Project Guideline

**Steps Followed in the Analysis**

## **1. Importing Data:**

- The dataset is imported and the top 5 rows are inspected to understand the structure and initial content.

## **2. Inspecting Dataset Shape:**

- The shape of the dataset is checked to determine the number of rows and columns.

## 

## **3. Dataset Information:**

- Detailed information about the dataset, including the number of rows, columns, data types, and memory requirements, is retrieved.

## 

## **4. Checking for Null Values**:

- The dataset is checked for any missing values in the columns.

## **5. Overall Statistics:**

- Descriptive statistics of the dataset are obtained to get an overview of the data distribution.

## **6. Data Processing:**

- An additional column is added to describe the age of the car.

- The 'Year' column is dropped after calculating the age.

## **7. Outlier Removal:**

- Outliers in the 'Selling\_Price' column are identified using a boxplot.

- Rows with outliers are examined and removed to improve model accuracy.

## 

## **8. Encoding Categorical Data:**

- Categorical variables are encoded using one-hot encoding to convert them into numerical format.

## 

## **9. Splitting Data into Features and Targe**t:

- The dataset is split into features (X) and target variable (y).

## 

## **10. Splitting Data into Training and Testing Sets:**

- The data is divided into training and testing sets for model validation.

## 

## **11. Feature Scaling:**

- StandardScaler is used to standardize the feature values.

## 

## **12. Model Training:**

- Four different regression models (Linear Regression, RandomForestRegressor, GradientBoostingRegressor, and XGBRegressor) are trained on the training data.

## 

## **13. Prediction on Test Data:**

- Predictions are made on the test data using the trained models.

## 

## **14. Evaluating Model Performance:**

- The performance of each model is evaluated using the R^2 score.

- The scores are compared to identify the best model.

## 

## **15. Saving and Training the Best Model:**

- The best model (XGBRegressor) is re-trained on the entire dataset and saved for future use.

## 

## **16. Prediction on New Data:**

- Predictions are made using the best model on new data.

Findings & Recommendation

**Findings and Recommendations**

**# Findings:**

**1. Dataset Overview:**

- The dataset consists of 8 columns and a significant number of rows, providing a robust foundation for the analysis.

**2. Data Integrity:**

- No null values were found, indicating a clean dataset ready for analysis.

- Outliers in the 'Selling\_Price' column were identified and handled appropriately.

**3. Feature Engineering:**

- A new feature 'Age' was created to better capture the impact of a car's age on its selling price.

- Categorical variables were successfully encoded for model compatibility.

**4. Model Performance:**

- Four models were trained and evaluated:

- Linear Regression: R^2 score = 0.65

- Random Forest Regressor: R^2 score = 0.86

- Gradient Boosting Regressor: R^2 score = 0.88

- XGBRegressor: R^2 score = 0.89

**- The XGBRegressor outperformed other models with an R^2 score of 0.89, indicating it is the best model for this dataset.**

**5. Model Deployment:**

- The XGBRegressor model was retrained on the entire dataset and saved for production use.

- The model was tested on new data and performed accurately.

**# Recommendations:**

**1. Model Adoption:**

- Implement the XGBRegressor model in production for car price prediction tasks due to its superior performance.

**2. Feature Expansion:**

- Consider adding more features related to car condition, market trends, and economic factors to further enhance model accuracy.

**3. Regular Updates:**

- Periodically retrain the model with new data to maintain its accuracy and relevance.

**4. Monitoring and Maintenance:**

- Establish a monitoring system to track the model's performance and make necessary adjustments.

**5. User Training:**

- Train relevant team members on how to use the model for making predictions and interpreting results.

By following these steps and recommendations, we can ensure that the car price prediction model is effectively integrated into our processes, leading to more accurate and reliable predictions.