**'Fresh & Fitness Co'**

**Business Introduction**

**'Fresh & Fitness Co'** is a private UK based company specialized in producing 100% fresh fruit/vegetable juices and beverages and selling them primarily in thier fitness facilities. It has presence in over 25 locations in London, Liverpool and Manchester in UK with annual revenue of £40 million and over 400 employees.

**Market segmentation**

It targets mostly top 25% of the population with an annual income of £75,000 and ages between 25-55. **'Fresh & Fitness Co'** has been quite successful in terms of retaining higher turnout rate of its members (93.5%) compared to other private fitness halls (85%) in the market.

**Products and Facilities**

**'Fresh & Fitness Co'** only uses organic fruits and vegetables produced by local farmers only for its juices and protein beverages. In addition to regular gym facilities and swimming pool, it offers wide range of services such as private cross-fit and yoga classes, oriental meditation and cardio dances for all its premium members.

**Business Problem (International Expansion)**

**'Fresh & Fitness Co'** has been planning to expand its business internationally for few years already. Its first primary target been the state of New York, US where it is planning to open a test facility. Despite the fierce competetion in US fitness market, the company is quite confident in its success and ability to differentiate itself from other market participants.

**However, the company has following questions/problems to analyze before it decides to enter US market:**

*1. Which area/neighborhood to enter in New York, US?*  
*2. What is the most popular venue in the neighborhood in the area chosen?*  
*3. What is the demographics/overall health condition of the area?*  
*4. How healthy is the diet of the population?*  
*5. How likely is it have a high member turnover rate?*

**Methodology and Data**

1. **Population by county in the state of New York**

I have started the analysis with the demographics to see the 10 most populous counties in NY. I pulled the data from [NY Health Data](https://www.health.ny.gov/statistics/vital_statistics/2009/table02.htm), cleaned it to show only County, Population (from highest to lowest), Area and Density which will be used for our analysis in later sections.

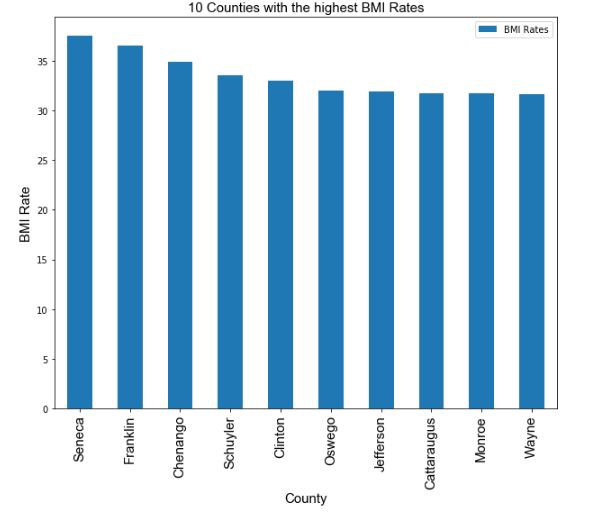
**2. NY State: Age-adjusted percentage of adults obese (BMI 30 or higher)**

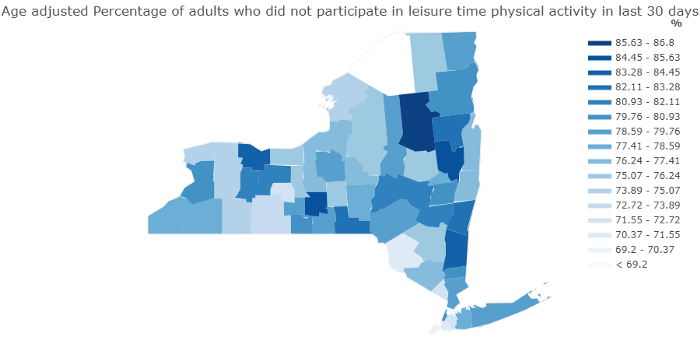
One of the most important factors in determining the area would be to analyze the overall health well-being of the population. I decided to check the Body Mass Indicator (BMI) index > 30 for adults (obesity rate) each county. The [data](https://health.data.ny.gov/Health/Community-Health-Age-adjusted-percentage-of-adults/mmzn-r7ff) was cleaned to show the 10 counties with the highest percentage BMI index above30 and vice versa.

**3. Age-adjusted Percentage of adults eating 5 or more fruits or vegetables per day**

Another point I took into consideration is the percentage of adults who consumed 5 more fruits/vegetables per day. The most healthy eating county was Onondaga (33.7%) and only 6.3% of the population consumed 5+ fruits/vegetables in the Bronx.

**4. Age-adjusted Percentage of adults who did not participate in leisure-time physical activity in the last 30 days**

Finally, I cleaned data for physical the percentage of people who did not participate in the leisure-time physical activity in a month and combined health data into one data frame. The most active area was Bronx (30.8%) while the most inactive county was Hamilton (13.2%)



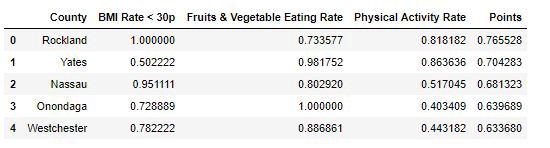
# ****Combined health data heatmap****

After I consolidated the health data into a single data frame, I normalized the data using [MinMaxScaler method of Scikit-learn](https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.MinMaxScaler.html" \t "_blank) because numbers I was going to use for the heatmap was distinct and in different ranges.

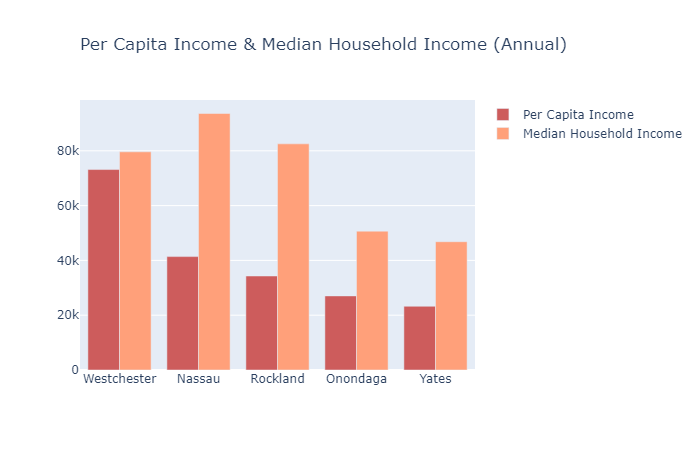
BMI Rate < 30p | Fruits & Vegetables Consumption | Physical Activity in last 30 days



Based on the data illustrated in the heatmap, I created points for each county which is the equally-weighted average of BMI, Fruits and Vegetables and Physical activity data and selected **5 counties with the highest points**.



One of the key factors for the business is the financial health of the region. ‘fresh and fitness Co.’ primarily targets individuals with above-average annual income and prices its products and services accordingly. I created a stacked bar chart using [Plotly](https://plot.ly/python/bar-charts/" \t "_blank) to portray the annual income statistics for the five counties above.



As the bar chart shows, Westchester has the highest per capita annual income around (75k) which is the clear winner in this selection.

I used [**Foursquare API**](https://developer.foursquare.com/)to get the most popular venues in 13 largest cities & towns of Westchester, NY to see which area has sports facilities as one of the most favorite spots within 20,000 m (12.5 miles) radius of the city center.

Using Foursquare API, I pulled 1300 different venues and 137 distinct venue types of the areas chosen and displayed them **10 Most Common Venues**in a dataframe.

As the venue types are somehow similar in their nature, it is slightly difficult to see a clear distinction between cities. To partition data and come up with the final decision for the business, I used [**k-means clustering**](https://towardsdatascience.com/understanding-k-means-clustering-in-machine-learning-6a6e67336aa1)which is one of the simplest and popular unsupervised machine learning algorithms.I chose 3 Ks, meaning I decided to cluster data into 3 different groups and plot it in the map using [**Folium**](https://python-visualization.github.io/folium/).

## **Results**

**Cluster 0 (Red):** White Plains, Rye, Eastchester, Harrison

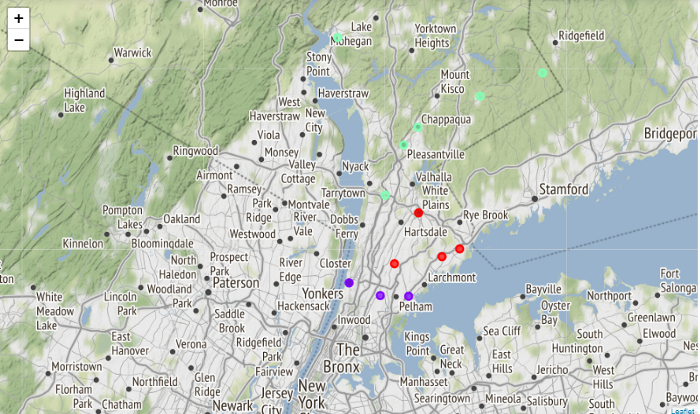


**Cluster 1 (Purple):** Yonkers, New Rochelle, Mount Vernon



**Cluster 2 (Green):** Peekskill, Bedford, Greenburgh, Lewisboro, Pleasantville, Chappaqua





# Conclusion and suggestion to the ‘fresh & fitness Co’

The report suggests that the business should consider one of the cities of Cluster 1 (Yonkers, New Rochelle and Mount Vernon) in Westchester County, NY as an initial location for the following reason:

1. It has one of the highest overall well-being of the population in terms of a healthy diet, low obesity index and staying active
2. It has the second-highest per capita income in the state of NY which indicates people are most likely afford higher than average services and products of the business
3. According to Foursquare, cities being proposed for the entry have more sports facilities (gym & golf centers) and recreational spots such as parks in the most common venues than other areas.