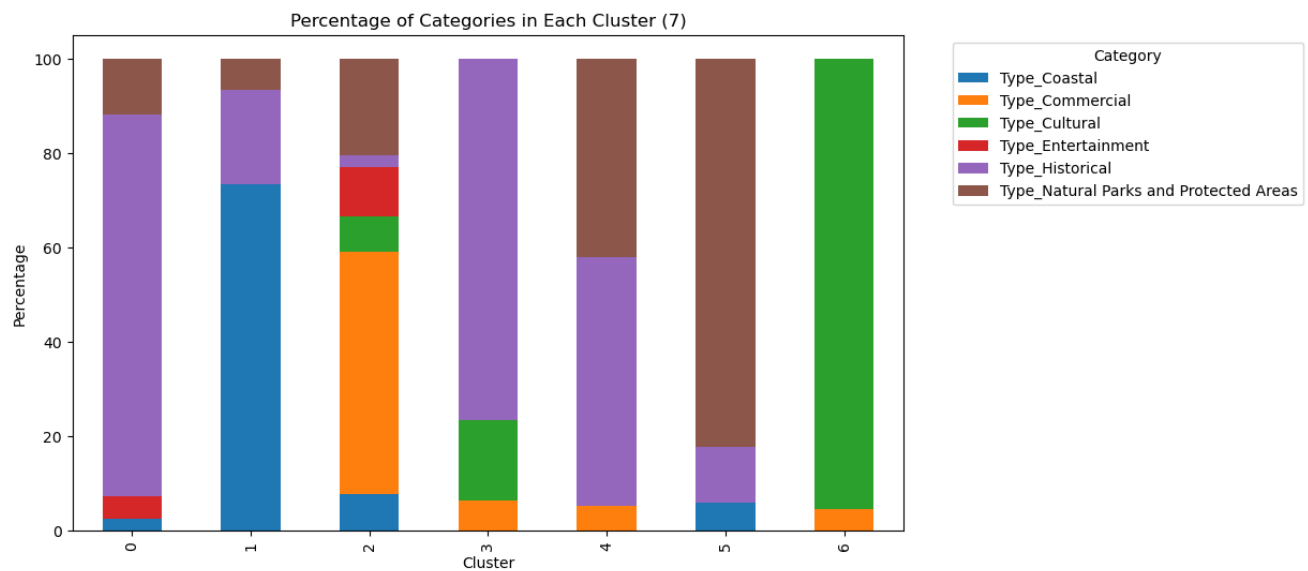
➤ **5 Clusters:**



➤ **6 Clusters:**



➤ 7 Clusters:

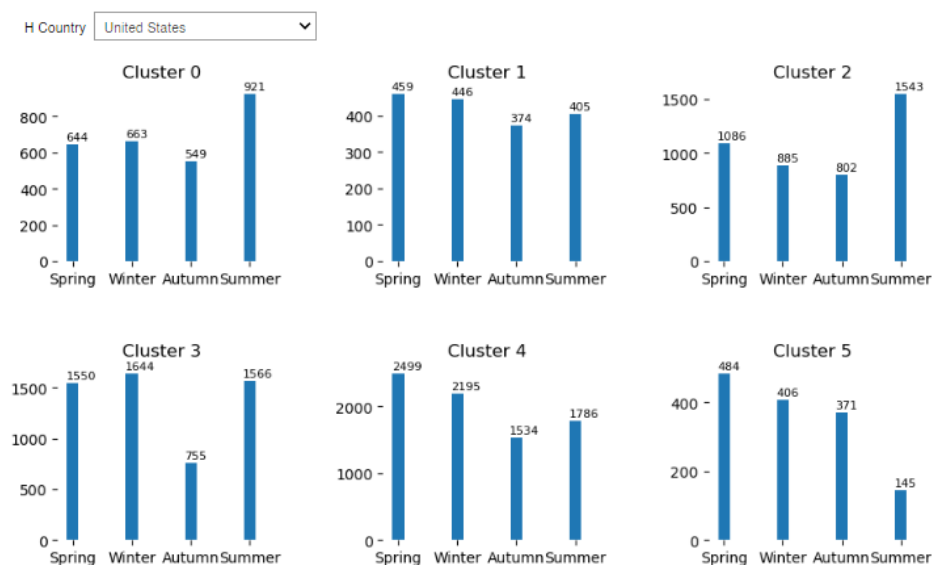


6. Results

Following the above comprehensive analysis, we noticed that categories are more dispersed for clusters 4, 5 & 7 than for cluster 6. Moreover, considering the Elbow method and silhouette outcome, we concluded that the optimal number of clusters for our study is 6-clusters.

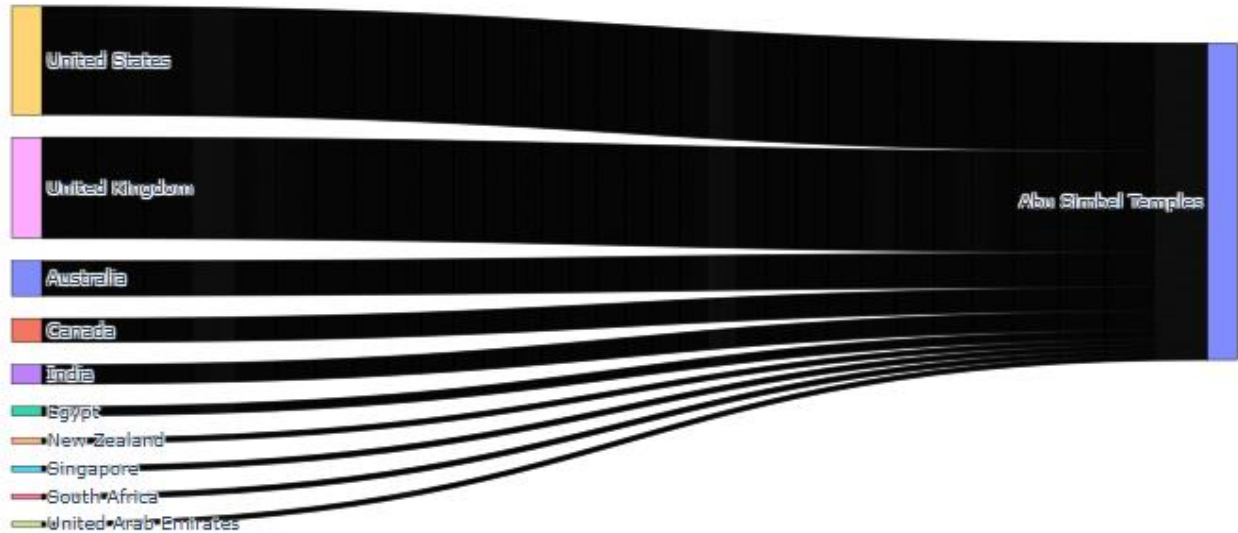
Building on this conclusion, we have created 3 interactive visualizations which can be used to answer the questions we started with.

1. The first visualization displays data for the top 20 source countries (Countries with the highest number of tourists travelling to different places). For each country, we can display the number of travelers per season. This is displayed for each of our 6 clusters separately.

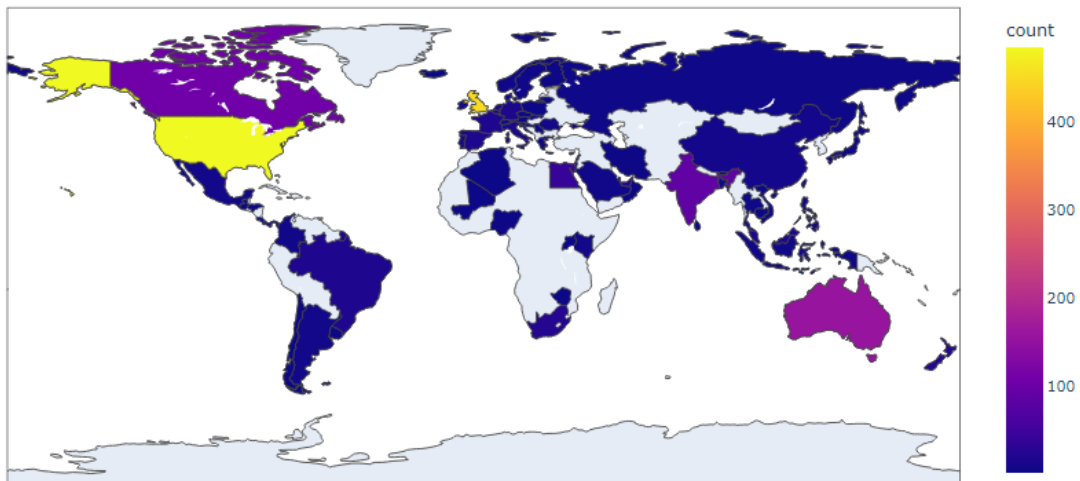


2. In the second visualization, you can choose any cluster, and the place of interest, and then the chart will display the home countries of tourists visiting this place and the number of visits per country. This can help us understand tourists' preferences.

Cluster: 3
Place: Abu Simbel Temples



3. For the last visualization, we added a map that uses color shades to represent the number of visitors from each source country. It corresponds to the options selected in chart # 2.



7. Recommendations:

We recommend using our visualizations to analyze the travelling preferences of different nationalities and use it to boost the tourism industry in Egypt.

For example, knowing that American tourists are the highest in expenditures, we can target them by following the below steps:

- a. Use chart # 1 to select the USA and find the cluster with the highest number of US tourists. The chart shows it is cluster 4.
- b. Use chart # 2 to select cluster 4 and then check the places in Egypt that falls within the same cluster but has low number of US visitors (I.e. Al-Muizz street), these are the places we need to target. We can launch marketing campaigns directed to US tourists to market these places as being a good travelling destination due to the low cost of travelling as well as having other interesting places to visit in Egypt. We can also improve these places by adding facilities that attract US tourists (i.e. special types of gifts, transportation methods, snacks and Cafés.... etc.).
- c. Use the sentiment analysis table to view the % of positive reviews for the places we want to promote, and work on improving this percentage in case it was lower than a given threshold.