# High Income US Neighbourhoods Category Analysis

## *Applied Data Science Capstone - Final Report*

-Edoardo Romani- March/April 2019

# Introduction

## Topic of interest:

Exploring the distribution of venue categories for the 65 most affluent Neighbourhoods in the USA.

## Target Audience

The aim is to map each high-income neighbourhood according to the relative presence of each venue category bucket. The following analysis can then help identify over/undeserved neighbourhoods with respect to each category.

Ultimately, this can help small-business owners' value relative segment competition, and can also be supportive in the identification of target segments, for particular areas, in which a business launch or store opening would make sense, given the venue category distribution of the segment and the ability to serve an affluent pool of customers.

# Data

## List of highest-income urban neighborhoods in the United States

Source: Wikipedia ->**List of highest-income urban neighborhoods in the United States**

WIKIPEDIA
*The Free Encyclopedia*

Data contains name of neighborhoods, race distribution, income data (mean), as well as location references (Metropolitan Area, State).

## Venue and Location data for high income neighbourhoods

Source: Foursquare API ->

FOURSQUARE

This data will be fetched from the API based on data source number 1. The resulting dataset is venue data (50 venues for each neighbourhood) of the recommended locations in the most affluent USA neighbourhoods:

# Methodology: 3 step process

| | |
|---|---|
| Restaurants | 124 |
| Shops | 92 |
| Enterntainment | 68 |
| Cafes | 33 |
| Sports | 25 |
| Art | 8 |

## Exploratory data analysis

Assessment of the various categories that are automatically assigned/rendered to the data points by Foursquare

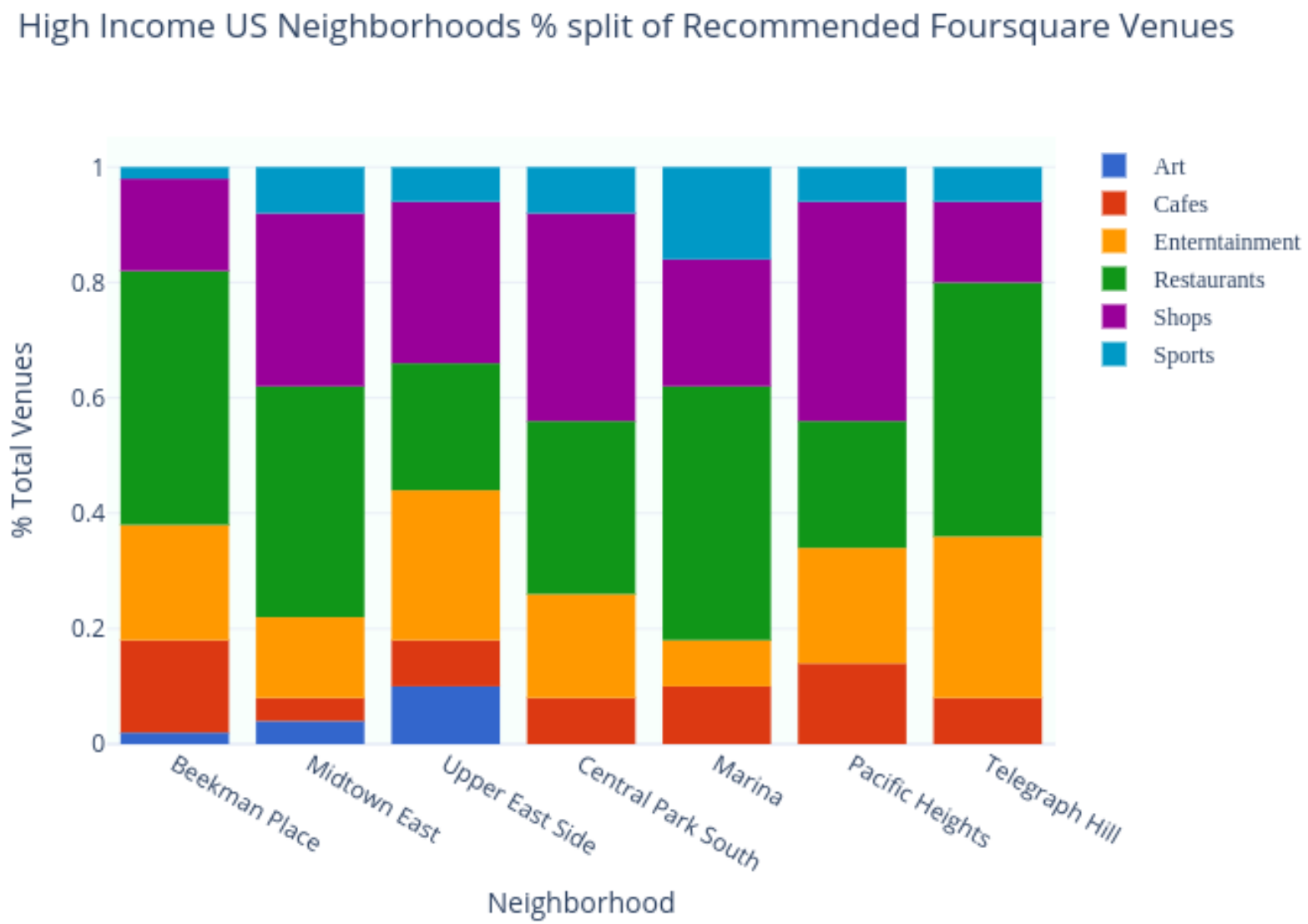
## Data cleaning and codification

Re-codification of each data point's Category Value to appear in exclusively one of the following categories (Restaurants, cafes, shops, art, sports, Enterntainment)

## Penetration Rate calculation

Knowing that, for each neighbourhood, 50 recommended venues were fetched from the API (the maximum amount allowed under the basic/free account, one can calculate the proportion of each category within a certain neighbourhood.

# Results




Stable representation across the Restaurant, Cafes, Shops and Entertainment categories

Skewed results across the Art, Sports and Cafes (especially under-represented in Midtown east)

# Recommendations

## Topic of interest:

Potential Business implications (for simplicity, I am assuming this analysis as the only reason for taking an implications, without further investigating into the structural nature of those neighbourhoods, where I realize other confounding factors are here not accounted for)

- Offering Artistic venue solutions to neighbourhoods underrepresented in the Art Venue category as a viable source of revenue with initially low market competition

## Additional areas for analysis:

To further investigate penetration rate differences from a statistical perspective, one could opt for the following as additional model recommendations:

- ANOVA (Analysis of Variance)
- Clustering or Classification Algorithms

# Conclusion

In this final assignment I have tried to show that it is possible to utilize Foursquare Venue data to identify potentially attractive business opportunities in high-income areas.

The API is very powerful and easy to use, and it is definitely a go-to tool for any geo-based analytical project.

FULL ANALYSIS LINK (NOTEBOOK):

https://github.com/Eromani1/Location_Data/blob/master/Venue_Analysis_%20Top%2010%20US%20Neighborhoods.ipynb