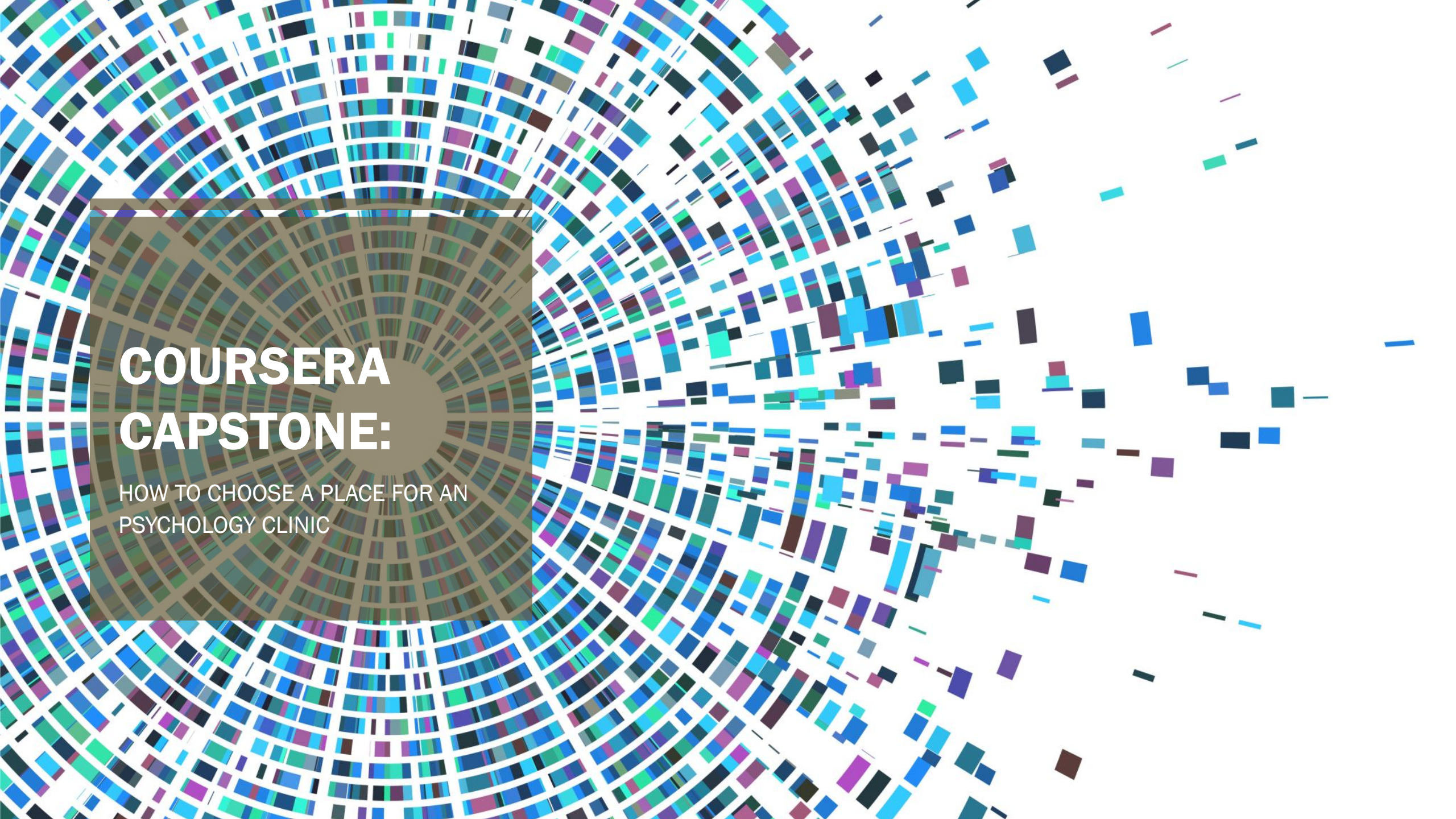


COURSERA CAPSTONE:

HOW TO CHOOSE A PLACE FOR AN PSYCHOLOGY CLINIC

INTRODUCTION

- Almost every business depends on location. At first, the location has an immediate effect on fixed costs and it can also be very important for variable earnings. In this problem, it will be chosen as the best location for a Psychology Clinic. The main criteria adopted are the relation between some variables such as the number of banks, hospitals, and the number of results of 'psy' results in the Foursquare database. Also some analysis will be made based on property evaluation of Rent Cost per SqFt and Market Value.
- Customer C is a psychologist who is also owner of one clinic of psychology in New York. C wants to know what areas of New York are good to install your clinics. Right now C has only one clinic in Manhattan but C wants to open new clinics in other regions too.

BUSINESS PROBLEM

- C intends to scale your business and also to dilute some fixed costs by using the same client target for all your clinics in New York.
-
- C intends to use similar as possible furniture, paints and customer psychology challengers. C believes that beyond one-by-one therapy, group therapy is also a great tool to providing to your clients better health and quality of life, so it is interesting to deal with clients of similar backgrounds and interests.
-
- In C experience as clinical psychology, there is a strong correlation between customer psychology profile and house/work neighborhood area. It's known by the client experience that Manhattan is a great place but frequently over too overpriced. So, the client wants to know other areas that are similar to Manhattan that should be also investigated.

METHODOLOGY: PROBLEM SOLUTION FRAMEWORK

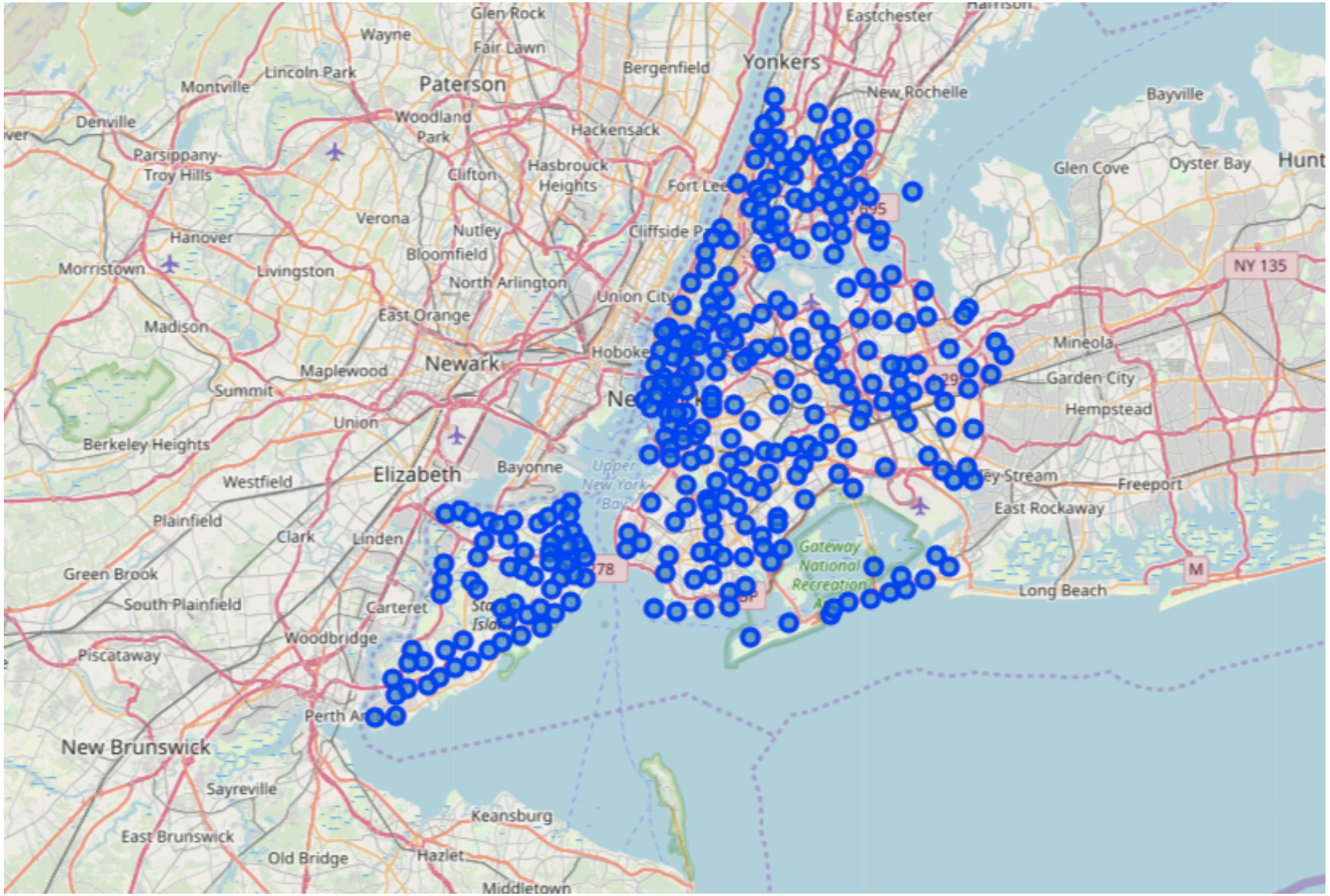
- C needs customer clusterization. As C is pretty sure about psychology profile and neighborhood, the first approach is definitely to try cluster neighbors in New York.
-
- It's necessary more information about his actual clinic in Manhattan:
- - C said in his actual clinic there are a lot of psychologist clinics, hospitals, and banks: it should be examined as evidence of good places;
- - it is supposed that the correlation is strong between place and psychological profile;
- - C wants to know which are the potencial neighborhoods; and
- - C wants a recommendation of potencial property with cost analysis.
-

DATA REQUERIMENTS AND APPROACH

- It will be used data from Foursquare API and New York Open Data of all properties. The steps admitted to solving the clusterization are:
-
- - Read New York Json file;
- - Add Latitude and Longitude information by Borough and Neighborhood;
- - Data Extract by Foursquare's API;
- - Find data about New York Properties;
- - Analyze correlation between Market Value and Rent Cost; and
- - Generate a list of Properties Recommendation.

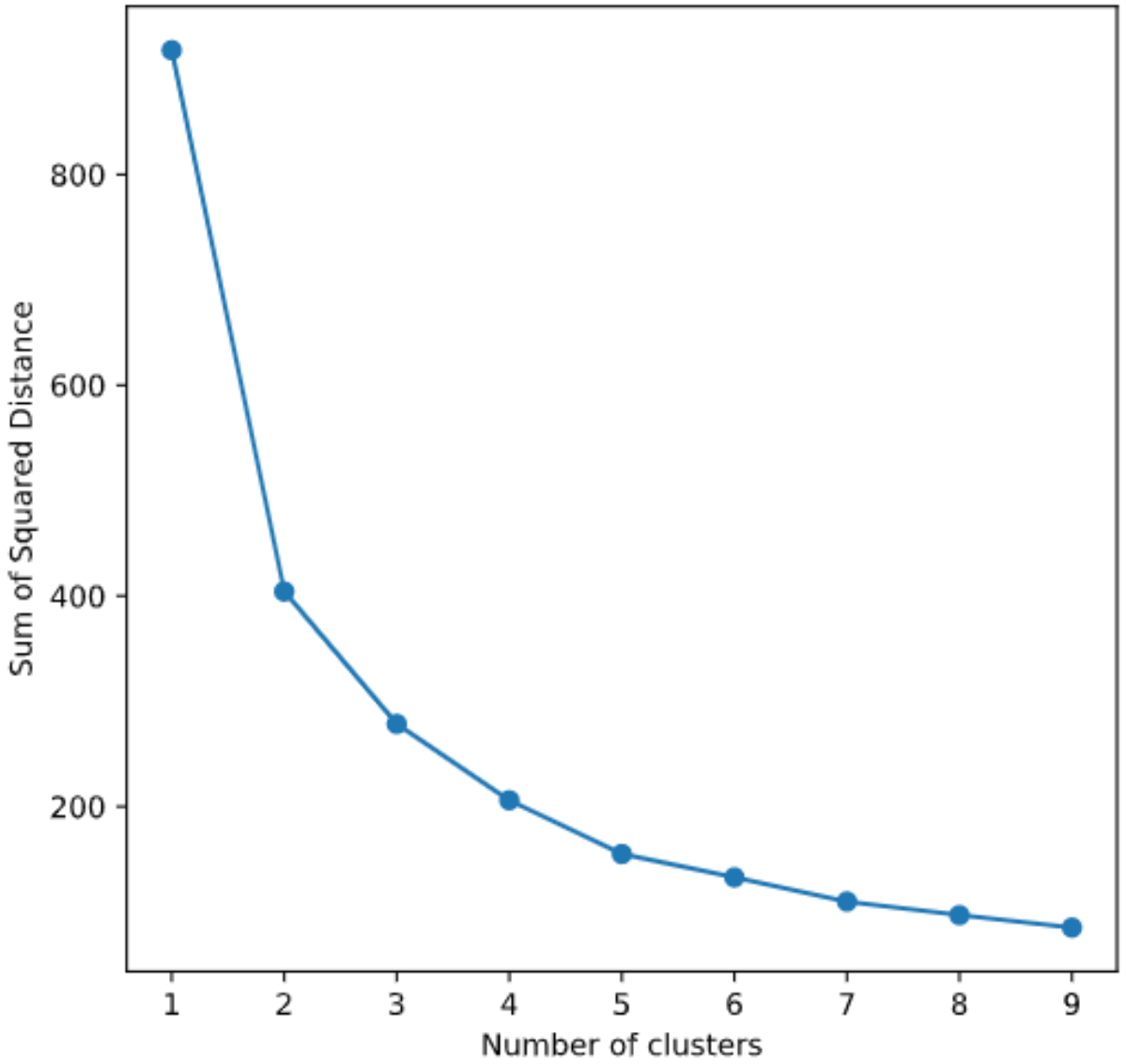
ANALYSIS: PROBLEM SOLVING

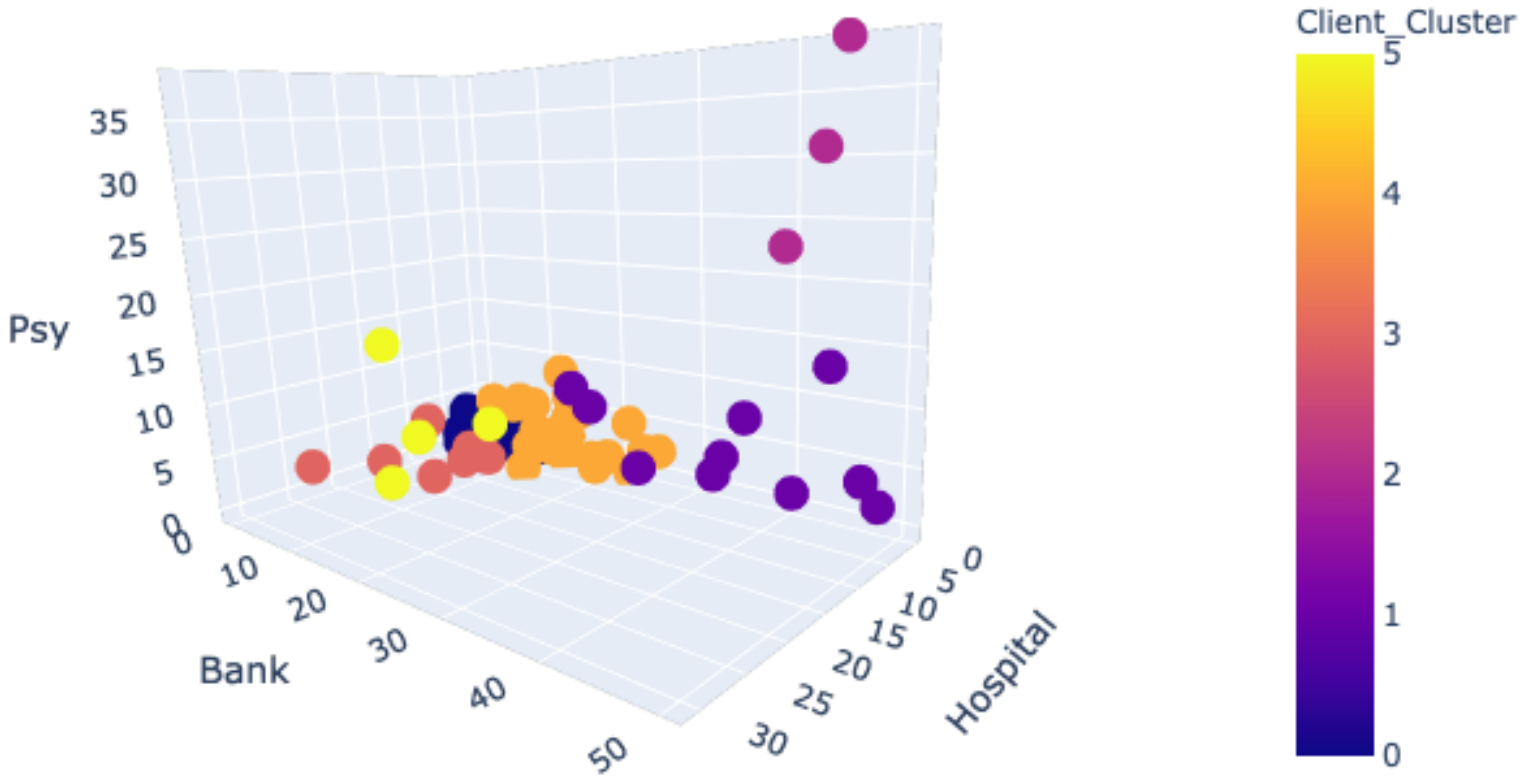
- At first, k-means will be used to cluster regions of New York based on venues amount, because it has a strong relation with psychology profile as suggested by C. To perform that it will be observed a strict radius to minimize the incidence of the same results for different searches. So, k-means will clusterize psychology profile of potencial clients in New York.
-
- After clusterization, the main cluster will be evaluated based on Rent Cost by Square Feet and Market Value. Properties with great Market Value is good, because it usually means better infrastructure, better location and more value companies arounded. Nevertheless, Market Value usually pulls Rent Cost by Square Feet which is bad.
-
- So, Properties with good clients cluster, high market value and low rent cost is better. At the end it is provided a list of properties recommendation in decedent order for C.
-

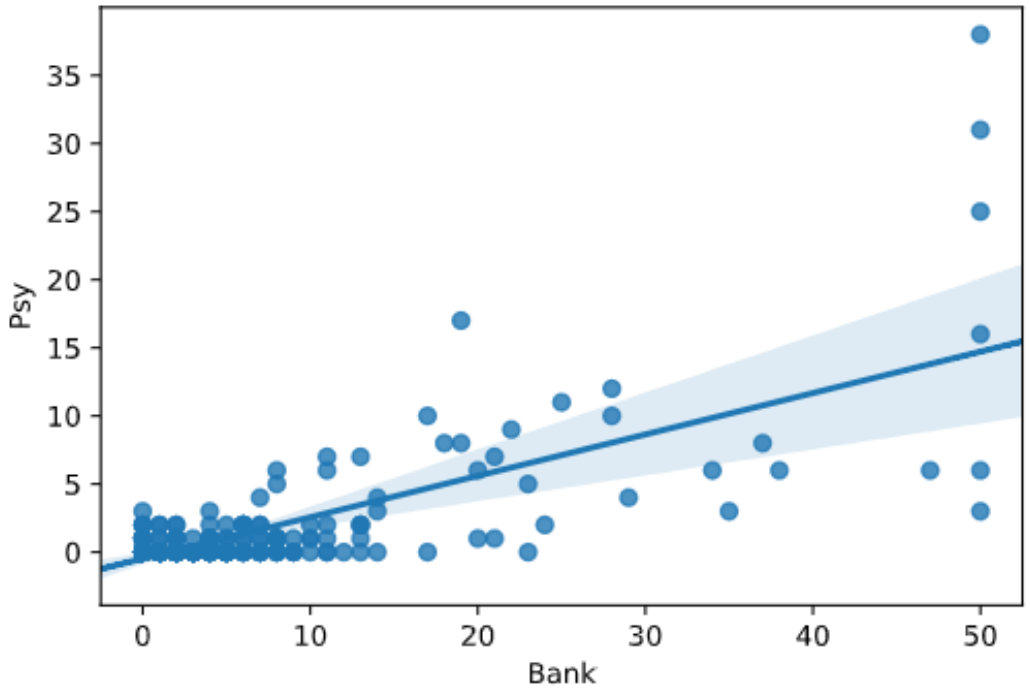

Glen Rock
Wayne
Fair Lawn
Eastchester
Yonkers
Bergenfield
Paterson
Lincoln Park
Montville
New Rochelle
Bayville
Denville
Woodland Park
Hackensack
Parsippany-Troy Hills
Hasbrouck Heights
Clifton
Glen Cove
Oyster Bay
Fort Lee
Verona
Nutley
Cliffside Park
Hanover
Bloomfield
Morristown
Livingston
North Arlington
Union City
NY 135
East Orange
Madison
Mineola
Hoboken
Newark
Maplewood
Garden City
Summit
Hempstead
Union
Berkeley Heights
Bayonne
Elizabeth
Westfield
Freeport
East Rockaway
Plainfield
Clark
Linden
Gateway National Recreation
Green Brook
Long Beach
South Plainfield
Carteret
Woodbridge
Piscataway
Perth Amboy
New Brunswick
Sayreville
Keansburg
East Brunswick
Old Bridge
Hazlet
Middletown

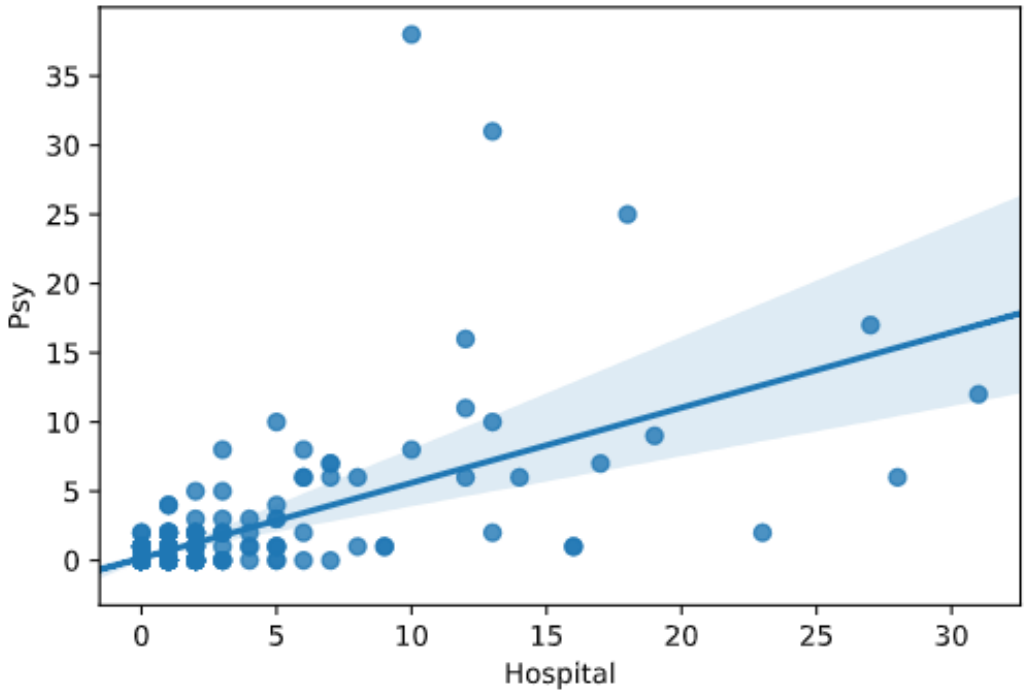
CLUSTERING REGIONS

- The clusterization are using three independent variables, it's highly desirable to seggregate all cube vertices (8). If we assume each axis are one variable, the cube represents the normalized variables. So, it is interesting use 8 clusters to analyze all cases separatelly. The image bellow illustrate a cube 2x2 where each vertice is a unique block.


Sum of Squared Distance
800
600
400
200
1
2
3
4
5
6
7
8
9
Number of clusters

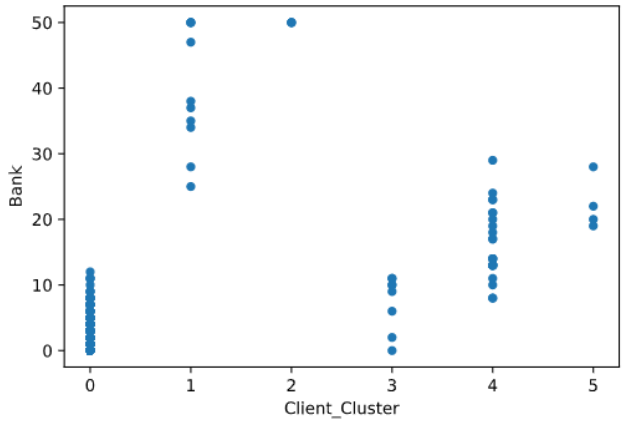

Psy
35
30
25
20
15
10
5
0
Bank
0
10
20
30
40
50
Hospital
0
5
10
15
20
25
30
Client_Cluster
5
4
3
2
1
0

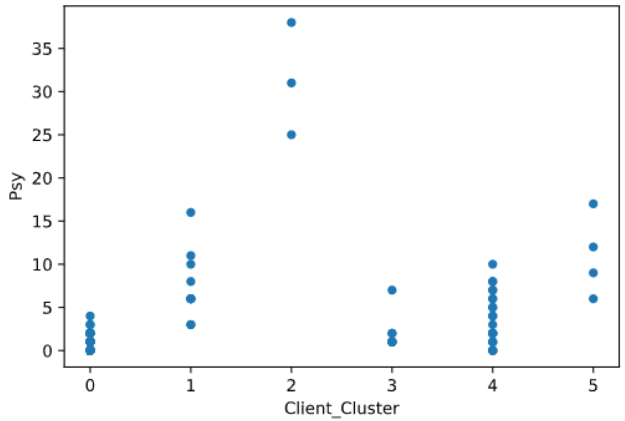


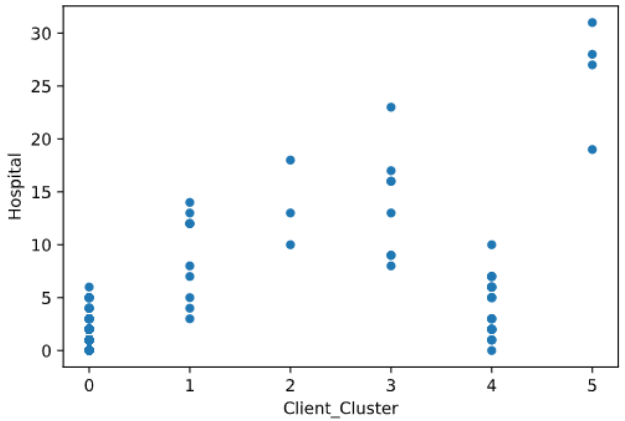


LINEAR REGRESSION

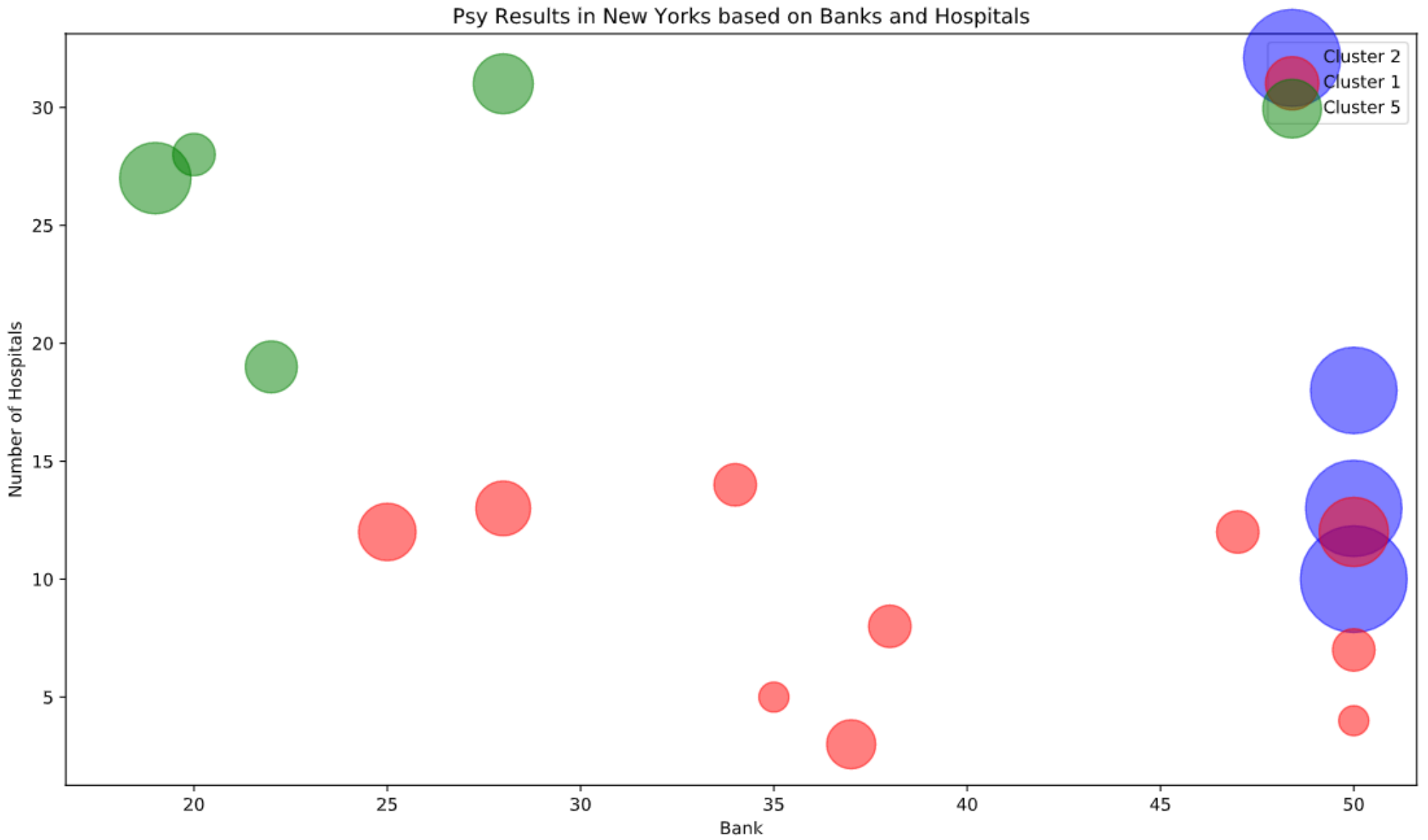
- Plot for Banks and Hospital relation in Psy Results





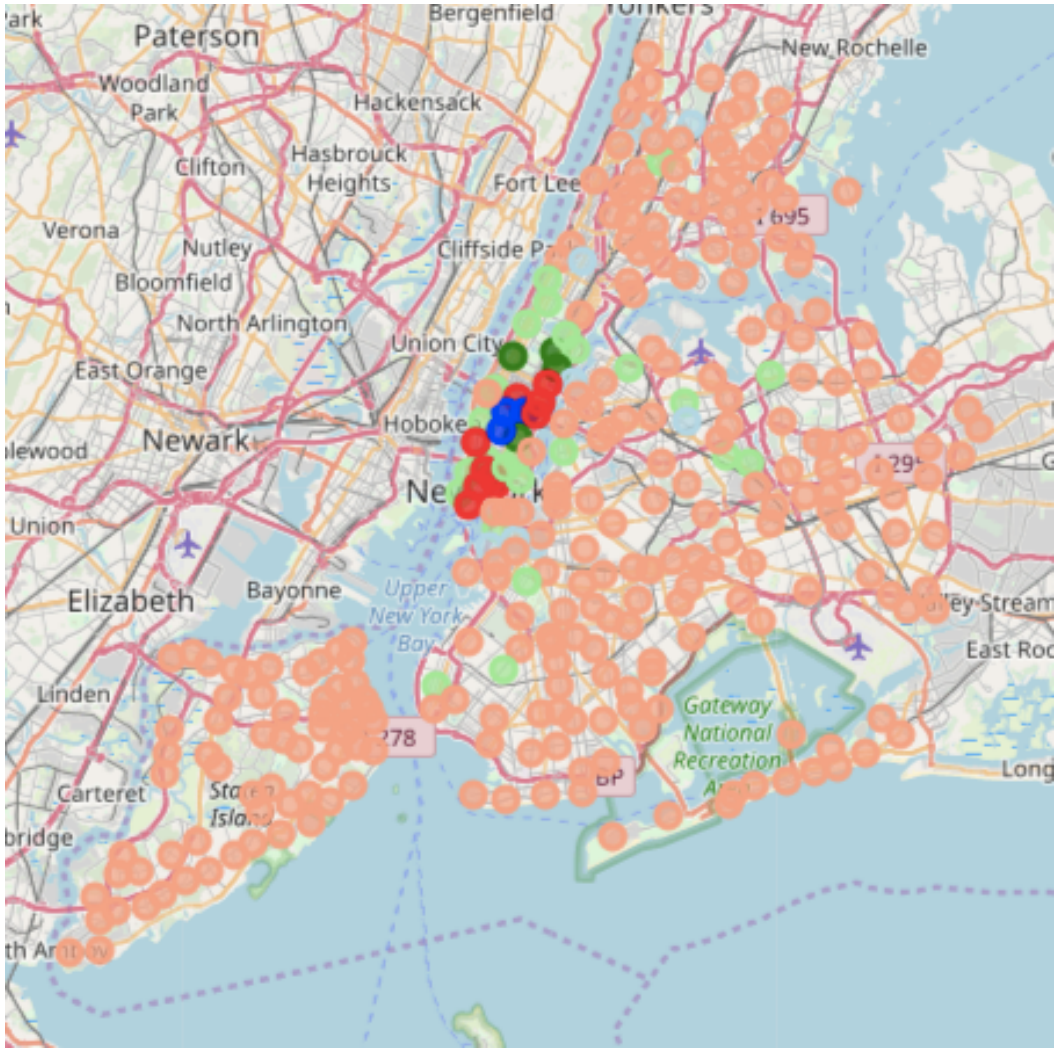


CLUSTER CHARACTERISTICS BASED ON VARIABLES

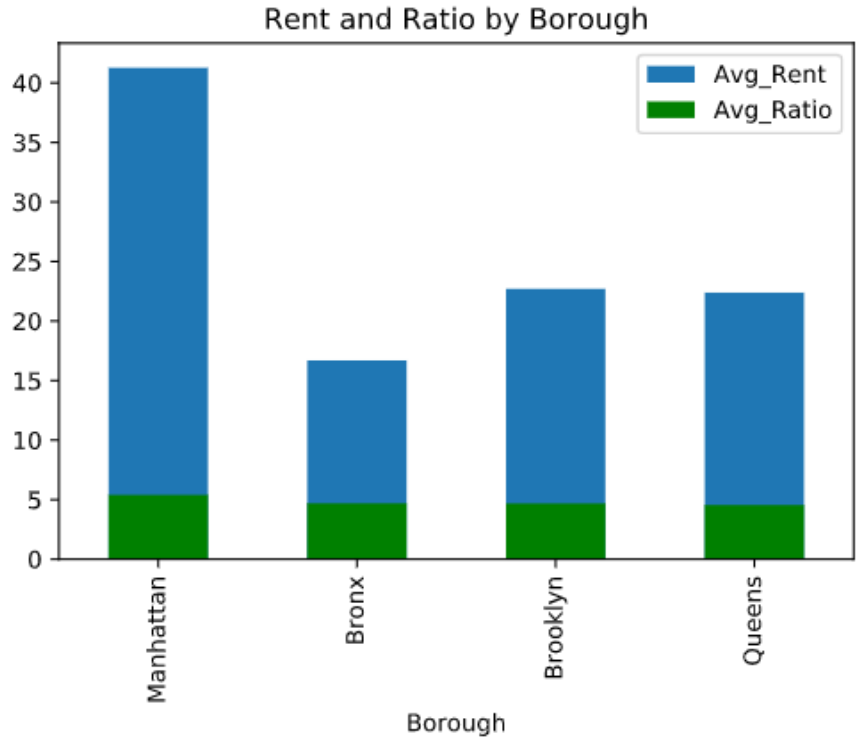

Psy Results in New Yorks based on Banks and Hospitals
Cluster 2
Cluster 1
Cluster 5
Number of Hospitals
30
25
20
15
10
5
20
25
30
35
40
45
50
Bank

NEW YORK MAP OF CLUSTER

- Clusters made based on variables psy results, hospitals and banks.



RENT AND RATION AVERAGES IN BOROUGH



CONCLUSION

- If Manhattan is a option: there is places inside Manhattan better than others based on criteria used;
- Manhattan is way more expensive than Bronx, Brooklyn and Queens;
- Maybe it is interesting add actual clinic location to add distance as parameter if Manhattan is an option for another clinic.
- A lot more can be done like add Sq Feet requirement, costs restrictions, maximize distance, et coetera.