









Final Examination: PGR110 Visual Analytics



Part One: Dataset Description


Dataset 1:


Sales and Shipment data for a company operating in 6 European countries. This particular dataset has 19 columns and 4,100 rows of data. The following are a few notable characteristics for each column in the dataset.

1. **Row ID:**  *Numerical value*. There are a few of rows missing from the dataset. The maximum number in the Row ID is 10,000; whereas the actual number of completed data rows is only 4,100. There are jumps in Row IDs. This is indicative that the dataset has been “trimmed” down from a much larger dataset.
2. **Order ID:**  *Text value*. This is NOT a unique value, that is one Order can appear in multiple records (rows) depending upon the number of different products in the particular order.
3. **Order Date:**  *Date value*. The date on which the order was placed.
4. **Dispatch Date:**  *Date value*. The date that the corresponding products are shipped to the customer. The speed at which this process is carried out is based on the delivery modes.
5. **Delivery Mode:**  *Text value*. Contains how the products are to be delivered. There are 4 alternatives; 1) Same day delivery, 2) First class, 3) Second class, and 4) Standard class. There is no additional information of the constraints on the service parameter on speed for the different classes, with exception for the “Same Day” delivery.
6. **Customer ID:**  *Text value*. Contains a unique ID for each customer.
7. **Customer Name:**  *Text value*. Contains the name of each customer.
8. **Segment:**  *Text value*. This is a classification of customer segments. This company categorizes its customer into 3 groups with the number of the customers in each group shown.

Customer by Segment

Consumer	Corporate	Home Office
383	222	137
9. **City:**  *Text value*. However, Tableau recognises that it is  *a geographic value*. This indicates the name of the city where the customer resides. (There are 337 cities without geographical location, we need to make adjustment in Tableau.)

10. State/Province: *Abc Text value.* However, Tableau recognises that it is  a *geographic value*. The name of the State where the customer resides. (There are 33 entries without geographical location and we need to make adjustment in Tableau.)

11. Country/Region: *Abc Text value.* However, Tableau recognises that it is  a *geographic value*. The name of the Country/Region where the customer resides.



12. Region: *Abc Text value.* Geographical region that the country is located on the globe. Germany is in “Central (Europe)” and the rest of the countries in the dataset (Denmark, Norway, Sweden, Finland and United Kingdom) are in the “Northern (Europe).”

13. Product ID: *Abc Text value.* Contains unique ID for each product.

14. Category: *Abc Text value.* A major grouping of products. There are 3 groups: Office Supplies, Technology, and Furniture.

15. Sub-category: *Abc Text value.* The sub-group of products. Each of the 3 major groups above comprises of multiple products as shown in the figure.



16. Product Name: *Abc Text value.* Contains the name of product.


17. Sales: *# Numerical value.* The total price paid in Euros.

18. Quantity: *# Numerical value.* The number of unit(s) of the product sold.

19. Profit: *# Numerical value.* Profit or loss generated from the transaction in the row.

Dataset 2:

This dataset contains information of 140 countries on 8 indexes over the period of 2015-2022. There are 11 columns and 1,043 rows. The following are a few notable characteristics for each column in the dataset.

1. Country: *Abc Text value.* However, Tableau recognises that it is  a *geographic value*. This contains the country name. There are 1 unknown (Somaliland) in Tableau, we google searched the latitude and longitude, the entered in Tableau.

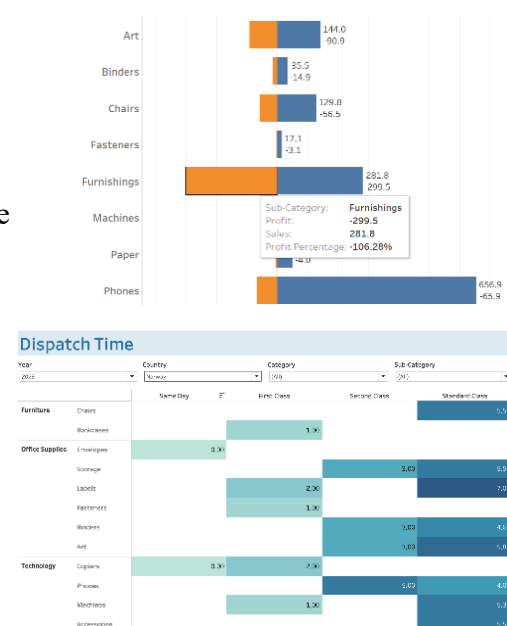
2. Happiness Rank: *# Numerical value.* The ranking of the country based on the happiness score.

3. **Happiness Score:** # Numerical value. The indicator of how happy the population of the country is. The higher the number, the happier the citizens.
4. **Economy:** # Numerical value. A measurement of economic health of a citizen in the country (similar to GDP per capita). The higher the score for a country, the better life expectancy of its populations.
5. **Family:** # Numerical value. Family score indicates the level of social support one can expect to get should one struggle. The availability of family, relatives or friends one can count on to help out. The higher the score, the better social support.
6. **Health (Life Expectancy):** # Numerical value. This is the estimated life expectancy of the population in the country. The higher the score, the longer people live.
7. **Freedom:** # Numerical value. This refers to the ability to make choices in one's life. This measures the level of satisfaction of the level of freedom over the life of the population. The higher the score, the higher the level of freedom the citizens enjoy, have over their own lives.
8. **Trust (Government Corruption):** # Numerical value. This measures the level of trust in the society. The higher the score, the lower level of perceived corruption in the society, particularly in the government and business.
9. **Generosity:** # Numerical value. This index measures the level of generosity of the people in a country. How generous citizens are.
10. **Year:** 📅 Date value. Refers to the year that data was collected.
11. **Region:** 🌐 Text value. Refers to geographical location on the globe of the country.

Data Transformation

We performed two data transformations in our analysis. The first one is creating one variable called “Profit Percentage”, ($Profit / Sales$). This new variable was used in calculating profitability of products and aggregate to the profitability of the whole category, displaying the information in Tooltips.

The second data transformation was done to create the time it takes for the stores to ship product to customers, $FLOAT([Dispatch Date]-[Order Date])$. This new variable was used in creating the following visualization.

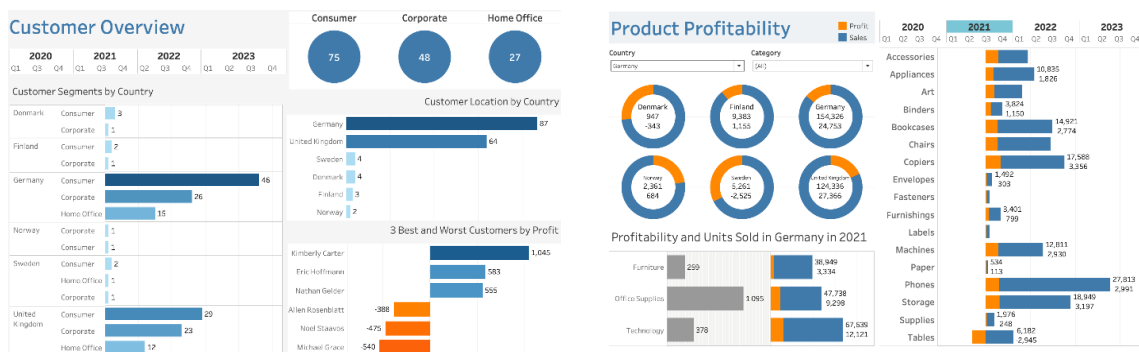


Part Two: Dashboard Design and Development:

Dataset 1: We thought that there would be a few dashboards for this dataset. This is primarily governed by the Sales Manager being the intended user. There should be one dashboard dedicated to customer details, one for product details, a third one for profitability, and one last one for efficiency of dispatching products. In addition to the aforementioned dashboards, there will be a couple of other self-contained visualizations as well.

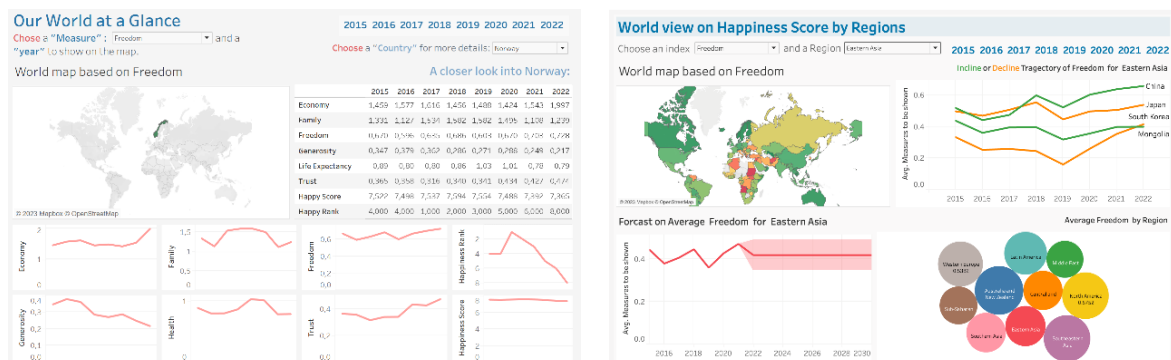
Dataset 2: There should be one main visualization which is the world map showing the 8 indexes (happiness rank is logically reverse; The higher the ranking, the worse off the country is). In order to deal with the proper ranking and its order, we decided to create a separate self-contained visualization for it.

Having explored and understood the two datasets, we developed concepts for what the end users need from our visualizations. These are some of the major ideas for each dataset.



We have opted to use the bar graphs in the majorities of our visualisation, as shown in the screenshots above. This because, they are not only very intuitive for users when comparing data, but also effective to convey a message and insights. Bar charts and its variation (Stacked bar charts) are best used to display and compare categorical data (E.g.: 3 “Segments” of customers in Dataset 1 and “Region” of countries in Dataset 2). This kind of graph emphasises the relationship between different groups. It is particular effective in showing different data points. Additionally, it is one of the most commonly used graph type and, therefore, it is easier for users to interpret and understand.

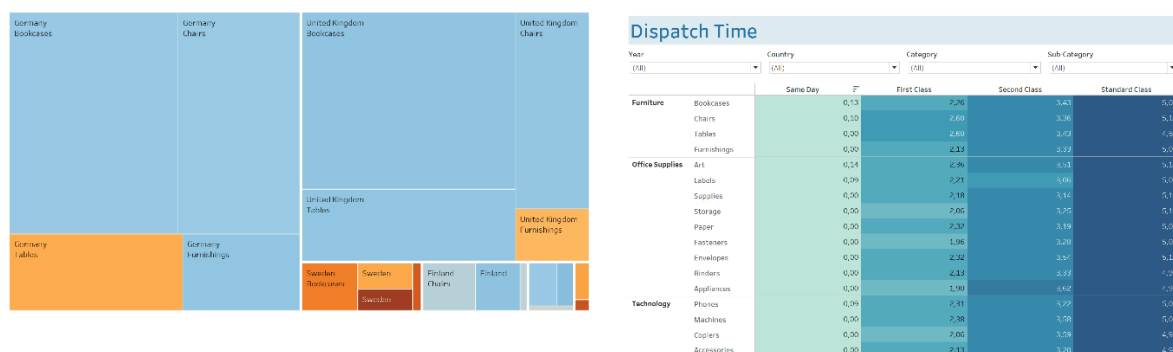
Pie Charts, or in this particular case Doughnut Charts, were used to show information where the proportion to the whole is the main message. In the above screenshot, we used Doughnut charts to depict the proportion of Profit to the Sales figures for each of the markets. A pie chart works best when there are a few (preferably not more than 5) distinct categories, only 2 in our case.



We chose to use Line charts to show a series of data over time; Sales and Profit data in Dataset 1, and all of the 8 indexes in Dataset 2. The above screenshots illustrate how we used the line charts in our analysis. The line charts on the left-hand screenshot show changes of the 8 indexes over time. Tableau's forecast feature was also used in the screenshot on the right-hand side to show the possible value of the "Freedom" index for East Asian countries.

Maps have also been used in our visualizations in both datasets, as exemplified in the above screenshot from the visualizations of Dataset 2. These particular visualizations show two distinct behaviours of the Maps in our visualizations. In the screenshot on the left, the map is zoomed in to showing one selected country (Norway), while the one on the right it shows all the countries that we have data on.

The decision to use a simple table in displaying data, as shown in the above screenshot, was made based on its simplicity and its ability to offer the users to extract numerical data out of the dashboard with ease. In addition, users can perform their own analysis, and draw their own conclusions based on the filtered data.



Treemap was used in one of visualizations in Dataset 1 to show proportions (Sales figures) in hierarchical relations (Countries and City, Category and Sub-category of product). In the above screenshot, profitability is overlaid as colouring in the treemap.

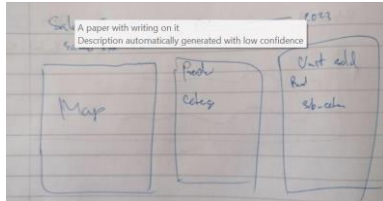
Evolution of our dashboard designs

We started the process of sketching with diverse ideas in mind, regarding how to put the visualizations together and creating informative dashboard. One of the key decisions for the group is whether we should construct our dashboards with the aim to answer the exam questions or to do so with the end users in mind. We decided to create our visualizations, dashboards for the Sales Manager of the company.

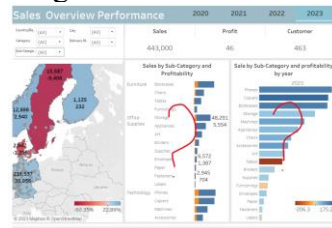
This process helped us to gather all of ideas and put it into paper, it helped us to understand and envision what we wanted. Thereafter, we began to work on creating visualizations on Tableau and further refined the design until we reached the final designs.

Below are selected three dashboard designs beginning its infancy of extremely rough sketches, to working dashboards, finally to functioning Dashboards.

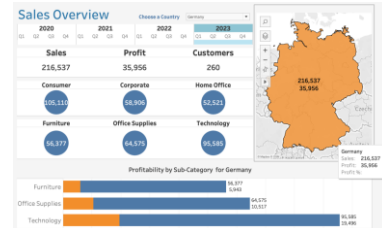
Sketch



Progress



Final versions

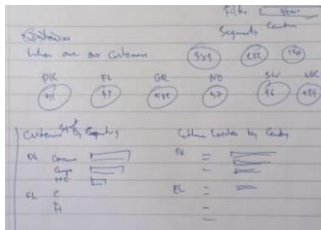


Initially we wanted to have a map showing all countries the company have business in, product categories and the number of sales.

The idea was quickly changed to having KPIs and including filters to make the dashboard more powerful. We also realized we wanted to put more visualization onto the Sales Dashboard than the space would allow. So, we decided to have a couple of more powerful visualizations in the Story on its own.

The final version consisting of KPIs, ability to filter by years and quarters. The bar chart at the bottom is used to filter the Sales by Country for the selected year (and quarters). All of the KPIs will be filtered

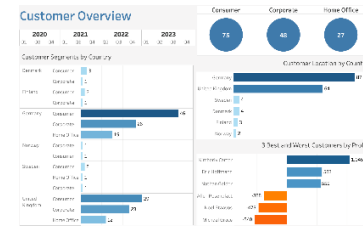
Sketch



Progress



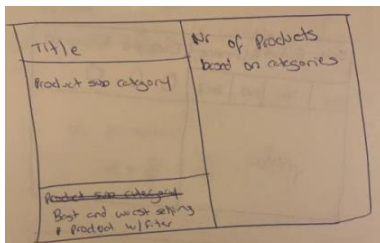
Final versions



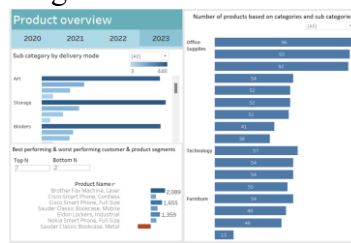
The second dashboard we would like to show the evolutions on is our Customer Overview dashboard. It started of with a simple design with the intension to show the summary of who and where the customers are.

This dashboard, then, transitioned to including the ability to filter the number of customers by country and year. Then we decided to include the 3 best and worst customers based on the profit/loss.

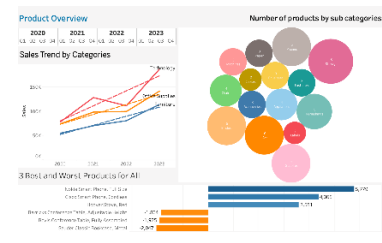
Sketch



Progress



Final versions



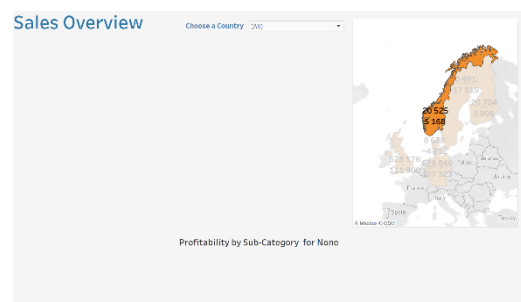
The Product Overview dashboard started off with the focus on the number of products by category. Then, different filters entered the discussion and transformed the design and visualizations to include the 3 best and worst products based on profit and loss as well. We removed the Deliver Modes element from this dashboard and changed the type of graph to using line chart which allows us to include a trend line for each of the three product

Challenges faced

- *Obtaining new skills for the project.* Thought we have learned a lot from the class, there were new skills required for this project. Therefore, we found ourselves compelled to find other resources online.

- *Making Doughnut chart.* We learned to create doughnut charts so that we can show the proportion of profit to sales and not using pie charts.

- *Using Map as filter.* We tried to use map as a filter for countries. It worked well for the visualizations in Dataset2. However, we could not use the map as filter in the Sales Overview dashboard in Dataset 1. When we turned map as filter on, all but the map itself disappeared. It took us a while, but we overcame the issue in the end (screenshot is in the next section).



Part Three: Answering the question posted in exam.

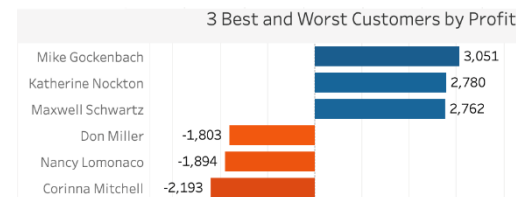
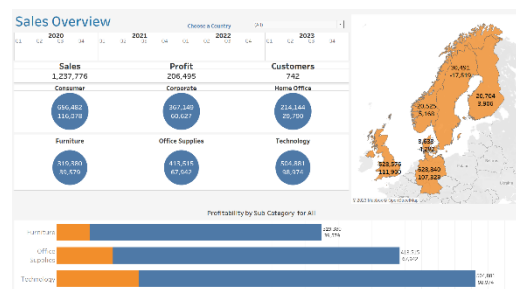
Dataset 1:

- a) Which segments make the most sales (revenue) and profit over time?

The market segment that generated the high revenue and the most profit “Consumer”.

- b) Identify their best performing and worst performing customer

The 3 best customers based on Profit / loss generated from 2020 – 2023 are Mike, Katherine and Maxwell, while the worst customers were Don, Nancy and Corinna. This graph is included in “Customer Overview” where the list can be filtered by year and country.

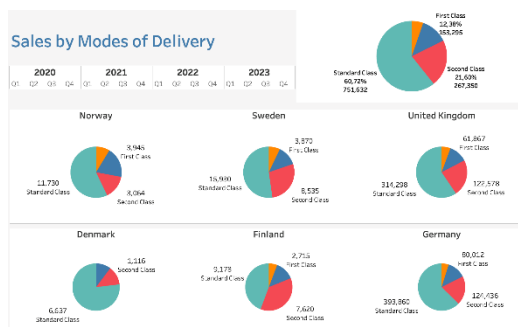
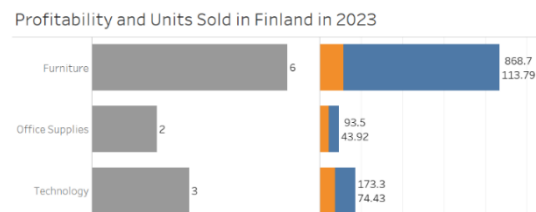


Identify their best performing and worst performing product segments



The best performing customer based on profit generated over all 3 years were Nokia Smart Phone, Cisco Smart Phone and Hoover Stove. While the 3 worst performing products were Barricks Conference Table, Bevis Conference Table, and Sauder Classic Bookcase.

The best and worst performing product segment can be found in the “Product Profitability” dashboard. The graph has been filtered to Finland for 2023.

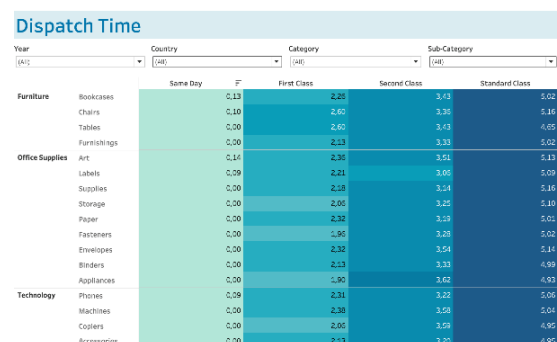


- c) Drill down on regions and delivery modes.

We have created one dashboard called “Sales by Modes of Delivery” where the user can have a quick view of the sales by country, and it can be filtered by year.

- d) Time takes to dispatch orders across products.

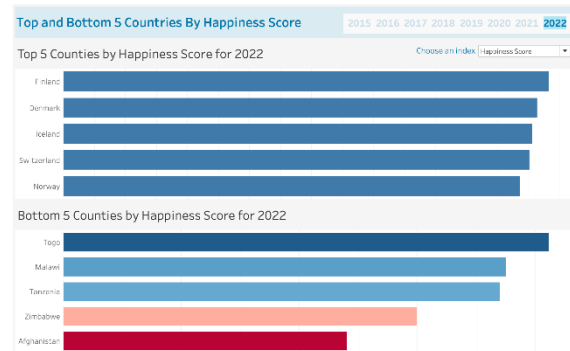
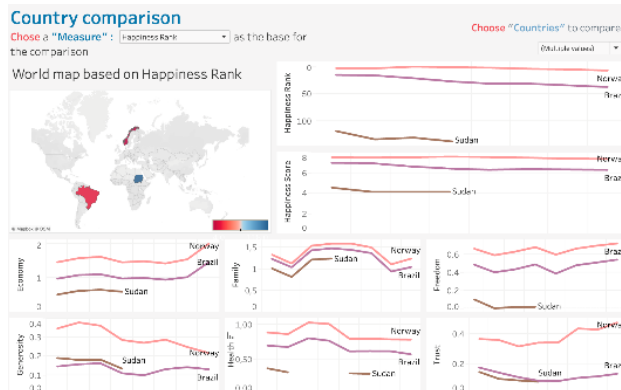
This dashboard allows the users to explore the time it takes for stores to send orders to their customers. And it can be filtered by year, country, category and sub-category.



Dataset2:

a) Ranking and happiness scores of countries over time.

The ranking of countries based on “Happiness Score” is shown in the screenshot. We decided to show top 5 and bottom 5 countries. Users can fill the indexes and filter the data by years.



b) Compare and classify different countries.

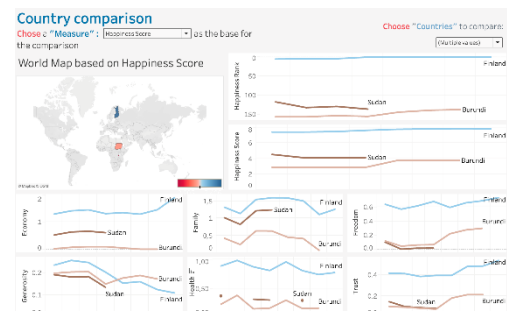
The screenshot shows comparison on basis of “Happiness Rank” over time for Sudan, Brazil and Norway.



Classification of countries based on Happiness Rank can be shown in the world map via colouring.

c) Explore the factors or variables (e.g., economy, Family, Freedom, Trust in Government, etc.) to explain the happiness of countries.

The happiness score of a country is influenced by the overall trajectory of the other 6 indexes. The above screenshot shows a comparison on Finland, Burundi and Sudan. However, it is not possible to conclude with absolute certainty without performing a correlation analysis.

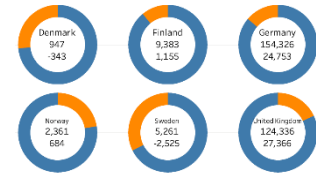


Part Four: Ten insights

Dataset 1:

- **Insight 1:** All transactions in Denmark and Sweden generated loss.

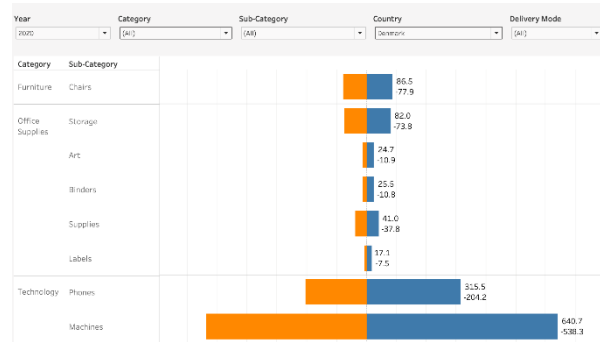
As shown in the screenshots produced from “Product Performance” dashboard. The portion the loss is on the left side on the doughnut charts, and the profit are in the negative.



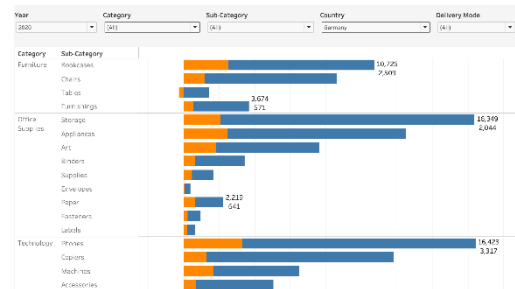
Upon further investigation, we see that the company is losing money across all product categories in these two countries.

- **Benefits and actions users can take.**

The Sales Manager can look into the operations in these two countries to figure out the real root cause. This could reveal possible misfits in Product/Pricing strategy and others.

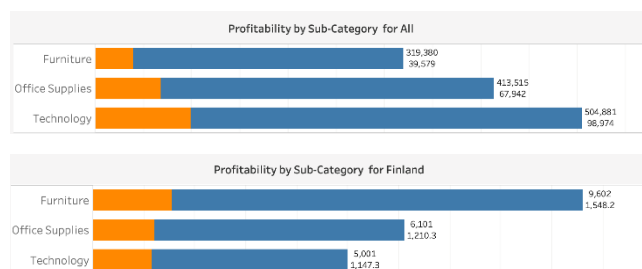


Having the profit shown on the negative side of the graph, as shown in the screenshot for the Table sub-category, can visually show the loss, as opposed to having the making profit shown in the other bars. The use of bar chart makes it easier to visually identify the product with highest sales/profit. Manager can also identify lowest performing products by delivery mode as well.



- **Insight 2: Furniture is the worst performing product group overall.**

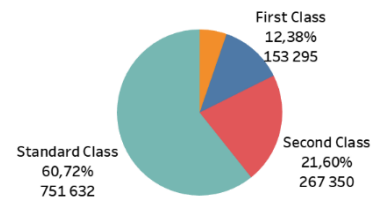
The screenshot shows a summary of sales and profit by product groups over the four years. The sales for Furniture was € 319.4 thousands, with € 39.5 thousands profit. However, Furniture was the best performer in Finland where the total sales was € 1,548, with a profit of € 9,602.



▪ **Benefits and actions users can take.** This shows that furniture has a group of products has the potential to actually make profit. The Sales Manager should look closer into the Product/Pricing Strategy. Maybe draw a lesson from the operation in Finland, then apply it across other countries.

- **Insight 3: Standard delivery is the most preferred by customers**

“Standard Class” is the most common mode of delivery chosen by the customer. It accounts for 60.72% over the four years period. It is also worth noting that the company does not offer Same-day delivery in Denmark, making the size of the yellow slice of the pie small.



- **Benefits and actions users can take.** Understanding the most common mode of delivery enables the Sales Manager to optimize the cost of the shipment by negotiating better transportation rates and consolidate shipments or explore bulk discounts. Being able to leverage the volume of business would allow the company to utilize the benefits of economy of scales.

- **Insight 4: The company is doing well in shipping the products to its customers.**

The majority of the same-day delivery orders were shipped within the one-day limit, with only shipments in the UK and Germany that took longer than 1 day to be shipped out (using country filter on the same dashboard).

Year	Country	Category	Sub Category	Same Day	First Class	Second Class	Standard Class
(All)	(All)	(All)	(All)				
Furniture		Bookcases		0.13	2.26	3.43	5.02
			Chairs	0.10	2.60	3.35	5.16
			Tables	0.00	2.60	3.43	4.95
			Furnishings	0.00	2.13	3.33	5.02
				0.14	2.36	3.51	5.13
Office Supplies		Art	Labels	0.09	2.21	3.06	5.09
			Supplies	0.00	2.18	3.14	5.16
		Storage		0.00	2.06	3.23	5.10
			Paper	0.00	2.32	3.19	5.01
		Fasteners		0.00	1.96	3.28	5.02
			Envelopes	0.00	2.32	3.54	5.14
		Binders		0.00	2.13	3.33	4.99
			Appliances	0.00	1.90	3.62	4.93
		Technology	Phones	0.09	2.31	3.22	5.06
			Machines	0.00	2.38	3.58	5.04
			Copiers	0.00	2.06	3.50	4.95
			Accessories	0.00	2.13	3.20	4.95

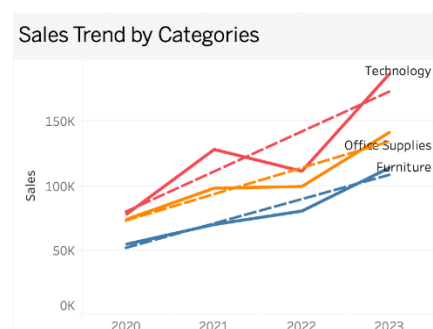
Even with bigger sized items such as furniture, the company managed to have them shipped to their customers with an average of 4-5 days for Standard Class delivery.

- **Benefits and actions users can take.** The company can drill down to country level to see the shipping performance and further improve the time to dispatch products. The company can better understand the customers’ demand patterns in each country which will assist stock keeping and inventory management. This will help avoid stockouts and reduce overall inventory carrying costs and improve inventory turnover.

- **Insight 5: All product category has an upward trend despite short-term fluctuation.**

Despite short-term fluctuations in the volume of sales in each of the product groups in 2022, the overall trend is still positive.

- **Benefits and actions users can take.** This is to show that all of the product groups still have potential to generate revenue and hopefully profit, the company would have to focus more on the cost and pricing strategy, as well as targeting the right target group of customers (corporate customer generate higher profit percentage than the final consumers).

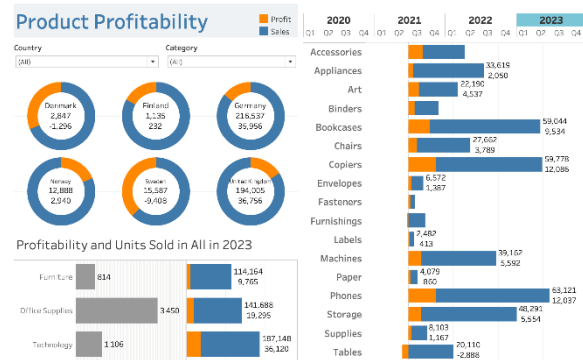


- **Insight 6: Technology product group offers the highest profit spearheaded by Phones.**

Technology accounts for the highest sales figures and profit over the four years, with Phone being the leader of the pack.

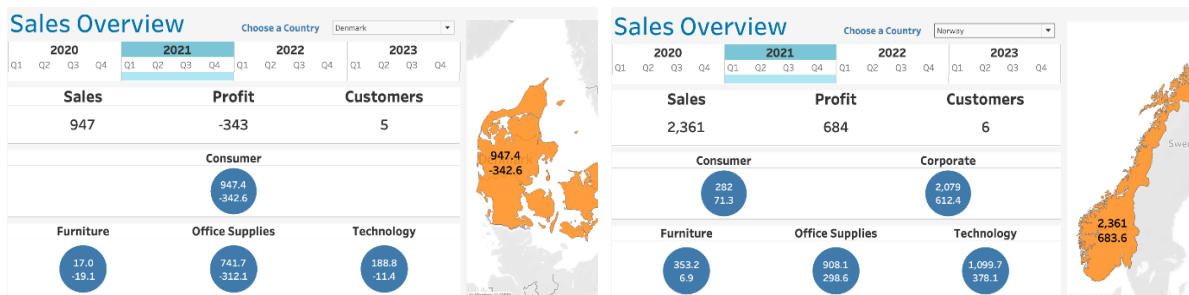
- **Benefits and actions users can take.**

The company can utilize product line extension strategy by adding more products into the existing profitable product line (Phone and Technology). In addition, the company can also use cross-selling techniques and product bundling strategy to increase the total amount of sales per existing customer.



- **Insight 7: The company doesn't have the same market segments in different countries.**

Despite having three major market segments, the company seemed to have shifted the marketing strategy resulting in reduced market segments. As shown in the screenshots below, in 2021 there was only Consumer market in Denmark (after having gained access to all three segments in 2020), then added corporate market in 2022 and went back to having three segments in 2023. The similar situation took place for Norway where the company did not have Home Office in both 2021 and 2022.



- **Benefits and actions users can take.** The company could do market research in each country to better understand the customer and their needs. An example could be to understand why furniture were sold more in Finland than in other countries. Is it because there has been a bigger demand from the customer there then in other countries, or is it the situation where the has been an untapped market which represents an opportunity for growth.

Benefits would be both market and product diversification, by which the company can better target different segments in different countries. These diversifications help avoid risks that come from fluctuations in a particular segment or country.

By understanding of markets, it offers the company competitive advantages through clearly identified market segments in each country, and better serve customers' needs better than competitors.

Dataset 2:

- **Insight 8: Countries in the African continent are poorer (citizens has least healthy economy)**

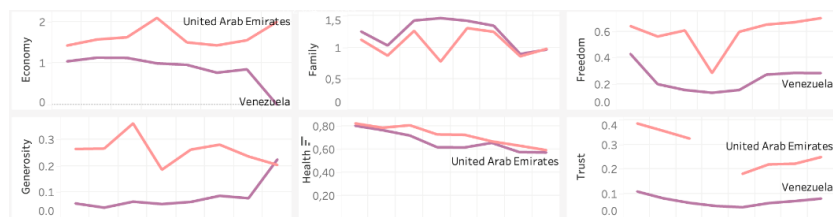
It may not come as a surprise, but this reiterate the fact that people in countries in Africa are generally economically impaired than any other countries in other regions.

- *Benefits and action users can take.* The map only shows that the poorest region is Africa, it does not show reason behind it. This could spark an interest to dive deeper into the underlying cause for this region. The attention on the issue would be useful in raising awareness and understanding, prompting discussions and encouraging the public to learn more about the underlying causes and factors that contribute to poverty. It can also garner more compassion and empathy which could encourage the public to support more charitable organizations.



- **Insight 9: The level of Generosity seems to have an inverse relationship to Economy.**

Comparing
Venezuela to United Arab
Emirates, we could see an
emerging relationship



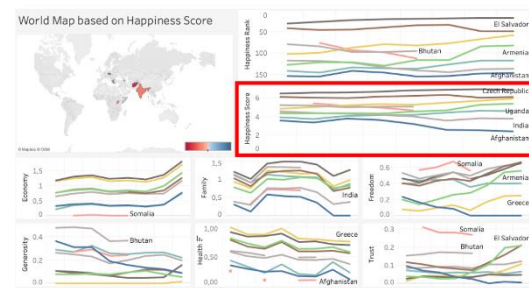
between “Economy” and “Generosity”. There seems to be an inverse relationship between the two indexes. When the economy is going well (in the case of UAE) the citizens become less generous. As opposed to the situation in Venezuela where the economy took a nose dive, but its citizens became more generous.

Human are generally kind and good to others. See the case of Venezuela where the country is experiencing economic downturn while generosity score has a steep incline. (Add screenshot)

- *Benefits and action the user can take.* This could be a spark of interest in further research to better understand the underlying factors, and to see if there is a correlation or causal relationship between economic prosperity and the level of generosity. Additionally, this could raise awareness and generate self-reflection, and advocacy for social change. The citizens could be inspired to evaluate their own levels of generosity and consider how they can contribute positively to their own country.

- **Insight 10: Happiness Score changes gradually, fluctuations in an index is offset by the effects of the others.**

Happiness score tends to share trajectory with other indexes. It changes rather gradually over time compared to the fluctuations in other indexes as seen in the screenshot. Specifically, when most of the indexes for a country has a positive incline, then the happiness score is also on an increase, NOT necessarily the happiness ranking of the country.



- **Benefits and actions users can take.** Equipped with the understanding that the level of “happiness” is subjective and complex and highly based on culture. Happiness score is affected by a combination of factors, some of which may not be included in the dataset (e.g., religions). People are encouraged to explore what factors contributing to their own happiness, particularly indexes such as family and trust. This should also serve as a reminder that people should prioritize happiness and well-being, rather than increase the score of one particular factor.

- **Insight 11: We are living in the world of relatively high perceived corruption in the government and business.**

The results of using Trust as a proxy of perceived corruption in the government and business reveals a rather scary world we live in. There is relatively high level of perceived corruption in the majority of countries. With western European countries leading the pack of countries with lowest perceived corruption in the society.



- **Benefits and actions users can take.** The realization could serve the public to strive to do better, be more transparent and enforcing the check-and-balance of power. The public and advocate for policies and practices that promote freedom of the press, freedom of information which would encourage the citizens to participate in public policy decisions. Higher level of participation and accountability can lead to an increased in Trust in government.



References:

Few, S. (2009). Data Visualization for Human Perception. [online] The Interaction Design Foundation. Available at: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception>. [Accessed 16th May 2023].

- Healy K, (2019). Data visualization: a practical introduction [Book, Online PDF]. Princeton, N.J.: Princeton University Press. Available at file:///Users/mithi/Downloads/dokumen.pub_data-visualization-a-practical-introduction-0691181616-9780691181615.pdf. [Accessed 19th May, 2023]
- Penguin Analytics. Tableau in Two Minutes - Using Nested If Statements. <https://www.youtube.com/watch?v=Z5j0l3f6GOA> [Accessed on 19 May 2023].
- Tableau, (n.d.). Data Types. [online] Available at: https://help.tableau.com/current/pro/desktop/en-us/datafields_typesandroles_datatypes.htm. [Accessed 11th May 2023].
- Tableau, (n.d.). Finding the Top N Within a Category | Tableau Software. [online] Available at: <https://kb.tableau.com/articles/howto/finding-the-top-n-within-a-category> [Accessed 11th May 2023].
- Tableau. (n.d.). Bar Charts. [online] Available at: <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/bar-charts> [Accessed 11th May, 2023]
- Tableau. (Jeffery Shaffer). 5 Tips on Designing Colorblind-Friendly Visualizations. [online] Available at: <https://www.tableau.com/blog/examining-data-viz-rules-dont-use-red-green-together>. [Accessed 16th May 2023].
- Tableau. (n.d.). Understanding and Using Tree Maps. [online] Available at: <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/treemaps>. [Accessed 11th May, 2023].
- Sosulski, K. (2019). Data visualization made simple : insights into becoming visual. New York, Ny: Routledge [Book, Online PDF]. Available at: https://cicerocq.files.wordpress.com/2020/03/sosulski-kristen-data-visualization-made-simple_-insights-into-becoming-visual-2019-routledge.pdf [Accessed 18th May 2023].
- Wexler, S., Shaffer, J. and Cotgreave, A. (2017). The Big Book of Dashboards : Visualizing Your Data Using Real-World Business Scenarios [Book, Online PDF]. Hoboken, Nj: John Wiley & Sons. Available at: file:///Users/mithi/Downloads/The%20Big%20Book%20of%20Dashboards_%20Visualizing%20Your%20Data%20Using%20Real-World%20Business%20Scenarios.pdf [Accessed 15th May 2023].
- Ynteractive Training. How to Compare Data From Two Countries in Tableau, https://www.youtube.com/watch?v=kiKIF_NMe_E [Accessed on 15 May 2023].
- Ynteractive Training. How to Format a Number as a Decimal in Tableau, <https://www.youtube.com/watch?v=FF4zAmaA2Uw> [Accessed 19 May 2023].
- Ynteractive Training. How to Compare Data From Two Countries in Tableau. [online] Available at: https://www.youtube.com/watch?v=kiKIF_NMe_E [Accessed 21st May 2023].
-
-