Exploring Towns in the Hudson Valley

# Introduction

## Background

With travel restrictions put in place due to Covid-19, many are re-thinking their living location. As many people are allowed to work from home and will likely continue to do so, moving out of urban area or even suburbs into the great outdoor is now an option that not only become feasible, but welcome.

## Problem

As someone who has lived in New York for 15 years, I had never explored the outskirts of New York City (let’s be honest, living in NYC comes with the pride that says the city has everything I need). While I have heard of towns like Beacon and Kingston, I never know what each town has to offer and what sets them apart. Visiting each individual town will take considerable amount of time; looking up towns online seems to lack directed focus.

## Interest

With this project, I hope to find a way to compare towns along the Hudson Valley side by side and group them by cluster based on the town’s characteristics. Each cluster will have its own charm and uniqueness and thereby allows user to further explore towns in a certain cluster based on his/her own specific interest.

# Data acquisition and cleaning

## Data Sources

A quick Google search returns a website: Discover Upstate New York (<https://www.discoverupstateny.com/packages/?region=hudson-valley&category=cities-towns>). It has listed out the top 8 towns along the Hudson Valley.

## Data Cleaning

The link provided a short description for each town. I used BeautifulSoup to parse through the link to obtain the names of the towns. I also manually added other towns from the Google search. After the list of towns was compiled, I used GeoCoder to iterate through the list to obtain the coordinates of each town.

## Feature Selection

While using GeoCoder, coordinates were not returned for certain towns. Only towns with coordinates returned were selected for further exploration within this project. In the end, 13 towns were selected.



# Exploratory Data Analysis

## Foursquare

I used the “Explore” function in Foursquare to identify the different types of venues each town has to offer. Because my Foursquare credentials only allow the search to return 100 venues, after running an initial round of the function, it became clear that the search is too broad and returns remarkably similar results for each town.

I decided to limit the search to only **trending**, **nextVenues** and **topPicks**. This will allow the search to only return the most popular venues that best represent the towns.

# Classification Model

## Applying K-Means Clustering

The goal of the project is to identify clusters of towns that speak to different personal preferences – big outdoors, big suburban town, a mix of both, etc. I used K-Means Clustering to group the towns together based on the types of the venues that Foursquare returns.

After using the elbow method, K=3 is the most efficient number of clusters.

## Solutions to problems

After running K-Means Cluster, I have renamed the different classifications based on its characteristics of its top venues.

Table

Description automatically generated

# Conclusions

By using Foursquare and K-Means Cluster, it can be identified that the closer the towns are to New York City, the more likely they are commutable suburban towns. As we move further north, there are towns that offer primarily hiking trails and outdoor activities. And finally, if we move further north and get closer to the Catskills, towns offer a more variety of shops and restaurants, as well as hiking trails and outdoor activities.

# Future Directions

This project only studies the different characteristics of each town. As city dwellers contemplate moving out of the city, we would need to dive more into costs of living and the real estate market (both sale and return) in those towns.