Introduction

Background

The case study which I am going to present is the restaurant industry. This is an evergreen industry. There won't be a dearth of restaurants in any place. The expertise in starting a restaurant are aplenty. The issue under consideration is to help anybody who is going to start a restaurant in any city. The city being considered here is the Kansas city. The current restaurant details are available and that can be used to predict and help the customer about feasibility and profitability of the business. There are many aspects that we can think of when considering this problem: Location, type of restaurant, menu offered, price list, type of employees, age group of employees, chefs etc. There are many to consider.

Problem

The customer want to start a restaurant in a given City(Kansas here). We as data scientist need to help the customer to help find the best location. We need to find the type and location for the customer. We can take into consideration different aspects: like menu, location, target customers etc. The issue we consider here is only the aspect of the location and type of the restaurant(whether it s Japanese, Italian, Thai, Asian etc). We need to analyze the competition as well. Need to find the neighboring restaurants and their types and need to confirm if this is a good fit for the customer to start the business.

Data Acquisition and Cleaning

Data Source

The Data for the issue is obtained from the following website,

https://data.world/datasets/restaurants

This describes the different location ins Kansas city. This is an exhaustive list of all places in the area. The data under consideration is only for Kansas city. The data is in the form of json file and it is uploaded into the Watson Studio,

Data Cleaning

The data needs some clean up required which will be done using the pandas. We will use different data frame techniques to correct the data. Will be filtering out the records it the restaurant stopped operation recently.

Feature Selection

The features selected will be name and the coordinates of the location. This together with the foursqaure dara can be used for further exploration.

## Exploratory Data Analysis

Predictive Modelling

This is the classic case of usage of K means Clustering algorithm. Clustering is a method of unsupervised learning and is a common technique for statistical data analysis used in this case. We apply a clustering algorithm, to get some valuable insight into the data. Along with K-Means clustering we will be using the foursquare data to find out the location coordinates as well.

Analyse each neighborhood with the results obtained.

Results

The criteria for the final selection depends on the customer. What this project shows the customer is to help him decide the preference of location. Some of the customer want the restaurant to be opened where other restaurants are also available. Some want to open it in a place where there are no other restaurants. Some of them wat to decide the type by looking at the types of the other restaurants. That can be decided by yet another project considering the people frequenting the area of interest. This project give an insight into the location and types of venues around in (Kansas)

First map the locations on the Map using Folium package from the location jsaon file.
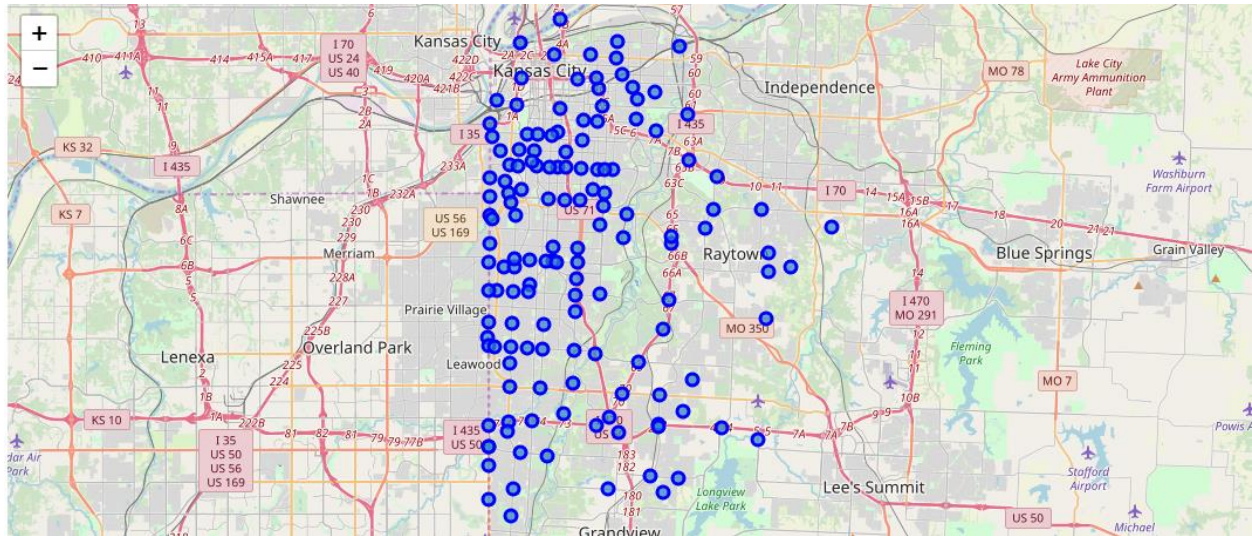
Then used the foursquare API calls to find out the categories of establishments in the neighborhoods.

	Neighborhood	ATM	Airport Terminal	American Restaurant	Antique Shop	Art Gallery	Art Museum	Arts & Crafts Store	Assisted Living	Automotive Shop	BBQ Joint	Bakery	Bar	Baseball Field	Baseball Stadium	Basketball Court	Beer Bar	Beer Garden	Big Box Store
0	Verona Hills	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	Marlborough Heights-Marlborough Pride	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	Marlborough Heights-Marlborough Pride	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Marlborough Heights-Marlborough Pride	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	Marlborough Heights-Marlborough Pride	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Verona Hills	38.920223	-94.608006	The Barstow Cafeteria	38.918134	-94.605530	American Restaurant
1	Marlborough Heights-Marlborough Pride	38.973438	-94.576572	BB's Lawnside BBQ	38.972825	-94.574602	BBQ Joint
2	Marlborough Heights-Marlborough Pride	38.973438	-94.576572	Save-A-Lot	38.973401	-94.580415	Grocery Store
3	Marlborough Heights-Marlborough Pride	38.973438	-94.576572	Dollar General	38.974383	-94.581020	Discount Store
4	Marlborough Heights-Marlborough Pride	38.973438	-94.576572	China Star	38.973228	-94.579279	Chinese Restaurant

The build a data frame with all the details collected to find out the venues and how frequenting it has been.

```
----Armour Fields----
                 venue  freq
0  Other Great Outdoors  0.33
1          Beer Garden   0.33
2                 Café   0.33
3                  ATM   0.00
4            Pet Store   0.00


----Armour Hills----
                 venue  freq
0      Ice Cream Shop    0.2
1            Wine Bar    0.2
2               Trail    0.2
3       Breakfast Spot   0.2
4   American Restaurant  0.2
```

Then use pandas data frame to list out the common places in that order.

	Neighborhood	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Armour Fields	Café	Beer Garden	Other Great Outdoors	Wings Joint	Fast Food Restaurant	Exhibit	Event Space	Ethiopian Restaurant	Electronics Store	Dry Cleaner
1	Armour Hills	Wine Bar	American Restaurant	Ice Cream Shop	Trail	Breakfast Spot	Wings Joint	Drugstore	Filipino Restaurant	Fast Food Restaurant	Exhibit
2	Ashland Ridge	Sports Bar	Hotel	Outdoor Sculpture	Sandwich Place	Wings Joint	Dog Run	Event Space	Ethiopian Restaurant	Electronics Store	Dry Cleaner
3	Bannister Ares	Fast Food Restaurant	Sandwich Place	Discount Store	Liquor Store	Dog Run	Exhibit	Event Space	Ethiopian Restaurant	Electronics Store	Dry Cleaner
4	Beacon Hills	Fast Food Restaurant	American Restaurant	Juice Bar	Discount Store	Performing Arts Venue	Wings Joint	Donut Shop	Exhibit	Event Space	Ethiopian Restaurant

Then used the k means clustering algorithm to find the 5 cluster so that we can decide which cluster is to be chosen. Here I have used 5 cluster only.

Verona Hills & Marlborough Heights-Marlborough Pride are the best places to choose for a food place as seen obviously below.

	Name	Latitude	Longitude	Cluster Labels	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	6th Most Common Venue	7th Most Common Venue	8th Most Common Venue	9th Most Common Venue	10th Most Common Venue
0	Verona Hills	38.920223	-94.608006	1.0	American Restaurant	Wings Joint	Donut Shop	Filipino Restaurant	Fast Food Restaurant	Exhibit	Event Space	Ethiopian Restaurant	Electronics Store	Dry Cleaner
1	Marlborough Heights-Marlborough Pride	38.973438	-94.576572	2.0	Chinese Restaurant	Discount Store	Grocery Store	Bus Stop	BBQ Joint	Wings Joint	Drugstore	Fast Food Restaurant	Exhibit	Event Space
2	Oldham Farms	38.967442	-94.519940	1.0	Moving Target	Hotel	Gym / Fitness Center	Italian Restaurant	Donut Shop	Fast Food Restaurant	Exhibit	Event Space	Ethiopian Restaurant	Electronics Store
3	Ward Parkway Plaza	38.978758	-94.608447	0.0	Historic Site	Coffee Shop	Juice Bar	Chinese Restaurant	Convenience Store	Dessert Shop	Dry Cleaner	Filipino Restaurant	Fast Food Restaurant	Exhibit
4	Stanford Gardens	39.012761	-94.607761	1.0	Other Great Outdoors	Lake	Sculpture Garden	Wings Joint	Donut Shop	Exhibit	Event Space	Ethiopian Restaurant	Electronics Store	Dry Cleaner