|  |  |  |
| --- | --- | --- |
| Checkpoint II | Checkpoint II: Data Cleaning & Processing | |
| Group: | G16 |
| Date: | 2023/09/25 |
|  |  |

# Initial Dataset

# The objective of our information visualization project revolves around electric cars and harnesses data from various sources, primarily focusing on information related to electric vehicles (EVs) and their attributes.

# These datasets have been meticulously sourced to ensure accuracy and relevance for our visualization. Kaggle’s have combined all the data in one dataset, on a tabular format, providing an easier interpretation.

The initial dataset has 102 rows and 15 columns (total size: 102x15=1530).

# Uma imagem com texto, captura de ecrã, Tipo de letra, número Descrição gerada automaticamente

# Selected/Derived Data

With every quantitative attribute, we can find the average value for each one.

|  |  |
| --- | --- |
| **Selected Attributes** | **Derived Attributes** |
| BodyStyle, TopSpeed\_Kmh | \*avg\_top\_speed\_per\_body\_style |
| FastCharge\_Kmh, PowerTrain | \*fast\_charging\_rate\_per\_power\_train |
| Efficiency\_WhKm, BodyStyle | \*avg\_efficiency\_per\_body\_style |
| Efficiency\_WhKm, Segment | \*avg\_efficiency\_per\_segment |
| TopSpeed\_Kmh | \*avg\_top\_speed |
| Efficiency\_WhKm | \*avg\_efficiency |
| Battery Pack | \*avg\_acceleration\_speed |
| Acceleration Speed | \*avg\_battery\_pack |
| Range | \*avg\_range |
| Price | \*avg\_price |
| Fast Charging | \*avg\_fast\_charging |

# Data Abstraction

|  |  |  |  |
| --- | --- | --- | --- |
| **Type** | **Scale** | **Field Name** | **Semantics** |
| Quantitative (ratio) | Sequential | “avg\_top\_speed\_per\_body\_style” | Average top speed per body style, in km/h |
| “avg\_efficiency\_per\_body\_style” | Average efficiency per body style, in Wh/km |
| “avg\_efficiency\_per\_segment” | Average efficiency per segment, in Wh/km |
| “fast\_charging\_rate\_per\_power\_train” | How powertrains affect fast charging rate |
| “avg\_top\_speed”, “avg\_efficiency”, “avg\_price”, “avg\_fast\_charging”, “avg\_range”, “avg\_battery\_pack”, “avg\_acceleration\_speed” | Average of each quantitative attribute |
| “AccelSec”, “TopSpeed\_KmH”, “Range\_Km”, “Battery\_Pack Kwh”, “Efficiency\_WhKm”, “FastCharge\_KmH”, “Price” | Specifications of each electric vehicle |
| Nominal |  | “Brand”, “Model”, “RapidCharge”, “PowerTrain”, “Body Style”, “Segment” | Specifications of each electric vehicle |
| Does not exist hierarchy between data attributes. | | | |

# Data Processing

# It was determined that "Seats" and "PlugType" attributes could be safely removed from the dataset. As the dataset doesn’t contain significant missing values, there was no need for imputation or removal of rows with missing values.

# Once the data processing and attribute removal steps were completed, the cleaned dataset was exported in CSV format for compatibility with D3.

df = pd.read\_csv('ElectricCarData\_Clean\_Me.csv')

df.drop(columns=["PlugType", "Seats"], inplace=True)

df.to\_csv('cleaned\_dataset.csv', index=False)

# Mapping (Data sample/Questions)

1 – What is the relationship between a car's acceleration speed and its price in euros?

{“AccelSec”: “4.6”, “PriceEuro”: “55480”}

2 – Which car brands offer electric vehicles with the longest range (in kilometers) and the highest top speed?

{“Brand”: “Tesla”, “Range\_Km”: “460”, “TopSpeed\_KmH”: “233”}

3 – What proportion of electric cars have a top speed (Km/H) exceeding 150 Km/H, and does this proportion change based on the body style of the vehicles?

{“TopSpeed\_KmH”: “233”, “BodyStyle”: “Sedan”}

4 - What is the distribution of electric cars' fast charge capabilities across different power train types?

{“FastCharge”: “Yes”, “PowerTrain”: “AWD”}

5 – Can we identify any recurring design trends in body styles among electric cars with the highest efficiency (Wh/Km) and do these trends vary by market segment?

{“BodyStyle”: “Sedan”, “Efficiency\_ WhKm”: “161”, “Segment”: “D”}