House-Price-Prediction

House price prediction using Multiple Linear regression and Keras Regression

This is a famous data set for beginners practicing regression. In this program, I will implement multivariate linear/keras regression to predict the "Sale prices" of houses.

❯ Which DataSet is this?

Bangalore Housing Dataset is an alternative to commonly used Boston Housing dataset.

❯ What you would need to run .ipynb file?

Jupyter Notebook from Anaconda distribution

❯ What libraries are used in the code?

scikitlearn, pandas, numpy and matplotlib.

❯ Dataset

The data contains information from the Bengaluru census. The columns are as follows:
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
housingMedianAge: Median age of a house within a block; a lower number is a newer building
totalRooms: Total number of rooms within a block
totalBedrooms: Total number of bedrooms within a block
population: Total number of people residing within a block
households: Total number of households, a group of people residing within a home unit, for a block
medianIncome: Median income for households within a block of houses (measured in tens of thousands of Dollars)
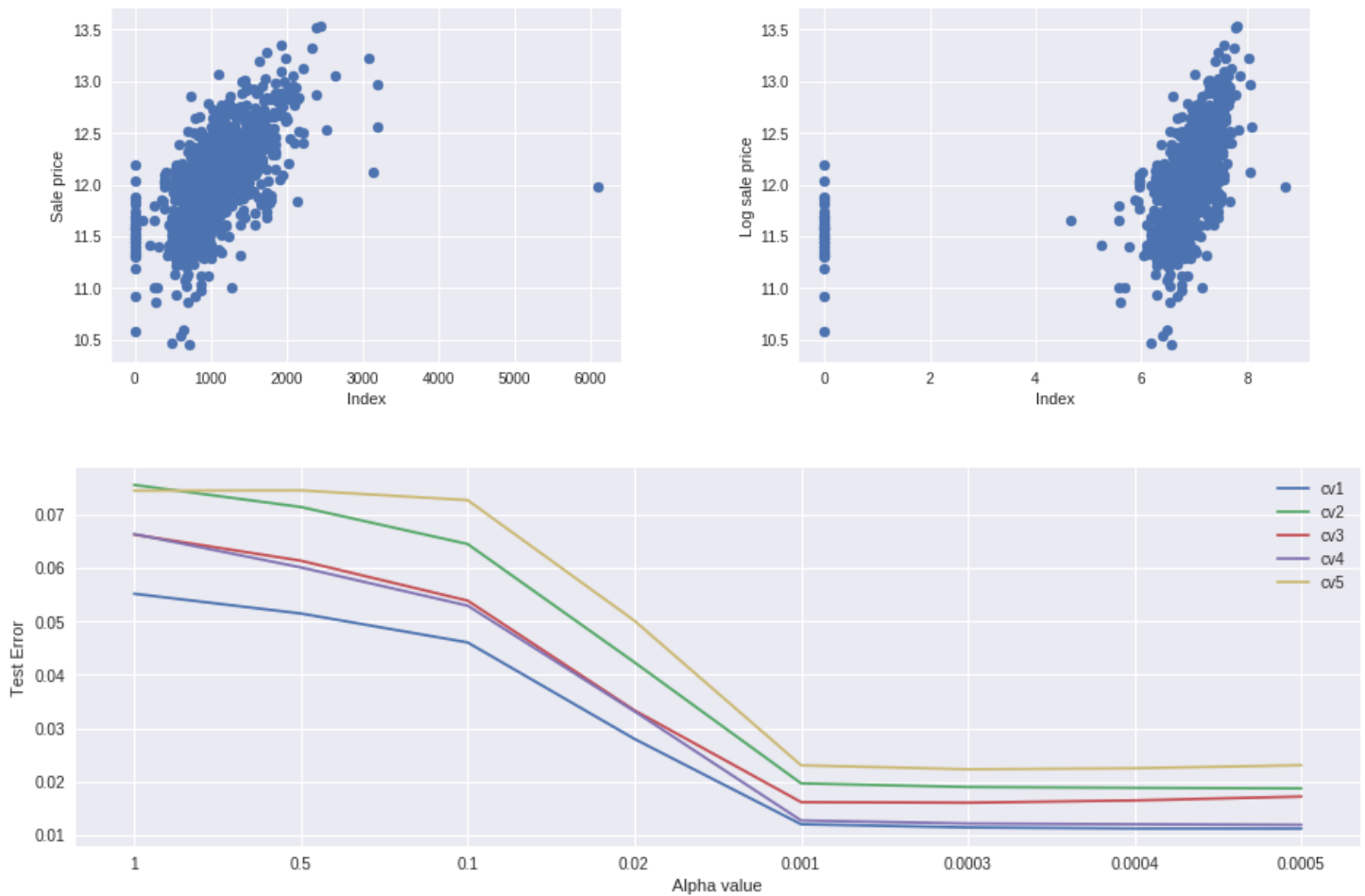*medianHouseValue:* Median house value for households within a block (measured in Dollars)
oceanProximity: Location of the house w.r.t ocean/sea

❯ Regression Used

I have used Linear regression and keras with gradient descent to find the best coefficient values of the predictors.

Output of Multiple Linear Regression


Tensorflow and Keras

TensorFlow is an end-to-end open source platform for machine learning. It has a comprehensive, flexible ecosystem of tools, libraries and community resources that lets researchers push the state-of-the-art in ML and developers easily build and deploy ML powered applications.

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, R, Theano, or PlaidML. Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible.

How was the score improved? How are the hyperparameters selected?

The only hyperparameter in case of lienar regression is the regularization paramter. The best tregularization paramter is estimated as the one that performed best across cross validation datasets.

Output of keras regression
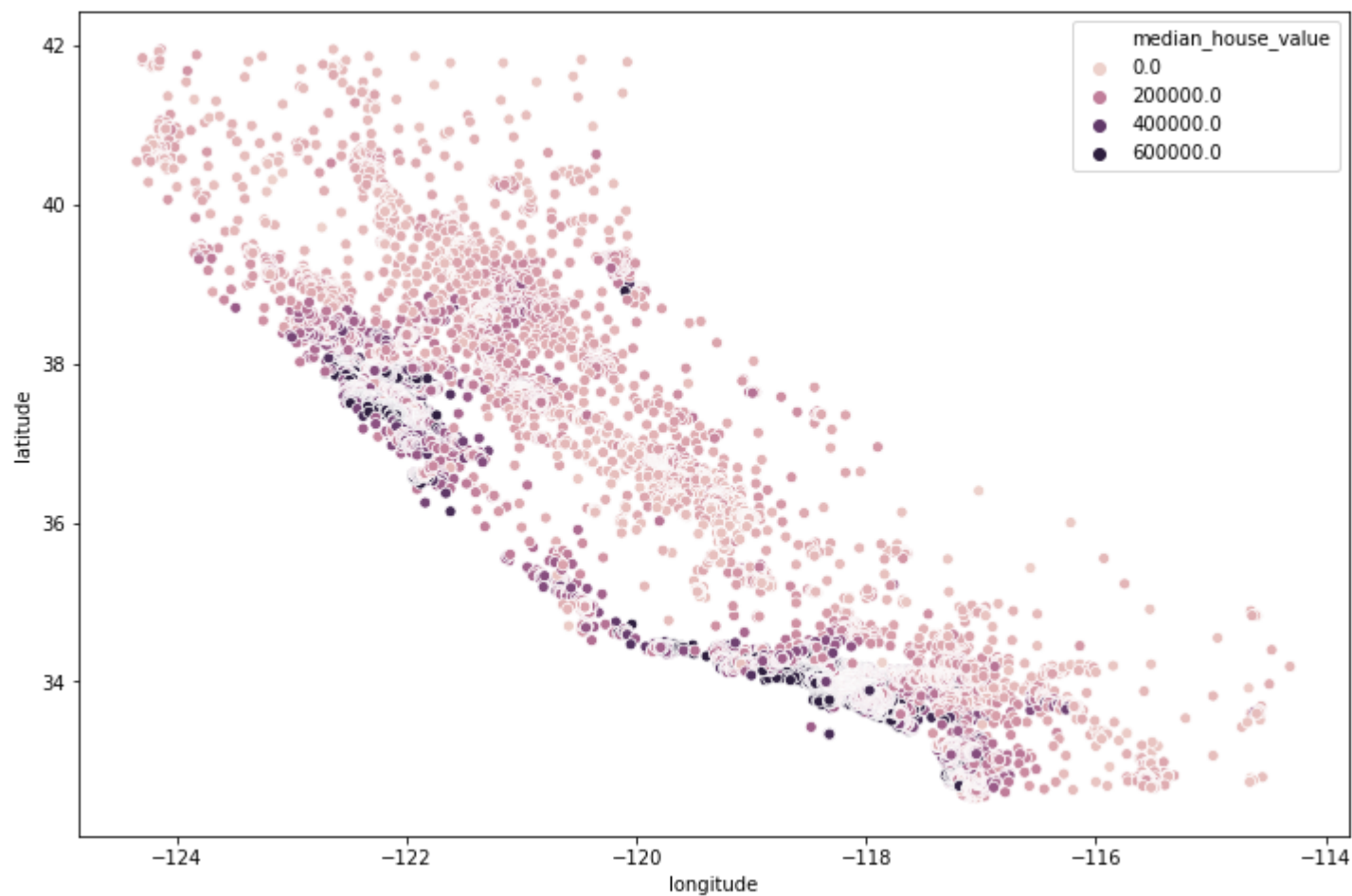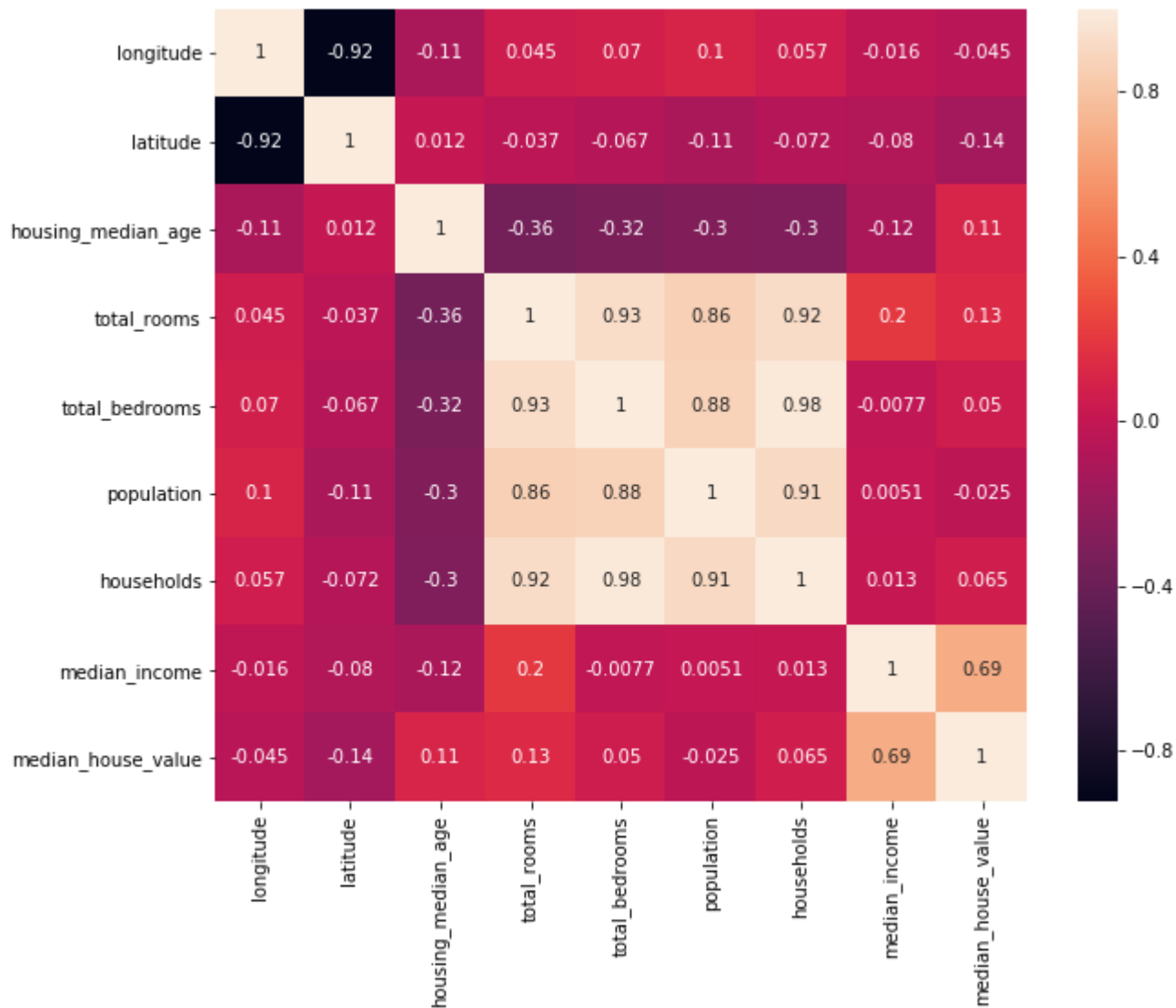
What did we achieve from running this model?

Low Overall quality index and less pool area negatively effect the sale price of the house.

High Overall quality rating of 10,9 and neighborhoods of Crawfor and StoneBr are postively associated with high sale prices.

By how much these variables effect the house sales price can be calculated by substituting our estimated coefficients values in the regression equation.