As part of my final capstone project for the IBM Data Science Professional Certification, I tried to determine the best possible location, to open an upscale restaurant.

1. Introduction

For many investors, the idea of opening **a premium restaurant** is always daunting. There are a plethora of reasons why it could fail.

However, the probability of success can be leveraged by simply avoiding the mistakes executed by others. However, what are these lessons, and can we determine a preferred location to open a premium restaurant?

2. The Type of Data to be Collected

2.1 Using the data contained on Wikipedia:

I obtained a list of the most visited museums in the world, as it contained the essential data attributes to build my dataset.

The essential data elements included :

- Names of Individual Museums
- Country of Origin
- Number of Visitors per year

Using BeautifulSoup (a python package to scrape website data), I scraped the Wikipedia webpage to extract the necessary data/tags and convert the raw data into a pandas Dataframe for additional analysis.

2.2 Obtaining the location coordinates:

The next step entailed obtaining the geographic coordinates (latitude and longitude) for each museum.

Using the Google API, I created a request to obtain the geo-coordinates for each museum. For e.g., the request for the first museum, which included the latitude and longitude coordinates of the Louvre in Paris were obtained as follows:

- https://maps.googleapis.com/maps/api/geocode/json?address=Louvre +Paris.
- The API request specifies the output in JSON format (open-standard file format that uses human-readable text to transmit data)

The detail steps for the scrapping process can be obtained here.

The combination of the Museums and the geo-coordinates allowed me to develop a meaningful model.

3. Methodology

3.1 Data Cleaning and Preprocessing

Once we had the data in a tabular format, which included the latitude and longitude) for each museum, I generated valuable features using Foursquare to obtain the nearby venues surrounding each museum.

Access to the data was obtained by signing up for Foursquare and obtaining a Client ID, Client Secret Key. The signup information was appended to API Request, which included the version, venue, and search term.

The data was further analyzed to include the restaurants surrounding each museum within a specific radius. Access to the Notebook can be found here.

3.2 Cleaning Data

The cleaning of data in this notebook is completed as follows in two phases, one for the Top Museums selection and using the Google API to obtain the latitude and longitude.

The notebook for Top museum is available here.

The notebook for obtaining the latitude and longitude data is available here.

3.3 Visualization

Using Folium, I created different visualizations to obtain a better understanding of the data. The objective was to create interactive and beautiful maps using Latitude and Longitude.

3.4 Machine Learning

To gain further insights, and using the python package Machine Sci-Kit Learn, I applied the K Means Clustering Algorithm to generate unique clusters based on the restaurants surrounding each major tourist venue. These clusters were visualized using Folium.

The clustering process was informative, as it better helps our decision where to open an upscale restaurant. To gain further understanding, please refer to the notebook here.

4. Results:

- For the first cluster, which included multiple locations across the globe, there is no distinct difference between a fast-food restaurant and a premium restaurant, as they both co-existed

- However, in Cluster 5, Paris was individually distinct and identified a set of unique premium restaurants at specific locations
- Chinese preferences were more distinct, as many preferred Fast Food at specific locations, followed by Chinese food
- Ideally, Cluster 1 and Cluster 5 are favored for higher premium restaurants

5. Discussions:

- A premium restaurant at a famous museum location appears to be a good idea; however, one needs to consider what type of cuisine to serve.
- Over and above choosing the preferred location, the following points should be considered:
- Marketing is critical to any business, and success is established by having a robust product and obtaining customer feedback
- Due to high foot traffic, premium rent costs will have to be taken into consideration
- Food costs and labor costs will be different at each location

6. Conclusion:

From the analysis, not all locations are created equal. Tourist prefers different cuisines at different locations.

Quote, "Oliver added, "In the future, I might spend a bit more time getting in second and getting it right."

Opening a premium restaurant near a tourist venue, with existing restaurants appears to be a smart decision.

7. References

Access to the Python Notebook containing the data and calculations for my analysis above can be accessed here .