**CALIFORNIA HOUSE PRICES**

California is a state located in the Western Coast of United States of America. It is located on the shores of Pacific Ocean.It is the most populous and third largest state in United States of America. Los Angeles is the most populous city in California. In this study , we have taken the Housing prices data from the 1990 Census of California. This data was initially published in the paper “Sparse spatial autoregressions." Statistics & Probability Letters 33.3 (1997): 291-297” by Pace, R. Kelley, and Ronald Barry.The main objective of this study is to identify the various factors that affect the House Prices. It also aims to identify the income groups which own houses. This data set contains 9 columns and 20,640 rows. The following are the columns.

* Longitude: A measure of how far west a house is; a higher value is farther west
* Latitude: A measure of how far north a house is; a higher value is farther north
* Housing\_median\_age: Median age of a house within a block; a lower number is a newer building.
* Total\_rooms: Total number of rooms within a block
* Total\_bedrooms: Total number of bedrooms within a block
* Population: Total number of people living in a block.
* Households: Total number of households, a group of people residing within a home unit, for a block
* Median\_income: Median income for households within a block of houses (measured in tens of thousands of US Dollars)
* Median\_house\_value: Median house value for households within a block (measured in US Dollars)
* Ocean\_proximity: Location of the house w.r.t ocean/sea

**Objective 1**: To identify the location where maximum number of Households are situated.

**Objective 2:** To identify the Features of High Priced Locality and Low Priced Locality.

**Objective 3:** To identify the income groups who own Houses.

**Objective 4:** To identify the factors that lead to determination of House Prices.

**Methodology**

We constructed a Histogram of Longitude and Latitude in order to identify the location where maximum number of households are situated. To identify the locality where the house prices were highest and lowest, we applied the method of conditional filtering. Correlation and heatmaps were used to identify various factors that affect the house prices.

**Results:**

In this study the longitude measures how far west the house is. Higher the value of longitude, higher to the west while latitude measures how far north the house is. After analyzing the data, it was found that majority of Households are located towards south west region i.e -118 longitude and 34 latitude. It is found that majority of the Households are located within 1 hour from the ocean.

The Highest median House Value is $5,00,001. There are about 44,9664 Households in the area where median House Value is the highest. Most of these houses are located within 1 hour from the ocean. The Average income of people living in these are is $7,82,512. Around 1073860 people live in the locality where House were highest.

The Lowest Median House value is $14,999. There are about 435 households in the locality where house prices are lowest. Most of these Houses are located in the inland region. The Population is as low as 1221. Thus, it can be concluded that as population rises House Prices also start rising.

It can be inferred from the data that most of the Houses are owned by Households whose Household income lies between $2,00,00 and $4,00,00.

There is a strong positive correlation between income and House Prices. As income rises, demand for houses start rising leading to an increase in prices of Houses. There is a very weak positive correlation between number of households and House Prices. There is a weak positive correlation between population and house prices. As the number of bedrooms rises House Prices also starts rising. There is a negative correlation between latitude and house prices. This implies that as we go farther to the north, House prices start falling. There is a positive relationship between longitude and house Prices. Thus, prices of houses located towards the west coast are higher as compared to the east. There is no relationship between house prices and age of the building. Thus, the age of the building does not determine the value of the houses.

**Conclusion**

Most of the High-Priced Houses are located close to the west Coast. Majority of people reside in the South West region. Most of the people residing in these regions have an average income lying between $300,000 to $400,000 Income is one of the driving factors of House prices with a positive correlation of 0.69. House Prices start falling as we go towards the north. Thus income, longitude, age of the building, total number of bedrooms, total number of rooms have a positive relationship with House Prices.