The Battle of the Neighborhoods – Week 1

- Introduction/Problem Section
- Data Section

1. Introduction

1.1 Background

Californians by nature are fun loving people it does not matter what race, culture or region they belong to. They love to spend quality time with family and friends. Los Angeles (LA) is a densely populated metro city. To avoid the daily rat race, people live in the suburbs of LA and still like to be called as Angelenos. Lots of people live in the suburbs of LA and most of the suburbs in outskirt of LA fall under the vicinity of the largest valley in the world called San Fernando Valley (SFV). This is where the city of west hills is located.

1.2 Problem

A well-known Bowling chain is our audience and stakeholder. The sponsors/stakeholders are looking forward to open up a new location for their bowling alley in the SFV's West Hills city. *The problem is where should they open their Bowling Alley in the city in order to get minimum competition, maximum customer turnout in short find an optimal location?* This project specifically targets the stakeholders who are interested in opening a new location for **Bowling Alley** in the city of **West Hills of California**, USA.

This report will answer the above question and provide with possibly a best solution to the problem. We will use our data science skills to wrangle the data and analyze some of the areas of given neighborhoods, based on the above criteria. In order to support our findings, we will present the stakeholders with the best or optimal location to facilitate their decision making.

Some demographics of West Hills: The population of the city is almost around 50,000 with mean household income of $120,608, which is way higher than the national household income. 75% of the population has either Master's or Higher degree, Bachelor's degree or, some college degree as opposed of national percentage of 61%. Not only that 79% of the population is adult population. It is a fairly diverse city when it comes to racial diversity.

1.3 Interest

As mentioned in the above item 1.2 our client is a well-known Bowling and Venues company and this report is of great interest to them (executives of the company) as they plan on adding a new location to their chain of Bowling Alleys.

2.Data Acquisition and Cleaning

2.1 Data Sources

For resolving above problem, data for the West Hills city for the neighborhood was acquired from acquired from Redfin in the form of .CSV files. Few versions were downloaded for last 3 years data and the combined all the datasets in to one .CSV data file. Also, we acquired some demographics about the city of West Hills from Niche

Data acquired through downloading csv files or scraped data have been combined into a single dataset, as a .CSV file. There were some missing values and some parts of the data were coded wrong, so we cleaned the data to suit our purpose. Apart from above Foursquare API has also been deployed.

2.2 Data Cleaning

We dropped some features from the data and kept some as they were more relevant for our solution. The combined data file has below features, out of which only the few were kept in the data and rest were dropped. We also renamed the columns

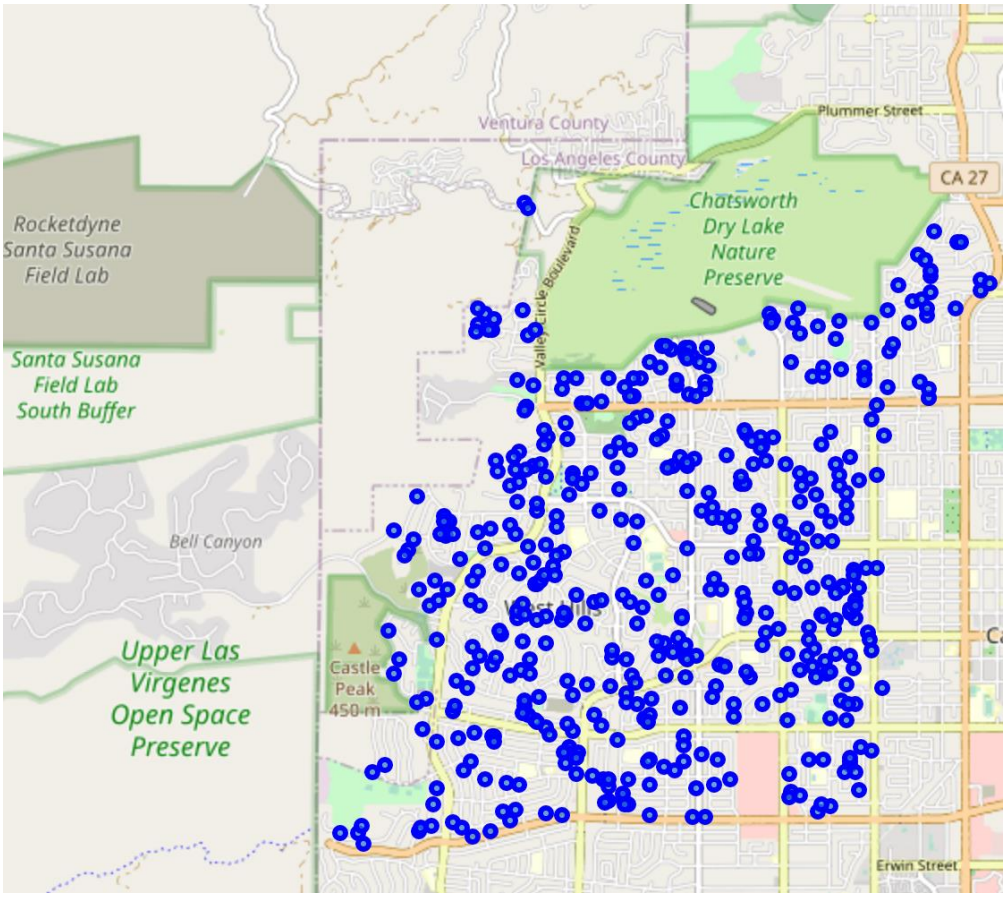
ID	Features	Kept	Renamed
1	SALE TYPE	FALSE	-
2	SOLD DATE	FALSE	-
3	PROPERTY TYPE	FALSE	-
4	ADDRESS	**TRUE**	Address
5	CITY	**TRUE**	City
6	STATE OR PROVINCE	**TRUE**	State
7	ZIP OR POSTAL CODE	**TRUE**	Zip
8	PRICE	FALSE	-
9	BEDS	FALSE	-
10	BATHS	FALSE	-
11	LOCATION	**TRUE**	Location
12	SQUARE FEET	FALSE	-
13	LOT SIZE	FALSE	-
14	YEAR BUILT	FALSE	-
15	DAYS ON MARKET	FALSE	-
16	$/SQUARE FEET	FALSE	-
17	HOA/MONTH	FALSE	-
18	STATUS	FALSE	-
19	NEXT OPEN HOUSE START TIME	FALSE	-
20	NEXT OPEN HOUSE END TIME	FALSE	-
21	URL	FALSE	-
22	SOURCE	FALSE	-
23	MLS#	FALSE	-
24	FAVORITE	FALSE	-
25	INTERESTED	FALSE	-
26	LATITUDE	**TRUE**	Latitude
27	LONGITUDE	**TRUE**	Longitude

Later the column state and location were dropped as information was redundant and we already had the data in City feature. Some rows were dropped as well since they were wrongly zip coded.

We also combined similar addresses with different unit number into one by ignoring the #unit number, like in the example below. Later we removed the duplicate records.

	Address	City	State	ZipCode	Latitude	Longitude
	22421 Sherman Way #8	West Hills	CA	91307	34.201300	-118.615225
	22421 Sherman Way #4	West Hills	CA	91307	34.201300	-118.615225
	22421 Sherman Way #1	West Hills	CA	91307	34.201300	-118.615225
	22525 Sherman Way #203	West Hills	CA	91307	34.201563	-118.618593
	22525 Sherman Way #702	West Hills	CA	91307	34.201563	-118.618593

We found the geographical co-ordinates of the West Hills and used Folium to create the map of the neighborhood with the latitude and longitude that we found. Below is a screen shot of the same which plots areas in West Hills.



Foursquare API

Foursquare is a city guide for local place of interests, it get you the access of global POI data along with their Venus, co-ordinate in short it provide geospatial analytical functions. We use the power of

Foursquare and we implemented Foursquare API. With the help of Foursquare API we explored the venues and business types around the neighborhoods.

Venues in the neighborhood and their Categories.

Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
7947 Cowper Ave	34.214568	-118.647376	Lazy J Park	34.212002	-118.644622	Park
7336 Asman Ave	34.203795	-118.616474	Go's Mart	34.200696	-118.613450	Sushi Restaurant
7336 Asman Ave	34.203795	-118.616474	Ginger Thai	34.200518	-118.613958	Thai Restaurant
7336 Asman Ave	34.203795	-118.616474	Sze-chwan Inn	34.202187	-118.613009	Chinese Restaurant
7336 Asman Ave	34.203795	-118.616474	Nico's Family Restaurant	34.200851	-118.614398	American Restaurant
7336 Asman Ave	34.203795	-118.616474	Del Taco	34.199975	-118.614415	Fast Food Restaurant
7336 Asman Ave	34.203795	-118.616474	Royal Delhi Palace	34.202205	-118.612549	Indian Restaurant
7336 Asman Ave	34.203795	-118.616474	Doner King	34.201850	-118.614446	Mediterranean Restaurant
7336 Asman Ave	34.203795	-118.616474	Sherman Plaza	34.200914	-118.613821	Shopping Mall
7336 Asman Ave	34.203795	-118.616474	Casa De Papi Mexican Grill	34.200565	-118.613960	Mexican Restaurant
7336 Asman Ave	34.203795	-118.616474	Supertans 24	34.202100	-118.611773	Spa
8563 Rudnick Ave	34.226878	-118.613329	Saturday Night Sessions Radio	34.225170	-118.615662	Music Venue
23540 Community St	34.221487	-118.640221	Valley Circle Canyon	34.219550	-118.644622	Scenic Lookout
23540 Community St	34.221487	-118.640221	The Bun Truck	34.218129	-118.640229	Food Truck
23540 Community St	34.221487	-118.640221	The Nail Oasis	34.219616	-118.645049	Cosmetics Shop
23540 Community St	34.221487	-118.640221	Al Italano	34.219624	-118.645065	Italian Restaurant
23540 Community St	34.221487	-118.640221	Silver Flask	34.219456	-118.644990	Liquor Store
7930 Mencken Ave	34.214276	-118.649154	Lazy J Park	34.212002	-118.644622	Park

Later we used Foursquare to explore our venues and categories furthermore, also determined the top 5 common venues in every neighborhood and

This concludes our Data Acquisition and Cleaning section as rest of the analysis will be part of the main report.

In this project we used the following libraries and packages.

- Pandas – For data analysis
- NumPy - For working with arrays/vector data
- BeautifulSoup – For Scraping (if any)
- Requests – For handling requests
- Geopy – For geocoding data
- Folium – For map plotting and markers.
- Sklearn – For kMeans clustering
- Foursquare API – For venues and getting place of interest.

Note: *This document has relevant information only for the **Introduction/Business Problem Section** and **Data Section** as mentioned in the submission page of Coursera. This document will be updated as Main Report for Final Delivery of the Course –The Battle of Neighborhood Report.*