**Santosh Singh**

**Project 3 House Price Prediction Model:**

In this analysis, we applied a linear regression model to predict house prices based on the lot area using the Ames Housing dataset. The dataset was first cleaned by handling missing values: numerical columns were imputed with their mean, while categorical columns were filled with the most frequent values (mode). Duplicates were removed, and categorical variables were converted to numerical form using one-hot encoding.

We selected the "Lot Area" as the feature and "SalePrice" as the target for the regression model. The dataset was split into training and testing sets, with 80% used for training and 20% for testing. A Linear Regression model was trained on the training data, and predictions were made on the test set.

Model performance was evaluated using three metrics: Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared. These metrics gave us an indication of how well the model was able to predict the sale price based on lot area.

Finally, the results were visualized using scatter plots. Training data was shown in blue, testing data in green, and the regression line was plotted in red. This allowed us to visually assess the relationship between lot area and house price, with the regression line representing the model's predictions.