

Section2: Data Collection

The data section can be divided in two parts:

- **Statistics of California housing:** This data is for analysing which cities have more costly houses, median housing price, latitudes and longitudes of various areas of California.

This data is fetched from <https://www.statista.com/> .

Statista provides statistics and data within 600 industries and 50+ countries.

It is fetched under the category of California Housing and the link to the same is:

<https://www.statista.com/search/?q=california%20housing&qKat=search>

Example can be seen below:

	longitude	latitude	population	households	median_house_value	Neighbourhood
0	-122.29	37.83	554	187	75700	Berkley
1	-122.29	37.82	86	23	75000	Berkley
2	-122.29	37.81	377	122	86100	Berkley
3	-122.29	37.80	492	147	81300	Berkley
4	-122.27	37.79	718	302	187500	Berkley

- **Geographical data using coordinates:** After cleaning the above data, Foursquare API is used to explore the specific areas of California, the nearby venues, analyse all the neighbourhoods and form clusters out of them. The Foursquare Places API provides location based experiences with diverse information about venues, users, photos, and check-ins. Additionally, Foursquare allows developers to build

audience segments for analysis and measurement. JSON is the preferred response format.

Example can be seen below:

