**Background of Problem Statement**

The US Census Bureau has published California Census Data which has 10 types of metrics such as the population, median income, median housing price, and so on for each block group in California. The dataset also serves as an input for project scoping and tries to specify the functional and nonfunctional requirements for it.

Data Description

# California Housing Data Set Description



Many of the Machine Learning Crash Course Programming Exercises use the California housing data set, which contains data drawn from the 1990 U.S. Census. The following table provides descriptions, data ranges, and data types for each feature in the data set.

| **Column title** | **Description** | **Range\*** | **Datatype** |
| --- | --- | --- | --- |
| longitude | A measure of how far west a house is; a higher value is farther west | * Longitude values range from -180 to +180 * Data set min: -124.3 * Data set max: -114.3 | float64 |
| latitude | A measure of how far north a house is; a higher value is farther north | * Latitude values range from -90 to +90 * Data set min: 32.5 * Data set max: 42.5 | float64 |
| housingMedianAge | Median age of a house within a block; a lower number is a newer building | * Data set min: 1.0 * Data set max: 52.0 | float64 |
| totalRooms | Total number of rooms within a block | * Data set min: 2.0 * Data set max: 37937.0 | float64 |
| totalBedrooms | Total number of bedrooms within a block | * Data set min: 1.0 * Data set max: 6445.0 | float64 |
| population | Total number of people residing within a block | * Data set min: 3.0 * Data set max: 35682.0 | float64 |
| households | Total number of households, a group of people residing within a home unit, for a block | * Data set min: 1.0 * Data set max: 6082.0 | float64 |
| medianIncome | Median income for households within a block of houses (measured in tens of thousands of US Dollars) | * Data set min: 0.5 * Data set max: 15.0 | float64 |
| medianHouseValue | Median house value for households within a block (measured in US Dollars) | * Data set min: 14999.0 * Data set max: 500001.0 | float64 |

==

Substantial variation in housing conditions exists across the state, although some common patterns emerge. Most notably, we find some indicators of housing demand outstripping supply: New housing production did not keep pace with population growth; already low vacancy rates in many counties declined even further in the 1990s, especially for rental units; real prices increased in most counties in the 1990s; and households in the state became more crowded in most counties. However, for many indicators, the trends of the 1990s were much less notable than those of the 1980s. For example, price increases were much more substantial in the 1980s than in the 1990s. Within the state, the Bay Area often stood out, with other housing markets in other regions of the state showing few if any signs of a lack of overall supply. The general picture is one of a tight housing market, but with the 1980s exhibiting more remarkable changes than the 1990s.

==

Measures of income and unemployment levels are usually incorporated in typical housing demand equations. As income and employment increase, the demand for new housing is expected to increase. Higher current housing prices decrease the current demand for housing, whereas expectations of higher future housing prices increase demand. Our model describes expected appreciation or expected housing valuation as a function of expected inflation (see Appendix A). Furthermore, because housing is both a commodity and an investment, real estate can be seen as a substitute asset for stocks, particularly when inflation is high. In periods of high economic growth, housing and stock prices tend to move together; however, in inflationary environments, houses have done better than stocks. Historically, stocks have better returns when inflation is lower (Wasserman, 1998). Thus, returns on stocks are also a determinant of housing demand because they are an alternative investment. Typically, housing demand or supply equations use either mortgage rates or the prime rate as a proxy for the cost of credit. The higher the interest rate, the lower the demand for housing. We looked at the difference between the long-term and short-term interest rate because it measures the tightness of credit better than the level of either interest rate separately. It is also a better measure because a large spread between the long- and short-term interest rates indicates expectations of rising inflation. The relationship between this spread and the demand of housing is expected to be positive.

==

Changes in population growth, the age structure of the population, and immigration also affect the demand for housing. As the population grows, more housing is needed, but a high proportion of children in the population leads to smaller increases in housing demand.

==

Final Summarty

Slower population growth in the 1990s compared to the 1980s, which slowed the demand for housing. The nature of that growth, too, explains much of the apparent lack of new housing in the state, with immigrants and children accounting for most of this growth, but consuming proportionately fewer housing units than other demographic groups

The severity and duration of the economic recession that took place in 1990 in California, leading to low housing valuation.

Price by Age - Older houses are costlier than the houses with a median age. New houses are also equally costlier than the median age houses.

Median House Price Distribution - Majority of the houses are comes under the price of 100k. Dataset is skewed towards right hand side. Also interms of houses based ocean proximity the dataset contains uneuqal distribution.

Average House Price @ Different Types of lands - House in Island is costlier than the house in Inland. House closer to the beach or ocean are costlier than the house that is there away from the ocean.