Data Science Capstone Final Project

Opening a Hotel in Miami

Introduction

According to growthink.com, there are 74, 372 hotel in the United States. The market for hotels has grown at a rate of 4.7% for the past five years. Nowhere is the need for a new hotel stronger than in Miami, Florida. Miami leads the nation in hotel occupancy, rates, and revenue. As a thriving cultural, musical, vacation, and financial hub, the influx of tourists, both domestic and international, and businessmen, Miami is a destination location for a variety of travelers. There is no better location to build a new hotel than the city which has the demand and need for more rooms. The question is: where to open a new hotel in Miami?

Business Problem

The objective of this final assignment is to select the best location in Miami to open a new hotel. Using data science methodology and machine learning techniques such as clustering, this assignment seeks to answer the question: in which Miami neighborhood should a hotel group choose to open a new hotel?

Target Audience

The target audience of this assignment is hotel groups and management services seeking a business opportunity to open a hotel in Miami. This project is timely because the demand for hotels is only increasing and competition between hotels will continue to increase as well. Selecting the ideal location to open hotel will be important in securing the success or failure of the hotel.

Data

We need data from reliable sources for analysis:

- Wikipedia List of Miami Neighborhoods: **https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Miami**

-This Wikipedia page contains a list of neighborhoods in Miami. Using BeautifulSoup and Python, we will conduct web scraping and extract the list of neighborhoods from this html. Then we will use the Geocoder package to get the geo-coordinates of each neighborhood.

- Foursquare Developers Access to venue data: https://foursquare.com/

- We will make calls to the Foursquare API to obtain venue data, specifically hotel data, for the neighborhoods in Miami.

Methodology

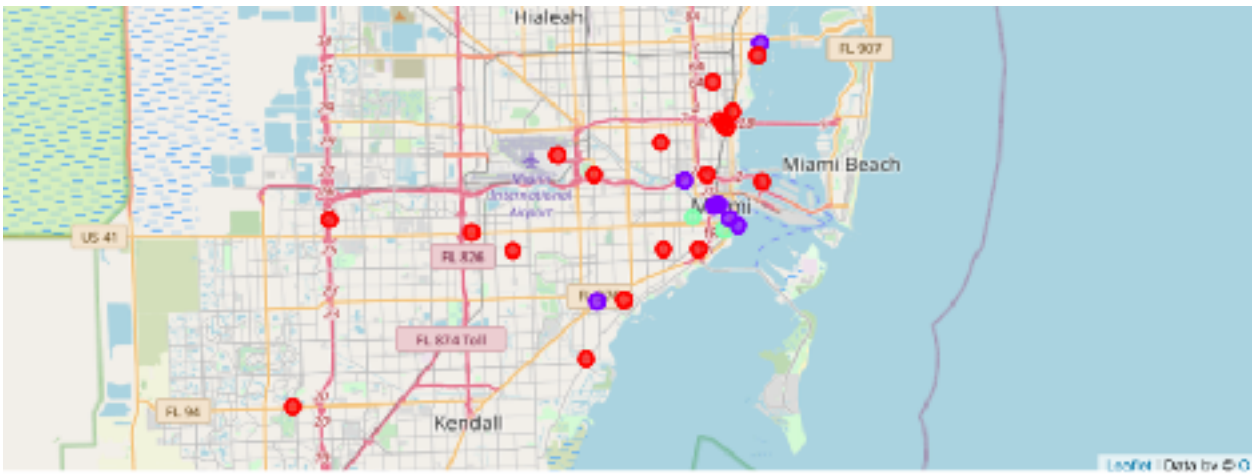
1. First, we will use the list of Miami neighborhoods provided the Wikipedia page "Category: Neighborhoods in Miami", located at url https://en.wikipedia.org/wiki/Category:Neighborhoods_in_Miami. Using BeautifulSoup and Python packages, we will scrape this webpage to extract the list of neighborhoods and put it into a data frame. Once in the dataframe, we will reindex the columns and then remove the first five rows because they were links to other wikipedia pages and not neighborhoods. Next, we will find the latitudes and longitudes of each neighborhood using the Geocoder package and merge the data frames so that neighborhoods have their corresponding lats and longs in columns. Once the data frame is complete with neighborhoods and coordinates, we will plot the neighborhoods on an interactive map with the folium package. Here, we can zoom in and out to get a closer look at the area of each neighborhood.

2. Secondly, we will make a call to the Foursquare API using our Foursquare Developer credentials to obtain a list of 100 venues. Foursquare will return the data in JSON format, so we will extract the data and put it into a new dataframe. Foursquare provided venue categories, venue name, and the lats and longs of the venue. We then count how many venues are in each neighborhood and get a list of the unique venues. We then group the rows by neighborhood and take the mean frequency of occurrence for each venue category to prepare the data for clustering. Hotels is our category of interest so we will filter hotels out and use that going forward.

3. Third, we perform k-means clustering on the data, which is an algorithm that places every data point with similar data points in a cluster for k number of

clusters. K-means clustering is an unsupervised machine learning algorithm that place like objects together in clusters. Each cluster contains similar data while the clusters differ from each other greatly. We used k=3 to create three clusters based on frequency of occurrence of hotels in those neighborhoods. Lastly, we used the folium package to plot the clusters to create the below visual.

Results



The screenshot below shows the results of the k-means clustering performed on the data:

-Red dots are low concentration of hotels. (Cluster 0)

-Purple dots are moderate concentration of hotels. (Cluster 1)

-Green dots are high concentration of hotels. (Cluster 2)

Discussion

By using the clustering technique, the breakdown in concentration of hotels in Miami was created to show three different clusters each representing different levels of hotel concentrations:

The red dots represent neighborhoods in Miami with no current hotel. These neighborhoods, such as Park West and Allapattah, are residential and are not likely to draw large tourist or business crowds.

The purple dots represent the greater downtown area of Miami. The hotels are more reasonably priced for tourists and visitors, and the hotel competition is not as stiff as the financial center. These neighborhoods provide a great business opportunity because there is less competition while the demand is still high for a hotel slightly outside downtown.

The green dots represent the neighborhoods with the highest concentration of hotels, Brickell and Riverside. Riverside and Brickell constitute the downtown are of Miami. Brickell is the Miami's financial center, so the influx of businessmen from all over the world require numerous hotels in various price ranges. Brickell also has waterfront views and a thriving art scene. The high number of hotels increases competition the area, but the high volume of visitors to this area warrant many hotels.

With less concentration of hotels and close proximity to the desired downtown area, this project recommends opening hotel in a neighborhood listed in Cluster 1, such as Government Center or Greater Downtown. There is potential for hotel group managers to place unique hotels in close proximity to the downtown attractions while having easy access to the bridges to Miami Beach. This project does not recommend placing a hotel in Cluster 1 due to stuff competition or Cluster 0 due to lack of visitors.

Conclusion

Using the data science methodology, we identified the business problem of where to open a new hotel in Miami, identified data required for analysis, cleaned the data, used machine learning to cluster the data, and created a visualization of the results of the analysis. We used a combination Wikipedia data and Foursquare data to analyze hotel presence in Miami neighborhoods and concluded that hotels built in neighborhoods in cluster 1 provide the greatest potential for success.

Data from other factors such as trends in domestic travel for business or pleasure and consumer willingness to spend money on travel should be analyzed to see what effect these data sets have on selecting the next hotel location. It is important for stakeholders to understand the demand for the new hotel and how demand can change in reference to the economy.