**Capstone Project 1: Exploratory Data Analysis**

The Housing pricing dataset from the kaggle contains 1500 rows with 81 columns including target variable Price. After filling the missing values and removing the outliers. The below statistical inferences are made on the train dataset.

Performed hypothesis test to check the significant relationship between target variable Price and No. of bathrooms. The p-value for the hypothesis test is less than the level of significance 0.05, so we reject the null hypothesis. So we support that there is a correlation between a number of bathrooms and price.

I also conducted a hypothesis test to check the correlation between a total area and price. The p-value for the hypothesis test is less than the level of significance 0.05, so we reject the null hypothesis and suggest that there is a correlation between a total area and price.

Similarly, I conducted a hypothesis test to check the correlation between overall quality and price. The p-value for the hypothesis test is less than the level of significance 0.05, so we reject the null hypothesis and suggest that there is a correlation between overall quality and price.

Conducted a hypothesis test to check the correlation between a three season porch area in square feet and price. The p-value for the hypothesis test is greater than the level of significance 0.05, so we accept the null hypothesis and suggest that there is no correlation between three season porch area in square feet and price.

I also conducted a hypothesis test to check if there is no statistical importance between mean house price and a number of bedrooms less than 3 and greater than 3. The p-value for the test was greater than the level of significance 0.05, so we fail to reject the null hypothesis. This suggests us that there is no statistical importance between mean house price and a number of bedrooms less than 3 and greater than 3.