California Housing Dataset



PROJECT TITLE:

California Housing Prices derived from 1990 census data.

PROBLEM STATEMENT

The project aims at building a model of housing prices to predict median house values in California using the provided dataset. This model should learn from the data and be able to predict the median housing price in any district, given all the other metrics. This dataset comes under finance and housing sector.

Districts or block groups are the smallest geographical units for which the US Census Bureau publishes sample data (a block group typically has a population of 600 to 3,000 people). There are 20,640 districts in the project dataset.

The data pertains to the houses found in each California district and some summary stats about them based on the 1990 census data. The columns are. Longitude, latitude, housing median age, total rooms, total bedrooms, population, households', median income, median house value, ocean proximity

Synopsis of the dataset:  
<https://www.kaggle.com/camnugent/california-housing-prices>

Expected Visualizations and outcomes:

1. Our project aims to find out the attributes that really matter on deciding the housing prices in the state of California.
2. Seaborn graphs of prices across the state to determine the price structure in the given neighborhood.
3. Other important geospatial Visualizations like distance b/w big cities, distance from the coastal proximity to a house to predict the price of that house etc.
4. Seaborn heatmaps for the check of correlation.