Master of Science in
Information Systems Business Analytics

Advanced Visual Analytics Exam

Candidate numbers: 29,27, 34
Date: 15.11.2024
Wordcount: 7697

Kristiania University College
Prinsens Gate 7-9
0152 Oslo
Norway

1. Introduction:

This report is about PostNord, a company that provides multiple product types for logistics, including parcel delivery and freight services, serving both a B2C and B2B market. In writing this report, we intend to highlight key information regarding products, industries, and their clients over time to illuminate important business factors, demand drivers, and interactions with the market. The end product of this report is this report, cleaned data in a CSV format, workflows and a dashboard file with 3 dashboards. Our goal is to discover operational understanding and highlight actionable insights for the management of PostNord.

To clarify how to read this report, we want to describe the following sections. First, in section 2, we deal with the methodology, a five-step process, incorporating the ETL process, using Alteryx. Second, in section 3, we explain the design process, expanding on the ETL theory, limitations, assumptions, dashboard iteration process and heuristics. Third, in section 4, we discuss the actionable insights, and final dashboards, and answer the report questions. Finally, in section 5, we conclude the report and provide some recommendations for the future of PostNord.

Background:

PostNord was forged by the merger of Swedens Posten AB and Denmarks Post Danmark A/S and has retained its position as a central company in the delivery sector. It is a government owned company, with a 60/40 equity position between Sweden and Denmark respectively. Post-merger, the company expanded to Norway. [1]

From the 2023 annual report of PostNord, there are 4 significant comments made by their CEO, Annemarie Gardshol, made four significant comments. We understand that the data we have been provided is from 2017 to mid-2019, yet we expect that these current comments were created through a multi-year process rather than a singular, crossectional atemporal perspective.

First, their goal isn't to improve fourth-quarter income. Second, they aim to become the market and cost leader for parcels in the Nordic region, expanding services to parcel lockers. Third, is the emphasis on financially sustainable mail business, and desire for regulatory relief. Fourth and final is their prioritisation of being ESG-friendly:

- "We have set a prioritized climate agenda with the goal of being fossil-free by 2030. In the quarter, carbon dioxide emissions decreased by 16 percent, compared with the fourth quarter in 2022."
  Annemarie Gardshol, CEO

Report questions:

From Mrs Gardshols comments questions regarding the business emerge:

What are the most popular routes? What can we do to decrease emissions? And how can we do so in an economically feasible way? These questions are united by the central question of this report:

If PostNord was to increase their operations, which market, product and sector would be the best to fulfull their corporate goals of parcel dominance, sustainability and financial growth?

2. Data Transformation and Workflows

Methodology:

In the methods section, we are using the five-step, Big-data pipeline (BDP), workflow model described by (Fisher, 2024)[2]:

[1] About Postnord

[2] Fisher, D. *et al.* (2012) 'Interactions with big data analytics', *Interactions*, 19(3), p. 54

1) *Acquiring data (in methodology)*
2) *Choosing an architecture (in methodology – expanded on in design and implementation)*
3) *Shaping the data to the architecture (in methodology)*
4) *Writing and editing code (Not represented - no-code solution)*
5) *Reflecting and iterating on the results (Represented in actionable insights)*

Given that we use Alteryx (no code), we naturally do not represent step 4 in our analysis. Additionally, step 5 (reflecting and iterating on the results) will be explained in section 4—Actionable Insights, and step 3 will be further developed in the subsequent section (section 3).

Big data pipeline workflow:

1) **BDP step 1 - Acquiring data:**
   Our data sources were the exam files and data downloaded from the GeoNames database. From the Exam documentation titled "Exam_MB221_AdvancedVisualAnalytics_2024", we ascertained the following types of files - (a) DDS Dropbox data, (b) Product files, (c) Master data file, (d) About the product – PowerPoint presentation.

   a) **Product files – (DatadumpProduct*.xls):**

      From the DDS Dropbox, there are folders for each of the products, structured as data dumps, for all 4 products we are to analyse. The products Innight, Groupage, and Pallet have data from January 2017 to August 2019. For these products, there are 64 files, all ordered as data dumps (for example – "DatadumpProduct G_2017_apr.xls"). The product road freight only has data from May 2019 to August 2019, with a singular file for loading. Additionally, we have the instructive Microsoft Word document "Description_of_Variables", which serves as one of two documents to describe the variables and their conditions in the dataset to be made (the second being the About product PowerPoint document). Please consult the graph below regarding the shared product variables, variable description and data types.

      [Appendix 1 - Product Data]

   b) **Master data files:**

      From the Master file data, one folder yielded geographical data files that were not complete and ultimately not used. Second was the ADS_DDS file, which had three sheets. Of these three sheets, we used only the first one, as the second sheet is for the courier (C) product (which we are not to analyze). Third and final is postcode zone data, which we did not need as we used it as we downloaded data from the Geo data webpage.

   c) **Geodata:**

      The geodata was downloaded, with each country's ISO-3166 alpha2 code (NO, DK, SE) selected separately from the geodata webpage.

   d) **About the product - PostNord slide deck:**

From the “about product” link, we attained the PowerPoint presentation which described the capacities and conditions of each product type. These conditions later proved useful when cleaning the data.

2) **BDP step 1 - Choosing an architecture:**
From (Fisher, 2024) we need a specific architecture to define the actions of our project. Because we were directed to download and tidy data, we have determined that the correct architecture would be an ETL-type method architecture. ETL stands for extract, load and transform, and is a method of combining data sources and files for analysis and storage in a central repository [3] [4]. Our data is explained above, our extraction was done from the exam drop box, the transformation was completed with Alteryx and the central repository we used was Google Disk whereby we uploaded the CSV. files after transformation. Hence, this project, as executed by us, has met all the prerequisite conditions to be defined as an ETL process. We will discuss the details of the ETL process execution in section 3 -design and implementation.
Additionally, we chose to execute this architecture using Alteryx for data tidying and Tableau desktop for visualisation. Our justifications are that although Python is excellent for tidying and has a lot of versatility as a tool, we find the large map-like workflow in Alteryx to provide some key overview of all the data steps and give a clear understanding of all steps in a relatively small space. Then we chose Tableau desktop due to our familiarity with the tool relative to other tools like Power BI, which would make the end product better.

3) **Shaping data into architecture - data cleaning steps:**
All steps are seen in the table for:

[Appendix 2 - data cleaning steps]

3. Design & Implementation:

This section will expand on the definition of the ETL process, and iterations of the dashboards, with comments on design, and design heuristics.

ETL Description:

As mentioned in the methods section of this paper we have elected to use the ETL process as an architectural base for our project operations. Hence, we desire to elucidate the components of the ETL process to provide clarity through greater granularity.
As previously mentioned, the acronym ETL stands for Extract, Transform and Load.

a) **Extracting process:**
The extraction process is defined as:
- “In data extraction, extract, transform, and load (ETL) tools extract or copy raw data from multiple sources and store it in a staging area. A staging area (or landing zone) is

[3] What is ETL
[4] Central Repository

an intermediate storage area for temporarily storing extracted data. " (Amazon Web Sercvices) [5]

Our extraction process was the process of gathering data of correct typology. The extraction process can be divided into two parts (a) exam data extraction, (b) geographical data extraction.
From the exam data, we simply navigate through the folder process and load the product data (G,I,R and L), and the master data file (DDS master data file). As mentioned, the master data has 3 sheets, whereby the first sheet was the only one to provide necessary information. Our way of determining this was by comparing the information between sheets. As they were conflicting we, chose the first sheet for extraction.

From the geographical data extraction, we first inspected the DK_Geonames, DK_Tableau, files, and found them lacking in information. Hence, we elected to extract the data from the GeoNames website, for more complete information regarding geographies.

The above files were primarily of xls format, and some in csv format.

Our staging area was our respective computers where we had our data, until we uploaded them to the google disk as our central repository after transformation.

b) Transform:

The second component is the transformation, which removes errors, fixes duplications, defines datatypes, creates new variables, and removes unnecessary ones. The transformation component is integrated with the tidy data concept from Wickham (2014).
The transformation is described in section X about data pipeline. In broad strokes, considered some key problems.

- Dealing with duplicated rows
- Deducting which columns are relevant
- Dealing with datatypes
- Excluding records based on predetermined product conditions.
- Types of joining/ unions to perform
- Dealing with NA values and 0 values.
- Dealing with column NA value in need of filling (SPL and PLL)

We attempt to make this into a tidy data format with considerations regarding the basis of tidy data as described below:

- "Tidy data is a standard way of mapping the meaning of a dataset to its structure. A dataset is messy or tidy depending on how rows, columns and tables are matched up with observations, variables and types. In tidy data:
    1. Each variable forms a column.
    2. Each observation forms a row.
    3. Each type of observational unit forms a table." (Wickham, 2014, page 54)[6].

[5] What is ETL
[6] Wickham, H. (2014) 'Tidy Data', *Journal of Statistical Software*, 59(10). doi:10.18637/jss.v059.i10.

From this definition we experienced we had multiple suggestions of how to tidy the data, and we further discuss these variants in the following section on limitations and assumptions.

c) Load:

As the dataset is transformed, and cleaned to the desired format, we proceed to upload it to the desired central repository. The definition of a central repository is:

-“ A central repository is a location where all relevant documentation, such as policies, procedures, and case information, is stored and easily accessible to team members.” [7]

From this reading we determine that our central repository would be the teams shared google disk folder for Advanced Visual Analytics. Another reading would be that the WiseFlow delivery would be the central repository.

Limitations and assumptions:

This section consists of (a) limitations and comments on the data tidying process, and (b) the list of assumptions we made regarding the data tidying, followed by our augmentation justifications.

a) Assumptions and justifications:

i) We assume that each row of product type L implies a singular pallet. This is because if the user wanted multiple pallets, they would have chosen a different product, namely Groupage or Road Freight, which markets multiple pallet services. Hence, by choosing either of these product types, we assume that there can be multiple pallets compared to I and L. The idea is that for product type I, there is no mention of pallets; hence, we assume that they only send parcels.

ii) We presumed that all rows exceeding their weight requirements should be excluded. This is because the weight requirements of G and L stipulate certain maximal conditions (2500 and 1000 KGs, respectively), which means that if the truck is overweight, it is not feasible to drive, according to PostNords' own documentation.

iii) We assume that for parcel count equal 0 in the road freight product category, rather than 0 goods being sent, the entire truck space was ordered. This assumption comes from conversing with our professor and can be confirmed by Lester.

iv) We assume that for all Price_Paid columns equal 0 we assumed that the transaction was invalid and removed them from the data frame. (check)

v) For parcel count equal 0, and consignor weight equal 0, we assume that these items were simply not counted or weighted. The reason is that there might have been some time crunch, logistical error or simply human error.

vi) For main customer, system, allocated SAP number, Industry level 1 and 2 NA values we assume that they are not material to the analysis and we keep them as they are.

vii) We assume that completely duplicated rows would be, non material and should be removed since the shipping tracker ID should be unique for each transaction, and there would be no reason to think that these revenues would actually be collected.

[7] Central Repository

viii) We assume that since each row should be a transaction that we should not melt the columns ConsignorCountryCode and ConsigneeCountryCode (aswell as Consignor Location Nr, Consignee Location Nr) should not be melted together. this is explained further in the limitations section below.

b) Limitations and comments:

i) First, from the description of variables, we do not know precisely if a product delivers parcels or pallets selectively. This is simply one example of uncertainties relating to the data. Another example would be the 0 value.

ii) Second, is the dk_geonames and dk_tableu data, which was not complete, and exhaustive, which caused us to use supplementary data for geographies.

iii) Third, there would be the lack of ETL documentation available freely on the internet, referenced in the syllabus from the exam paper. To discuss the formal ETL structure we therefore used additional external papers.

iv) Fourth would be the time limitation.

v) Fifth is that we do not possess all the costs (only surcharges, toll and freight). Not knowing the full cost is detrimental because we cannot know whether a high revenue customer is the same as a high profit margin customer.

vi) Sixth, is that the road freight data is only available for the last couple of months, which makes any visualization of revenue sum skewed, because road freight has fewer observations hence a lower total.

vii) Moreover, It would be great to know the customer names rather than customer 1, 2..n, because it would allow us to compare direct competitors. Insight into the customers in general is quite hard due to the lack of columns.

viii) Eitght we were limited when graphing by we found that the amount of categories was too large, and the graphical representations where too squeezed together.

ix) Ninth, we did not have complete data from year start to year end in 2019, which meant that quarterly analysis would not be as useful as it could have been.

x) Parcel/pallet type is a bit obscure because one is a standard meassure of space whilst the other one is simply a package. Because the variable is described in categoricals of fraction of a pallet, then the parcel component of this variable is not expressed and should in our estimation not be a part of the name of the variable.

xi) Are there any rows without postal codes which do not work in a company who's operation is to deliver packages to specific destinations.

xii) We do not know the size and types of fuel for the different trucks and products. Additionally we don't know the routes of the different trucks which means that it is impossible to determine the emissions per trip.

xiii) Packages are not in uniform size such as the pallets. This is a limitation because we see that the maximal amount of packages in a trip was 3333, however we have no way of determining whether or not this is reasonable to drop this row or not.

xiv) In python NA values are possible to impute in a numeric (float or integer value), however in tableau desktop NA values, would make the entire variable into a string. Therefore, we had to make values into 0, and then filter them out when making the visualizations.

xv) We assume that given the definition of Wickham, as "each each observation is a row", would make it wrong to melt the data, for the columns consignors country code, and consignee country code in addition to the Consignee Location Nr, Consignor Location Nr. In a perfect world, we would love to separate rows into consignor or consignee.

xvi) We are limited by not knowing which orders are in which trucks, since we would then have a better understanding of efficiency between the different product types, and the emissions per package.

Process and iteration:

Our professor informed us (although not explicitly on the exam sheet) that it would be instructive to describe our process of iteration regarding dashboards and their creations. This section is, therefore, something of a storybook, where we display sketches (see the appendix) and their dashboards. Moreover, we critique these dashboards, using the body of literature provided to make better advanced visual analytics.

We decided that we were going to have 3 dashboards. We named them (a) Freight & products, (b) Product and Industry, and (c) Revenue over time, and will now review their iteration processes.

a) Routes - DB process:

[Appendix 3 & Appendix 4]

Given the ESG sentiments expressed by the CEO of PostNord, we wanted to start with the highest freight paths, which happened to be in Denmark. This made us understand that we needed an overview of the freight operations of the three countries to explain how dominant Denmark's current operations were. Moreover, with this display, we could also see the markets where PostNord has infrastructure (current routes, drivers, etc), but not a large amount of business. This first graph aids us in determining which country could be great for expansion within the Nordics. This dashboard only contains two graphs, but they therefore have more space and allow for a good overview.[8]

Then we have consistent colouring between graphs, which is good, although we still need appropriate colour pallets for the colour blind. [9]

Our rationale for choosing bar graphs in all sections of the dashboard is that they again signal quite clearly differences in routes for products in a cross-sectional, determined period of time. They are easy to understand and make the eyes focus on graphs line by line rather than zig-zagging too much right to left, which could be a cause of distraction.

b) Industry and Products DB process:

[Appendix 5 & Appendix 6]

Iteration 1 - Product & industry:

Our main changes were removing the average revenue chart, resizing the bar graph, and updating parameters. Our academic justification for making these changes were primarily related to formatting design and attention management. We decided to remove the Average revenue per product because it did not fit very well, design-wise, on the current

[8] Shneiderman, B. (1996) 'The eyes have it: A task by data type taxonomy for information visualizations', *Proceedings of the 1996 IEEE Symposium on Visual Languages*, p. 336

[9] Bera, P. (2016) 'How colors in business dashboards affect users' decision making', *Communications of the ACM*, 59(4), p. 50

dashboard, and it drew a lot of attention. We find that charts in their current format seem to not generally manage attention well. As explained in page 3 of Toreni (2022), we must aim to create an attentive UI designed to assist users of this dashboard in managing their attentional resources. This concept is further expanded on as Design Principle 1. [10]

Moreover, we would be breaking with the principle of "Overview first, zoom in, then details on demand", as explained by Shneiderman (1996) since we were since we started with a bar graph, which would act as some kind of overview and then moved to a tree graph which again was a type of overview.[11]

Regarding colours we realised that for people with colour blindness there might be some difficulty in observing visualisations, hence we decided to commit to a colour-blind pallet, in later versions, as emphasised by Bera (2016)[12]

[Appendix 7 & Appendix 8]

Iteration 2 – Product & industry:

We wanted to display the KPIs or tables at the top of the page as presented in class, with sharper formatting, which separates the sections and allocates attention better.

Moreover, we were a bit dissatisfied with the insight from the tree graph and opted to focus more on the sub-industries rather than products. This also helped us to improve storytelling by creating a clear background and context and then giving more granular information, as explained by Knaflic the quote:

-"Context is essential in storytelling with data, enabling clearer communication and better decision-making." [13]

By creating two graphs with the guided drill-down action path, we use individualized visual attention feedback. The products were made to a parameter to make the dashboard more interactive and give feedback, which improves allocation of attention (Toreni, 2022).

[Appendix 9 & Appendix 10]

Our rationale for selecting the bar graphs is that we want to compare the levels of different industries and sub-industries at singular points in time, and they allow for straightforward comparisons. Moreover, we turned them sideways because the highest categories (i.e., the biggest industry served) are so much larger than the other industry values that one could not observe the other values in a vertical graph. Moreover, the dashboard has an overview and zooms, which makes grasping the information more accessible.

---

[10] Toreini, P. et al. (2022) 'Designing attentive information dashboards', *Journal of the Association for Information Systems*, 22(2), p. 528

[11] Shneiderman, B. (1996) 'The eyes have it: A task by data type taxonomy for information visualizations', *Proceedings of the 1996 IEEE Symposium on Visual Languages*, p. 336

[12] Bera, P. (2016) 'How colors in business dashboards affect users' decision making', *Communications of the ACM*, 59(4), p. 50

[13] Knaflic, C.N. (2015) *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, p. 19.

c) Revenue over time - DB process:

[Appendix 7 & Appendix 8]

Iteration 1 – Revenue over time:

Our first critique was that the dashboard provided a information overload. We were mostly satisfied with the stand-alone graphs, but we realized that we needed to remove unnecessary visuals.[14]

Again, we faced some difficulties with the colour pallet department, but elected to finish detailed formatting in the later versions, save time. The sentiment of simplicity, and clarity of dashboards was echoed in the "Visualization Design Practices in a Crisis".[15]

Then we had some difficulties with the information seeking mantra, where we managed to display mainly overview graphs. We found that this was due to the nuances of revenue by country, and differences in average and total revenue. However, we needed to improve these graphs to have a more succinct financial insight. [16]

Moreover, we were somewhat overwhelmed by the amount of data and could use some context before getting into the graphs themselves. This is also highlighted by some clutter in the graphs.

- " Clutter is your enemy! Remove unnecessary elements that do not contribute to the narrative."[17]

We see that the curn rate is a bit uninformative, because the majority of the bars are completely uniform to each other. It is a ratio which is hard to visualize, although customer churn is an interesting component of the business. It is however very important to be able to gather something from a graph. In this case the churn worked on paper but not in reality, and without any drilldown options. Therefore, the dashboard becomes a bit fragmented and lacks some user centric design component. Therefore we decided to change the format more dramatically, and focus more on the revenue over time as a overview graph and have some smaller graphs inspecting specific periods of time.

[Appendix 9 & Appendix 10]

Our rationale for choosing these graphs is that we here represent the time component on the x-axis, which makes this a time series, hence the line graph is the appropriate graph, as it makes the graph less cluttered. This is true for all the graphs, except the very left graph, describing a singular point in time three months before and three months after, which are

[14] Zhang, Y. et al. (2022) 'Visualization design practices in a crisis: Behind the scenes with COVID-19 dashboard creators', *IEEE Transactions on Visualization and Computer Graphics*, p. 1037
[15] Zhang, Y. et al. (2022) 'Visualization design practices in a crisis: Behind the scenes with COVID-19 dashboard creators', *IEEE Transactions on Visualization and Computer Graphics*, p. 1037
[16] Shneiderman, B. (1996) 'The eyes have it: A task by data type taxonomy for information visualizations', *Proceedings of the 1996 IEEE Symposium on Visual Languages*, p. 336
[17] Knaflic, C.N. (2015) *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, p. 71

singular time points, hence crossectional with two buckets, and are best represented by bar graphs as suggested by Knaflic.[18]

Heuristics use:

Jacob Nielsens General principles:

Jacob Nielsen's 10 general principles for interaction design were used as a broad rule of thumb for designing our dashboards (Nielsen, 2024)[19].

1. **Visibility of system Status**

The response for using the filters on the route dashboard is a bit slow since it is filtering through approximately 3 million rows of possible routes. Therefore, the response time is slower than what is wanted for this specific dashboard.

2. **Match between the system and the real world**

Since we are making the dashboards for the sales/product manager we have used terms they are familiar with like revenue or country Codes like DK for Danmark etc. We also have used the same control layouts to make the user familiar with the different dashboards. The filters are common drop-down menus where the user can click on one or multiple values.

3. **User control and freedom**

The dashboards where the user may click on multiple filters at once have a reset filter button. This button works like an emergency button if the user wants to go back to the original layout. The buttons are visible and easy to see and interact with.

4. **Consistent and standards**

Every dashboard follows the same layout for colors, positions, and variables. This keeps the users from needing to learn something new from the other dashboards, making it our standard throughout the dashboard process

5. Error prevention

We made sure no filters or other actions the users could make would lead to any errors. However, if this is not the case for the user, the “Reset All Filters” button will reset the dashboard to its original form.

[18] Knaflic, C.N. (2015) *Storytelling with data: A data visualization guide for business professionals*. John Wiley & Sons, p. 134

[19]Nielsen, J. (2024) *10 usability heuristics for user interface design*, *Nielsen Norman Group*. Available at: https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed: 15 November 2024

6. **Recognition rather than recall**

   The dashboard has a consistent design to support recognition. Every dashboard has an information button that clearly describes the functions and graphs used within it.

7. **Flexibility and efficiency of use**

   We recognize that users have varying levels of expertise with interactive dashboards. Initially, we aimed to allow expert users to interact directly with the graphs, by passing on the need for filters. However, this approach introduced issues, as it risked compromising the visuals or even causing entire graphs to disappear if multiple variables were selected. To maintain dashboard functionality and ensure a consistent experience for all users, we ultimately decided to remove this feature.

8. **Aesthetics and minimalist design**

   We minimalized the number of variables in our graphs. We started with having Top 10 Routes but quickly understood that 5 would be better.

9. **Help users recognize, diagnose and recover from errors**

   As mentioned earlier, the dashboards should not display any errors, and the reset button would recover any errors if they should occur.

10. **Help and documentation**

    Tooltips and the information button are available if the user needs any help or more context.[20]

Guidelines for Dashboard Designing

We also included some specific guidelines from (Kaufman, 2004) for dashboard design since Nielsen's 10 rules of thumb are more general to further improve our dashboards.

We incorporated specific guidelines from Kaufman (2004) for dashboard design to complement Nielsen's more general "10 Rules of Thumb" and enhance our dashboards.

Guideline *G1.3* has been applied throughout all dashboards. In the Industry and Revenue dashboards, key figures such as customer count, revenue, and average revenue are scaled larger with a more vivid background color to help them stand out. Similarly, in the third dashboard, the revenue line graph occupies the top of the layout as an overview, with the graphs below serving as drill-downs for more detailed insights.[21]

---

[20] Nielsen, J. (2024) *10 usability heuristics for user interface design*, *Nielsen Norman Group*. Available at: https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed: 15 November 2024).

[21] Ware, C. (2012) *Information visualization*. 3rd edn. Morgan Kaufmann, p. 15

Guideline *G4.1* was implemented in the Industries and Products dashboard. The background color was adjusted to a less saturated tone to make the numbers stand out more prominently. A light gray background was also used in the revenue-over-time graph to highlight the COVID-19 period.[22]

For the top 5 vs bottom industries graph, the colour step was reduced from 5 to 3 to make value changes clearer, in alignment with *Guideline G4.7*.[23]

We limited each dashboard to 3-4 graphs or visualisations to maintain clarity and avoid visual clutter. This approach follows *Guideline G5.1*, ensuring an average viewing angle of about 5 degrees, as recommended for readability.[24]

The information button was styled in blue with a white background to make it distinct from other dashboard elements, as were the navigation buttons for switching between dashboards and filters. This design choice aligns with *Guideline G5.3* for visual differentiation. [25][26]

4. Actionable Insights:

Results and key findings:

Central questions:

From the annual statement comments by the PostNord CEO, we deduced clear informational priorities, of financial viability, sustainability and parcel market share. Our central questions are, therefore as follows:

Dashboard – Freight:

1. How is the business operations divided between the countries (Norway, Sweden, and Denmark)?
2. What routes are the most important to the company?
3. Which are the biggest routes (by product) for the company, that are not using Innight, and could we be better advocates for innight purchases on these routes, to be more ESG compliant?

Industries and products -dashboard 2:

1. What are the core industries that PostNord serves?
2. What are the top Sub-industries that PostNord serves?
3. What products are the most often used by industry level 1 and industry level 2?

[22] Ware, C. (2012) *Information visualization*. 3rd edn. Morgan Kaufmann, p. 108

[23] Ware, C. (2012) *Information visualization*. 3rd edn. Morgan Kaufmann, p. 117

[24] Ware, C. (2012) *Information visualization*. 3rd edn. Morgan Kaufmann, p. 141

[25] Ware, C. (2012) *Information visualization*. 3rd edn. Morgan Kaufmann, p. 149.

[26] Kaufman, M. (2004) *Information visualization: Perception for design*. San Francisco, CA, California: Morgan Kaufman.

Revenue over time – dashboard 3

1. How has the average and total revenue per country been month by month?
2. What does the difference in average and total revenue over months per country imply?
3. What was the effect of covid-10 on parcels delivered?
4. What was the growth rate of revenue by country in the latest full year?
5. What was the effects of higher handling costs implemented in march 2018?

Primary question -all dashboards:

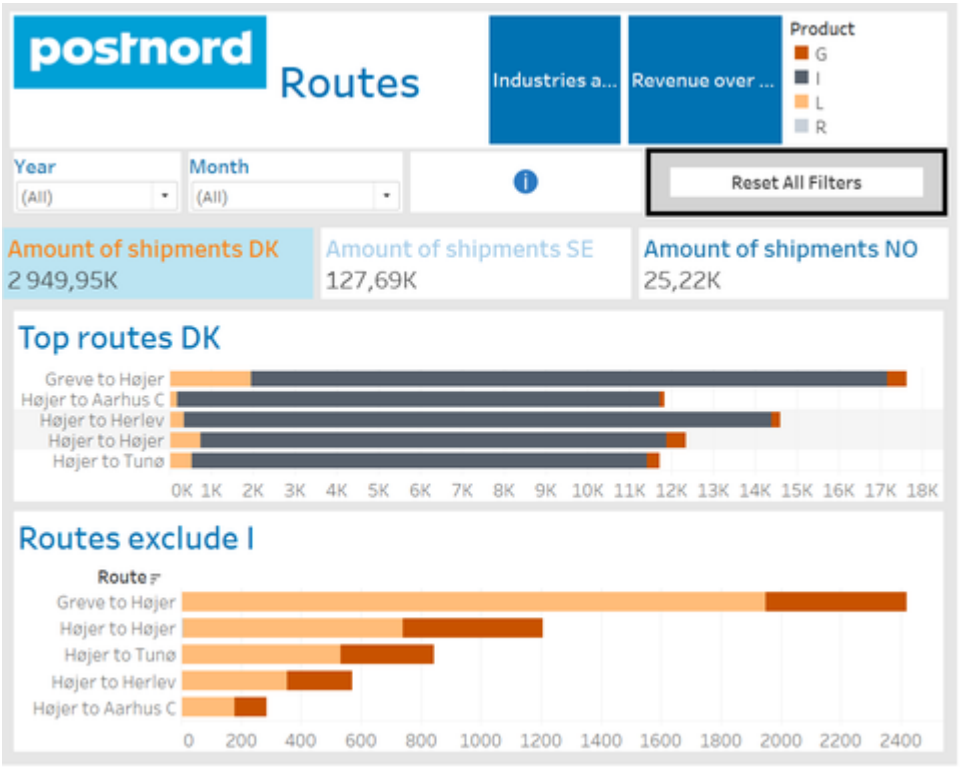
If PostNord were to increase its operations, which market, product, and sector would best fulfil its corporate goals of parcel dominance, sustainability, and financial growth?

Final dashboards & actionable insights

1. **Actionable insight - Routes DB (FP):**

For this dashboard, we observe the top routes taken (from A to B in latitude and longitude), by country and then by product. The overview graph on top is the stacked bar graph, showing how Denmark is the country where most routes occur. Naturally, the top routes in Denmark are also equal to the top overall routes, which is the second graph. Additionally, the second graph shows the product distribution by route. TWe can see that the top product is innight, which is good for PostNord's goal of becoming more eco-friendly. The final graph, is a chart of top routes
excluding the Innight product type. This drill-down graph allows us to discover which routes have the most potential to be converted to the insight product type. If NordPost desires to have a cleaner footprint, they could contact recurring these top 5 rutes and suggest using the Innight option.

Key insight 1: This graph's key insight is that the top routes primarily use the Innight service and are in Denmark.



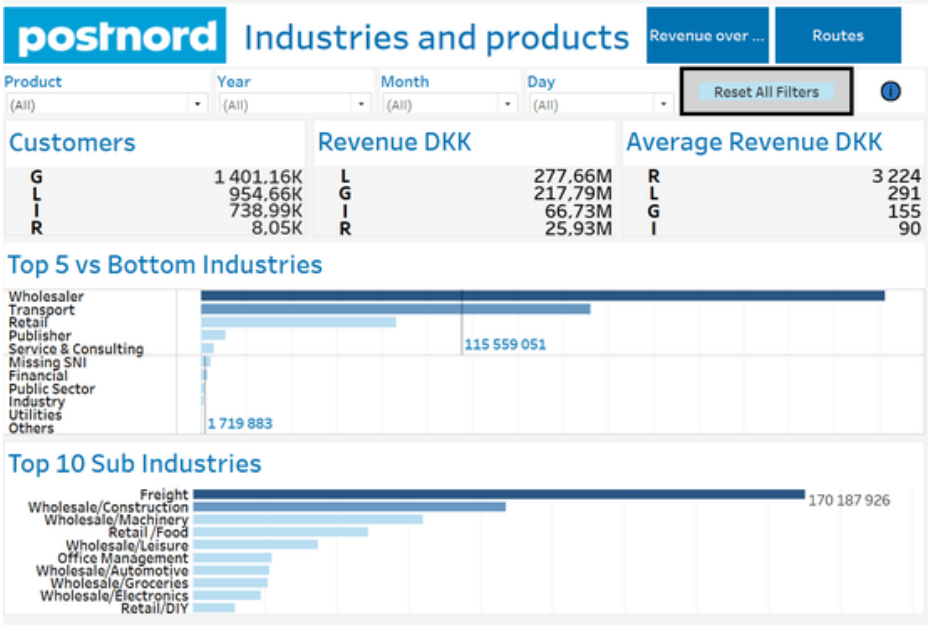
2. Actionable insight - Industry and Products DB (PI):

From the PI DB, we observe some key KPIs regarding the customer base, such as total revenue, average revenue and number of customers per product type. Here we see that Innight has very low revenue and average revenue. Moreover, Innight seems to meb only used in Wholesale, transport and alittle bit public sector. From this we see that there is a massive amount of industries that have not even tried Innight over these years, which could imply a lack of concerted effort by the sales department to push/ market this service. Conversely, roadfreight has less than 8 thousand customers, yet it has the larges revenue per customer. This would imply that if we were purely interested in financial gain, the focus should be on selling roadfreight services. The biggest industry is Transport, (and sub industry Freight). Pallets and Groupage have more similar customer numbers and revenue numbers -with both earning the business more than DKK 480 mn in aggregate between january of 2017 and september 2019. Top industries for Pallets, are freight, construction and leisure. Top industries of Groupage are freight, construction, and food. The described categories show that the common denominator is the freight subclass. After brief research, we found that freight is a catch-all category for delivering large quantities of any type of good, from merchandise to food. Given this restriction, we have met a dead end, industrially speaking. Keep in mind that road freight is a type of product that has a lower number of data points (n) than the other product types, and has gained massively from the COVID-19 crisis demand increase. Therefore, an intellectually honest conclusion would point to pallets as the better product to push given pre-covid circumstances.

In short PostNord, is primarily moving, transport and freight goods in addition to wholesale. Additionally, from the categories where freight, is the highest there is a higher revenue per customer. Therefore, PostNord makes the most from large scale movement of good, implying that they should double down on B2B, wholesale, large deliveries, where there is more to be gained.

Key insight 2: The second key insight is that Innight, although very popular, generates little absolute and relative revenue, while Road Freight has the highest relative revenue and pallets the highest total revenue.

Key Insight 3: Third, the main industries are transport and Freight, which implies that B2B customers are best for the business and should be the focus for future expansion.



3. **Actionable insight - Revenue over time DB (RT):**

First, for the revenue over time dashboard, we ended up with a more temporally focused dashboard. In terms of insights, we see that the sum of revenues by country(RT.a), highlights how Denmark has the higher revenue. However, if we see the average revenue, per county (RT.b), we see that Norway and Sweden have higher revenues. Given the assumption that the profit margin is higher, then the natural
insight is that PostNord is leaving money on the table and should seek to expand their customer base in Norway and Sweden.

Second, the stipulated lines on the graph indicate events impacting the freight industry. There are 4 such events:

PostNord increased costs due to competition law compliance and invested in training operational and client-facing personnel to keep compliance and legal knowledge high internally. This caused higher costs, which were passed on to the customers. [27] We found no clear, actionable insight.
In April 2018, PostNord created a handling fee for parcels coming from outside of Europe. This was needed to cover the admin-related costs incurred from dealing with the packages and VAT. [28]

[27] PostNord Annual statement 2017 - Follow-up of compliance section, page 34
[28] PostNord Website - part of "Global competition" - the 6th section, not including the header.

In 2018 there was an urban fee levied in Oslo, Norway to decrease the amount of cars in the city center. This naturally impacted the delivery services to the capital of Norway and, therefore PostNord. [29]
COVID-19 started in late February 2020 in the Nordic countries. These countries had different lockdown policies, as Sweden allowed people to retain their right to free movement. However, the pandemic impacted the demand for packages.
In the instance of increased handling fees, the following period has a slight increase in revenue. Moreover, the revenue and packages from the start of the COVID era saw a dramatic spike. For parcels sent from February of 2020 to May of the same year, there was more than a +50% increase in packages sent. This sentiment of growth in both parcel amount and revenue total, as the start of the year, showed great growth. Interestingly, the average revenue only increased after April of 2019. This implies that the average revenue has some lagging effect from positive exogenous demand shocks. In short, top-down compliance has affected revenues very little, but demand shocks greatly impact the business.

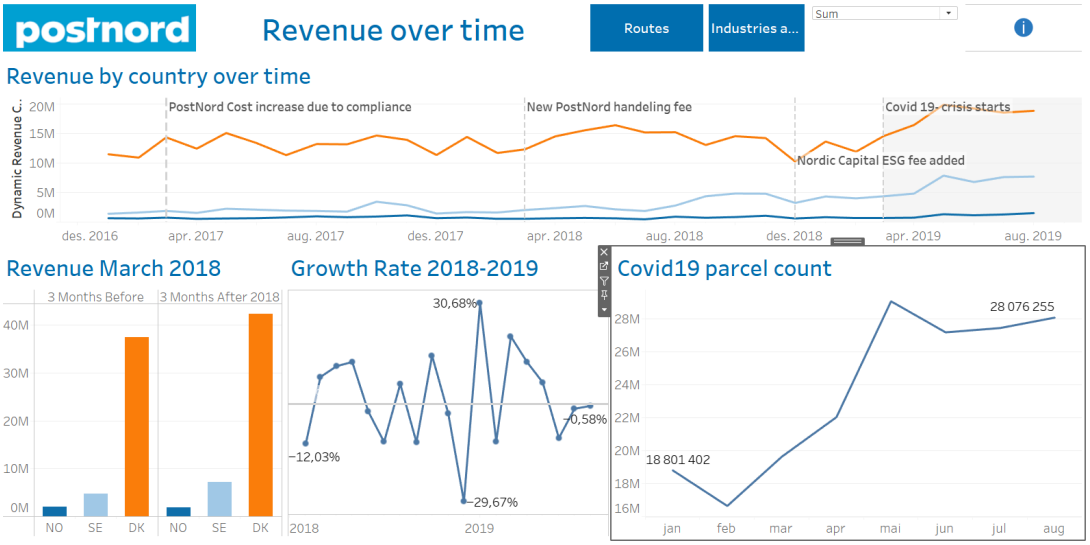
Third, a key insight from the time-series graph is that the long-term growth of YOY seems quite low, which implies that this is a defensive business that is not massively affected by the business cycle. One might point out that this seems to break with the graphical representation, as the covid era saw massive fiscal stimuli from centralised government, however, we denote that the economic growth between 2016 and 2019 was very high, and this does not seem to have disproportionally affected the total revenues. The influx of spending is due to the lockdown and the fact that consumers cannot be outside to the same degree as before; instead, they can order goods online.

Fourth, the key insight from the 3-month before/after foreign handling fee institution bar graph shows very slight movement in both total revenue and average revenue. The total revenue was X million before and X million after - which shows some decline in the customer base's low margin high, high quantity component. Conversely, the higher revenue customers show no discernable movements. The implication is that this handling fee only impacted people with very high price sensitivity whilst not affecting the blue chip clients.

Key insight 4: The fourth key insight is that compliance-related costs have not impacted sales nearly as much as exogenous demand shocks to the economy. Moreover, this is primarily a demand-side business, where deliveries and mailing are commodities with little price discovery.

Key insight 5: The fifth key insight is that total revenue by country is highest in Denmark, but lowest in Denmark for average revenue, implying that the Danish market is primarily high quantity, low revenue customers, whilst the inverse is true for Sweden and Norway.

[29] Nordic Council of Ministers report



4. **Actionable insight -Final question:**
   Finally, concern ourselves with the core question of this report, which is:
   " If PostNord was to increase their operations, which market, product and sector would be the best to fulfil their corporate goals of parcel dominance, sustainability and financial growth?"

   From the advanced visual analytics we have an answer. From the total revenue per country, we see that Denmark is the leader. However, when we switch to average revenue, we find the leaders to be Norway and Sweden. Moreover, from the parcels sent, graphs in the Freight and products DB we see that the vast majority of parcels are sent to Denmark, from Denmark. From these two insights, we determine that PostNord should increase operations in Norway and Sweden, gaining more revenue, and capturing more of the Scandinavian market. Because of PostNord's desire to be eco friendly, they would choose to emphasise their Innight services. Moreover, the high revenue clients demand mostly freight and wholesale and product type R. From a financial perspective the best option would be to expand these services and increase the revenue. Geographically the Consigner locations are all in Denmark, and if they are using ships to cross the Skagerakk, or Kattegatt, they have to make a long way to Copenhagen, across the bridge, and up through the west coast of Sweden to start their travels across Norway. Therefore we find that the truly strategic business move would be to set up distribution centres (Consignor locations) in a geographically advantaged place such as on the border between norway and sweden towards Fredrikstad or Svinesund. Given incoming goods are flown in, we would see shorter trips, more eco friendly opportunities, and potentially higher revenue from accessing the Norwegian/ Swedish market in a new way.
   Our dashboards improve decision making because they highlight the geographic leaders, the industries providing highest revenue, and the types of exogenous shocks that should worry decision makers.

Further research:

Further research should focus on the collection of new variables such as shipment truck (which parcels and packages are in which trucks), types of fuel (diesel, biodiesel, gasoline, electric), type of vehicle (size of truck), energy needed per km, fuel price at that time (for all kinds of fuel), routes taken to travel (could be a series of latitude/longitude combinations that would make a more accurate route

description), and many more. We find that the aim of net-zero emissions to be excellent, yet in our current state, we do not have to capacity to faithfully calculate those emissions precisely without a long list of assumptions. Moreover, research should be done to determine the feasibility of new consignor centres, and determine the complete costs for the company, to understand average profit margins, as it is more relevant than revenue alone.

5. Conclusion:

PostNord is an essential business in the Nordic logistics industry due to its high shipment count in the Nordics. This analysis displays the company's dominance in the Danish market and its opportunity to use current infrastructure to increase operations in Norway/Sweden and revenue while achieving its net-zero emissions goal.

In this paper, we have taken a complete analysis approach. We started by reviewing company information and selecting a method of analysis and an architecture. We cleaned our data using ETL principles. Then, we explained our assumptions and limitations, heuristics, and iterations of our three dashboards. Afterwards, we explained the actionable insights of our final dashboards and the primary question of our report.

The first key insights drawn from the dashboards show Denmark as the only consignor country with the highest total revenue but low average revenue per customer. This implies that Danish customers are, on average, high-volume, low-margin businesses. Norway and Sweden offer opportunities for expansion with higher average revenues per customer, indicating the potential for more profitable engagements in these markets. The Insight product generates low revenue per customer while having a green upside, making it a tradeoff between revenues and social responsibility.

From additional insights, the compliance, red tape, and regulation applied to the company through various top-down actors have caused little change in revenue. However, the COVID-19 crisis massively grew both parcel count and total revenue. Although the average revenue increased, it did so some months after the COVID crisis first occurred.

The data suggests that PostNord should expand operations in the Norwegian/Swedish market and continue to serve the high-volume freight sub-industry, as clients from that industry provide the highest average revenue. Given the long path from Denmark to Sweden (geographically), either across the sea or through Sweden, additional consignor locations strategically placed somewhere in Sweden/ Norway are needed, and perhaps even expansion of parcel lockers for more dynamic deliveries.

Future research and data collection are needed to determine profits, route optimisation, emissions, and knowledge regarding potential expansion. With these actions, PostNord is in an excellent position to achieve its goals of being the leader of Nordic parcel transportation, net-zero emissions, and strong financial growth.

6. References:

< Harvard Style>

1. Bera, P. (2016) ‘How colors in business dashboards affect users’ decision making’, Communications of the ACM, 59(4), pp. 50–57. doi:10.1145/2818993.

2. Fisher, D. et al. (2012) ‘Interactions with big data analytics’, Interactions, 19(3), pp. 50–59. doi:10.1145/2168931.2168943.

3. Knaflic, C.N. (2015) Storytelling with data: A Data Visualization Guide for Business professiona. John Wiley & Sons.

4. Shneiderman, B. (1996) ‘The eyes have it: A task by data type taxonomy for information visualizations’, Proceedings 1996 IEEE Symposium on Visual Languages [Preprint]. doi:10.1109/vl.1996.545307.

5. Toreini, P. et al. (2022) ‘Designing attentive information dashboards’, Journal of the Association for Information Systems, 22(2), pp. 521–552. doi:10.17705/1jais.00732.

6. Ware, C. (2012) Information visualization, 3rd Edition. Morgan Kaufmann.

7. Wickham, H. (2014) ‘Tidy Data’, Journal of Statistical Software, 59(10). doi:10.18637/jss.v059.i10.

8. Zhang, Y. et al. (2022) ‘Visualization design practices in a crisis: Behind the scenes with covid-19 dashboard creators’, IEEE Transactions on Visualization and Computer Graphics, pp. 1037-1047. doi:10.1109/tvcg.2022.3209493.

9. https://www.postnord.no/en/about-us

10. https://aws.amazon.com/what-is/etl/

11. https://www.sciencedirect.com/topics/computer-science/central-repository

12. https://www.postnord.com/globalassets/group/documents/investor-relations/financial-reporting/annual-and-sustainability-reports/2017_en_postnord.pdf

13. https://www.postnord.com/insights/e-commerce-in-the-nordics-six-month-report-2018

14. https://pub.norden.org/temanord2023-520/preface.html

15. Nielsen, J. (2024) 10 usability heuristics for user interface design, Nielsen Norman Group. Available at: https://www.nngroup.com/articles/ten-usability-heuristics/ (Accessed: 15 November 2024

7. Appendix:

Appendix 1: Product data - variable descriptions

Products data (R,L, I, G)		
Column names	**Description**	**Data type**
Customer_ID	It is unique ID identifying the corporate customer paying for the service.	V_W
ConsignorCountryCode	Country from where the shipment originates (gets picked up). E.g. DK for Denmark, SE for Sweden, NO for Norway.	V_W
Consignor Location Nr	A particular postal code from where the shipment originates (gets picked up). E.g. 2000 would mean that the shipment is get picked from Frederiskberg.	V_W
ConsigneeCountryCode	Country from where the shipment gets finally delivered. E.g. DK for Denmark, SE for Sweden, NO for Norway.	V_W
Consignee Location Nr	A particular postal code from where the shipment gets finally delivered. E.g. 1363 in Norway would mean that the shipment is get delivered in Høvik Norway.	V_W
Product	Short code for Product. E.g. I for Innight, L for Pallet, R for Road Freight and G for Groupage	V_W
Parcel/Pallet type	This is relevant to Product "Pallet" or "L". A "PLL" is a full pallet, HPL is half Pallet, QPL is quarter Pallet and SPL is for shipments with a special shape. If the value is left blank, then assume it to be a PLL. Groupages are a collection of pallets	V_W
Shipment Tracking Number	Unique tracking number for every shipment	V_W
ShipmentDate	Day the Shipment was picked from Consignor Location and scanned using PostNord systems.	DATE
ConsignorParcelCount	Total Number of items in a particular Shipment.	
ConsignorWeight	Weight of the Shipment.	float
Price_Paid	Total Price in DKK the customers pays for the delivery. It is the sum of Freight and all 9 surcharges.	float
Base_Price	Price in DKK offered from the PostNord standard sheet.	float

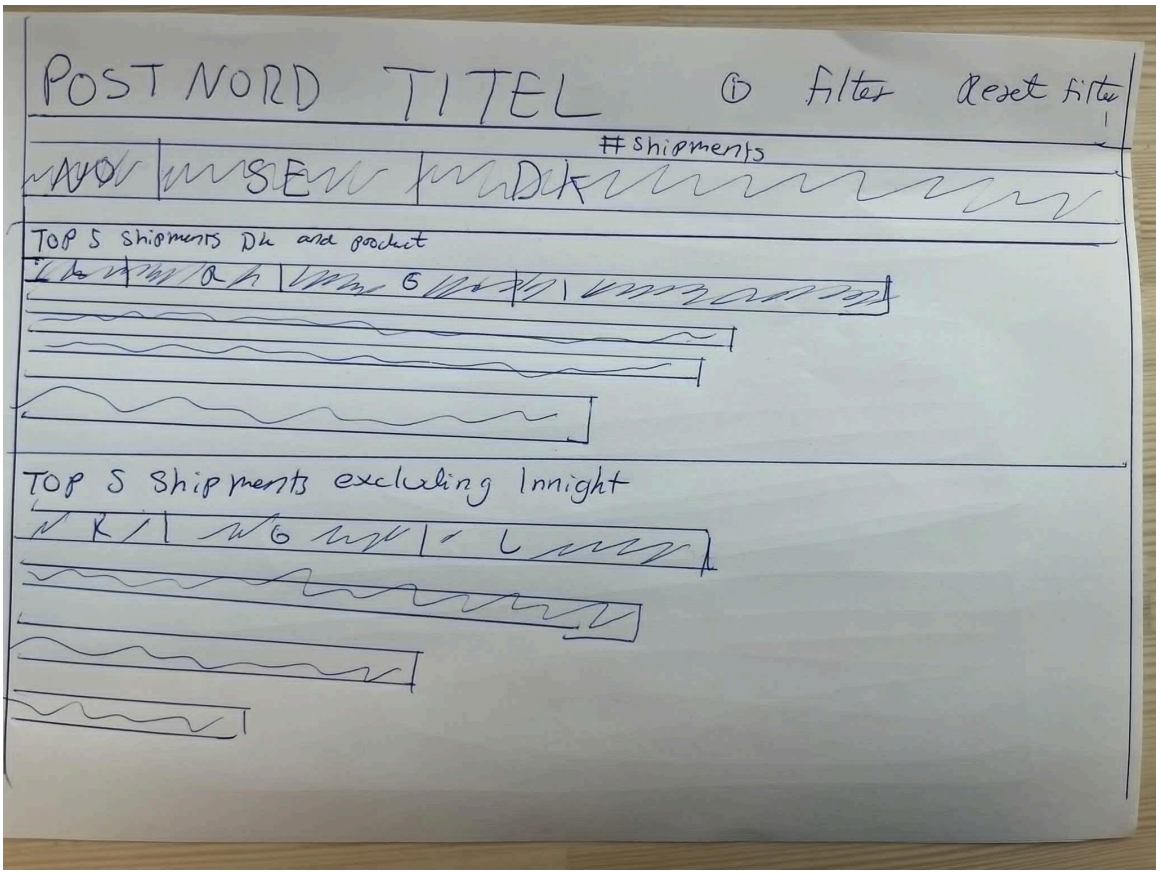
Appendix 2: Shaping data - tidying steps:

Step	Task Description	Action/Details
Step 1: Product Dataset Cleaning		
DDS Groupage	Replace pallet types with N/A	Replace all pallet types with 'N/A'
DDS Groupage	Exclude parcels that weigh over 2500 KGs	2378 rows removed
DDS Innight	Replace pallet types with N/A	Replace all pallet types with 'N/A'
DDS Pallets	Replace N/A values with PLL	Replace all 'N/A' values with 'PLL'
DDS Pallets	Replace PKK values with SPL	Replace all 'PKK' values with 'SPL'
DDS Pallets	Drop rows with pallet overweight	Remove rows based on specific weight criteria
DDS Pallets	Drop rows with PLL and weight > 1000 KGs	Remove rows where 'PLL' > 1000 KGs

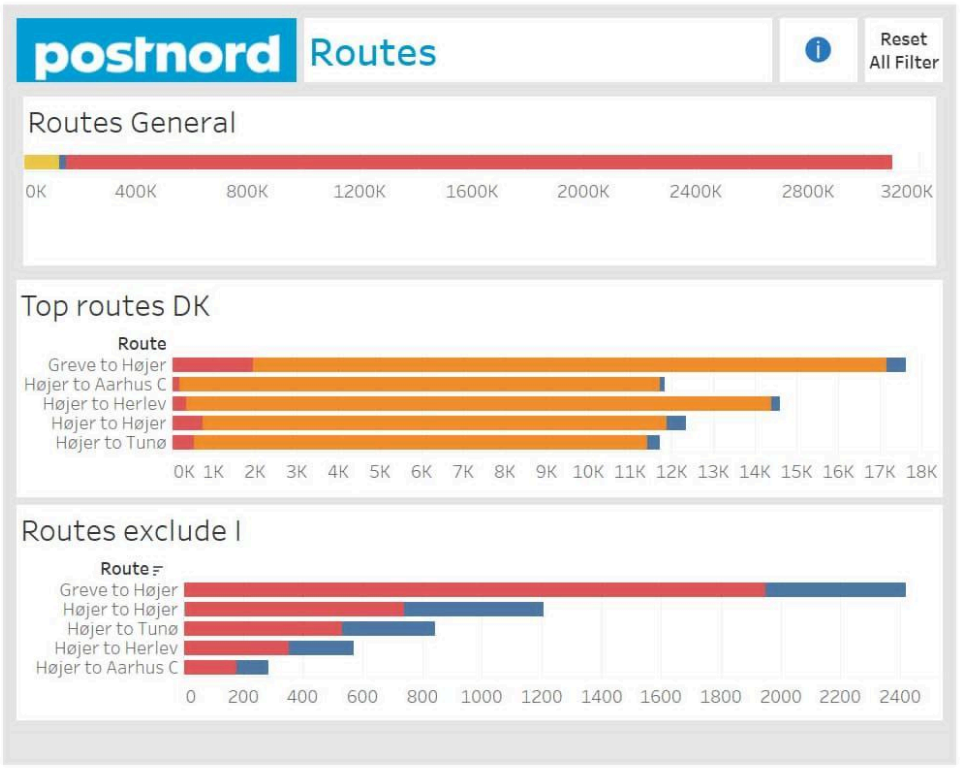
DDS Pallets	Drop rows with HPL and weight > 500 KGs	Remove rows where 'HPL' > 500 KGs
DDS Pallets	Drop rows with QPL and weight > 250 KGs	Remove rows where 'QPL' > 250 KGs
DDS Pallets	Drop rows with SPL and weight > 1000 KGs	Remove rows where 'SPL' > 1000 KGs
DDS Pallets	Total rows removed: 117,634	
DDS Road Freight	Replace pallet types with N/A	Replace all pallet types with 'N/A'
Step 2: Union of Datasets		
Union of Datasets	Combine DDS Pallets, Innight, Groupage, Road Freight	Full join datasets and remove duplicates
Union of Datasets	Check for duplicates	Ensure no duplicate rows
Step 3: Clean Geo_datasets		
SE.txt	Remove unnecessary columns	Keep only required columns
SE.txt	Remove spaces from SE postal codes	Trim spaces in SE postal codes
NO.txt	Remove unnecessary columns	Keep only required columns
DK.txt	Remove unnecessary columns	Keep only required columns
Step 4: Union Geo_datasets		
Union Geo_datasets	Full join (union) SE.txt, DK.txt, and NO.txt	
Step 5: Join Master/Geo Data on Product Dataset		
Join Master Data	Left join Master data to Product dataset	Join on Customer ID
Step 6: Join Geo Data on Product Dataset		
Join Denmark Geo	Left join DK Geo data on Product dataset	Match on Consignor Country Code and Consignor Postal Code

Append Geo Data	Concatenate NO.txt, SE.txt, DK.txt as geo_data	Create combined geo_data dataset
Join Geo Data	Left join geo_data on Product dataset	Match on Consignee Country Code and Consignee Postal Code
Step 7: Filter Out Duplicates		
Remove Duplicate Rows	Remove any duplicate rows	Ensure no duplicates remain
Check Shipment Code Duplicates	Verify unique shipment codes	Ensure no duplicate shipment codes

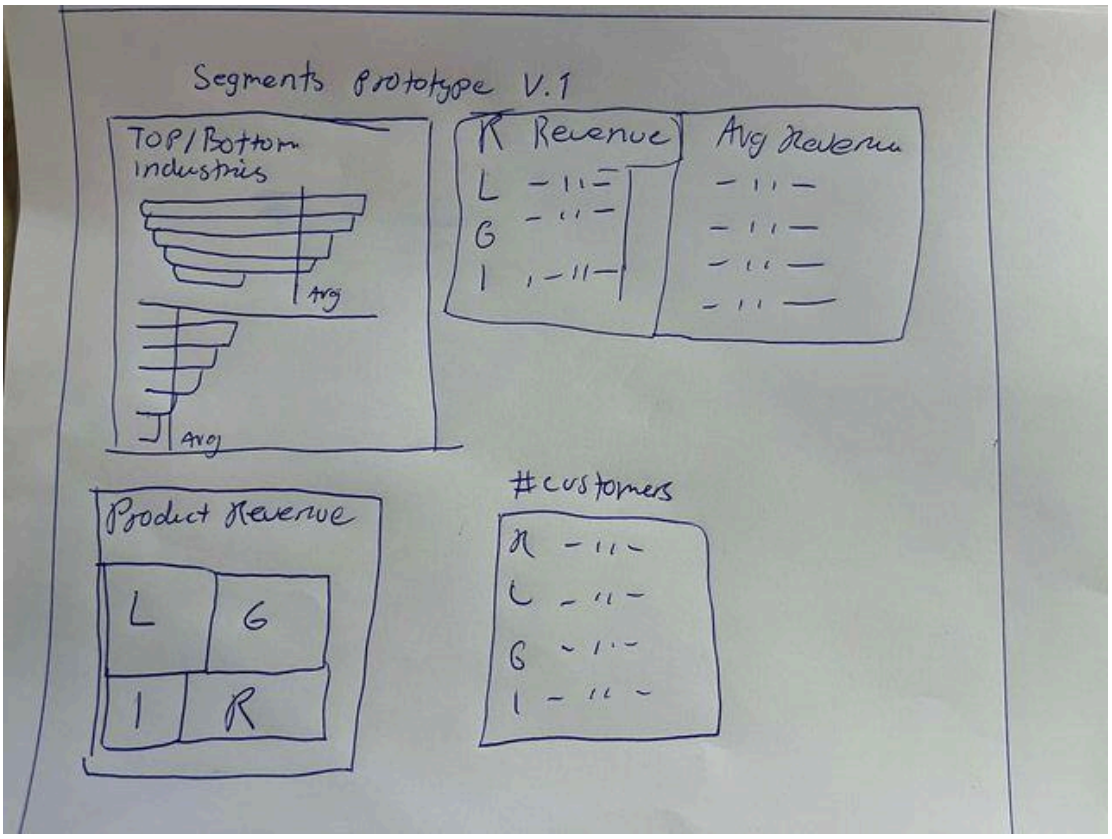
Appendix 3: Freight product– sketch 1



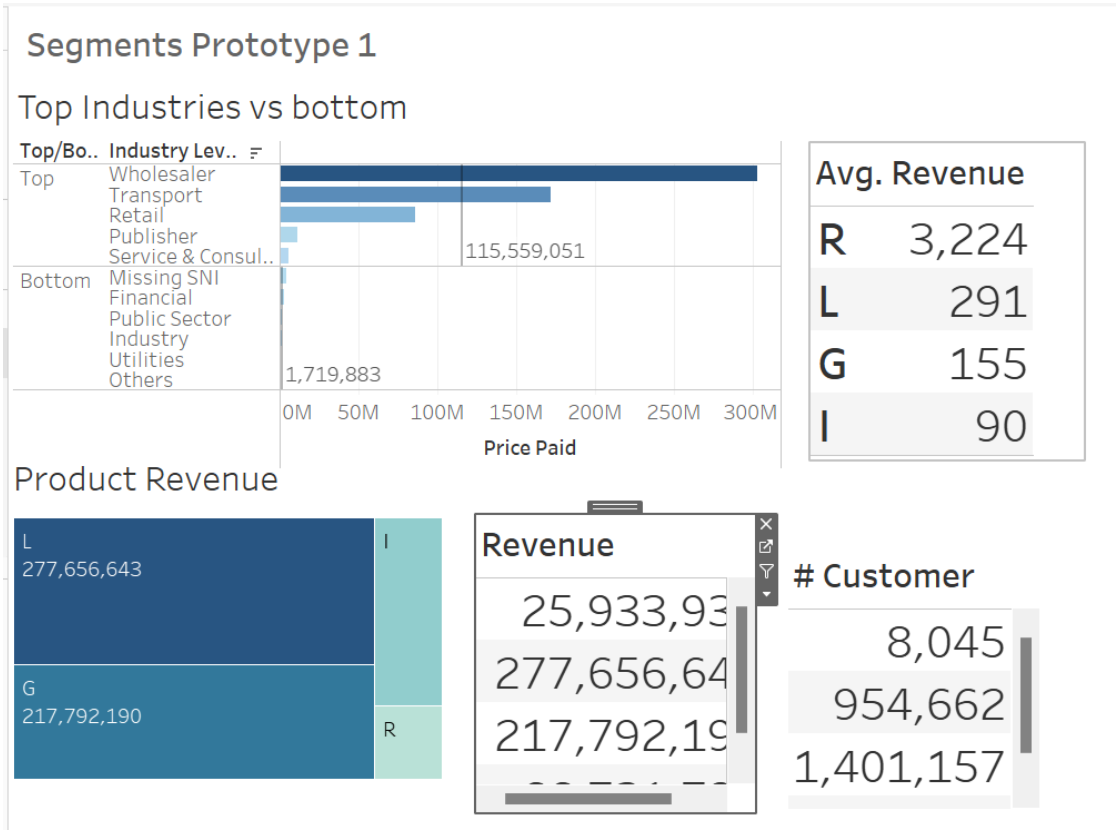
Appendix 4: Freight Product – dashboard 1



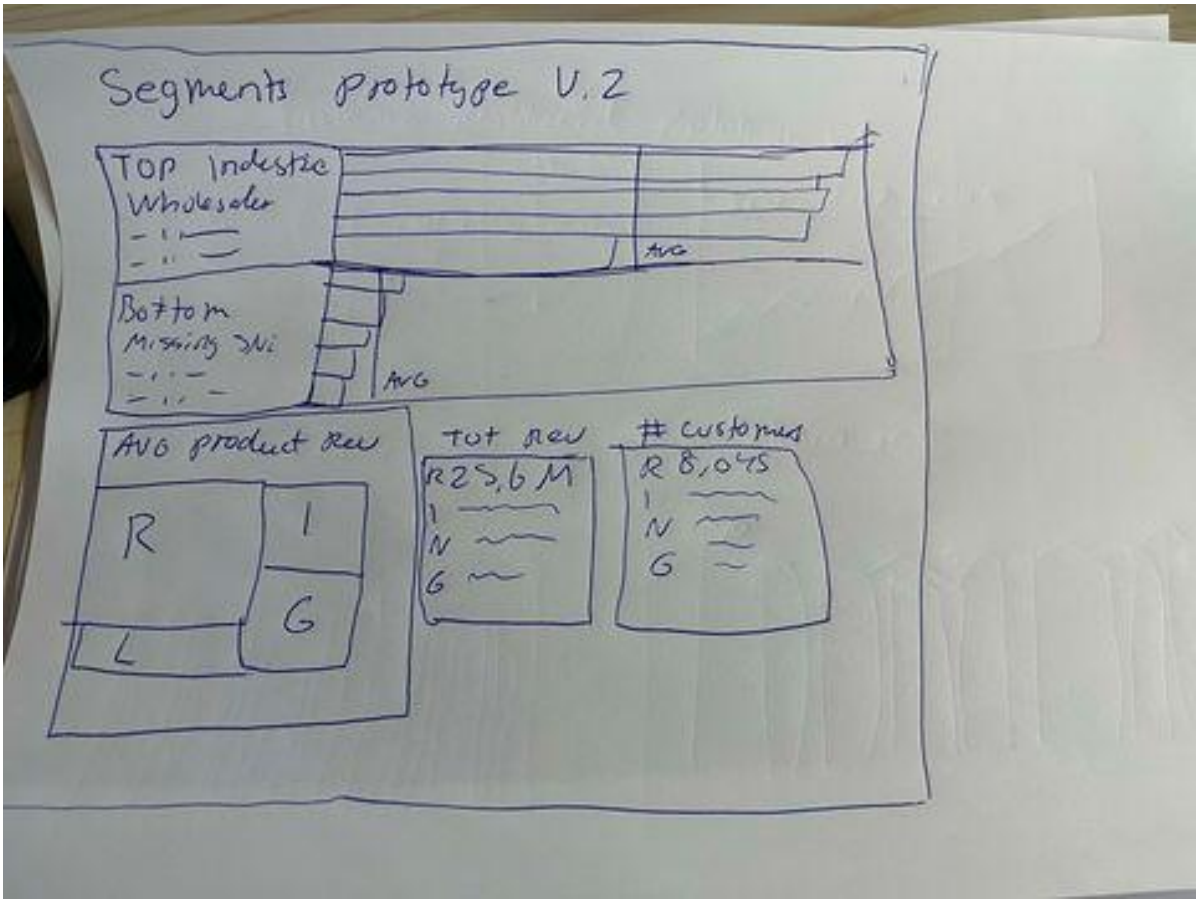
Appendix 5: Industry Product – sketch 1



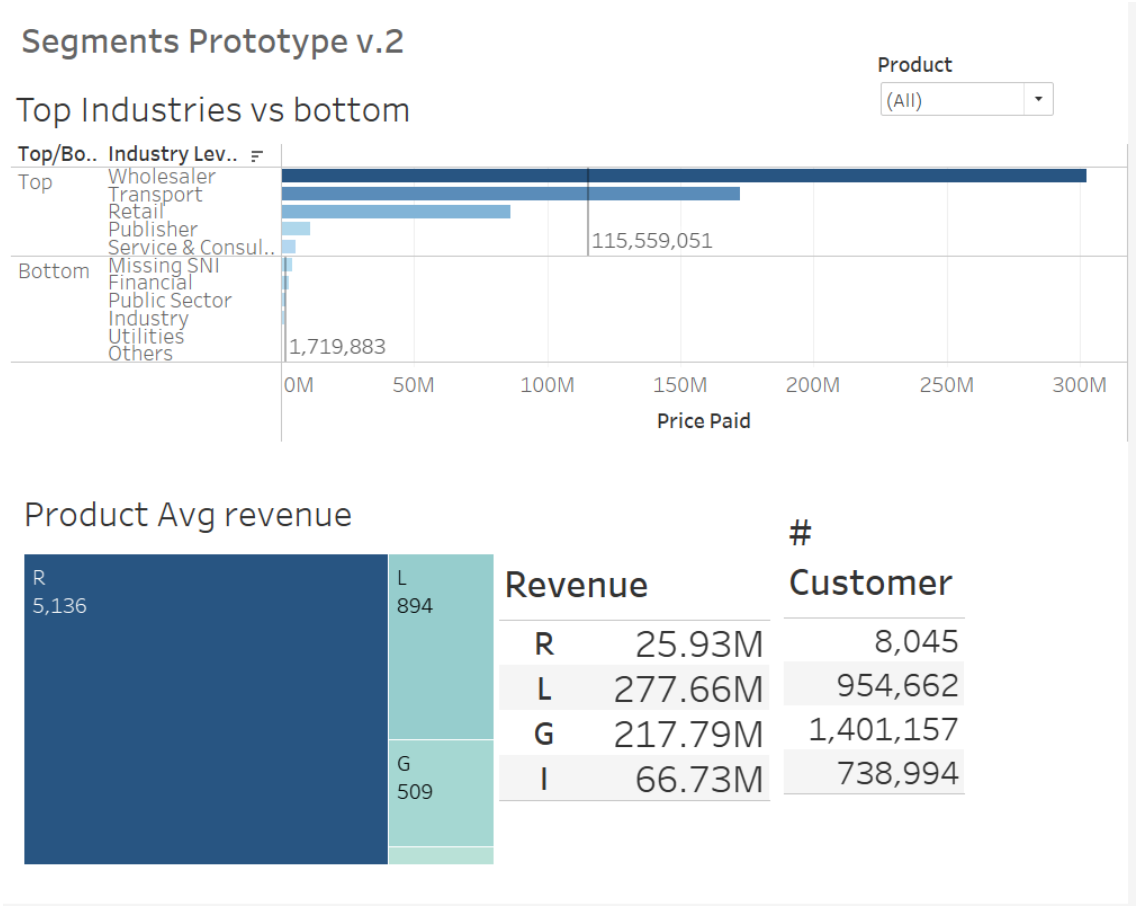
Appendix 6: Industri Product - dashboard 1



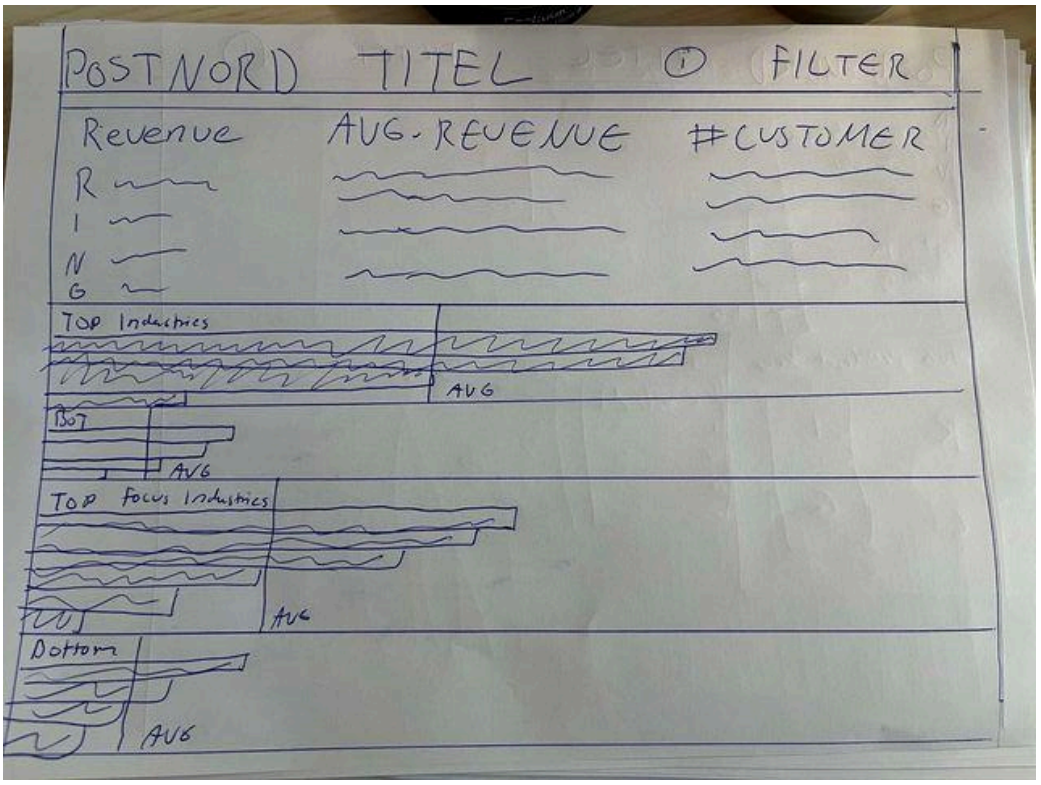
Appendix 7: Industry Product – sketch 2



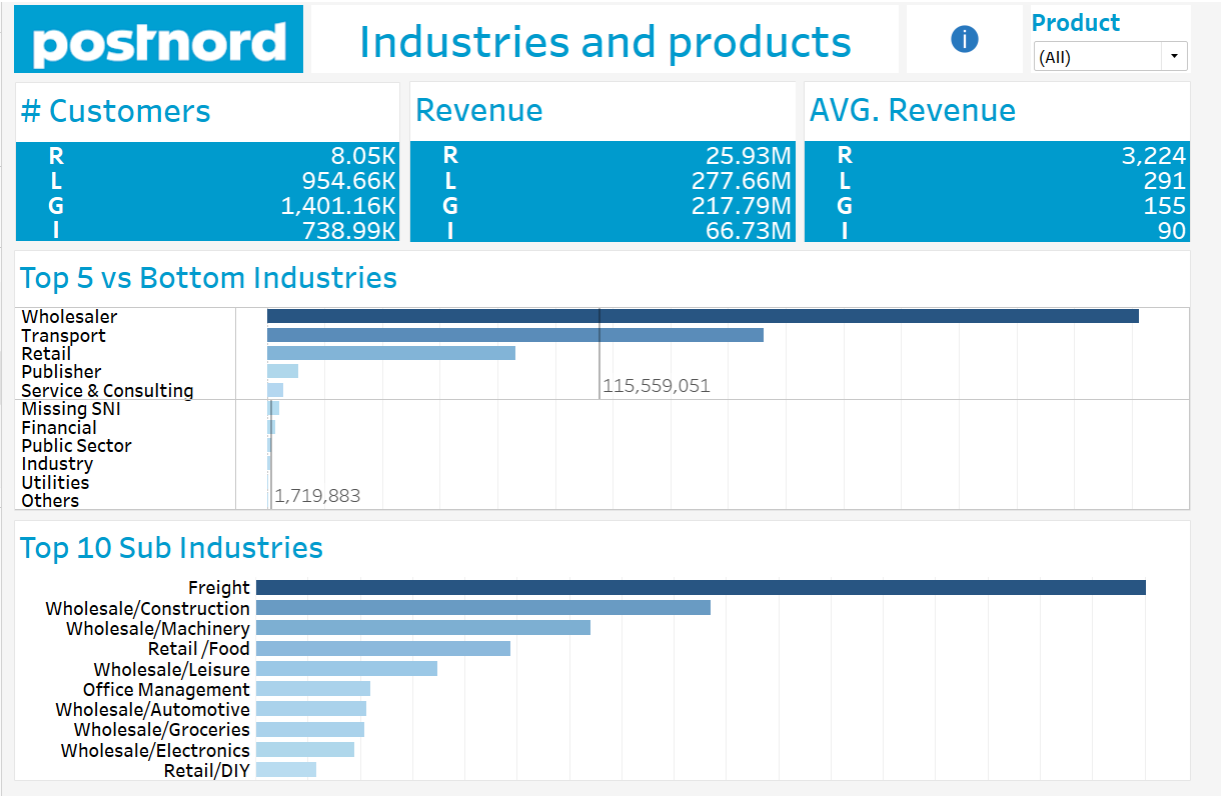
Appendix 8: Industry Product – Dashboard 2



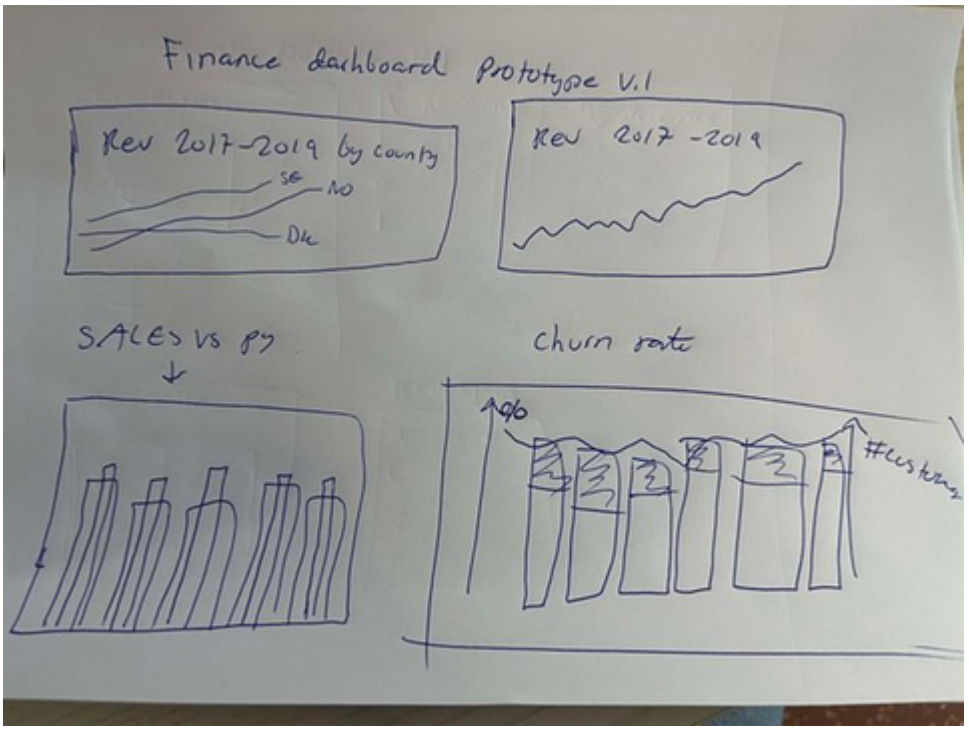
Appendix 9: Industry Product – Sketch 3



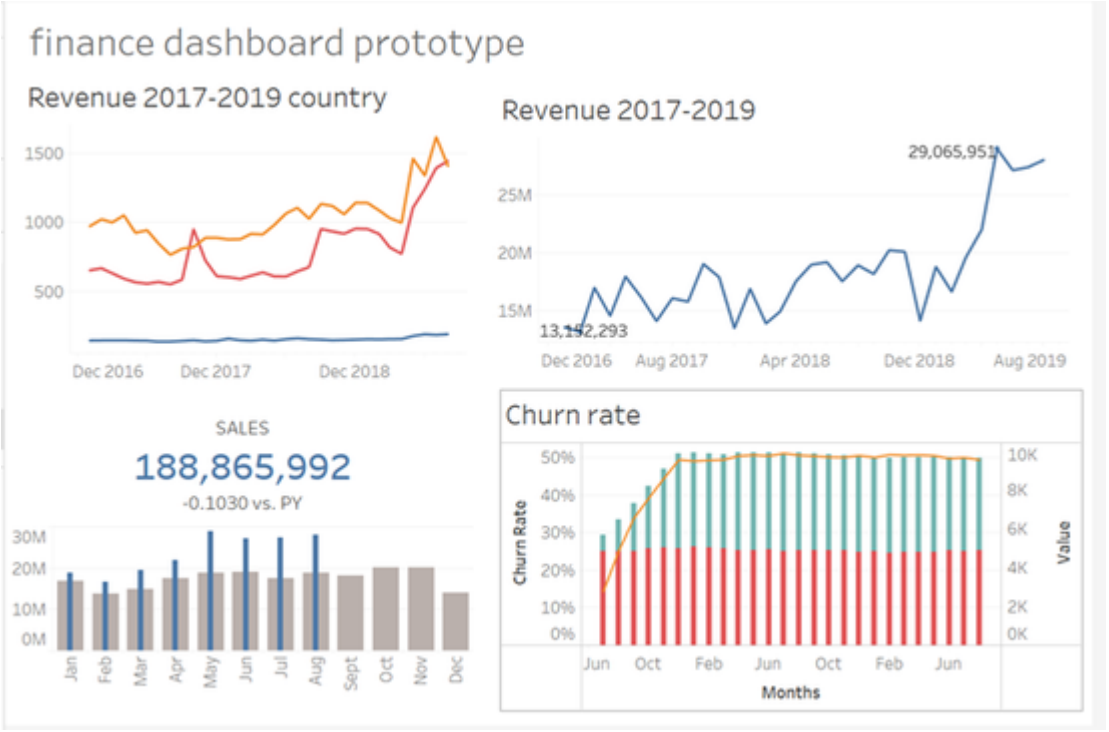
Appendix 10: Industry Product – dashboard 3



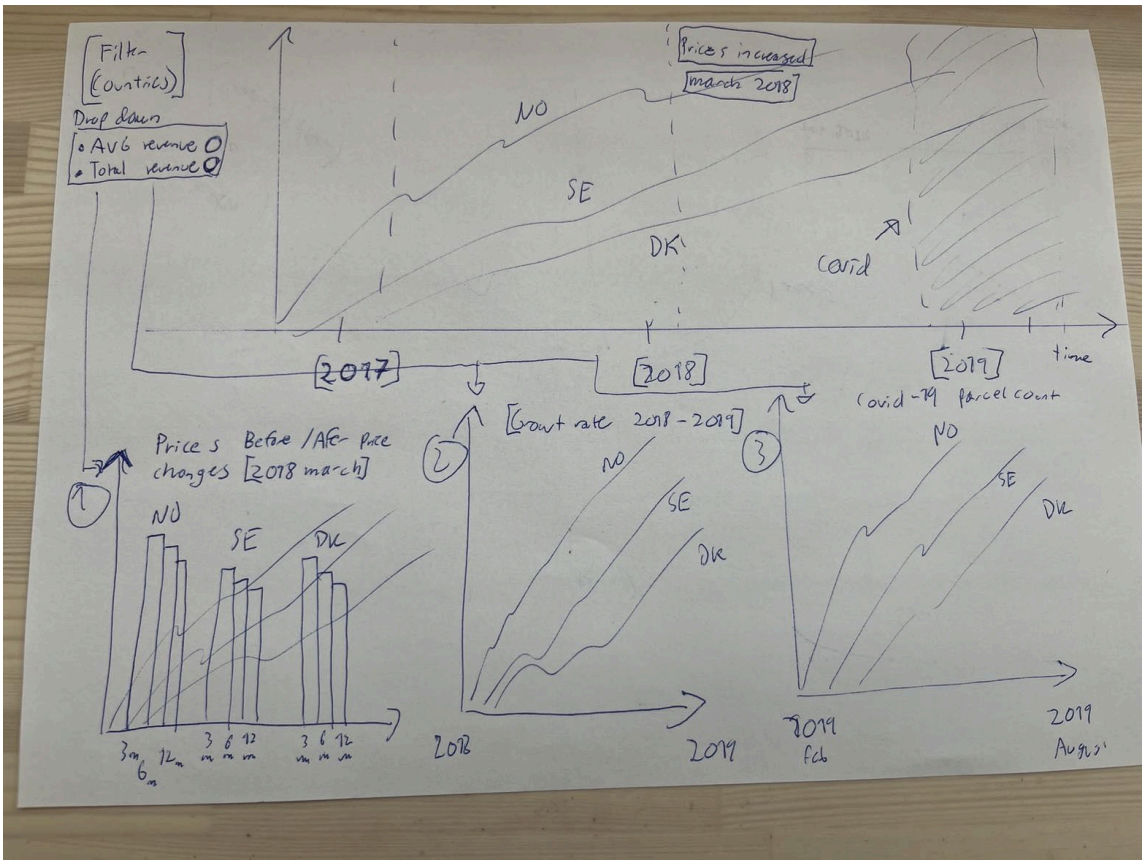
Appendix 11: Revenue over time – sketch 1



Appendix 12: Revenue over time – dashboard 1



Appendix 13: Revenue over time – sketch 2


Filter (countries)
Drop down
• Avg revenue
• Total revenue
Prices increased
march 2018
NO
SE
DK
covid
2017
2018
2019
time
Prices Before/After price changes [2018 march]
1
NO
SE
DK
3m
6m
12m
3 m
6 m
12 m
3 m
6 m
12 m
[Count rate 2018 - 2019]
2
NO
SE
DK
2018
2019
covid-19 parcel count
3
NO
SE
DK
2019 feb
2019 August