Coursera Capstone Project ~ Thesis Report

The Battle of Neighborhoods

Author: Cassin Thangam Edwin

1. Problem Statement

People of all kinds from around the world come to California, USA with some of these them aspiring to make this commercial place a home of their own. With varying budgets and needs, people find it very hard to find a suitable place and neighborhood to accommodate them and their families. Due to high cost of living and other multiple issues, California housing has been struggling. With the Silicon Valley having its base here, the problem has emerged on its peak. A potential client aspiring to buy a suitable property would like to become knowledgeable about the ongoing pricing to make a conscious decision. Further, he/ she would like to consider several factors like proximity to age of the residents, number of bedrooms in each house, number of houses, population and price.

With US government provided authentic data on California properties coupled with data science techniques, one can make derive the useful information about current pricing in different localities of California while considering other factors of his choice. This would help the potential client to make an informed decision about buying a suitable property.

Target audience: -
Potential clients looks to buy suitable property in California but are skeptical due to lack of knowledge and volatile market conditions.

Stakeholders: -
1. Government of US
2. Sellers
3. Buyers
4. Real estate agents

2. Data Section

Following sources of data are used while executing the Capstone Project: -

- **Data title: -**
  Open Data found in "Sparse spatial autoregressions." Statistics & Probability Letters 33.3 (1997): 291-297. published by Pace, R. Kelley, and Ronald Barry.

  Type of data: -
  Dataset in form of CSV file

  Time Period: -
  1990 California Census

  Description of the dataset: -
  House Pricing Data includes information on all properties in California, USA that are sold for full market value.

  The data set contains 100 rows of valuable information with clarity and veracity.

  This dataset was hosted on https://www.kaggle.com/camnugent/california-housing-prices for ease of use. This data is available for open source and is approved by the US Government.

  Source: -
  https://www.kaggle.com/camnugent/california-housing-prices

- **Data title: -**
  Batch Reverse Geocoding

  Type of data: -
  JSON

  Duration: -
  N/A

  Description of the data: -
  Location coordinates obtained by reverse geocoding.

  Latitude and Longitude Information obtained from California Pricing Dataset is used to obtain the House number,Address,City columns.

  Source: -
  https://csv2geo.com/

- **Data title: -**
  Foursquare location data

  Type of data: -
  JSON

  Duration: -
  N/A

  Description of the data: -
  Location coordinates obtained by Foursquare API calls.

  To determine the proximity of various amenities as per the client's requirement, Foursquare location data is used.

  Source: -
  https://foursquare.com/

3. Methodology

California Housing Dataset contains the prices of the houses in California's cities taken from a 1990 census. This dataset was hosted on https://www.kaggle.com/camnugent/california-housing-prices for ease of use. This data is available for open source and is approved by the US Government.

This project is focused on investigating the most recent market prices of the houses in the city of London so that a potential client can buy a property based on his/her budget.

The automated script developed as a part of this project does the following:
- Initially, the address information is extracted using Batch Reverse Geocoding.
- Parsed the necessary data from the dataset which includes the median income and housing prices of the cities.
- Data is pre-processed and unnecessary /invalid information is dropped.
- The Average housing values of all the cities are found.
- Based on the client's budget the average housing price is plotted on the map of California.
- Foursquare credentials are entered and the recommended locations are further led into API calls to determine various venues in proximity to them. All the tabulated responses are presented to the client to get clarity while making a decision to buy a house.
- It is found that Coffee shops and Cafes top the list in the count followed by restaurants, liquor shops and lots of food stores are in this Vicinity which means it's a good residential area.

4. Results

Upon running the exploratory data analysis for state of California to find the average pricing of houses of the cities, we get the following as the result.

	City	Average Housing Value of the City
0	Berkeley	197800.000000
1	Emeryville	82033.333333
2	Oakland	154352.188406
3	Philatelic	127175.000000

Further following venues are enlisted there by helping the potential client to make a good decision while buying the house.

	Street	Street Latitude	Street Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Upper (Jordan) Trail (Uc)	37.88	-122.23	Grizzly Peak	37.880218	-122.230049	Scenic Lookout
1	Upper (Jordan) Trail (Uc)	37.88	-122.23	Berkeley Hills (View Of The City)	37.880493	-122.230206	Mountain
2	Upper (Jordan) Trail (Uc)	37.88	-122.23	Grizzley Peak Picnic Area, Tilden Park	37.881304	-122.231551	Park
3	Upper (Jordan) Trail (Uc)	37.88	-122.23	Fish Ranch Rd.	37.880437	-122.230110	Scenic Lookout
4	Upper (Jordan) Trail (Uc)	37.88	-122.23	The Wall	37.881412	-122.234198	Scenic Lookout
5	Upper (Jordan) Trail (Uc)	37.88	-122.23	Tilden Regional Park - Lupine Trail	37.880581	-122.225146	Trail
6	Tunnel Rd	37.86	-122.22	GCH Remodel, LLC	37.857931	-122.221924	Construction & Landscaping
7	Tunnel Rd	37.86	-122.22	Air Duct Cleaning Oakland	37.857444	-122.219733	Home Service
8	Tunnel Rd	37.86	-122.22	Wild Turkey Hill	37.856666	-122.219858	Trail
9	Tunnel Rd	37.86	-122.22	Parkwoods Pool	37.856500	-122.220740	Pool
10	Tunnel Rd	37.86	-122.22	Parkwoods Gym	37.856478	-122.220779	Gym
11	Tunnel Rd	37.86	-122.22	Parkwoods Hot Tub	37.856443	-122.220774	Water Park
12	Golden Gate Ave	37.85	-122.24	Little Chabot Field	37.850620	-122.241240	Baseball Field
13	Golden Gate Ave	37.85	-122.24	Claremont Path	37.849738	-122.240864	Trail

5. Discussion

Based upon the findings in the results section, the user can take a confident decision about choosing the city, street in California for making his/her residency.

As we have no listed requirements or constraints from the client we can provide him/her some insights upon assumption:

- If he/she belongs to a high earning background, Berkeley or Oakland suits him/her the best in terms of housing.
- If he/she is a Coffee lover or a person who loves to spend time in a Café, California suits the best having Coffee Shop/Café at every street almost.