IBM DATA SCIENCE CAPSTONE:

WHERE TO LIVE FOR A YOUNG PROFESSIONAL IN CHICAGO, IL

By: Shelley Leung

June 2020

BUSINESS PROBLEM

- Retaining and attracting young professionals and tourist is an important part of being able to develop the Windy City's economy and allow it to sustain growth.
- Objective: Where a person stays is an important part of the overall experience of the city.
- In this research project, we will target the new city transplants looking to settle into life in a new city.
- Business question: Where should a new transplant settle in the city that will have a diverse and easy access to amenities and entertainment?

DATA SOURCES

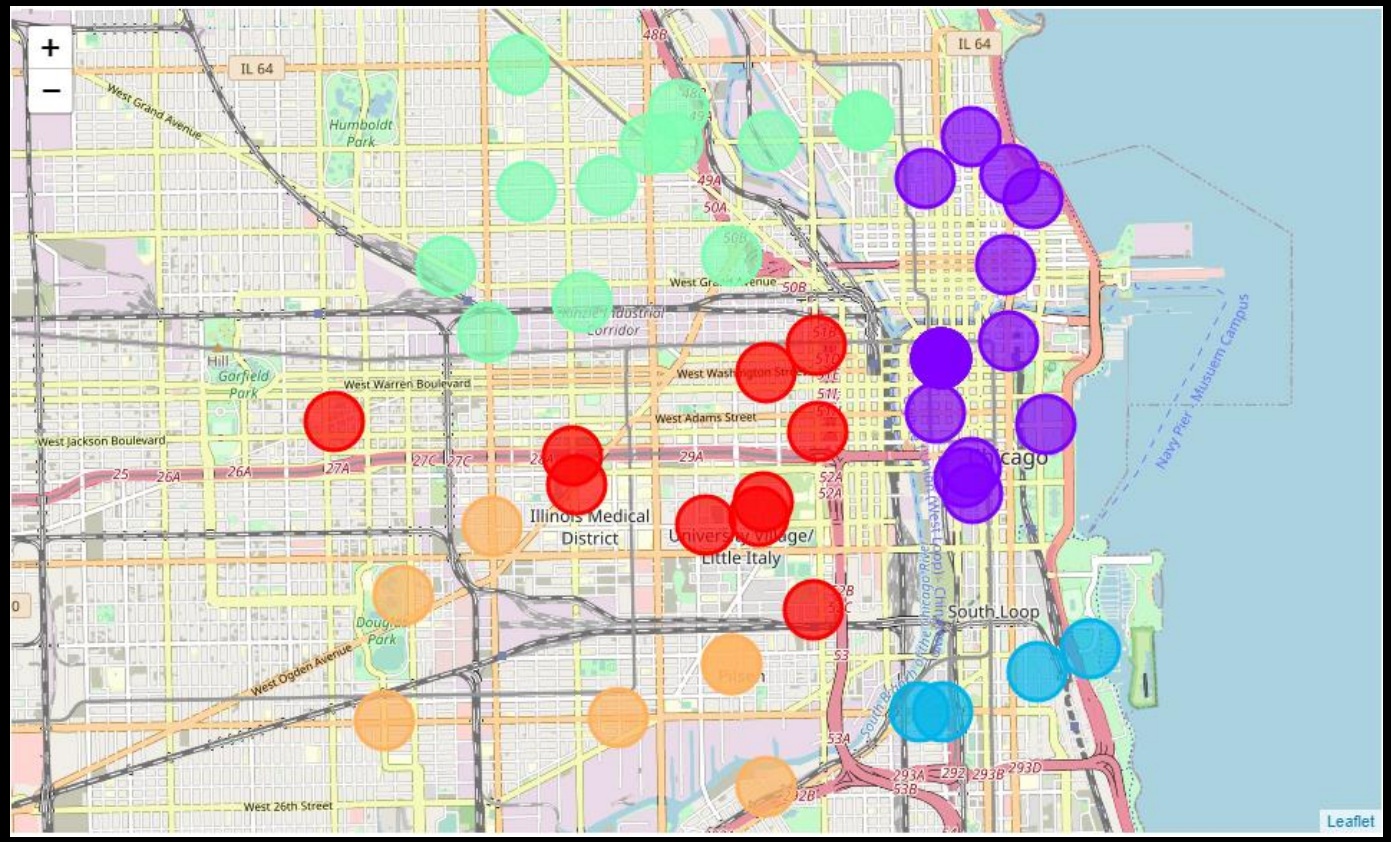
- Data required:
  - List of neighborhoods in the Greater Chicago area
  - Latitude and Longitude coordinates of the neighborhoods and venues in the area
  - Venue data for each neighborhood
- Data sources:
  - Wikipedia list of neighborhoods: (https://en.wikipedia.org/wiki/List_of_neighborhoods_in_Chicago)
  - Geocoder package for coordinates
  - Foursquare API for venue data

METHODOLOGY

- Build a dataframe of neighborhoods in Chicago, IL by web scraping data from a Wikipedia page.
- Get the geographical coordinates of the different neighborhoods.
- Explore and cluster the neighborhoods (k-Means clustering) to determine which neighborhoods are within city center.
- Obtain the venue data for the cluster that contain city center neighborhoods from Foursquare API.
- Determine the top 10 venues within each neighborhood and perform k-means clustering for each.
- Create a visual for the clusters using folium maps.
- Select the best neighborhood in each cluster that contains the most venues.

RESULTS

- The final results created 5 different clusters:
- ***Cluster 0 (Red):*** Contains the most Italian, Mexican, fast food, and New American restaurants. There is also an abundant number of coffee shops located in this cluster.
- ***Cluster 1 (Purple):*** Contains the most hotels and theaters.
- ***Cluster 2 (Aqua Blue):*** Contains the most Chinese restaurants and parks.
- ***Cluster 3 (Light Green):*** Contains the most diverse number of venues that range from bars, restaurants, gyms, coffee shops, grocery stores and more.
- ***Cluster 4 (Orange):*** Contains the most Mexican restaurants, Sandwich places and bars.



CONCLUSION

Observations & Recommendations

- Clusters 0,2,4 seem homogenous with the types of venues they have in the area.
- Cluster 1 contains the most hotels and theaters which signify a tourist/business area.
- Cluster 3 is ideal due to the different choices of venues and amenities located in the area.

Limitations & Suggestions for further research

- Foursquare limited venue data to 100 per neighborhood.
- Crime rates of the neighborhoods was not included.
- Affordability of the neighborhoods was not included.
- Census statistics for the age groups residing in the neighborhoods were not included.

THANK YOU