**TEAM 6**

Yashaswini Reddy

Manya Gupta

Mitesh Patel

Vijay Chandra Atheli

MISM6212

**Business Question:**

Purchasing a used car can be a daunting experience for many customers, especially without the technical knowledge or access to historical market data, buyers often have to rely solely on the seller's word, which may lead to overpaying or falling for overpriced deals.

This project aims to develop a Car Price Prediction Model that assists used car buyers in making informed decisions. By leveraging factors such as the car’s mileage, fuel type, transmission type, engine size, and year of manufacture, the model predicts a fair market price for a given vehicle. Customers can then compare this predicted price to the dealer’s offer and judge whether they are getting a reasonable deal or being overcharged.

The goal is to build a reliable, data-driven support tool that brings transparency to used car pricing and empowers customers with confidence in their purchasing decisions.

**Data Sources:**

100,000 UK Used Car Dataset: This dataset includes used car listings from multiple brands, featuring attributes such as year, price, mileage, fuel type, transmission, and more.

[Used Car Dataset - vw, audi, bmw, ford, hyundai, mercedes, skoda, Toyota and vauxhall](https://www.kaggle.com/datasets/adityadesai13/used-car-dataset-ford-and-mercedes)

**Data Quality Concerns:**

**Potential Issues:**

* Missing or Incomplete Data – Some listings may lack information on mileage, price, or fuel type.
* Outliers & Anomalies – Unusually high or low prices may distort predictions.
* Duplicate Listings – Multiple records of the same vehicle may introduce bias.
* Categorical Data Encoding – Transmission types and fuel types need encoding for machine learning models.

**Resolution Plan:**

* Data Cleaning: Remove duplicates, fill missing values using median imputation, and handle outliers with statistical techniques.
* Feature Engineering: Convert categorical variables (e.g., fuel type, transmission) into numerical representations using one-hot encoding or label encoding.
* Normalization/Scaling: Standardize numerical features like mileage and engine size to improve model performance.
* Exploratory Data Analysis (EDA): Identify patterns in pricing across different car features.

**Methods:**

**Example:**

**Data Mining Techniques:**

* Descriptive Statistics: Summarize key trends in car pricing, mileage, and fuel efficiency.
* Correlation Analysis: Identify relationships between price and features such as mileage, engine size, and fuel type.
* Clustering: Group similar cars based on pricing and mileage patterns to identify optimal pricing strategies.

**Machine Learning Models:**

* Linear Regression: Predict car prices using key features like year, mileage, and fuel type.
* Decision Trees: Identify key factors influencing price variations and classify cars into price brackets.
* Random Forest: Improve price predictions by aggregating multiple decision trees.

**Tools:**

* Python Libraries: Pandas, Scikit-learn.
* Data Visualization: Matplotlib, Seaborn for dashboard creation.

**Preliminary results:**

Not yet available – The dataset is being preprocessed, and exploratory analysis is in progress.

**Challenges:**

Handling Missing Data – Using imputation techniques to fill missing values.

Feature Selection & Engineering – Identifying the most important features affecting car prices.

Overfitting in Models – Using cross-validation and hyperparameter tuning to prevent overfitting.

Changing Market Trends – Ensuring the model adapts to fluctuations in used car prices.

**Proposed Solution:**  
Data Imputation for missing values (median for numerical, mode for categorical).

Feature Engineering to create additional relevant predictors (e.g., car age).

**Team Member Contributions:**

Yashaswini Reddy: Data collection and formatting.

Manya Gupta: Python data mining (descriptive statistics, correlation analysis).

Mitesh Patel: Machine learning model development (Linear Regression, Decision Trees).

Vijay Chandra Atheli: Data integration, Data cleansing, Application development and UI/UX design.

**Conclusion:**

This project aims to develop a data-driven pricing model to assist customers in making informed pricing decisions. By leveraging machine learning techniques, the model will help customers understand historical pricings and make decisions.