

Capstone Project: Coursera Applied Data Science

Open a Supermarket in Greater Manchester



ZIYANG LIU

2020/07/24

Introduction

- ▶ Greater Manchester is a metropolitan county and combined authority area in North West England, with a population of 2.8 million; the second largest in England after Greater London.
- ▶ Today, Greater Manchester is the economic centre of the North West region of England and is the largest sub-regional economy in the UK outside London and South East England. Greater Manchester represents more than £42 billion of the UK regional GVA.



Introduction

- ▶ Supermarkets are the most important aspect of food production and distribution as they are the interface between supply and demand. It increase local consumption and also provide food with lower prices to residents.
- ▶ The success of opening a supermarket hinges on various aspects, one of the most important, is to choose the ideal location.

Business Issue

- ▶ This project aims to use machine learning techniques to find the best place to open a shopping mall in Philadelphia.
- ▶ The target people are businessmen, investors or even local government, who want to make profit via building a supermarket. It will also benefit residents.

Data

- ▶ Neighborhood Information
- ▶ Geographic Coordinates
- ▶ Venue Data
- ▶ Existing Supermarkets



Data Source

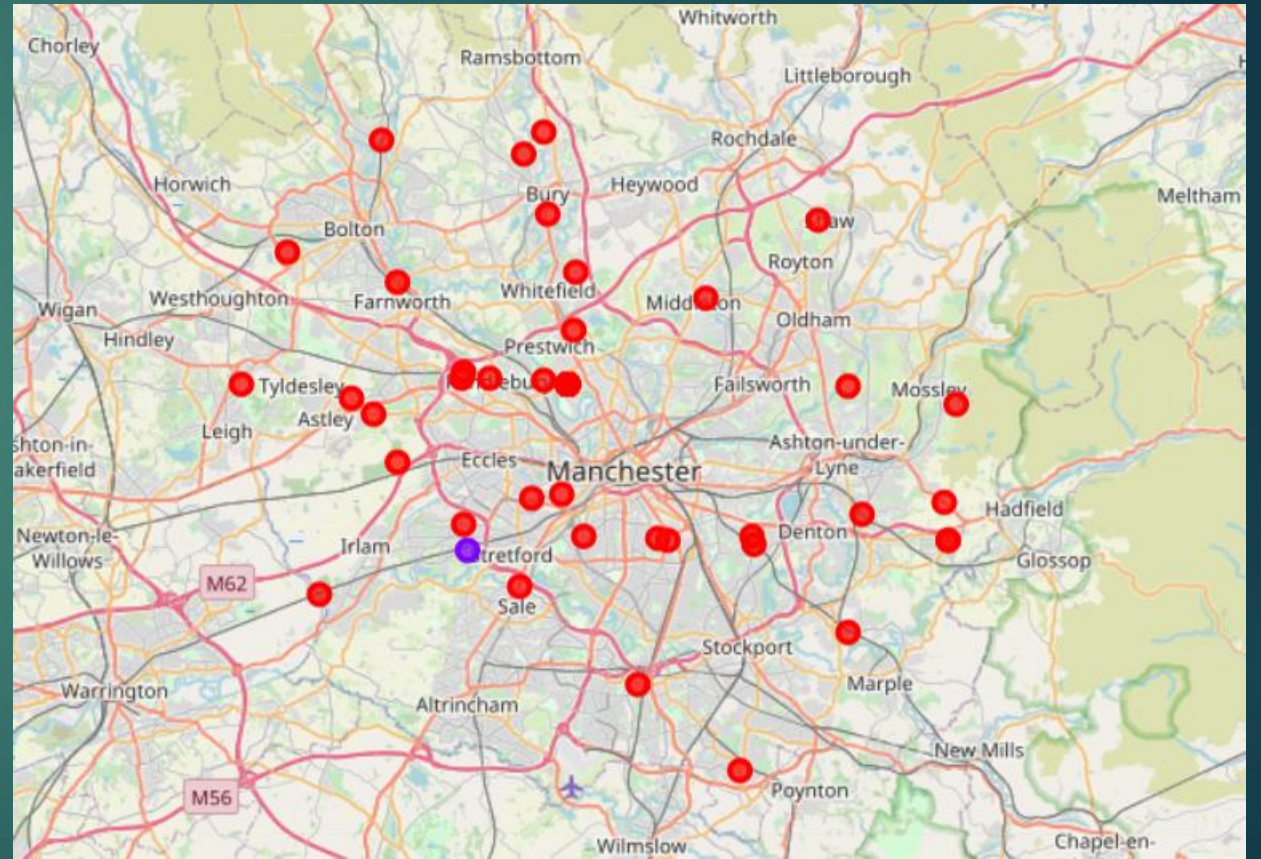
- ▶ The neighborhood data is available on https://en.wikipedia.org/wiki/Category:Areas_of_Greater_Manchester
- ▶ We use Foursquare API to access venue data, coordinates and other data.

Method

- ▶ Wiki page to find Neighborhoods
- ▶ Folium map and geocoder to get coordinate and visualize
- ▶ Foursquare API to access venue
- ▶ Group data and find frequency of each venue
- ▶ Use K-Means to build clusters
- ▶ Visualize using a map

Result

"Urmston is the best place to open a shopping mall."



Limitations

- ▶ In getting the venue data, 4 neighborhoods are missed automatically and may influence the accuracy of the result.
- ▶ The shopping mall frequency might be biased as the Foursquare query shows only few shopping mall.
- ▶ This project only focused on the frequency of shopping mall, other factors may also plays important roles here.

Conclusion

- ▶ In this capstone project, I worked through identifying business issues, reviewing K-Means algorithm's, using BeautifulSoup to get data from html and so on. It really helped me learn a lot and improved my analytical skills

Thanks

