Location Search for New Coffee Shops

APPLIED DATA SCIENCE CAPSTONE PROJECT

Introduction

International Coffee Chain is planning to enter UK market

Numerous challenges and risks are anticipated

- Strong competition
- High costs

Objectives

- Team responsible for the expansion would needs insights regarding the following:
  - Are there any London locations where a new coffee shop can fit?
  - If the answer to the above is yes, what are the characteristics of these areas?
  - In terms of competitors what are the price range and ratings for existing coffee shops?
- Task assigned to a data analyst who exploited various data sources to provide answers

Data

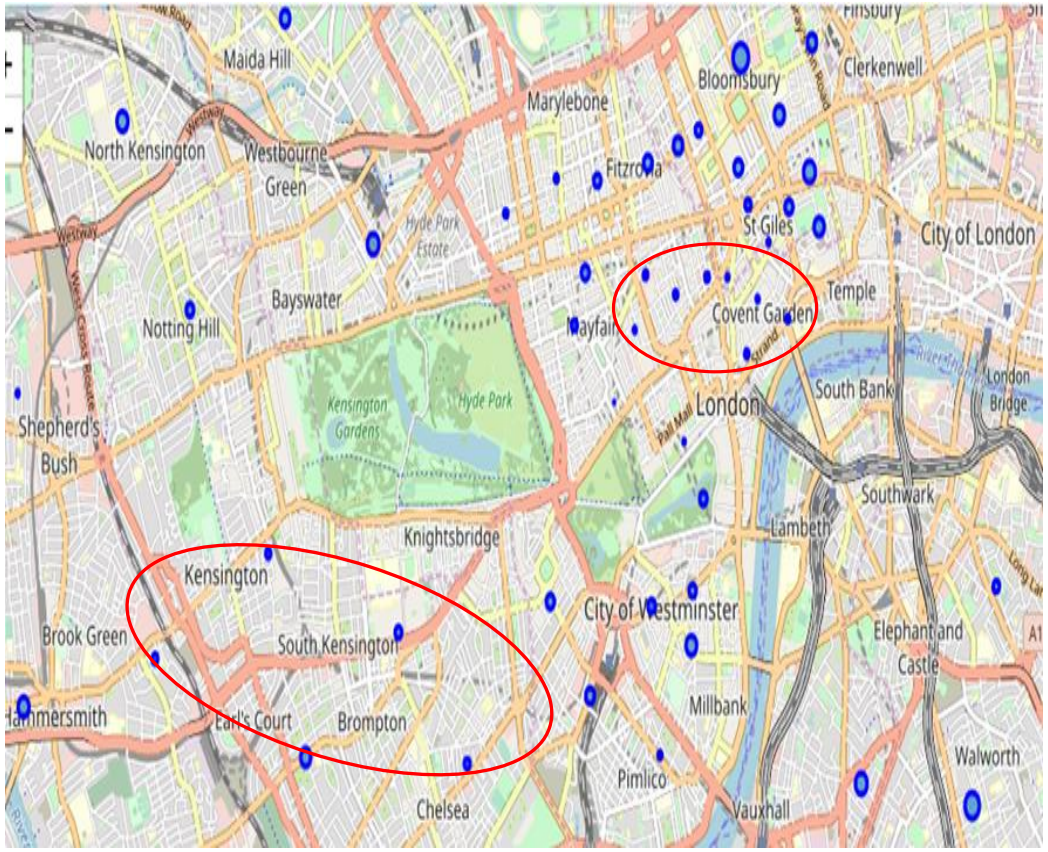
- Primary data source Foursquare API
  - Used mainly two endpoints: Explore and Details
- Geographical information scraped from
  - Wikipedia – London districts and postcode (out codes)
  - Free Map Tools – Coordinates
- Used pandas for extracting, transforming and loading data

Methodology

- Exploratory data analysis
  - Descriptive statistics using pandas
- Geospatial visual analytics
  - Maps and bubble maps using folium
- Segment districts
  - Apply clustering with k-means
  - Used elbow technique to find the optimum k

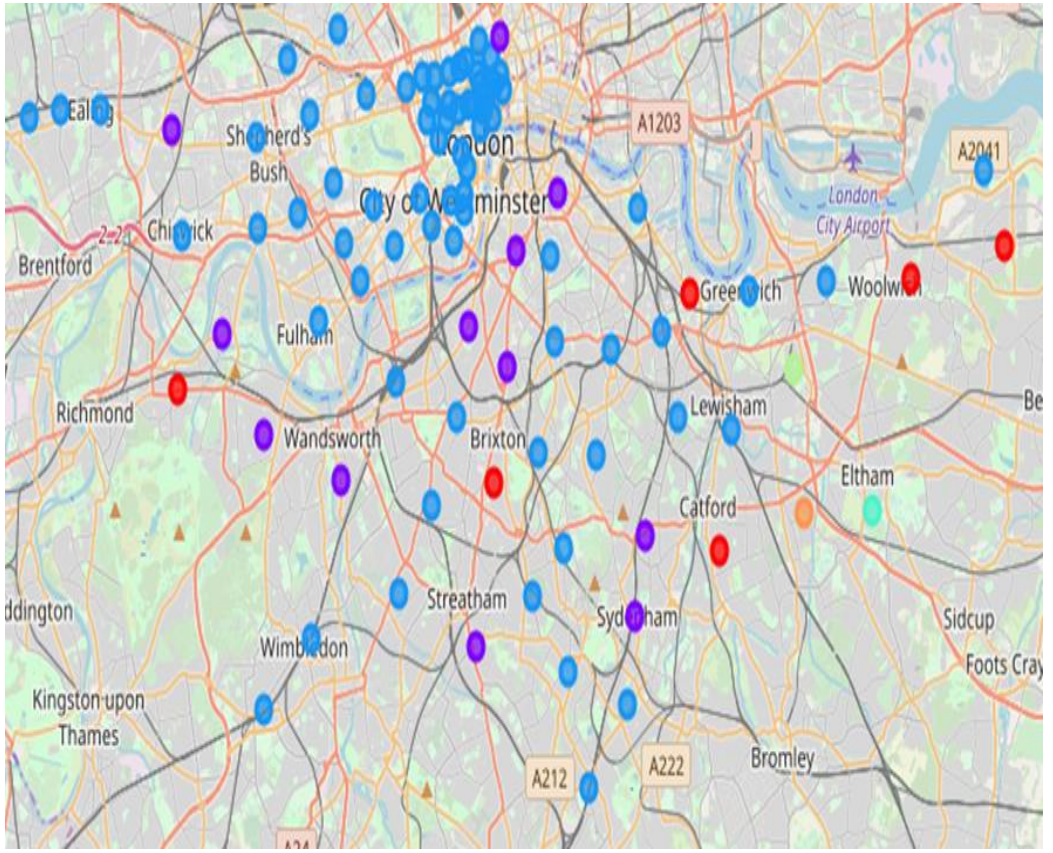
Candidate Locations

- Areas with low coffee shops frequency
  - Soho
  - Covent Garden
  - Kensington
  - South Kensington
  - Chelsea
- Potential candidates for new coffee shops
- Part of the biggest cluster (k-means)



Districts Segmentation

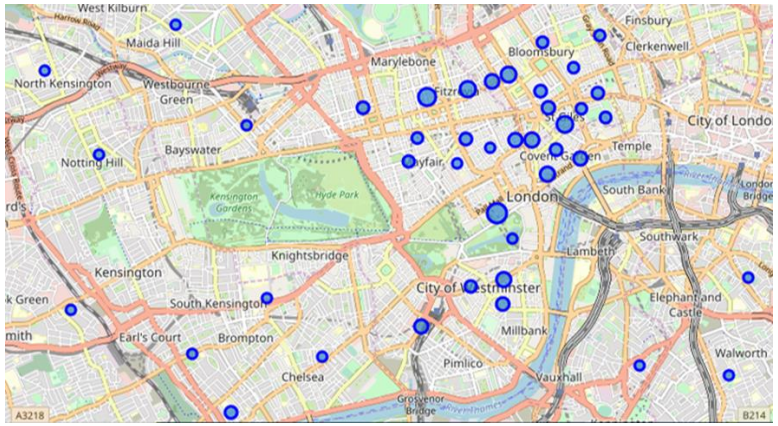
- Used k-means with k = 6
- One very big cluster (blue circles)
  - Many similar top-10 most common venues
  - Most common venues are hotels theatres, coffee shops, restaurants
- The pubs districts (purple)
  - Pubs is the 1st most common venue
- The grocery stores (red)
  - High frequency of grocery stores
- Three single district clusters



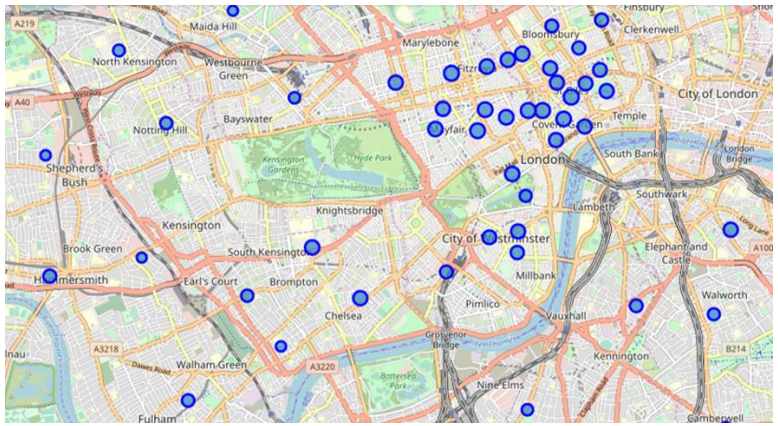
Competition

- Coffee Shops in candidate locations
  - Low to medium price range
  - High ratings
- These are Important challenges for the company

Price Range



Ratings



Future Directions

- Get data from other location data providers
- Expand analysis across other dimensions
  - Cost factors e.g. labour and property costs
  - Population demographics e.g. age, income, spending
  - Conduct surveys about personal time and leisure preferences

Conclusion

- London is a rather competitive environment thus
  - Carefully plan all steps
  - Pay attention to the findings of this project
  - Extend analysis to other aspects not included in this project

References

- Jupyter notebook with the analysis

https://github.com/constantinoslazarou/Coursera_Capstone/blob/master/Applied_Data_Science_Project_Final.ipynb

- Project final report

https://github.com/constantinoslazarou/Coursera_Capstone/blob/master/Applied_Data_Science_Capstone_Project_Report.pdf