**Car Price Prediction Project Report**

**1.Introduction**

Brief overview of the project objective predicting the car price based on the various features like Engine type, transmission type etc

**2.Data cleaning**

Data cleaning ensures consistency and quality of data, which is crucial for model performance.

**Steps involved**

* **Loading data**
* Load the dataset using pandas
* **Removing Duplicates**
* Remove the duplicate rows, if any
* **Handling missing values**
* Identify missing values using data. isnull (). sum ().
* Decide on an imputation strategy, such as filling missing values with the median or mean for numerical columns, and the mode for categorical columns, or dropping columns if appropriate.
* **Outlier Detection and Handling:**
* Use boxplots to identify outliers in numerical columns.
* Decide whether to remove, cap, or transform outliers based on their impact on model performance.
* **Data Type Conversion:**
* Convert columns to appropriate data types (e.g., integer, float, categorical) for consistency and performance.

**3. Data Preprocessing**

Data preprocessing involves preparing data machine learning model by encoding categorical variables, scaling numerical features, etc.

**Steps Involved:**

* **Encoding Categorical Variables:**
* Label Encoding for binary categories.
* One-Hot Encoding for categorical variables with multiple categories using pd.get dummies ().
* **Scaling Numerical Features:**
* Use StandardScaler to standardize features for algorithms sensitive to feature scale, like regression or KNN.

**4. Exploratory Data Analysis (EDA)**

EDA is used to analyze patterns, detect anomalies, and check assumptions before proceeding with machine learning.

**Steps Involved:**

* **Data Summary and Structure**:
* Show summary statistics and data types.
* **Distribution Analysis**:
* Plot histograms for numerical columns to understand distributions.
* **Correlation Analysis**:
* Use a correlation matrix to understand relationships between features and the target variable.
* **Correlation Analysis**:
* Use a correlation matrix to understand relationships between features and the target variable.
* **Outlier Analysis**:
* Use boxplots to identify outliers in key numeric variables.
* **Feature-Target Relationships**:
* Analyze how features like Fuel Type, Transmission Type, etc., relate to the target Car Price.

**5. Machine Learning**

The machine learning section involves training

and evaluating models to predict car prices.

**Steps Involved:**

* **Train-Test Split**:
* Split data into training and testing sets.
* **Model Selection**:
* Try different regression models (e.g., Linear Regression, Decision Tree, Random Forest, Gradient Boosting).
* **Training and Evaluation**:
* Train the models on the training set.
* Evaluate model performance using metrics like RMSE, MAE, and R² on the test set.
* Use cross-validation for more robust evaluation.

**6. Building the Streamlit Application**

Finally, deploy the model using a Streamlit application to make it accessible to users.

**Steps Involved:**

* **Install Streamlit**:
* Install Streamlit using pip install streamlit.
* **Create a Streamlit App**:
* Add input fields for all features (e.g., fuel type, transmission type, etc.).
* Preprocess inputs from the user to match model requirements.
* Display the predicted car price based on user inputs.