Model

$$\begin{aligned}\ln(price) =& 10.2082 + 0.3618 \cdot \text{waterfront} - 0.0160 \cdot \text{bedrooms} - 0.0153 \cdot \text{bathrooms} + 0.1400 \cdot \text{sqft\_living}^{0.3} \\ &+ 0.0088 \cdot \text{floors} + 0.1494 \cdot \text{view}^{0.5} + 0.0105 \cdot \text{grade}^2 + 0.1187 \cdot \ln(\text{sqft\_living15})\end{aligned}$$

Explanation of the Model

Before I begin explain the coefficients, I notice that $P > |t|$ for the *floors* variable is 0.155, which makes *floors* insignificant for the analysis.

1. The model has the sample intercept of 10.2082. If we assume that all explanatory variables are zeros, this would mean that the price would be $e^{10.2082} \approx 27,124$

2. 0.3618 is the coefficient for *waterfront*. *Waterfront* is a categorical variable coded as 0 or 1, a one unit difference represents switching from one category to the other. 10.2082 is then the average difference in *price* between the category for which *waterfront* = 0 (no waterfront) and the category for which *waterfront* = 1 (the house has a waterfront). So compared to $\ln(price)$ of the house with no waterfront, we would expect the $\ln(price)$ for the house with waterfront to be 0.3618 higher, on average, if we fix all other explanatory variables.
   Let y_2 = price of the house with waterfront and y_1 = price of the house with no waterfront, then

   $$\ln y_2 - \ln y_1 = 0.3618 \implies \ln \frac{y_2}{y_1} = 0.3618 \implies \frac{y_2}{y_1} = e^{0.3618}$$

   To understand the change in *price* in percents, if we switch from the house with no waterfront to the house with waterfront while keeping all other variables the same, will use the following formula:

   $$100 \times \left[\frac{y_2 - y_1}{y_1}\right] = 100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[e^{0.3618} - 1\right] \approx 43.59\%$$

   Thus, switching from the house with no waterfront to the house with waterfront while keeping all other explanatory variables fixed, will increase the price by 43.59%.

3. -0.0160 is the coefficient for number of *bedrooms*.

   Let y_2 = price for the house with x_2 number of bedrooms and y_1 = price for the house with x_1 number of bedrooms, then

   $$\ln y_2 - \ln y_1 = -0.016x_2 - (-0.016x_1) = -0.016(x_2 - x_1) \implies \ln \frac{y_2}{y_1} = -0.016(x_2 - x_1) \implies \frac{y_2}{y_1} = e^{-0.016(x_2 - x_1)}$$

   If we increase the number of bedrooms by 1 while keeping the other variables fixed and use percents, we will get the following

   $$100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[e^{-0.016} - 1\right] \approx -1.58\%$$

   Thus, increasing the number of bedrooms by 1 while keeping the other variables fixed, will decrease the price of the house by 1.58%. Which is a strange result.

4. The same explanation we have for -0.0153 which a coefficient for number of bathrooms.
   If we increase the number of bathrooms by 1 while keeping the other variables fixed, will decrease the price of the house by 1.51%. Which is a strange result.

5. 0.1400 is the coefficient for *sqft_living*. Let $y_2 =$ price for the house with x_2 *sqft_living* and $y_1 =$ price for the house with x_1 *sqft_living*, then

$$\ln y_2 - \ln y_1 = 0.1400 x_2^{0.3} - 0.1400 x_1^{0.3} \implies \ln \frac{y_2}{y_1} = 0.1400\left(x_2^{0.3} - x_1^{0.3}\right) \implies \frac{y_2}{y_1} = e^{0.1400\left(x_2^{0.3} - x_1^{0.3}\right)}$$

   If we increase the *sqft_living* from $1000ft$ to $1100ft$ while keeping the other variables fixed, we will get the following change in price in percents:

$$100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[e^{0.1400\left(1100^{0.3} - 1000^{0.3}\right)} - 1\right] \approx 3.28\%$$

   If the *sqft_living* is $1000ft$ and we increase it to $1100ft$ while keeping the other variables fixed, we will get the change in price of 3.28%. In this particular example 10% change in *sqft_living* starting from $x_1 = 1000ft$ forces 3.28% change in price.

6. 0.1494 is the coefficient for the *view*. The coefficient for the *view* has the same explanation as the *sqft_living*.

   If we increase the *view* by 1 unit from 2 to 3 while keeping the other variables fixed, we will get the following:

$$100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[e^{0.1494\left(3^{0.5} - 2^{0.5}\right)} - 1\right] \approx 4.86\%$$

   In this particular example if the *view* will increase from 2 to 3, the price will increase by 4.86%.

7. 0.0105 is the coefficient for the *grade*. The coefficient for the *grade* has the same explanation as the *sqft_living*.

   If we increase the *grade* by 1 unit from 2 to 3 while keeping the other variables fixed, we will get the following:

$$100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[e^{0.0105\left(3^2 - 2^2\right)} - 1\right] \approx 5.39\%$$

   In this particular example if the *grade* will increase from 2 to 3 while other variables stay the same, the price will increase by 5.39%.

8. 0.1187 is the coefficient for the *sqft_living15* Let $y_2 =$ price for the house with x_2 *sqft_living15* and $y_1 =$ price for the house with x_1 *sqft_living15*, then

$$\ln y_2 - \ln y_1 = 0.1187 \ln x_2 - 0.1187 \ln x_1 \implies \ln \frac{y_2}{y_1} = 0.1187 \ln \frac{x_2}{x_1} \implies \frac{y_2}{y_1} = e^{0.1187} \frac{x_2}{x_1}$$

   The change in price in precent will be:

$$100 \times \left[\frac{y_2}{y_1} - 1\right] = 100 \times \left[e^{0.1187} \frac{x_2}{x_1} - 1\right]$$

   If we increase the *sqft_living15* by 100 units from 1000 to 1100 while keeping the other variables fixed, we will get the following:

$$100 \times \left[e^{0.1187} \times \frac{1100}{1000} - 1\right] \approx 23.87\%$$

   Thus, 10% increase in *sqft_living15* will lead to 23.87% increase in *price*.

R^2

The model has $R^2 \approx 0.5$. This means that our model explains about 50% of the variation by using *sqft_living* as independent variable.

ANOVA

Is our model with many explanatory variable better than the model with zero explanatory variables?
Our model has $F - statistic = 1.737 \times 10^4$ and $Prob > F$ is 0.000.
The Null Hypothesis: The slope$= 0$
The Alternative Hypothesis: The slope$\neq 0$
Our p-value for this model is $p = 0.000 < 0.05 = \alpha$. Thus, we have enough evidence to reject the Null Hypothesis at 5% level of significance and we conclude that the Test tells us, that at least one of the coefficients is not 0. Since our p-value is 0, there is a 0% probability that the improvements that we are seeing with our independent variables model are due to random chance alone.