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.

- Miami has a thriving cultural, musical, vacation, and financial hub

- Miami sees an influx of tourists, both domestic and international, and businessmen

- 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

- 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

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

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

Methodology

- Acquire data from Wikipedia List of Miami Neighborhoods
- Use web scraping techniques in Python to extract list of neighborhoods
- Use Geocoder package to obtain geo-coordinates for each neighborhood
- Populate neighborhood data into a data frame
- Plot the neighborhoods on a map using Folium
- Call the Foursquare API to obtain venue data and see how many venues returned for each neighborhood
- Group the venues by neighborhood and take the mean frequency of occurrence for each type of venue, focusing on hotels
- Cluster the data using K-Means clustering to see which area has higher or lower number of hotels

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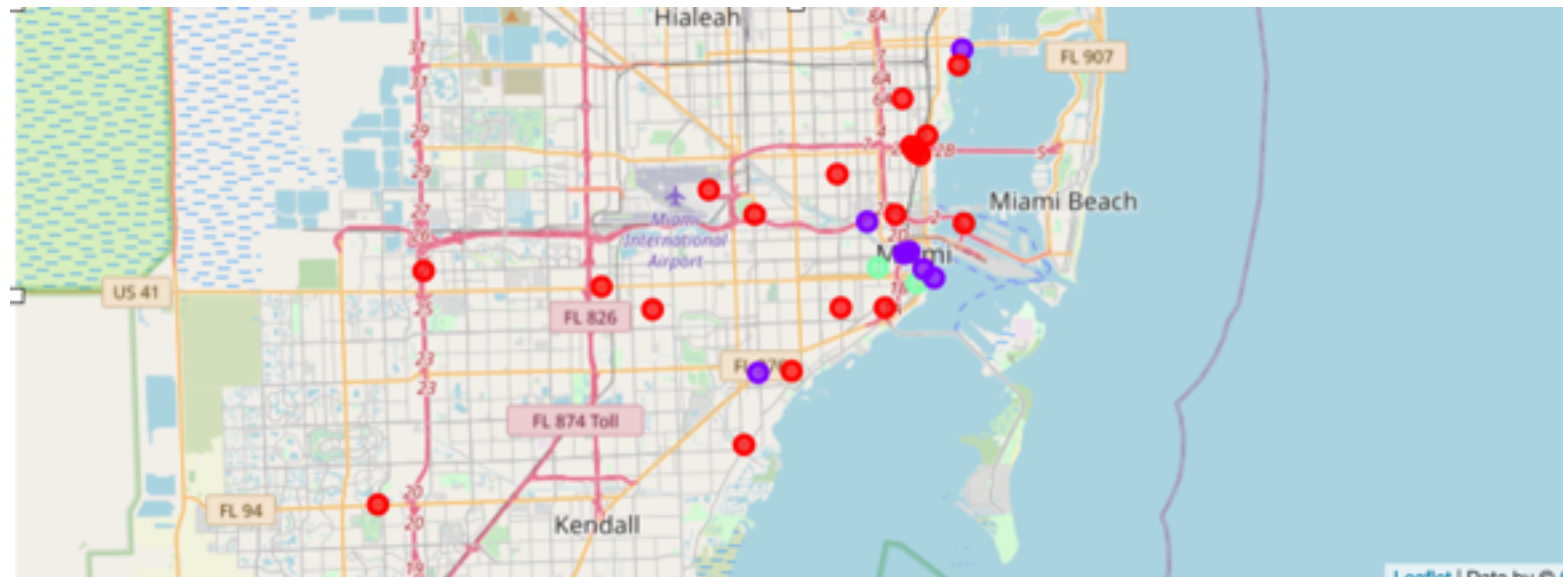
- Red dots are low concentration of hotels. (Cluster 0)

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

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

Discussion

- The red dots represent neighborhoods in Miami with no current hotels are residential and are not likely to draw large tourist or business crowds.

- The purple dots represent the greater downtown area of Miami. 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. The high number of hotels increases competition in 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, the purple cluster, such as Government Center or Greater Downtown.

Conclusion

- We used a combination of 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.