Advanced Data Science Capstone

Opening a restaurant in San Francisco

Use Case

- There is always a demand for restaurants to open in San Francisco as there is a large variety of cuisines from around the world that people are willing to try.
- San Francisco is a large world class destination in the entire world and the USA with attractions such as the Golden Gate Bridge.
- It is the largest tech hub in the USA with around 800,000 people living in it and with a decent amount of public transportation such as BART and MUNI.
- The vast amount of residents are very high income, well educated, diverse and mostly work in the tech industry.

Data Set

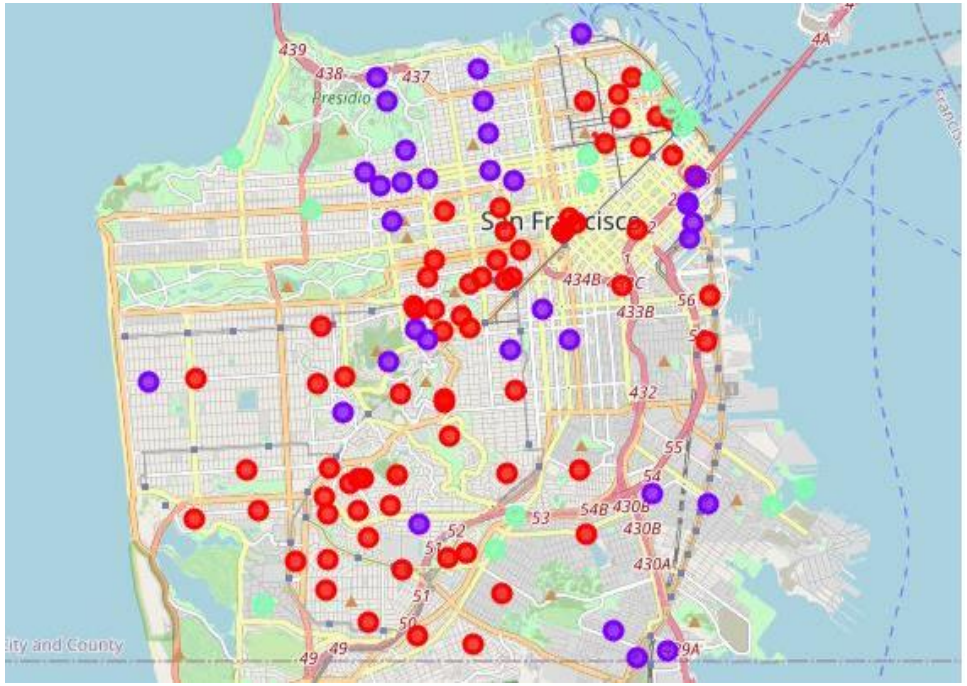
- The list of neighborhoods was gathered from Wikipedia.
- The list of venues for each venue was gathered from the Foursquare API.
- The coordinates for each neighborhood was gathered using geocoder.
- The python code was written in Jupyter Notebook using Anaconda and used built in libraries such as Pandas, Numpy, Matplotlib, and SKLearn as well as BeautifulSoup, Geocoder, and Folium.

Data Quality Assessment

All the coordinates appears to be located accurately on the map, so the data is extremely accurate, and the neighborhood coordinates are attached to the venue locations properly.

Data Exploration

Data Visualization



Feature Engineering

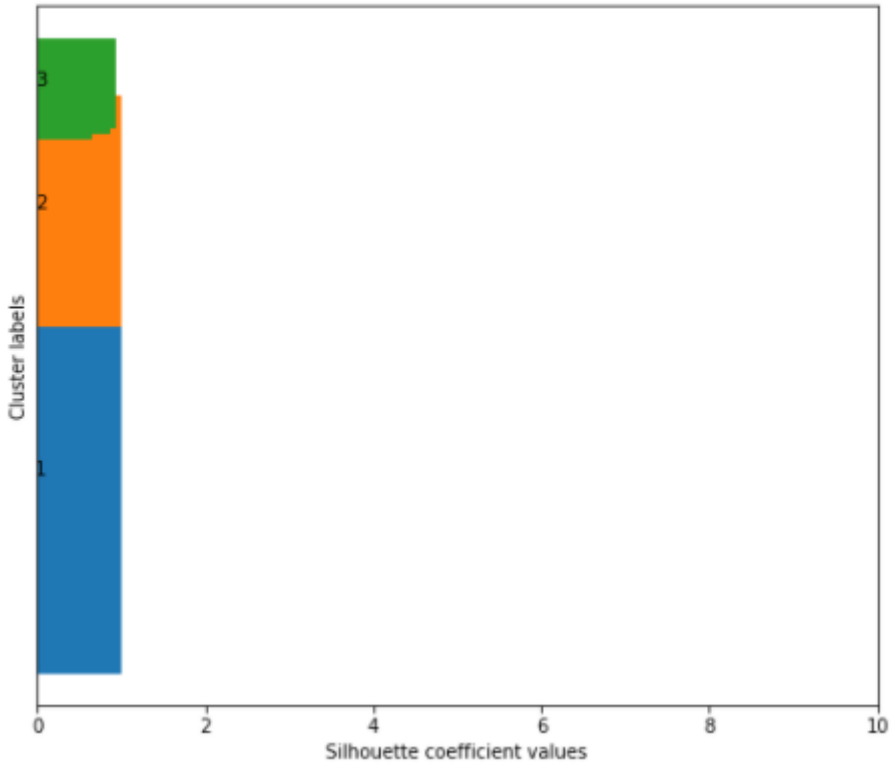
- One Hot Encoding

	Neighborhoods	Zoo	Zoo Exhibit	Accessories Store	Adult Boutique	African Restaurant	Alternative Healer	American Restaurant	Antique Shop	Aquarium	Arcade	Argentinian Restaurant	Art Gallery
0	1 Alamo Square	0.00	0.00	0.00	0.00	0.000000	0.00	0.010000	0.00	0.00	0.00	0.00	0.000000
1	10 Butchertown (Old and New)	0.00	0.00	0.00	0.00	0.000000	0.00	0.010000	0.00	0.00	0.00	0.00	0.010000
2	100 South End	0.00	0.00	0.00	0.00	0.000000	0.00	0.000000	0.00	0.00	0.00	0.00	0.010000
3	101 South of Market	0.00	0.00	0.00	0.00	0.000000	0.00	0.000000	0.00	0.00	0.00	0.00	0.010000
4	102 South Park	0.00	0.00	0.00	0.00	0.000000	0.00	0.000000	0.00	0.00	0.00	0.00	0.010000
5	103 Sunnydale	0.00	0.00	0.00	0.00	0.000000	0.00	0.022222	0.00	0.00	0.00	0.00	0.000000
6	104 Sunnyside	0.00	0.00	0.00	0.00	0.000000	0.01	0.000000	0.00	0.00	0.00	0.00	0.000000

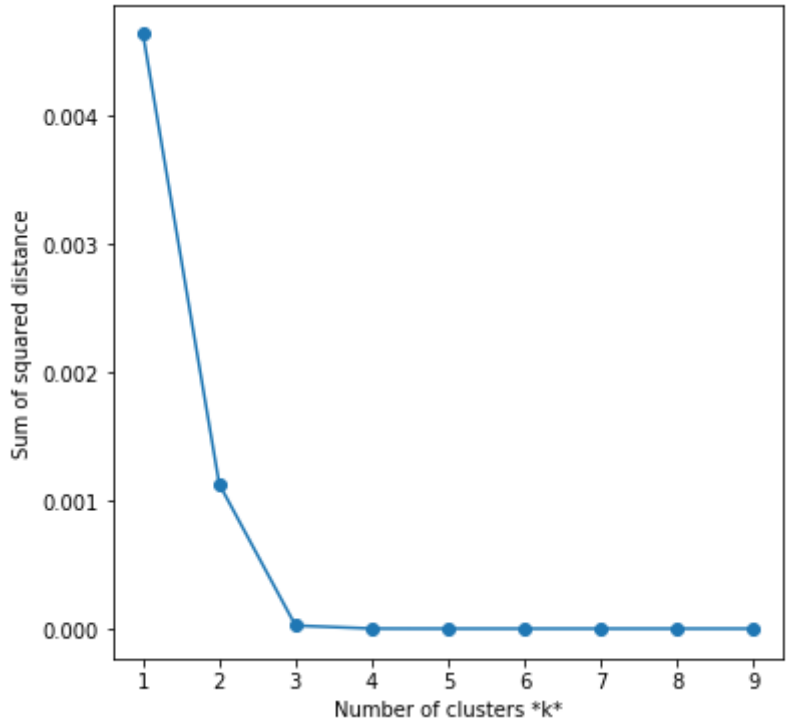
Machine Learning Algorithm

- K means clustering with 3 clusters and no deep learning algorithm required because this is an unsupervised clustering and using neural networks such as hierarchical clustering or autoencoders would be difficult.
- BeautifulSoup was used to parse the html file of the Wikipedia page for the list of SF neighborhoods to get the list of neighborhoods of SF and put it in a Pandas data frame for analysis.
- All the Python code was written in Jupyter Notebook.
- Geocoder was used to obtain the coordinates of the SF neighborhoods.
- The maps of the coordinate and clusters was made using Folium.
- Kmeans Clustering using SKLearn was used to create the clusters and analyze which neighborhoods had the most restaurants in which clusters.
- The foursquare API was used to find the top 10 most popular venues for each neighborhood of SF.

Model Performance Indicator and Evaluation



Model Performance Indicator and Evaluation



Model Performance Indicator and Evaluation

- The Calinski-Harabasz index of 11481.744 shows how well the clusters are separated and the high number means the clusters are dense and well separated and the number was calculated very fast.

Results

- The most restaurants appear to be in the green cluster which is located near the ocean and parks.
- The least amount of restaurants are in the middle neighborhoods of SF which are mostly residential in the red clusters.
- The best place to open a new restaurants would be in the red cluster where there are the least amount of restaurants.
- The foursquare api was used to calculate the top 10 most popular venue for each neighborhood in SF.
- The best place to open up a restaurant in San Francisco is in the residential areas which are not in the tourist areas because there is a large competition for restaurants to operate over there and residents of SF would like to eat in restaurants which are close by and would like to enjoy a large variety of cuisines.

References

- https://en.wikipedia.org/wiki/List_of_neighborhoods_in_San_Francisco
- https://en.wikipedia.org/wiki/San_Francisco
- https://www.niche.com/places-to-live/san-francisco-san-francisco-ca/
- https://developer.foursquare.com/