Identifying Suitable Neighborhoods for Employee

To assist International Assignment Department

Chiew Seng Chun

May 27th, 2020

Introduction

- Big corporations need to relocate employee
- Tricky and costly if not managed correctly
- Can use past experience – sound approach *but*
- Past experience may not suit all
- **How can the department provide more useful recommendations of where to live?**
- Audience – International assignment department and employee

Scenario

- For the purpose of this assignment, the following scenario was created:

"Imagine John Doe, a Microsoft employee who has worked in London, UK for a long time, has recently been reassigned to San Francisco, California in the United States. Due to the reassignment, John Doe and his family will be relocating to San Francisco. John Doe has a specific requirement to stay in an area in San Francisco that is similar to where their family is staying now in London. John Doe currently stays near the Richmond Underground tube station and the Microsoft Headquarters in San Francisco is at 555 California St 200, San Francisco, CA 94104, United States. John Doe would also like to stay somewhere close to the new office. The international assignment department now has to provide a list of recommended neighborhoods in San Francisco for John Doe."

Data

- Data Sources
  - Neighborhood data for San Francisco - https://data.sfgov.org/Geographic-Locations-and-Boundaries/SF-Find-Neighborhoods/pty2-tcw4.
  - Foursquare API Venues data
  - Geodata
- Data Cleaning
  - Geojson of neighborhood data parsed and centroid computed using shapely library
  - Json results from Foursquare API parsed to obtain venue name, category and geodata
  - Geopy used to obtain geodata for Richmond station and Microsoft Headquarters in San Francisco

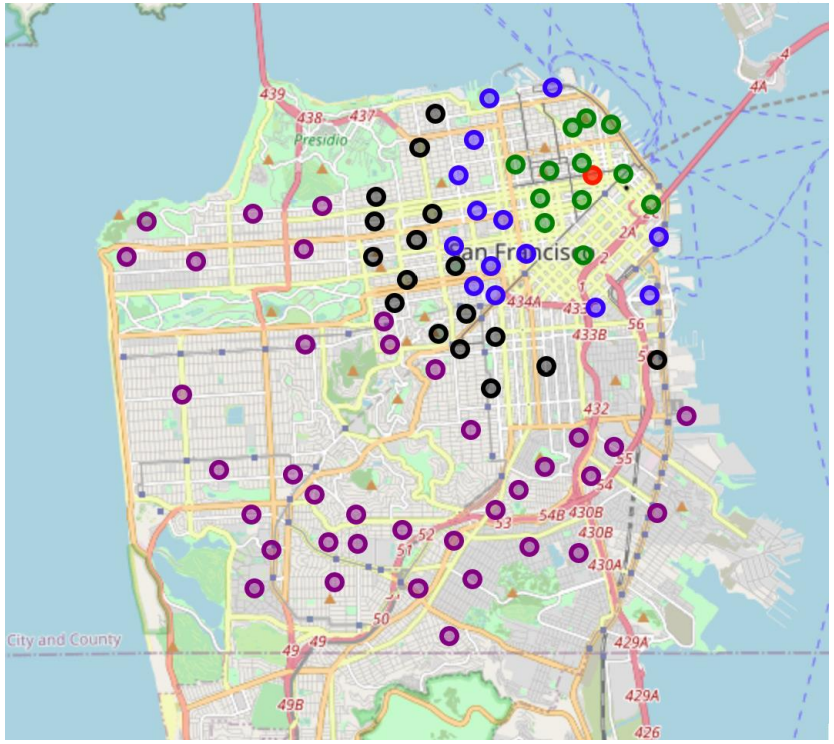
Methodology

- Obtain location of Richmond Underground Station and explore the area within 500m of the station
- Obtain location of neighborhoods for San Francisco.
- Explore each neighborhoods and add the explored data from Richmond.
- Cluster the neighborhoods and highlight which San Francisco cluster is similar to Richmond.
- Within that cluster, sort neighborhoods by distance to Microsoft Headquarters.

Analysis

- Richmond returned 100 venues with 53 unique venue categories (top 3 are pubs, coffee shops and Italian restaurants)
- 117 San Francisco neighborhoods were found and using Foursquare API, 4632 venues were found with 368 unique venue categories. (top 3 categories are coffee shops, cafes and parks)
- K-means clustering is used and 8 clusters were used
- 80 San Francisco neighborhoods are in the same cluster as Richmond
- The neighborhoods are further segmented by distance from Microsoft Headquarters

Analysis (cont'd)



Map showing neighborhoods similar to Richmond colour-coded based on distance from Microsoft Headquarters

Results and Discussion

- Not all San Francisco neighborhoods have Richmond's top ten venue categories in their top ten.
- Perhaps due venue categories that exist only in San Francisco neighborhoods.
- 12 San Francisco neighborhoods are within 1 mile of Microsoft Headquarters
- 14 San Francisco neighborhoods are within 1-2 miles of Microsoft Headquarters
- 17 San Francisco neighborhoods are within 2-3 miles of Microsoft Headquarters
- 37 San Francisco neighborhoods are more than 3 miles of Microsoft Headquarters
- Recommendation can now be provided to John Doe
- Further refinement can be made if John has further requirements such as preference for parks

Conclusion

- Possible to use Foursquare data to find neighborhoodsin other areas that is similar to a neighborhood in another area.
- Further parameters such as availability of properties (rent/sale), nearby school with specific ranking, crime rate can be added to further refine the recommendations.
- Reduction of dimension of category venues may be required to ensure better clustering or use of other clustering algorithms.