Problem Statement

Background & Context

This dataset contains listings of used cars and their associated attributes, including price, brand, model year, condition, mileage, fuel type, and location. It is typically used for market analysis, price prediction, and demand-supply insights in the used automobile sector. The data reflects real-world car listings and can support applications in business analytics, machine learning models for price estimation, and consumer behavior studies.

Objective

The objective of this assignment is to develop hands-on proficiency in data analysis using the Pandas library in Python. Learners will work with a real-world car listings dataset to perform essential data wrangling tasks such as loading, cleaning, filtering, grouping, and summarizing data. By the end of the assignment, learners will be able to extract meaningful business insights, visualize key trends, and apply foundational techniques for exploratory data analysis (EDA) in a practical context.

Tasks

1. Data Ingestion & Quality Profiling

1. Load & Inspect
   - Read car_prices.csv into a pandas DataFrame.
   - Display the first 5 rows.
   - Display data types and record count.

1.2. Understanding the Data Structure

- Check the shape of the dataset (rows and columns).
- Display column names and data types.

1.3 Missing & Anomaly Detection

- Quantify nulls per column; visualize with a bar chart or heatmap.
- Resolve null values by appropriate strategy based on the datatype and percentage of null values.
- Count number of duplicate records and delete if any present

2. Data frames Queries:

2.1 Calculate the average, minimum, and maximum car price?

2.2 List all unique colors of cars?

2.3 Find the number of unique car brands and car models?

2.4 Find all car information having selling prices greater than $165000?

2.5 Find the top 5 most frequently sold car models?

2.6 What is the average selling price of cars by brand (make)?

2.7 What is the minimum selling price of cars for each interior?

2.8 Find highest odometer reading per year from highest to lowest order?

2.9 Create a new column for car age (assuming the current year is 2025)

2.10 Find the number of cars having a condition greater than or equal to 48 and odometer greater than 90000?

2.11 Which state consistently has higher car prices for newer cars(year>2013)?

2.12 For cars with excellent condition (top 20%), which makes have the lowest average price (value for money)?

3. Data Visualization and insights:

3.1 Show the correlation of all the features(columns) having numerical values(e.g. selling price, odometer).

3.2 Plot a graph to show the average selling price by year and explain the pattern you notice from the graph. Which plot will you use bar or scatter?

3.3 Plot a graph to show average selling price by odometer and explain the trend you notice from the graph.

3.4 Plot a graph to show the number of cars sold in each state. Find the top three highest car selling states by having a look at the graph?

3.5 Plot a bar graph of average selling price by condition score ranges of size 5. Give a summary of the insights.

3.6 Plot a bar graph of no. of cars sold by condition ranges of size 10. Give some insights from the graph.

3.7 Plot a box plot to show the distribution of car selling prices grouped by color. Extract insights and if required, remove outliers and plot the graph again.