Project Proposal: Car Price Prediction

1. Introduction

The objective of this project is to predict car prices using linear regression and hypothesis testing. By analyzing the factors that influence car prices, we aim to build a predictive model that can estimate the price of a car based on its features and specifications.

2. Data

We will utilize a dataset containing information about various factors affecting car prices, such as brand, model, horsepower, mileage, safety features, and GPS. The dataset will also include other relevant variables, such as car age, fuel type, and transmission type.

3. Methodology

3.1. Exploratory Data Analysis (EDA)

Conduct EDA to understand the distribution of car price data and explore the relationships between car features and prices using visualizations and summary statistics.

3.2. Linear Regression

Build a linear regression model with car features as the independent variables and car prices as the dependent variable.

Evaluate the significance of car features in predicting car prices.

3.3. Hypothesis Testing

Formulate hypotheses to test the significance of the relationship between car features and car prices.

Perform hypothesis testing using appropriate statistical tests to validate the hypotheses.

4. Expected Outcomes

The project outcomes will include a predictive model that can estimate car prices based on car features and specifications. Additionally, the analysis will provide insights into the key factors that influence car prices, enabling the company to make data-driven decisions to optimize pricing strategies and improve customer satisfaction.

5. Conclusion

In conclusion, this project aims to provide a comprehensive analysis of car prices using linear regression and hypothesis testing. By leveraging these analytical techniques, the company seeks to gain actionable insights that can drive improvements in pricing strategies and customer satisfaction.

PROMPT:

statistical test for car price prediction

ChatGPT:

When developing a car price prediction model, you typically use regression analysis to establish a relationship between the independent variables (features) and the dependent variable (car price). However, after building the model, it's essential to assess its performance and the statistical significance of the features. Here are some statistical tests and metrics you might use in the context of car price prediction:

1.Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE):

These metrics are not tests but are crucial for evaluating the accuracy of your prediction model. Lower values indicate better performance.

2.R-squared (R²):

R-squared measures the proportion of the variance in the dependent variable (car price) that is predictable from the independent variables. A higher R-squared indicates a better-fitting model.

3.F-Test:

The F-test assesses the overall significance of your regression model. It compares the fit of your model with a model that has no predictors (null model). A significant F-test suggests that your model is better than no model at all.

4.t-Tests for Individual Coefficients:

Conduct t-tests on individual coefficients to assess the statistical significance of each predictor variable. A low p-value indicates that the variable is likely a meaningful addition to the model.

5.Durbin-Watson Test:

This test checks for autocorrelation in the residuals. Autocorrelation can be an issue when observations are not independent over time or space. The Durbin-Watson statistic ranges from 0 to 4, with 2 indicating no autocorrelation.

6.Heteroscedasticity Tests (e.g., Breusch-Pagan Test):

Heteroscedasticity refers to non-constant variance of residuals. The Breusch-Pagan test can help determine whether the variance of residuals is consistent across all levels of the independent variables.

7.Jarque-Bera Test:

The Jarque-Bera test assesses whether the residuals of your model have a normal distribution. Deviations from normality may impact the reliability of statistical inference.

8.VIF (Variance Inflation Factor):

VIF helps identify multicollinearity among independent variables. High VIF values suggest a high degree of correlation between predictors, which might affect the model's stability and interpretability.

9.Feature Significance Testing:

Depending on the nature of your data, you might want to conduct feature significance testing to determine which features significantly contribute to predicting car prices.

Remember that the choice of statistical tests should align with the assumptions of the regression model and the characteristics of your data. Always validate the assumptions, interpret the results, and consider the practical significance of your findings in the context of car price prediction.

Car accidents:  
  
https://www.kaggle.com/code/kenethm/accidents-in-kenya/notebook