Project diagram

Dataset information

This data set contains 90,000+ car data from 1970 to 2024, the dataset contains the following columns:

- Model: The model of the car
- Year: The manufacturing year of the car.
- Price: The price of the car.
- Transmission: The type of transmission used in the car.
- Mileage: The mileage of the car.
- FuelType: The type of fuel used by the car.
- Tax: The tax rate applicable to the car.
- MPG: The miles per gallon efficiency of the car.
- EngineSize: The size of the car's engine.
- Manufacturer: The manufacturer of the car.

Project Steps:

Suppose we work for a company that imports and exports cars throughout Europe. The company is considering a significant investment in purchasing used cars for resale. Among these cars, some are vintage, and there's a particular interest in their profitability. The project manager has approached us to research this matter:

Step 1: First, we need to understand the problem:

- What does the company want?
- How can we solve the matter?
- What do the stakeholders state as a problem?

Step 2: We need to prepare our tools and decide what we need to do.

- What research is needed?
- What needs to be figured out how to solve this problem?
- Where is the data located (files, database, external system, internal system)?

Step 3: We need to process our data

- We have to study our dataset and process it, by cleaning, fixing, or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
- We have to identify if our data are bias
- We need to use proper tools to find incorrect and incomplete data.
- We also need to remove inconsistencies in data.

Step 4: Now we can start the analysis:

- We have to perform different calculations and get additional metrics
- We can create different views for the data. Like tables with our results, filter and pivot them.
- Make charts and viz
- What story is our data telling us?

Step 5: Share our data!

- After the feedback we have to improve the general outcome. From one angle, the decision will most likely be more informed and better, but also the transparency will grant that there is more support to the findings.
- What would help us understand the viz if we were the listeners?
- We need to wait for the company's feedback

Step 6: Act! We know our results, let's start!

- What potential solutions to the outlined problem could there be?
- Is this problem worth solving? (Yes, that is also a potential outcome)
- How can the feedback received during the sharing phase (step 5) be used to meet the stakeholder’s needs and expectations?

Project Steps in-depth:

As we said before, the company is considering a significant investment in purchasing used cars for resale. This dataset contains some important columns that we can use to extract our results: Some of them are, the model of the cars and the year which the car was manufactured. Another important element is the current price of the cars that are listed in the market. Furthermore, there is the type of transmission used in the cars. We have to consider that, this dataset represents data from the UK market, thus in EU, manual transmissions are preferred because they are more suitable because for the small, narrow roads and crowded cities of EU countries.

Step 1: First, we need to understand the problem:

- What does the company want?
- How can we solve the matter?
- What do the stakeholders state as a problem?

A. The company is considering a significant investment in purchasing used cars for resale
B. We need to see if we can find more datasets that can assist that solve the problem
C. The problem that the stakeholders state, is that they are not sure what cars should consider purchasing. They want to see how many cars are available and “worth” buying.
    1. Which manufacturer?
    2. Do people want big size engines or small?
    3. How much is the tax for the car?
    4. How much does vintage car costs? We have already defined that vintage cars as those that are at least 25 years old from the current year (2024), so cars manufactured in 1999 or earlier.

Step 2: We need to prepare our tools and decide what we need to do.

- What research is needed?
- What needs to be figured out how to solve this problem?
- Where is the data located (files, database, external system, internal system)?

A. They made the research, searching in public datasets and we decide that the current dataset is best suited for our project.
B. We need to figure out what we need to calculate to reach our objective.
C. We know that our data located to Kaggle. We downloaded the dataset and it’s in .csv file.

Step 3: We need to process our data

- We have to study our dataset and process it, by cleaning, fixing, or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.
- We have to identify if our data are bias
- We need to use proper tools to find incorrect and incomplete data.
- We also need to remove inconsistencies in data.

A. The have study our dataset and we have imported in db browser, we are going to use SQlite and Python-pandas for our analysis. The dataset is cleaned. We have removed duplicates and fixed incorrectly formatted or incomplete data within our dataset.
B. We know which data are bias and what we can use, to solve our problem.

C. We used Excel, SQL and Python(pandas) to find incorrect and incomplete data.
D. We removed inconsistencies in data and replaced missing values.

Step 4: Now we can start the analysis:

- We have to perform different calculations and get additional metrics
- We can create different views for the data. Like tables with our results, filter and pivot them.
- Make charts and viz
- What story is our data telling us?

A. Some of them are:

1a. Vintage Cars

This query selects all cars manufactured in or before 1999. Analyzing vintage cars separately can reveal trends and preferences specific to older vehicles, including their valuation, rarity, and characteristics that appeal to collectors and enthusiasts.

2a. Average Summary Metrics

By calculating rounded averages and extreme values for price, mileage, MPG, engine size, and the year, along with the minimum and maximum prices, you gain a comprehensive statistical overview of your dataset. This helps in understanding the general pricing, efficiency, and usage patterns of the cars in your dataset.

3a. Average Car Prices Over Years

Tracking the average price of cars over the years allows you to identify pricing trends, such as inflation effects, changes in consumer buying power, and shifts in manufacturer pricing strategies. This can be crucial for understanding market dynamics over time.

4a. Counts of Car Manufacturers

Counting the number of cars per manufacturer gives insight into market dominance and the diversity of offerings. It helps identify which brands are most prevalent in your dataset, which can be an indicator of popularity, brand loyalty, or market saturation.

5a. Counts of Car Models per Manufacturer

This query provides a deeper dive into each manufacturer's range, showing which ones offer a wider variety of models. This can indicate a manufacturer's market strategy, targeting different consumer segments or preferences.

6a. Average Price, MPG, and Engine Size by Manufacturer

Comparing manufacturers on the basis of average price, fuel efficiency (MPG), and engine size offers insights into their market positioning (e.g., luxury vs. economy), technological focus (e.g., fuel efficiency vs. performance), and potential consumer demographics.

7a. Transmission Counts

Understanding the distribution of transmission types (e.g., manual, automatic) across your dataset can highlight consumer preferences, technological advancements, and regional trends in driving habits or vehicle functionality.

B. Why do we perform these analyses?

1b. Strategic Insights: These queries help automotive businesses, analysts, and enthusiasts understand market trends, consumer preferences, and competitive landscapes.

2b. Historical Trends: Analyzing data from 1970 to 2024 provides valuable insights into how the automotive industry has evolved over more than five decades.

3b. Decision Support: For businesses, these insights can support strategic decisions related to inventory management, marketing strategies, and future investments. For consumers, it can inform buying decisions.

4b. Market Segmentation: Understanding different segments (e.g., vintage cars, high MPG vehicles) enables targeted analysis, which is crucial for marketing, restoration projects, and more.

5b. Technological and Environmental Impact: Analyzing trends in engine size, MPG, and the popularity of fuel types can provide insights into the automotive industry's impact on the environment and its response to technological advancements and environmental regulation

Step 5: Share our data!

- After the feedback we have to improve the general outcome. From one angle, the decision will most likely be more informed and better, but also the transparency will grant that there is more support to the findings.
- What would help us understand the viz if we were the listeners?
- We need to wait for the company's feedback

Step 6: Act! We know our results, let's start!

- What potential solutions to the outlined problem could there be?
- Is this problem worth solving? (Yes, that is also a potential outcome)
- How can the feedback received during the sharing phase (step 5) be used to meet the stakeholder’s needs and expectations?