Coursera Capstone

IBM Applied Data Science Capstone

A Recommendation System for The Foreigners Want to Buy House in Pakistan

Abstract:

This project is basically a recommendation system for a foreigner. This system will analyze the Zameen.com provided data for the properties in different cities of Pakistan and then ask the client or foreigner what's his or her budget to buy a house and in which city he wants to buy. According to the Zameen.com there are total 168K houses , flats , rooms , upper and lower portions are for sale in different cities of Pakistan. Our system will first clean the data provided by the Zameen.com then make a cluster of the data in the asked city then based on the clusters which cluster will have lower average sale value than the system will use Foursquare API to analyze each house in particular cluster. Finally our recommendation will be based on number of important daily life venues like Banks, ATM, Schools,

Introduction:

There are total millions of houses, plats , separate rooms , upper portions etc are sale on different websites in Pakistan. Whenever a new foreigner wants to buy a house or for rent he or she will get confused where to look for the data or info. The best thing he or she can do is to ask any real state agent for the prices and details but asking a real state agent is not beneficial why? Because real state agent will have information only for houses available under his or hers supervision. And going to each of these real state agent is very difficult. So the best way is to get the data , clean it and then analyze it according to the instructions of that foreigner.

According to the data there are total 168k houses under Zameen.com including houses, plats , separate rooms , upper portions and also farm houses for rents and sale both. Prices may very if someone is desired to buy in a particular city.

In the case of audience, these foreigners may want to directly contact for the recommendations to the Real State agents and this project will be very helpful for each of these real state agents because now every agent will have every information of the houses in Pakistan. Beside this a Real state agent can also recommend to the local buyers and sellers. So, in our case Stakeholders will be Real State agents and audience will be local citizens and foreigners.

Foursquare API will be very useful because at the end of the project we have to analyze each houses based on the latitude and longitude present in the dataset. Our recommendation will purely based on the number of valuable venues like , Hotels , Resturants, Schools , ATM , Medical Stores , Masajids etc.

But as for the first step will be grouping our data based on the city to check which is have avg low prices and then will further dig down to include only house , flat or any other categories. After doing these analysis we will be using K-Means clustering to cluster our houses based on the prices and number of venues etc.

Data/Source:

The data that I'm using is available on the internet and based on the validation it's a valid data. It is provided by the Zameen.com one of the best digital property dealers in Pakistan as of now. The dataset contains many features of the houses. There is feature present area which is either kanal or Marla so our first attempt will be remove these units and then use only one unit. And futher more agency for each of the property is also present so we will only explore those houses where there is an agency otherwise we will delete our sample.

link for the dataset: https://www.kaggle.com/huzzefakhan/zameencom-property-data-pakistan