

Section 2: 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>

- **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.