# Open a small coffee shop in Manchester

Eric Liu

# Executive Summary

- Methodologies
  - Web scraping was performed to extract data from various sources as well as retrieval of geographical coordinates
  - A series of algorithms & methodologies are tested to apply to the model, i.e. K-Means (K-5), One-Hot Encoding

- A suggestion for location of a coffee shop can be found through the model

# Introduction

- During COVID-19 lockdown, social-distancing measures were implemented, all shops were closed, and business activities almost stopped.
- Vaccination rate is reaching high recently, business activities restart and rent rate is relatively low.
- Business opportunities can be seen right now and in coming months
- Opening a small coffee shop can be a good business in UK.
- Study any neighbourhood in one of major cities in United Kingdom - Manchester

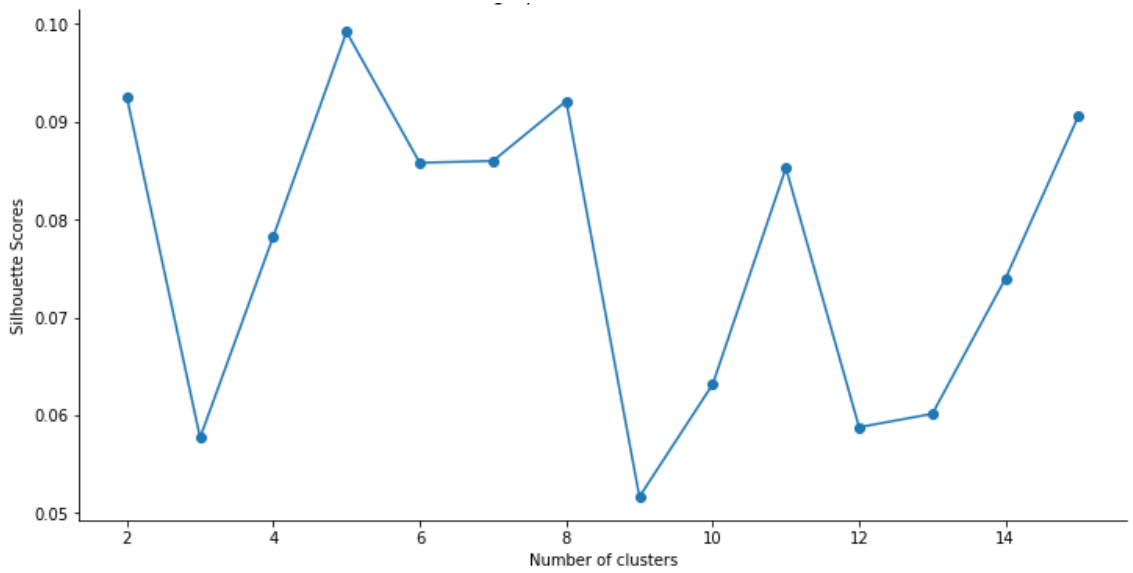
# Data Collection & Data Wrangling Methodology

- Data Requirement:
  - Explore and search via Internet through data scraping method
  - A dataset of neighborhoods / districts likely to start a small business in Manchester city and streets

- Data Wrangling
  - Clean, re-structure and enrich raw data collected from Internet and dataset from tools
    - geographical data like coordinates of neighborhoods by GeoPy and
    - Map data by FourSquare location platform
    - Calculate distance between cafes in neighborhoods

# EDA and interactive visual analytics & Predictive analysis methodology

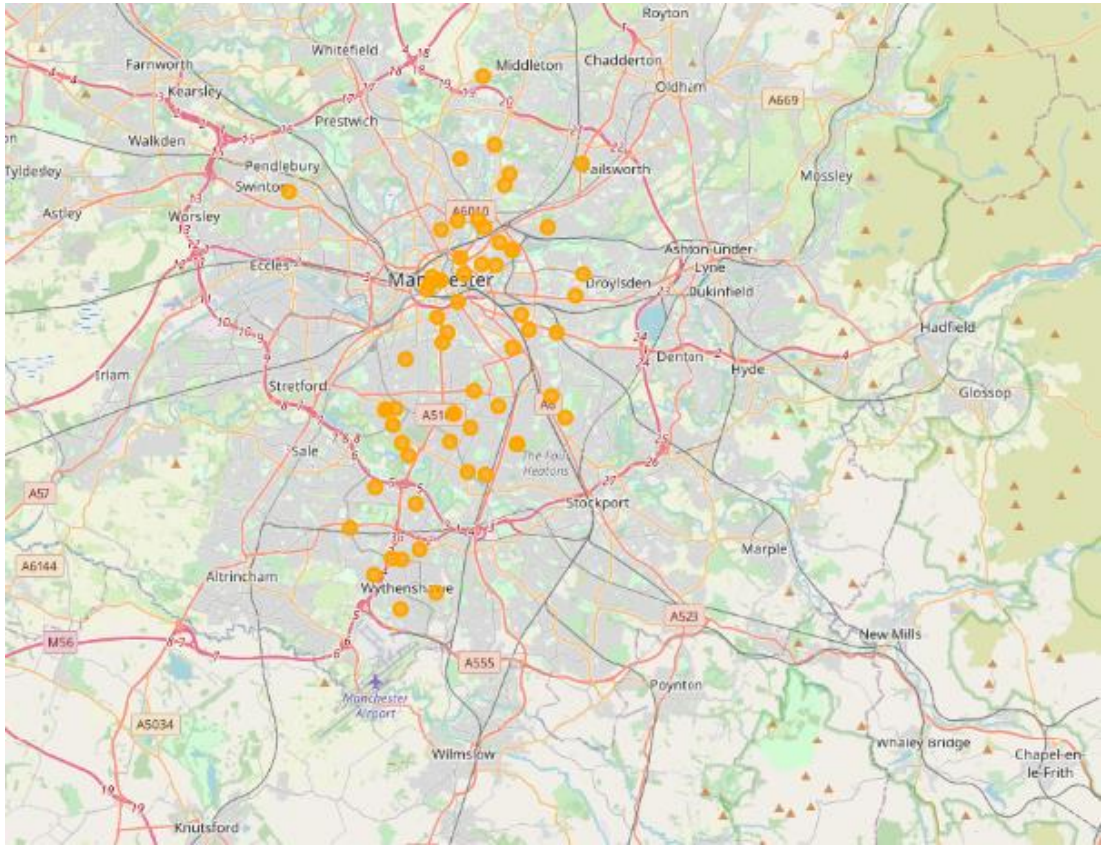
- Get connected with Foursquare API
- Work with Geopy.geocodes to import Nominatim to get coordinates latitudes and longitudes of cafes in Manchester
- Plot map by Folium with labels and colors
- One-hot encoding - all the venues are grouped by the neighborhoods and the category belongs to the venue
- Clustering by K-Means methodology, unsupervised learning technique was used to cluster the neighborhoods based on the category of venues near the neighborhoods.



# EDA with data visualization

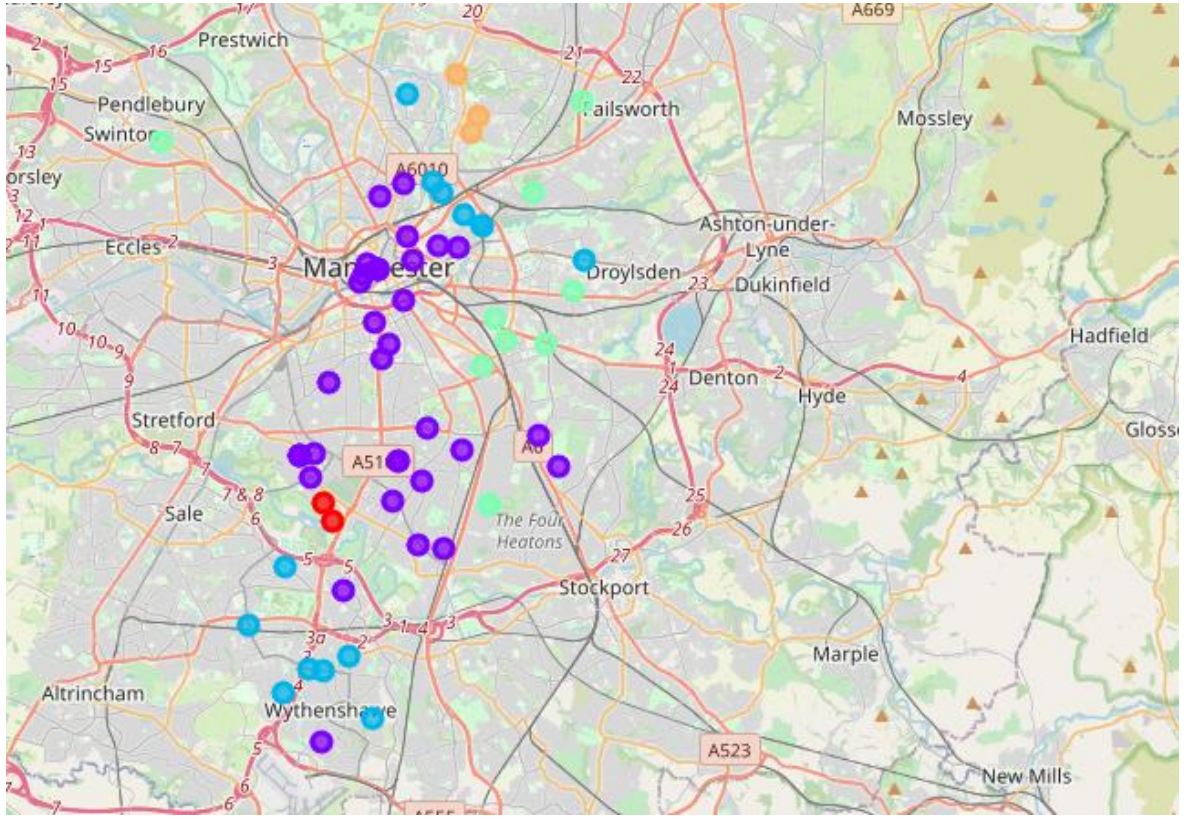
- Using Folium for Python draw on the map with similar cafes in clusters in neighborhoods. The chart helps to visualize the distribution of cafes and differentiate which cluster is preferred to open a coffee shop with less competition.

# Build an interactive map with Folium




- Using folium, a map was plotted to show how the different neighborhoods are distributed in Manchester City

# Build a Dashboard with Plotly Dash




Each colour represent its cluster group for easy classification

# Predictive analysis (Clustering)

- One-hot encoding was performed to analyze and to narrow down the most common venues in each of the districts. Given all the venues surrounding them, subdistricts were clustered using *K*-means algorithm.
- One important aspect of the k-means model is to determine the number of clusters to use in model development. This was determined by the Silhouette score which was calculated for a range of clusters from 2 to 15. Given that there is a peak at $k = 5$, the *K*-Means clustering was proceeded with that value

- The Python libraries and dependencies are used as follows: pandas, NumPy, BeautifulSoup, GeoPy.geocoder, Nominatim, JSON, Folium, Matplotlib, etc

# Results

- The results show that districts in Ladybarn and Barlow Moor are the best options for opening a new coffee shop with less competition by K-means algorithm.

| Cluster | Neighbourhood | Common Venue 1 | Common Venue 2 | Common Venue 3 | Common Venue 4 | Common Venue 5 | Common Venue 6 | Common Venue 7 | Common Venue 8 | Common Venue 9 | Common Venue 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Barlow Moor | Grocery Store | Bus Station | Indian Restaurant | Hotel | Lake | Forest | Tram Station | Fast Food Restaurant | Outdoor Supply Store | Gym |
| 2 | Belle Vue, Manchester | Convenience Store | Supermarket | Gym | Electronics Store | Racetrack | Restaurant | Pharmacy | Fast Food Restaurant | Climbing Gym | Park |
| 14 | Chorlton Park (ward) | Indian Restaurant | Park | Grocery Store | Pizza Place | Japanese Restaurant | Gym | Music Venue | Spanish Restaurant | Café | Coffee Sho |
| 29 | Harpurhey | Hotel | Flea Market | Supermarket | Gym / Fitness Center | Tram Station | Shopping Mall | Park | Pastry Shop | Museum | Music Venue |
| 32 | Ladybarn | Hotel | Pub | Bar | Fast Food Restaurant | Asian Restaurant | Park | Grocery Store | Racetrack | Gym / Fitness Center | Sandwich Place |
| 33 | Levenshulme | Indian Restaurant | Grocery Store | Café | Antique Shop | Gym / Fitness Center | Fast Food Restaurant | Farmers Market | Sandwich Place | Pub | Track Stadium |

# Conclusion

- Principally investors search for opportunities to open a business in where competitions are not strong.
- All of the districts were explored and then clustered based on the similarity of their surrounding venues using *K*-Means clustering algorithm.
- However, factor of competition of similar business may be a positive, i.e. it creates a special street or community like pub zone which customers will immediately think about that spot if want to enjoy happy hour with friends.
- This topic is out of scope in this assignment