One of my mom's friends says her daughter, who is a realtor, is super busy at work, and has closed many deals recently. I am surprised that even during the pandemic, there are so many people buying houses. Naturally, I want to expose the ways to predict the ups and downs of the housing market, and maybe later apply the same logic to predict the stock market price.

Dataset

The dataset for my project is taken from https://www.kaggle.com/camnugent/california-housing-prices#, which pertains to the houses found in a given California district, and some summary stats about them based on the 1990 census data. Although it is not the latest, it does give me a chance to learn how people make economic predictions.
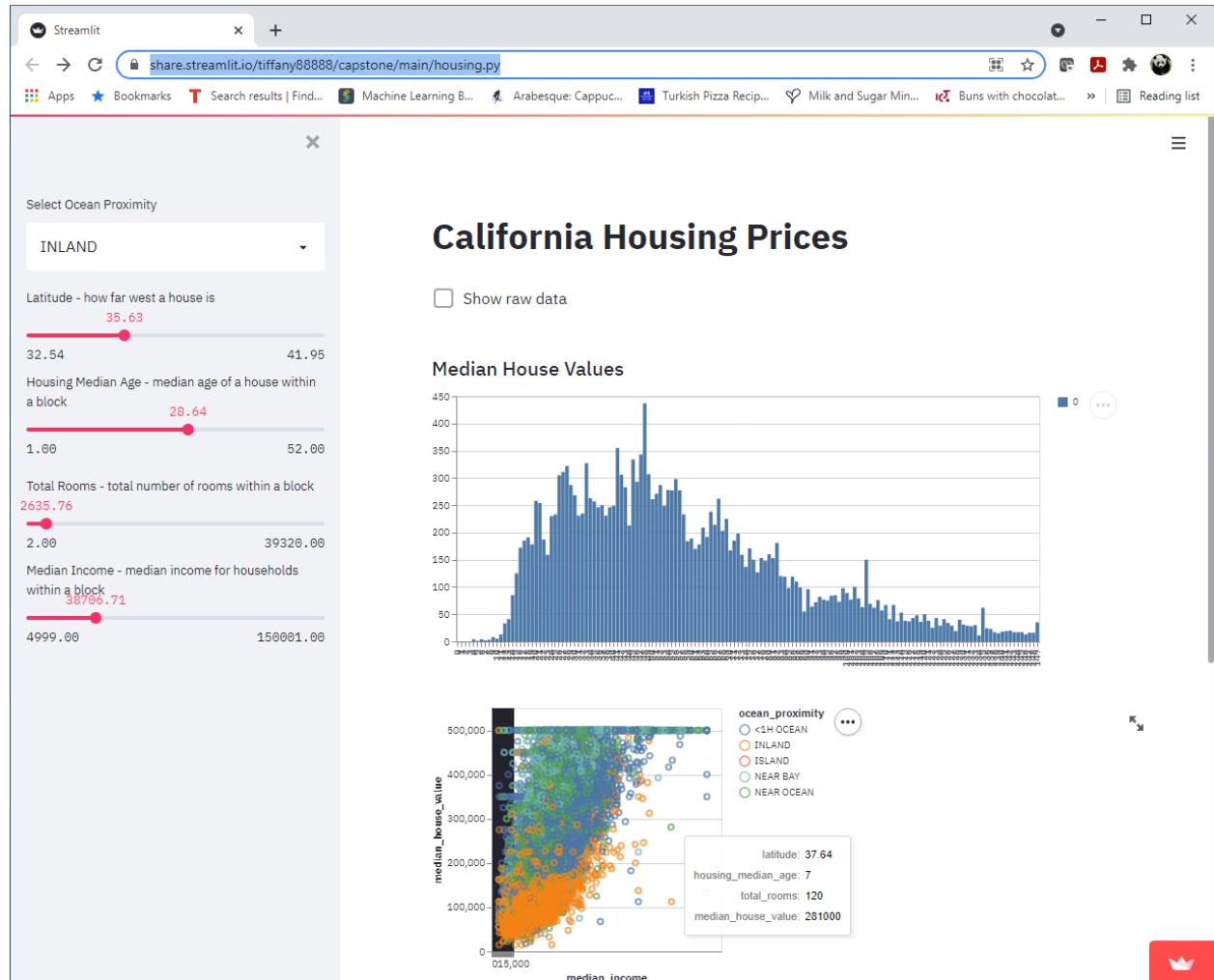
Frontend Web Interface Location

The frontend web interface is hosted on https://share.streamlit.io/tiffany88888/capstone/main/housing.py

Source Code Location

The source code is at: https://github.com/tiffany88888/capstone

Graphical User Interface

Show Raw Data

Under the headline, California Housing Prices, is the selection box Show raw data. If selected, it would list the raw dataset.

Charts

A picture is worth a thousand words. I have included 2 charts to demonstrate the dataset.

Histogram

It is the overview of median house values, divided over 150 slots.

Altair

It consists of the median_income(x-axis) and median_house_value(y-axis). As the cursor moves over a spot, it would display the latitude, housing_median_age, total_rooms, and median_house_value in the tooltip.

Set User Parameters

Users may set the parameters which have a major influence on the housing price on the left Sidebar. The selections, minimum, maximum, and mean values are all from the dataset. Initially, all sliders are set to their mean values.
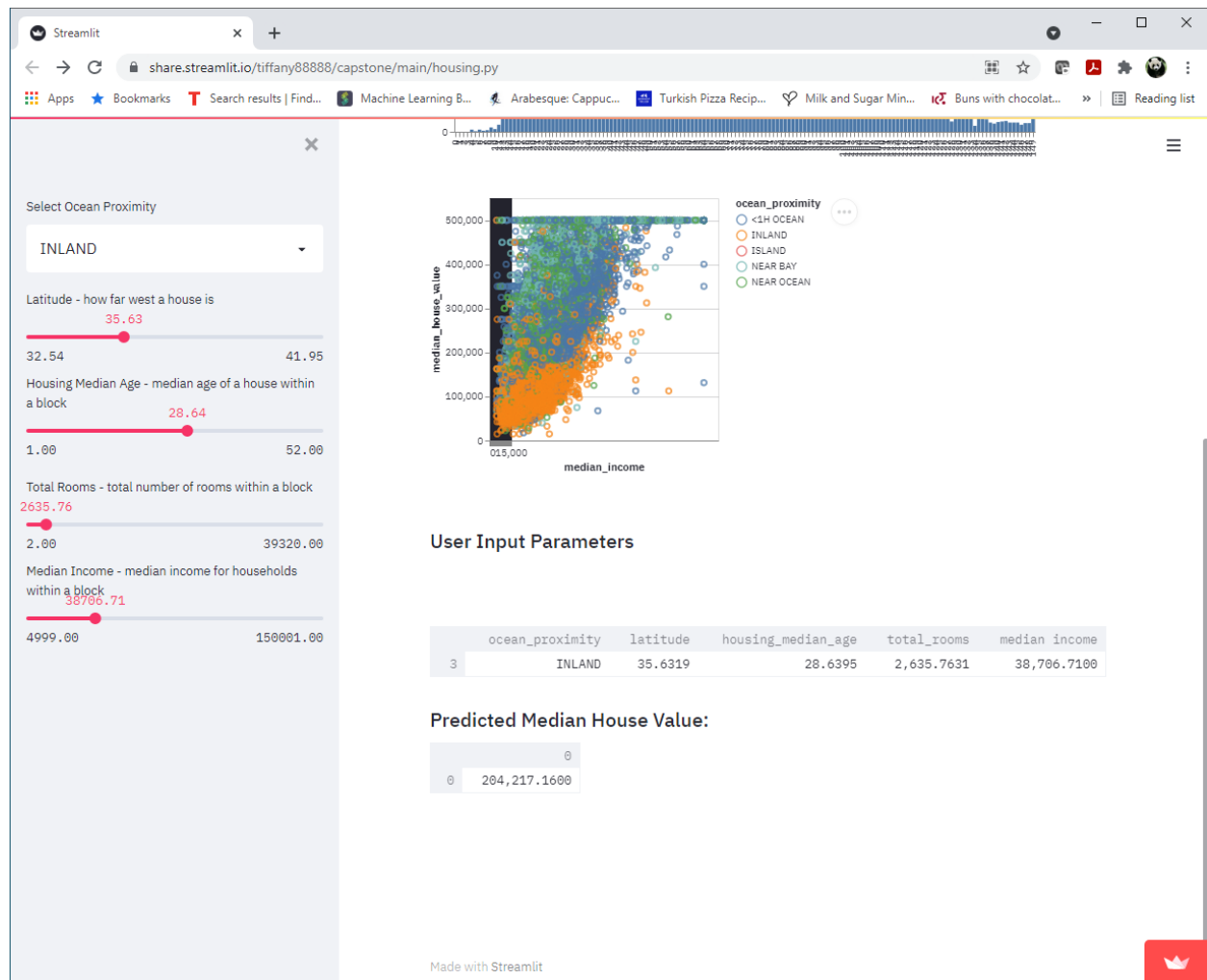
- Ocean Proximity - select from INLAND, ISLAND, NEAR OCEAN, NEAR BAY, <1H OCEAN
- Latitude - indicate how far west a house is.
- Housing Median Age - the median age of a house within a block.
- Total Rooms - the total number of rooms within a block.
- Median Income - the median income for households within a block.

User Input Parameters

The user selections are reflected in the User Input Parameters section.

Predicted Median House Value

The Predicted Median House Value is listed at the bottom.

Conclusion

The web tool might come in handy when a realtor or potential customer wants to identify the potential areas to settle in.

The project is interesting. Although the machine model training takes a long time, it is only a one-time thing. However, I am truly amazed at how easy it is to deploy using streamlit, and update the prediction based on user selections automatically. I have not tried other deployment methods, but I would definitely keep on using streamlit.