Locating an Ideal Neighborhood for Small Business in Atlanta, GA

Sravan Neela

Applied Data Science Capstone

IBM Data Science Professional Certificate

Dec 2020

# Executive Summary

Small business often struggles during their initial years (1-5yrs) due to several factors that eventually lead them to failure. Some reasons are more common than others such as market needs, location, demography (Otar, 2018). This analysis will try to increase the odds of business success by considering safety, income, population for ideal placement of a small business from a range of options.

# Introduction: The (Small) Business Problem

The objective of this analysis is to find relatively safe and potentially profitable location in the city of Atlanta, GA to start a small business. Venues for small businesses such as restaurants, sports bar, coffee shops, saloon or a spa will be explored. Small business entrepreneurs who are considering opening such venues will be target audience for this report.

The report will try to make a recommendation to identify locations by analyzing key statistics such as crime reports, demographics, such as income, population and real-estate values from datasets made available city’s government (APD, 2020) (City of Atlanta, 2020).

Since the target audience are small entrepreneurs to reduce the startup costs, report will try to use publicly (for free to public) available government sources to make location recommendations. Analysis of these datasets will help provide key insights to make informed decisions such that the business is in relatively safe neighborhood and can provide higher business turnover.

The analysis does not assume the individual has a specific venue in mind but will provide several options from the information available from the analysis.

The analysis will aim to provide key insights to potentially avoid leading causes of what causes a small business to fail in their initial year of startup (Otar, 2018).

# Data

Per the objectives set for this analysis, the following datasets will be gathered to make location recommendations by analyzing

1. Atlanta city crime data to identify safe locations
2. GIS data from Atlanta Department of City Planning to obtain geographical information
3. Census data from City of Atlanta to get neighborhood demographics such as median income, population and median home values by (Neighborhood planning units) NPU.
4. Finding most common and least common venues in proximity of an ideal location using Foursquare API.

To make these decisions the analysis will use data from Atlanta's official public data sources listed below:

1. Atlanta Police Department (APD): Data repository containing historic crime reports dating back to 2009. These data sets contain information relevant to analysis they are: -
2. Report Number: PD's Record locator.
3. Occur Date: Date the crime occurred.
4. Report Date: Date the crime was reported or recorded in PD's System.
5. Location: Street address where the crime was reported.
6. UCR Literal: Nature and type of crime.
7. Neighborhood: Neighborhood where the crime was reported.
8. Latitude & Longitude: Geographical location of reported crime.

For this analysis data was extracted from APD website for years 2017 to 2020.

Location of Dataset extract: <https://raw.githubusercontent.com/sravann/Coursera_Capstone/master/COBRA_2017-2020.csv>

APD official Source:

<https://www.atlantapd.org/i-want-to/crime-data-downloads>

1. Atlanta Department of City Planning: official GIS information from containing Neighborhood Planning Unit (NPU) information with Neighborhood name and NPU code. These data sets contain information relevant to analysis they are:
2. NAME: Name of the neighborhood
3. NPU: NPU code from A-Z, groups neighborhoods from A to Z.

Location of Dataset:

<https://dpcd-coaplangis.opendata.arcgis.com/datasets/npu>

1. City of Atlanta- official Website: Website with information on neighborhood demographic information. The data set contains information on:
2. NPU: Atlanta's Neighborhood Planning Unit (NPU) identifier.
3. Region: Relative location of neighborhood group in Atlanta city.
4. Median Age: identify general age group of the neighborhood to cater targeted goods and services.
5. Median Income: To asses income of general population that will help with pricing of goods and services.
6. Median Home Values: To provide general information on housing and real estate.
7. Population: Total population of the neighborhood.
8. - Growth: population growth of the neighborhood to asses growing or declining neighborhoods.

Location of consolidated dataset: <https://raw.githubusercontent.com/sravann/Coursera_Capstone/master/COBRA_2017-2020.csv>

1. Foursquare API: Foursquare API will be used to get comprehensive list of venues located in the ideal neighborhood from which the report provides insights into most and least common venues in the neighborhood. Foursquare API will help extract:
2. The list of venues in the proximity of a narrowed down ideal neighborhood list.
3. List of venues such as restaurants, ice cream parlors, bars etc.

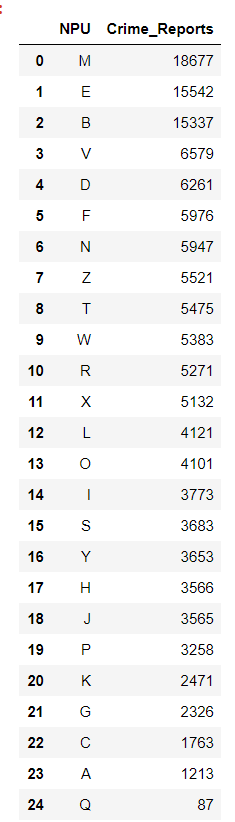
# Methodology

Methodology used for this analysis will include Exploratory Data Analysis and Modelling.

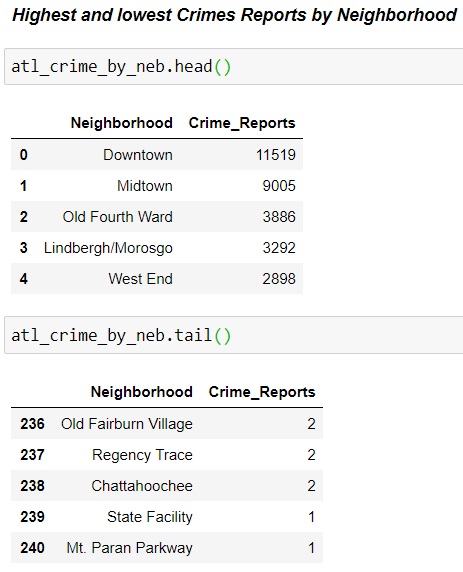
## EDA

Exploratory Data Analysis was performed to help investors with some key insights into the city. This was done by counting the crime reports across the city's NPUs and their respective neighborhoods.

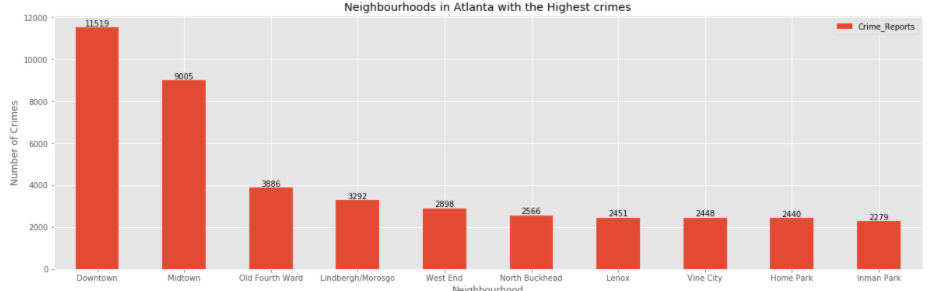
From the datasets gathered we can look at neighborhoods grouped by NPUs and crime reports.

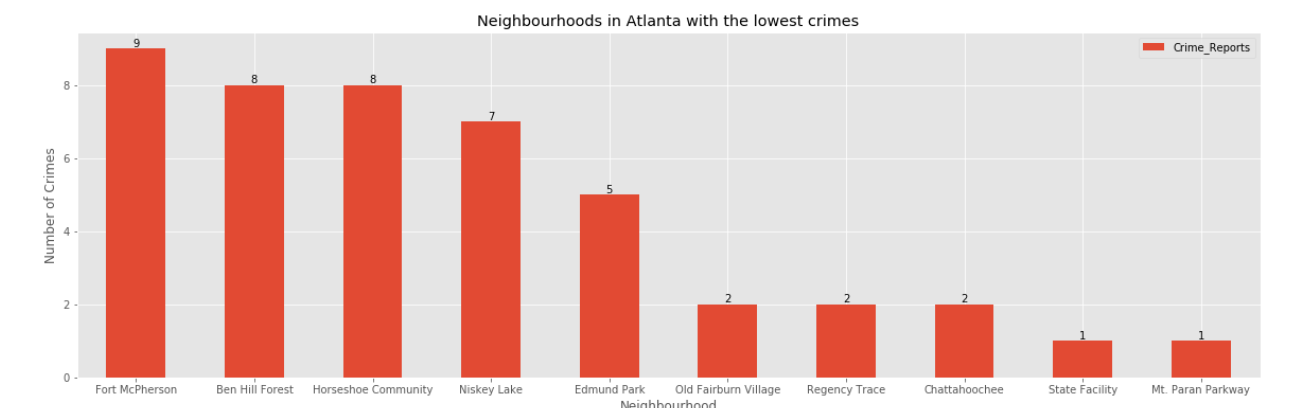


And then highest and lowest Crimes reported in neighborhoods

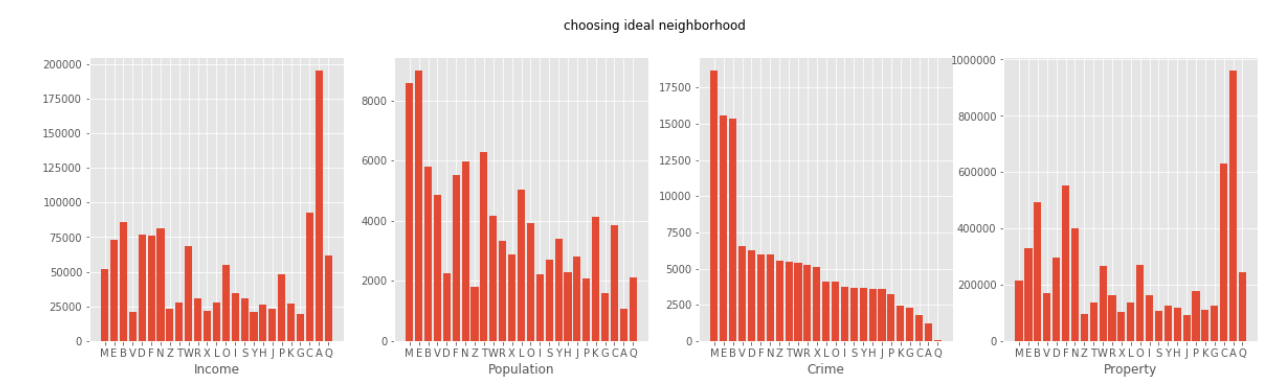


Visualizing Crime Reports





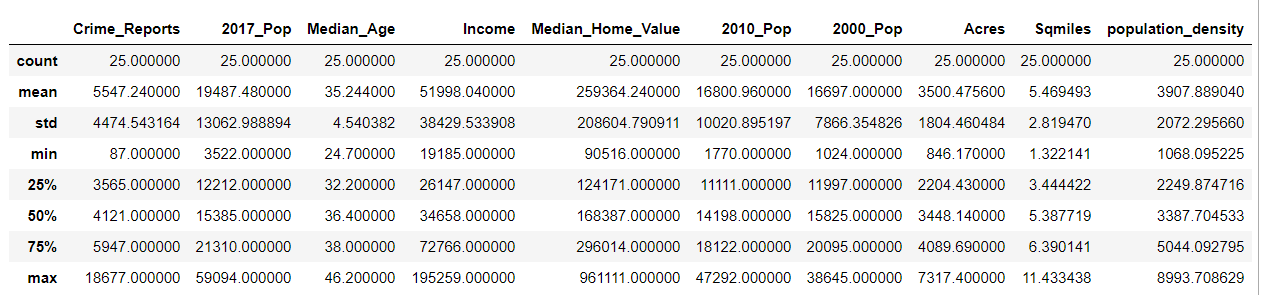
Importing census data to analyze and visualize demographic information



## Modelling

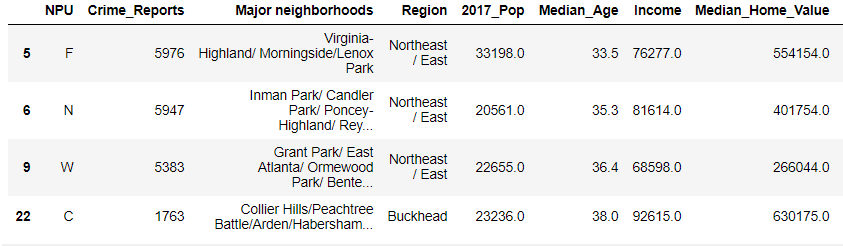
Also, to help investors use the right neighborhood within an NPU we will be clustering similar neighborhoods using K - means clustering using datasets based on predefined cluster size. K-Means clustering will be used to a group data based on existing venues which will provide insights to make informed decisions.

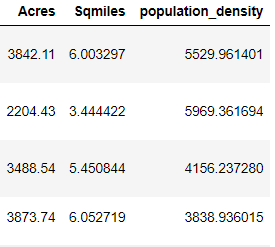
Preliminary data description



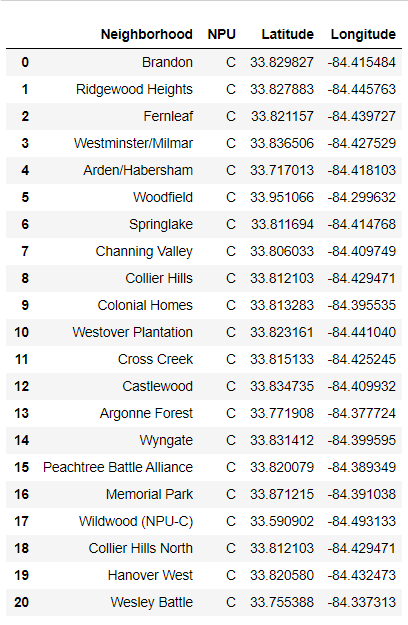
Once required datasets were imported, merging and then sorting the data set with the following criteria

1. Median income >$65000/yr
2. Population density > 3000/sq-mile
3. Crime Reports for the last 3 yrs. <6000

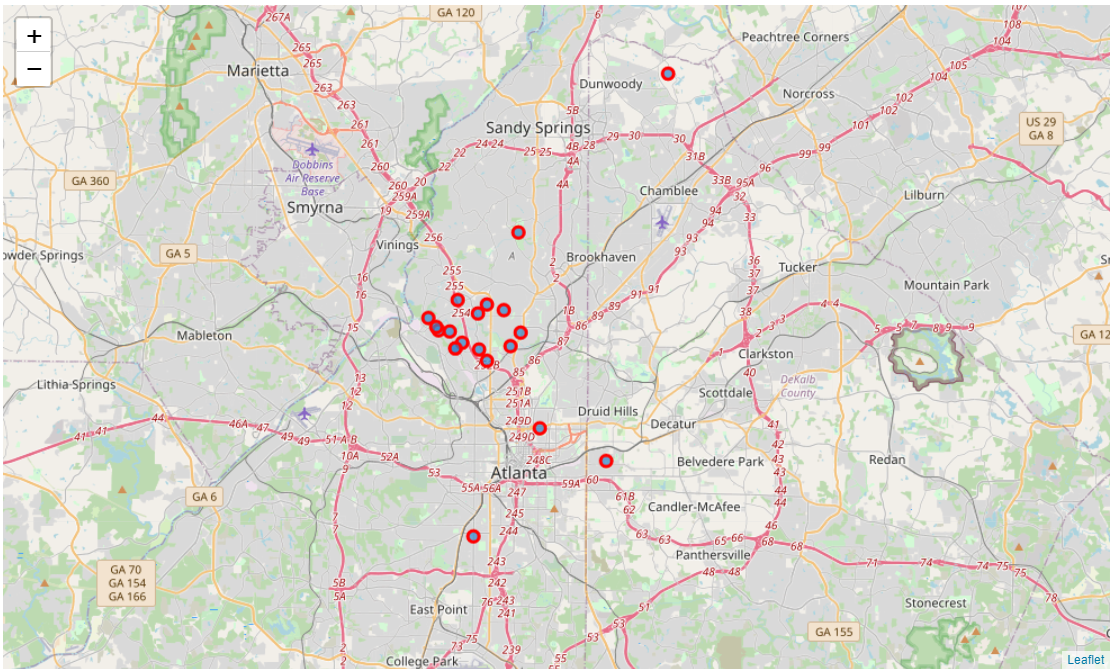




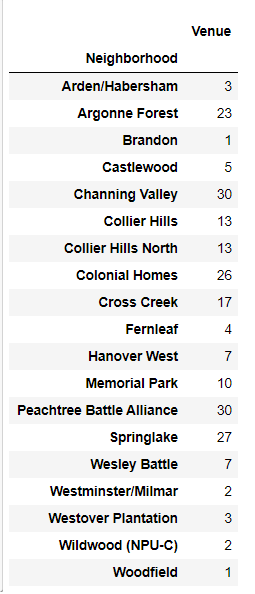
After applying the criteria, NPU C was selected due to lowest crime reports in the group.



Plotting the neighborhoods in NPU C



Number of venues in each neighborhood



Filtering the top five venues in each neighborhood.

----Arden/Habersham----

venue frequency

0 Juice Bar 0.33

1 Scenic Lookout 0.33

2 Food Truck 0.33

3 American Restaurant 0.00

4 Pharmacy 0.00

----Argonne Forest----

venue frequency

0 Southern / Soul Food Restaurant 0.09

1 Ice Cream Shop 0.04

2 Brewery 0.04

3 Food Truck 0.04

4 Gay Bar 0.04

----Brandon----

venue frequency

0 Ice Cream Shop 1.0

1 American Restaurant 0.0

2 Pet Store 0.0

3 Rental Car Location 0.0

4 Recreation Center 0.0

----Castlewood----

venue frequency

0 Dog Run 0.2

1 Wine Shop 0.2

2 Ski Trail 0.2

3 Boat or Ferry 0.2

4 Gym 0.2

----Channing Valley----

venue frequency

0 Food Truck 0.10

1 Mexican Restaurant 0.07

2 Department Store 0.07

3 Sandwich Place 0.07

4 Pub 0.03

----Collier Hills----

venue frequency

0 Italian Restaurant 0.15

1 Taco Place 0.15

2 Pizza Place 0.08

3 Cupcake Shop 0.08

4 Gas Station 0.08

----Collier Hills North----

venue frequency

0 Italian Restaurant 0.15

1 Taco Place 0.15

2 Pizza Place 0.08

3 Cupcake Shop 0.08

4 Gas Station 0.08

----Colonial Homes----

venue frequency

0 Pizza Place 0.08

1 Pharmacy 0.04

2 Doctor's Office 0.04

3 Financial or Legal Service 0.04

4 Food Truck 0.04

----Cross Creek----

venue frequency

0 Italian Restaurant 0.12

1 Golf Course 0.06

2 Playground 0.06

3 Gym 0.06

4 Cajun / Creole Restaurant 0.06

----Fernleaf----

venue frequency

0 Dog Run 0.25

1 College Track 0.25

2 Pool 0.25

3 Speakeasy 0.25

4 American Restaurant 0.00

----Hanover West----

venue frequency

0 Gym 0.14

1 Shop & Service 0.14

2 Wine Bar 0.14

3 College Track 0.14

4 Gym / Fitness Center 0.14

----Memorial Park----

venue frequency

0 Park 0.2

1 Golf Course 0.1

2 Gym / Fitness Center 0.1

3 Tennis Court 0.1

4 Theater 0.1

----Peachtree Battle Alliance----

venue frequency

0 American Restaurant 0.07

1 Mexican Restaurant 0.07

2 Mediterranean Restaurant 0.07

3 Pharmacy 0.07

4 Tanning Salon 0.03

----Springlake----

venue frequency

0 Pizza Place 0.07

1 Massage Studio 0.07

2 Sandwich Place 0.07

3 Salon / Barbershop 0.07

4 Pet Store 0.07

----Wesley Battle----

venue frequency

0 Grocery Store 0.14

1 Automotive Shop 0.14

2 Recreation Center 0.14

3 Construction & Landscaping 0.14

4 Park 0.14

----Westminster/Milmar----

venue frequency

0 Jewelry Store 0.5

1 Beer Garden 0.5

2 American Restaurant 0.0

3 Pharmacy 0.0

4 Residential Building (Apartment / Condo) 0.0

----Westover Plantation----

venue frequency

0 Baseball Field 0.33

1 Pool 0.33

2 Outdoors & Recreation 0.33

3 American Restaurant 0.00

4 Pharmacy 0.00

----Wildwood (NPU-C)----

venue frequency

0 Rental Car Location 0.5

1 Pool 0.5

2 American Restaurant 0.0

3 Latin American Restaurant 0.0

4 Residential Building (Apartment / Condo) 0.0

----Woodfield----

venue frequency

0 Campground 1.0

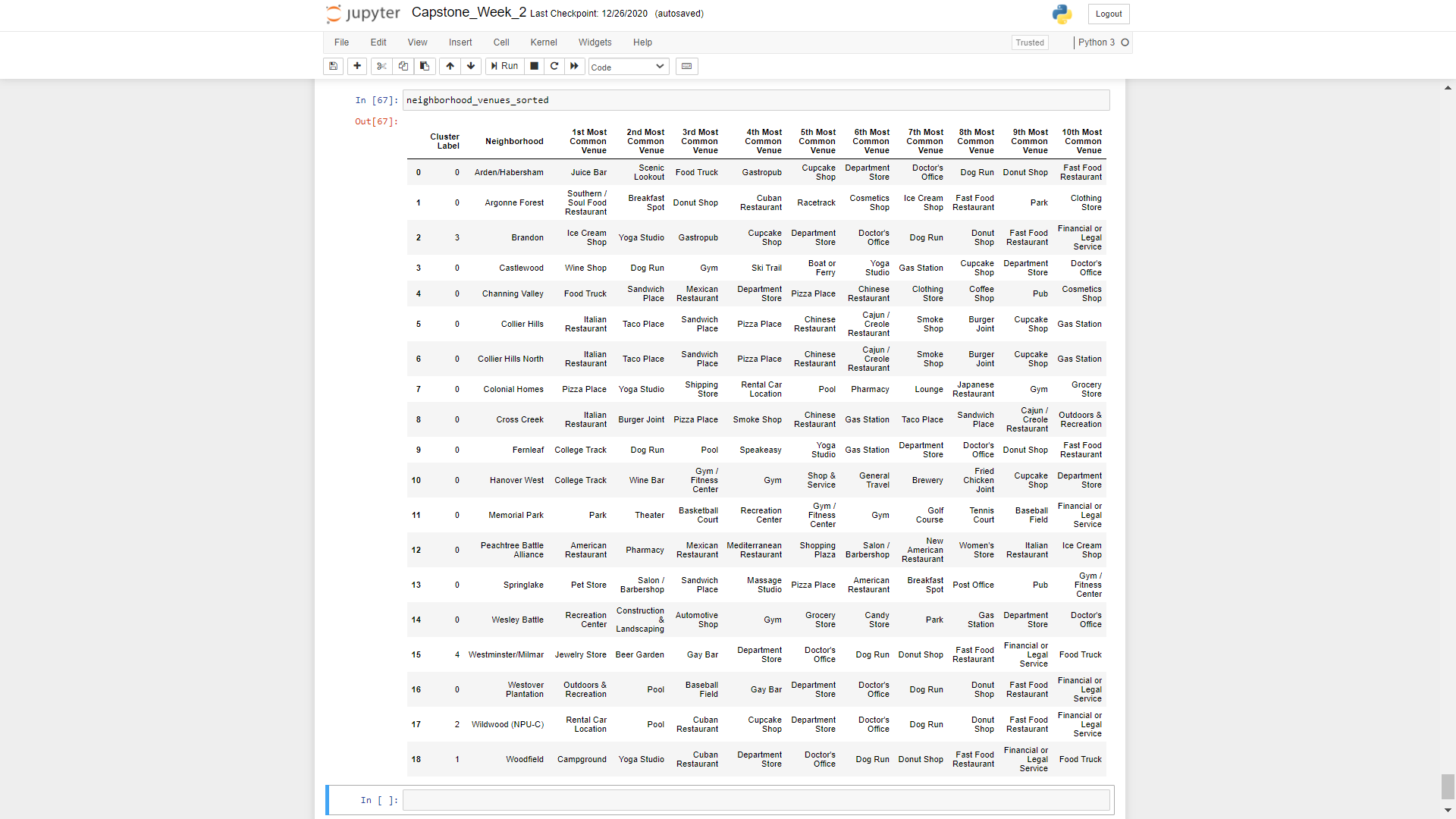
1 American Restaurant 0.0

2 Pharmacy 0.0

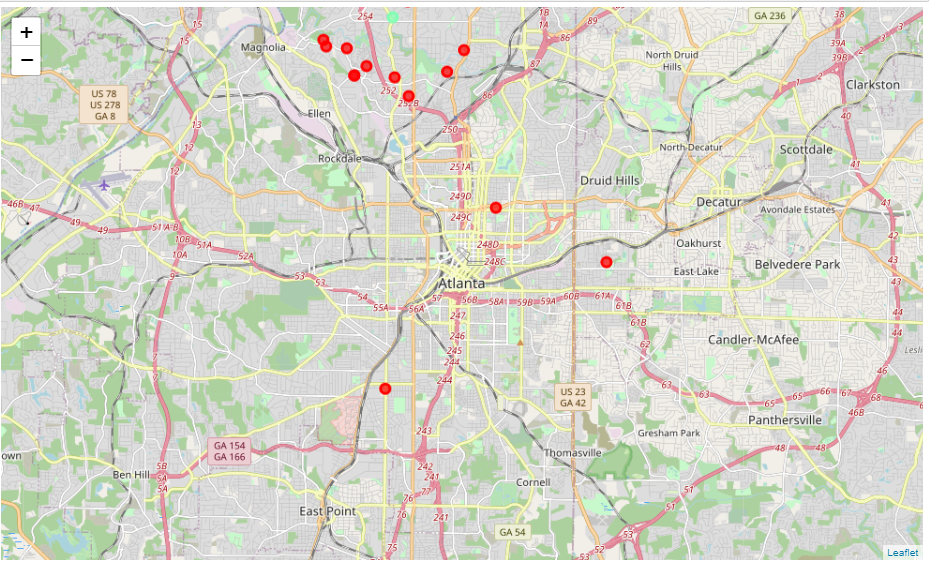
3 Residential Building (Apartment / Condo) 0.0

4 Rental Car Location 0.0

Top 10 venues in NPU C neighborhoods.



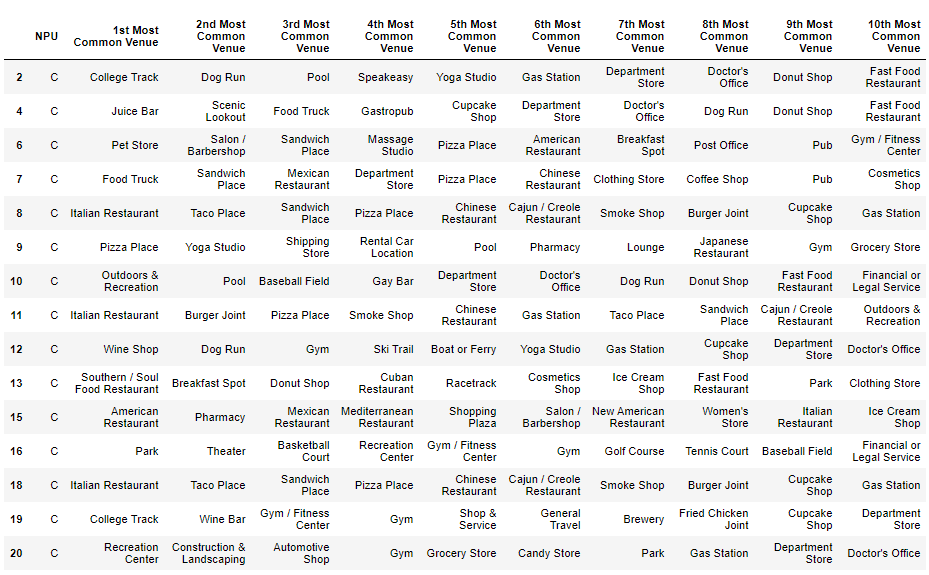
Now using K means Clustering using K=5, create clusters for neighborhoods in NPU C



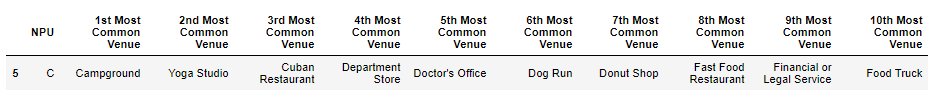
# Results

Examining the cluster results

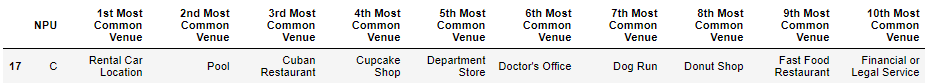
## Cluster 1



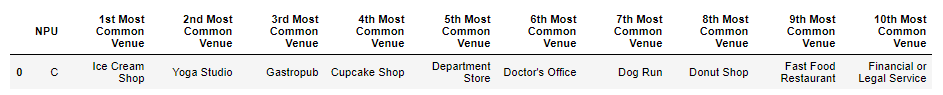
## Cluster 2



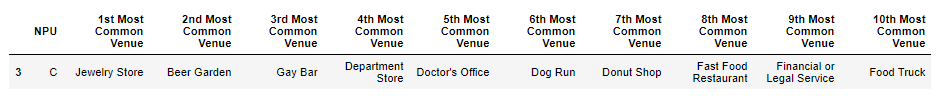
## Cluster 3



## Cluster 4



## Cluster 5



# Discussion

The objective of the analysis was to help small entrepreneurs to make data based informed decisions. The analysis here provides several insights some of the key aspects, but specifically identifying NPU in the following order of priority

1. Lowest crime to ensure safety of business and customers.
2. Neighborhood with median income over $65000/year to cater premium goods and services.
3. Highest population density in neighborhoods with lowest crime and Income to ensure business revenue.

Neighborhood venues were then identified from Foursquare API and other data sources.

K-Means clustering further enabled choosing the right venue based on most and least common venues in the area.

This information can be used to identify in which neighborhood and which venue, the business can be the most successful.

# Conclusion

Through the analysis, insights were offered into the city to help small entrepreneurs with recommendation as to which business will be most successful in the safest neighborhood based on income, population and crime criteria. The report was presented with exploratory data analysis, then picking the ideal neighborhood using machine learning techniques. Several business options are available, a specific venue can be decided on least common venue available from the clusters presented.

# References:

Atlanta Police Department (APD)(2020). Crime Statistics- Crime Data Downloads [Data Set]. APD. https://www.atlantapd.org/i-want-to/crime-data-downloads

City of Atlanta- Department of City Planning (2020) *NPU* [Data Set]. City of Atlanta. https://dpcd-coaplangis.opendata.arcgis.com/datasets/npu

City of Atlanta (2020). *Office of Zoning and Development- Population - 2017 Census Estimates by NPU.* City of Atlanta. https://www.atlantaga.gov/government/departments/city-planning/office-of-zoning-development/population

Otar, C. (2018, October 25). *What Percentage Of Small Businesses Fail -- And How Can You Avoid Being One Of Them?.* Forbes. https://www.forbes.com/sites/forbesfinancecouncil/2018/10/25/what-percentage-of-small-businesses-fail-and-how-can-you-avoid-being-one-of-them/?sh=1698a8a543b5