Determining the feasibility of 'BreMiEgg', a bread, milk and egg delivering service in Manhattan

A capstone project by Nilesh Ade

Outline

- Introduction
- Description of data
- Methodology
- Results
- Discussion
- Conclusion

BreMiEgg

Bring me my bread, milk and eggs

(hypothetical brand)

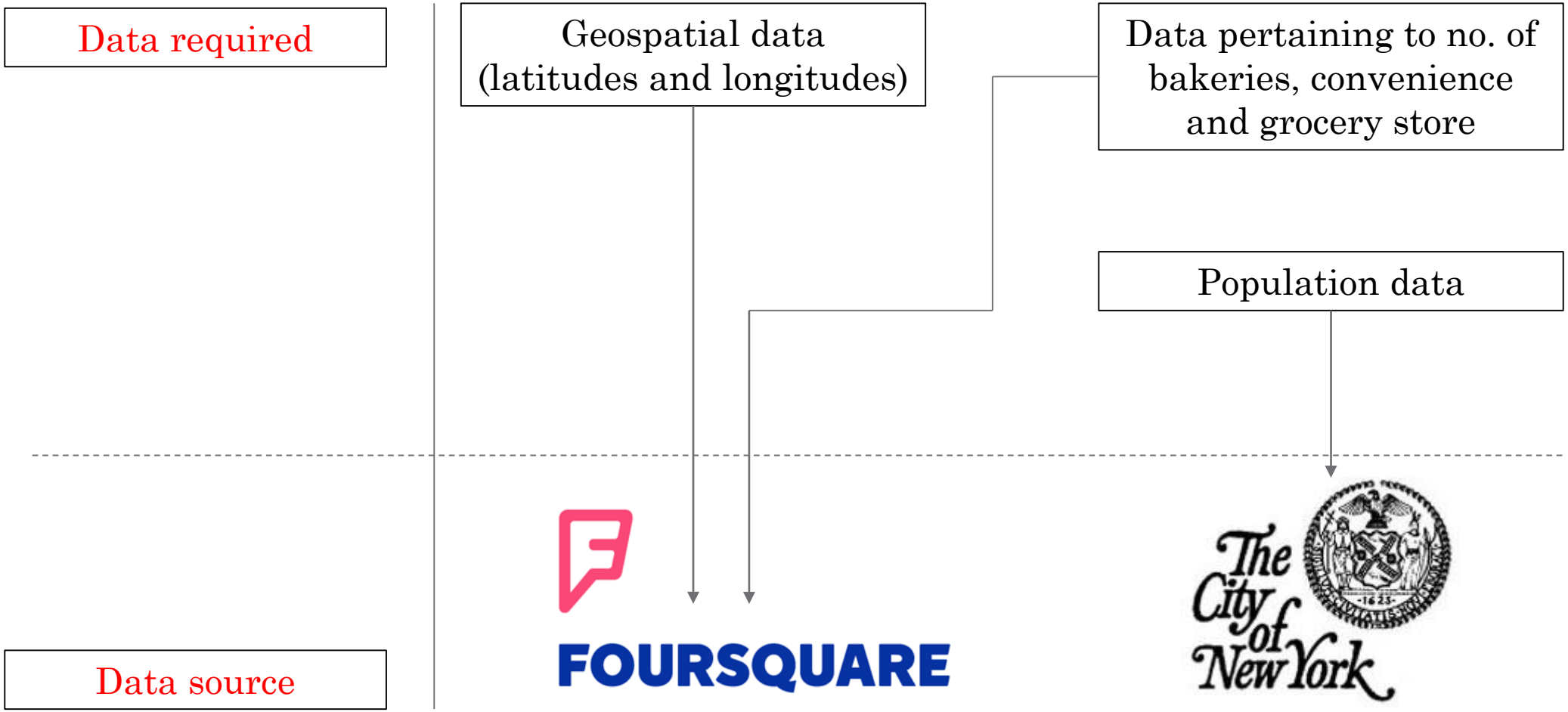
Introduction

- Bread, milk and eggs are staple components of American diet
- Sales of USD 24.7, 50.1 and 10.17 billion respectively
- Unfeasible to buy these products every day from grocery stores
- **BreMiEgg** – An online service that delivers bread, milk and eggs

- **Problem statement**: Find a suitable location for a warehouse for this service in the neighborhoods of Manhattan

Reference for statistics: https://www.statista.com

Description of data



Methodology

1. Normalization of data
2. Definition of Business Value:

   $$aP - bNb - cNc - dNg$$

3. Coefficients determined based on personal judgment
4. Clustering by k-means algorithm based on business value and geospatial data
5. Visualization of clusters by Folium

	Neighborhood	Population	Latitude	Longitude	Bakery	Convenience Store	Grocery Store
0	Marble Hill	0.353125	40.876551	-73.910660	0.0	0.0	0.0
1	Central Harlem	0.568690	40.815976	-73.943211	0.0	0.0	0.0
2	Hamilton Heights	0.366526	40.823604	-73.949688	0.5	0.0	0.0
3	Manhattanville	0.173367	40.816934	-73.957385	0.0	0.0	0.0
4	Morningside Heights	0.422495	40.808000	-73.963896	0.0	0.0	1.0

Normalized dataframe example

Variables: population (P), number of bakeries (Nb), convenience (Nc) and grocery stores (Ng) in different Manhattan neighborhoods

Results

1. Neighborhoods clustered in **four groups**

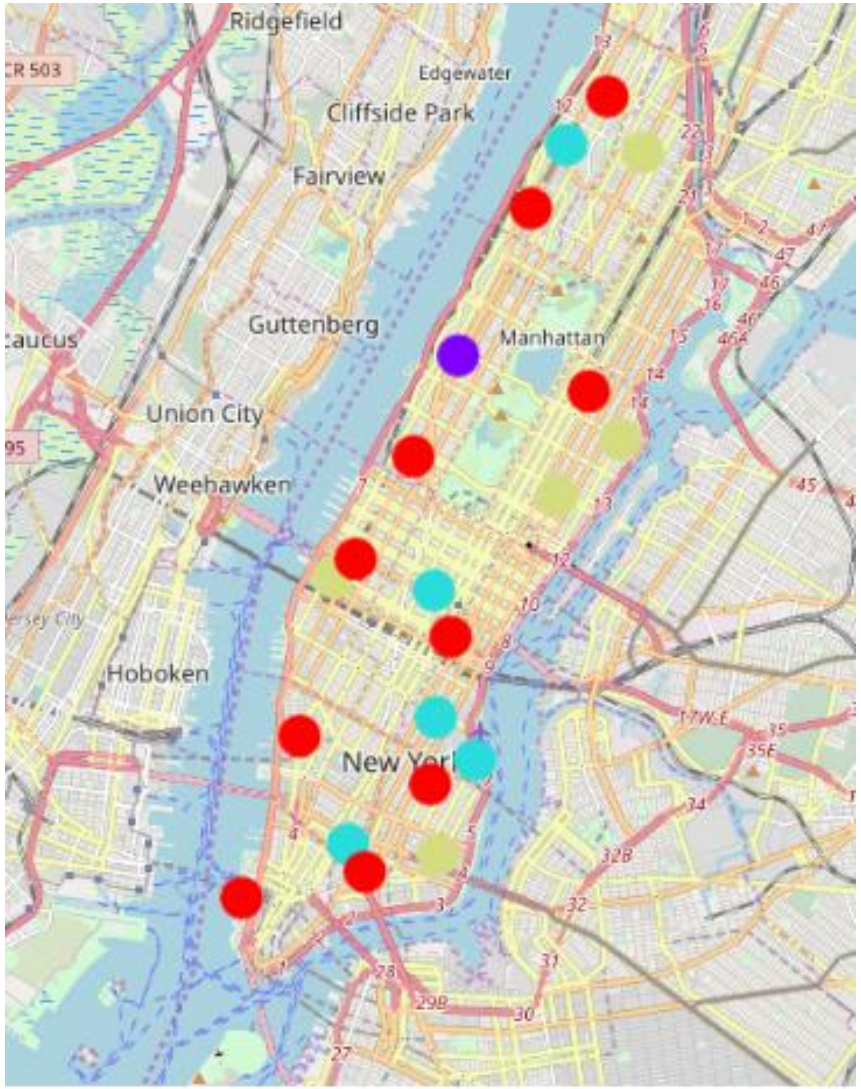
Cluster 0: red

Cluster 1: Purple

Cluster 2:Cyan

Cluster 3: Yellow

2. Clusters 0,1,2 and 3 have 11,1,5 and 6 neighborhoods respectively



Discussion

1. Described approach is successful
2. Cluster 1 is ideal location for warehouse (Upper West Side neighborhood)

Reason:

a) Itself has high average business value

b) closer to other clusters with high average business values

Cluster	Color	Average business value
0	Red	1.03
1	Purple	2.70
2	Cyan	0.39
3	Yellow	1.56

Discussion

Limitations:

1. Few neighborhoods neglected due to lack of population data
2. Population data is not recent (from 2010)
3. Coefficients in business value function assigned based on personal judgment
4. Number of clusters arbitrarily chosen as 4

Conclusion

- Capstone project consisting of Population data collection/wrangling, foursquare API data collection/wrangling, K-means clustering, and folium map visualization is successful in attaining the required goal.
- Can be improved by using recent data and including various other parameters such as commercial rents, utility costs etc.
- Approach can be implemented to address a variety of similar problems such as determining the location for a retail delivery service, location of a supermarket etc.

BreMiEgg

Bring me my bread, milk and eggs

Note: Refer the report for detailed description at:
https://github.com/nilesh149/Coursera_Capstone/blob/master/Report.ipynb

Thank you