PROJECT_2: Housing "Price" Prediction Model

By; Chris Kamau kinyanjui

•OVERVIEW

In this project I take the role of a data scientist tasked to model the formula of predicting a price of a house in the northwestern county, King County in Washington State in the United States of America.

BUSINESS & DATA UNDERSTANDING.

The stake holder for this was a real estate firm found in the King county to assist them in price determination. Using the data "kc_house_data.csv" I am to make the predictive model.
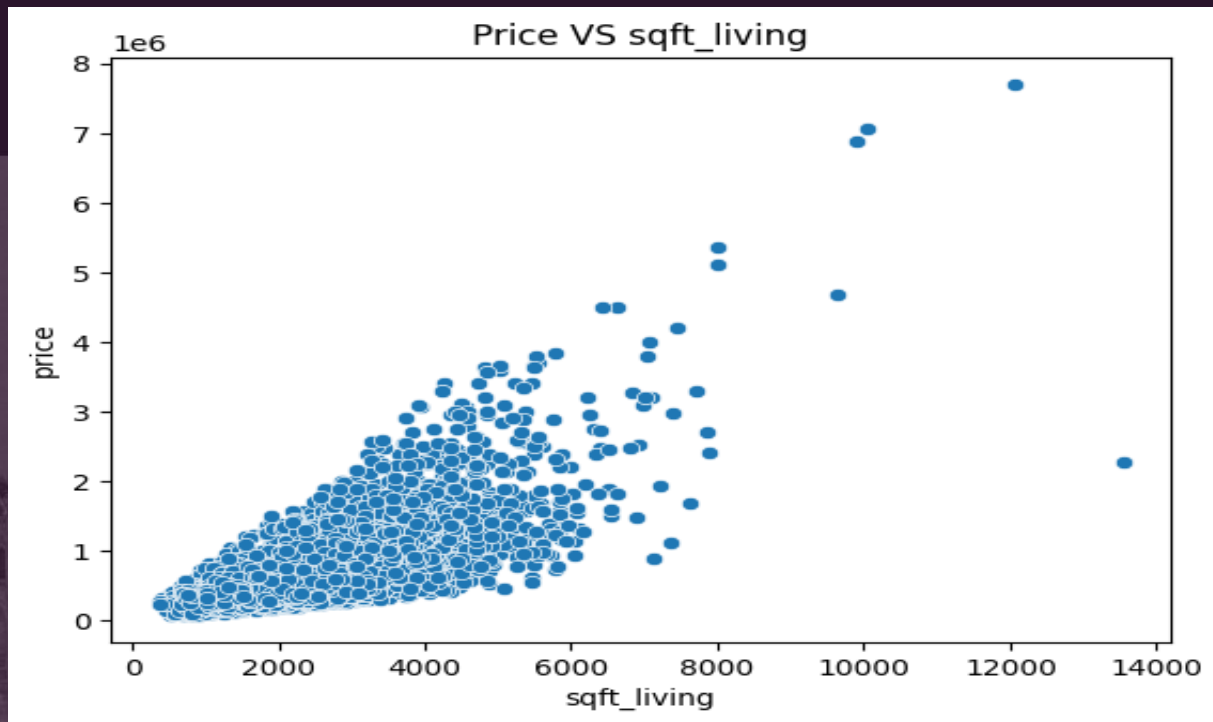
DATA UNDERSTANDING.

It has been loaded in with pandas as kc. Short for King County. It has 21597 rows and 20 columns.

No duplicate rows were observed. Only 3 columns were observed with missing values;

- **yr_renovated** 17.7% missing

- **waterfront** 11.0% missing

- **view** 0.29% missing

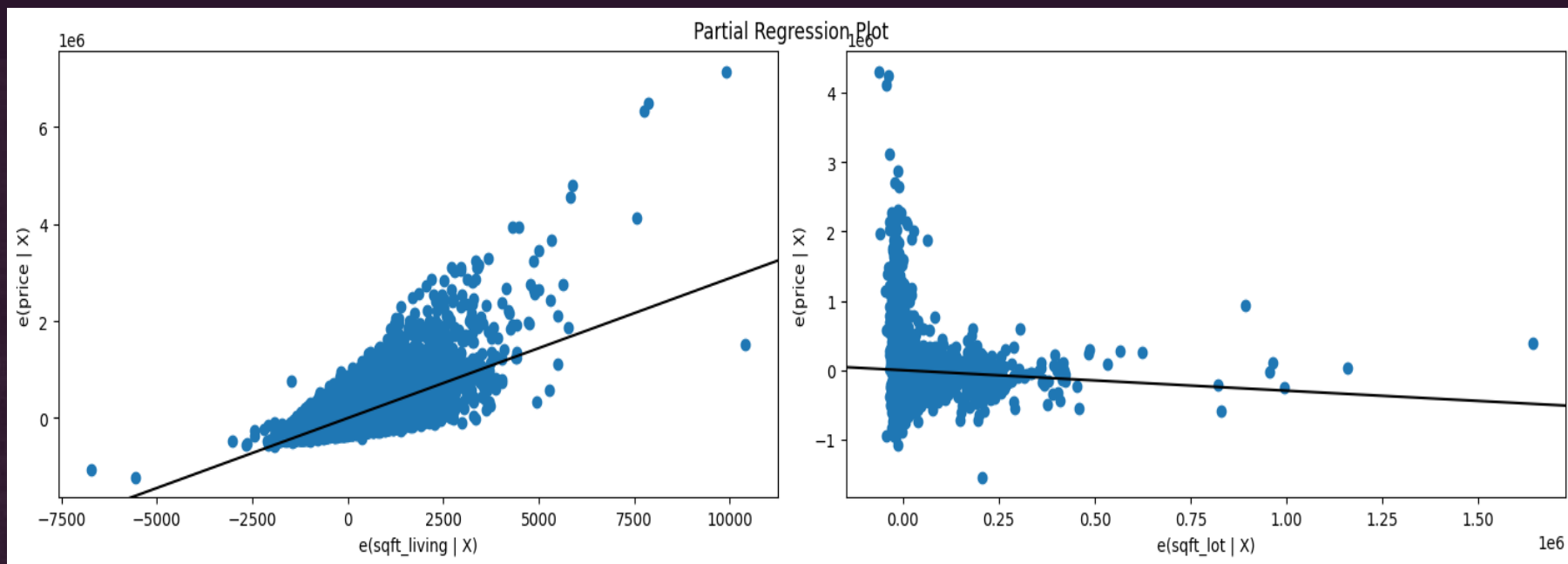# Linear relationship between price and sqft_living

- **MODELING**

The base model formula achieved up-to 95% confidence ;

$$y = ([282 - 291])Sqft\_living + ([-65,400 - -45,000]) \text{ USD} + 175,000 \text{ USD}$$

Multi-linear model of sqft_living with sqft_lot.

$$y = ([282 - 291])\text{Sqft\_living} + ([-0.3 - -0.05])\text{sqft\_lot} + ([-65,400 - -45,000])\ \text{USD} + 175,000\ \text{USD}$$
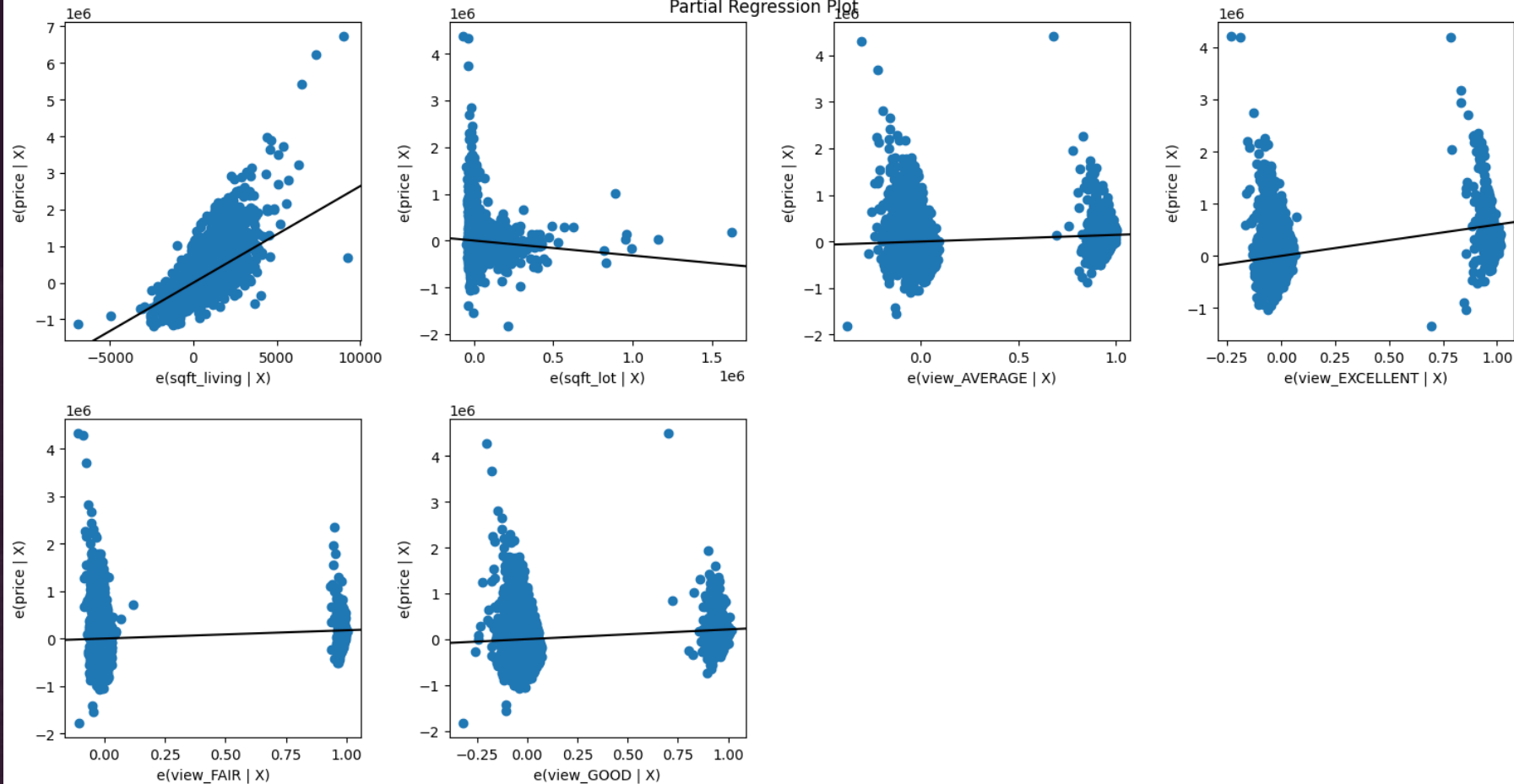
Partial Regression Plot

Multi-linear model of sqft_living with sqft_lot with views.

$$y = ([259 - 268])Sqft\_living - ([-0.41 - -0.23])Sqft\_lot + ([143,200 - 9,880])view\_Average + ([174,900 - 16,200])view\_Fair + ([212,400 - 13,800])view\_Good + ([600,900 - 16,600])view\_Excellent - 25,800 \ USD + 167,100 \ USD$$

Partial Regression Plot

- ## REGRESSION RESULTS

From the models created , I have decided to favor *Multi-linear model of sqft_living with sqft_lot with views.*

Since it explained the largest percentage, about 60% of the variance observed in price variable.

Recomendations

Despite the formula I favor, the errors are still probable and for a more accurate prediction of the price I recommend the real estate firm to create models for each zip uniquely as they have different pricing per zip-code with the highest being **Medina,Washington (98039)** with a mean price of **2,215,069 $** and the lowest zip-code Auburn, Washington (98002) with a mean price of **233,924 $.**

THANK YOU