Battle of Neighborhood

Asian Restaurants of Pennsylvania

Coursera Capstone Project

Pazhani Subramaniam Sivathanu Pillai

Mar 01, 2020

Contents

[**Introduction**: 2](#_Toc34081208)

[**Problem** 2](#_Toc34081209)

[**Stakeholders** 2](#_Toc34081210)

[**Methodology Highlights** 3](#_Toc34081211)

[**RESTful API Calls to Foursquare** 5](#_Toc34081212)

[**Clustering and Segmentation** 15](#_Toc34081213)

[**Insights of analysis** 18](#_Toc34081214)

[**Conclusion** 19](#_Toc34081215)

[**References** 19](#_Toc34081216)

**Introduction**:

Pennsylvania is one of the most populous state of united states and the 9th most densely populated state of America. Undoubtedly, it is a most diverse state with over 13.5 million people from various ethnicity.

**Problem**

Food Diversity is an important part of an ethnically diverse metropolis. The idea of this project is to categorically segment the neighbourhood of Pennsylvania into major clusters and examine their cuisines. A desirable intention is to examine the neighbourhood cluster’s food habits and taste. Further examination might reveal if food has any relationship with the diversity of a neighbourhood. This project will help to understand the diversity of a neighbourhood by leveraging venue data from Foursquare’s ‘Places API’ and ‘k-means clustering' unsupervised machine learning algorithm. Exploratory Data Analysis (EDA) will help to discover further about the culture and diversity of the neighbourhood.

**Stakeholders**

This quantifiable analysis can be used to understand the distribution of Asian cuisines over different counties of Pennsylvania. Also, it can be utilized by a new food vendor who is willing to open his or her restaurant. Or by a government authority to examine and study their city’s culture diversity better.

**Methodology Highlights**

In this analysis, we shall take different counties in Pennsylvania and to categorically segment its neighbourhoods into major clusters and examine their Asian cuisines. The intention is to examine the neighbourhood cluster's

Import the Pennsylvania population and projection information map from <https://harrisburg.psu.edu/>

Import GEOJSON file for the different counties at Pennsylvania from <https://www.pasda.psu.edu/>

Mark those counties in the Pennsylvania map

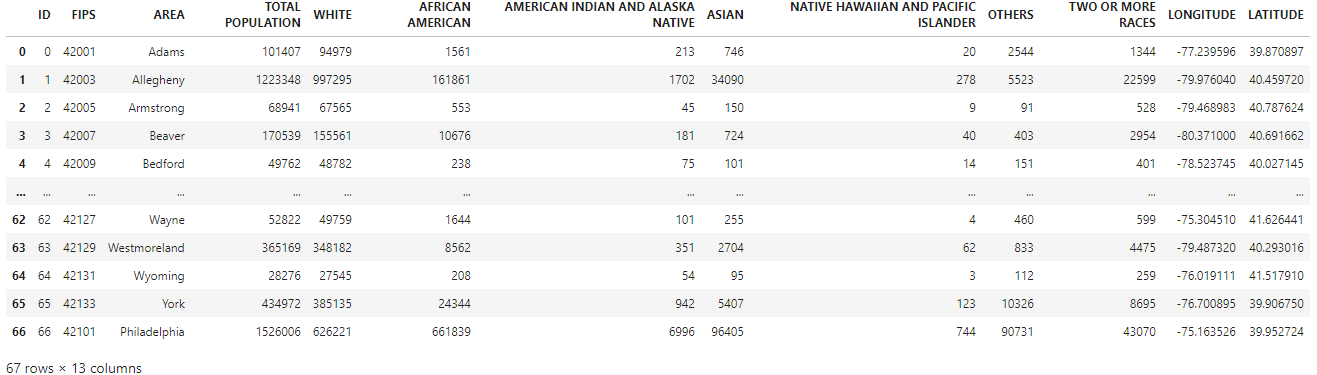
Fetch the availabilities of different restaurants using foursquare API

Analyse different Asian cuisines available in the counties and derive the relationship with the population ethnicity.

Refer the detailed code in the notebook

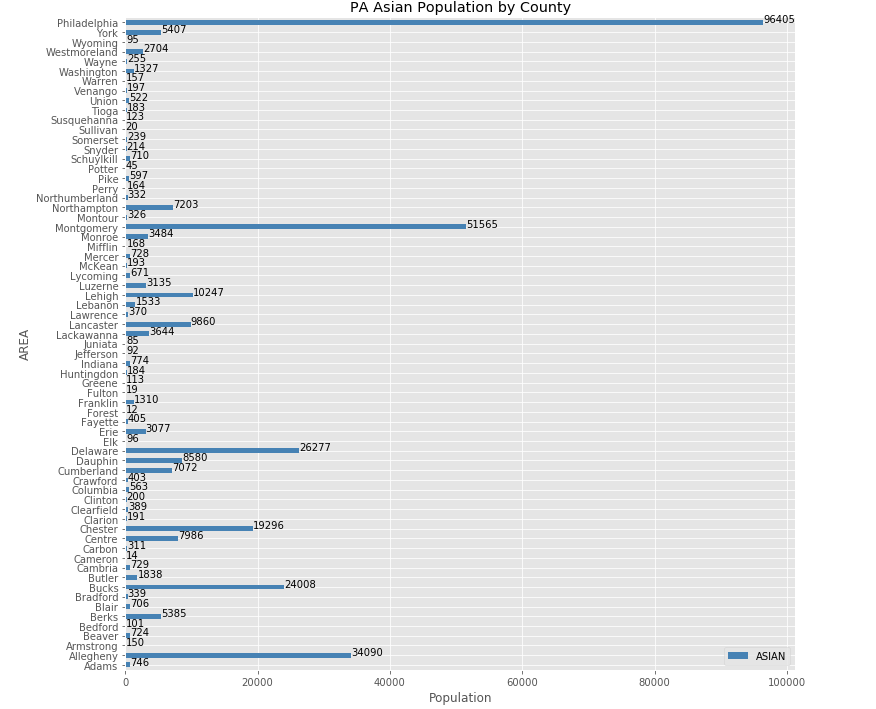
Download the Pennsylvania population dataset





Plot Asian population by county





**RESTful API Calls to Foursquare**

The Foursquare API is used to explore the neighbourhoods and segment them. To access the API, ‘CLIENT\_ID’, ‘CLIENT\_SECRET’ and ‘VERSION’ is defined.

There are many endpoints available on Foursquare for various GET requests. But, to explore the cuisines, it is required that all the venues extracted are from ‘Food’ category. Foursquare Venue Category Hierarchy is retrieved and returned request is further analyzed.

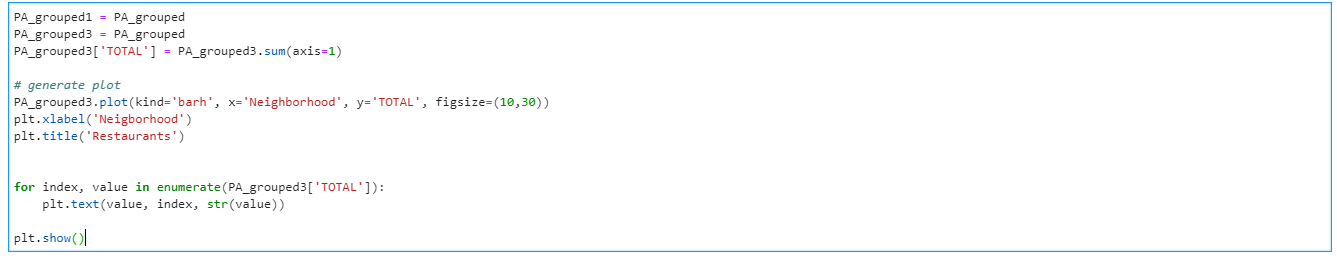
Upon analysis, it is found that there are 2 major categories ‘Asian’ and ‘Indian’ in order to get all Asian restaurants in Pennsylvania.

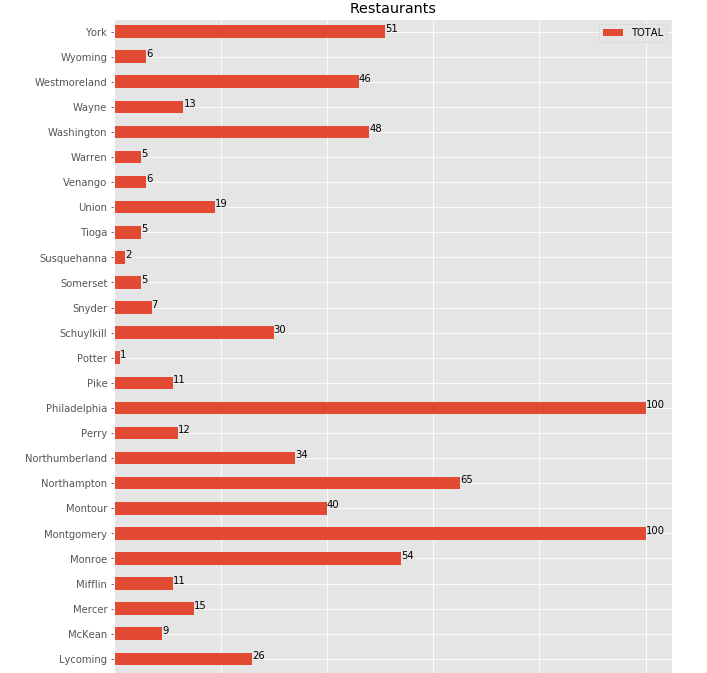


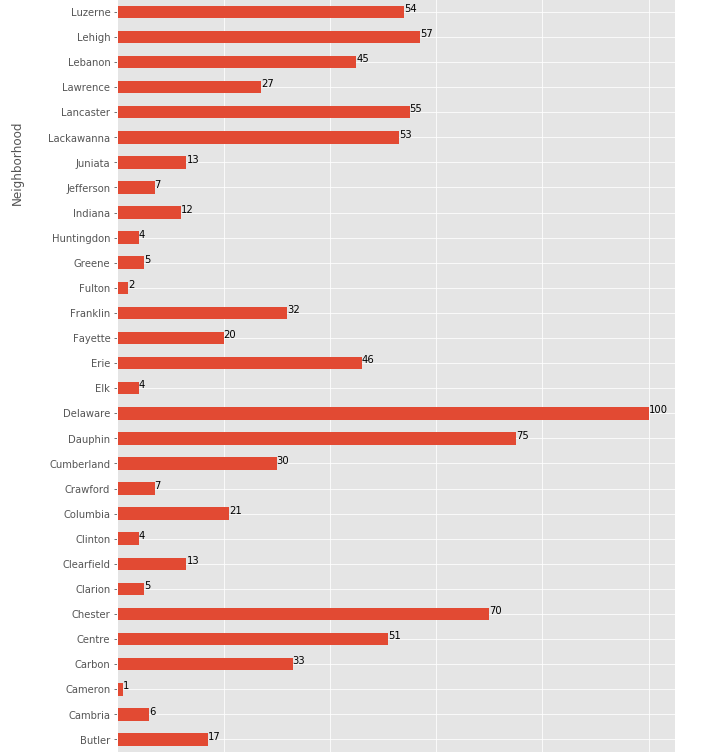


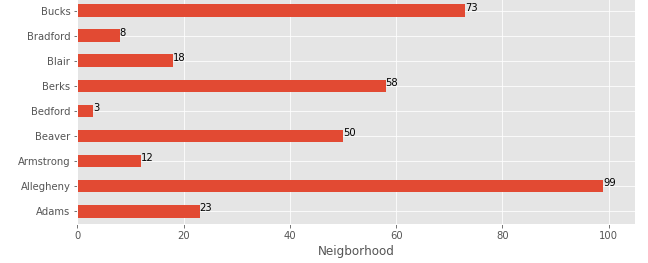


Lets see the distribution of Asian population across Pennsylvania

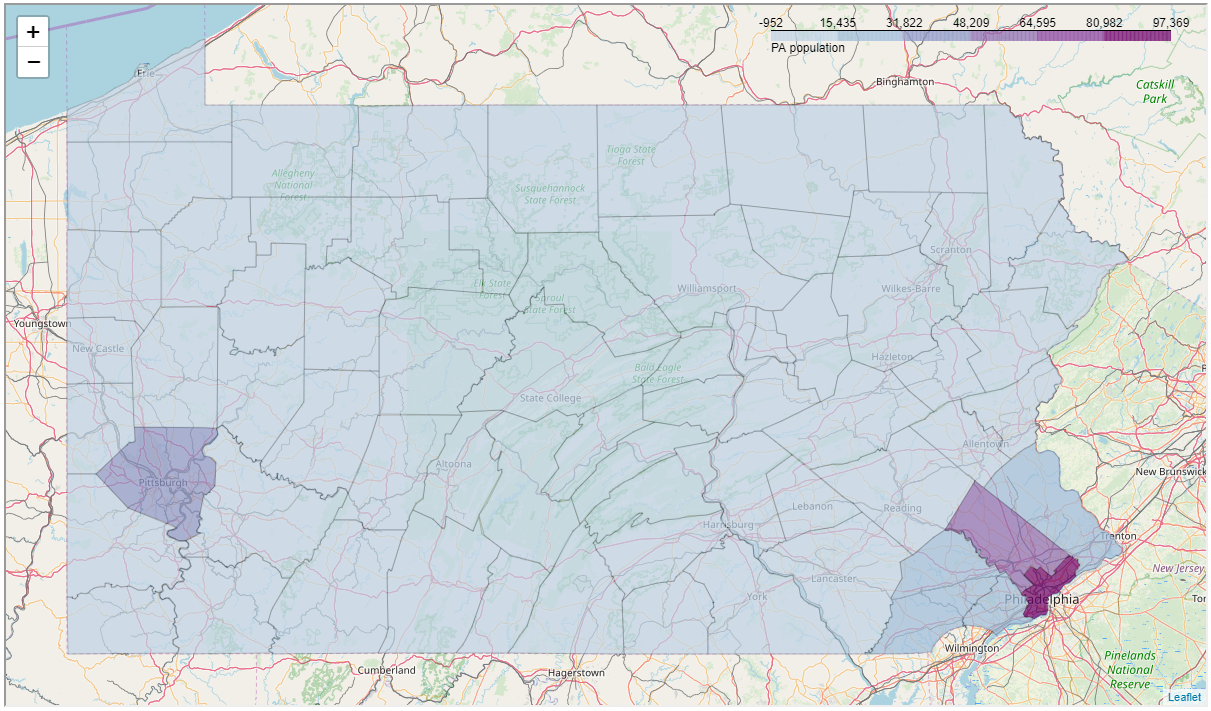




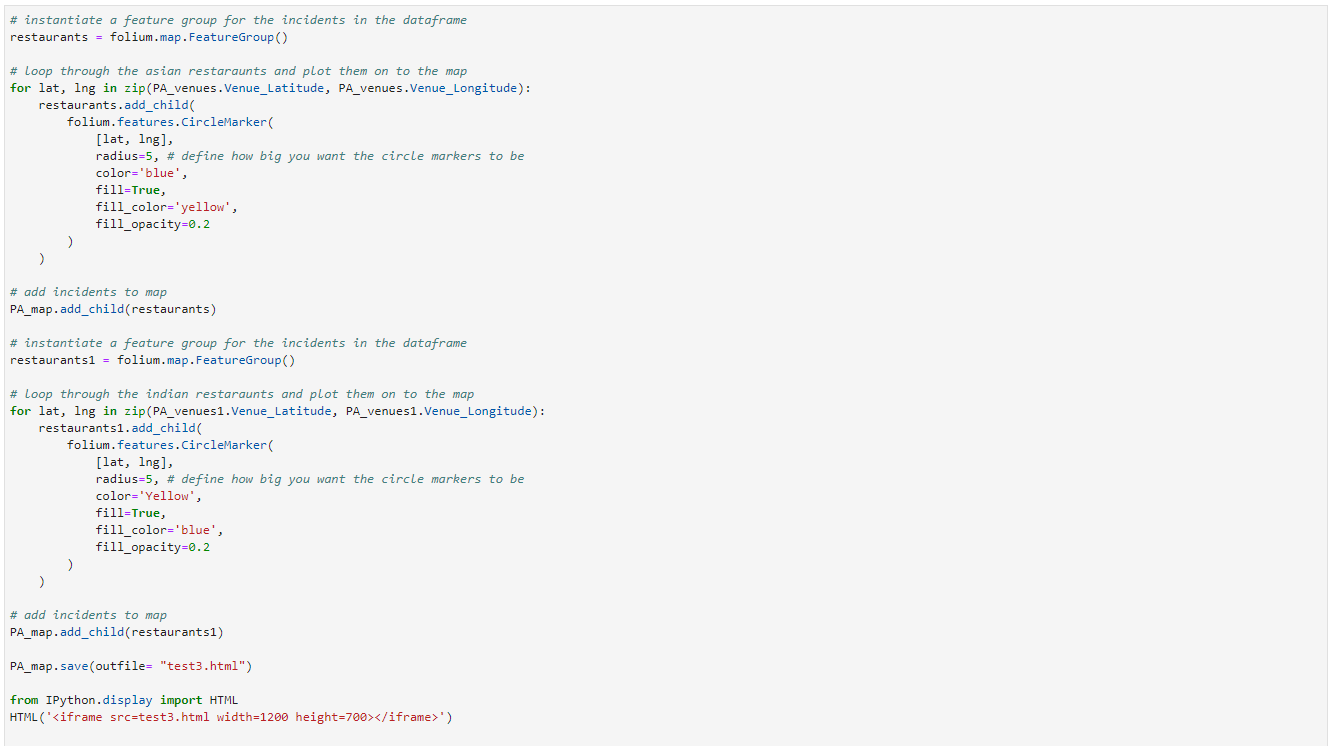


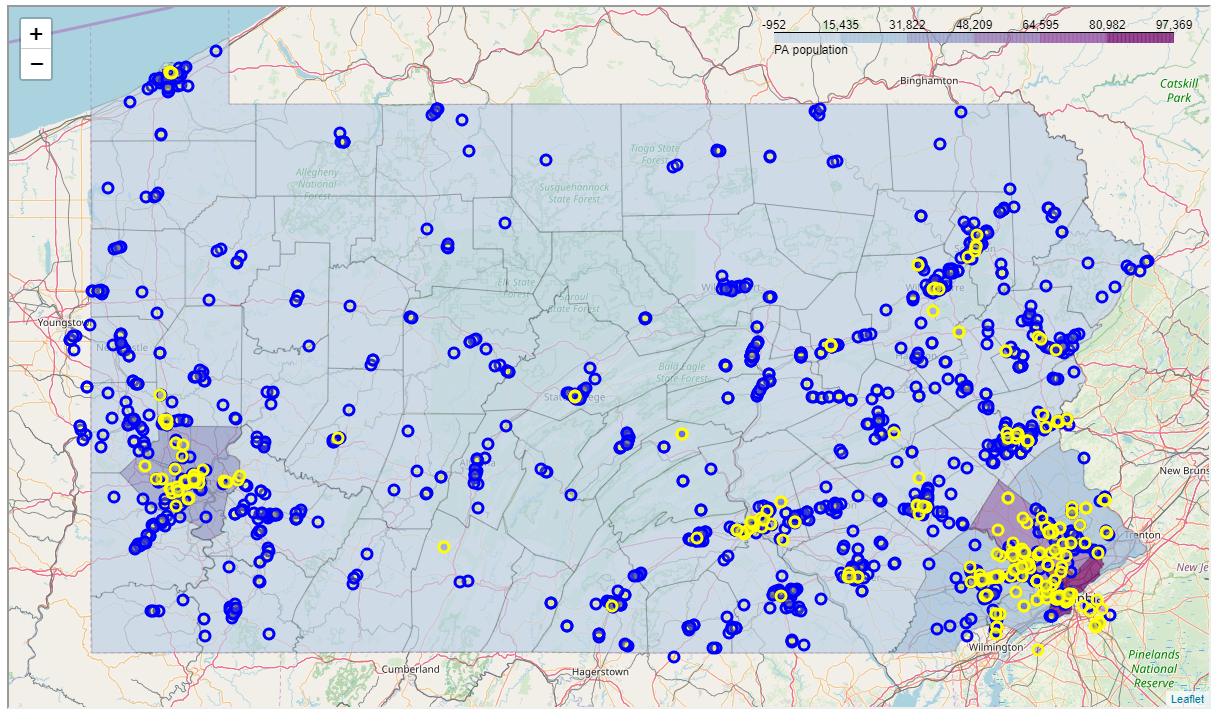






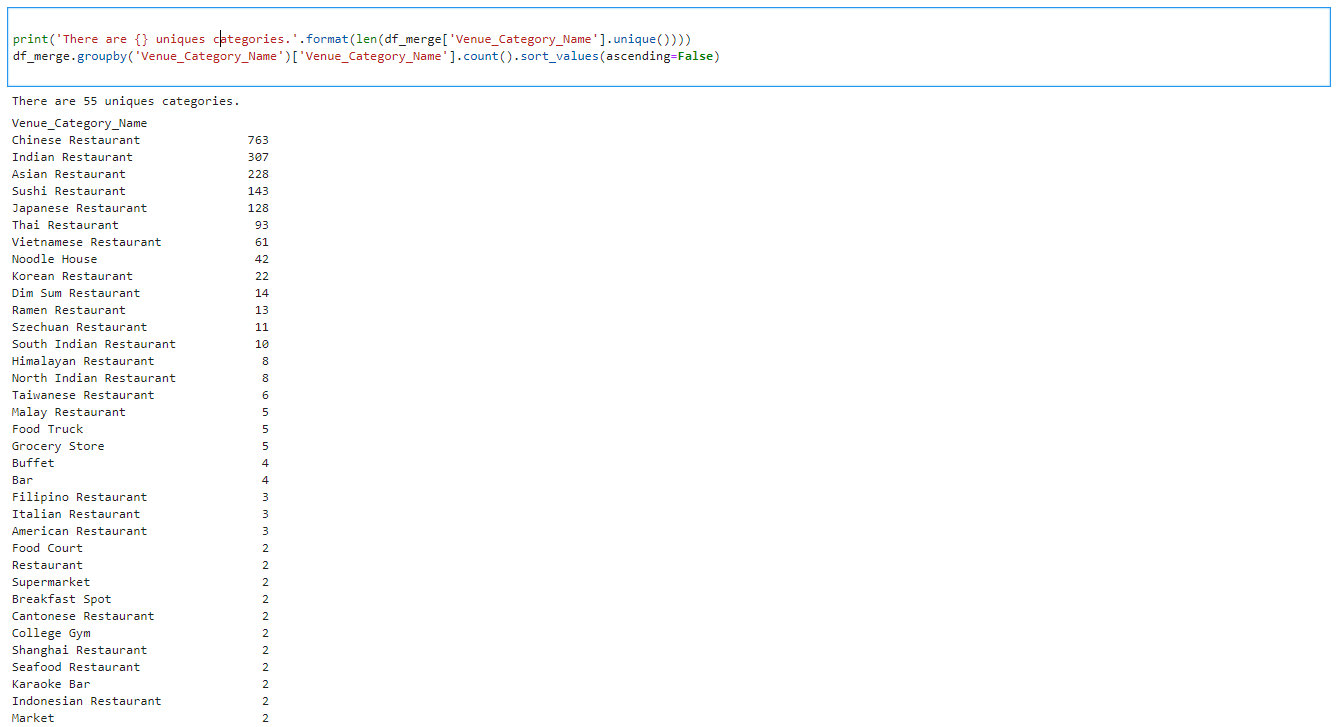
Lets plot the distribution of Asian and Indian restaurants across Pennsylvania





The distribution of restaurants are high on highly populated counties and there are counties where there are no Indian or Asian cuisines.

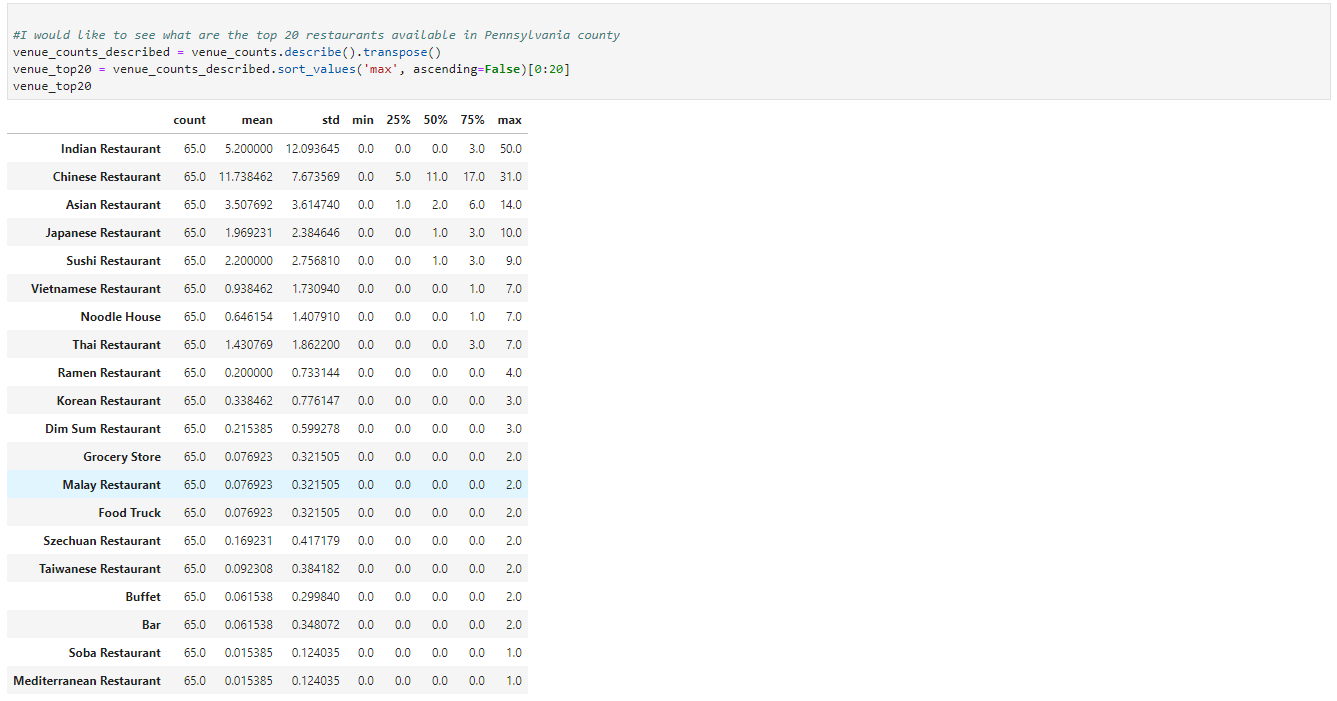
Let’s examine the different sub-categories of the Asian restaurants and their count in Pennsylvania.





There are 55 unique restaurant categories in Pennsylvania.

The top 20 restaurants by its total count are



The scatter plot of top 10 restaurants by Asian population is



The Indian restaurants averages around 40 restaurants in counties with population over 20k followed by Chinese restaurants with around 15 restaurants in those counties.

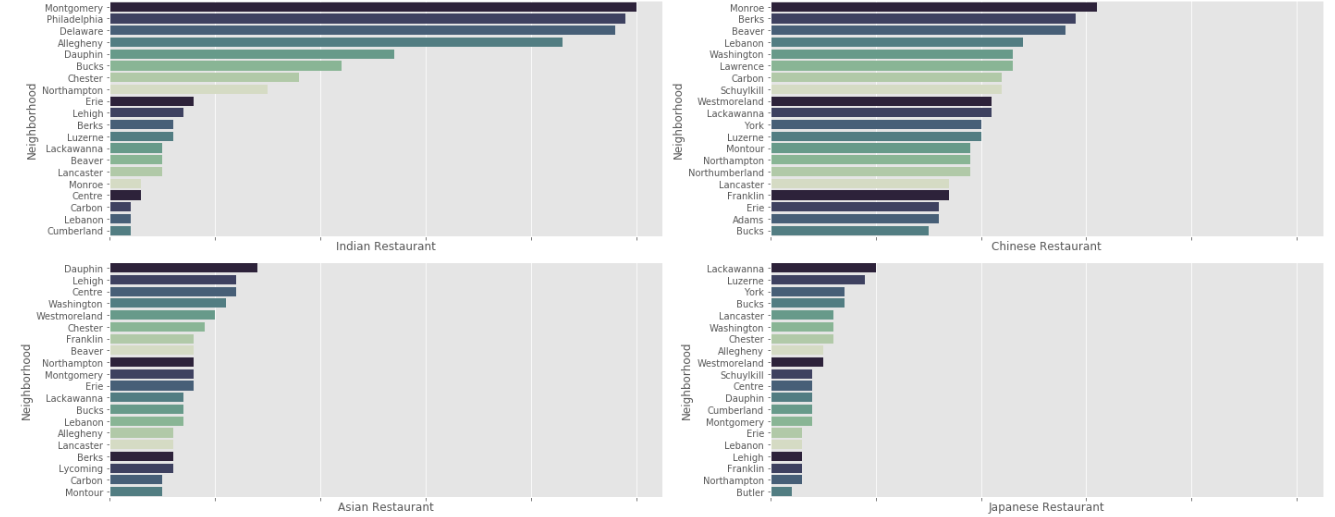
Let’s plot the distribution for counties with population less than 20k.

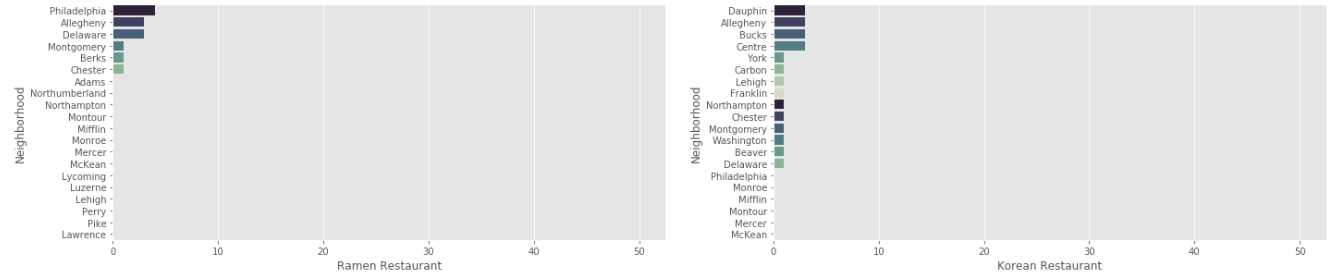
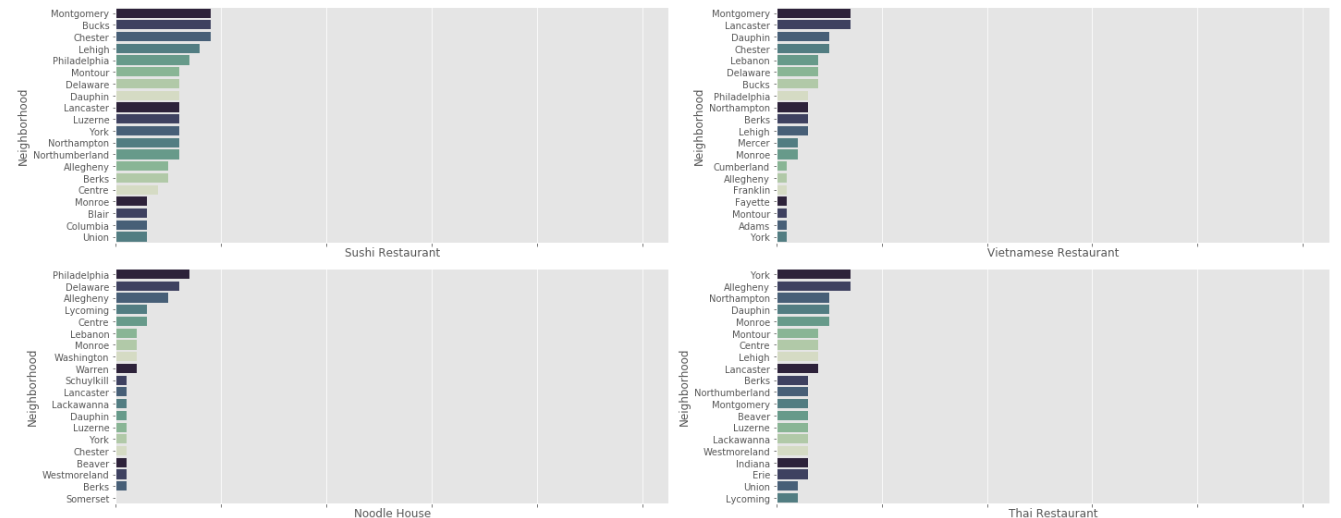


The Chinese restaurants shows upper hand in the counties with less than 20k population.

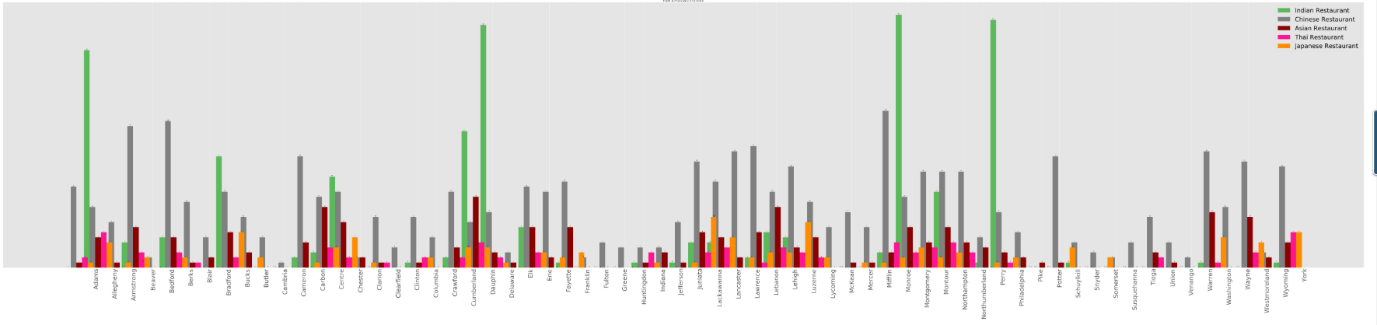
The distribution of the top 10 restaurants across counties.





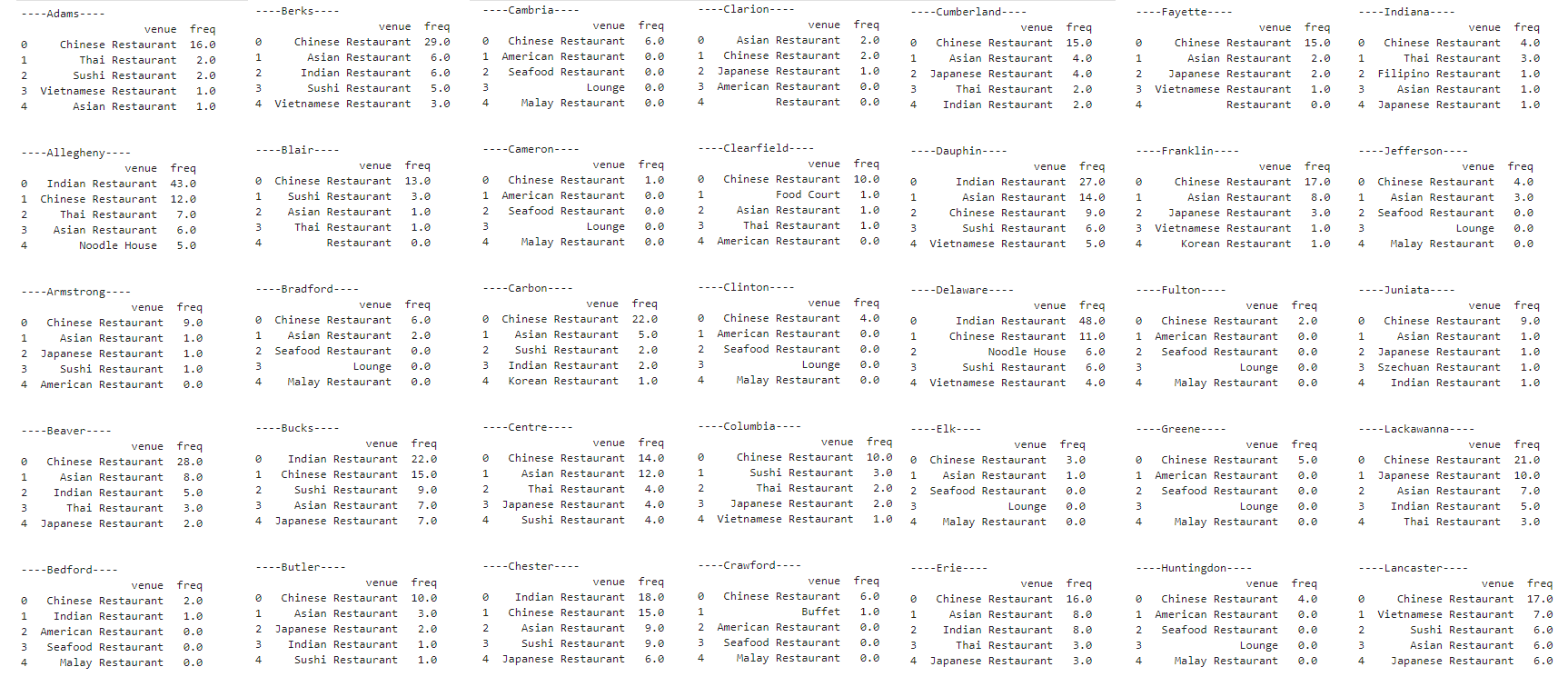


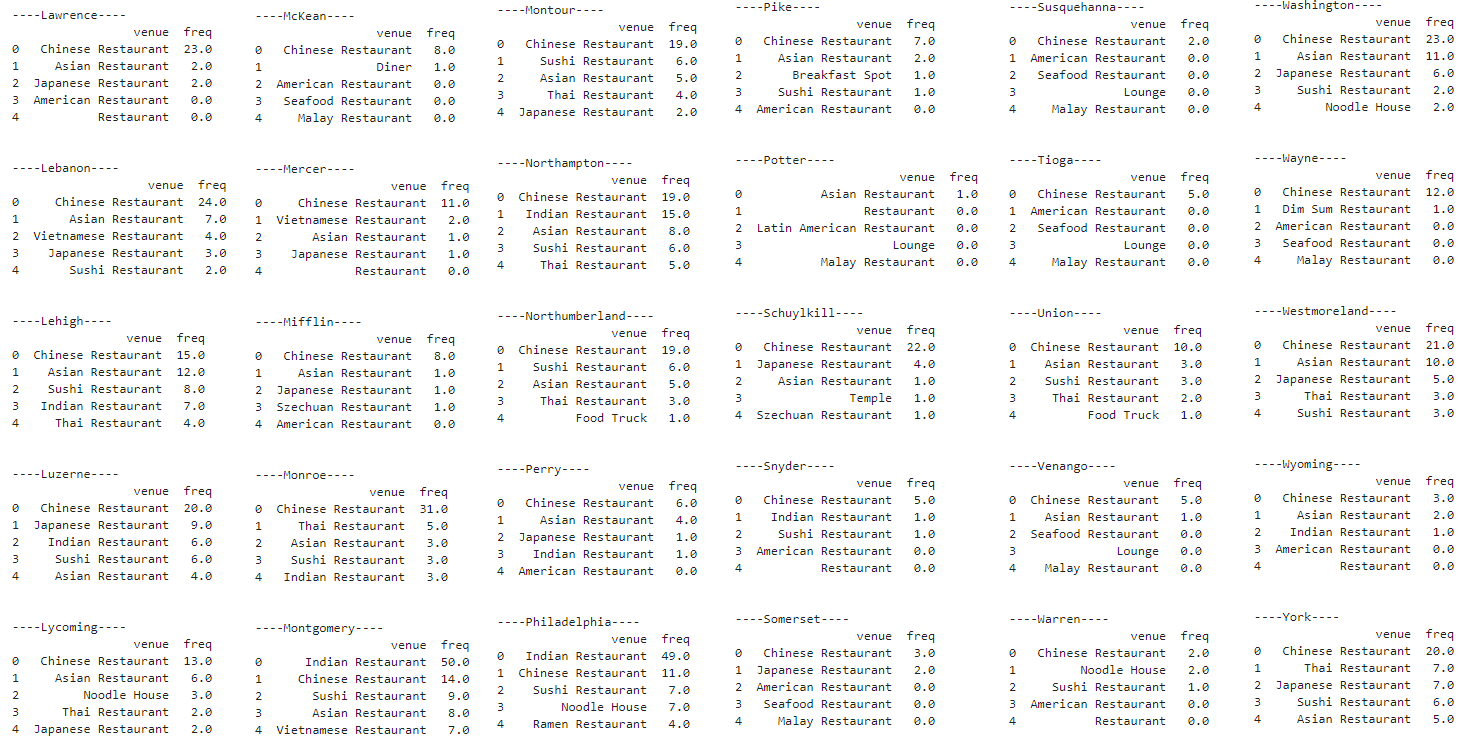
Let’s have a closer look at the distribution of top 5 restaurants across counties.



Lets find the most common venues across counties

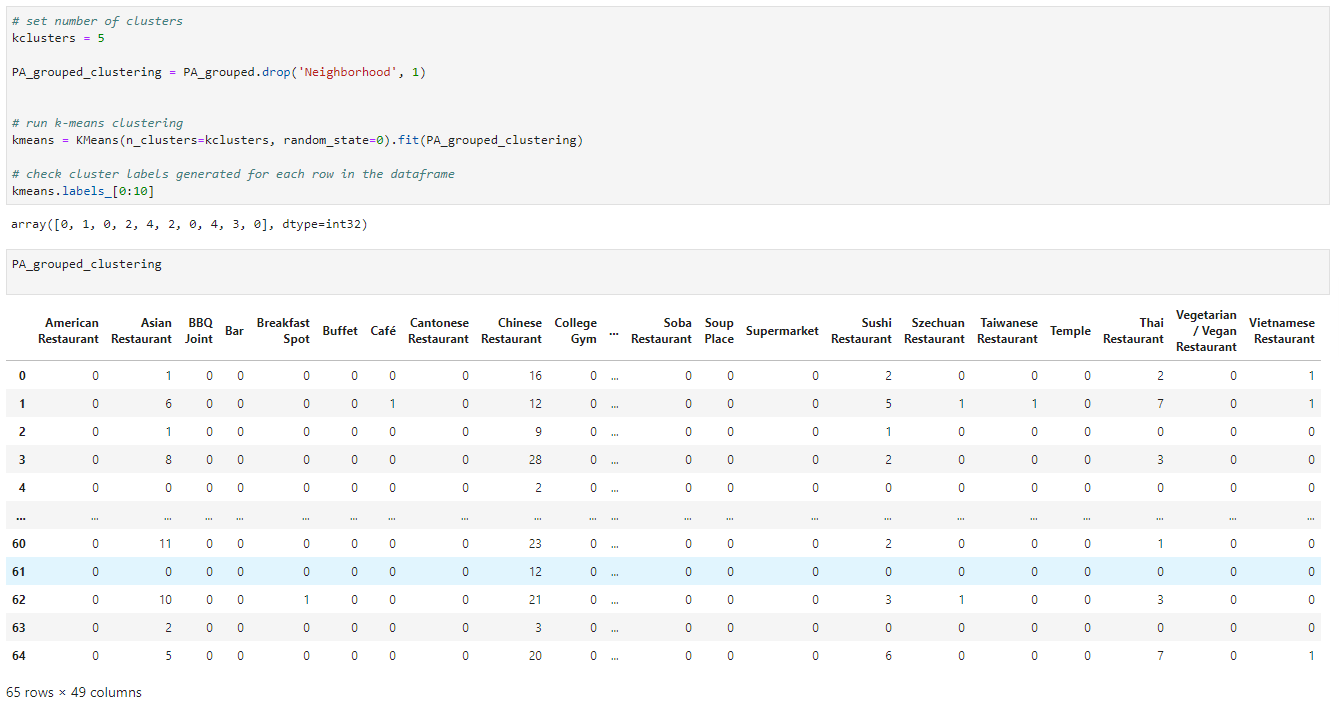






**Clustering and Segmentation**

Run k-means to cluster the neighborhood into 5 clusters.

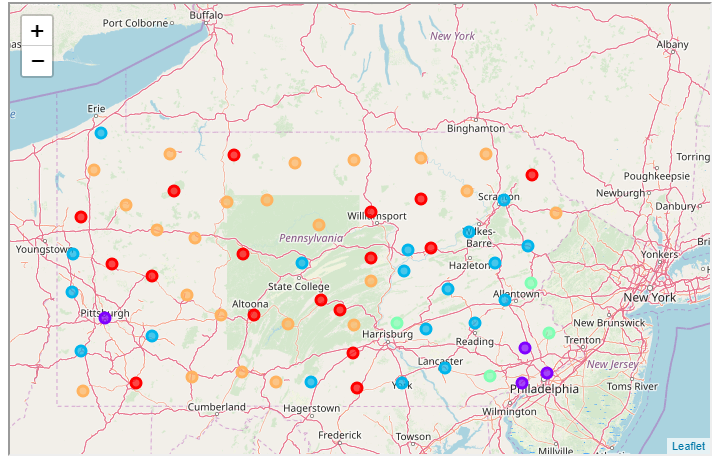


Add clustering label and merge with PA counties population dataset.

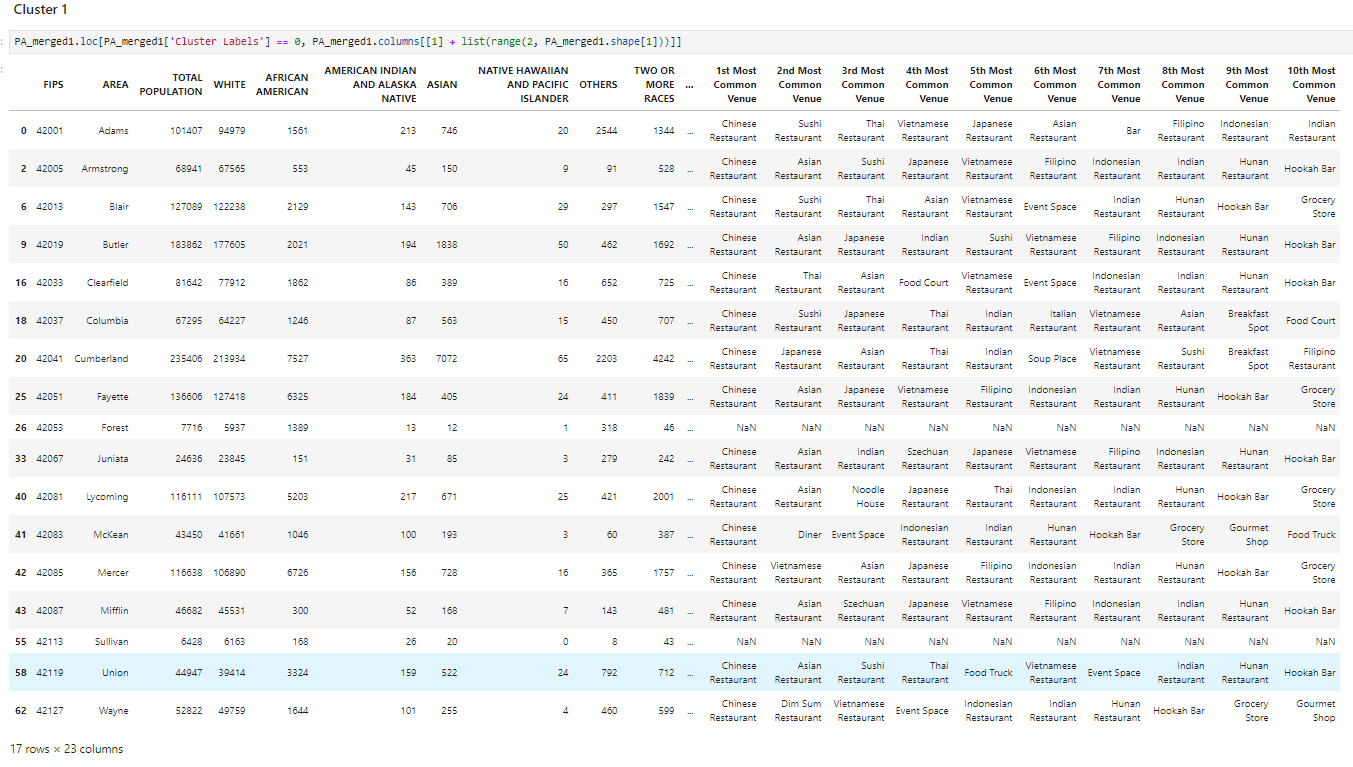


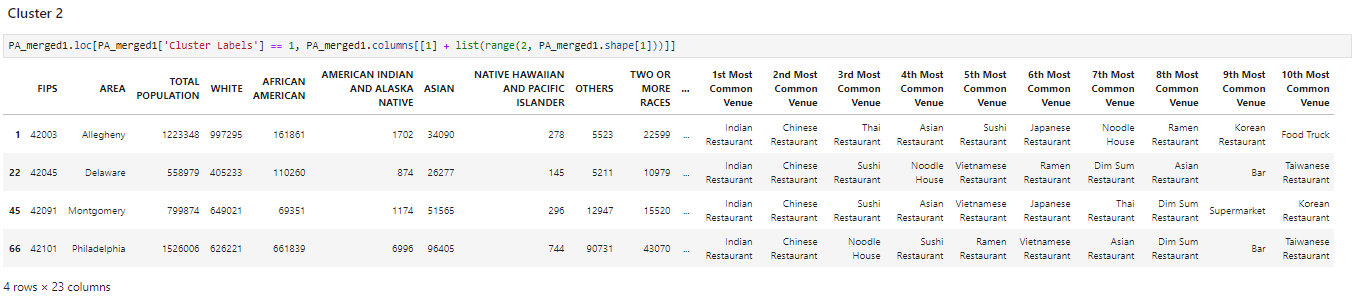
Lets plot the clusters in PA map.

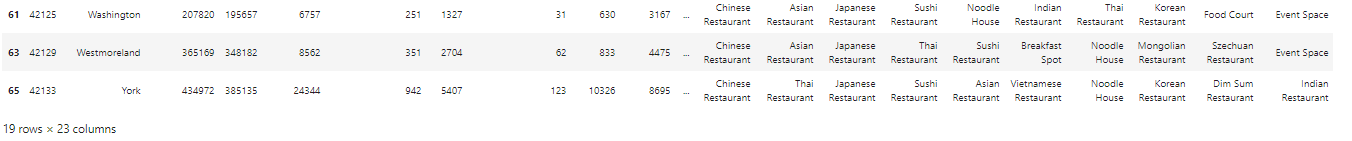
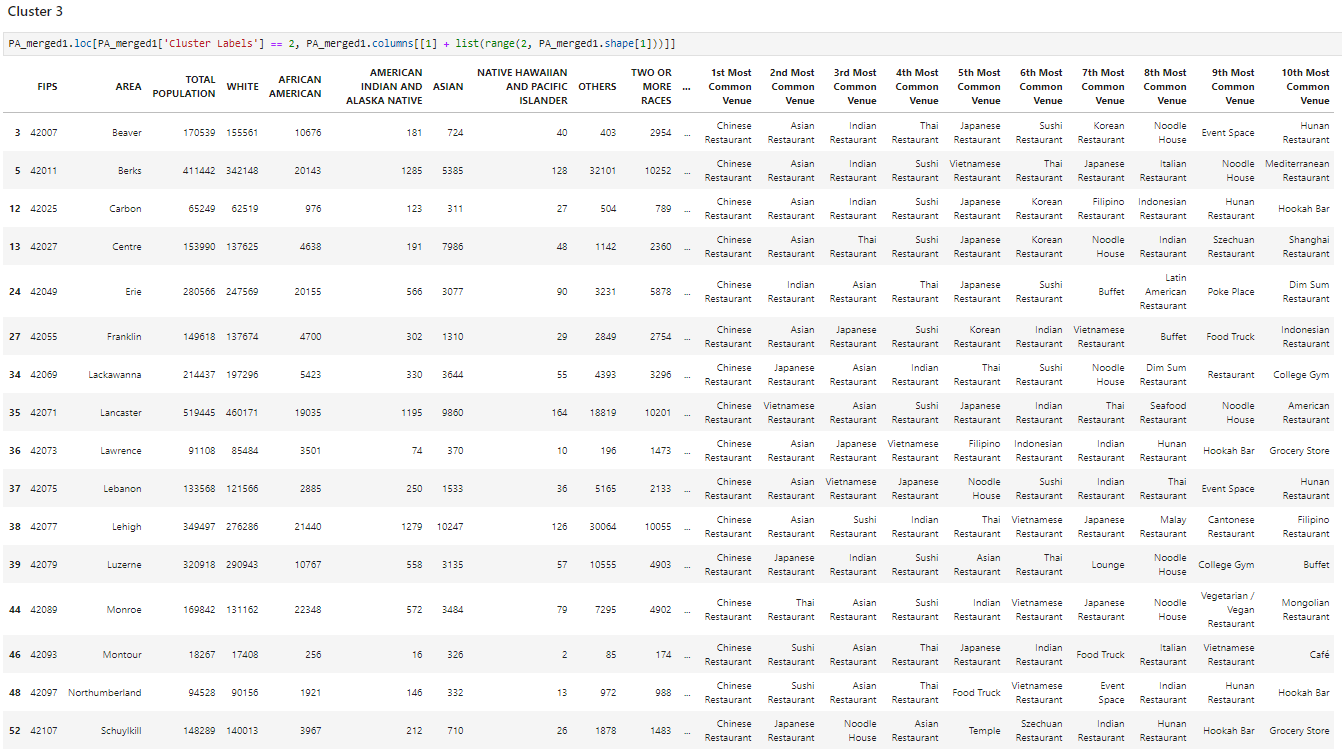




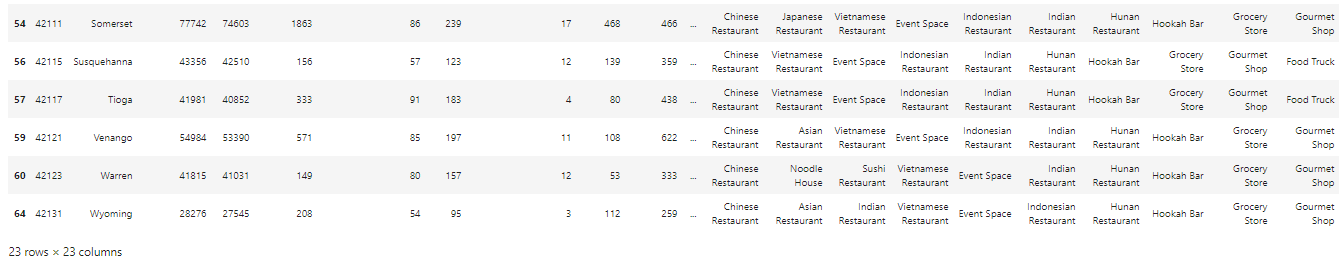
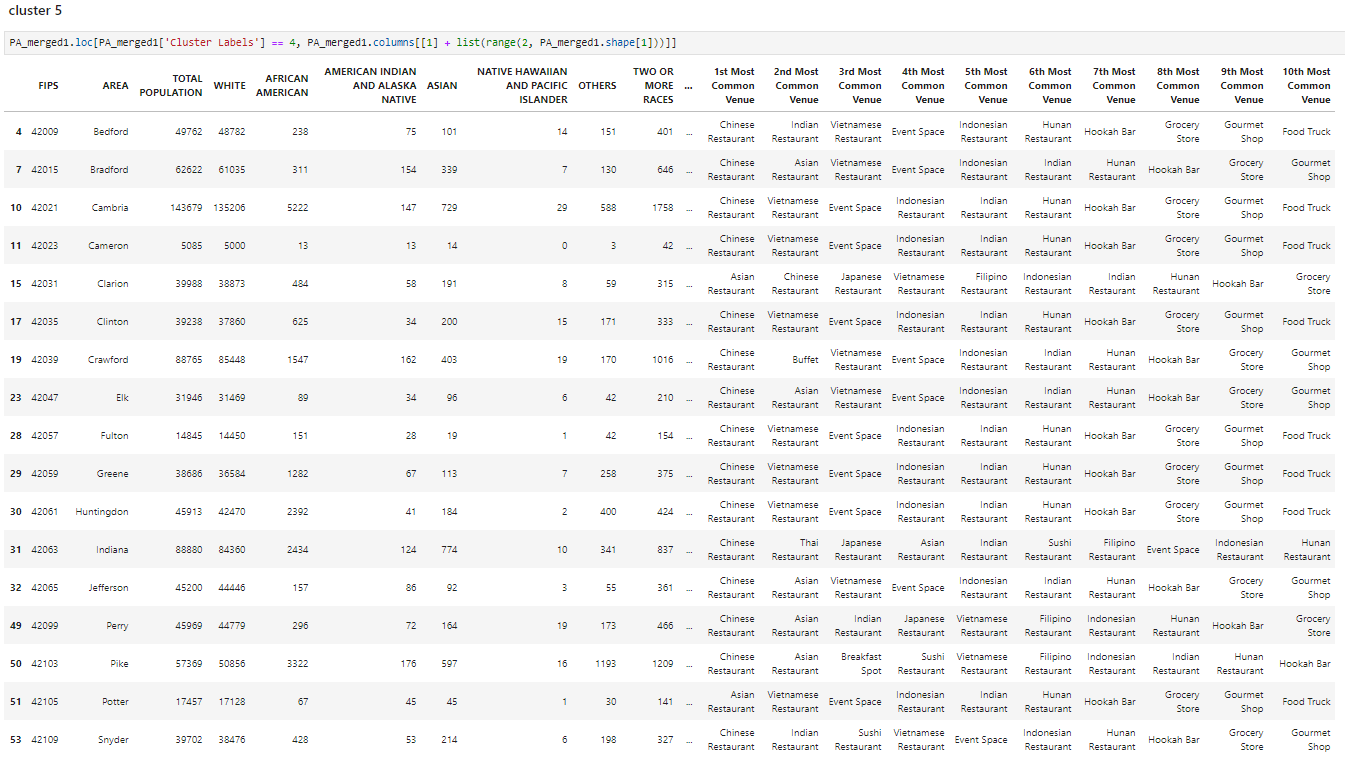
Examine clusters











**Insights of analysis**

* The top 5 Asian restaurants are
  + Chinese Restaurant - 764
  + Indian Restaurant - 305
  + Asian Restaurant - 228
  + Sushi Restaurant - 142
  + Japanese Restaurant - 127
* Chinese and Indian restaurants are most popular Asian cuisines.
* There are 40+ Indian restaurants on counties with more than 20k ASIAN population while the Chinese restaurants averages 11 on those counties.
* In counties with less than 20k population Chinese restaurants have upper hand and averages around 15 while Indian restaurants averages around 7.
* The Asian restaurants are popular even in the counties where the Asian population is less than a few hundreds.
* Indian restaurant counts 50 which is the highest in all counties by an Asian restaurant.
* Chinese restaurant counts highest in Pennsylvania holding 40% of all Asian restaurants followed by Indian restaurants with 20% share.
* Two least populated counties Forest and Sullivan have no Asian restaurants.
* The counties in cluster 1 and cluster 3 were currently dominated by Chinese restaurants, but promises better opportunity for Indian, Japanese and Sushi restaurants.
* The counties in cluster 2 and cluster 4 were currently dominated by Indian restaurants followed by Chinese restaurants.
* Asian cuisines are popular at counties in cluster 5 irrespective of its least Asian population.

**Conclusion**

Chinese and Indian restaurants are the most popular Asian restaurants in Pennsylvania. With a least projection of 10 percentage increase in total population on various counties, there are more opportunities for Indian, Japanese and Sushi restaurants.

This analysis can be further enhanced with population projection, food preference surveys to find best county to start any specific restaurant.

**References**

* <https://harrisburg.psu.edu/>
* <https://www.pasda.psu.edu/>
* <https://foursquare.com/>