

# Identifying Turkish Restaurant Traffic in Seattle

David V Bryan

16 March 2021

## 1. Introduction

### 1.1 Background

The restaurant business is highly competitive; most restaurants close their doors within the first few years. Location is one of the most important factors, as it carries a large weight in the quantity of customers and many expenses of the business.

### 1.2 Problem

For this project, it will be necessary to identify which Turkish restaurants in Seattle get the most customers. Accomplishing this will require learning what restaurants are in the area and a general idea of how much traffic they bring in.

### 1.3 Interest

This information would be useful to anyone looking to open a new restaurant of this type in Seattle. The methodology could also be expanded and used for any new physical business in any location.

## 2. Data

### 2.1 Data Used

#### FourSquare Data:

- Restaurant Categories

- List of Restaurants in Seattle (within designated categories)

- Details of Each of Restaurant

- Location (Longitude, Latitude)

#### Yelp Data:

- Number of Reviews / Restaurant

## 2.2 Data Cleaning:

The data obtained through FourSquare and Yelp is filtered and added to a list called **res\_data**. This is a list of restaurants containing the following information for each:

[restaurant\_latitude, restaurant\_longitude, restaurant\_name, kmeans\_cluster, average\_reviews]

## 3. Methodology

### 3.1 Restaurant Categories

After obtaining a list of categories from FourSquare, categories that are likely to serve Turkish food will need to be identified manually. The FourSquare API returns a list of dictionaries like the following:

```
{'id': '4f04af1f2fb6e1c99f3db0bb',  
  'name': 'Turkish Restaurant',  
  'pluralName': 'Turkish Restaurants',  
  'shortName': 'Turkish',  
  'icon': {'prefix':  
'https://ss3.4sqi.net/img/categories_v2/food/turkish_',  
  'suffix': '.png'},  
  'categories': [{'id': '530faca9bcb57f1066bc2f3',  
    'name': 'Borek Place',  
    'pluralName': 'Borek Places',  
    'shortName': 'Borek',  
    'icon': {'prefix':  
'https://ss3.4sqi.net/img/categories_v2/food/turkish_',  
    'suffix': '.png'},  
    'categories': []}]}
```

For this project, the following FourSquare restaurant categories are used:

*Mediterranean Restaurant*  
*Middle Eastern Restaurant*  
*Turkish Restaurant*

### 3.2 List of Restaurants

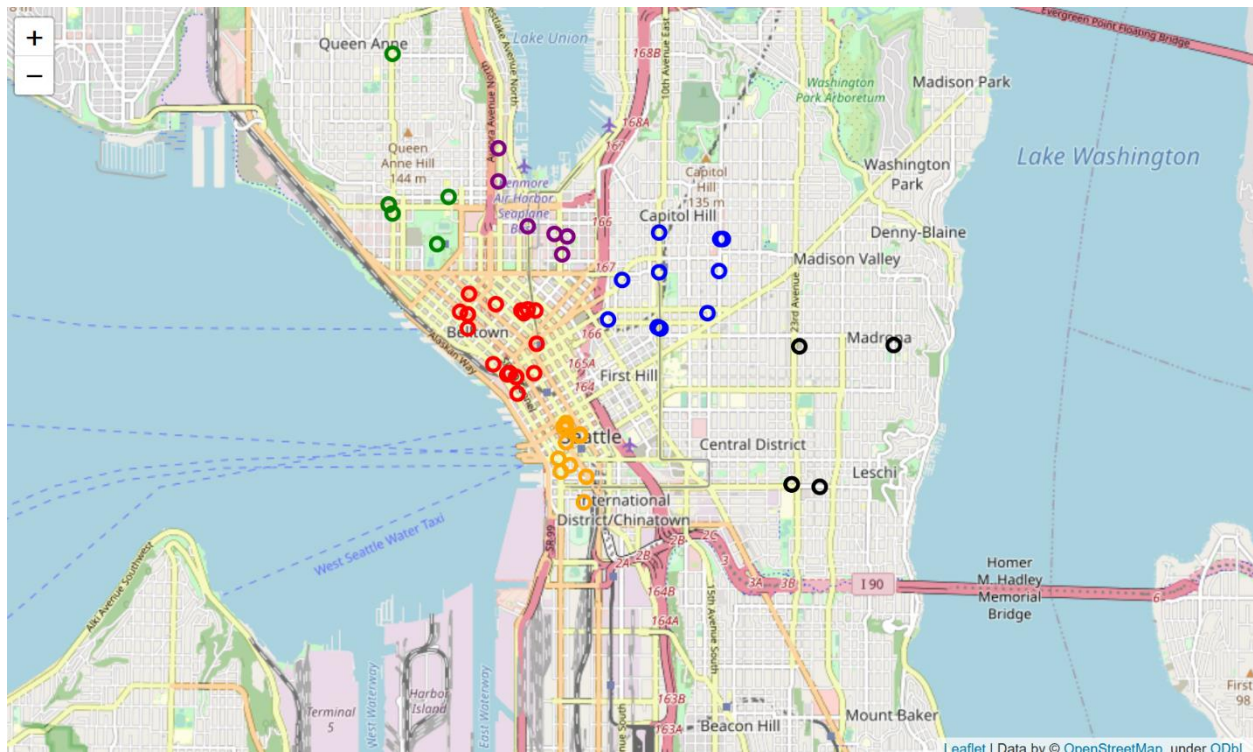
The ids of the categories are then used to make API calls to FourSquare in order to obtain the **Name** and **Location** of each restaurant in the area.

Using the category id to pull a list of restaurants results in a list of restaurants like the following:

```
{
  'id': '5112de23e4b0fe8b80b23fbe',
  'name': 'Petra Mediterranean Bistro',
  'location': {
    'address': '1933 7th Ave',
    'crossStreet': 'Westlake',
    'lat': 47.6151873930063,
    'lng': -122.33761837545067,
    'labeledLatLngs': [
      {
        'label': 'display',
        'lat': 47.6151873930063,
        'lng': -122.33761837545067
      },
      {
        'label': 'entrance',
        'lat': 47.614817,
        'lng': -122.337268
      }
    ],
    'distance': 1178,
    'postalCode': '98101',
    'cc': 'US',
    'city': 'Seattle',
    'state': 'WA',
    'country': 'United States',
    'formattedAddress': [
      '1933 7th Ave (Westlake)',
      'Seattle, WA 98101',
      'United States'
    ]
  }
}
```

### 3.3 Location Clusters

The location data is then used to sort the list of restaurants into **six clusters** using KMeans, which appears as such on the map:



### 3.4 Reviews per Restaurant per Cluster

The name of each restaurant is then used to find it on Yelp and retrieve the number of reviews.

The total number of reviews for all restaurants in each cluster is then divided by the number of restaurants to obtain the average number of reviews per restaurant in each cluster.

## 4. Results

Calculating the average # of restaurant reviews for the six clusters results in the following:

Color	Avg_Reviews
green	292
red	401
blue	294
orange	142
purple	120
black	181

## 5. Discussion

It can be seen that restaurants in the downtown area, despite there being more restaurants, obtain a greater number of reviews than other areas. **However, this can be misleading at first glance.** Although there is increased traffic at these locations, there is bound to be increased costs associated as well, such as rental and insurance rates.

It may come as a surprise that **restaurants in the blue cluster on average received twice as many reviews as those in the orange cluster.**

Upon a closer inspection, this may be due to a combination of two factors: the lower density of restaurants in the blue area, and the high number of middle-class residences in that area. There

may be less competition and more customers per restaurant, indicating a potentially good location for business.

There are obviously other factors that need to be considered for an informed decision, such as rental and insurance costs for starters. This project provides a piece of the information though that should be evaluated when making a decision.

## **6. Conclusion**

To verify or refute these findings, it would be necessary to develop a more in-depth model that further refines the sorting of locations. It would also be beneficial to implement factors such as rental costs, crime rate, and income of customers. Doing so would be extremely time intensive though and is therefore beyond the scope of this project.