**Locations to Open Two New Chain Restaurants**

**1. Introduction**

**1.1 Problem location**

The business of new restaurant “Sunny Lantern” is very successful at Coventry, in the State of Rhode Island. There is funding to invest two more “Sunny Lantern” chain restaurants in other towns of Rhode Island.

**1.2 Who would be interested in this project**

Stakeholders for this invest need help from Data Science to identify the two cities that the restaurants will be. The two cities will have same neighborhood as Coventry. This neighborhoods have high chance to make sure the two chain restaurants as popular as it in Coventry, RI.

**2. Data Source**

Data obtained from “https://www.unitedstateszipcodes.org/mo/” include all cities in the United States, the coordinate of each city, and which state these cities belong to. For example, Coventry, coordinate “41.68, -71.66”, and it is in the state of Rhode Island. Foursquare location data will be used to explore neighborhood of cities in in Rhode Island. For example, the communities in Coventry include Bar, Yoga Studio and so on. The cities in Rhode Island also have the similar communities as in Coventry will be selected to open “Sunny Lantern” chain restraint.

**3. Methodology**

**3.1 Download packages**

The package of “numpy” was installed to arrange data for data frame. Package of “pandas” were installed to read the csv data into data frame and process data. Nominatim from geocoders was used to look up geographic coordinate of the city, state, or country we are interested in, such as the State of Rhode Island. Folium was installed to show map of Rhode Island and labels on map. The packages from “matplotlib” such as cm, colors, and pylot were used to show the colors of different clusters on map of Rhode Island. Request package was used to get the nearby venues from Foursquare. KMeans from Sklearn.cluster was installed to make clusters according to nearby venues of each city in Rhode Island.

**3.2 Data Cleaning**

Data were downloaded from original source “https://www.unitedstateszipcodes.org/mo/” as an csv file and then were read into data frame (“df”) in python. “df” includes columns of “Postcode”, “Neighborhood”, “Borough”, “Latitude”, and “Longitude”. “df” were cleaned and only included the data regarding the state (“Borough” as column) of Rhode Island. The cleaned data were read to new data frame “RI\_df” and contained 95 rows.

**3.2 Map of Rhode Island**

Coordinate of Rhode Island was obtained through “Nominatim”. Folium map, RI\_map were created according to the coordinate and each city is labeled with blue before clustering.

**3.3 Obtain the neighborhood information**

Foursqure account was created with “Client\_ID” and “Client\_Secret”. The function of “getNearbyVenues” was created to obtain the nearby venues of each city in Rhode Island according to their coordinate in the data frame of “RI\_df”. Nearby venues were read into new data frame “RI\_venues”. The Venue Categories were collected and grouped from “RI\_venues” by each neighborhood. Top 8 and top 10 most common venues for each neighborhood were ranked and collected in new data frame “neighbourhoods\_venues\_sorted”.

**3.4 K-cluster**

“K-means clustering” were used for machine learning to generate eight clusters. The cluster labels were added to data frame RI\_merged by using the function of “insert” and “join”. The “Neighborhood” without cluster label was discarded by using “dropna”. The neighborhoods in same cluster were pickup out. There are 8 groups of neighborhoods in total.

**3.5 Map cluster**

Cluster labels that are in the format of float were converted to integer by “astype”, because color scheme did not recognized float but integer. Then color scheme for the clusters were set with unique colors. These unique colors were applied to the label of Rhode Island map.

**4. Result**

All packages were successfully installed. All data belong to Rhode Island were picked and these “Neighborhood” were mapped to the state of Rhode Island before clustering (Fig.1).

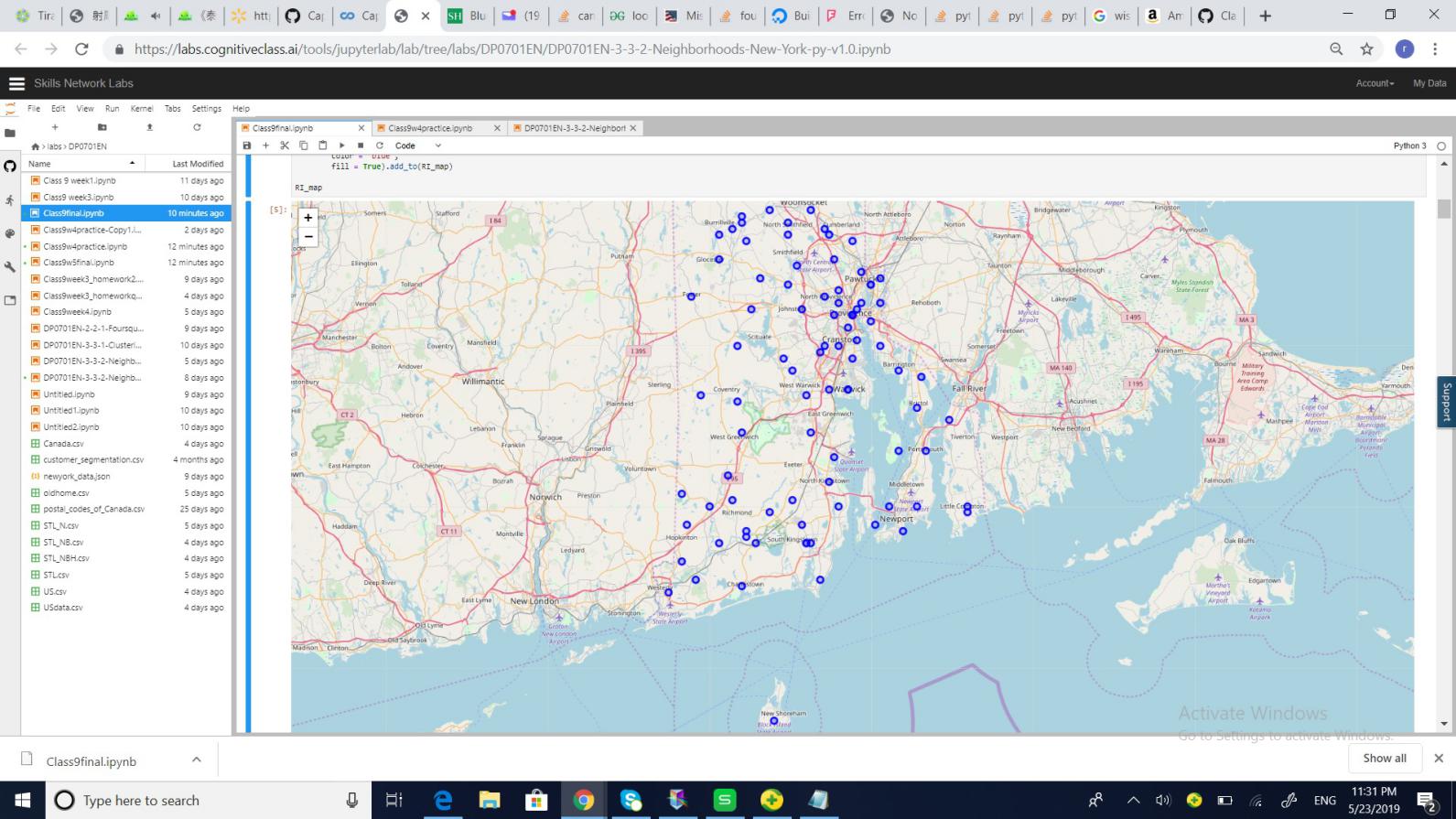


Figure. 1 Map of Rhode Island labeled with each city.

Venues of neighborhood for each city in Rhode Island were successfully obtained from Foursqure. Venues with 740 values and 44 neighborhoods in total were returned from Foursque for neighborhood of Rhode Island including the columns of the nearby venue, their coordinates, as well as venue categories. There are 150 columns with unique venue categories. The venue categories with same neighborhood were grouped and averaged. Top eight venues were picked for each neighborhood according to their frequencies and printed out. Results are verified but not in the data frame format. Top ten venues for each neighborhood were collected and read into “neighbourhoods\_venues\_sorted” with 44 rows and 11 columns. There are 8 clusters applied to these 44 neighborhoods. Cluster labels were added to “neighbourhoods\_venues\_sorted” and generated new data frame “RI\_merged”. However, the neighborhood information for some cities are not available, so the cluster labels showed “NaN”. Therefore, the neighborhood with “NaN” were dropped. Then all 8 clusters were demonstrated. For example, the top eight venues of Coventry are Bar, Yoga Studio , Diner, Fast Food Restaurant, Farmers Market, Falafel Restaurant, Donut Shop, Dive Bar, Discount Store, and Dessert Shop. Map of Rhoda Island with cluster labels were obtained (Fig. 2).

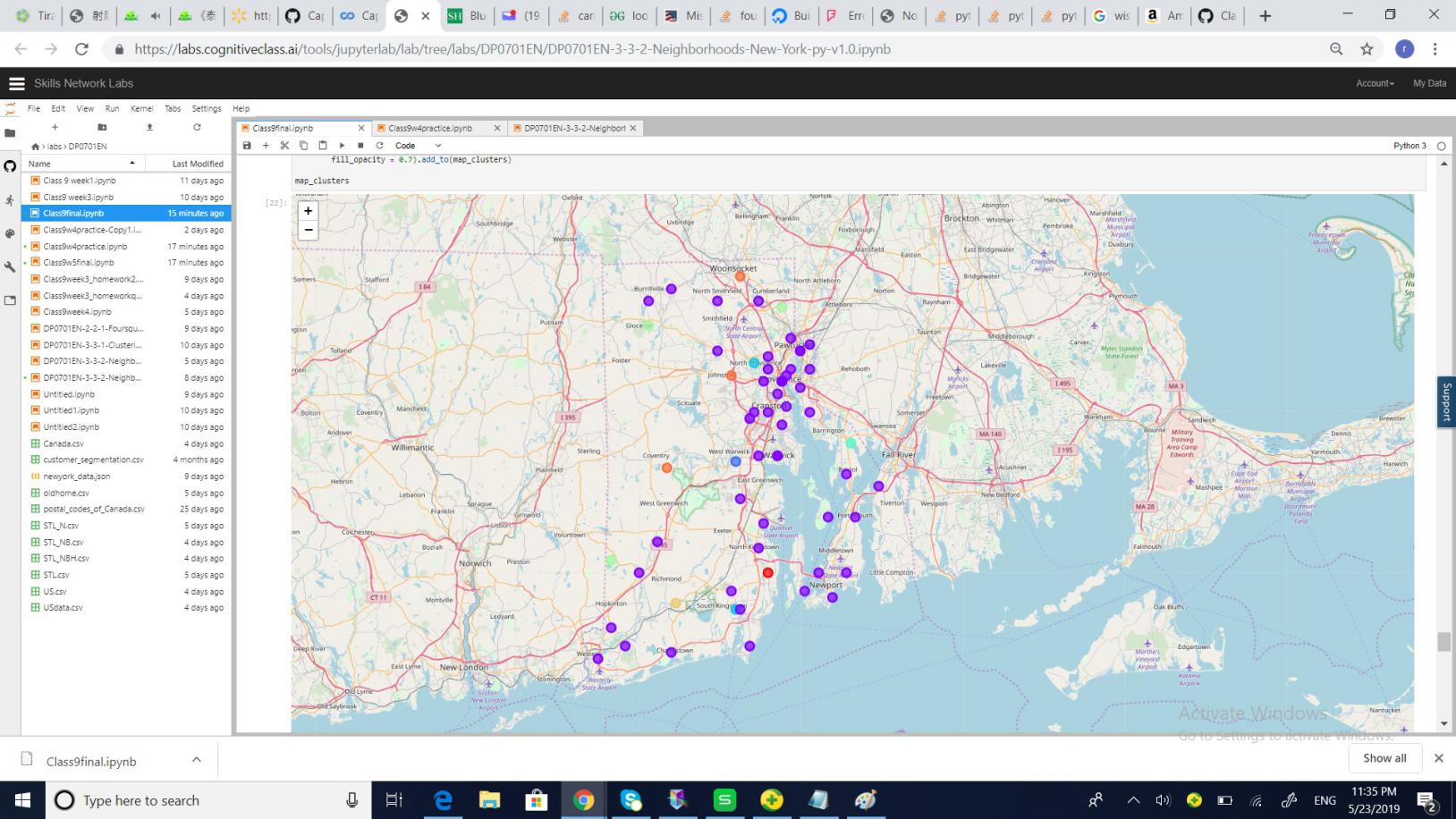


Figure. 2 Map of Rhode Island labeled with cluster labels.

**5. Discussion**

The two cities fall in the same cluster as Coventry are Woonsocket and Johnston. The cities to open new chain restaurants are Woonsocket and Johnston, since they have the similar neighborhood as Coventry.

**6. Conclusion**

This report successfully explored all cities in Rhoda Island in addition to the city for current successful restaurant, Coventry. Woonsocket and Johnston were recommended to open new restaurant of “Sunny Lantern”.