Don't get stressed – Study on Finding a perfect spot to open a Yoga Studio in Manhattan

Applied Data Science Capstone Project

1. Introduction

Motivation

- In our stressful time, self awareness has become more and more important to get resilient and to recover from the daily challenges. In the last couple of years several strategies like sport , meditation and yoga has seen a risen request amongst people.
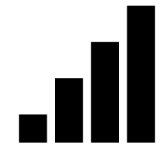
- Manhattan is one of the busiest and most stressful cities in the world. Recreation should always be of high demand. Since time is of an essence here as well. Long distances are not acceptable so the venues must be easily to reach from home.
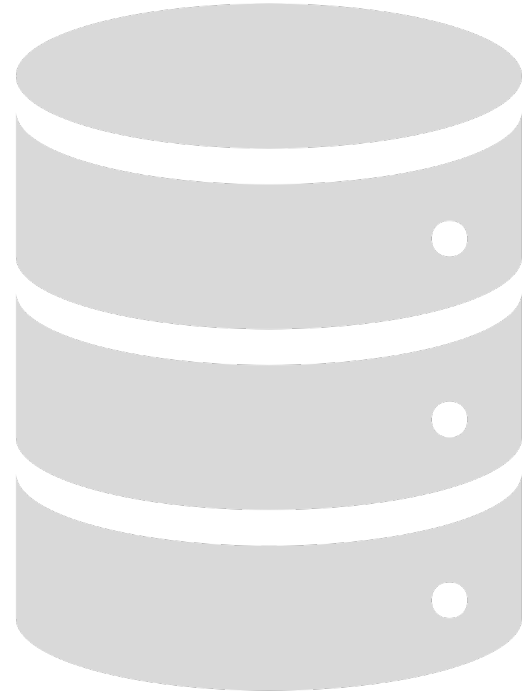
Business Case

- Our customer wants to open a new yoga venue in Manhattan. Special interest should be paid to the already available venues and population and income of the area.

- Basic researches on market size and price should also be conducted.
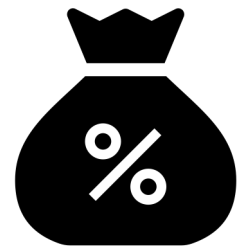
Data Source

Geographical Data

- By using the foursquare API, it will be possible to visualize the location of yoga studios in Manhattan and to identify a suitable place.

- A .json data was used to get the neighborhoods of Manhattan

- In a first step the geographical data was extraxted from a .json file and directed to a dataframe. We further narrowed the dataset to Manhatta as the place of interest.

- Since Yoga courses could not also held in Yoga studios but also in other sport venues, the search was broadened to gyms, sport clubs, gym, spa, etc. so the neighborhoods with a high density of this venues could be easily identfied.

Population and Financial Data

- In order to make first estimation on the market size, the population of each found neighborhood was analyzed and the average "Yoga Studio per inhabitant" will be calculated.

- For the commercial estimation the median income of the Neighborhood was assessed as well.

- Since neighborhoods are not an official classfication for areas in the US, the data for the population and median income was gathered manually from the www.wikipedia.com.

- The population data on wikipedia were gathered by the US census bureau in 2010.
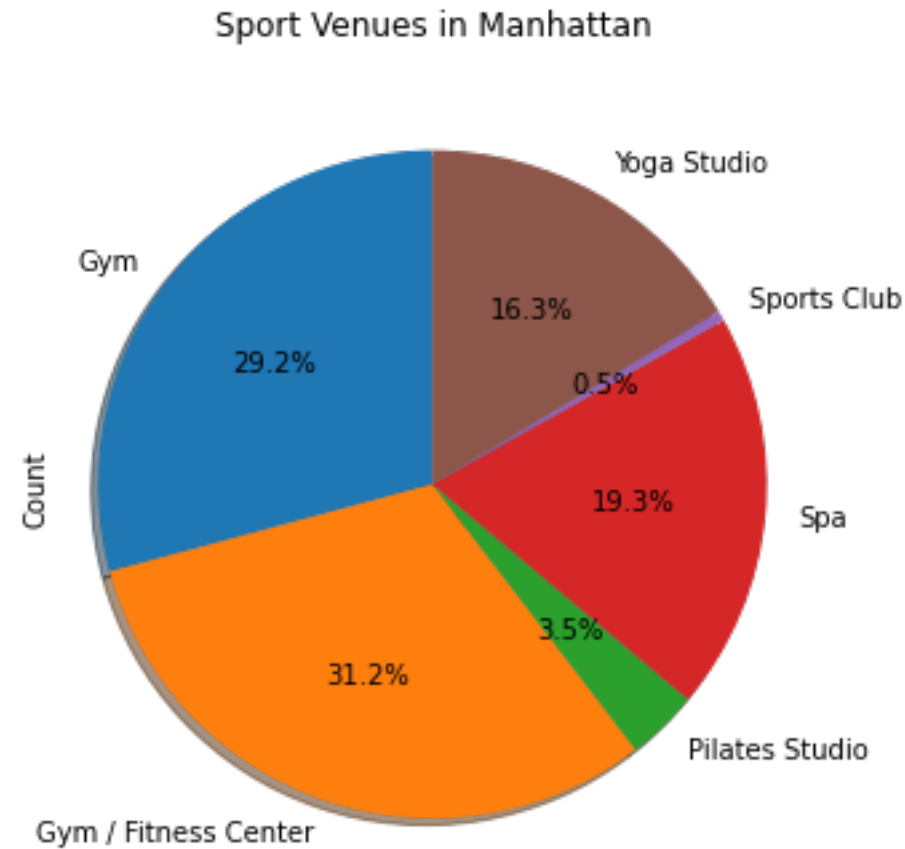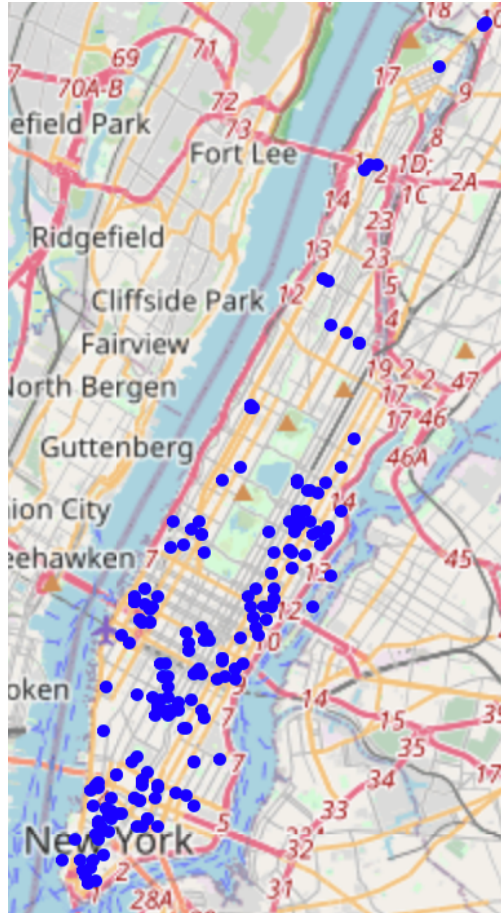
Data Use

The data will be used to:

1. Identify neighborhoods with low population to yoga venue ratio.

2. Identify neighborhoods with high income to earn more money.

3. Determine any correlation between poupulation/income to no. of Yoga venues.

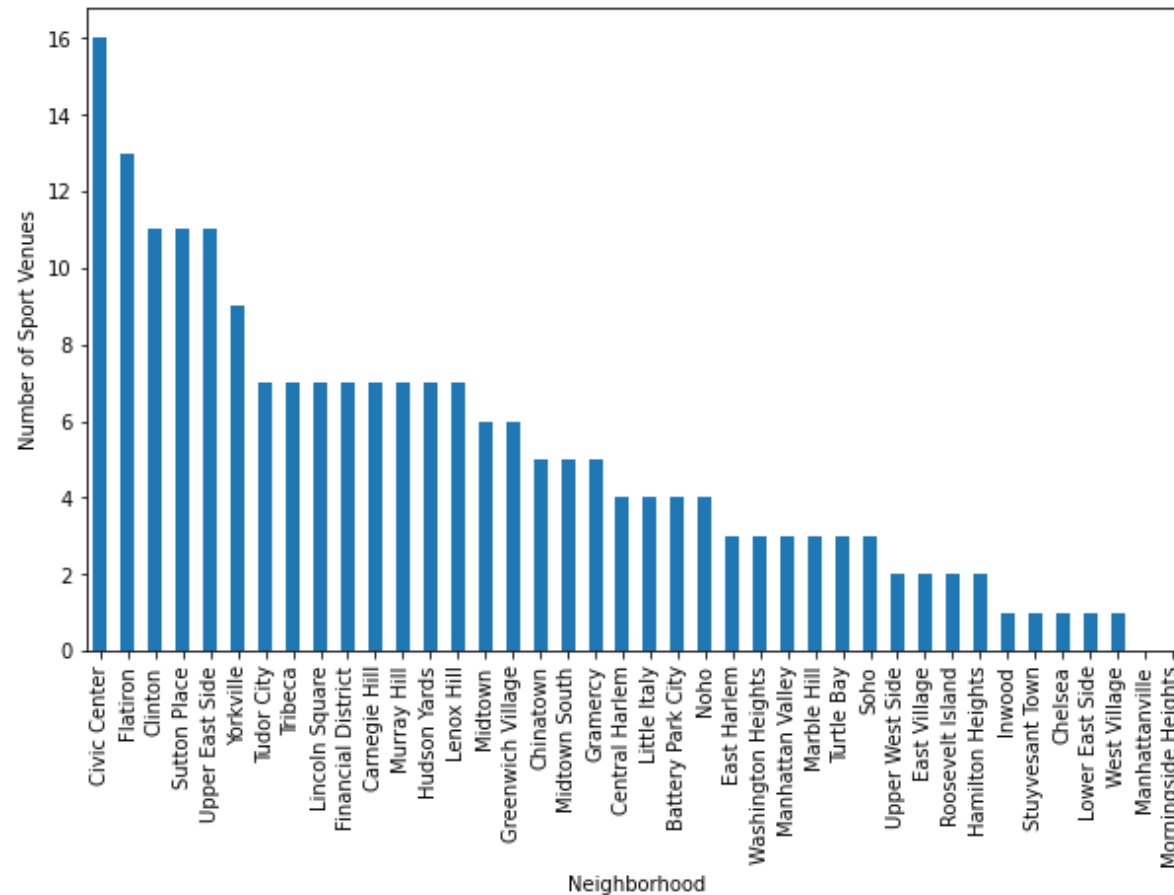4. Visualize the venues on a map.

Data Analysis
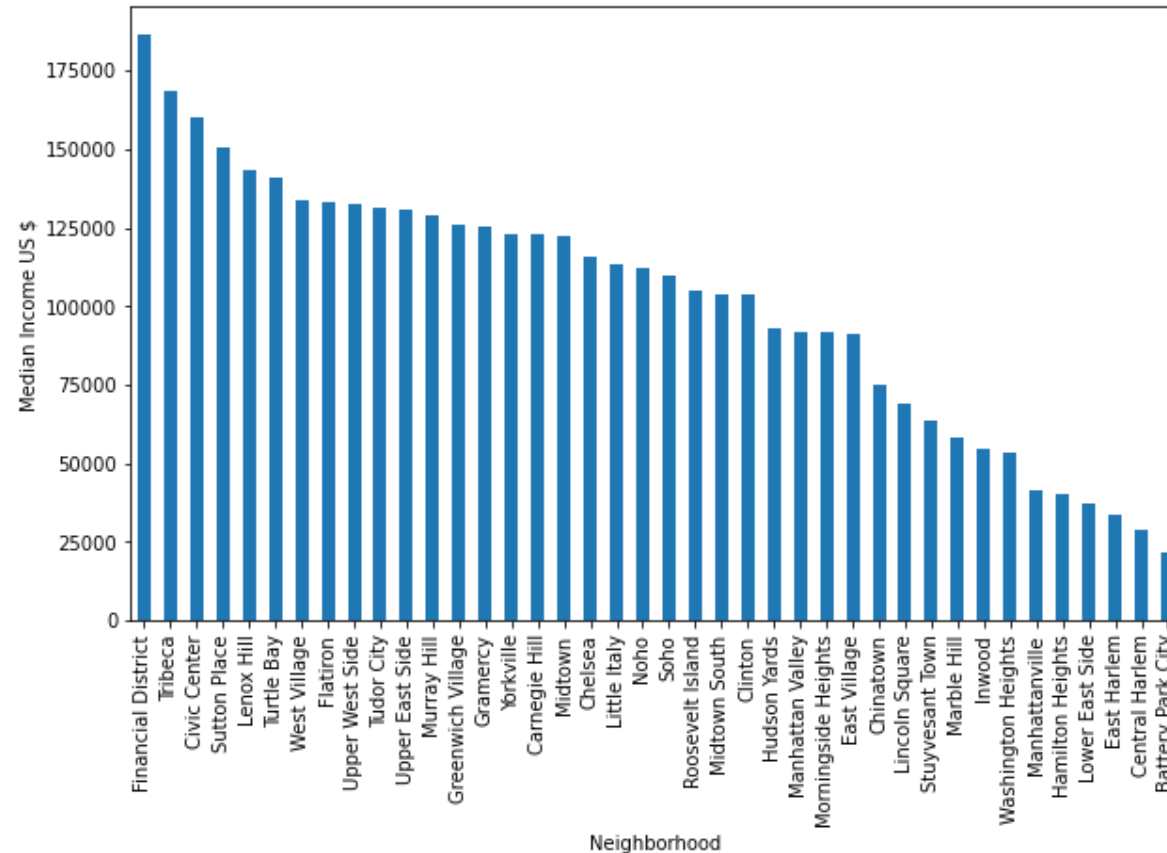
Overview: Sport Venues in Manhattan

Sport Venues in Manhattan

Venue Category	Count
Gym	59
Gym / Fitness Center	63
Pilates Studio	7
Spa	39
Sports Club	1
Yoga Studio	33

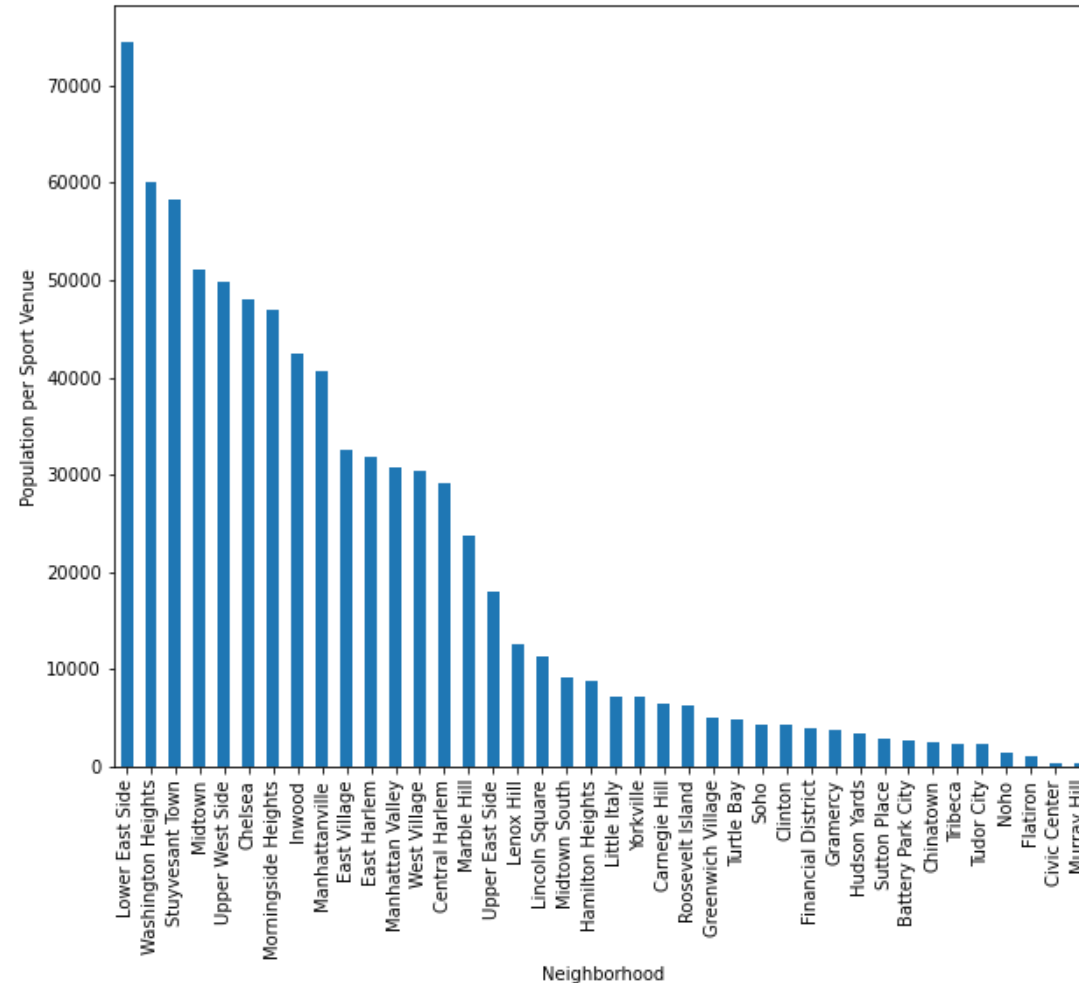# Overview: In which neighborhoods are the sport venues located?

- The most sport venues can be found in Civic Center, Flatiron and Clinton

- Manhattanville and Morningside Heights has no sport venues at all.

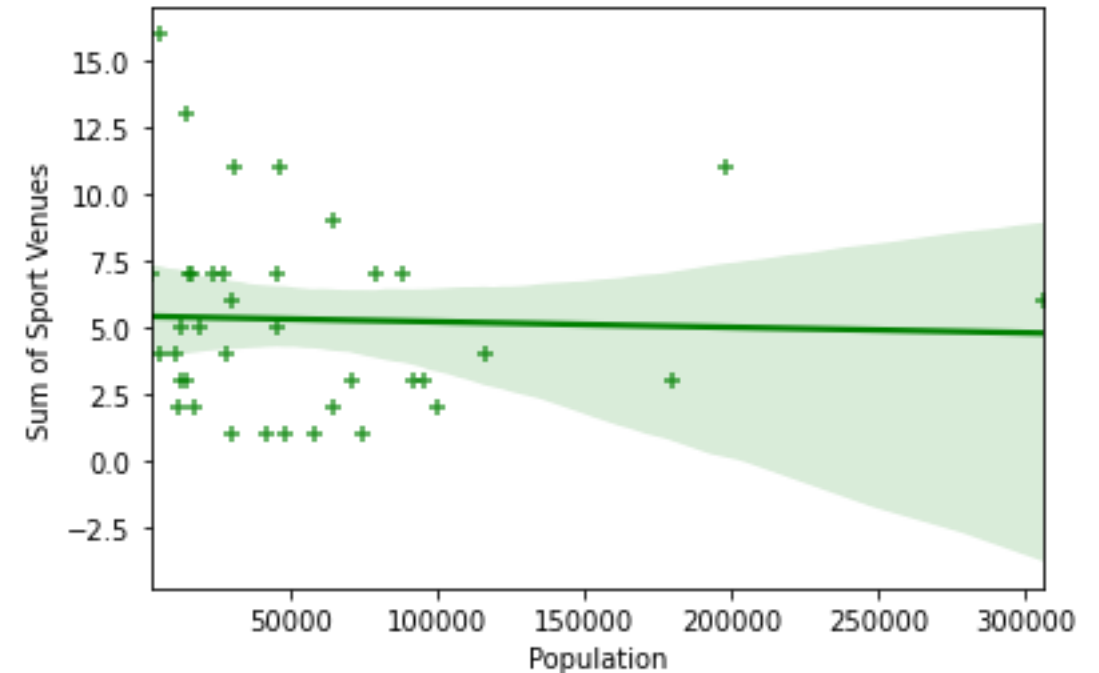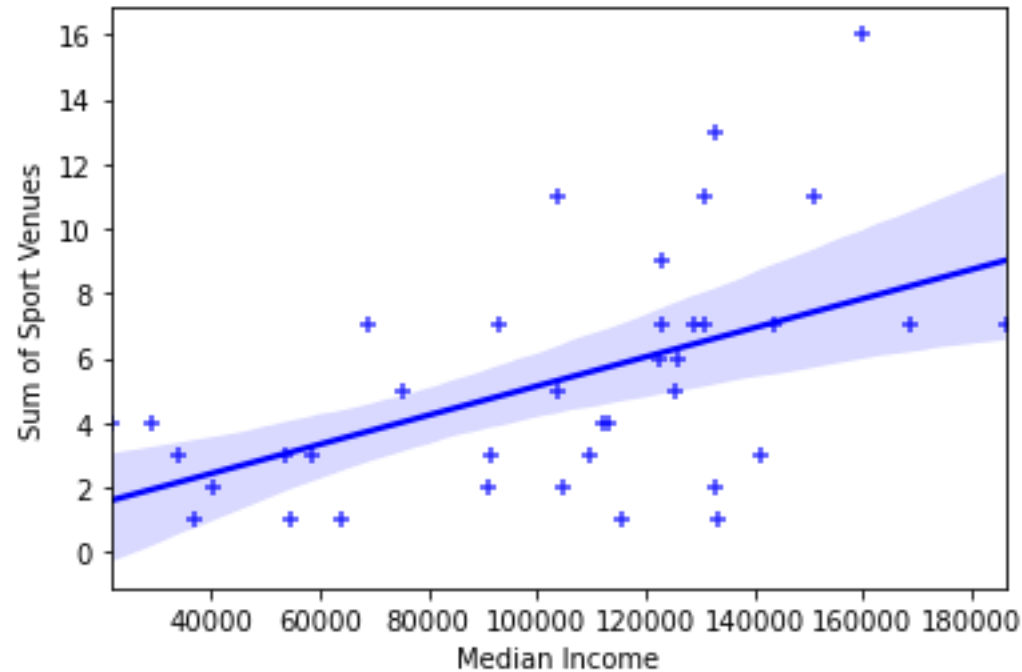# Overview: Median Income in Manhattan by Neighborhood

- The Median Income was used instead of the average income, because it is less affected by "super-rich" people.

- Customers in Tribeca and Civic Center have the highest Median Income.

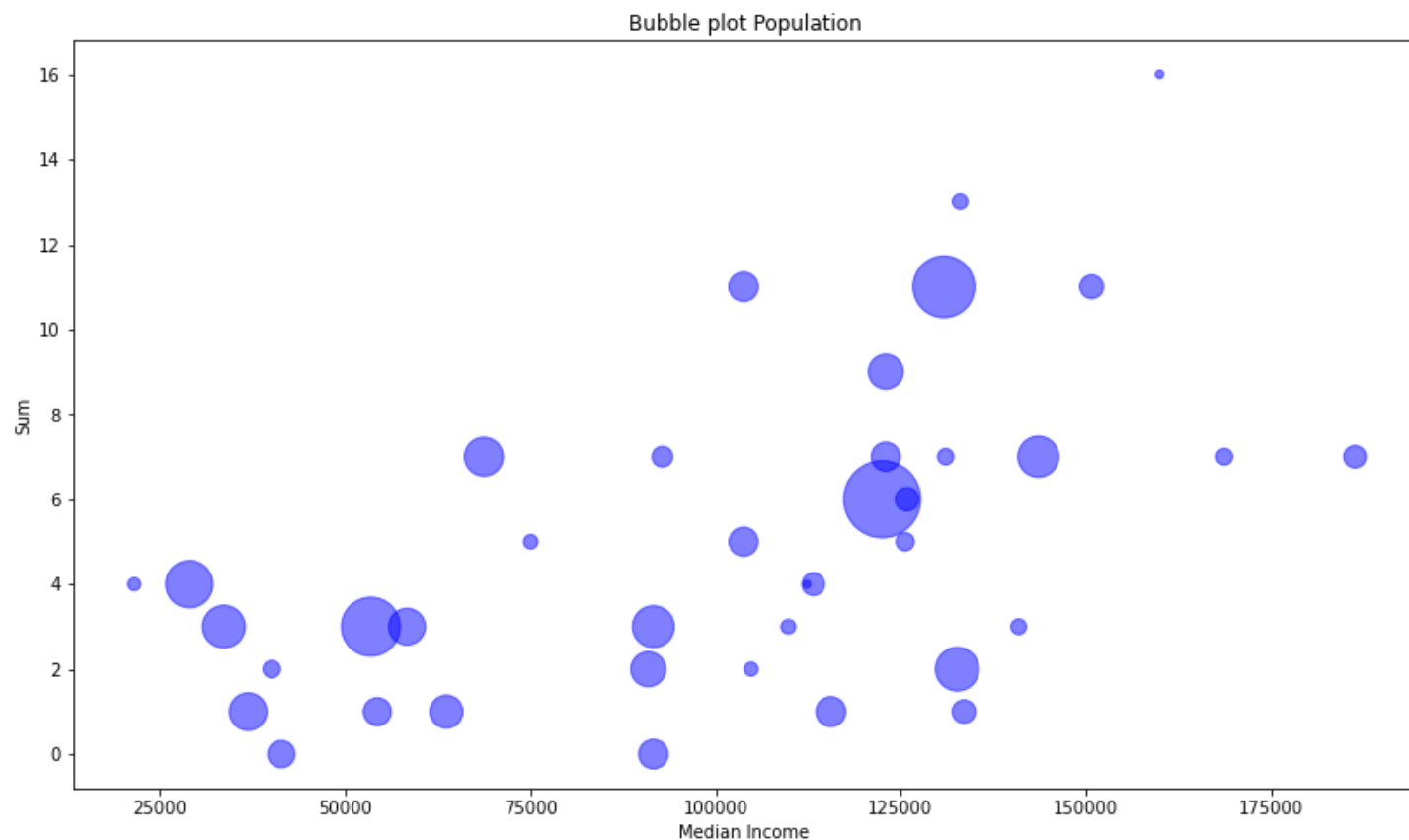# Overview: Population per Sport Venue

- Here we see the population of each neighborhood divided by the number of Sport venues in the area.

- For areas with no sport venues (Manhattanville and Morningside Heights) the total population was used.
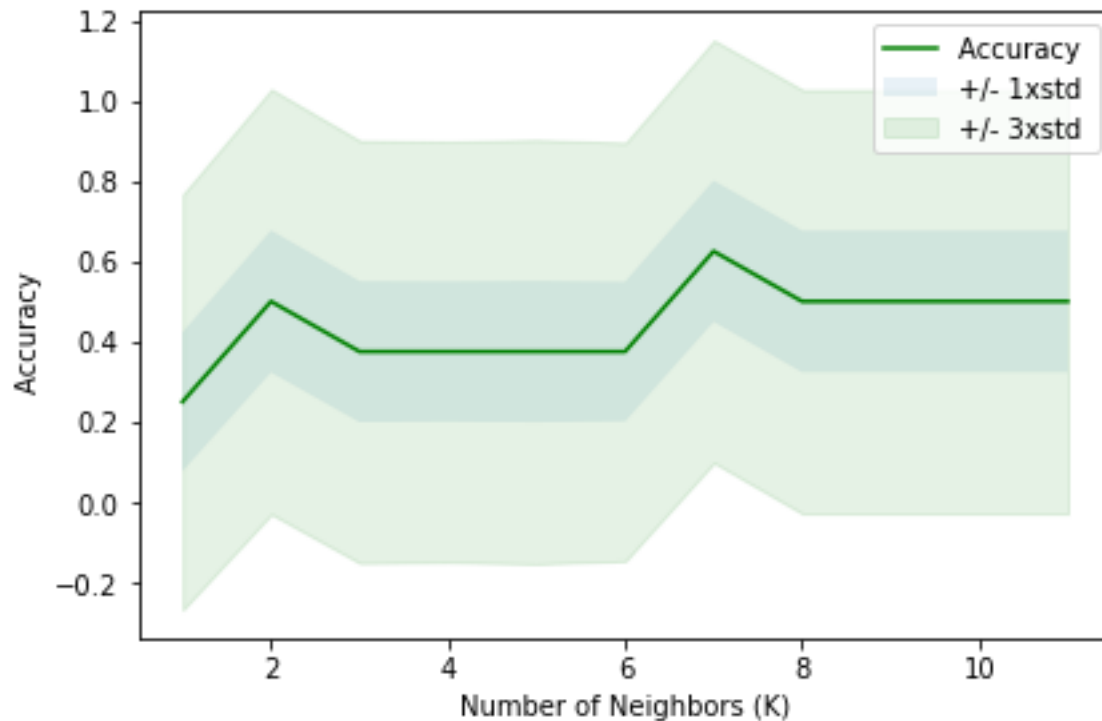
What is the best Predictor?

- The Median Income of a neighborhood show a positive correlation with the number of Sport Venue in this area.
- The population does not correlate with the number of sport venues

Bubble plot: A deeper insight in the data
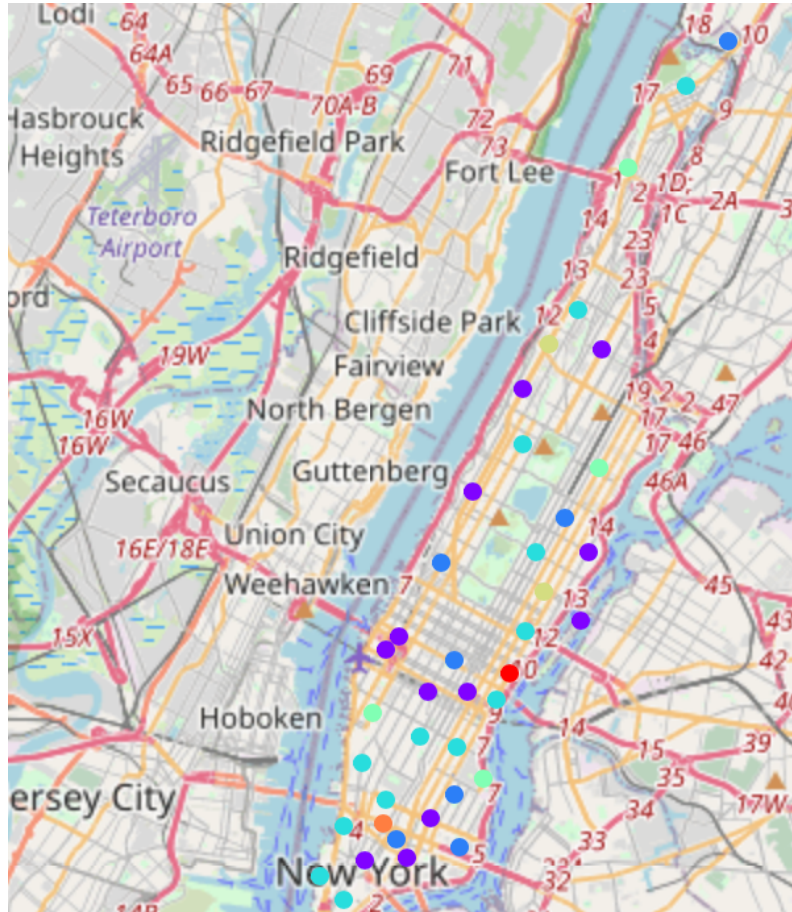

Bubble plot Population

- The scatter plot was upgraded to a bubble plot by adding the information of the population in the marker.

- The size of the markers represent the population in the neighborhood.

- Large circles represent highly populated neighborhoods.

Machine Learning: Finding the best K-Value for the K-nearest Cluster

- The best k-Value is 7. With an accuracy value of 0.625.

- The K-Value was used to calculate the clusters that are displayed in the folium map.

Clustered Venues: Overview on Manhattan

Cluster	Color
0	Red
1	Blue
2	Light blue
3	Turquoise
4	Light green
5	Yellow
6	Orange

- The seven cluster were displayed on a folium map.
- The key features are median income population and number of sport venues.
- Cluster 0 represents only Tribeca. One of the neighborhoods with the highest median income, but low population.
- Cluster 6 contained Soho with a high number of sport venues and a low population.
- The other cluster included 2-12 Neighborhoods.

Conclusion

- The neighborhoods of Manhattan were analyzed according to the feature median income, population and number of sport venues.

- Yoga studios are in areas with high income rather than high population

- We recommend to open a yoga studio at Morningside Heights since here is no competition and the inhabitants have a reasonably high income to afford yoga lessons.