Capstone Presentation

Name: Tang Ka Kit
Contact: tangjay58@gmail.com

Link: https://github.com/jaytang0508/Coursera_Capstone/blob/master/Cousera_Capstone%20(3).ipynb

Table of Content

Introduction

- **Target:** compares the consumer behaviours among income groups in New York City (NYC)

- **Method:** observing the citizens' favourite types of venues

- **Beneficiaries:** Entrepreneurs

- At the end, you will know:

1. What kind of business should we establish?

2. Which income group should we target?

3. Simple program designed for letting you know which exact neighborhoods to set up the business

Data

- Median Income of the neighborhoods in NYC

  - https://www.renthop.com/study/assets/new-york-city-cost-of-living-2017/nyc-2br-median-rent-and-income-table.html

- Longitude & Latitude data of the neighborhoods in NYC

  - https://geo.nyu.edu/catalog/nyu_2451_34572

- Common Venue data of the neighborhoods in NYC

  - Foursquare

Methodology

- Divided into 4 parts…

1. Import and Clean data

2. Explore the Neighborhood

3. Analyze the relationship between income group and common venues

4. Create a tool to display where the venues are famous to visit

Import and clean data and explore the neighborhood

- All data are merged into a dataframe from the mentioned sources
- Result:

	Neighborhood	Borough	Median Income in dollar	Latitude	Longitude	1st Most Common Venue	2nd Most Common Venue
0	Long Island City	Queens	28,378	40.750217	-73.939202	Hotel	Coffee Shop
1	Williamsburg	Brooklyn	21,502	40.707144	-73.958115	Bagel Shop	Coffee Shop
2	Lower East Side	Manhattan	31,273	40.717807	-73.980890	Coffee Shop	Café

How to divide the group? (1)

- Manually seems impossible

The median income in each neighborhood in New York
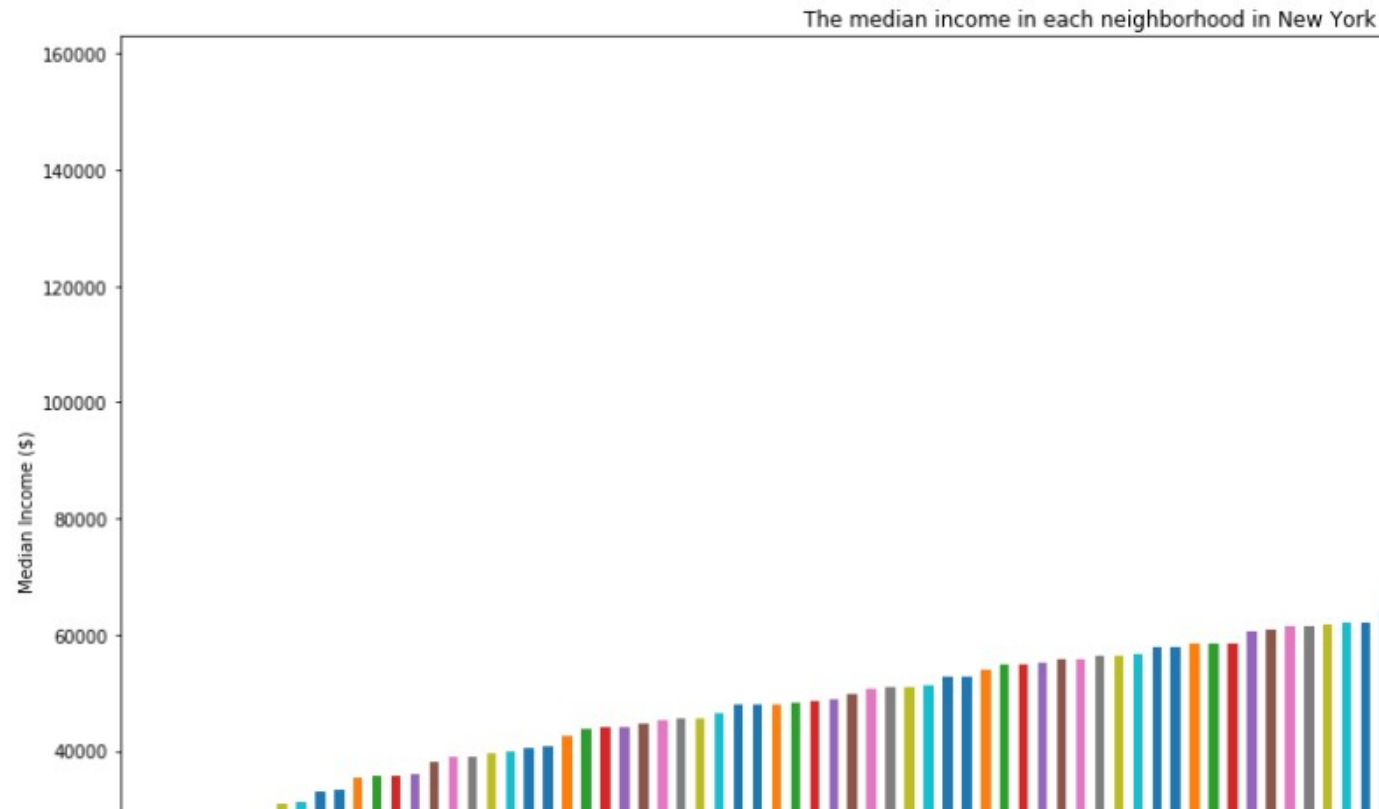
How to divide the group? (2)

- Use K means clustering on their median income

- **Result (USD per year):**
- Low Income (20334 ~ 52696): 12 Neighborhoods
- Moderate Income (53836 ~ 85496): 45 Neighborhoods
- High Income (88868 ~ 155213): 45 Neighborhoods

Result – intra-group comparison

- Answers the question: 'what kind of business should one establish in each of the income group'

- Outstanding venue in each group:

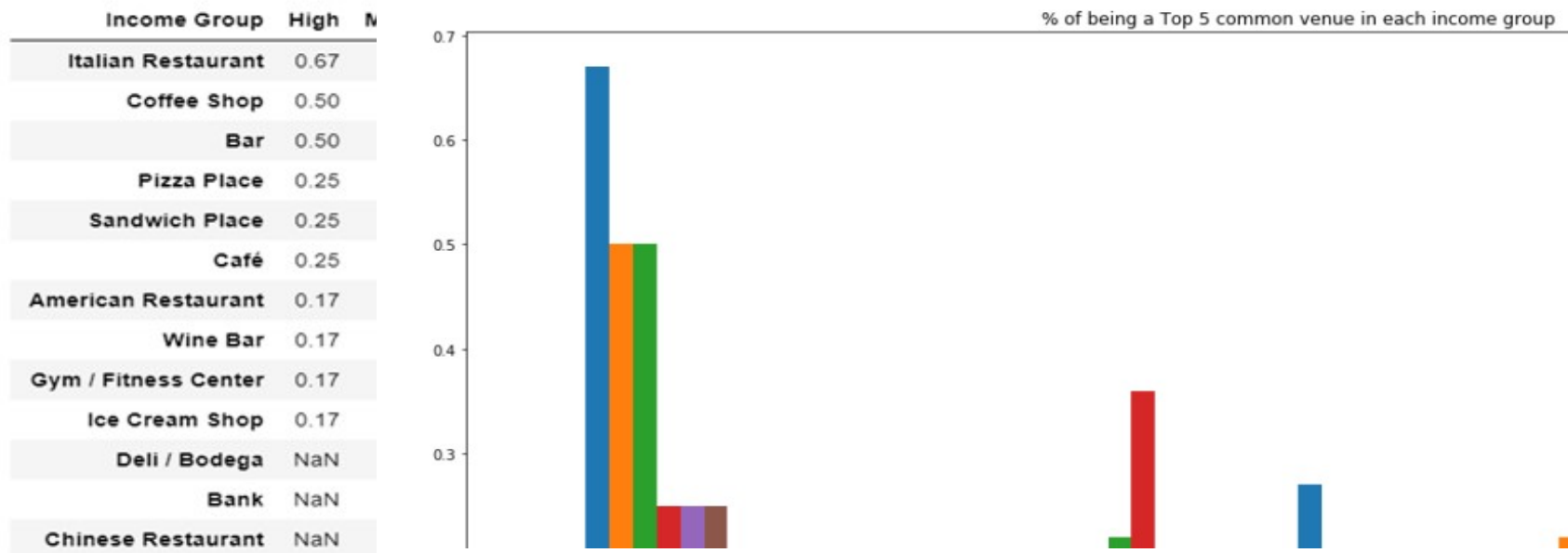{High: Italian Restaurant, Moderate: Pizza Place, Low: Deli/ Bodega and Pizza Place}

Income Group	High
Venue	
Italian Restaurant	8
Coffee Shop	6
Bar	6
Pizza Place	3
Sandwich Place	3
Café	3

Income Group	Moderate
Venue	
Pizza Place	16
Deli / Bodega	12
Bar	10
Bank	9
Chinese Restaurant	9
Park	9

number of being a top 5 common venue in each
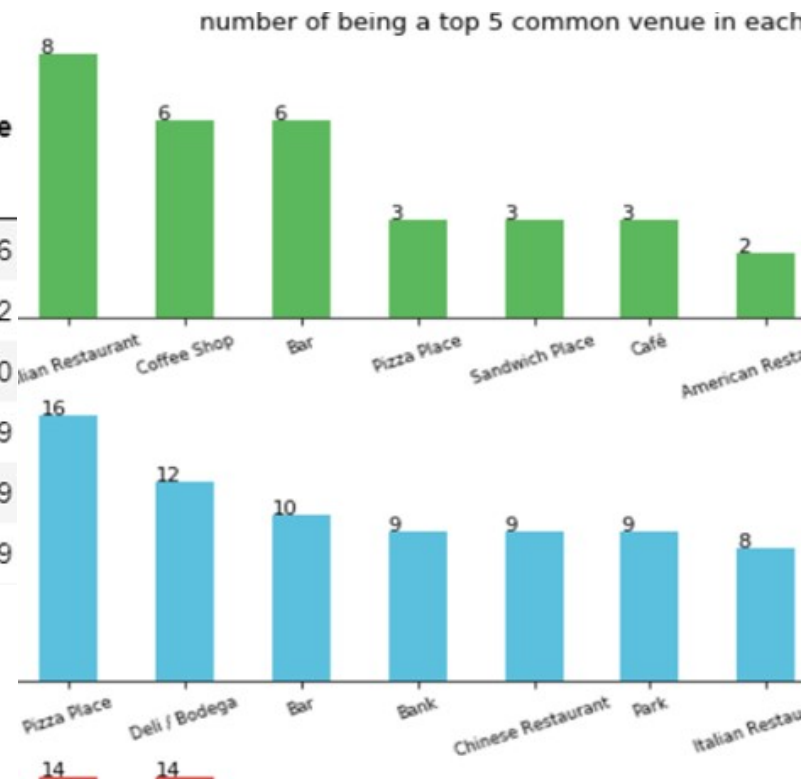
Result – inter-group comparison

- Answers the question: 'which income group one should target'

- **Formula**

- A color bar appears once = must do business targeting that group

- A color bar appears twice or above = choose group with higher % at first, expand to the second group or the third later

Income Group	High	M
Italian Restaurant	0.67	
Coffee Shop	0.50	
Bar	0.50	
Pizza Place	0.25	
Sandwich Place	0.25	
Café	0.25	
American Restaurant	0.17	
Wine Bar	0.17	
Gym / Fitness Center	0.17	
Ice Cream Shop	0.17	
Deli / Bodega	NaN	
Bank	NaN	
Chinese Restaurant	NaN	

% of being a Top 5 common venue in each income group

The tool: *search_fame('venue name')*

- show the neighborhood and income group being a top 5 common venue
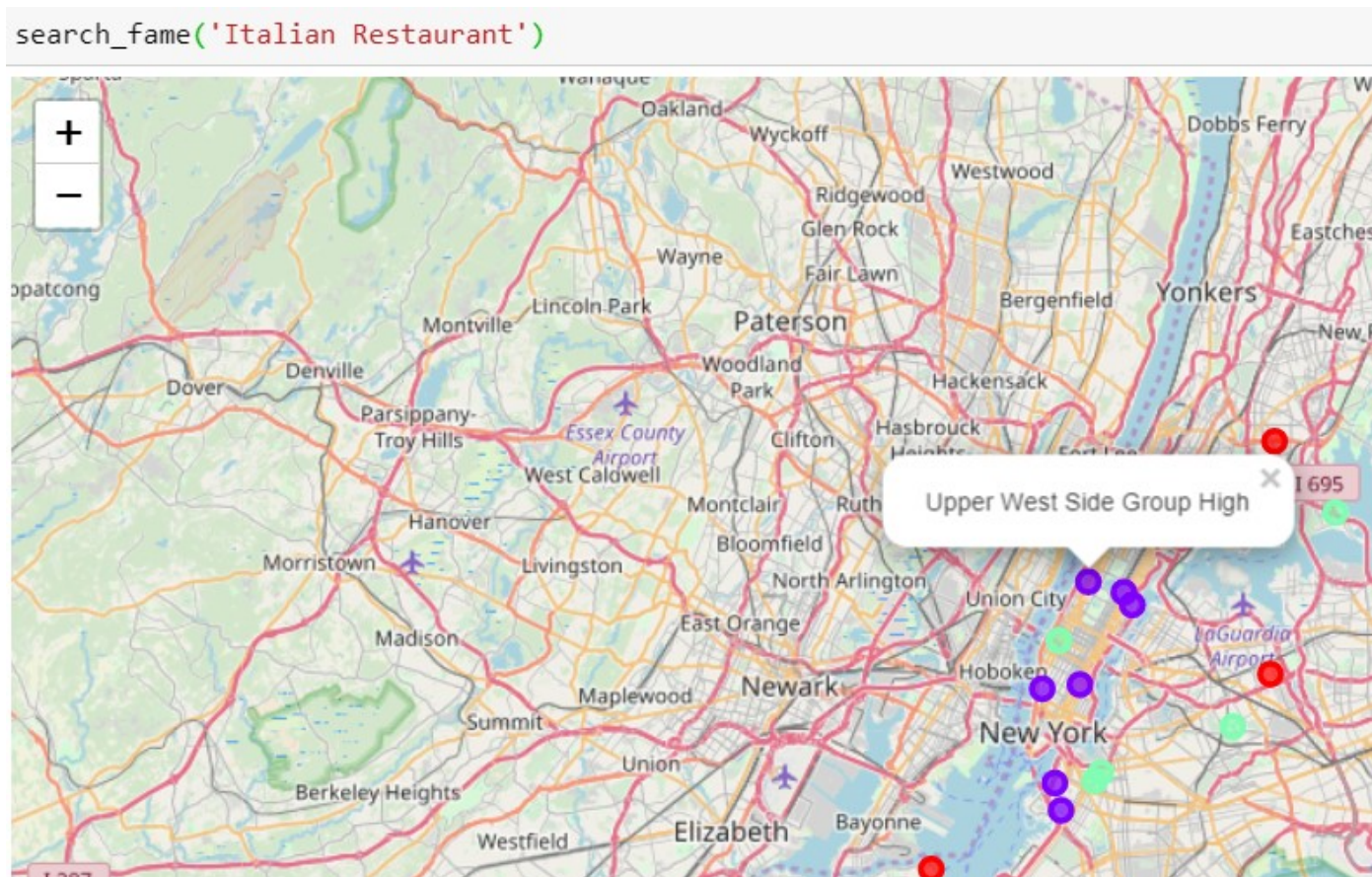
How to improve?

- Enlarge the data set for the high income group
- Append the data set for the turnover of all venues
- Apply logistic regression for the part of inter-group comparison

Conclusion

- Understand the market situation in NYC

- Earn insights about the consuming properties of different income groups

- Helps entrepreneurs to target the places and groups they should start or expand their businesses