

[G0R72A] Data Visualisation Project Report

Part 1. Metadata

- Version: implementation (10/5/2024)
- Students:
 - Raisa Carmen - s0204278 - KU Leuven (Master of Statistics & Data Science, blended track)
 - Steven Huygens - r0624888 - KU Leuven (Computer Science)
 - Tamaya Van Crielinge - s0214219 - KU Leuven (Master of Statistics & Data Science, blended track)
- Group number: group_07
- Dataset: SunCharge

Part 2. Project description

2.1. Data description

The dataset provides information of SunCharge's supply chain, a manufacturer of car and home batteries, covering its distribution centers (DCs), production plants (PPs), customers, suppliers, and inventory. This allows us to analyze operational efficiency and key metrics like inventory levels, order processing times, and transportation times. With vendor, customer, and distribution center locations included, we can also optimize logistics, transportation routes, and market expansion strategies geographically.

2.2. Feature description

The following features represent the key aspects of the SunCharge dataset:

- **Geographical locations:** Addresses for all vendors, plants, and customers.
- **Product quantities of forecasted and sold products:** Sold and purchased product quantities from Jan 2022 to Dec 2024, forecasted sales from Jan 2022 to Dec 2025 per DC, including product type, customer details, order and shipment dates, and DC data.
- **The downstream supply chain and production line:** Information on raw material vendors and bill of materials for car/home batteries. SunCharge's 3 PPs produce these batteries, data details their shipment to DCs.
- **Inventory and stock quantities:** End-of-month inventory data from December 2022 to January 2024, covering on-shelf and in-transit quantities, as well as safety stock quantities.

Mapping vendor, plant, and customer addresses helps analyze regional demand patterns for forecasted and sold quantities, identifying operational issues and shipment opportunities. This data also aids in forecasting material and final product needs, optimizing inventory management.

2.3. Guiding questions

1. **What does the upstream supply chain look like and how can SunCharge improve?:** How diversified is the upstream supply chain? Do some vendors have a monopoly on raw materials? How fast can the raw materials be delivered?
2. **What does the material flow within SunCharge and all its plants (PP and DCs) look like and how can SunCharge improve?:** Are PP to DCs shipments timely? If not, where's the delay? Can DCs receive timely supplies during PP outage? Which DCs have high inventory? Where to adjust inventory? Which DCs go below safety stock? Where to open new PPs for demand?
3. **What does the downstream supply chain look like and how can SunCharge improve?:** Which DCs serve which customers? Can orders reach customers on time? Which customers face longer delivery times? What sales patterns emerge? Where to open new DCs for sales?

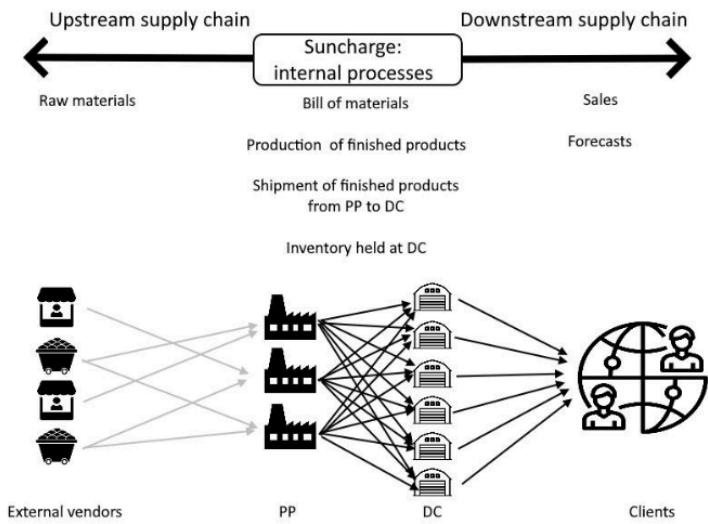


Figure 1: Diagram of the supply chain

Part 3. Visual design

3.1 Process description

Materials: A [Miro board](#) was created for collaborative sharing of individual sketches of the SunCharge dataset. See Appendix A for our Miro board overview. Additionally, a Google Drive was set up to share individual data exploration reports, meeting minutes and task lists.

Before meeting 1: Every member of the group individually completed the *Diverge phase*. Everyone made some designs before the first meeting took place.

Meeting 1: Afterwards, an initial meeting was convened for presenting individual visualizations. During this session, we initiated the *Emerge phase* by discussing the strengths in each other's visualizations and exploring combinations to address specific questions about optimizing efficiency in the SunCharge dataset. As a result, we concluded the meeting with 8 visuals for further development, either as combined or optimized visuals. Two team members divided the creation of new visuals, while another created the feature-by-feature matrix (see Miro) to ensure comprehensive coverage of dataset aspects. This feature-by-feature matrix, interestingly, also included some designs for features that are currently not available in the data but would help in answering the guiding questions described above (question 3 of part 2).

Meeting 2: A follow-up meeting was scheduled one week later to review the new visuals. During this discussion, we identified additional features to incorporate, resulting in the addition of one more visual and the optimisation of others, bringing the total to 9 combinatory visuals.

Meeting 3: The objective for the final meeting was to outline the questions each visual could address regarding the dataset. This resulted in a prioritized list of three visuals to answer our guiding research questions discussed previously.

3.2. Illustrative sketches

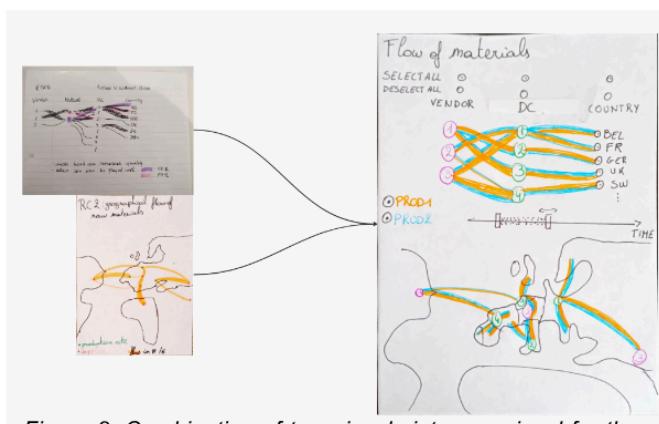


Figure 2: Combination of two visuals into one visual for the flow of materials

Figure 2 demonstrates the merging of two visuals from the diverge phase into one during the emerge phase. The top visual of the diverge phase shows the flow of the materials from vendors to DCs to the country of the customers. The wider a line, the higher the quantity of the materials. The bottom visual shows the same, but here the vendors and plants are placed on their geographical locations. These two visuals are then combined in a single visual which can be seen on the right in Figure 2. In this visual, orange lines represent the flow of car batteries, while blue lines represent home battery flows. The top and bottom parts are similar to the top and bottom visuals created in the diverge phase,

respectively. However, the bottom part now includes the countries of the customers as well. In the top half of the figure, the user can select which PP's, DC's and countries' flow they do (not) want to see and for which time period.

In Figure 3 two visuals of the diverge phase are again combined in a single visual. The visual at the top from the diverge phase shows the amount of sales per day. The color per day changes depending on how many sales are made. With this visual, certain trends in the sales can be detected, for example if there

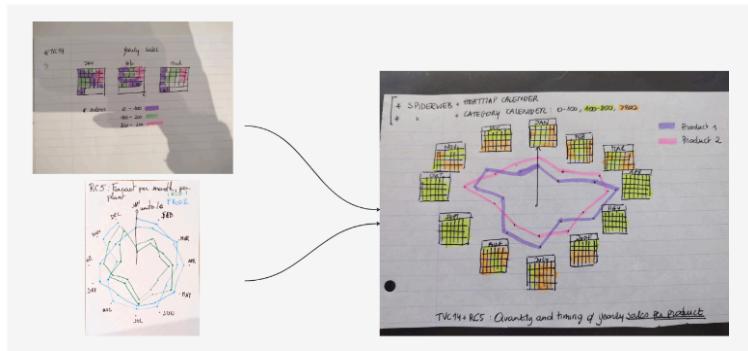


Figure 3: Combination of two visuals into one visual for the amount of sales

shown in purple and those of home batteries in pink. The top visual from the diverge phase is integrated into this final visual as a glyph, with the addition of the calendar for each month placed on the corresponding place in the spider web chart. Instead of using color to show the number of sales, lightness is chosen to display the amount of sales, since this is a better visual to show quantitative data.

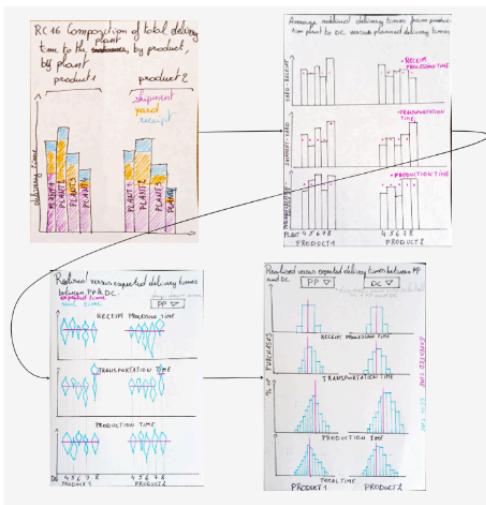


Figure 4: Different phases of a visual to display the delivery times

while the blue histograms show the actual times.

3.3. Choice of implemented sketches

There are three visuals we would like to have implemented which help answering the questions proposed in section 2.3. The first visual we would like to have implemented, is the visual shown in Figure 2. The encoding of this visual is explained in the corresponding part in section 3.2. For this visual, you would be able to select from which products, PPs, DCs and countries of customers you would like to see the data. There would also be a slider to select the time period for which you would like to see the data from. This visual can give insights into the internal processes as well as into the downstream supply chain which is related to our guiding questions one and three. For example, more specific questions that can be answered with the visual could be: Has the flow of materials (volume & route) changed over time? Which DCs get supplied by which vendors? Which vendors' products end up in which countries?

are more sales on mondays. The visual at the bottom from the diverge phase plots the forecasted number of sales per month in a spider web chart. In this visual, the forecasted number of sales of car batteries are shown in green and those of home batteries in blue. These two visuals are then combined in a single visual to show the number of sales of a certain plant. In this visual the number of sales per month are plotted in a spider web chart. The number of sales of car batteries are

Figure 4 compares the actual and expected time it takes to ship a product from the PP to the DC. The first figure is from the Diverge phase, the second and third from the emerge phase while the third is from the converge phase. From the first to the second figure, we expanded the graph since each PP can ship to each DC. Instead of colors for the different times, we use small multiples. We then chose violin charts (beeswarsms were considered) to show the variance in the real delivery times instead of bar charts for the average delivery times. Lastly, upon inspecting the data, we saw that the real and expected times are expressed in days instead of hours and minutes. With an expected receipt processing time of only two days, a violin plot will not look good since we likely only observe times of either 1, 2 or 3 days. Histograms seemed like a better fit. Total delivery time was also added and drop-down menus can be used to choose for which DC and PP the times need to be shown so as not to clutter the figure too much. The pink line represents the expected time

The second visual we wish to implement is the visual shown in Figure 3 at the right. The encoding of the visual is explained in the corresponding part in section 3.2. For optimized encoding, we are also considering a dropdown menu to include years 2022 to 2024 in this figure, along with the option to hover over the glyphs to enlarge them and display the actual quantities per day or week. This visual provides insights into the downstream supply chain which is related to our guiding question three, revealing temporal sales patterns throughout the year. For example, more specific questions that can be answered with the visual could be: Are there clear periods of high and low product demand yearly? Are there seasonal sales patterns, influenced by factors like holidays? When do customers prefer to purchase specific products each month or week?

The last visual is displayed in Figure 5. This visual shows the gross inventory of a plant over time. On the positive y-axis the inventory of car batteries are plotted, while on the 'negative' y-axis the inventory of home batteries are displayed. In blue the in transit inventory is shown and in purple the on shelf inventory. There is also a red line for each product to show the safety stock quantity threshold. There would be a drop down menu to select from which DC you would like to see the data displayed. This visual provides insights into internal processes, particularly the management of inventories by each plant, aligning with our guiding question number 2. For example, more specific questions that can be answered with the visual could be: What are the current product inventory levels compared to safety stock at various plants? What is the proportion of in-transit and on-shelf inventory per plant? Are there any temporal patterns in plant inventory levels?

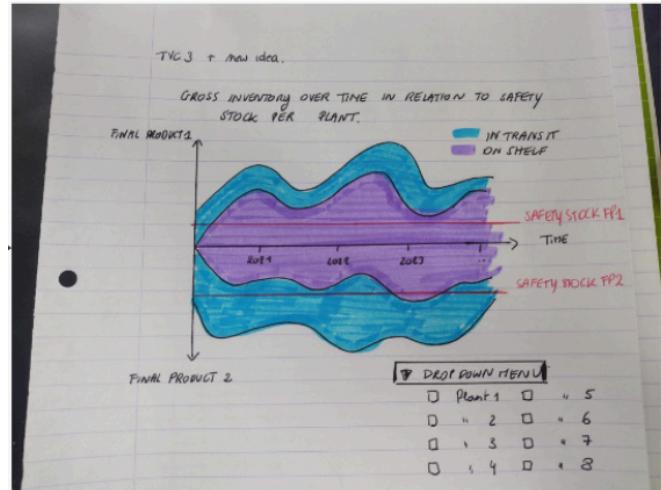


Figure 5: Visual to show the inventory of a plant over time

Part 4. Implementation

For the implementation, we used Svelte(kit) for the first visualization and Svelte(kit) and D3 for the second one. The landing page of the Suncharge dataset visualizations features a prominent header providing direct access to our two main visualizations: App 1 and App 2. Additionally, in the top right corner, you'll find a convenient link to our GitHub page, where you can access the code behind the visualizations. A video demonstrating the functionality of our visuals can be accessed with this link: <https://www.youtube.com/watch?v=Jvh7Eph-Ovw>. Furthermore, a link to our GitHub page to access the code behind these visualizations can be found here: https://github.com/raisac/data_visualisation.git.

In designing **App 1**, depicted in Figure 6, our goal was to provide insights into the downstream supply chain revealing temporal sales patterns throughout the year (guiding question n°3). For example, more specific questions intended to be answered with the visual were: Are there clear periods of high and low product demand yearly? Are there seasonal sales patterns, influenced by factors like holidays? In accordance with further details provided in section 3.2 of the project documentation, we proposed several interactive features, including a dropdown menu for selecting years, a hover functionality allowing users to enlarge glyphs, and a display of quantities per product. We're pleased to announce that all these elements have been successfully integrated into our radar plot with calendar-glyphs.

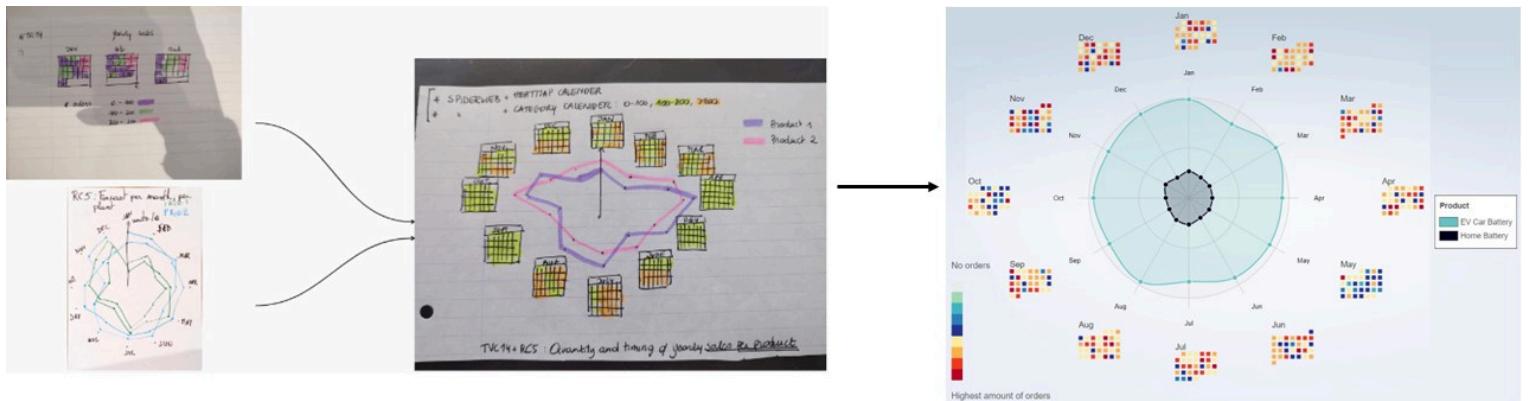


Figure 6: The design and implementation process of App 1

Furthermore, we have extended the original proposal by introducing the capability to distinguish between plants, along with an interactive legend that appears only when both products are displayed. This enhancement adds depth to the original design, offering users more comprehensive insights. This comprehensive visual empowers users with interactive tools to delve into sales trends across various dimensions, including product, week, month, year, and plant, effectively addressing the second guiding question concerning annual product demand and seasonal sales patterns. More specifically, the following visual encoding and interaction are employed in our visualization, as seen in Figure 7:

- The dropdown menu empowers users to select the product, year, and plant, enabling **interactive features that dynamically alter the glyphs and radar plot presentations** (*in green*).
- The heatmap calendar glyphs, seamlessly integrated within the radar plot, visually represent each month, offering **interactive functionality such as enlarging upon hovering** to provide detailed insights into sales trends. Additionally, a color legend accompanies the heatmap calendar for the amount of sales per day. (*in blue*)
- The radar plot not only presents sales data per month but also incorporates **interactive elements including hovering over a month to display sales volume** (*in blue*) and a legend that

distinguishes between product lines which will only be present when both products are displayed (in red), enhancing user engagement and data exploration.

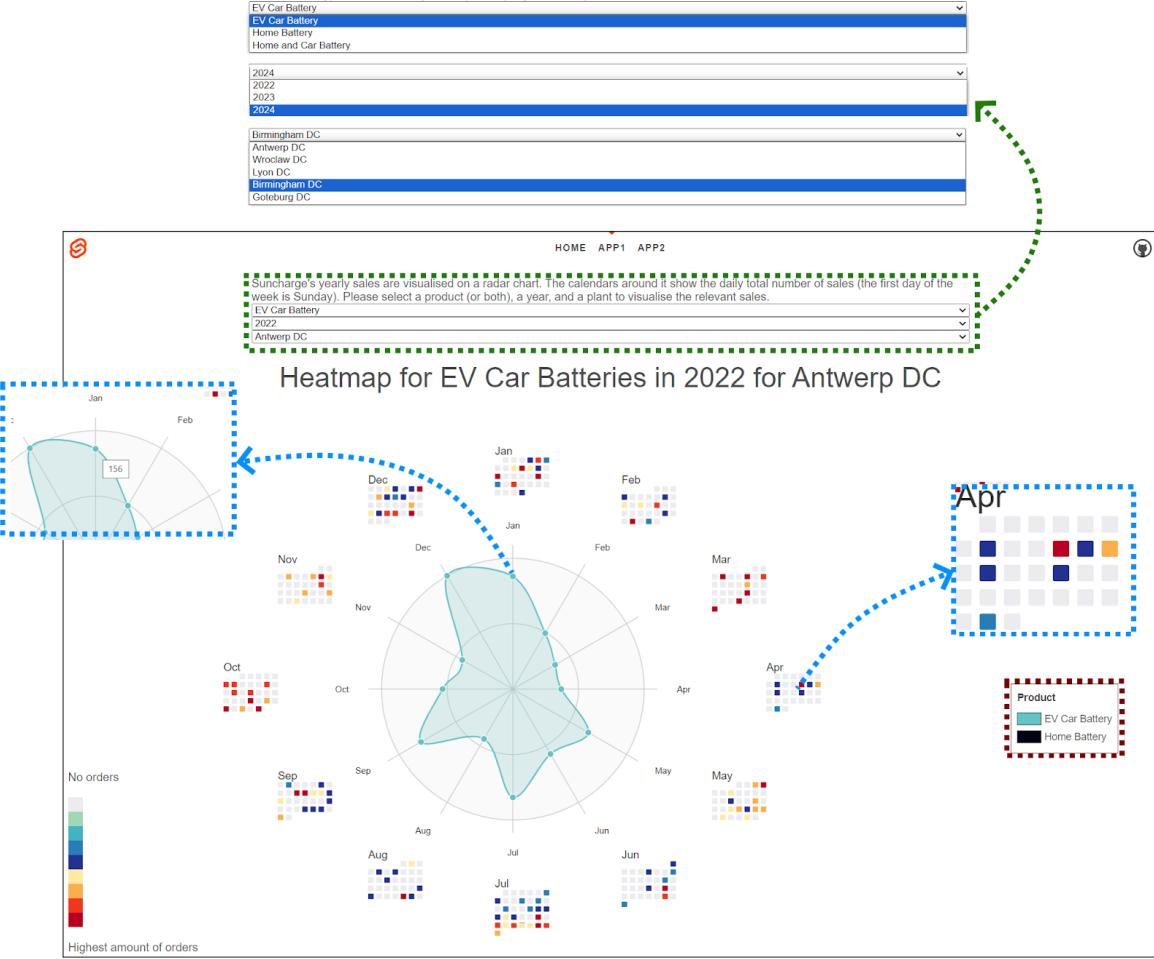


Figure 7: Dynamic effects in the first app. Drop-down menus to filter the data are indicated in green. The blue boxes show effects when you hover over the figure; enlarged calendars and the total number of sales in a month. The legend in red is only visible if both products are selected to appear in the figure.

In our design for **App 2**, as seen in Figure 8, we aimed to give insights into the internal processes as well as into the downstream supply, referring to guiding questions one and three. For example, more specific questions we intended to answer with the visual were: Has the flow of materials (volume & route) changed over time? Which DCs get supplied by which vendors? Which vendors' products end up in which countries? As outlined in section 3.2 of the project documentation, our intention was to incorporate several interactive features allowing users to visualize the flow of various products. We intended for the users to have the ability to select specific production plants (PPs), distribution centers (DCs), and countries for viewing, as well as customize the time period of interest with a slider. Unfortunately, we were not able to implement all intended features because there is no data concerning the external vendors.

Therefore, for the second application, we decided to focus on the geographical location of the customers, the quantities purchased by a customer and the delays for deliveries. As can be seen in Figure 9, There are three different sections for this application. The three different sections have been framed with different colors to make it more clear. There is a section to filter the data, framed by the green rectangle. The second section is to visualize the distribution centers and the cities of the customers on a map

geographically. This section is framed by the red rectangle. Lastly, there is a section to give more details on the data shown on the map. This section provides more details and is framed by the orange rectangle. At all times, the three sections correspond to the same filtered data.

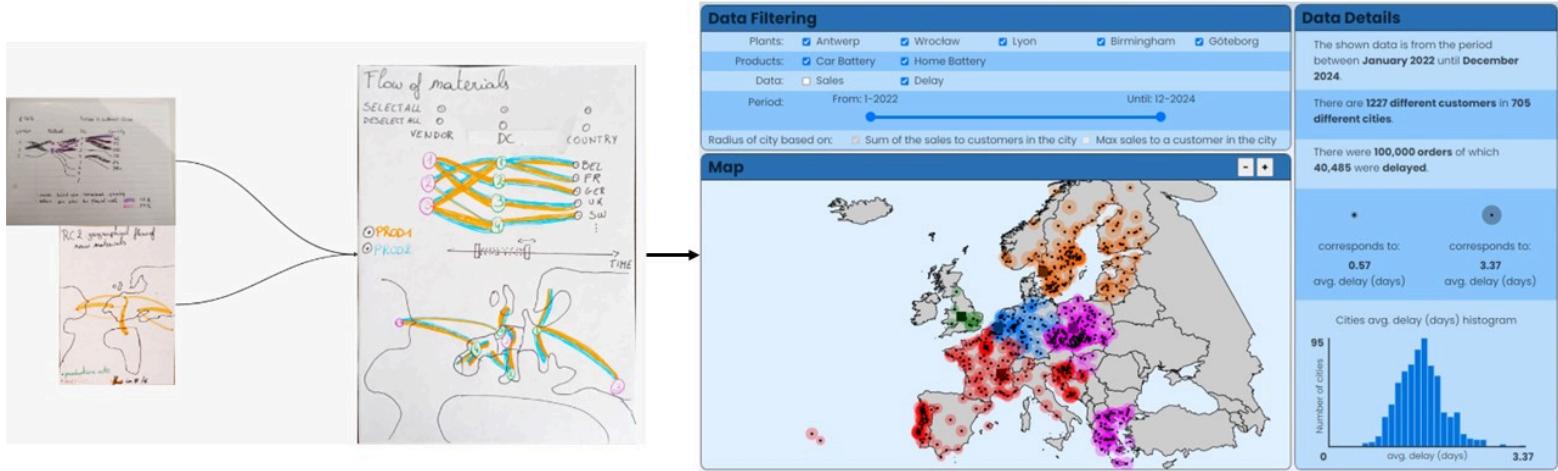


Figure 8: The design and implementation process of App 2.

The “**Data Filtering**” section can be used to apply certain filters for the data you want to visualize. On the first row you can **select which plants** you want to include in the data. On the second row you can select if you want the data of **car batteries and/or home batteries**. On the next row you can select if you want the data about the **amount of sales** or if you want data about the **delays of the sales**. On this row, only one checkbox can be checked. The following row contains a **slider to select the period** you want to consider for the data. The **last row is only available if you have selected the “Sales” checkbox** on the third row. If you have checked the first checkbox, the radius around a city on the map will correspond to the sum of the sales to all customers in that city. If you have checked the second checkbox the radius around a city will correspond to the maximum sum of sales to one customer in that city. The reason for this distinction is because, in Berlin, there are a lot of customers and in most other cities there is only one customer. So with the first checkbox checked, Berlin will have the max radius around it while nearly all other cities will have the smallest radius around it. While, with the second checkbox checked, the radii around the cities will be somewhere between the smallest and biggest radius.

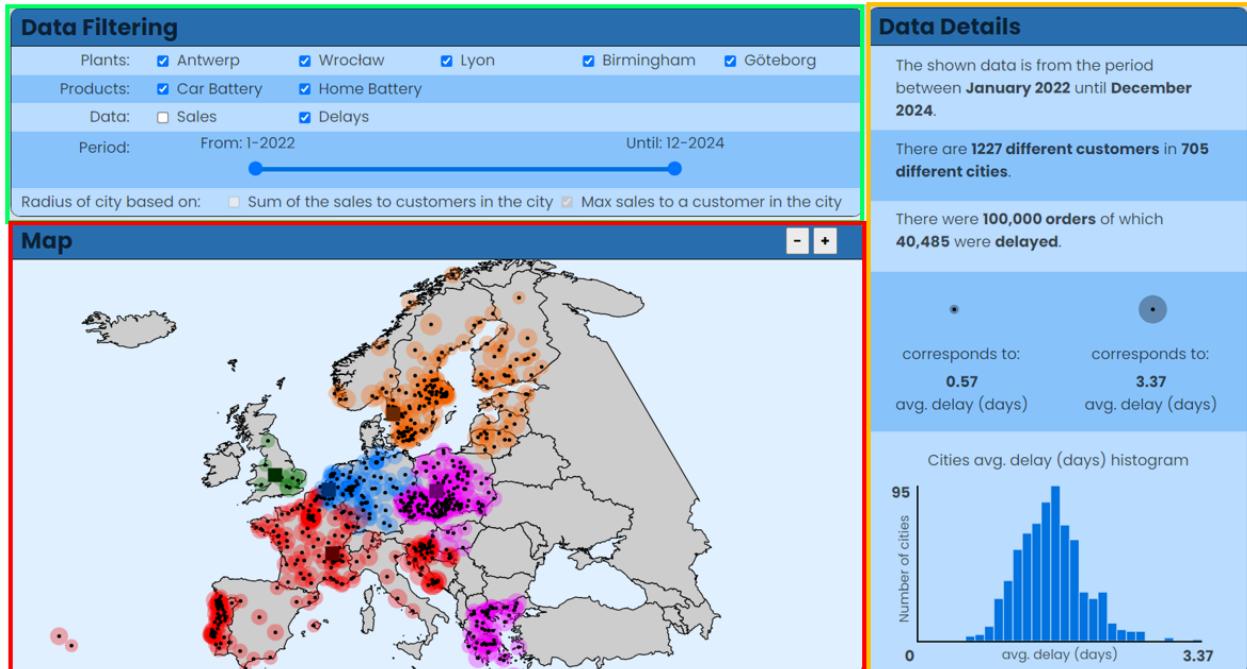


Figure 9: The different interactive sections of App2.

The “Map” section shows the cities of the customers. Each city which has data according to the filter settings is shown on the map with a black circle. Around this circle, another circle is drawn with a low opacity, a certain color, and a certain radius. The **color corresponds to which distribution center** delivers the product to the customers in that city. The **radius corresponds to the value of the filtered data** belonging to that city. For example, if you have selected the “delays” checkbox, like in Figure 9, then a small radius around a city means that the city has low average delays for deliveries, while a large radius around a city means that the average delay is bigger for that city. This section has dynamic functionality as well. You can **drag the map around and you can zoom in or out** with the buttons located at the top right in this section. You can also **hover over a city and data about the city will be shown**, depending on the type of data you have selected in the “Data Filtering” section. On the left of Figure 10 you can see the hover effect if the “Delays” checkbox is selected. On the right you can see the effect if the “Sales” checkbox is checked.

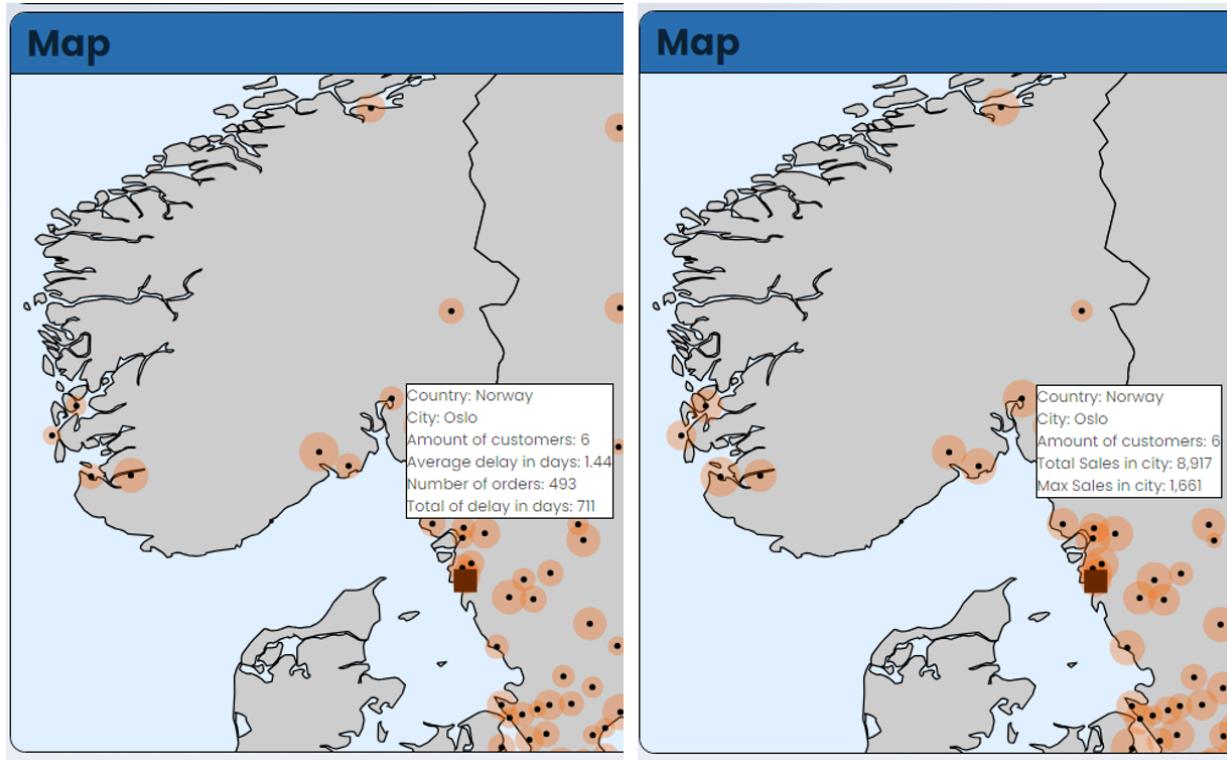


Figure 10: Hover effects of the “Map” section for App 2.

The last section in Figure 9 is the “Data Details” section. In this section, some extra details are provided for the filtered data. Firstly, the **period that is considered** for the data is stated. Then, **the amount** of different customers and the amount of different cities the customers are located in is shown. Then the data shown in the **following row depends on what “Data” checkbox you have selected**. If the “Sales” checkbox is selected this row will show how many car and/or home batteries are purchased, depending on the product checkboxes. If the “Delays” checkbox is selected then it shows the amount of orders and the amount of delayed orders. The next row shows what minimum and maximum value the minimum and maximum radius correspond to. These radii are the radii drawn around the cities. Every radius around a city falls between these minimum and maximum radii depending on the value the city corresponds with. Lastly, There is a **simple histogram to show the distribution of the data of the cities**. There are also

hover effects for the histogram. If you hover over a bar there is information about the bin ranges of the corresponding bar and the value of the bar.

Part 5. Findings

Suncharge demanded an intuitive monitoring system to detect supply chain issues and identify areas of improvement to optimize their logistics. The current visualizations mainly delve deeper into Suncharge's sales data.

App 1: Radar chart with calendar glyphs

The first app visualizes trends in customer demand. The user can visualize the monthly and daily sales for the product, plant and year they like. A couple of things stand out.

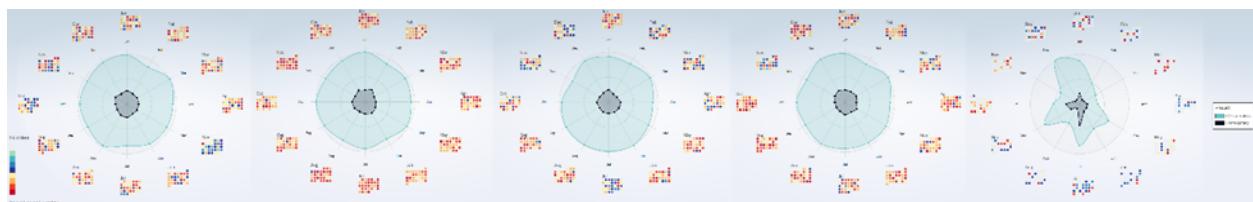


Figure 11: Sales for both products in 2024, over all plants (from left to right; Goteborg, Antwerp, Wroclaw, Lyon, Birmingham)

Firstly, as shown in Figure 11, it can be seen that the **sales for home batteries are consistently lower than the car battery sales** for all plants, in 2024. Using the app, it can be seen that this is true for 2022 and 2023 as well. Though it is not clear in the print screen in Figure 11, the dynamic visual allows users to hoover over the dots in the radar and thus reveal the monthly sales. This shows that the **Antwerp plant sells the most by far**, followed by Lyon and Wroclaw who sell approximately the same amount, then Goteburg, and lastly **Birmingham with the least amount of sales**. In fact, the Birmingham sales are consistently about 1 or 2% of the sales of the Antwerp DC.

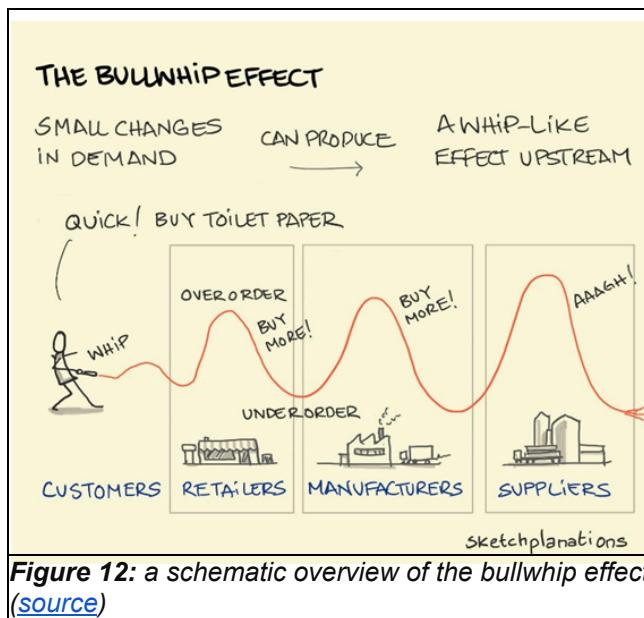


Figure 12: a schematic overview of the bullwhip effect ([source](#))

Birmingham's sales are very variable. This is likely partially due to the low amount of sales in general, coming from a small group of customers. However, the bullwhip effect is a well-known phenomenon in supply chain management where unpredictable demand may lead to serious and costly problems upstream. Figure 12 shows this schematically; each partner in the upstream supply chain needs to take some precautions to be able to work with variable demands. The retailer may have to hold more inventory since the demand is so unpredictable. The manufacturer and the supplier may invest in

extra production capacity (eg. machines) since demand may be unexpectedly large in some months but this expensive extra production capacity will not be used in periods with low demand. Since the total sales of Birmingham are low, this unpredictable demand may not pose a big problem yet as their demand is

pooled together with the more predictable demands of the other plants. However, it may be worthwhile to look into what is causing this unpredictable demand (e.g. badly timed promotion campaigns, unreliable resellers) and adjust accordingly. Additionally, Suncharge should reconsider whether the operating costs of the Birmingham plant are worth the meager sales and possible disruptive supply chain effects. It is also interesting to see that there tend to be more sales on Monday and less on Sunday. However, exceptions exist. In fact, we were surprised to see that there does not seem to be a very strong weekly pattern in the sales (Figure 13).

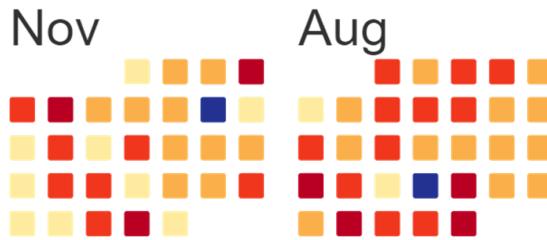


Figure 13: Antwerp DC, 2024 sales. November seems to have a weekly pattern with less sales on Sunday (first day of the week) and more on Monday or Tuesday. But the same pattern cannot be unveiled in the month of August, for example.

App 2: Geographical sales and delays

This visual was intended to serve as a comprehensive analysis of SunCharge's supply chain, focusing on both upstream and downstream processes. It aimed to address our first and third guiding questions concerning the efficiency and effectiveness of SunCharge's operations. Our first guiding question aimed to investigate the upstream supply chain, by delving into the structure and diversity of suppliers, assessing whether any monopolies exist on raw materials and the speed at which these materials are delivered. Our second guiding question on the downstream side examines the distribution network, identifying which distribution centers cater to specific customer segments, assessing the timeliness of order deliveries, and analyzing sales patterns to optimize future distribution strategies. While we were unable to address the first guiding question due to the absence of external vendor data in the SunCharge dataset, our analysis has shed light on several other critical aspects.

Firstly, we identified the distribution centers (DCs) serving specific customer segments and proposed strategic locations for new DCs based on geographic demand patterns. Linked to specified guiding questions: **Which DCs serve which customers? Where to open new DCs for sales?** For instance, upon selecting the sales option in the data filtering section, a clear geographical division for each plant became evident (Figure 14). The Antwerp distribution center emerged as a key player, serving 505 customers across Belgium, the Netherlands, and Germany. Meanwhile, the Lyon DC efficiently caters to customers in France, Spain, Portugal, Italy, Switzerland, Slovenia, and Croatia (n=269). The Wroclaw DC serves Poland, Czech Republic, Hungary, and even Greece (n=232) and has customers with the largest orders per capita. The Goteborg DC serves Denmark, Sweden, Norway, Finland, Estonia, Latvia, and Lithuania (n=208), while the Birmingham DC serves only 7 customers in the UK. This analysis underscores the significance of the Wroclaw DC in SunCharge's operations, while highlighting the limited contribution of the Birmingham DC to overall sales. Notably, Wroclaw's extensive coverage, including Greece—a geographically distant region with a significant customer base—suggests potential optimization by closing the Birmingham plant and establishing one in Greece. We also observe in the data details of Figure 14 a higher sales volume of home batteries compared to car batteries, indicating the former's popularity among customers.

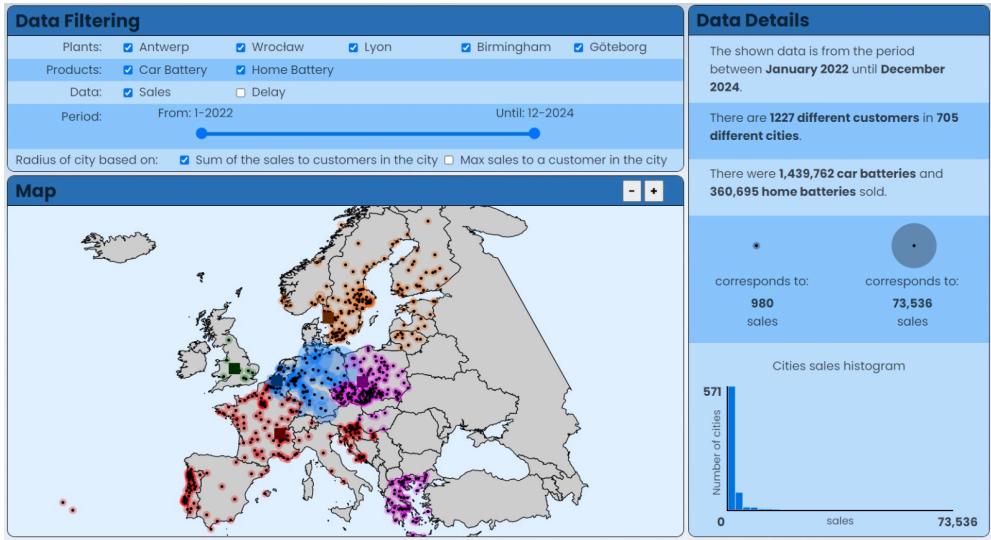


Figure 14: Geographical division of customer sales for each plant

In terms of delivery efficiency, we delved into order fulfillment timelines and pinpointed customers facing longer delivery times. Guided by the pre-specified questions: **Can orders reach customers on time? Which customers face longer delivery times?** By selecting ‘delays’ in the data filtering section, we uncovered varying delay durations across different DCs and products, as seen in Figure 15. Our analysis reveals that Wroclaw and Goteburg experience the largest delays, attributed to their extensive coverage across multiple countries. Notably, Wroclaw faces particularly significant delays due to its customer base in Greece. Based on both analysis of delays and sales, we recommend optimizing operations by potentially closing the Birmingham plant, which serves only seven customers, and establishing a new distribution center in Greece. The Antwerp distribution center can efficiently serve UK customers, while Wroclaw, with the highest delays, would benefit from decreased average delay times. Additionally, we observed that home battery deliveries are notably more delayed than car batteries, with differences ranging from +1 to +4 days, particularly pronounced in Goteburg and Lyon. Given these findings, it may be prudent to prioritize optimization of the home battery supply chain before addressing car batteries, particularly since customers in Goteburg and Lyon ordering home batteries face the longest delivery times, potentially impacting overall customer satisfaction.

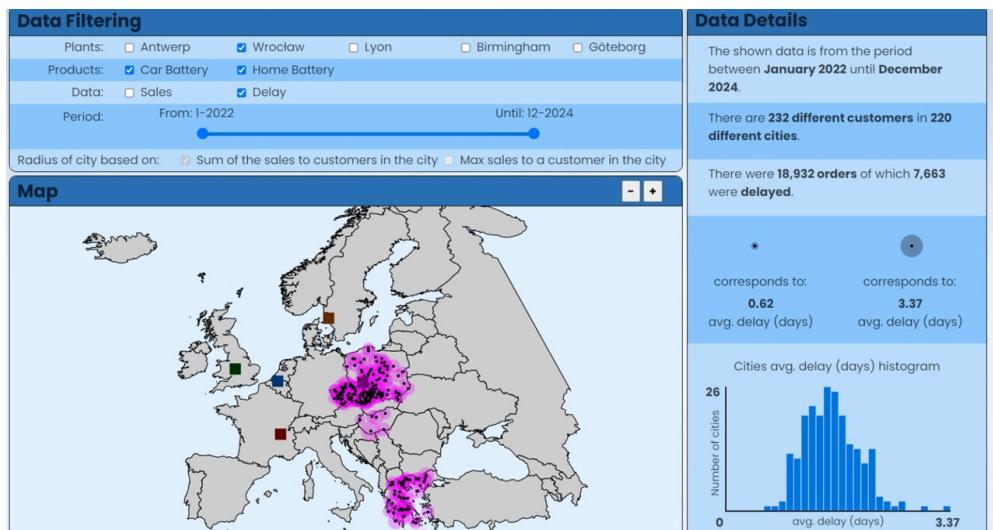


Figure 15: Geographical division of the plant (Wroclaw) with the highest delivery delays both products

When examining emerging patterns (based on a pre-specified question: **What sales patterns emerge?**), we observe distinct temporal differences between the years 2022, 2023, and 2024. Notably, 2023 stands out as the worst year in terms of sales performance, characterized by low quantities per customer. This trend is marked by a decrease in sales between 2022 and 2023, followed by an increase from 2023 to 2024. However, it's noteworthy that overall, sales decreased between 2022 and 2024 across both products. In contrast, the analysis of delays presents a different narrative, indicating a general pattern of process optimization from 2022 to 2024, with decreased delays overall. Nevertheless, a closer examination reveals deviations from this pattern across various plants and for both products. For instance shown in Figure 16, while Antwerp and Goteborg DCs experienced increased delays over the three years for home batteries, Lyon and Wroclaw demonstrated improvements in delays during the same period. Additionally, despite the overall trend of decreased delays, many plants exhibited their highest delay rates in 2023 compared to other years. Lyon, for example, improved delays for home batteries but faced larger delays for car batteries. For all plants, the delays for home batteries were larger than car batteries as seen in an example in Figure 16.

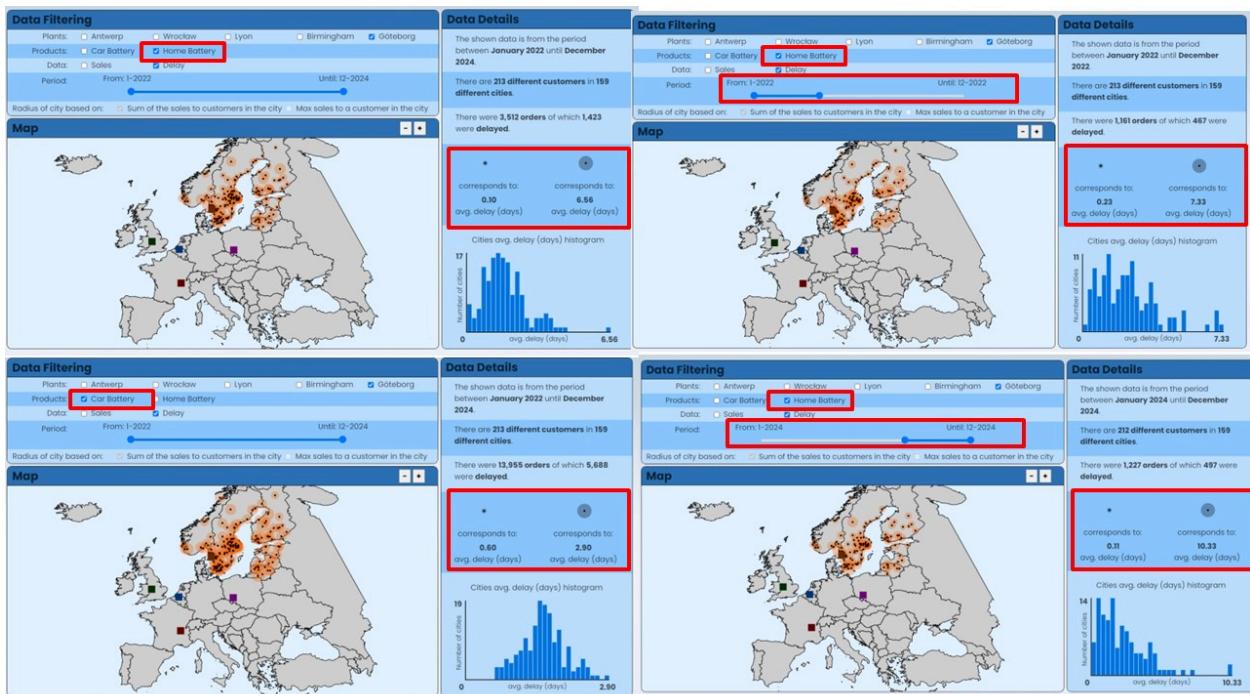


Figure 16: Left panel shows an example of higher delivery delays for home batteries than car batteries in the Göteborg DC. The right panel shows increased delivery times between the years 2022 and 2024 for the Göteborg DC.

In summary, optimization suggestions for results extracted from app 2 include:

1. Assessing the feasibility of expanding operations in key regions or consolidating operations elsewhere, such as closing the Birmingham plant and opening a new distribution center in Greece, based on delay patterns and sales distribution. Adjustments to the distribution network could enhance efficiency and customer satisfaction.
2. Prioritizing optimization efforts for home battery supply chains, especially in regions like Goteborg and Lyon, where customers experience prolonged delivery times, to improve efficiency and customer satisfaction.

3. Recognizing the higher popularity of home batteries over car batteries, as indicated by sales volumes and delivery times, and allocating resources accordingly to drive overall sales growth.

Part 6. Reflections

We are most proud of the fact that all three of us succeeded in visualizing the data using Svelte(kit). Neither of us had ever worked with Svelte before and only one of us had some programming experience in Javascript.

We are least proud of the limited breadth of topics that we could look into for the project. We only look at the downstream supply chain, mainly sales and delivery delay patterns while the dataset was much richer. The unavailability of some data on the upstream supply chain (such as purchases for all raw materials) prevented us from looking too much into this part of the supply chain. However, we could have also looked into SunCharges' internal processes such as the inventories, internal shipments between production plants and DCs, or the bill of materials. Additionally, since we had limited time to implement the visualization, we did not have much time left to clean up and structure the code scripts properly.

Part 7. Individual contributions

All team members contributed equally. The Table below shows each team member's contribution in detail.

	TASK	Steven	Tamaya	Raïsa
Diverge	Designing visuals	x	x	x
	Feature-by-feature matrix	x		
Emerge & converge	Choosing visuals to combine	x	x	x
	Combining visuals		x	x
Implementation	Visual 1: sales in a radar chart with calendar glyphs		x	x
	Visual 2: plants and customers on a world map with the flows between them	x		
Writing	Report 1: Implementation	x	x	x
	Report 2: Designs	x	x	x
Video	Report 2: Creation of video material	x		x

Appendix A. Sketches Overview

Please also check the [Miro board](#) for a better view of the designs (full link:

https://miro.com/welcomeonboard/eEtIM05FcXd2dXZjWTZEZDJtemY5eWVMy081UldHRFplaDdjOXN2d09XbWVLa3FDSzlybko5b3c2ajlIMFpCUXwzNDU4NzY0NTgxMjAyMjE3NDczfDI=?share_link_id=267801719843.

