[G0R72A] Data Visualisation Project Report

- *IMPORTANT: Report naming convention. Please name your file like this: group number + "_report_" + version (i.e. "design" or "implementation"). For example: "group15_report_design.pdf". So not "15-report-1.pdf" or similar.*

Part 1. Metadata

- Version: design (25/3/2024)
- Students: Jørgen Aleksander Arnesen, r-0978820, KU Leuven, Daria Małgorzata Plewa, r0976669, KU Leuven, Beste Karataş, r0973831, KU Leuven
- Group number: group_21
- Dataset: SunCharge

Part 2. Project description

1)

The provided dataset includes data on the supply chain of a battery manufacturer, SunCharge. Sun Charge has asked us to create a monitoring system to detect issues in the supply chain and identify improvements.

2)

The dataset inherits several key features. To understand the supply chain, it is essential to know the vendors of raw materials and which materials they inject into the supply chain. The supply of raw materials is then distributed in 3 different production plants. Key features in the production include production time, standard cost, and time values such as processing time and total inbound lead time. It is essential to know these features well to improve the supply chain. After production, the products are sent to different distribution centers. The inventory is kept here and sales to customers go from the distribution centers. Key features include gross inventory quantity, order quantities, and forecast of future sales.

3)

In this analysis, we will focus on two main questions. What makes purchase orders/sales orders late? Is there a correlation between late deliveries with material type, order quantity, or the plant they were requested to/from? Want to plot them all in the same plot? To achieve this, we will focus on both aspects of the supply chain:

One plot tracing the process from distribution center to customer, focusing on the sales data, columns referencing the expected and actual delivery date.

How do actual sales compared to predicted sales evolve? Although the question is easier to implement, we believe it still holds valuable insights into the company's profitability. By connecting the inventory and forecast data sets, the company will have an accurate picture of its expected orders and production.

One plot tracing from vendors to the production plants, focusing on the purchase data, additionally providing detailed information on which part of the process the delay is occurring giving insight into how it can be fixed.

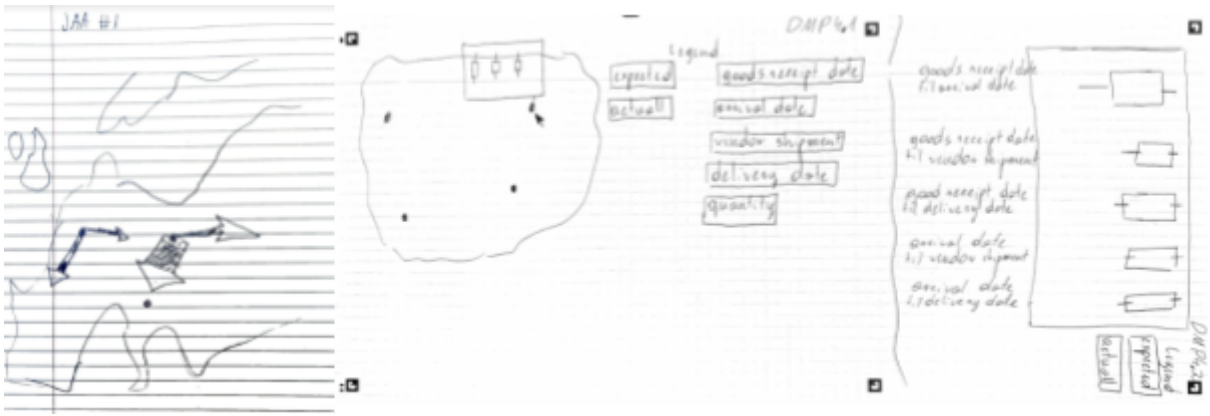
Part 3. Visual design

1-2)

The visual designs were created in collaborative sessions where early explorative sketches were refined into the final designs we wanted to implement. We started out sketching 5-10 small sketches for ourselves before comparing and grouping them. The early sketches are shown in Figure 1 in the appendix.

3-4)

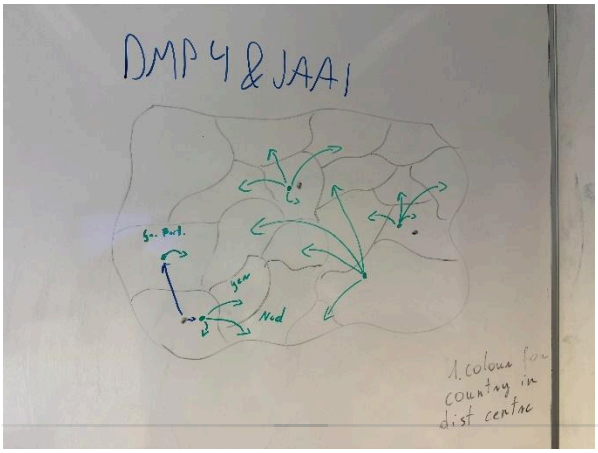
After initial designs were created and grouped, we continued discussing the designs and how to further develop them. The whiteboard sketches illustrate our discussions and how the final sketches emerged. JAA#1 shows an idea where a map is used to show distributions to different

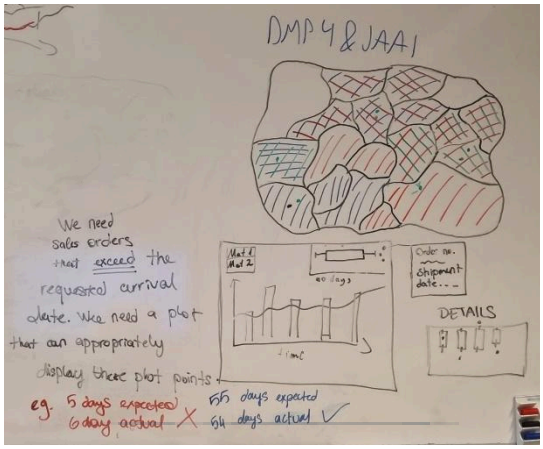


countries from the distribution centers. It also illustrates quantities of various products as the thickness of the arrow and their compositions. DMP4 shows the expected and actual dates for states of transport of goods to the client with visualization of boxplots that represent the amount of time needed for step – thanks to boxplots we can see the exceeding time values that could represent issues in transport.

The combination of these two sketches is shown in DMP4 & JAA1, which was further developed into a final sketch we would want to implement. The visualization shows a map of the different plants and distribution centers. The countries connected to the different distribution centers have the same color. Clicking on one of the distribution centers intensifies its color and opens a graph showing total transport time by quantity, and a boxplot showing the distribution of the time values for a of transport.

From here outliers and exceeding values are easily identified, which can further be explored by clicking on them. This gives all their quantified information in a journal and different boxplots showing the different transportation stages. It is then easy to see where the transportation problems are located.





The final implementation scrapped the boxplots, due to them not giving relevant information about the outliers. We therefore implemented a scatterplot instead, showing time differences as a percentage fraction that represents the differences between the actual transportation time and

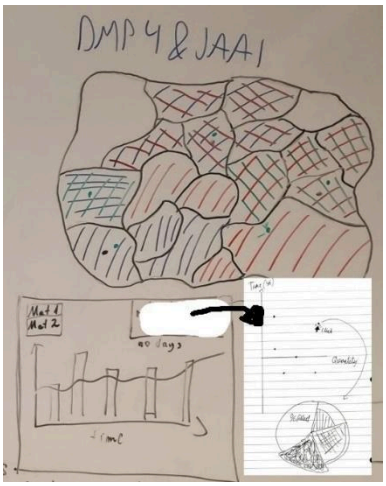
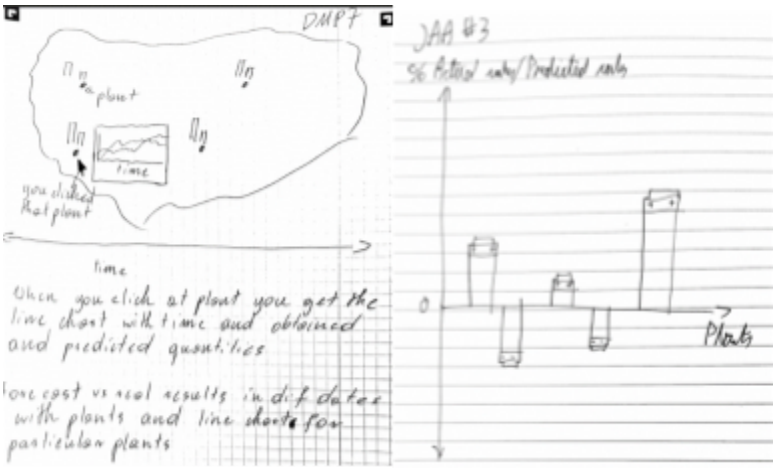
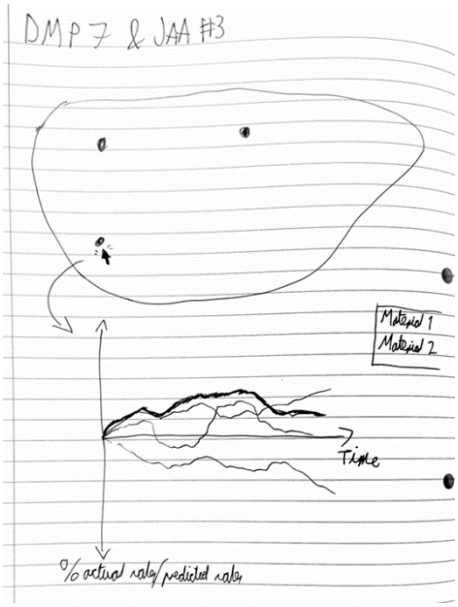


Figure 5: Final Implementation

the predicted transportation time, based on different shipment quantities. Clicking on one of the points shows a pie chart of the different specific transportation stages. The different pieces of the pie chart are then filled to show the percentage fraction of actual time compared to predicted time. In total, this will give an overview of both general and specific trends of transportation in the supply chain.

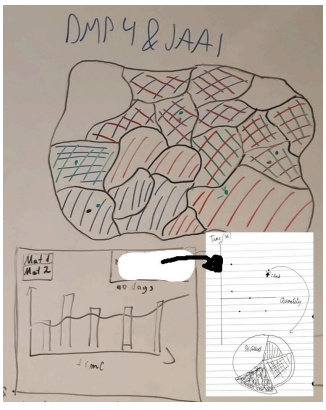
One design was developed from the two initial sketches DMP7 and JAA#3. DMP7 wanted to show how the predicted and obtained quantities developed in time at different plants by using an interactive map. JAA#3 wanted to show how the actual sales compared to predicted sales, with an output in percentage. We then combined these two plots to show how the actual sales compared to the predicted sales were for varied materials in different contribution centers. This is shown in a combined DMP7 & JAA#3 sketch.





The final DMP7 & JAA#3 sketch is one sketch we would implement if we were able to. We think it is important to show how predicted sales compare to the actual sales to discover potential problems in the forecasts. We also thought that a straightforward visualization would fit the problem best since it would be used as a graph that evolves whenever forecasts are updated.

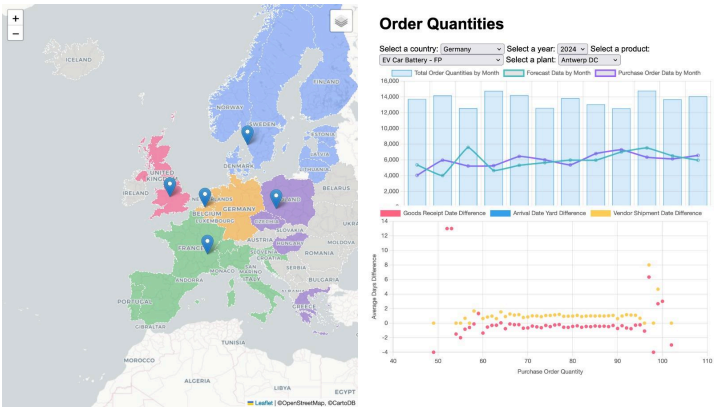
Part 4. Implementation



Visualization 1:

The intended design for our first visualization implemented a map of Europe as a backdrop and main hub. The map is divided into different

colors, where each color represents a distribution center. This makes it intuitive to grasp what distribution centers distribute to each country. We also wanted to add nodes on the map for the production plants and the distribution centers. Clicking on one of the distribution centers would open up a plot showing total transportation time by quantity and a scatterplot. The scatterplot would show transportation time as a percentage fraction that represents the differences between the actual transportation time and the predicted transportation time, based on different shipment quantities. The points in the scatterplot were intended to be piechart glyphs, and clicking on one of these would enlarge it. The pie chart would show the contribution in percentage of the different transportation steps.

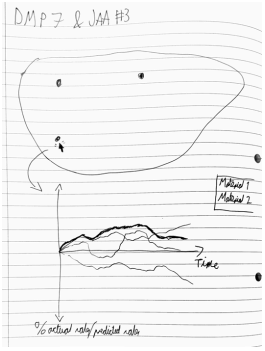


The actual implementation of our designs were done on a single main visualization hub. The first visualization is shown at the top right part of our main page. The left part shows the map, which is interactive and used for both of our visualizations. The map of Europe features different coloring based on connection to distribution centers. This was chosen so that the user gets an intuitive grasp of the distribution chain. Furthermore, the map features a dropdown menu to show production plants and distribution centers on the map Our first visualization on the top right shows a combined line chart and bar chart of order quantities. Toggle menus can be used in order to choose the import country, year, product and exporting plant. Clicking on a country or a plant on the map will also update the toggle menus automatically. The chart is also interactive, where you can hover over specific points to read out quantities. You can also toggle out or in sale order quantities, forecast by month and purchase orders by month, to make the visualization simpler if you do not need all the information. Scrolling down on the chart reveals the pie chart, which shows fractions of products sold. The scatterplot intended in our design is shown at the bottom right of our main page. It shows the goods receipt date, arrival date yard difference, and vendor shipment date difference as intended. The 3 different stages are toggleable so that the user can see the different ratios. Hovering over the points will also give the different shipment times.

Our actual design differs quite a bit from the intended design. This was done as a deliberate choice. Visualizations more similar to our intended design were created, but reading out the data was difficult. It was therefore decided to connect the data to a databank and create larger graphs that could be toggleable and update when you interact with the map. The intended functionality is still the same, but the overall layout of the visualization makes the graphs bigger. We were not able to create the piechart glyphs in the scatterplot as intended. We could not get it to work on our dataset, and when experimenting with some basic plotting the glyphs seemed too small to get any information from. We therefore scrapped the pie chart for shipment duration and made the different shipments toggleable instead. Hovering over the points in the scatterplot does almost the same as a pie chart would show, but the intuitiveness of the ratios might be lost.

If we had access to a skilled programmer we would like to implement the piechart glyphs as we intended in our designs. We were not able to implement good enough glyphs and therefore decided not to utilize them at all. We would also like to optimize our database. Right now it is not optimized for all our data. Having a skilled programmer to organize it and format the data nicely would have made our visualizations better.

Visualization 2:



The intended design for our second visualization is also based on a map of Europe as the backdrop. The map would show the different production plants. Clicking on these would open up a line chart showing the percentage of actual sales compared to predicted sales, for different materials. This is a simpler visualization, but important in order to discover problems with the prediction of sales.

It was decided to include the second visualization in the first one, as they could use the same skeleton. Interactivity on the page would toggle between visualization 1 and visualization 2. Visualization 2 therefore also implements the same map with different distribution areas in different colors. However, the graphs are togglable and can therefore display the graphs we wanted to include in our second visualization. As explained for the first visualization, the line chart intended for our second visualization is toggleable in the combined chart at the top right.

Since it was decided to implement both visualizations in one, the interactivity with the map differs from our intended design. We could create the intended design but decided that using a toggle feature seemed like a better idea. That gives the user more control over what is showing. We thought clicking on the map every time one wanted a different graph would be annoying to use, and therefore implemented the toggleable feature in this visualization as well. This makes changing the graphs intuitive Other than that the linechart is kept simple, as intended, to get a nice overview of sale predictions.

As for the first visualization having a skilled programmer to optimize our databank would be extremely helpful. If the database is working perfectly, our visualizations would be able to take in all the different data points from the CSV files. The line chart was supposed to be a simple visualization to check for sales predictions.

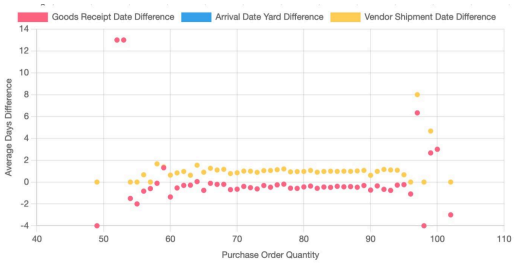
Here is the link to the youtube video, showcasing our visualizations:
https://www.youtube.com/watch?v=LTvtEN1qVi0
Here the link for GitHub repository with all codes
https://github.com/voira/data_visualisation

Part 5. Findings

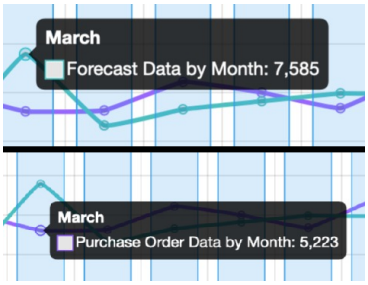
There were a lot of interesting findings in our visualization of the suncharge dataset. The map shows some outliers when distance to distribution centers are taken into account. First of all there is Greece. Greece almost exists on their own, far away from all distribution centers. Looking at shipment data incoming for Greece we see that…

Second of all, there are the baltic states, especially Lithuania. These are countries that are connected to the Göteborg distribution center, despite being closer to the Polish distribution center. We can see this in the shipment data for the Baltics…

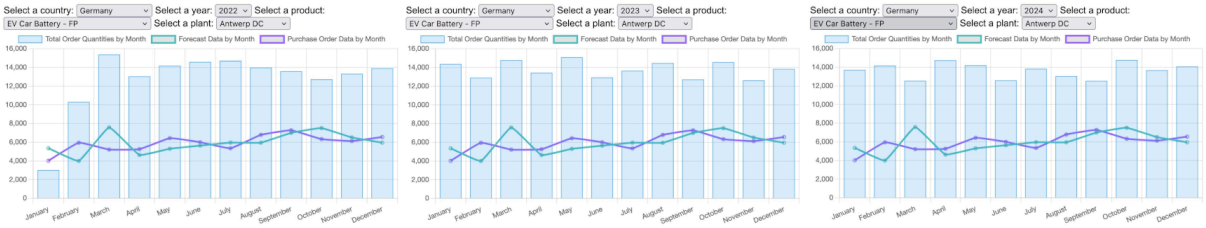
Generally looking at shipment dates we see that most shipments are on time or earlier/delayed by equal or less than 2 days. We do however see a lot of outliers when we go up in order quantity. There are shipments that come 4 days too early and more shipments that are up to 8 days late. This is a problem that should be looked at by the company. Goods arriving later, or even earlier, create problems for customers which may lead to negative responses to the company. It is quite logical that we see shipment problems with large quantities. It is however strange that we also see this with smaller quantities. Here we even see some shipments that were 11 days late. This is a huge problem for the company. Smaller quantities should be easier to handle and predict shipment times. That shipments are delayed this much shows a huge vulnerability in the supply chain that should be addressed. The company should look into their shipment methods and try to improve their shipment predictions to accurately predict shipments for smaller and larger quantities.



The predicted sales versus actual sales for the different products seem to show that the predictions are satisfactory. The largest gap in the predictions is in March. Here we see that the forecasted sales are 45% higher than the actual sales. This is certainly a concerning number, but the March predictions had the largest gap by far. The bad prediction might also be a result of February. We see that the actual sales are much higher than the forecasted sales in February, which might lead the firm to predict higher sales in March. Other than this outlier, the predictions are quite good. SunCharge should look into predictions in the earliest months since there seem to be inaccuracies there.



The most bought product is EV Car Battery in every country. Germany is the greatest buyer of all goods throughout all years. Since 2022 the sales obtained rather constant level in all countries. Always, the January 2022 obtains the worst results from the total order quantities for all products in all countries.



Part 7. Individual contributions

Daria: Ideas, sketching, backend database work, fetching data into code, report writing

Beste: Ideas, sketching, setting up initial map, front end work; js html css and formatting, report writing

Jørgen: Ideas, sketching, front end work, report writing

Appendix A. Sketches overview