Investor Analysis: Predicting which venues needed in NYC

MELVIN CAPITAL INVESTMENT INTO MANHATTAN, NYC

NNAMDI OSIA

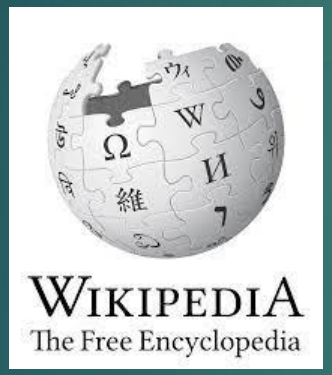
Introduction: Background

- Melvin Capital Management LP is an American investment management firm based in New York City.
- Melvin Capital invests primarily in tech and consumer stocks and is reported to have $8 billion in assets under management (AUM) as of January 2021..
- Melvin Capital is looking to produce a list of venues by popularity so that they are able to determine which venue are needed and which venues are commonplace in New York City Burroughs.

Introduction: Problem &Interest

- Data that might contribute to determining which venues are lacking and which are commonplace include New York City data that contains borough, neighborhoods along with their latitudes and longitudes, population data from scraping Wikipedia, and venue information from foursquare API.
- Hedge funds and venture capitalists would be interested in which venue are needed and which venues are commonplace
- Also, individuals looking to startup business in New York City may also be interested in this project and data.

Data: Data Sources



- Data for NYC can be found online at https://cocl.us/new_york_dataset or https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/labs/newyork_data.json.
- We can get the population data from scarping Wikipedia: https://en.wikipedia.org/wiki/Neighborhoods_in_New_York_City We are going to go through each of the links of neighborhood and find the population of each of them.
- For additional insights, we can fetch venue information from foursquare API (https://api.foursquare.com/v2/venues).

Data Acquisition and Cleaning

- Numpy is used as a library to handle data in a vectorized manner.
- Pandas library is used for data analysis.
- Json library is used to handle JSON files.
- Nominatim is imported from geopy to help convert addresses into latitude and longitude values.
- The requests library is imported to handle requests used in FourSquare API.
- The json notmalize library is used to tranform JSON file into a pandas dataframe.
- Matplotlib is used for associated plotting modules and sklearn is used to import KMeans clustering.
- Finally, folium is used for map rendering.

Analyze Nearby Venues

- We begin by finding nearby venues for each neighborhood and create a new dataframe.
- We create a function that will find this information within a predetermined radius using the name of the neighborhood and latitude and longitude of the neighborhood.
- After we have found nearby venues for each neighborhood, we then check how many venues are returned for each neighborhood.
- Finally, we find out how many unique categories exist.

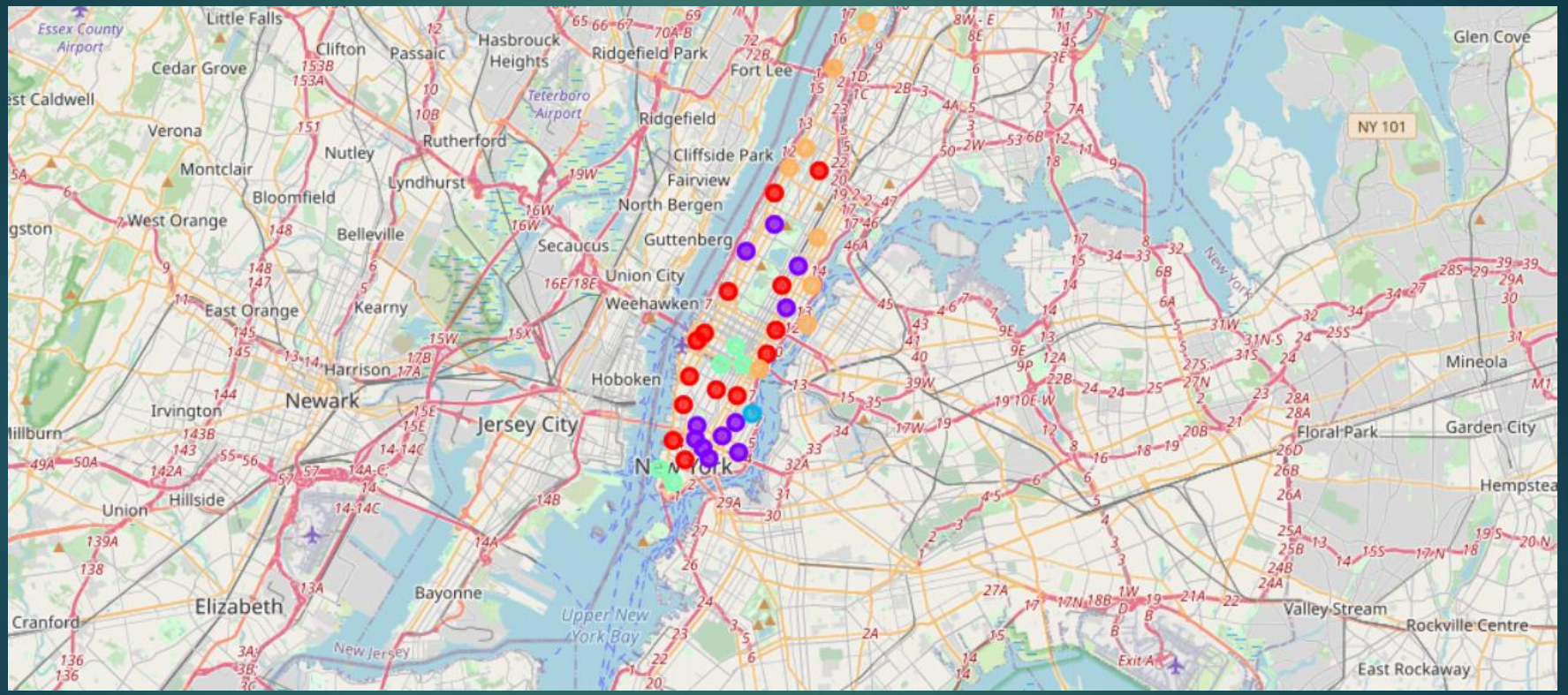
Exploratory Data Analysis

- One Hot Encoding
  - One Hot Encoding is a process in the data processing that is applied to categorical data, to convert it into a binary vector representation for use in machine learning algorithms.
- Kmeans clustering is one of the most popular clustering algorithms and usually the first thing practitioners apply when solving clustering tasks to get an idea of the structure of the dataset.
- The goal of kmeans is to group data points into distinct non-overlapping subgroups.

Results

- Further analysis shows us a total number of about 3235 venues in Manhattan alone.
- Out of the 3235 venues, 325 were found to be unique. For each neighborhood, we print the top 5 most common venues.
- Our analysis shows us that Hotel, Coffee Shops, Cafes, Italian and American Restaurants were the most popular venues.
- Doing a cluster analysis give us some additional information and takeaways into which neighborhoods are similar and unique

Results


Essex County Airport
Little Falls
Cedar Grove
Clifton
Passaic
Hasbrouck Heights
Ridgefield Park
Fort Lee
Glen Cove
Verona
Nutley
Rutherford
Teterboro Airport
Ridgefield
Cliffside Park
Fairview
North Bergen
Montclair
Bloomfield
Lyndhurst
West Orange
Belleville
Secaucus
Guttenberg
Union City
Weehawken
East Orange
Kearny
Harrison
Hoboken
Newark
Irvington
Jersey City
New York
Union
Hillside
Elizabeth
Bayonne
Cranford
Upper New York Bay
New Jersey
NY 101
Mineola
Floral Park
Garden City
Hempstead
Valley Stream
Rockville Centre
East Rockaway

Conclusion

Based on our analysis, I think that we can conclude that a great investment idea for Melvin Capital would be one of the following:

- Bar
- Park
- Coffee Shop
- Baseball Field
- Pet Service
- Gas Station
- Farmers Market
- Bistro
- Gym / Fitness Center
- Cocktail Bar.