

Exam Report - Visual Analytics

Dataset 1

Description

Dataset 1 contains data on orders, customers and sales for the fictive Company X. Each row represents a product ordered from what we assume is an online store providing office supplies such as stationery, furniture and electronic devices. We will describe some of its columns, and attempt to explain our interpretations of them. There is a lot of information in this dataset and we have been determined to narrow it down, while being aware of the limitations our views can hold. Since this is a store, our motivation has been to assist the company by communicating links between revenue/loss and trends in the different customer groups and countries.

The column *Row ID* contains a unique ID for each row. This data is numerical and interval. Its values can not be added or sorted, since their values are not chronological. Each *Row ID* will only show once in the dataset, as its unique number is used to mark each row. We used either *Order Date* or *Dispatch Date* when we wanted to view data over time in our sheets. These ordinal values can be categorized into years, quarters, months, weeks, days and we also used the values in calculations, for example when wanting to find the difference between order date and dispatch date. Unlike *Customer Name*, where different customers potentially can hold the same names, *Customer ID* is each customer's unique number. These values can show up in several rows, as each customer can place several orders. There is no benefit to calculating these numbers as they only serve to characterize another value in the dataset. *Customer ID* is nominal because it cannot be ranked or categorized. Each row number with a *Customer ID* can be counted together (Distinct Count), but the values themselves can not be counted. *Country* holds the data declaring from where each order is placed. This data provides important information about the geographic distribution of the company's orders. Understanding where orders are coming from can help make informed decisions regarding the company's future marketing focus. As with *Customer ID*, *Product ID* holds a unique value for each product. This data can provide insightful information about which products make the most sales or potential shortfalls of some products. Dataset 1 also has the column *Product Name*, but like in the case of *Customer ID* and *Customer Name*, we

relied on *Product ID* when extracting our data.

Transformations

By making a calculated field with the calculation “ $SUM(Profit)/SUM(Sales)$ ” we created *Profit Ratio*. Profitability Ratio is a calculation used to measure revenue per sale, this measurement can thereby withdraw potential costs related to each sale and expose what amount of the profit is left from that. In this assignment, we must take into consideration that we are missing data on expenses, if they do exist. This could be storage, internet domain costs, salaries etc.

Profit Ratio

$SUM([Profit]) / SUM([Sales])$

Another calculated field we made is *AVG days of packaging* (“ $AVG([Dispatch Date]-[Order Date])$ ”) which we made by withdrawing the day of order from the dispatch date. This lets us view the days between the order date and the day the order got dispatched.

AVG Days of packaging

$AVG([Dispatch Date]-[Order Date])$

Dashboards, Design

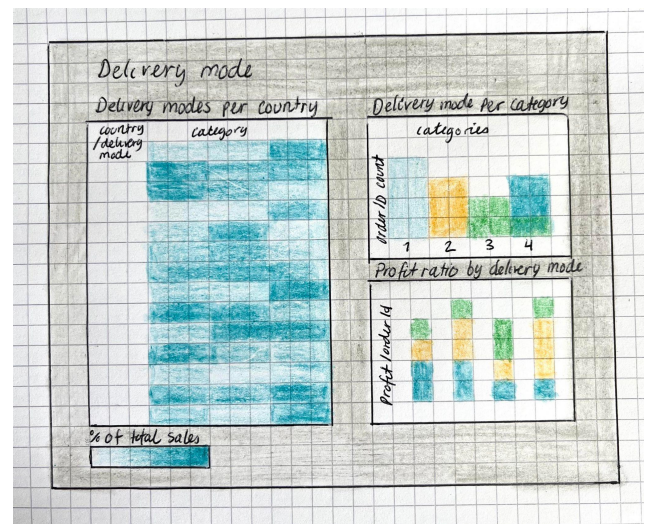
When figuring out the design, it was important to us to try to depict our data in a neutral and tidy way, avoiding potential confusions. We also wanted our assignment to possess a coherent, well developed style, for it to be easily consumed by its viewers. We early agreed on using blue as our main chart value color, as it is known to be a color signaling trust and reliability. We used green and orange when needing to separate data. Green often stands for vitality and nature, orange is associated with energy and optimism (Mørstad,E. & Tschudi-Madsen,S. 2023). We made a conscious choice not to use green and red together, since they can be confused with each other by colorblinds. Blue, green and orange became our bar chart-colors, we found they complimented but also amplified each other (Mørstad,E. & Tschudi-Madsen,S. 2023). The background is gray on all the dashboards in order to reduce contrasts and make the visualizations easier to read and more digestible. We used bar charts at most times, as they are good for comparing and measuring monetary values and showing financial data. Quote from *Knaflic, C,N. 2015, s.50*: “Bar charts are easy for our eyes to read. Our eyes compare the end points of the

bars, so it is easy to see quickly which category is the biggest, which is the smallest, and also the incremental difference between categories.”

Dashboards, Development

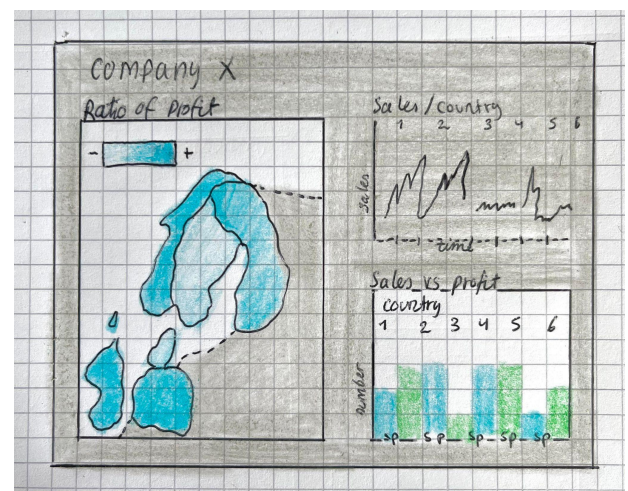
We started out by playing with and testing what we could do with the datasets in Tableau. A lot of sheets were made and then we came together and shared what insights the group participants had gotten from these examinations. This was done before any design had been developed, we

found that for us, this was the most purposeful order to do it, as the design could be made to highlight or even out some of the findings we did, as well as the fact that we wanted to establish a story or narrative and then discuss the design we were going to apply to tell it. Agreements were being made on important aspects or perspectives we wanted to include, such as focusing on revenue since it is a store’s



dataset, and further focusing on revenue and customer groups. We built our dashboards on dataset 1 on task A-D in the exam paper and categorized our chosen data into three dashboards centering around countries/overview, product categories and delivery. Since our main focus was showing revenue, we planned our first Dashboard to show a map and sales/profit linked to the different shipping locations from the dataset. Maps are very powerful visualizations for displaying spatial data and they help explain geographical patterns.

We probably could have made an understandable visualization without using a map, since Company X most likely is well acquainted with their customers' countries, but maps are still impactful and general ways of portraying geographic data (S.Few, 2009, p.4). Another reason we chose to



include a map was to use it as an interactive tool, being able to alter or customize surrounding data by clicking on it.

The first dashboard is *Sales Vs. Profit* and shows a map chart of countries *Company X* is shipping to, The list is accompanied by three sheets. When clicking each country, the data in the other charts will change to revolve around the chosen country. On the top to the left lies Sales Activity Per Quarter, using a line chart to draw the sales per year, grouped by country. Line charts are the most comprehensible way to portray continuous data, where time is visualized by the lines starting and ending point, and its position along the way can show additional values (Knafllic, C,N. 2015, s.45). Below, we find the bar chart *Sales And Profit*, this bar chart shows an interesting side by side comparison of sale numbers and profit numbers, per country. From this visualization we can for example see that Sweden makes more sales than Norway, but their profit is in minus while Norway's is in plus. Bar charts have been our standard chart type, exceptions have been made when demanded by data types like spatial data and time data. Bar charts are ideal in many cases, because they serve as standards for showcasing data and in that way stand out as more intuitive for viewers and demand less (Knafllic, C,N. 2015, s.50). The lowest chart is sales per segment, and shows the amount of profit and sales in the consumer groups, consisting of consumer, home office and corporate. We created two buttons in the upper right corner on each of the dashboards, with the purpose of navigating through our three dashboards.

The second dashboard, *Product Data*, gives an overview of profitability and sales per category, per country. First we used a map where we could click on each country to alter the surrounding sheets, *List of Countries*. After some consideration we decided the map visualization was excess in this context and that a text list would fulfill this task. *Sales In Category And Subcategory* shows how many sales are made in each category and sub-category. *Profit Ratio Per Category In Each Segment* is a bar chart that lets us view how each product category performs regarding profit ratio, also divided into customer segments.

We wanted to include data on delivery mode, related to popularity, profit and product category. We found this convenient in order to track customer activity and

preferences when ordering from the company. The third Dashboard, *Delivery Modes* is made up of three charts. A chart called *Overview of Delivery Mode Per Country* is visualized by a heatmap where the cells are saturated by their popularity/Order ID distinct count and grouped by product category and country. Text Table Charts can be replaced by Heatmaps to improve the viewers visual reception, since color can give a more immediate understanding of visualizations (Knafllic, C,N. 2015, s.42). The bar chart to the upper right, *Number of Orders per Delivery Mode in each Segment*, gives a more general view of which delivery modes are most frequently used per category, not divided by country. From these together, we can see that there are no big gaps between types of delivery used in the UK and Germany, which we can see from dashboard 1 are the most profitable countries, whereas in the less profitable countries, usage of delivery mode is more polarized. *Profit Ratio in each Delivery Mode* was first made to show sales per delivery mode but was then changed to show both profit and sales per delivery mode, illustrating some interesting lack of symmetry between profit and sales. Using a stacked bar chart allowed us to put the values of each consumer group into the bars, letting them make up the subcomponent pieces in the bars. (Knafllic, C,N. 2015, p.54-55).

Insights

1. We found the calculation "Profit Ratio" to be very useful to show revenue but it was also clear to us that it can only work when provided with all data on costs. Such as salaries, storage for products, advertisements, web domains etc. We do not have any data on this but chose to use it anyways.
2. When we found out that there was a big decline in Company X's sales in May 2022, we were compelled to find the reason for this. Since we are provided with limited data, not all patterns can be visible, and we learned that we can only work from our given data but Company X or other potential viewers may possess relatable information to help them form a bigger picture.
3. Both Denmark and Sweden have negative revenue. One significant customer drags Sweden's profit remarkably lower (Customer ID: NB-18580). We have speculated that some of the loss in the profit data might stem from returns.

We also discussed that these numbers may stem from customers returning their orders, or they could even be very high priced shipping fees paid by Company X. We considered deleting this data since they seemed like special cases that did not reflect normal orders, but decided on not deleting them. We concluded our job was processing the data given to us, not changing it.

4. Norway has the best profit per sales, but the best overall profit is for the UK. This Gives the impression that British customers use less money per order, but place orders more often than Norwegian customers. Best ratio of sales Average price per item is in Finland (100 Euros), and the lowest is in Denmark (39 Euros).
5. Technology is both the best selling category within all segments, and also generates the most ratio of profit, with phones being the highest selling product.

Dataset 2

Description

Dataset 2 is an examination on experienced happiness among inhabitants grouped by country. The score is on a scale from 1 to 10, the examination also asks inhabitants to rank their sense of freedom, generosity, family, economy and trust towards their government. Each data contributes to the overall evaluation of happiness in the different countries. The dataset contains world happiness surveys over eight years, we have used the column *Year* to display time.

The geographical data in this dataset has been our base when extracting data. Common for all the rankings and scores are that they are divided into geospatial spaces, divided into *Region* and *Country*. Without ordering the data by country or region, the data's purpose would change to only depict the world's overall happiness and the visualizations would then lose loads of potential insight on the geographic differences we have found.

Our interpretation of the survey provided in dataset 2 is that the happiness score is not an inquiry in itself but made up by multiplying the answers on perceived generosity, family, economy and trust towards government (World Happiness Report. 2023).

All the score data is numerical, they can be ordered and also added together to make up a sum or an average. The numerical *Happiness Rank* data, that puts each country on a world wide ranking after happiness score is ordinal. The values can be ordered but cannot be added together. *Health(Life Expectancy)* depicts expected years of life in each country, this data is also ordinal.

Transformations

When opening the dataset in Tableau, Year was set to String, and Region was too. These were changed to Date and Region's geographical role was set to Map. The column "Data" seems to have been made by Tableau as it did not show up in the Excel sheet. We assume this column was made on the purpose of storing values for each row in the dataset, like the column *Row ID* in dataset 1.

When viewing our map, Somaliland would show up as an "unknown" value. By this, we understood that it had an unknown geographic location. Since we could not make use of this data in the map, we used *Filter* to exclude it.

Dashboard, design:

In the world we know, there are clear positive and negative values associated with ranking happiness. We decided early on that red and blue would be proper colors to illustrate if scores were on the "good" or "bad" side of the happiness scale. Red is often associated with danger, caution, loss, while blue symbolizes calmness, stability, serenity (Ceratto, H. 2012). A good choice for the positive color could also have been green but considering this would have been an issue when presenting dashboards to colorblinds, we chose blue.

Experiences of color associations are also culturally conditioned, in China red is associated with prosperity and good luck. This may have been taken into consideration if showcasing our assignment in other parts of the world and is

important to be aware of since colors are powerful and can change the crowd's perception of the visualizations (Mørstad,E. & Tschudi-Madsen,S.2023). The happiness score could also have been measured in only one color, and the scale could for example have increased in saturation from white towards green. We opted for two colors to show a more nuanced picture. One of our goals was keeping these dashboards free of clutter. We tried making the dashboards interactive, as to keep the number of sheets to a minimum. This way we packed the dashboards with a lot of information, but still neat and easy to consume by viewers (Knafllic, C,N. 2015, s.84).

Dashboard, Development

Dataset 2 holds a lot of data on each of the countries, we had to make choices on whether to include these as separate data, and then eventually how to include them. We agreed that our opening visualization would group the countries into their regions and that way show the assembled, average data of each region. Since there are great variations within the countries in each continent or even region, we had to find a way to visualize this too.

The world happiness report is used worldwide by community organizers and researchers as well as the general public. The report provides defined indicators on happiness, and this data can be used to show correlations between these indicators and the overall happiness score. We decided our task would be to give an overview of the trends per geographic area, and possibly point out correlations between the indicators. We meant for these overviews to be simple and comprehensible to both academics and the general public. Dashboard one consists of a map displaying each country's happiness score and rank by color and bar charts displaying each region's average score and rank, and their average score compared to different indicators of happiness such as Trust in Government and Economy. To make these comparisons across years, countries and regions we chose to use several stacked bar charts, both in the main and second dashboard. When hovering above each country in the map, one can examine each country's specific rank and score, as well as what region the country belongs to. Clicking on one country will highlight the country's statistics in the bar charts. The filter in the top left corner allows for the viewer to

show statistics for one year at a time. If choosing all years, the dashboard displays an average score from years 2015-2022.

In dashboard two we chose to narrow our data down to look at both the top and bottom ranking countries through a filter based on years. The dashboard consists of two bar charts displaying the 10 most happy countries and the 10 least happy countries by rank. As previously stated this dashboard also allows for filtering by year, and choosing all years will display an average here as well. A button in the upper right corner lets us navigate between the two dashboards in dataset 2.

From these bar charts we see that the top ranking countries in the world are mostly located in either Western Europe or Australia and New Zealand no matter what year, while the bottom ranking countries are mostly located in Sub-Saharan Africa. There is also a higher focus on the key indicators of happiness, such as economy, freedom and health, in this dashboard. In the top ranking countries we see a high and steady score in Economy and Family, while the bottom countries have a low score in ec. Trust in Government and Generosity. It is worth mentioning that the higher scoring countries also display a low score in Trust in Government and Generosity, and a somewhat low score in Freedom.

Insights:

1. Surprisingly, most countries have an average score below 0.5 in Trust in Government. Overall Singapore scores highest with an average of 0.5, while Bosnia-Herzegovina scores the lowest with an average of 0.0026.
2. In the years 2021 and 2022, the top ranking countries' health score drops significantly, while we see a high and steady score in Economy and Family throughout the years.
3. Looking at the lowest ranking countries, Family is one indicating factor that scores relatively high compared to the other measurements in the average overall score, such as Trust in Government.

4. Throughout the years there are several countries missing from the report. This will naturally affect each region's ranking as well as the country itself in its average score.
5. Below is an average overall score of each indicator of happiness' top and bottom country. Although not necessarily landing a top or bottom rank, the countries in this table display the variety in each continent/region.

	Trust	Health	Generosity	Freedom	Family	Economy
↑	Singapore	Singapore	Myanmar	Uzbekistan	Island	Qatar
↓	Bosnia-Herzegovina	Central African Republic	Greece	Angola	Central African Republic	Somalia

References

- Ceratto, H. (2012) '*The Meaning of Colors*', p.4-20. Available at: <https://bloccs.xtec.cat/gemmasalvia1617/files/2017/02/the-meaning-of-colors-book.pdf> (Accessed: 08.05.2023).
- Edge, P & Few, S. (2009, Mars/april) '*Introduction to Geographical Data Visualization*', Visual Business Intelligence Newsletter. p.4. Available at: https://www.perceptualedge.com/articles/visual_business_intelligence/geographical_data_visualization.pdf?fbclid=IwAR2kOoWC1_tQjvYe1TOrKH9vrjnP6495No-lv6YF0-V_EyAaLhG6X08jgE (Accessed: 10.05.2023).
- Knafllic, C,N. (2015) '*Storytelling with Data*', chap.2: '*Choosing an effective visual*' p.42 - 55. Available at: <https://www.asrepayesh.com/assets/asrepayesh.com/repo/file/storytelling-with-data-cole-nussbaumer-knafllic.pdf> (Accessed: 11.05.2023).
- Knafllic, C,N. (2015) '*Storytelling with Data*', chap.3: '*clutter is your enemy!*' p.71-84. Available at: <https://www.asrepayesh.com/assets/asrepayesh.com/repo/file/storytelling-with-data-cole-nussbaumer-knafllic.pdf> (Accessed: 11.05.2023).
- Mørstad, E. & Tschudi-Madsen, S. (2023, 23. februar). '*Fargesymbolikk*'. Store Norske Leksikon. <https://snl.no/fargesymbolikk> (Accessed: 09.05.2023).
- World Happiness report (2023). FAQ. WHR. Available at: <https://worldhappiness.report/faq/> (Accessed: 15.05.2023).