**Essentials of Statistics and Math for Data Science - Capstone 2**

To proceed, I'll focus on entries with 2 bedrooms and include **PRICE** as the dependent variable, while **PRICE\_SQFT**, **NO\_OF\_BATHROOMS**, and **FLOOR\_SIZE** will serve as independent variables. I’ll conduct a multiple regression analysis to examine the relationships. To proceed, I'll focus on entries with 2 bedrooms and include **PRICE** as the dependent variable, while **PRICE\_SQFT**, **NO\_OF\_BATHROOMS**, and **FLOOR\_SIZE** will serve as independent variables. I’ll conduct a multiple regression analysis to examine the relationships.

The multiple regression analysis results show that all three independent variables—**price per square foot**, **number of bathrooms**, and **floor size**—significantly impact the **price** of a high-floor, spacious 2-bedroom house. Here are the main findings:

In the regression analysis, we explored how **price per square foot**, **number of bathrooms**, and **floor size** influence the **price** of a 2-bedroom property. Here’s a breakdown of the regression model and the statistical results.

**1. Model Specification and Fitting**

We used an **Ordinary Least Squares (OLS) regression model** to analyze the relationship between the independent variables and the dependent variable (price). This approach estimates the coefficients of each predictor to minimize the difference between the predicted and actual values of the dependent variable.

**Model Equation**

The regression equation based on the analysis is:

Price=β0+β1×Price per Square Foot+β2×Number of Bathrooms+β3×Floor Size+ϵ

Where:

* β0​ is the intercept.
* β1​, β2​, and β3 are the coefficients for the independent variables.
* ϵ is the error term.

**2. Interpretation of Key Results**

**Coefficients**

The coefficients provide insights into the relationship between each independent variable and the dependent variable (price):

* **Intercept**: The intercept (−97,560-97,560−97,560) represents the baseline price when all independent variables are zero. Although a zero floor size or no bathrooms is unrealistic, the intercept provides the foundational level of the model.
* **Price per Square Foot (PRICE\_SQFT)**: With a coefficient of **1368.67**, this variable has the strongest effect on price. A one-unit increase in price per square foot corresponds to an average increase of **$1,368.67** in the property price. This shows that properties with a higher price per square foot are, unsurprisingly, more expensive overall.
* **Number of Bathrooms (NO\_OF\_BATHROOMS)**: Each additional bathroom increases the property price by **$496.99**. This indicates that buyers tend to value additional bathrooms but that it has a relatively modest effect on price compared to other factors.
* **Floor Size (FLOOR\_SIZE)**: With a coefficient of **69.70**, each additional square foot of floor space increases the price by about **$69.70**. This result suggests that floor size is a critical driver of price, with larger properties commanding a higher price.

**Statistical Significance (p-values)**

The **p-values** for each predictor help determine whether each independent variable has a statistically significant effect on price:

* **PRICE\_SQFT** and **FLOOR\_SIZE** have p-values close to zero, which means they are highly significant predictors of price. These results support the idea that these features strongly influence property prices.
* **NO\_OF\_BATHROOMS** has a p-value of **0.044**, which is below the threshold of 0.05. This confirms that, although its effect on price is smaller, the number of bathrooms still significantly impacts price at a 95% confidence level.

**3. Model Performance**

* **R-squared**: The R-squared value of **0.985** indicates that 98.5% of the variance in price can be explained by the model. This high R-squared value suggests an excellent model fit, meaning the independent variables together provide a strong prediction of the price.
* **Adjusted R-squared**: At **0.985**, the adjusted R-squared remains nearly identical to the R-squared value, meaning that the model doesn’t suffer from unnecessary complexity or overfitting with the chosen variables.
* **F-statistic and Prob(F-statistic)**: The F-statistic of **7839** and an associated p-value near zero (indicated by **Prob(F-statistic) = 2.96e-323**) further confirm that the model as a whole is statistically significant. This supports the claim that the independent variables collectively provide substantial predictive power.

**Summary and Hypothesis Testing**

* Given these results, we **reject the null hypothesis (H0)**, which posited that there is no significant effect of the independent variables on price. The evidence strongly supports the **alternative hypothesis (H1)**, which suggests that **price per square foot, number of bathrooms, and floor size all have a statistically significant effect on the property price**.
* This model can now provide meaningful insights for pricing predictions in similar property markets and assist in understanding how various features drive the price.

**Conclusion**

Given the significant coefficients and p-values, we reject the null hypothesis (H0) and conclude that **price per square foot, number of bathrooms, and floor size significantly affect the property price** for a high-floor, spacious 2-bedroom house.