Electric Cars Details

Eduardo Lobo
99213
IST – Alameda
eduardo.lobo@tecnico.ulisboa.pt

Margarida Estrela
99271
IST – Alameda
margaridaestrela@tecnico.ulisboa.pt

Pedro Rodrigues
99300
IST – Alameda
pedro.dias.rodrigues@tecnico.ulisboa.pt

INTRODUCTION

Electric cars have become a fundamental aspect of modern transportation, reshaping the automotive industry and influencing our environmental sustainability efforts. Furthermore, the adoption and understanding of electric vehicles (EVs) play a crucial role in our daily lives, and their impact extends beyond individual convenience. Electric cars contribute to combating climate change, making them an essential component of a sustainable future.

Analysts may inquire about the factors influencing electric vehicle efficiency or seek correlations, for example, between acceleration and battery capacity. They might also investigate how fast charging infrastructure affects EV adoption or delve into the pricing dynamics across different vehicle segments. The dataset that we study offers a wealth of opportunities to draw conclusions regarding environmental benefits derived from energy efficiency.

Through data visualization and analysis, our goal is to uncover insights into the relationship between electric vehicle adoption, trade-off battery versus prices and outcomes generated for each manufacturer. By studying electric vehicle environments, we aim to contribute to a better understanding of the role electric cars play in shaping a sustainable future for Europe and the world. With this approach, we are expecting to identify which is the brand that more contributes to sustainability.

Example Questions

This visualization allows us to answer the following example questions:

1. What is the relationship between a car's acceleration speed and its price in euros?
2. Which car brands offer electric vehicles with the longest range (in kilometres) and the highest top speed?
3. What proportion of electric cars have a top speed (Km/H) exceeding 200 Km/H, and does this proportion change based on the body style of the vehicles?
4. What is the distribution of electric cars' fast charge capabilities across different power train types?
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?

RELATED WORK

We extensively utilized a variety of websites and scientific articles, to gather an extensive pool of information and data in support of our project. One of the websites that provides us more information and some visualizations is the same that provide us the datasets. In Kaggle's website (References 4) they try to take some conclusions to questions related to if Range of vehicle is proportional to Battery Pack Capacity, if High performance EV's have lower efficiency and so many others, but they cannot produce dynamic visualizations with high interactivity. So, after we detect this problem, we elaborate an interactive solution that achieve the user comprehension.

As you can see in the following figure, to answer the first question that is if Range of vehicle is proportional to Battery Pack Capacity, they combine a bar chart (representing the 'Range (Km)' variable) and a scatter plot (representing 'Battery Pack Kwh') to analyze and visually represent the relationship between these two variables.

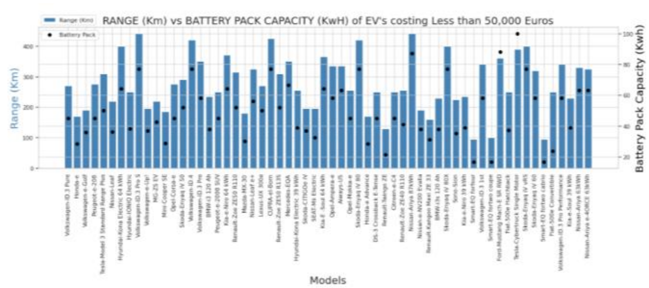

Figure 1. Example of visualization of range vs battery pack capacity of EV's in Kaggle's website

Another website that helps us starting an idea to the project is the study elaborated by the ACEA (European Automobile Manufacturer's Association. In fact, this study marks an important point of the development of this project because it is here that we decide that electric cars details will be our project theme. After this apart, this ACEA study shows the growth of the electric cars market making the comparison to different cars fuel types.

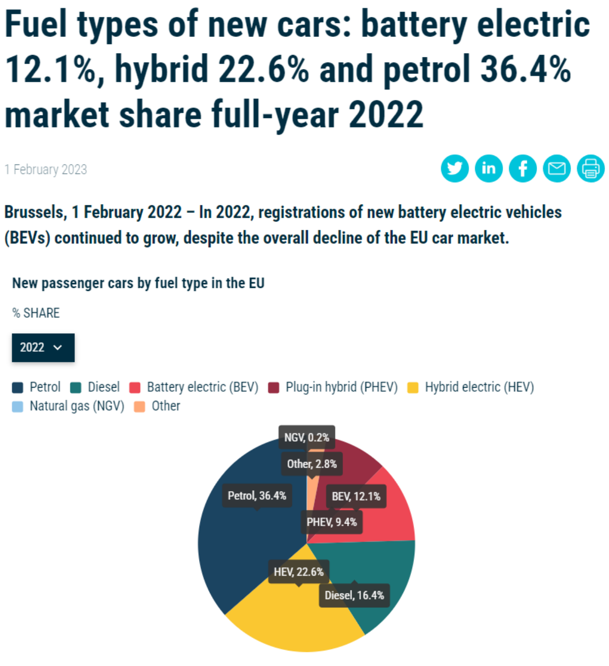

Figure 2. Fuel types of new cars – ACEA study about market growth

After some steps they concluded that alternatively powered vehicles (APVs) accounted for more than a half (53.1%) of the EU car market during the last quarter of the year.

One additional website that provides us some ideas is the US Department of Energy. The main idea is that the number of electric vehicle models grew consistently from 1991 to 2002 as flexible-fuel vehicles (FFVs) gained popularity.

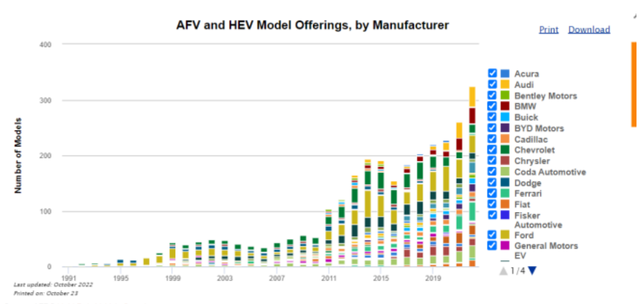

Figure 3. Electric Cars Models Offering by Manufacturer – US Department of Energy

After all this examples we got different points of electric cars details that will be helpful to create a good visualization with good user experience and usability

THE DATA

Our main source of data was the Kaggle's website, which is one website that provides various dataset related to different themes.

Initially, datasets have been meticulously sourced to ensure accuracy and relevance for our visualization. Our dataset was composed by two different .csv files, but Kaggle's have combined all the data in one dataset, on a tabular format, providing an easier interpretation.

We faced some challenges when acquiring and processing the data since at the beginning we are using a smaller dataset and becomes more difficult to establish relations between attributes. To solve this problem, we search about new datasets and found this extended 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.

VISUALIZATION

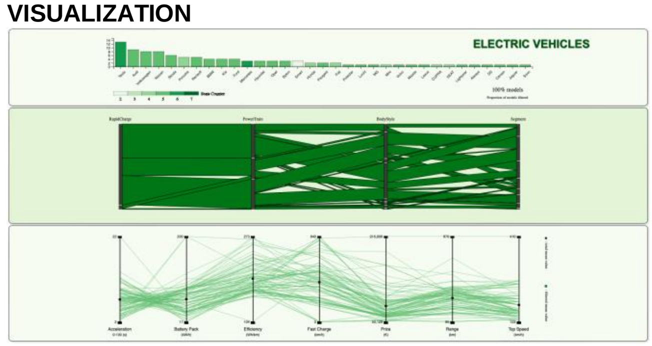

Figure 4. Layout Overview

Overall Description

Our final visualization layout consists in three different idioms, an alluvial diagram on the top, a bar chart in the middle and a parallel coordinates plot on the bottom. The layout of the visualization is shown above in figure 4.

Alluvial Diagram

The Alluvial Diagram, designed for visualizing data relationships, offers a unique perspective on the dataset.

This idiom is a crucial element of our visualization strategy, as it effectively represents the quantitative attributes within the dataset.

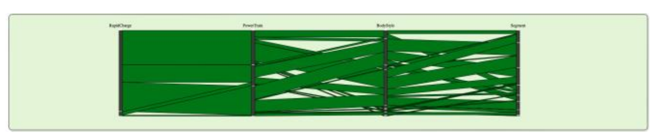

Figure 5. Alluvial Diagram for nominal attributes

This diagram provides a comprehensive view of the connections between various attributes, allowing users to explore data interactions and trends. Each colored polygon represents a distinct connection between attribute values. The stroke is the main key used to distinguish different relationships, and their saturation and opacity reflects the relative importance of each connection. The tooltips are a very useful tool that we choose to show trends and proportions of data flow.

The user can interact with this idiom by clicking on different values of attributes and highlighting them as is shown in next figure.

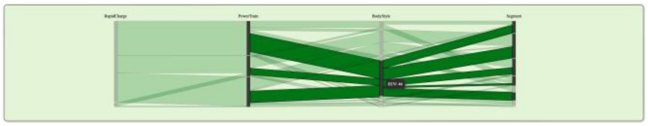

Figure 6. Alluvial Diagram – Tooltip and Select Attributes

This Alluvial Diagram is particularly valuable for tracking how attributes evolve as filters are applied, making it an essential tool for in-depth data analysis and informed decision-making. It offers an intuitive way to visualize complex data relationships, enabling users to gain insights into how different attributes are interrelated and change as the dataset is filtered.

Bar Chart

The Bar Chart is designed to provide a summarized view of the data, making it simpler to grasp key insights and trends.

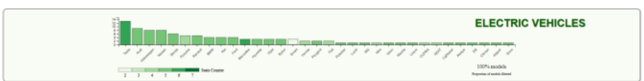

Figure 7. Bar Chart for number of models per brand.

The bar chart shows the number of models per brand and the percentage of models depending on the filters set by the other idioms. This idiom presents a straightforward visual encoding of data. As you can observe in figure 4, each bar encodes a unique brand, and their colors are thoughtfully scaled to correspond to the average number of seats offered by each one. In fact, is only used six different color to represent this attribute. At the near end of the bar chart's X axis, where a percentage value of models is displayed, you'll find a concise explanation that clarifies the purpose of this SVG text element. Essentially, it indicates the proportion of models we're filtering and analyzing in the chart.

One of the most important roles of this bar chart is enabling user to observe how brand attributes change when filters are set, making it an indispensable tool for comparative analysis and data-driven decision-making.

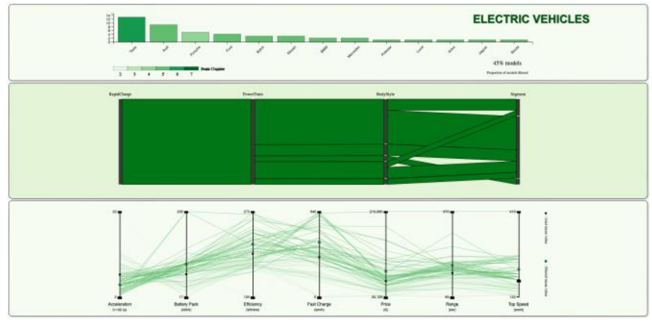

Figure 8. Analyzing Bar Char after a filter being applied

By selecting one column and after a mouseover event being called, it's possible to observe the color spectrum on the Parallel Coordinates plot lines. In blue lines, we can observe the models that correspond to mouse over event, and, in orange lines, we can observe the models that correspond to selected bar.

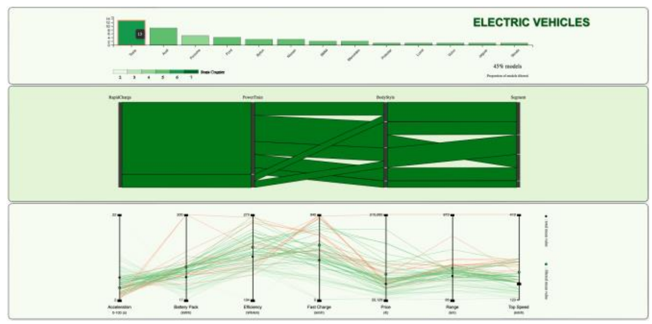

Figure 9. Bar Chart after selecting one column and mouser over event

When hovering over one of the bars a tooltip will appear a number that corresponds to the average number of seats (figure 9). Since it is not easy to understand the Y axis coordinate, the tooltip is the technique choose by us to facilitate the user comprehension.

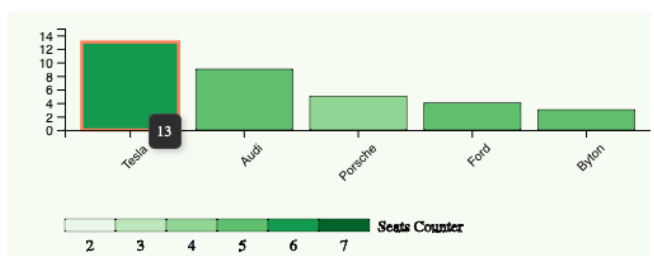

Figure 10. Bar Chart Tooltip – Number of Models per Brand

Parallel Coordinates Plot

This idiom is one of the most important since it's the focus of our subject. This graph is designed to offer insights into the relationships and trends in the data, allowing for easy comparison and analysis.

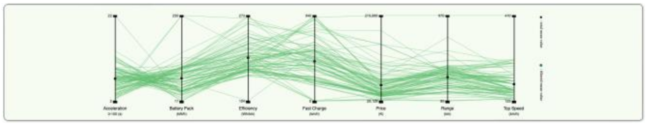

Figure 11. Parallel Coordinates to overview ordinal data

This plot is composed of six different axis each one representing an ordinal attribute. Each line encodes a model, so here the focus is about global cars details and not concrete brands details.

As we can observe in figure 8, we have implemented the ability for users to mouse over a bar in the bar chart, highlighting the corresponding lines in the parallel coordinates. We though in implementing a mouse over event on parallel coordinates but it becomes harder to select which line we want to mouse over because of lines intersections.

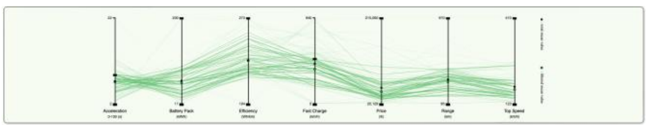

Figure 12. Parallel Coordinates filtering ranges of each Axis

In the Parallel Coordinates plot, lines representing data instances with opacity that changes dynamically based on whether they fall within the specified range of values for each filter. The user can seamlessly adjust the filters by moving the sliders up and down, instantly narrowing or broadening the scope of the data.

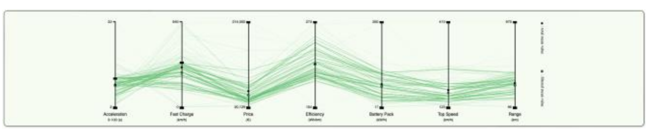

Figure 13. Parallel Coordinates Axis Switching

Additionally, for a deeper understanding of data relationships, the user has the freedom to drag and drop axes, altering their order to reveal intricate connections between attributes. These features put the user in control, allowing each one to interactively explore and analyze the dataset with ease.

Rationale and Custom Visualization

When we began developing our layout sketch, one of our primary tasks was to evaluate and select the most appropriate visualizations for our data. As a team, we engaged in numerous discussions and exchanged ideas to ultimately determine the best layout for our visualizations.

Upon comparing our initial visualization prototype (Figure 13) with the final layout version (Figure 4), we observed significant differences. Some visualization elements disappeared, others went through complete transformations, and a few new elements were introduced.

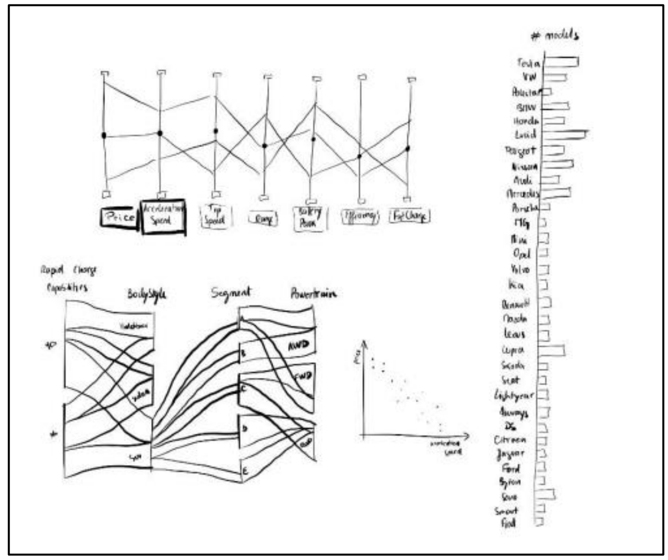

Figure 14. Initial Layout low fidelity prototype - Sketch

For instance, initially, we attempted to implement a scatter plot to establish connections between two different ordinal attributes, such as a car's price and its range. However, following feedback during the third checkpoint, we decided to eliminate the scatter plot. This decision was made because it could only correlate two attributes at a time and did not provide an easy means for users to identify clusters and other common patterns found in automotive details projects. The final position of each idiom has been changed due to preferring most information idiom located on the bottom of visualization.

The encoding we employed in our visualizations was as follows:

Variable	Data Type	Visual Encoding
“AccelSec”, “TopSpeed_KmH”, “Range_Km”, “Battery_Pack Kwh”, “Efficiency_WhKm”, “FastCharge_KmH”, “Price”	Quantitative	***Parallel coordinates:*** with sliders for each attribute **Marks:** lines, dots **Channels:** hue, stroke, opacity ***Bar chart*** **Marks:** bars **Channels:** hue, area
“Brand”, “Model”, “RapidCharge”, “PowerTrain”, “Body Style”, “Segment”	Nominal	***Alluvial:*** **Marks:** area **Channels:** hue, opacity

We chose the Alluvial Diagram as it provides a unique perspective on data relationships, particularly for categorical attributes, allowing users to explore data interactions and trends. It was chosen because it effectively represents the connections between various attributes, making it essential for in-depth data analysis. In fact, in our visualization, it was good to understand the flow of data but when one axis has so many values it could not be so user friendly.

The Bar Chart simplifies data summarization, displaying the number of models per brand and their attributes, facilitating comparative analysis and data-driven decision-making. This choice was made due to the need for a summarized view of the data, as it offers a straightforward visual encoding of information and a clear representation of how attributes change with applied filters. The percentage of models and the size of each bar are essential numbers to contextualize and facilitate user comprehension.

Finally, the Parallel Coordinates Plot allows for easy comparison and analysis of data relationships and trends, offering a global overview of ordinal data. Users can dynamically adjust filters, highlighting the connections between attributes. We selected this visualization to enable users to interactively explore and analyze the dataset with ease and discover intricate connections between attributes, making it a crucial element in our layout.

Demonstrate the Potential

To demonstrate the potential of our visualization, we will study two of the questions introduced in example questions and proof how our visualization has a solution.

1. ***What proportion of electric cars have a top speed (Km/H) exceeding 200 Km/H, and does this proportion change based on the body style of the vehicles?***

The question’s value was changed from 150Km/H to 200Km/H since the first value didn’t filter enough models.

We begin by examining the Parallel Coordinates Plot, where we focus on the "Top Speed" axis and apply a filter to display values starting at 200 Km/H.

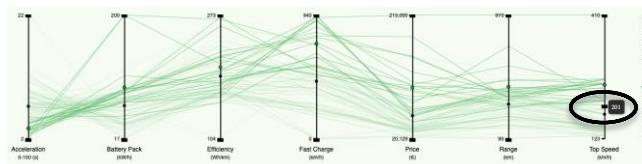

Figure 15. Parallel Coordinates – Top Speed > 200 Km/H

As we can see in Bar Chart percentage value, only 21% of total models have a top speed bigger than 200 Km/H that correspond to a total of five models that is the number of lines draw in parallel coordinates.

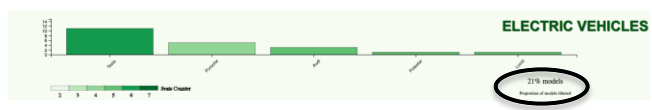

Figure 16. Bar Chart – % of Total Models with Top Speed > 200 Km/H

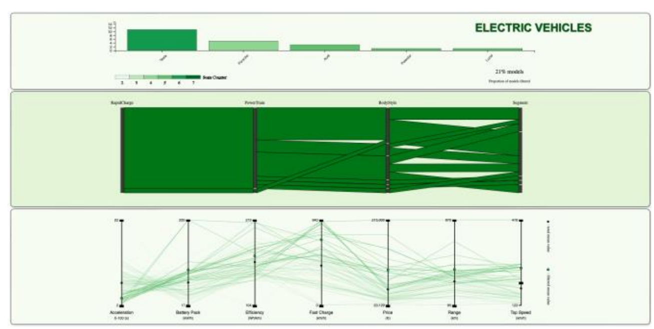

Figure 17. Visualization after filter applied

By looking at the alluvial diagram we can identify how different body styles exists and the proportion of each one. We can also click on each different value of each categorical attribute to highlight only the values we want to analyse.

In the following figures, we will present the tooltip that present the name of the body style and the number of models that each body style contains.

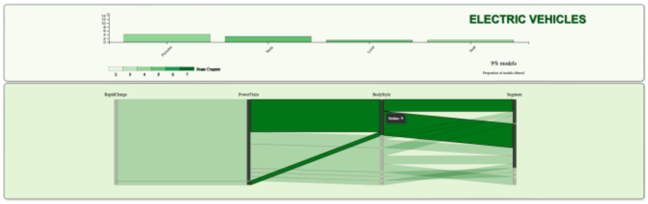

Figure 18. Alluvial Diagram Tooltip – Sedan (9 models)

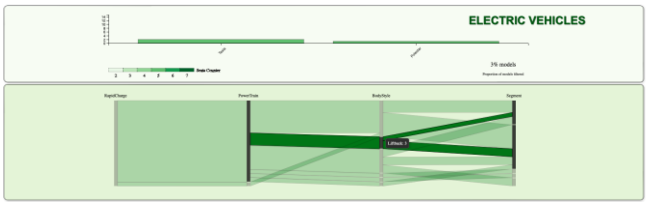

Figure 19. Alluvial Diagram Tooltip – Liftback (3 models)

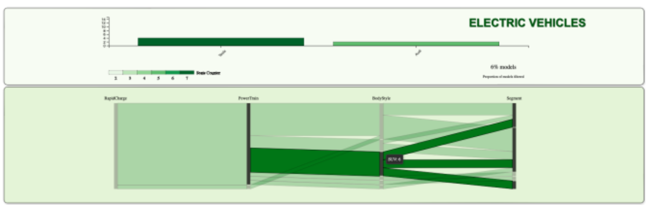

Figure 20. Alluvial Diagram Tooltip – SUV (6 models)

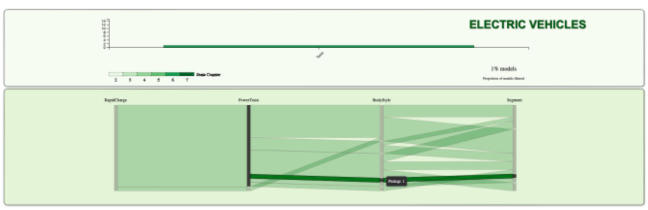

Figure 21. Alluvial Diagram Tooltip – Pickup (1 models)

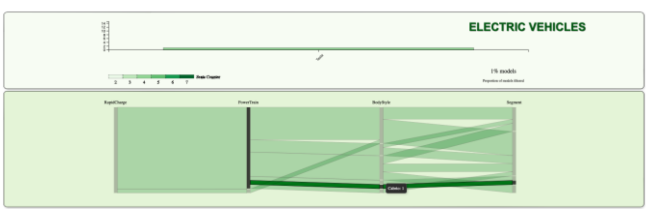

Figure 22. Alluvial Diagram Tooltip – Cabrio (1 models)

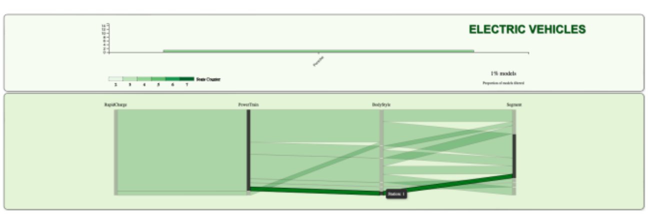

Figure 23. Alluvial Diagram Tooltip – Station (1 models)

2. ***Which car brands offer electric vehicles with the longest range (in kilometres) and the highest top speed?***

To start answering this question we must stablish some concepts due to facilitating the formulation of the answer. So, we will consider longest range as range bigger than average, and we will consider highest top speed as top speed bigger than average top speed.

So, the first step we need to filter parallel coordinates according to these considerations.

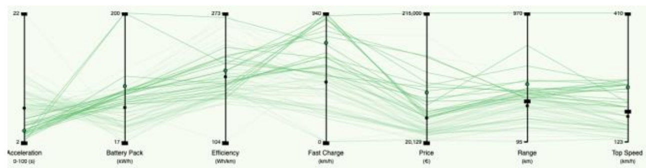

Figure 24. Parallel Coordinates – Top Speed && Range > Average

After it, we observe bar chart to verify which brands have more models with highest top speed and longest range.

Figure 25. Bar Chart – Brands with models with (Top Speed && Range > Average) filter

Now, we select one bar of bar chart, in this case, the Tesla bar because it is the one most representative. We also call mouse over event to show how many models which brand have in these conditions.

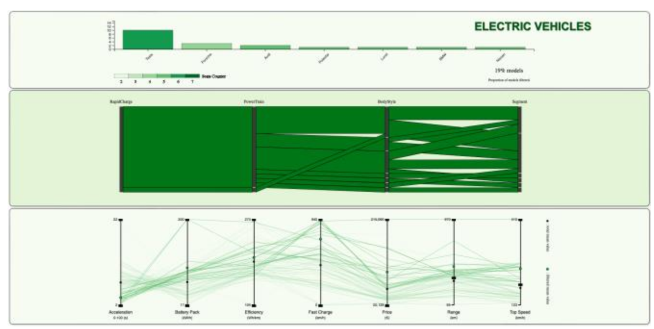

Figure 26. Visualization after all steps

Upon completing this analysis, it's evident that Tesla stands out as a brand with exceptionally high-top speed and extended range, as clearly depicted by the prominent orange lines in the Parallel Coordinates plot.

IMPLEMENTATION DETAILS

Overall Description

One of the requirements of this project was to use version 7 of d3.js, which is a popular *JavaScript* library that facilitates the production of dynamic and interactive data visualizations.

It is important to refer that all idioms were adapted from D3 ObservableHQ examples, except the alluvial diagram that has been created without following any example.

Throughout the creation of final implementation, we faced many challenges that forced us forget some initial ideas. Initially, we thought about the possibility of selecting two distinct bars that initiates a comparison between the attributes of each brand that are observable in other idioms. We need to change this implementation because it is harder to unselect the bars due to it is only possibly to changes the first one selected.

The implementation of a linking mechanism allows different visual components, such as Parallel Coordinates plots and Bar Charts, to work in harmony. By sharing a common data source, interactions in one component can trigger changes in others.

For instance, in a Parallel Coordinates plot, users can adjust filter settings, and this leads to dynamic updates in the Bar Chart. This synchronization of data and visuals is achieved through event listeners (drag and click) and call update functions. Smooth transitions and data filtering enhance the user experience.

Additionally, the user can switch the axes in the Parallel Coordinates plot. By doing it, the event drags, and drop is started, and the mechanism ensures that this change is reflected in other components.

Furthermore, when a nominal attribute (categorical) is selected by clicking on it in the alluvial diagram, we highlight the corresponding lines in the parallel coordinates and bar charts. This allows users to explore how a specific nominal attribute relates to brands and how it affects the measures represented in the parallel coordinates graph. It also shows which brands have the same selected nominal attribute.

These linking idioms provide a rich interactive experience, enabling users to explore and understand the relationships between brands, attributes, and measures intuitively and effectively in their visualization sketch.

CONCLUSION & FUTURE WORK

In the end, our project achieved the result we expected, and keeps like the initial version of the solution we had in mind.

We consider that the project was successful because we think we allow all the users the experience to easily answer the example questions we presented in Introduction.

The problems that appear during the different stages of project implementation had been a good opportunity for us to understand that three different persons had different perspectives but, after combined different ideas, the result visualization gets more attractive.

So, we can conclude that this was a useful experience, where we could learn, by empirical experience, how important is to draw a visualization that represents usability and, at the same time, utility.

FUTURE WORK

In our opinion this work could be improved by replacing the alluvial diagram with a visualization that would give more interaction for the user, and, in addition to this, promote the user to understand the correct portions of an attribute that is correlated to another attributes. As we can see in Body Style axis of alluvial diagram, if the axis has too many items it gets harder to understand the connections between them due to intersections.

REFERENCES

1. Kaggle's Dataset
https://www.kaggle.com/datasets/geoffnel/evs-one-electric-vehicle-dataset
https://www.kaggle.com/datasets/divyanshugupta95/cars-dataset-with-battery-pack-capacity
https://www.kaggle.com/code/divyanshugupta95/electric-vehicle-analysis
2. Parallel Coordinates Plot D3 Example - ObservableHQ
3. Bar Chart Plot D3 Example – ObservableHQ
4. Kaggle's Analysis Dataset
https://www.kaggle.com/code/divyanshugupta95/electric-vehicle-analysis/notebook#Data-Visualization
5. Fuel types of new cars – ACEA (European Automobile Manufacturer's Association)
https://www.acea.auto/fuel-pc/fuel-types-of-new-cars-battery-electric-12-1-hybrid-22-6-and-petrol-36-4-market-share-full-year-2022/
6. AFV and HEV Model offerings by manufacturer – US Department of Energy
https://afdc.energy.gov/data/10304