



A New Restaurant

Finding the right neighbourhood in Calgary, Alberta

Manny Gill

Location, Location, Location!

The right neighbourhood is a large part of success

- A large set of factors can define the success of a restaurant.
- Increased foot traffic, surrounding establishments and income profiles of communities can be part of a winning formula.
- Understanding the characteristics of communities and using data we can recommend a potential location in Calgary, Alberta, Canada.
- The information can help communities attract new restaurateurs to their communities, current restaurant owners, investors and governments.

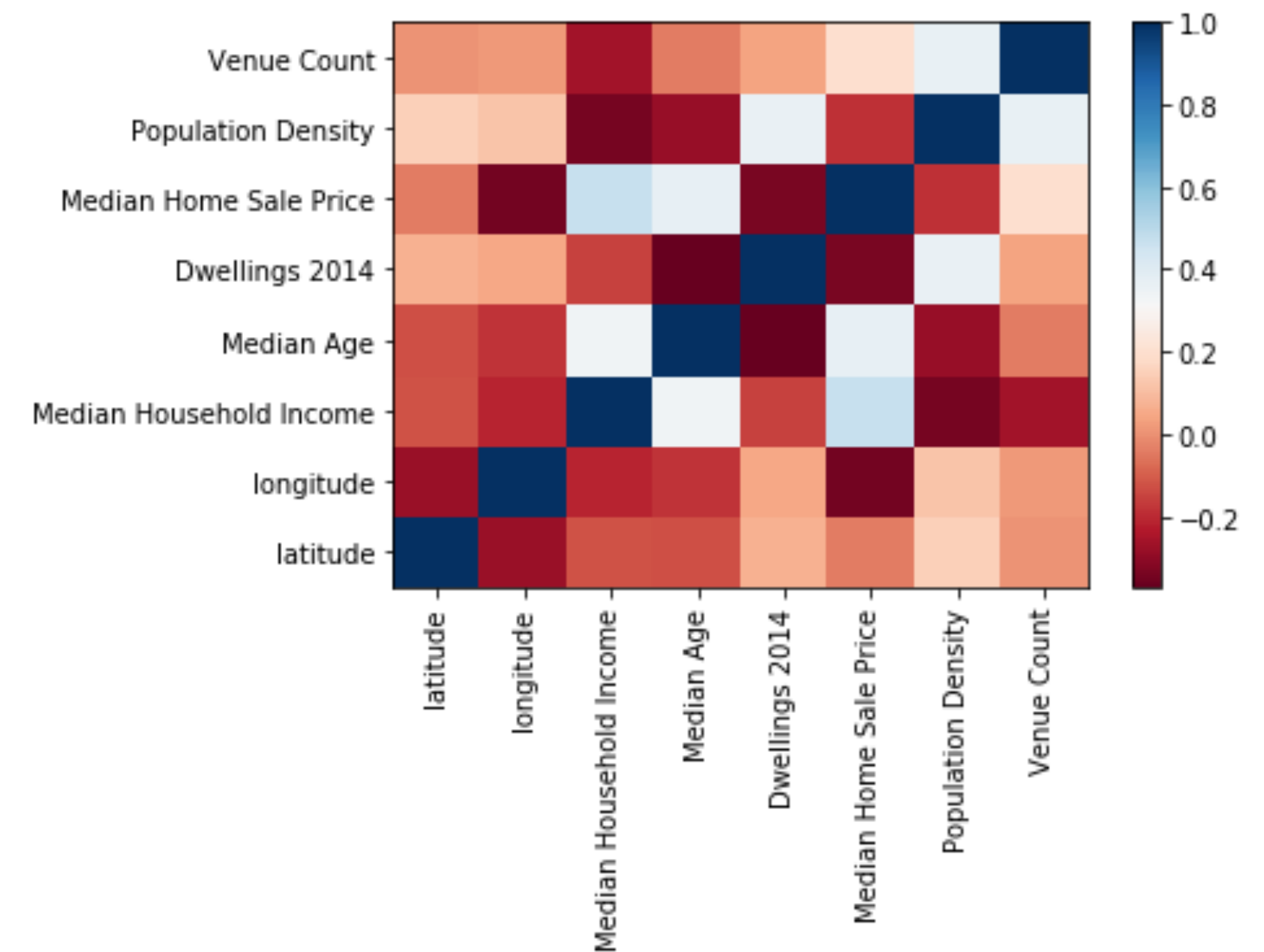
Data Acquisition and Cleaning

- Wikipedia(https://en.wikipedia.org/wiki/List_of_neighbourhoods_in_Calgary) : Using the data we will be able to get the neighbourhoods, population and population density.
- Foursquare API: Using the information from foursquare we will get the venues and restaurants for the neighbourhoods.
- Income Data: Data from Great News (<https://great-news.ca/demographics/>) and assume the information is representative of Venue Count in Calgary.
- Data was cleaned and missing or incomplete data accounted for.

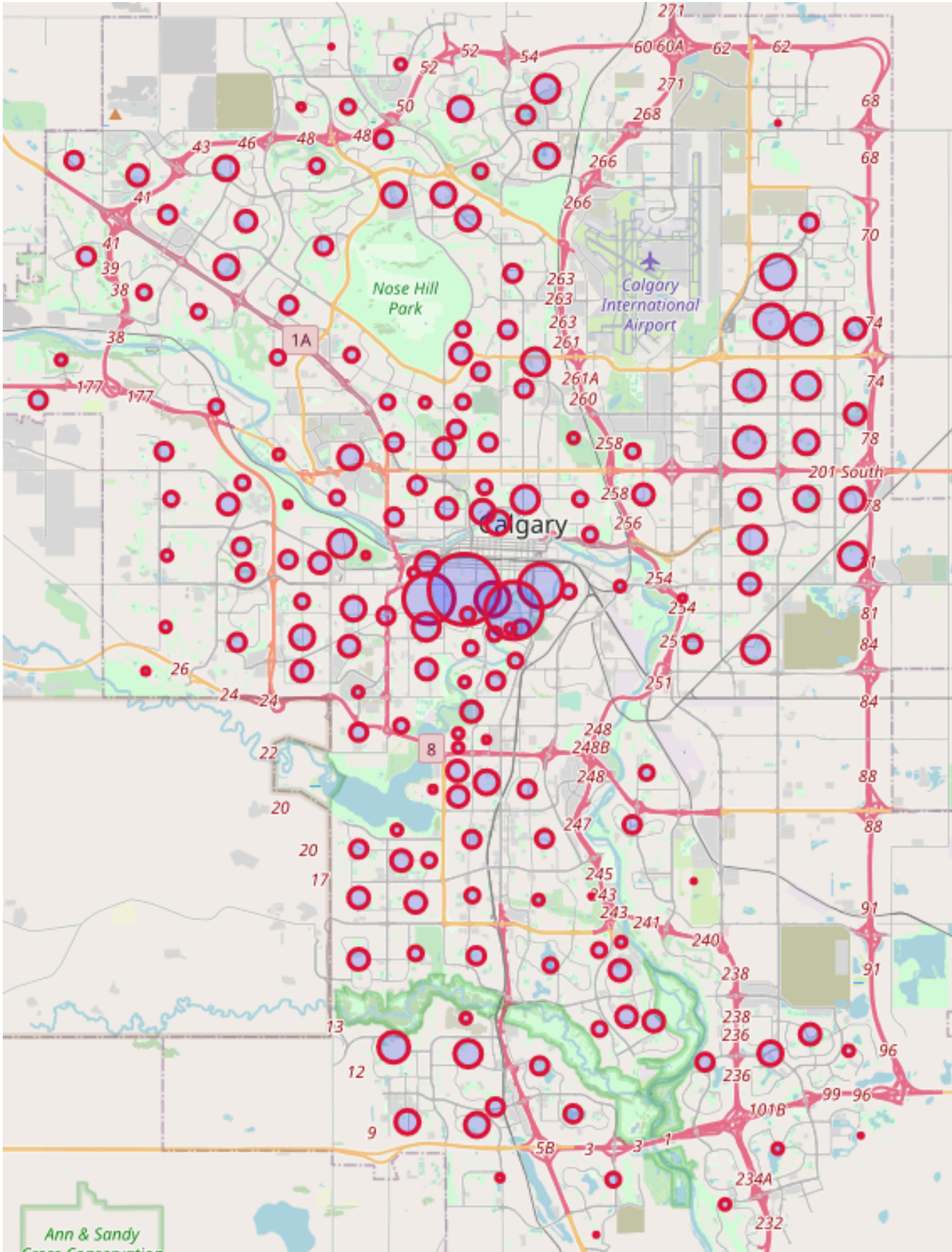
Exploring the Data

Becoming more familiar with the data

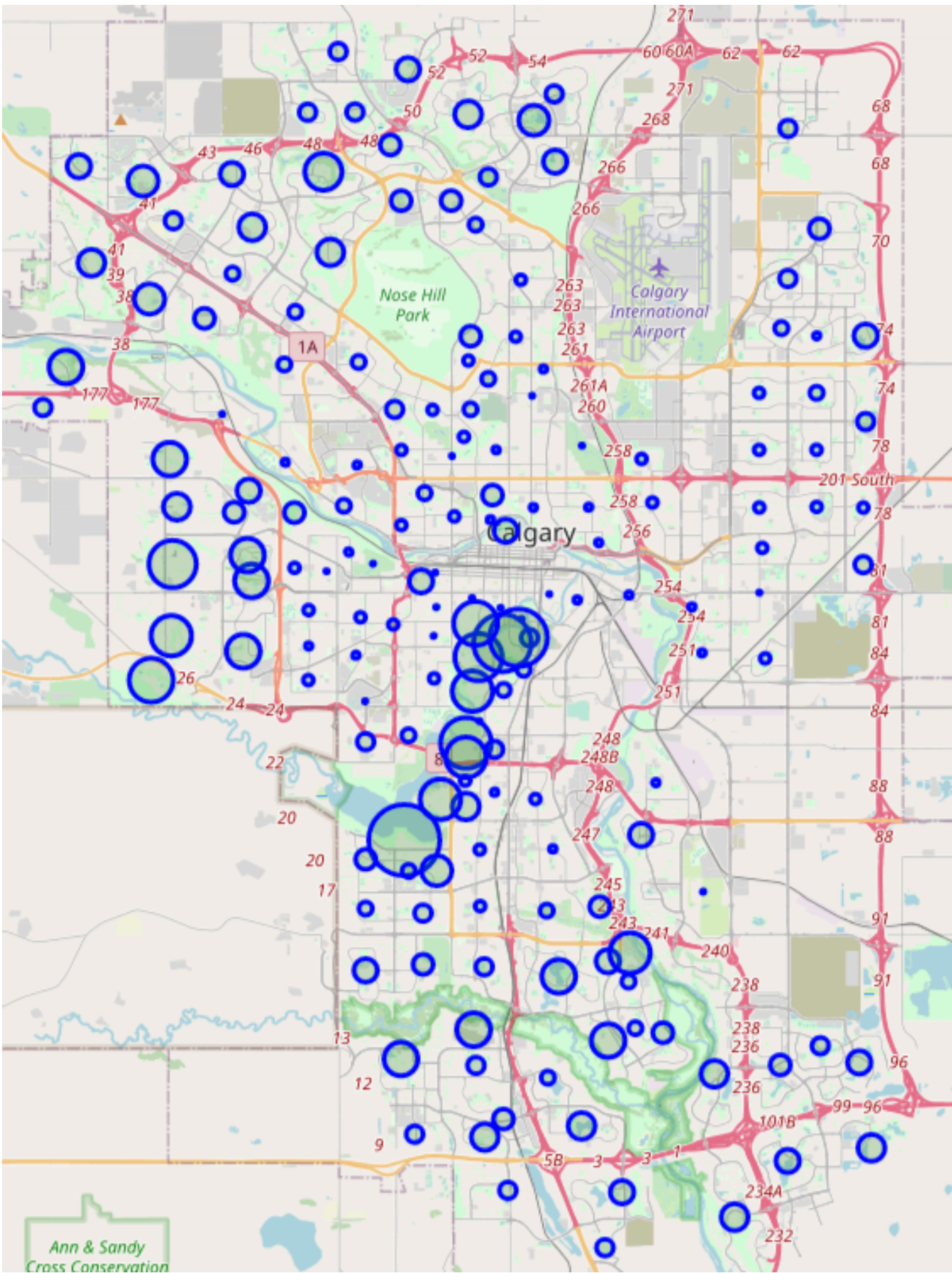
- Correlation studies, mapping the data, and viewing scatter plots helped to get more familiar with the data.
- Given the venue count is a target value as we looking to increase by adding a restaurant to a given community.



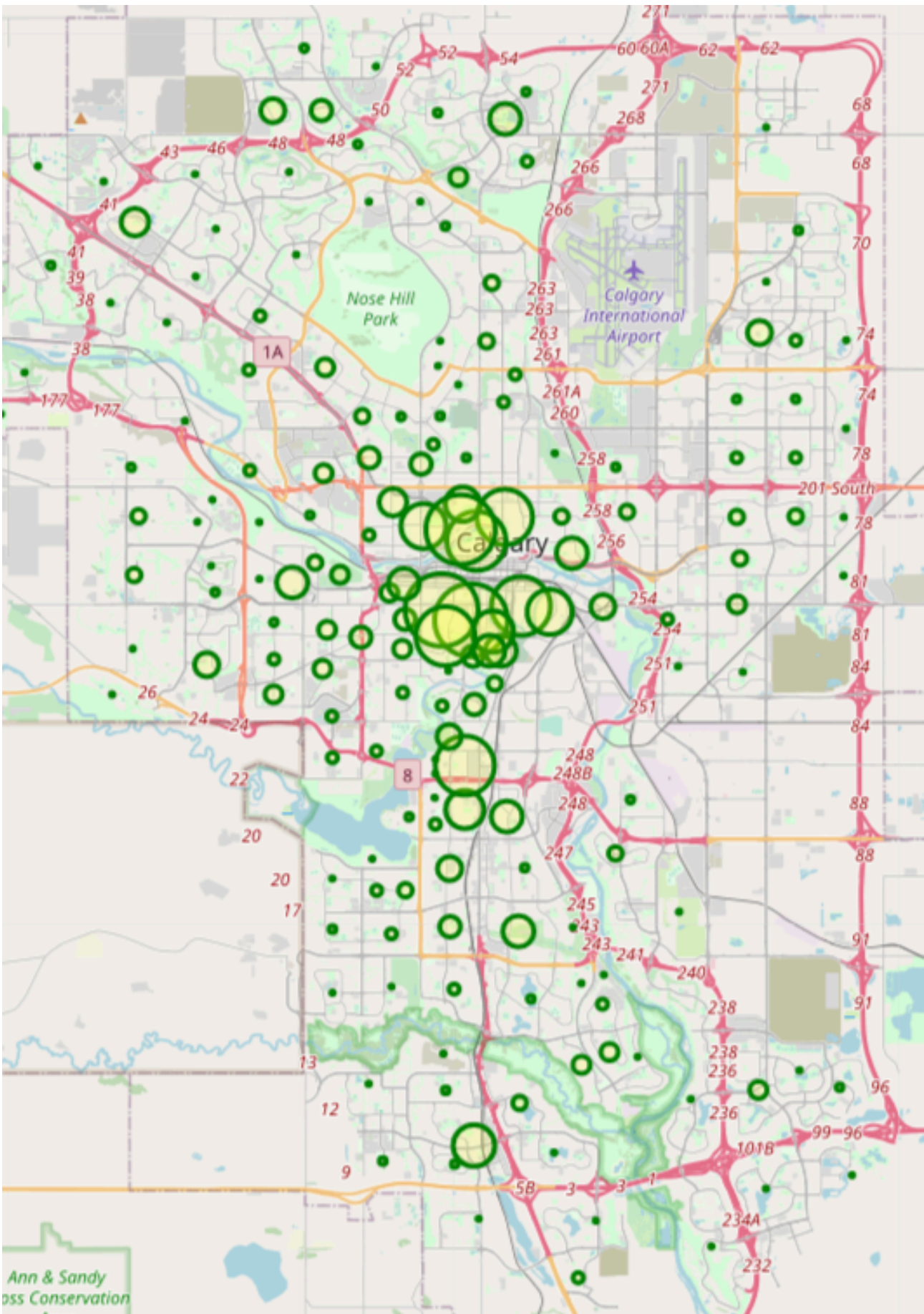
Population Density



Median Household Income

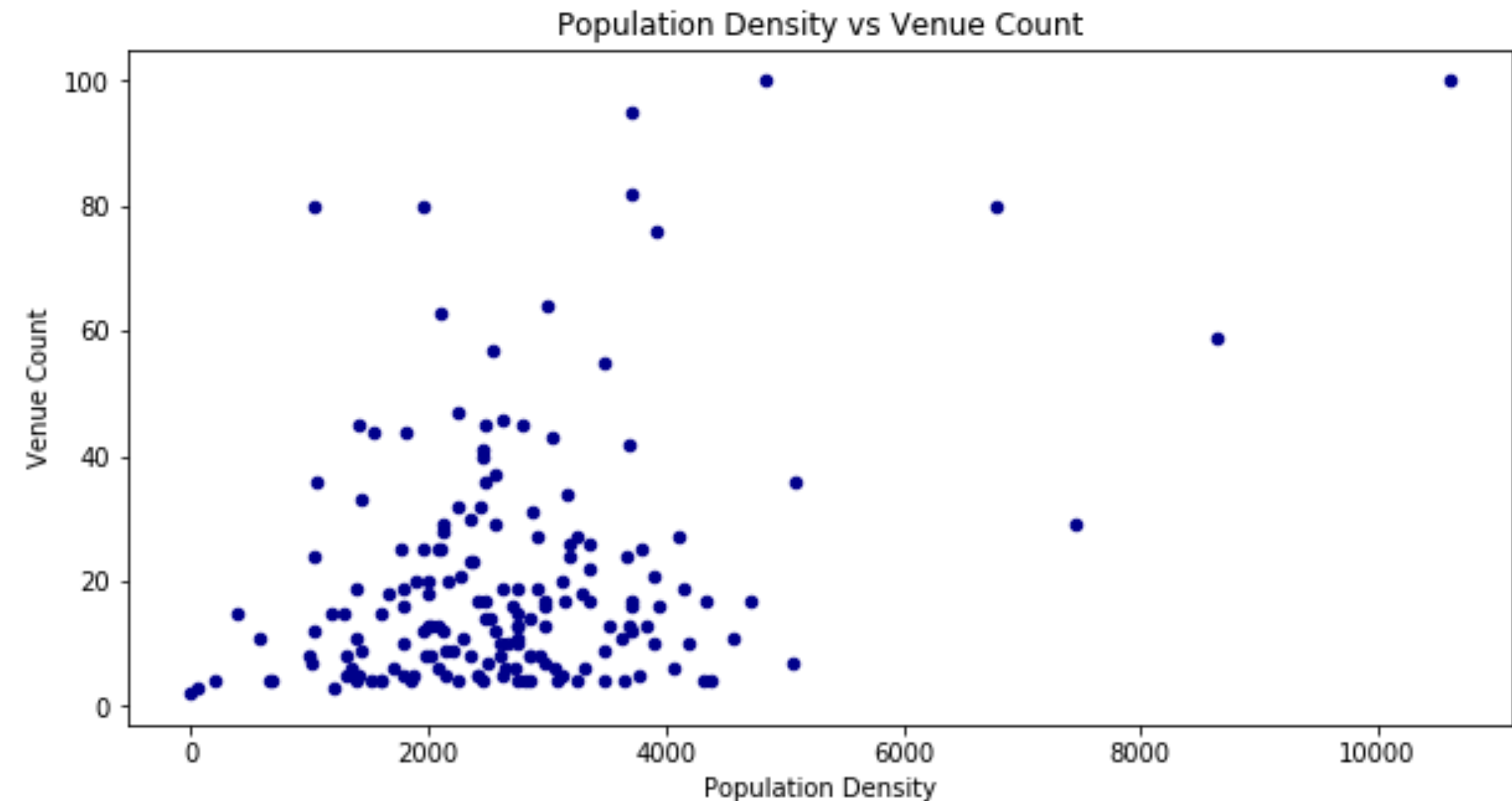


Venue Count Density



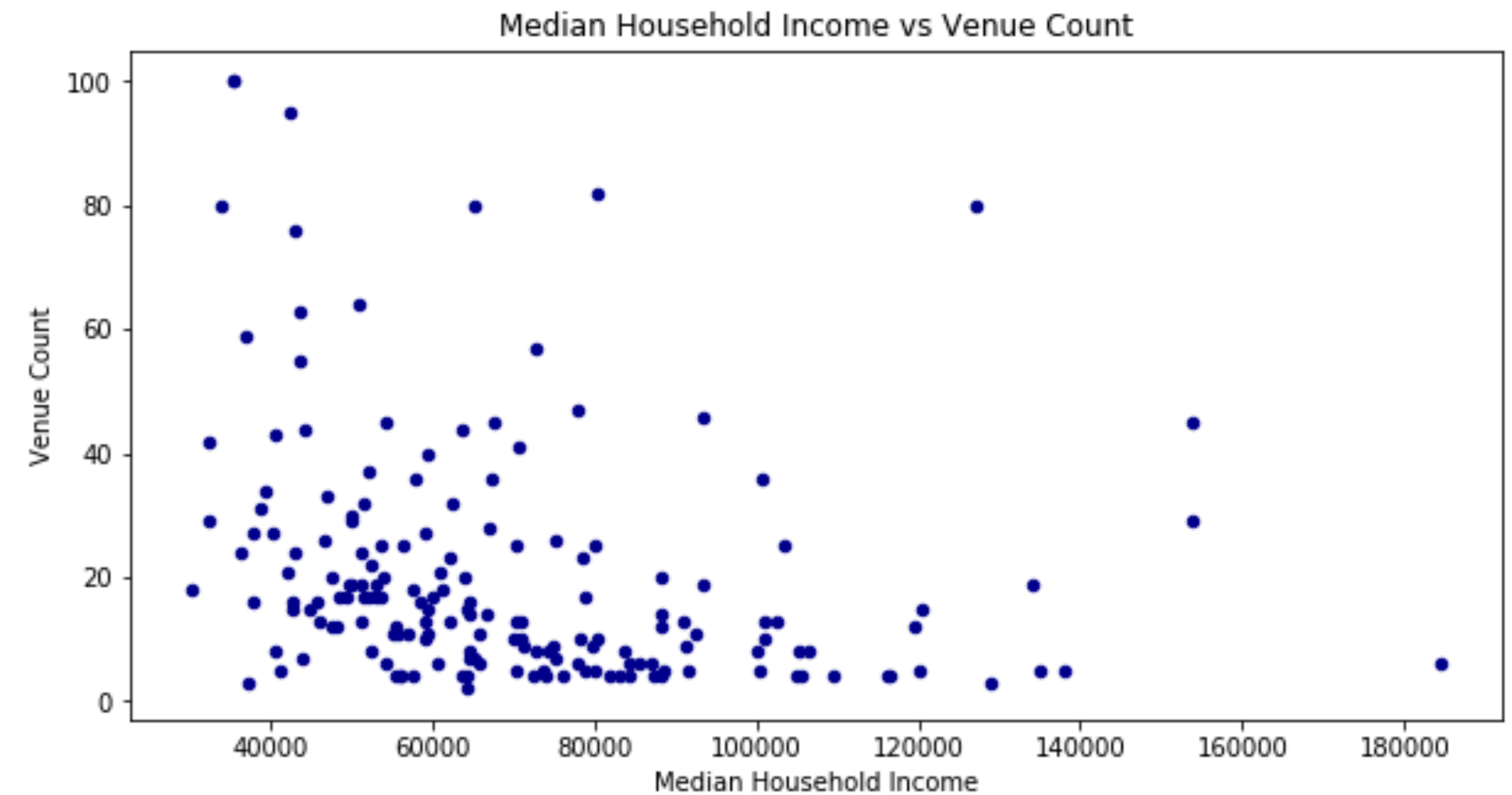
Population Density vs Venue Count

- The Pearson Correlation Coefficient is 0.3688238884606182 with a P-value of $P = 5.507276499715131e-07$ For Population Density vs. Venue Count
- Communities with shopping centres (Malls) we can see a large jump in a venue count.



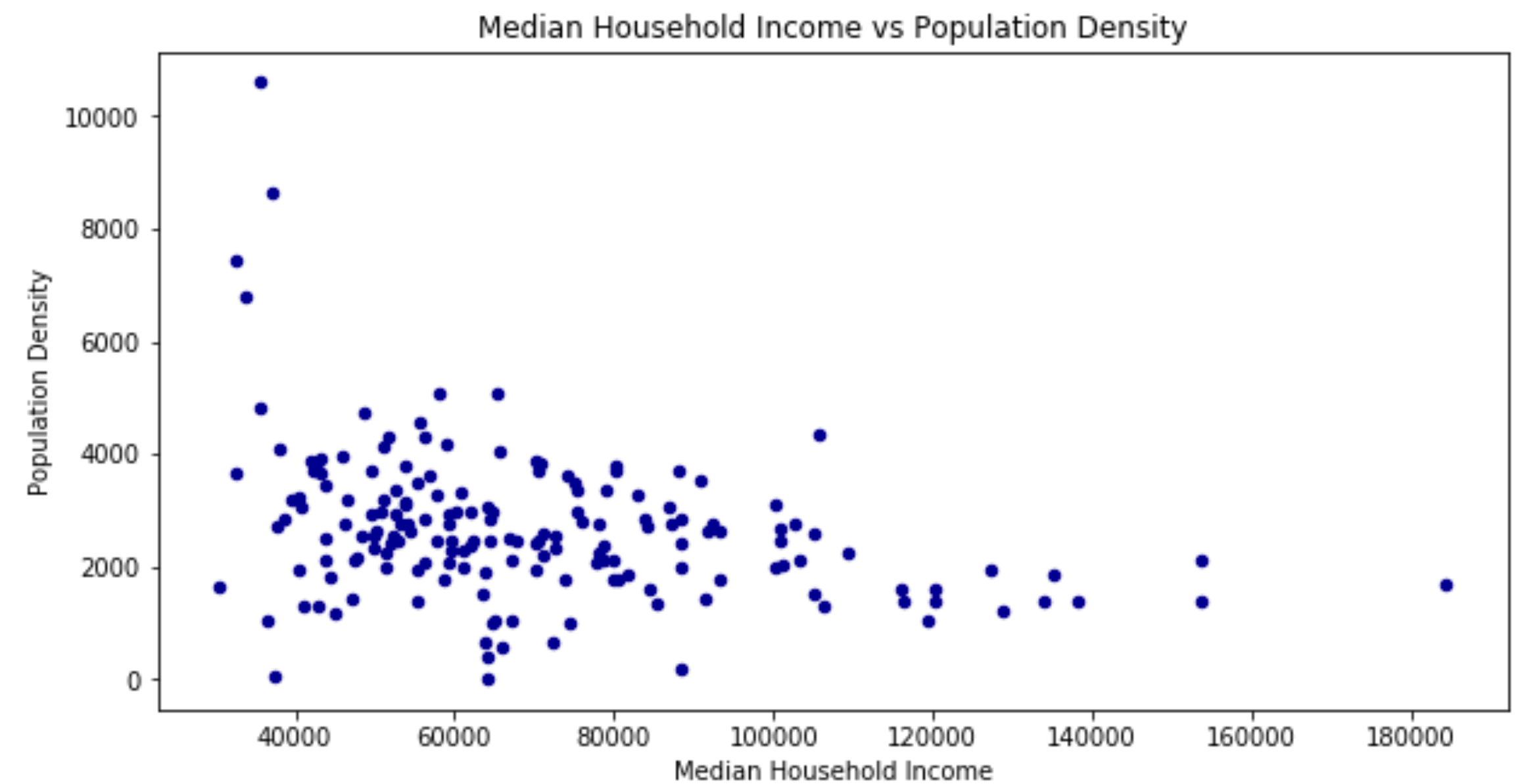
Median Household Income vs Venue Count

- The Pearson Correlation Coefficient is -0.26250677026798164 with a P-value of $P = 0.00046645289590050543$ For Median Household Income vs. Venue Count
- Median household income increases the venue count goes down



Median Household Income vs Population Density

- The Pearson Correlation Coefficient is -0.34131725946078717 with a P-value of $P = 4.046037043666337e-06$ For Population Density vs. Median Household Income
- Median household income increases the venue count generally goes down



Using DBSCAN to Classify Neighbourhoods

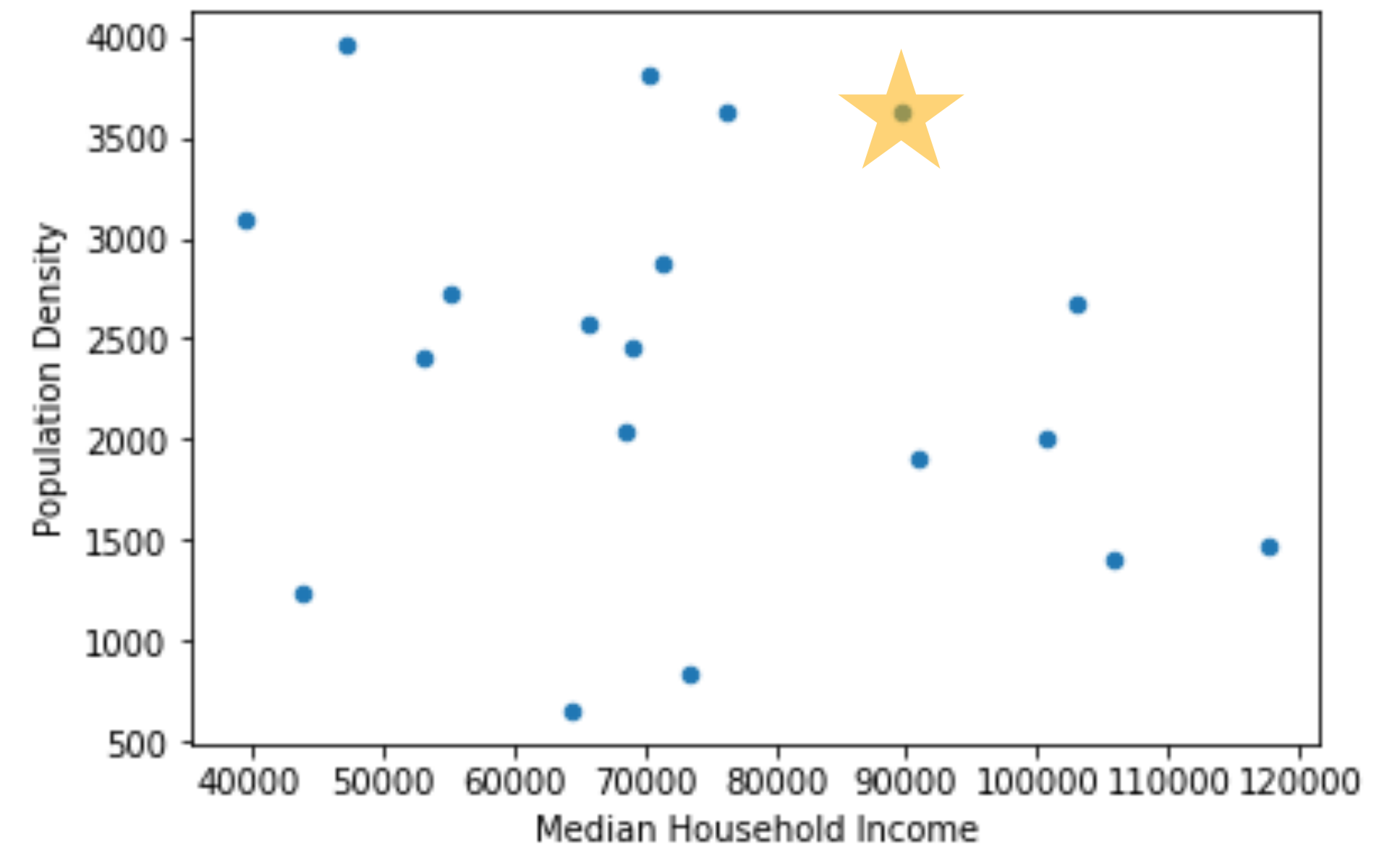
- Classify the communities with the given median household income, population density, and venue count into clusters

	latitude	longitude	Median Household Income	Median Age	Dwellings 2014	Median Home Sale Price	Population Density	Venue Count	Clus_Db
count	174.000000	174.000000	174.000000	174.000000	174.000000	1.740000e+02	174.000000	174.000000	174.000000
mean	51.034690	-114.081681	69529.178161	38.896552	2585.954023	5.976602e+05	2669.850575	20.678161	2.057471
std	0.074565	0.075531	26484.943723	5.823084	1853.100881	2.283349e+05	1340.214248	19.908167	4.890374
min	50.857617	-114.260486	30411.000000	29.000000	110.000000	3.050000e+05	0.000000	2.000000	-1.000000
25%	50.985098	-114.129630	51258.250000	34.250000	1153.750000	4.402500e+05	1956.750000	7.000000	-1.000000
50%	51.041900	-114.085021	64164.000000	38.000000	2402.000000	5.480000e+05	2544.500000	14.000000	0.000000
75%	51.084942	-114.036100	81422.500000	42.000000	3516.500000	6.876250e+05	3238.250000	25.000000	2.750000
max	51.175690	-113.924528	184383.000000	58.000000	15243.000000	1.701000e+06	10600.000000	100.000000	17.000000

	Median Household Income	Population Density	Venue Count
Clus_Db			
-1	71371.239437	2871.028169	30.690141
0	65774.925926	2569.425926	11.814815
1	69144.000000	2465.000000	43.000000
2	70481.000000	3806.666667	14.000000
3	39497.000000	3098.333333	31.000000
4	102903.666667	2673.000000	10.000000
5	76377.333333	3630.333333	6.333333
6	68637.500000	2041.000000	26.500000
7	117599.000000	1465.000000	4.666667
8	47176.666667	3956.333333	17.500000
9	53281.500000	2408.500000	34.750000
10	55145.500000	2722.750000	5.500000
11	100590.500000	2012.500000	10.000000
12	73366.000000	840.500000	6.500000
13	43851.500000	1236.500000	15.000000
14	89504.500000	3622.000000	12.500000
15	90844.000000	1899.500000	21.000000
16	64582.250000	655.500000	9.000000
17	105663.000000	1411.500000	6.000000

The Results

- Venues with average venue count are a good balance between optimal competition as well as gaining traffic to the restaurant.
- Higher the Median household income the more spending power the household has
- Panorama Hills: Population Density: 3,531, Median Household Income: \$90,851, Venue Count: 13



Conclusion and Future Discussion

- Once a new restaurant would be added to the neighbourhood we can see that venue count would be in the 50% percentage value of venue count per neighbourhood.
- Establish the bases for larger study in opening a new restaurant. Additional factors in establishing a successful restaurant.
- More data to take into consideration can include the leasing costs, restaurant types, demographics, established cultural district and so on.