Aim: The aim of this project is to analyze and visualize the median housing prices in California.

**About this data:**

1. longitude: A measure of how far west a house is; a higher value is farther west  
2. latitude: A measure of how far north a house is; a higher value is farther north  
3. housingMedianAge: Median age of a house within a block; a lower number is a newer building  
4. totalRooms: Total number of rooms within a block  
5. totalBedrooms: Total number of bedrooms within a block  
6. population: Total number of people residing within a block  
7. households: Total number of households, a group of people residing within a home unit, for a block  
8. medianIncome: Median income for households within a block of houses (measured in tens of thousands of US Dollars)  
9. medianHouseValue: Median house value for households within a block (measured in US Dollars)  
10. oceanProximity: Location of the house w.r.t ocean/sea

Step 1: Data pre-processing

* Loading data in python, understand data, if null or zero, and if yes what to do of those.
* Check data types of all (int, float, str)
* Plot a map using longitude and latitude (a location), and color scale (buying prices, no. of rooms, no of households = 3 maps) (G search: sale orders by region plot data in python)
* Map, tree map, pie chart

Step 2: Then create a dashboard in python (code)

Refer steps: <https://dimensionless.in/data-cleaning-categorization-normalization/>

12/11/2021:

-heatmap: less features

{Missing Data Heatmap When there is a smaller number of features, we can visualize the missing data via heatmap}

1. Filling the missing data with the mean or median value if it’s a numerical variable.
2. Table creation x4
3. visualization