Data Analysis

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 location of the restaurant, address and it's type. This is an exhaustive list of all restaurants in the area. The data under consideration is only for Kansas city. This is the initial consideration only. Later if I find something better I will make use of it.

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. The food outlets without in-dining option also will be removed from the data set. The other ones which should be removed is those without address.

Feature Selection

The features selected will be Address location and type of restaurant.

Exploratory Data Analysis

The target variables we use will be yes or no type.

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. Also relationship between different variables will be considered like price- chef, type-chef availability etc. When 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.