Coursera Capstone
IBM Applied Data Science Capstone

Exploring the opportunity for opening a new recreational center in Calgary

Praise Okwa, May 2020

Business Problem

Why bother?

- Recreational centres provide an opportunity for active living and recreation in a safe, inclusive environment
- By creating a positive atmosphere, these facilities become essential to personal health and wellness, thereby reducing reliance on healthcare and other costly social services.
- Current facilities face overcrowding as most of them are confined to central locations with the city and the outer neighbours are at a disadvantage

Data

Required data

- A list of all neighbourhoods within Calgary including residential communities, industrial areas, major parks and residual areas by electoral ward
- Datasets that show names and addresses for current recreation facilities, including amenities available at each location
- Longitudinal and Latitudinal data for location of neighbourhoods as well as current recreational facilities, as this will aid in collection more data relating to those locations as well as allow for plotting a map

Data sources

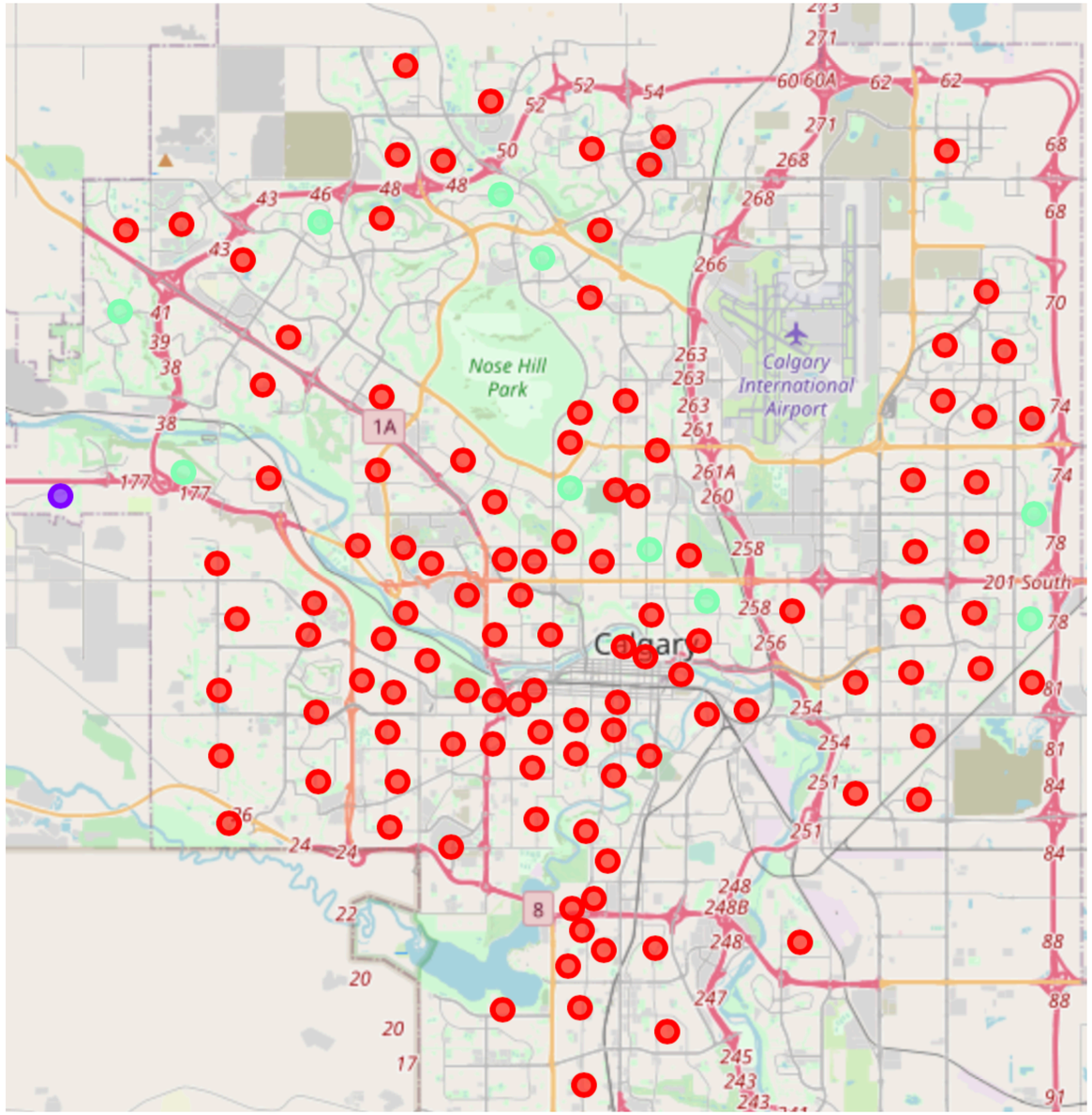
- Wikipedia provides list of all neighbourhoods within Calgary along with information regarding type, population, dwelling and % change in population for each neighbourhood. Datasets from the Calgary data page will also be used for information regarding current recreation centres.
- A Geocoder Package will be used to get coordinate information and LXML will be used to scrape the webpages
- The Foursquare API will be used for venue data

Methodology

- Web scraping of Wikipedia and Data Calgary for lists of neighbourhoods and recreational facilities
- Get Longitude and Latitude coordinates fusing Geocoder
- Use Foursquare API to get the venue data
- Group data by neighbourhood and then take the mean of the frequency of occurrence of each venue category
- Filter venue categories to recreational facilities
- Preform clustering using k-means
- Visualize clusters using Folium
- Perform additional exploratory data analysis

Results

- The neighbourhoods are placed in 3 clusters
    - Cluster 0 - Neighbourhoods with the highest concentration of recreation facilities
    - Cluster 1 - Neigbourhood with the lowest number of recreation facilities, right on the outside of the city
    - Cluster 2 - Neighbourhoods with low to no presence of recreational facilities



Discussion

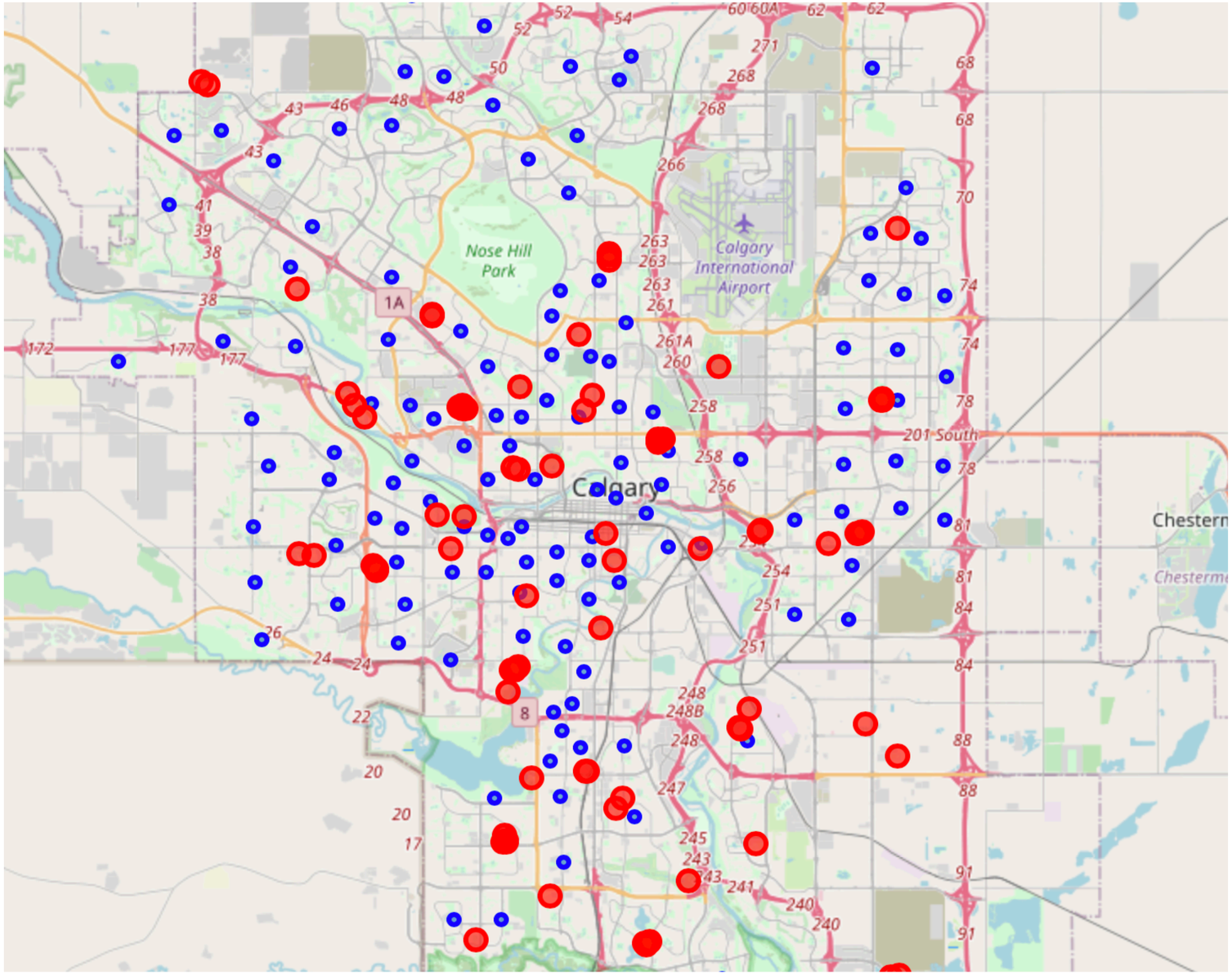
- Most of the recreation centres are concentrated in the central area of Calgary, with the highest number in cluster 0 .
- Cluster 2 offers the most promise for opening a new recreation centre
- Cluster 1 can be ignored as it represents neighbourhood right on the outskirts of the city.

Recommendation

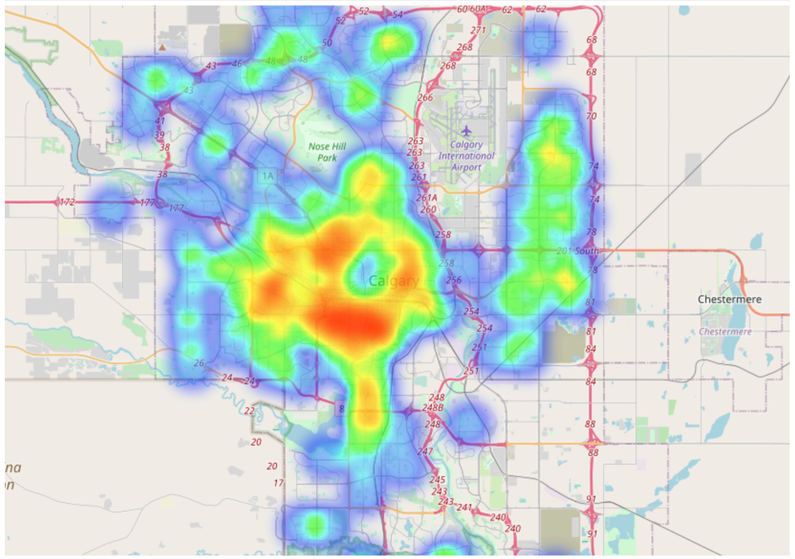
The neighbourhoods in cluster 2 offer the optimal location for a new recreational centre, specifically in the NW Quadrant

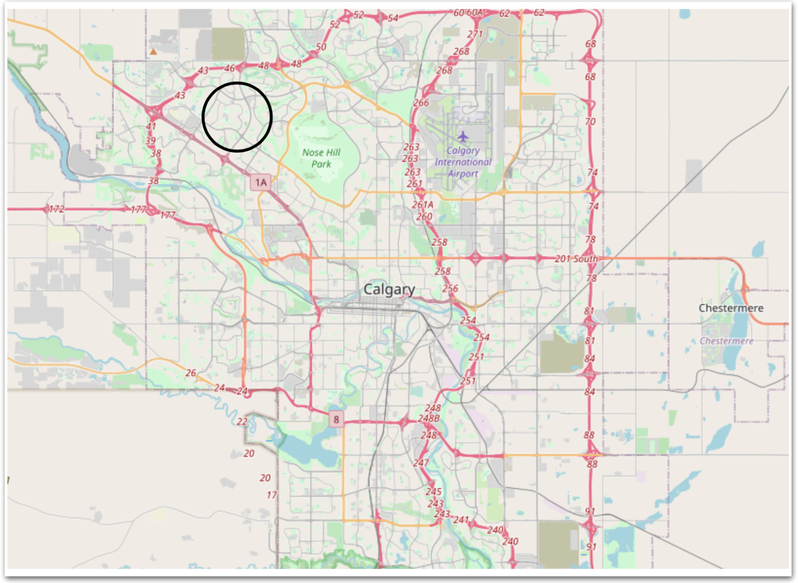
Map plots shown on next page

Plot of neighbourhoods (blue) against recreational centres (red)



Heat map showing concentration of recreational facilities





Proposed location for new centre

Conclusion

- The solution to the business problem of where a new recreational centre should be built is in the NW Quadrant of the city of Calgary
- These findings will assist stakeholders in making a decision in order to capitalize on opportunities in the new location while avoiding overcrowded central areas in the city.

Thanks!