Visual Analytics Report - Group 12

Vasco Cruz and Verena Grall

Graz University of Technology

Abstract. The final aim of this work is to provide an interactive dashboard to support strategic decision-making for a start-up focused on the transportation of lithium-ion batteries for electric vehicles (EVs). The understanding of battery demand, manufacturer location and their capacity is getting more important due to the increasing EV market. This project uses the Five Design-Sheet Methodology to design and develop an interactive dashboard using Tableau, supported by data preprocessing in Python. Multiple data sources, including for example Statista and the International Energy Agency, are combined in the dashboard to visualize global and regional trends in battery demands and manufacturing capacities. Features, which can be used to explore, forecast and identify opportunities in the EV-market, are an interactive map and dynamic plots. The data acquisition, preprocessing, and the dashboard development are discussed in this paper. Advantages as well as disadvantages, difficulties, like data gaps, and further improvements are stated. The resulting tool provides a user-friendly interface for the founders to make informed strategic decisions, with potential for future enhancements in data integration and feature interactivity.

Keywords: Data Cleaning · Data Imputation · Visual Analytics · Tableau · Python · Electric Vehicles (EVs) · Battery Capacity and Demand

1 Introduction and Motivation

Due to the growing market of Electrical Vehicles (EVs) also the need of lithium-ion batteries and their transportation is increasing. The founders of a start-up, which is addressing exactly this topic by developing and manufacturing battery cell transportation boxes, has to make strategic business decisions based on demands, the locations of manufacturer and their capacities. Therefore, different data about demands, manufacturer and capacities should be visualized in an interactive map, with features like zooming and selecting variables (years, countries, etc.). It should be possible for the founders to identify opportunities by means of a world map and different plots.

1.1 Process

The used design process is the Five Design-Sheet Methodology (FdS) which is shown in Figure 1. The process starts with the supplier and customer talking to each other and defining and considering the task. In the second step an idea

gets developed on a piece of paper. After that, three alternative designs are sketched. The costumer is met again in the fourth step, where the different sketches are discussed and the best one is selected. The last step is to implement the prototype.[?]

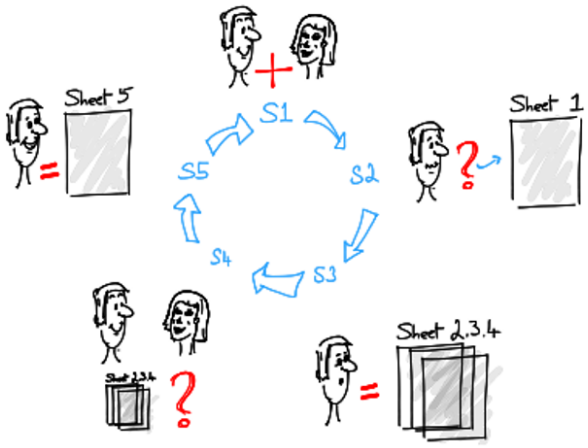

Fig. 1. Five Design-Sheet Methodology

This process was chosen because of the comprehensive planning, where different design ideas are explored and the best one can be taken, the iterative development, which allows adjustment based on feedback of the customer and even this user-centric focus.

1.2 Development Environment

As development environment two different levels are needed: Backend and Frontend.

Backend In the backend the exploratory data analysis and preprocessing is handled. The language used for this project is Python and the Softwares are JupyterNotebook and VisualStudioCode. The preprocessing is explained in the subsection 2.1.

Frontend The software used to visualize the dashboard in this project is Tableau. This is outlined in the section 3.

1.3 User - Task - Data

The intended **users** for the dashboard are the founders of the previous named start-up, which are the customers of this project. Furthermore, people, who are interested in EV-demand and capacity and the trend, are potential users of this visual analytics project. Skills and previous knowledge are basic knowledge about EVs and basic dashboard skills. When wanting to adjust the dashboard or the data, some basic knowledge about Tableau and Python would be helpful.

Tasks, addressed with this project are:

- exploring differences
- comparing countries
- predicting future demand / capacity
- identifying potential customers
- understanding production landscape
- analyzing trends

Data is treated in the next section.

2 Data Acquisition and Preprocessing - Python

2.1 Data Acquisition

The data used in this project comes from different sources. One dataset is from the International Energy Agency (IEA) about the historical and predicted demand of different types of electrical vehicles (EV). [3][4] This dataset consists of more than 9500 rows. A snippet of this dataset can be seen in 2. Further sources were Wiley[2] and Statista[5][6][7][8][9][10][11]. Different datasets about battery demands, manufacturer and their capacities were selected and merged. The third data source were the founders of the start-up themselves. They provided a data table about manufacturer in Europe and their capacities. Additionally, the population dataset which was obtained from the World Bank website[1] was used to construct a larger dataset with the included population values, needed to calculate the values per 10,000 inhabitants.

All in all qualitative as well as quantitative data was used for this project. The datasets were mainly stored in tables and therefore relatively well structured. Still, there had to be done some evaluation and preprocessing like checking for differences, merging, processing descriptive statistics, data cleaning, adding additional columns and predicting values.

region	category	parameter	mode	powertrain	year	unit	value
Australia	Historical	EV stock	Cars	BEV	2011	Vehicles	49
Australia	Historical	EV sales sh	Cars	EV	2011	percent	0.0065
Australia	Historical	EV stock sl	Cars	EV	2011	percent	0.00046
Australia	Historical	EV sales	Cars	BEV	2011	Vehicles	49
Australia	Historical	EV sales	Cars	BEV	2012	Vehicles	170
Australia	Historical	EV stock sl	Cars	EV	2012	percent	0.0028
Australia	Historical	EV sales sh	Cars	EV	2012	percent	0.03
Australia	Historical	EV stock	Cars	BEV	2012	Vehicles	220

Fig. 2. A subsection of what the main dataset looks like

Data Pre-Processing Data Processing is always necessary and can never be enough [2]. For this project the data pre-processing was conducted in Python JupyterNotebook and in VisualStudioCode, Since pre-processing largely determine the quality of the data and thus ultimately the dashboard, a considerable amount of effort in this project was dedicated to construct a dataset that included all the information necessary. The usual pre-processing steps were applied to all the datasets in the project, such as checking for duplicates, checking for null values and invalid data according to the domain knowledge possessed (for example negative age). Specially the merged dataset containing the prediction of capacity, was very flawed and has several missing data points. Considering this, the values were imputed according to the model that was the most accurate for each year. To discover which was the most accurate model, three different ones were implemented and their accuracy measured.

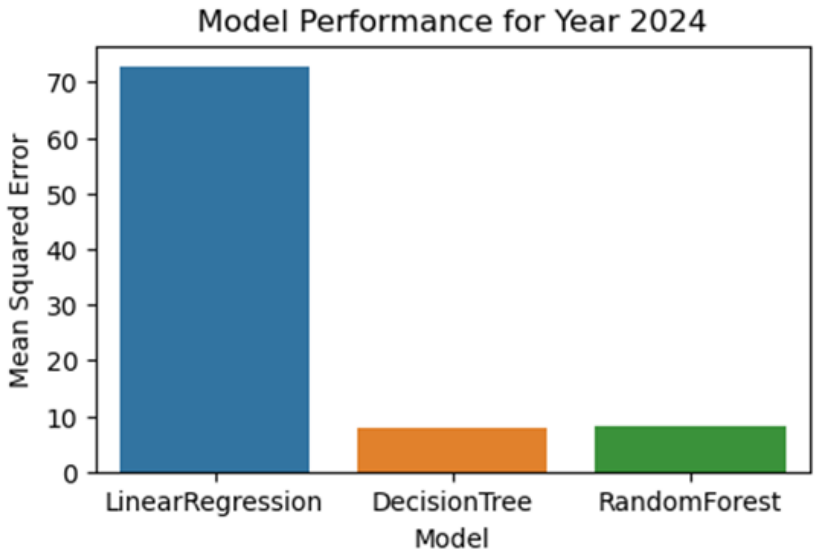

Fig. 3. 3 Models performance Evaluated. Linear Regression was the winner

As it can be seen in the example picture above, for predicting the values of the year 2024, Decision Tree Model was the winning model, due to the smallest mean squared error (MSE). To further refine and to be sure that the model selected was the best to impute the non-defined values, another plot was constructed with the predicted values and the actual values to get a visual idea of how exactly accurate the model is. The plot shows a quite accurate prediction.

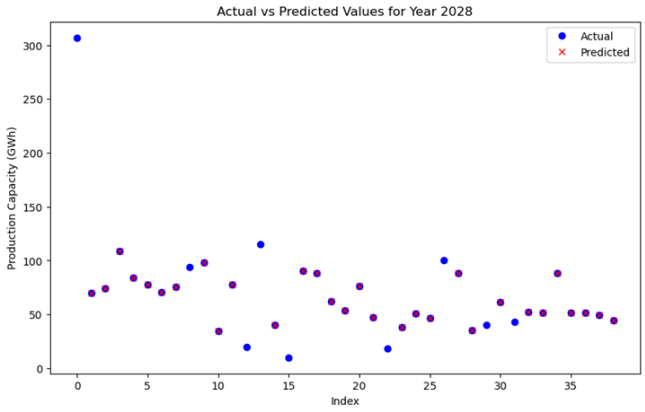

Fig. 4. Linear Regression predictions

Pre-processing another dataset included removing unnecessary columns for our analysis such as powertrain and mode, as well as joining the dataset with the world bank dataset for the population values between 2011-2022. As mentioned, this allowed the creation of a new dataset that was composed of a new column with the valuePer10000 inhabithants to be displayed in the final dashboard. A per capita value was also calculated, but it was extremely small for every country, thus not allowing for meaningful interesting comparisons between countries.

3 Data Visualization - Tableau

3.1 Tableau Overview

The data Visualization part was fully done using Tableau. Tableau is a software that allows for the construction of very interesting, clean and standardised dashboards. When importing data into Tableau, it is possible to preview it and even perform some useful operations. One of those, is to pivot the data, which means turning columns of data into rows. This was done in one dataset instead of turning over again to pre-processing it with Python

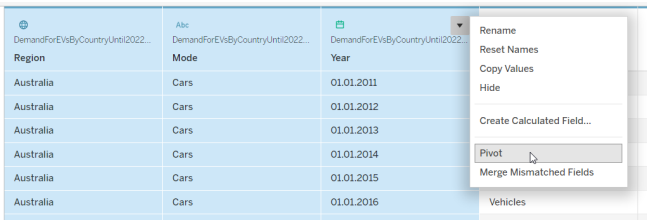

Fig. 5. Previewing Data and performing a data transformation technique

After this operation was done, it was time to start building the dashboard.

3.2 Tableau Dashboard building

Tableau has two different pieces that are stitched together in order to create a dashboard. Those are:

- "Sheets" where each piece of a dashboard is developed individually. This allows for better control, and fine tuning and bug hunting.
- "Dashboards" is where sheets are dragged into to create our final aesthetically pleasing dashboard.

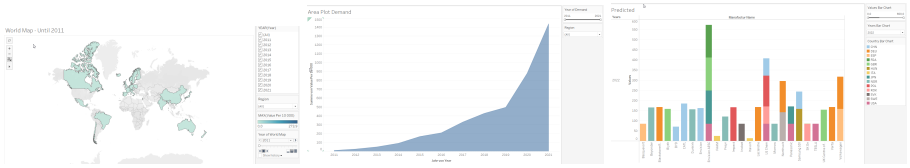

Fig. 6. Individual tableau sheets there will be combined into a final dashboard

In the "sheet" tab, there are several options of plots to chosen from depending on the available data, area plots, line plots, pie charts... and each one needs to be then constructed by the user by draging the data into the correct place. In the "dasboards" tab, it is possible to defined how each sheet interacts with eachother. This is extremely important to give the dashboards some interactivity. It is possible to define a behaviour when a country is selected in sheet1, then sheet2 is also updated with the new country.

The first sheet represents the world map and it will be the focus of our dashboard. It displays the geographical data and allows for the dashboard user to select a year and view the view the EV capacity for the specified year. The country colors are within a range, thus the darker the country is, the higher the value per 10000 inhabitants is.

The second sheet constructs a area line chart based on the group of countries selected by sheet1 in the world map. It gives the user an idea of how much the demand has been increasing over the years for the specified countries.

The third and final sheet, displays a stacked bar chart where each part of the column represents a country, while the x-axis has the manufacturer names and the y-axis the predicted capacity for the specified year. If the user selected a country that has manufacturers, the chart will also update to only display the projected capacity in that country by manufacturer.

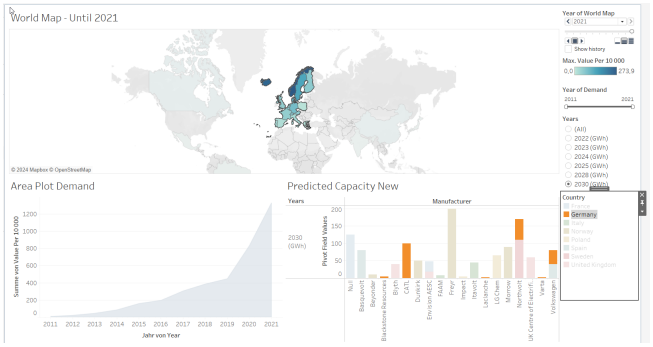

Fig. 7. Final Dashboard view with some options selected

The dashboard above has several chosen options and filters. First of all in the world map, it cam be clearly seen that European countries are highlighted. This will prove useful for the other two visualizations below the world map. On the bottom left, we have the area line chart, which displays the cumulative demand of those countries combined for each year. It is possible to see a slider filter called "Year of Demand" allowing for the visualization of the year only until 2020, or 2019, etc...Finally, we have on the bottom right, the stacked area chart which displays the different manufacturers in the group of selected countries. Furthermore, it was chosen to highlight only those who are present in Germany, thus the orange is highlighted. We can see an projection, for example for CATL, will have a capacity of around 100 GWh in the year 2030.

Its ease of use, and the fact that it can be worked with online in a browser with just an account were the main reasons why Tableau was chosen to visualize our data.

4 Discussion

Some advantages are, that the developed tool provides a holistic view of battery demand and manufacturing capacities in different countries and regions. Furthermore, the interactive design of the dashboard is intuitive and user-friendly.

Also the possibility to adjust this tool and extend it with additional data or features easily, makes it quite flexible.

Though, a major disadvantage is the lack of available data, which could affect the accuracy and completeness of the analysis.

This data limitations and gaps let the authors face major difficulties, which led to implementing the regression models for predicting missing values. Also the tool-selection, to select the right software for the frontend, was demanding.

On the other hand, the cooperation within the team and with the customers went well. Furthermore, this project was very informative and educational, which could be seen in the steep learning courve when implementing the prototype.

Further improvements are to expand the data sources, improve the dashboard interactivity and reduce features.

References

1. World Bank. (n.d.). World Bank Open Data: Population, total. Retrieved from https://data.worldbank.org/indicator/SP.POP.TOTL
2. Dasu, T., & Johnson, T. (2003). *Exploratory Data Mining and Data Cleaning.* John Wiley & Sons.
3. International Energy Agency (IEA). (2023). *Global EV Outlook 2023. Catching up with climate ambitions.*
4. IEA (2024), Global EV Outlook 2024, IEA, Paris https://www.iea.org/reports/global-ev-outlook-2024, Licence: CC BY 4.0
5. Statista. (June 28, 2021). Estimated market demand for lithium-ion batteries used in electric vehicles in 2019 with a forecast for 2020 through 2030 (in gigawatt hours) [Graph]. In Statista. Retrieved June 27, 2024, from https://www.statista.com/statistics/309570/lithium-ion-battery-market-in-electric-vehicles/
6. Statista. (June 23, 2021). Projected size of the global lithium-ion battery market from 2020 to 2026 (in billion U.S. dollars) [Graph]. In Statista. Retrieved June 27, 2024, from https://www.statista.com/statistics/1011187/projected-global-lithium-ion-battery-market-size/
7. Statista. (June 24, 2021). Projected global battery demand from 2020 to 2030, by application (in gigawatt hours) [Graph]. In Statista. Retrieved June 27, 2024, from https://www.statista.com/statistics/1103218/global-battery-demand-forecast/
8. Visual Capitalist. (August 18, 2019). Lithium ion battery production capacity by 2028, by company (in gigawatt hours) [Graph]. In Statista. Retrieved June 27, 2024, from https://www.statista.com/statistics/1103401/predicted-lithium-ion-battery-capacity-by-company/
9. whitehouse.gov. (June 14, 2021). Estimated production capacity of tier 1 to 3 lithium-ion battery factories worldwide in 2018 with a forecast for 2023 and 2028 (in gigawatt hours) [Graph]. In Statista. Retrieved June 27, 2024, from https://www.statista.com/statistics/1247625/global-production-capacity-of-lithium-ion-battery-factories/

10. McKinsey & Company. (January 16, 2023). Lithium-ion battery cell demand worldwide in 2022, with a forecast to 2030 (in gigawatt-hours) [Graph]. In Statista. Retrieved June 27, 2024, from https://www.statista.com/statistics/1419502/global-lithium-ion-battery-demand-forecast/
11. Hettesheimer, Tim; Neef, Christoph; Rosellón Inclán, Inés; Link, Steffen; Schmaltz, Thomas; Schuckert, Felix et al. (2023): Lithium-Ion Battery Roadmap - Industrialization Perspectives toward 2030. Unter Mitarbeit von Fraunhofer-Gesellschaft.
12. Roberts, Jonathan C.; Headleand, Chris; Ritsos, Panagiotis D. (2016): Sketching Designs Using the Five Design-Sheet Methodology. In: IEEE transactions on visualization and computer graphics 22 (1), S. 419–428. DOI: 10.1109/TVCG.2015.2467271.