# Real Estate Prices & Venues Data Analysis of Lahore, Pakistan

## Muhammad Tayyab

**Introduction**

**Background**

Pakistan is a captivating place due to its geographical significance and Lahore is one of most popular city to live and start a business. In last few years world has realized its importance and start investing in it. Since 2010 many international companies from different sectors like Automobile, Gas & Oil, Agriculture, fashion and construction establishing their businesses in Pakistan. CPEC a multi-billion project with China is about to complete and after that world will be attract towards Pakistan and everyone will be in hurry to invest at the right time and at the right place.

**Business Problem**

As we can see this is the right time to analyze the trends and invest your money so you will get the best out of your decision using machine learning approaches. Here we will work on real state sector as this is always considered one of the most profit generated area. We will see which place is best to buy according to your need. For example if someone wants to buy a place at non commercial area than where should he invest his minimum money to possess maximum land. Or if someone wants to start a business than definitely he wants a place which is surrounded by many people having interest related to their business.

**Data Collection**

Data is the base of every experiment. So, we will use Zameen.com dataset which include property type, price, location, area, bathrooms, bedrooms and some other features. You can download the dataset from kaggle '<https://www.kaggle.com/huzzefakhan/zameencom-property-data-pakistan'>. The dataset have many columns and some of them are not required in our analysis. We will remove and manipulate our data according to our need.

For location, tips and Venue services we will use Foursquare API and find the trends in the specified area.

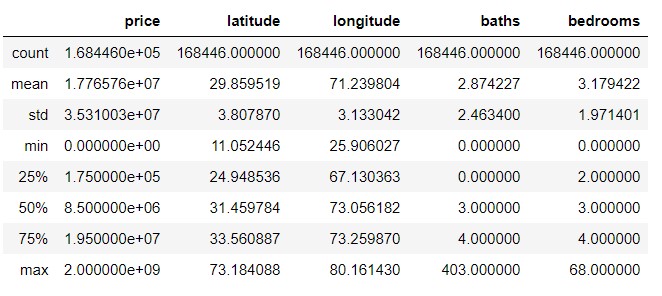
**Methodology section**

The Methodology section will describe the main components of our analysis and predication system. The Methodology section comprises four stages:

1. Collect Inspection Data
2. Explore and Understand Data
3. Data preparation and preprocessing
4. Modeling

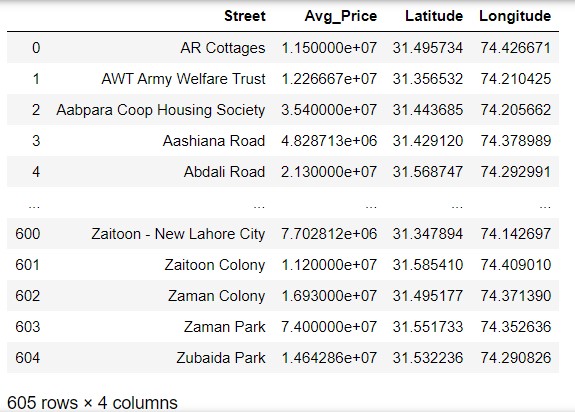
### 1. Collect and Exploring Data:

After importing the necessary libraries, i download the data from the Kaggle website and explore it.

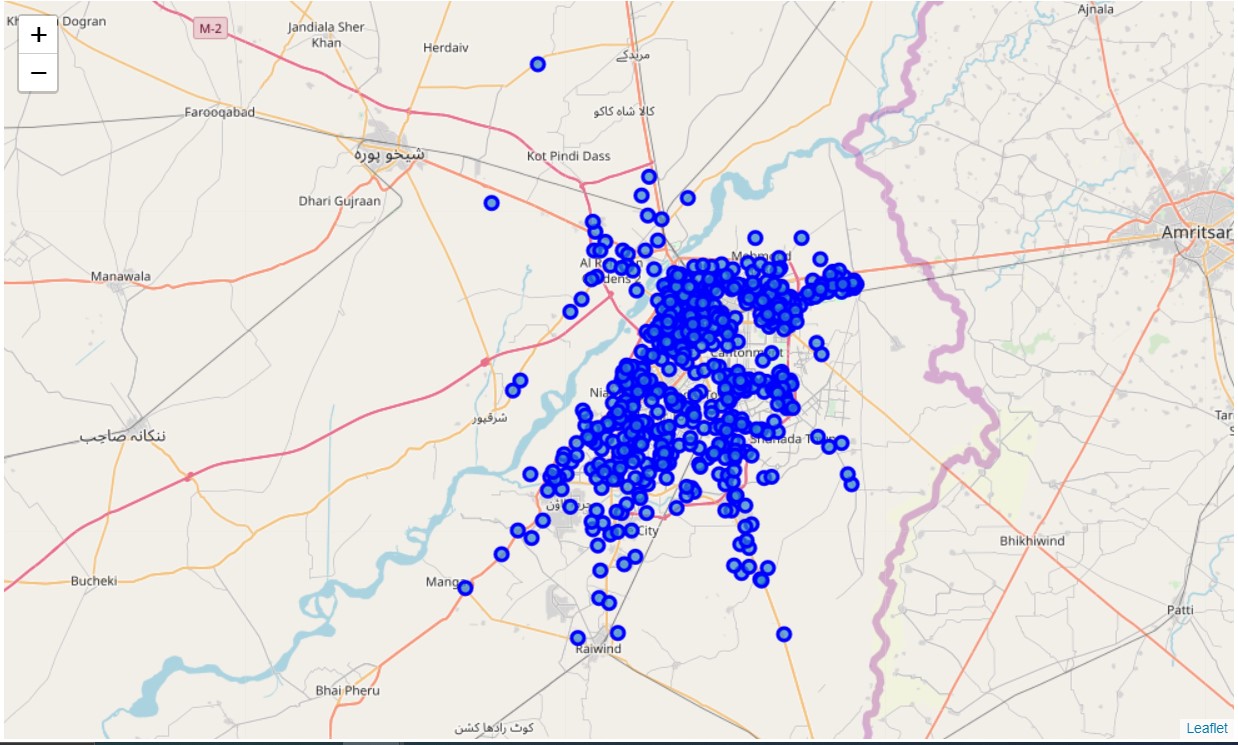


### 2. Data Preparation and Preprocessing:

I analyze that our data has outliers and some unnecessary columns. So i remove all the columns and outliers and set a threshold for price, property type and its purpose. The location was groupby and a mean price was placed in front of location along its coordinates.



After getting this data i used folium map to visualize the locations on map. I created a map of Lahore on top and used latitude and longitude values to get the visual as below:

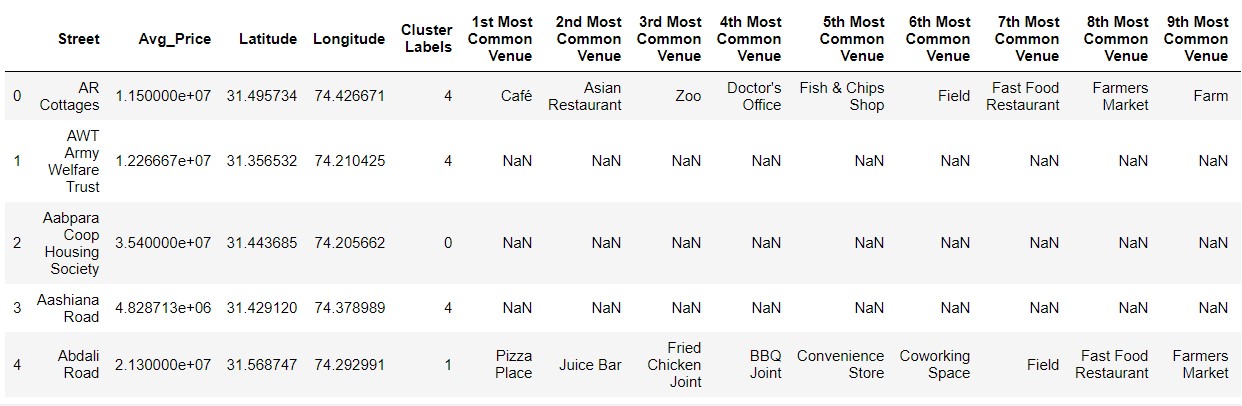


Using Foursquare API i generated **Top 10**trends in each location. The below table shows the result for a few location:



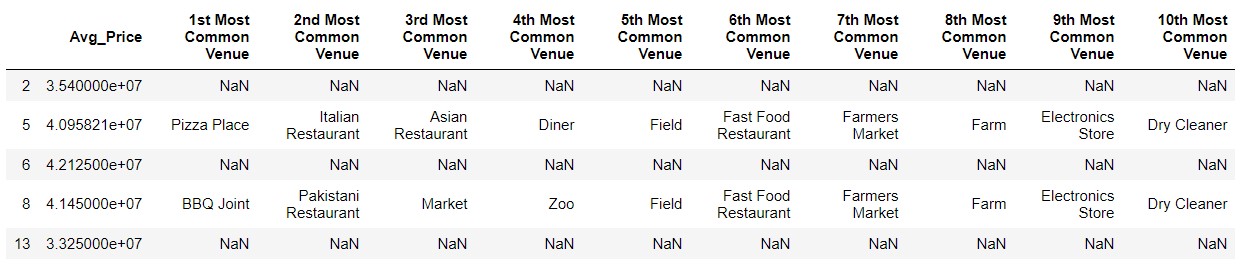
### 3. Modeling:

I have some common venue categories. In this reason I used unsupervised learning K-means algorithm to cluster the boroughs. K-Means algorithm is one of the most common cluster method of unsupervised learning. I will run K-Means to cluster the boroughs into 5 clusters and merge both lahore\_df and venue dataframe: (See Column #5)



Now we analyze each cluster separately.

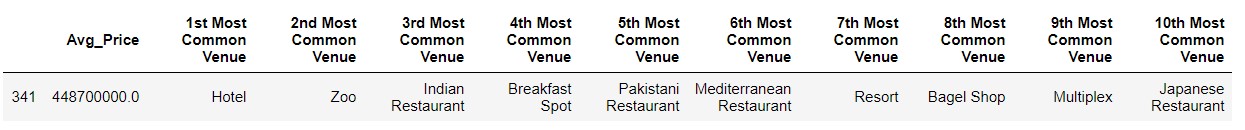
### Cluster 1



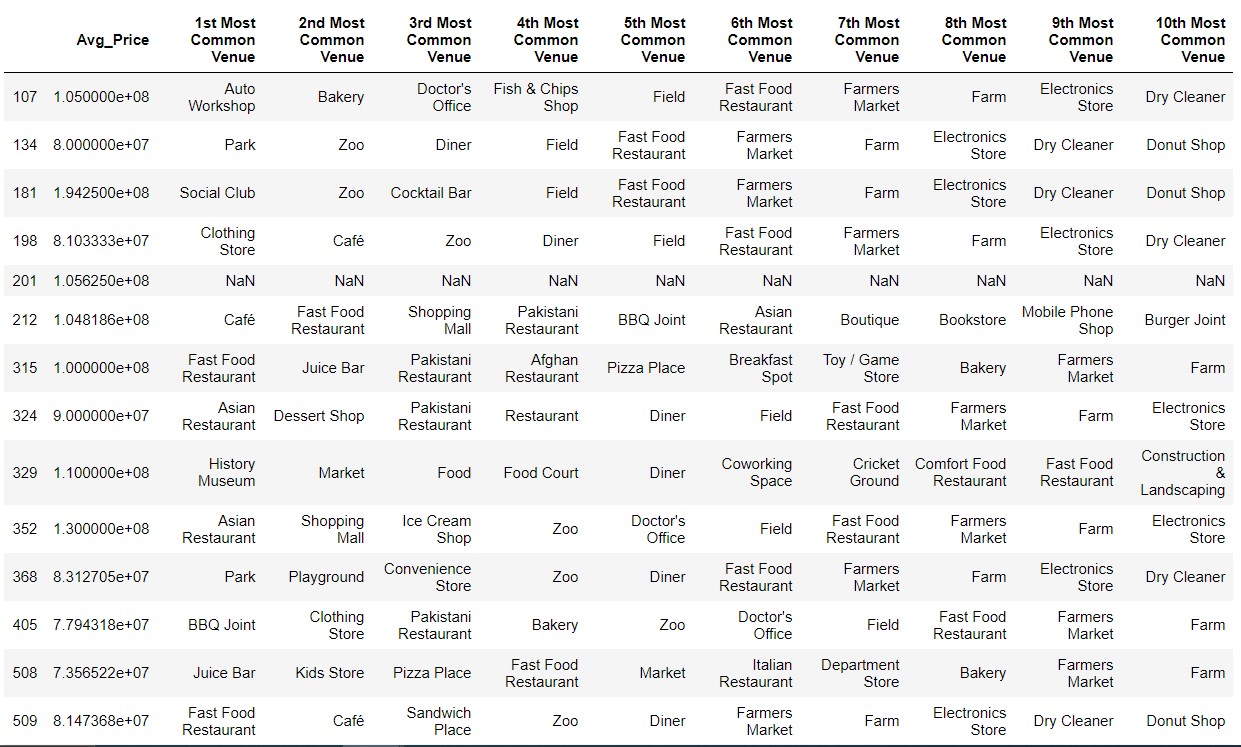
### Cluster 2



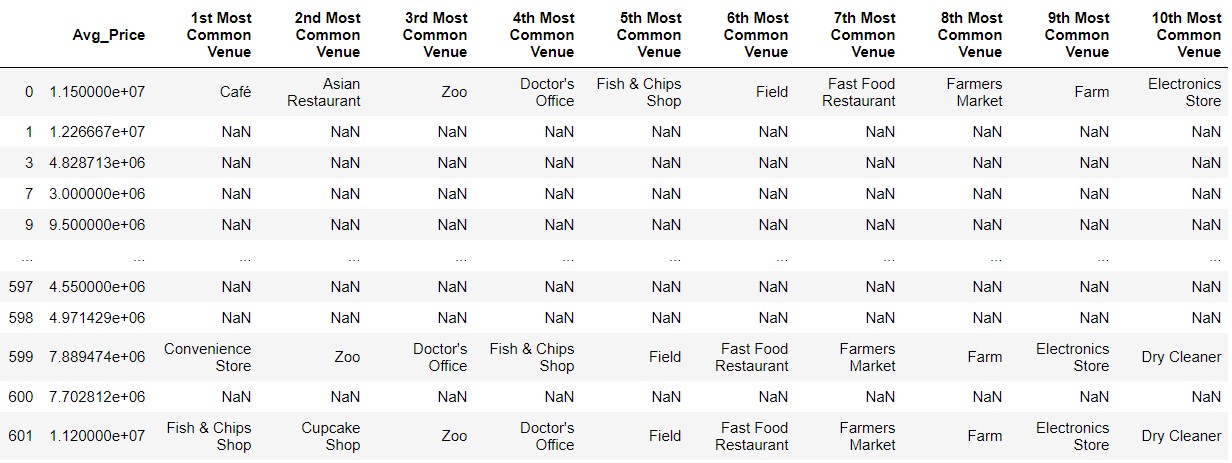
### Cluster 3



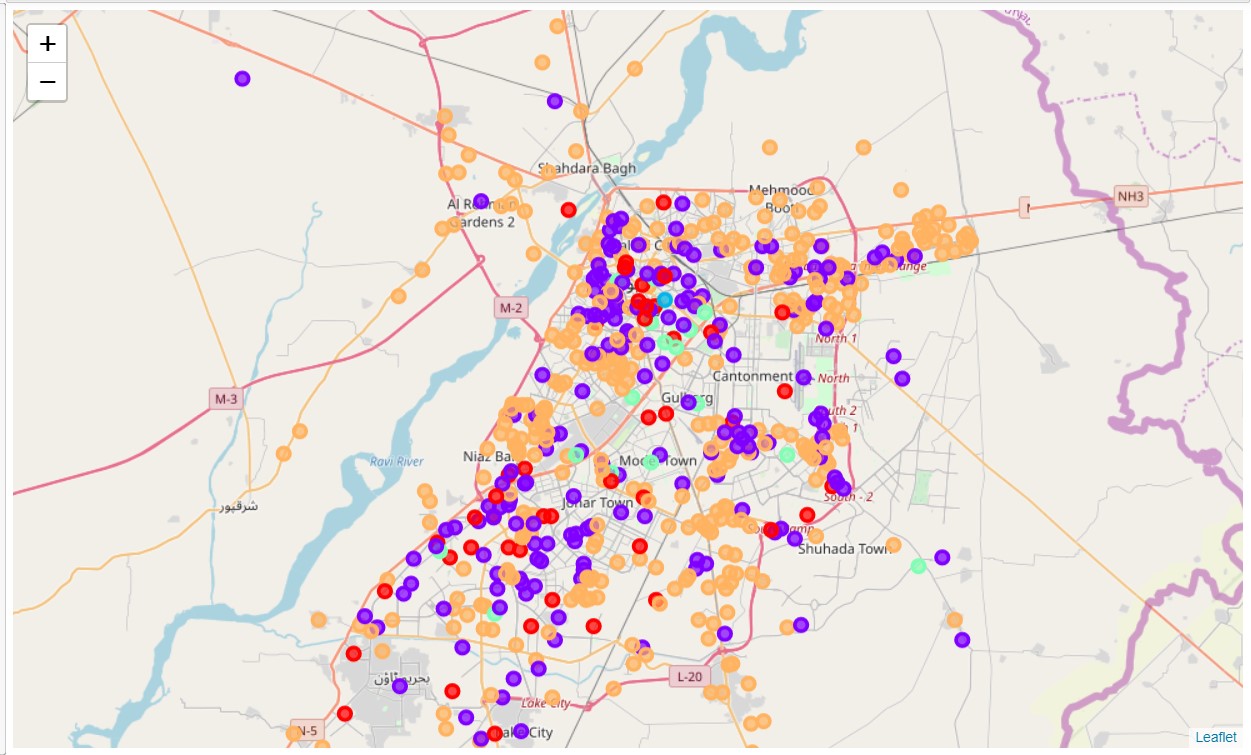
### Cluster 4



### Cluster 5



This is how our data is represented on map:



## Results

Finally we are done with our experiment and now it is the time to discuss our findings. We split our data into 5 big clusters. By analyzing that we can say that **Cluster 1** is good place to live or to start food business for those who want themselves surrounded by Food and cafes. You need average price from 29,900,000 to 67,500,000 PKR to buy 2 Marla to 6 Kanal place.

**Cluster 2** will be the best option for those who want entertainment (Cinemas, Multiplex, Art Gallery etc) as well as cannot compromise on shopping. You do not have to travel a lot to buy household stuff as you are surrounded by many supermarkets, stores, Book Shops etc. **Cluster 3** (Mall Road) is best and trendiest place for Hotel and Resort Businesses. Pakistan's most famous hotels are located on Mall Road e.g Pearl Continental, Avari Hotel and Flatties Hotel are the most renowned ones. So if you have a lot of money in other words averagely 450,000,000 PKR than you are all set to battle against these top notch hotel brands.

**Cluster 4 and 5** are the places which are usually the side areas and has less population and may attract those people who want to live in city without getting rid of luxury and also want to stick with nature but do not like the city traffic and noise pollution.