'How does population density influence business decisions?'

Zahra Adahman, MBS

IBM Data Science Professional Certificate-Capstone Project

Introduction

- An important business strategy is to understand the factor that could be important for maximizing profit.

- One factor is location of business enterprise.

- Higher traffic can drive up the chance of increasing and maintaining profits. High demand of people correlates with location with higher population density.

- An example of a city with this high population density is New York city. New York city is one of the top ten cities in the world with the highest population density per square mile.

- The understanding of the relationship between these population densities and the density of established businesses by neighborhood would enlighten business decision for potential stakeholders to determine what neighborhoods' needs for specific good or service.

Data acquisition and cleaning

- The data used for the study were gotten from different sources and via different techniques.

- The population data, the regular and polygon geoJSON of New York city containing the Manhattan Borough by Neighbourhood were downloaded from the internet.

- TThe data of the top places to go in Manhattan were scraped from the FourSquare site using the developer API access with a radius of 500 and limit of 1000 places.

- The data were sorted and combined into one dataframe for analysis by venue category, bakery.

Methodology

- Seahorse package - linear regression

- Matplotlib package - dot plots

- SciKit learn package – polynomial regression and modeling, R^2 score at the power of 2 and 6 .

- Independent variable – population.

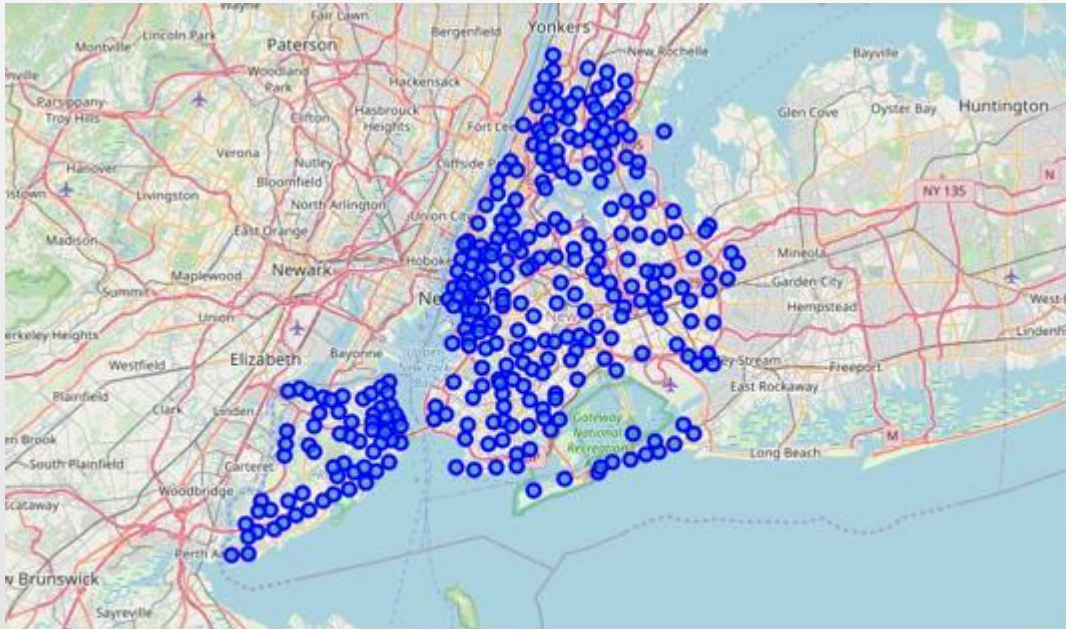- Dependent or target variable -bakery count.

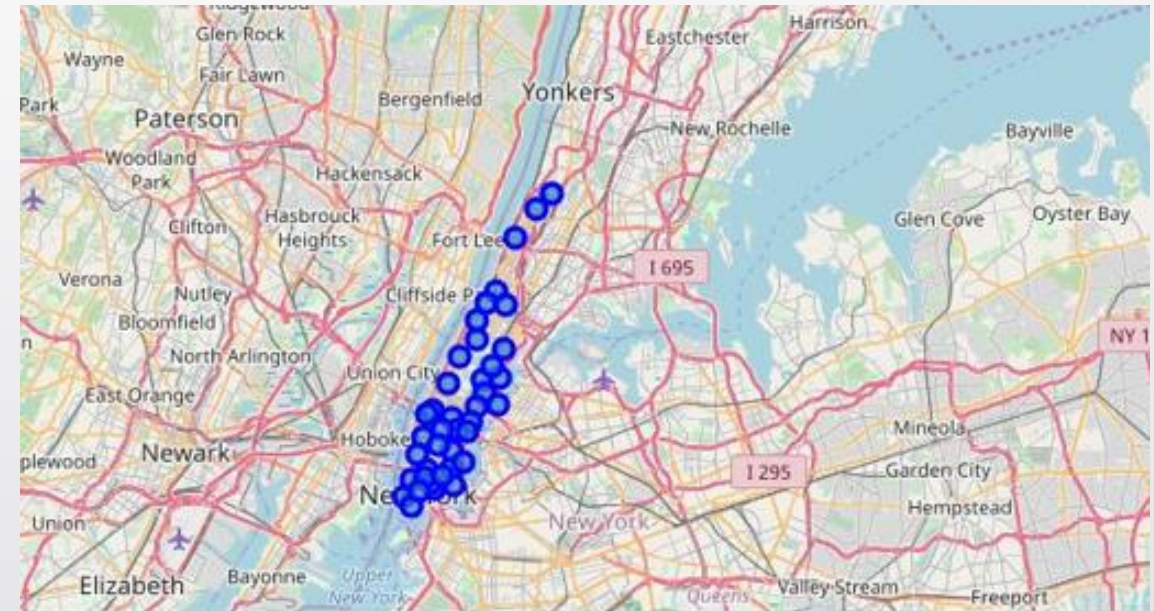**Fig.1** Blue dots represent all the New York City neighborhoods.

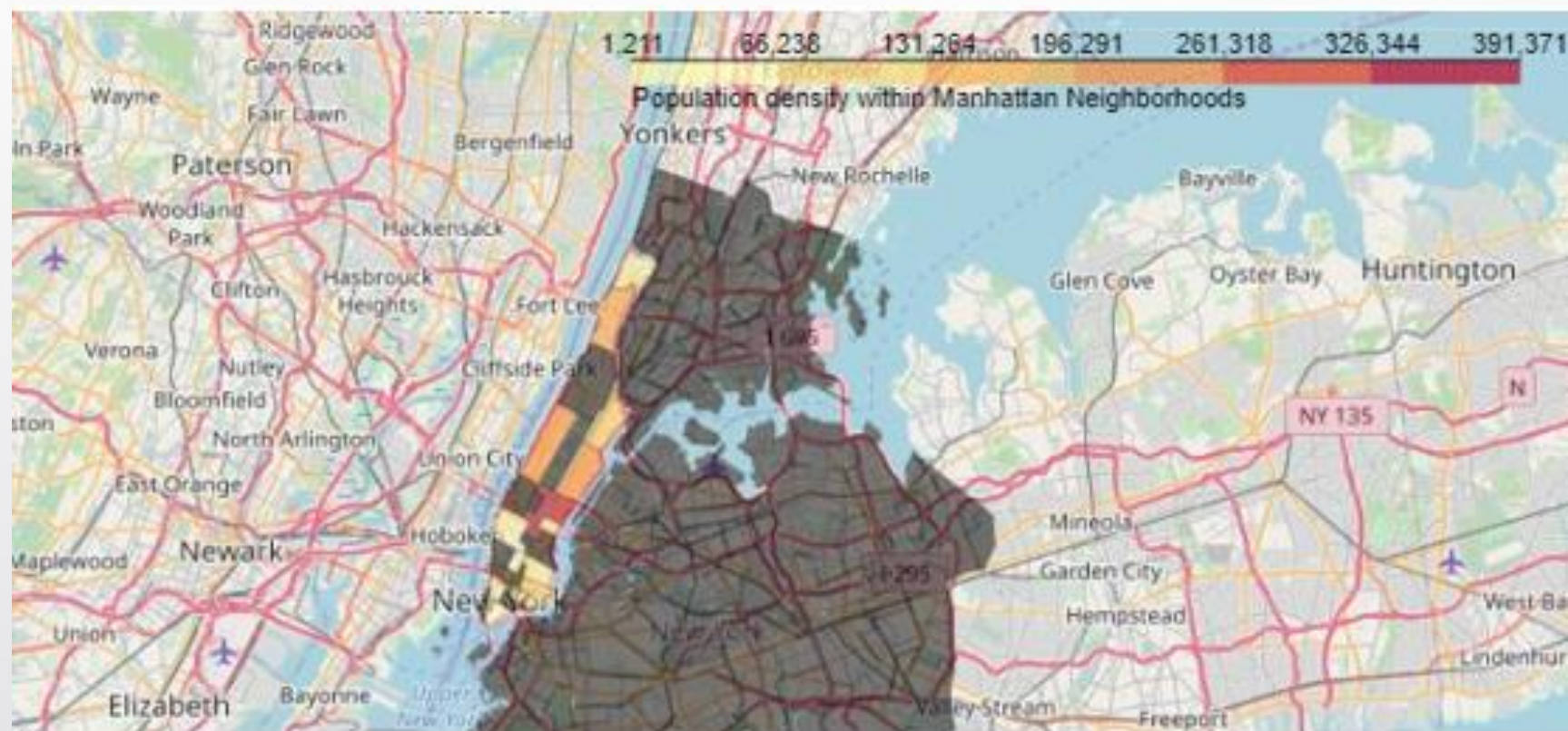**Fig.2** Blue dots represent all the New York City-Manhattan Borough neighborhoods.

Fig.3 Choropleth map dots showing the population density within the Manhattan Borough neighborhoods.

	Neighborhood	Population	Bakery count
0	Midtown	391371	3
1	Central Harlem	335109	1
2	Upper East Side	229688	4
3	Upper West Side	209084	3
4	Washington Heights	158318	4
5	East Harlem	115921	4
6	Chinatown	100000	4
7	Lower East Side	72957	2
8	East Village	62832	2
9	Lincoln Square	61489	2
10	Financial District	60976	1
11	Hamilton Heights	48520	2
12	Inwood	46746	2
13	Chelsea	38242	3
15	Yorkville	36221	1
16	Noho	24846	2
17	Greenwich Village	22785	2
18	Soho	19673	3
19	Tribeca	17362	2
20	Murray Hill	10284	1
22	Flatiron	8547	2
23	Little Italy	1211	6

Marble Hill
Chinatown
Washington Heights
Inwood
Hamilton Heights
Manhattanville
Central Harlem
East Harlem
Upper East Side
Yorkville
Lenox Hill
Roosevelt Island
Upper West Side
Lincoln Square
Clinton
Midtown
Murray Hill
Chelsea
Greenwich Village
East Village
Lower East Side
Tribeca
Little Italy
Soho
West Village
Manhattan Valley
Morningside Heights
Gramercy
Battery Park City
Financial District
Carnegie Hill
Noho
Civic Center
Midtown South
Sutton Place
Turtle Bay
Tudor City
Stuyvesant Town
Flatiron
Hudson Yards

Fig.4 . The final dataframe containing the bakery count and population for each neighborhood in Manhattan. B. The fourteen neighborhoods missing in the final dataframe highlighted in green.
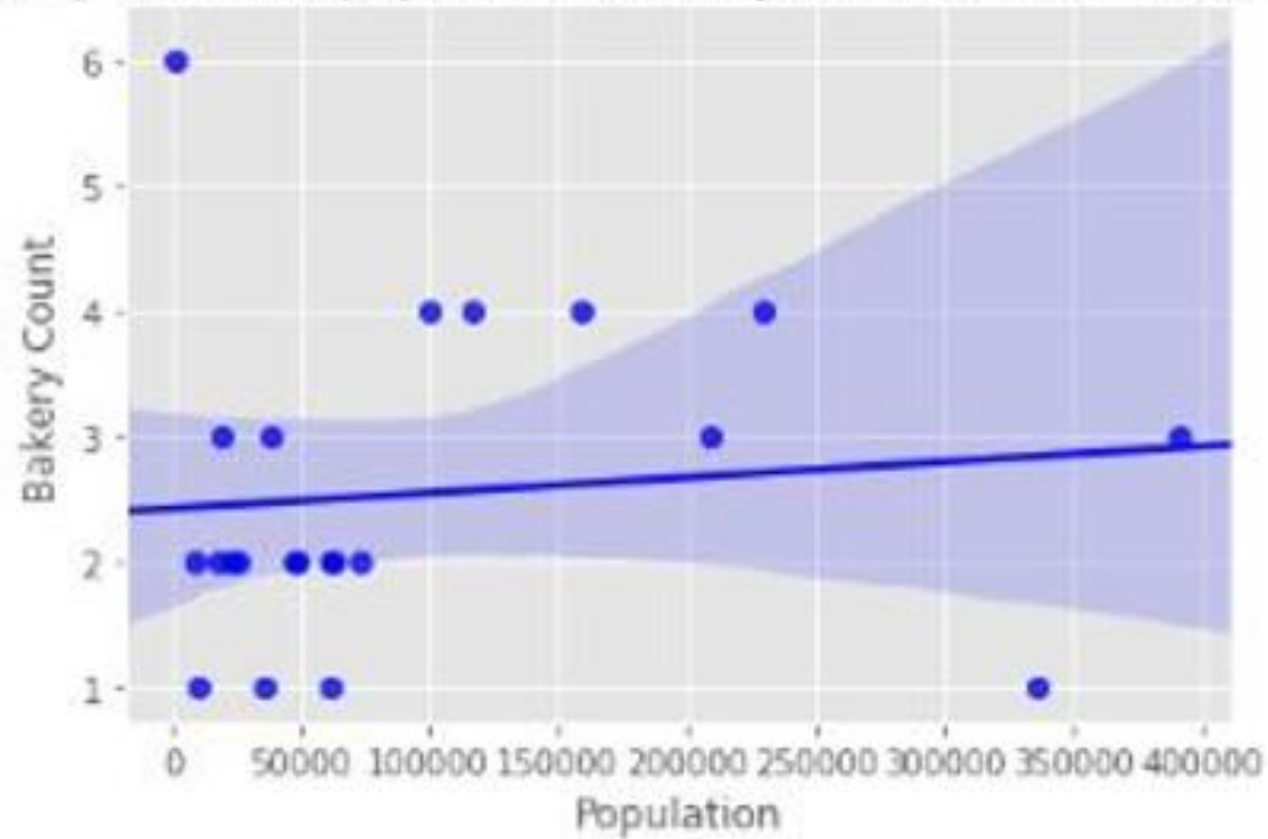
Fig.5 Seahorse regression plot of the relationship between population and bakery counts in neighborhoods is not linear.

Results

- The R^2 score of the polynomial regression at the power of 2 was negative 12. Showing there is no significant correlations between the two variables.

- The R^2 score when the data was fitted to a polynomial regression at the power of 6 improved, however, the score still was still negative, not close to 1.

Conclusion and discussion

- There is no significant correlations or relationship between the number of bakeries and population density.

- Business associates are advised to look at other variables such as access to public transportation, distribution of schools, parks and so on in different neighborhoods in a city to determine where to open a bakery for the highest profit possible.

Thank you