**CAPSTONE PROJECT REPORT**

**INTRODUCTION**

Location is one of the most influential factors that affect the revenue generated by a hotel. It is an important criterion that travelers consider when selecting a hotel and so the decision on where to site a new hotel establishment is very key. In view of this, this project aims at recommending the best neighborhoods for siting a hotel whose target market are tourists in a city. The recommendations are made based on the kinds of venues around different neighborhoods in the city and the number of such venues. Considering that this is a tourism-based hotel, a premium is placed on neighborhoods with a lot of tourist attractions such as museums, landmarks, historical sites and the likes. People looking to delve into hotel establishments for tourists would definitely find this project of great use as it provides a preliminary decision blueprint for where to site their hotels which can then be supplemented with further investigations into which of the recommended locations best suit their preferences.

**DATA USED**

The data to be used to solve the problem include:

1. A list of all the neighborhoods in the city (Los Angeles)
2. The coordinates (latitudes and longitudes) of each neighborhood
3. The venues around each neighborhood and their categories

The list of the neighborhoods was scraped from the Wikipedia page, <https://en.wikipedia.org/wiki/List_of_districts_and_neighborhoods_of_Los_Angeles>.

The coordinates of the neighborhoods were obtained using the geopy library. However, the coordinates of some neighborhoods were not returned by the library and so had to be manually sourced from the Wikipedia pages of the neighborhoods. The venues around each neighborhood and their categories were obtained by leveraging the location data provided by the Foursquare API.

**METHODOLOGY**

From the description of the problem, it is clear that a recommender system is what is required to solve it. The approach to solving the problem was divided into three distinct parts namely:

1. Data Acquisition and Cleaning
2. Exploring the Neighborhoods
3. Designing the Recommender System

**Data Acquisition and Cleaning**

First, the data from the Wikipedia page was fetched and parsed using BeautifulSoup. The data was then cleaned and stored in a list. A dataframe was then created from the list of the neighborhoods.

Next, the geopy library was used to get the coordinates of the neighborhoods. The coordinates of some neighborhoods were not found so those were obtained manually from Wikipedia. The final result was a dataframe of all the neighborhoods with their coordinates.

**Exploring the Neighborhoods**

In this section, the neighborhoods are explored by using the location data provided by the Foursquare API to get the venues around the neighborhood. The top 100 venues within a 1000-meter radius of each neighborhood were returned. The categories of the venues as well as their coordinates were also returned. The total number of venues, the number of unique venue categories, the number of venues for each neighborhood, the top 10 neighborhoods with the most venues, the most common venue categories, and the top 5 most common venues in each neighborhood were determined.

**Designing the Recommender System**

Firstly, the categories of interest were identified and then assigned weighting factors based on perceived importance as shown in the table below.

**Table 1: Categories and their Weighting Factors**

|  |  |
| --- | --- |
| **Category** | **Weighting Factor** |
| Attractions/Entertainments | 0.50 |
| Food Services | 0.15 |
| Medical Services | 0.15 |
| Transportation | 0.20 |
| Others | 0.00 |

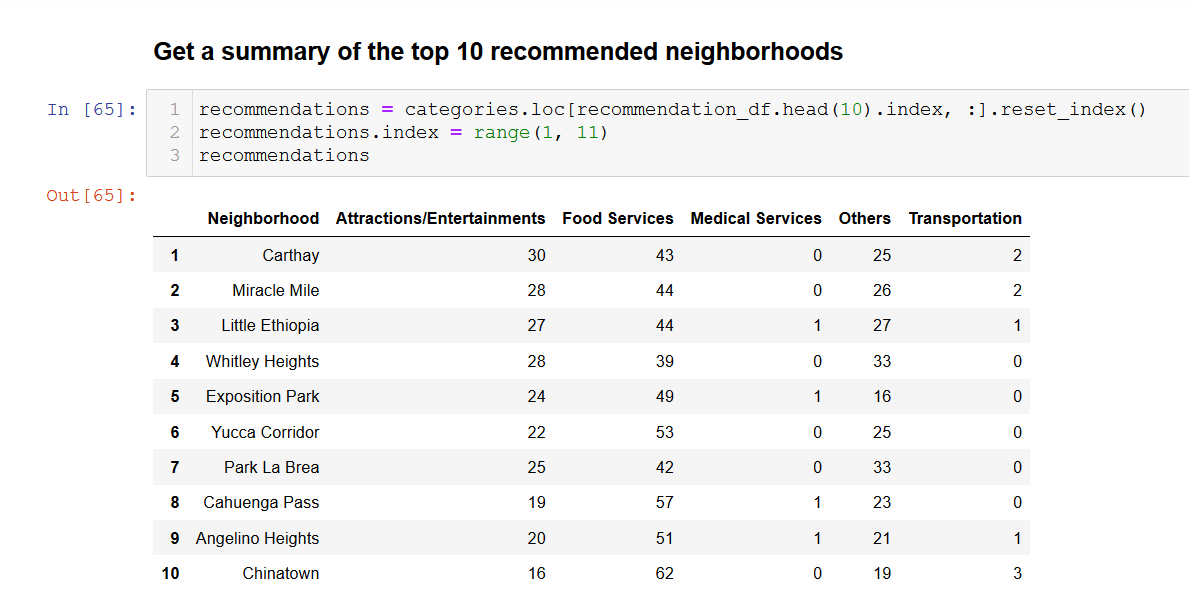
The above categories are then assigned to the various venues based on the venue categories. For instance, a venue whose venue category is Mexican Restaurant would be classified under the Food Services category.

Next, a profile for recommendation is created based on table 1. A dataframe of dummy variables for each venue based on its general category is then created. The dataframe is then grouped by neighborhood, where the final results are the neighborhoods and their total number of venues under each category.

Finally, a recommendation table is created by taking the weighted average of each neighborhood based on the defined profile. Recommendations are then made by checking the neighborhoods with the highest weighted averages.

**RESULTS AND DISCUSSION**

The final results of the project is shown below.



It is seen that Carthay Neighborhood is the number one recommendation for a hotel location in Los Angeles. The image above also shows the summary of the top 10 recommended neighborhoods which includes the number of venues under each category for each neighborhood.

Upon examination of the above results, it is seen that there is no one best option. The concept of trade-offs would have to be applied here. Questions such as, is having more attraction venues worth not having a hospital? Can we have less attraction venues and more transportation services? These are questions that can only be answered by the propective owner based on what they are looking for. Other factors such as zoning of the location, utilities like water, electricity, size of land, the number of hotels in the neighborhood currently can be investigated to decide on the best location from the recommended locations.

**CONCLUSION**

This project provided a blueprint for deciding on the best location for a hotel for tourists/travelers in a city. It is a data-driven, objective approach that can help simplify the decision-making process of business owners. The city of Los Angeles was used as a case study for the project and the top 5 recommended neighborhoods for establishing a hotel include Carthay, Miracle Mile, Little Ethiopia, Whitley Heights, and Exposition Park. Given the success of the project, the objective can also be changed to suit the owners’ preferences such as deciding a location to establish a business-based hotel or even a restaurant. In conclusion, the importance of Recommender Systems as a Machine Learning technique cannot be overemphasized as it can help business make decisions that can improve their profit margins and minimize losses.