**DATA VISUALISATION COURSE WORK**

**CS5803**

**Student ID: 2351211**

**Student Name: Nazeema Begum Rahman basha**

**2023-2024**

Contents

[1. INTRODUCTION 3](#_Toc164047756)

[1.1 DATA 3](#_Toc164047757)

[1.2 PRE-PROCESSING OF DATA: 4](#_Toc164047758)

[1.3 PERSONA AND QUESTIONS: 4](#_Toc164047759)

[1.4 REQUIREMENTS: 4](#_Toc164047760)

[2. DESIGN 5](#_Toc164047761)

[3. IMPLEMENTATION 6](#_Toc164047762)

[3.1 SALES PROFIT ACROSS DIFFERENT COUNTRY 6](#_Toc164047763)

[3.2 SALES VOLUME OF TOP 3 MARKETS: 7](#_Toc164047764)

[3.3 DISTRIBUTION OF SALES VS SUB\_CATEGORIES AMONG TOP 3 MARKETS 8](#_Toc164047765)

[3.4 AVERAGE SHIPPING COST ACROSS DIFFERENT REGION BY SHIP MODE: 10](#_Toc164047766)

[3.5 SECONDARY IMPLEMENTATION USING POWER BI: 10](#_Toc164047767)

[3.5.3 DISTRIBUTION OF SALES VS SUB\_CATEGORIES AMONG TOP 3 MARKETS 12](#_Toc164047768)

[3.5.4 AVERAGE SHIPPING COST ACROSS DIFFERENT REGION BY SHIP MODE: 13](#_Toc164047769)

[4. WALK THROUGH 15](#_Toc164047770)

[5. CONCLUSION AND DISCUSSION 17](#_Toc164047771)

[REFERENCE: 18](#_Toc164047772)

[Figure 1: Paper Landscape 6](#_Toc164048784)

[Figure 2: Interactive Dashboard 7](#_Toc164048785)

[Figure 3: sum of profit across country 8](#_Toc164048786)

[Figure 4: Sales volume of Top 3 country 9](#_Toc164048787)

[Figure 5: Sales Vs Sub\_categories among Top 3 markets 10](#_Toc164048788)

[Figure 6: Sales Vs Sub\_categories among Top 3 markets across Customer segment. 10](#_Toc164048789)

[Figure 7: Average Shipping Cost across different Region 11](#_Toc164048790)

[Figure 8: Choropleth Map in Tableau 12](#_Toc164048791)

[Figure 9: Packed Bubble chart 13](#_Toc164048792)

[Figure 10: Stacked bar chart for Sales Vs Subcategory 14](#_Toc164048793)

[Figure 11: Tree map for Average shipping cost across different region 14](#_Toc164048794)

[Figure 12: Interactive Dashboard Power Bi 15](#_Toc164048795)

[Figure 13: Tableau Story on Choropleth Map 16](#_Toc164048796)

[Figure 14: Tableau Story 2 on sales volume 17](#_Toc164048797)

[Figure 15: Tableau story 15 on sales performance and its percentage 17](#_Toc164048798)

[Figure 16: Tableau story 4 Average shipping cost by region 18](#_Toc164048799)

1. INTRODUCTION

Visualisation is an essential tool in the design process. It enables good design by allowing for exploration of both challenges and potential solutions (Chattopadhyay, A. and Gorn, G.J.,2001 et al). Visualisation designing process carried out by two ways. Firstly, we need to create a paper Landscape with the visuals experienced from our mental psychology which is based on imagination. Secondly, implementation was carried out using Visualization tool. The chosen dataset from Kaggle is a superstore order dataset in which the data on the sales of different products by the store[[1]](#footnote-1). The information also includes geographical location, product categories, shipment mode, type of customer, sales, quantity, discount, profit, order priority with respect to the given year etc, The currency for profit and sales is not mentioned in the dataset. So assuming USD as the currency.

## 1.1 DATA

The data dictionary of key variables used for the visualisation shown below:

|  |  |  |
| --- | --- | --- |
| **NAME** | **DISCRIPTION** | D**OMAIN** |
| Ship\_mode | The shipping mode chosen for the order (First class, Same day, Second class, Standard class) | Nominal |
| Segment | Segment to which the customer belongs to (Corporate, Consumer, Home Office | Nominal |
| State | State of the customer where the product was sold (Abia, Budapest etc.,) | Nominal |
| Country | State of the customer where the product was sold (Afghanistan, Hungary etc.,) | Nominal |
| Market | Market where the product sold (Africa, EU, US etc.,) | Nominal |
| Region | Region where the market belongs to (Africa, Canada etc.,) | Nominal |
| Category | Category under which the product falls (Furniture, Office Supplies, Technology) | Nominal |
| Sub\_category | Subcategory of the product (Storage, chairs, Accessories etc.,) | Nominal |
| Sales | Sales of the product in USD | Float |
| Quantity | Number of Quantity the product sold | Integer |
| Profit | Net Profit after selling the Goods/products in USD | Float |
| Shipping\_cost | Cost of shipping according to the shipping mode and priority in USD | Float |
| Year | Year of the | Integer (2011-2014) |

## 1.2 PRE-PROCESSING OF DATA:

The dataset consists of 51k rows which is unclean. Data cleaning has been carried out at the initial stage. In addition to that, a sample of data with 3k rows has been chosen for the visualisation process through Sampling technique of Machine learning.

## 1.3 PERSONA AND QUESTIONS:

The Persona is a supply chain manager of the store who manages the flow of items from suppliers to customers and optimizes supply chain procedures to maximize the profit. He needs to know the trend of sales, profit across different region and customer segment across the year 2011 - 2014 to improve the sales and profit margin of the store.He

1. How does the sales profit across different countries varies by year?

2. How does the sales volumes of the top three markets, categorized by customer segments and analysed over various years, what is the percentage of sales and how it varies with respect to the markets and customer segments?

3. Compare the sales performance of top 3 market with different product subcategories vary across customer segments, and how does it contribute to overall profitability?

4. What is the average shipping cost with respect to region and shipping mode, and how does the cost vary across different region by year?

## 1.4 REQUIREMENTS:

In this section, we outline the key needs for visualizing the relationships needed to answer each question. We also provide brief design concepts for representations and interactions.

1. To answer Q1, the persona needs to see sales profit across each country with variables such as **sum of** **Profit**, **Country, State** and **Year**.

* Density Map or Choropleth map of Country with the geographic role assigned to it, states, Sum of profit by Year.

2. To answer Q2, the persona needs to do a comparison of sum of sales with customer segment by top 3 markets based on sales value.

* Packed bubble chart Sum of sales by segment and market
* Included the market in the filter to visualise top 3 market by sum of sales and years.

3. To answer Q3, the persona wants to, compare the sales with subcategories of top 3 sales markets.

* Using Bar chart sales Vs product sub-categories/Market.

4. To answer Q4, the persona wants to explore, the Average shipping cost with respect to different region and ship mode.

* Using Tree map on Avg (ship cost) coloured by region.
* Year on the filter, region, year and Avg(Ship cost) on label card.

# 2. DESIGN

The Initial steps in the creation of the dashboard was paper landscape, in which all the questions were answered with suitable graph. At the initial phase, scatter plot and line chart was implemented Fig1.

A close-up of a graph

Description automatically generated

Figure 1: Paper Landscape

While implementing the graph the packed bubble chart makes more sense to the question at the time of creating a visualisation, the final product’s layout, design, and aesthetics has to be manage completely.

A screenshot of a computer

Description automatically generated

Figure 2: Interactive Dashboard

# 3. IMPLEMENTATION

**PRIMARY IMPLEMENTATION:**

Tableau is a robust data visualization application that enables users to build dynamic and intelligent visuals using a variety of data source. The Primary implementation of Tableau for the data source above mention is as follows:

## 3.1 SALES PROFIT ACROSS DIFFERENT COUNTRY

As mentioned above, this question is answered using choropleth Map with sum of profit, country, sales by year. (CS5703 Lab 19, (2024))

Load the Tableau with the data source. To create the map, drag the “country” to the data pane. Convert the geographic role of the “state” to country and add it to the label card. Drag and drop sum of profit, year, and country to the Detail on the mark card. Add the “year” as a global variable, that interact with the multiple worksheets on a dashboard, and it is the effective method for managing the data that is shown in different representations.

A map of the world

Description automatically generated

Figure 3: sum of profit across country

Sum of profit is added to the “colour” of the mark card and edited the range of the colour. Areas with high or low profitability can be easily identified by applying colour to distinct profit value ranges. Add the range value slider filter to the profit that lets users choose between a range of minimal and maximum profit values.

## 3.2 SALES VOLUME OF TOP 3 MARKETS:

Packed bubble chart is used to visualize this Sales volume of Top 3 markets. Drag the “sales” measure to the “size” of the mark card and “segment” to the “colour”. To visualize the labels , add the market, segment, year to the “Label” options in the mark card. To visualize the percentage of sales, drag and drop the sum of sales to the “Label” click on the drop down of the sales choose “quick table calculation” -> “Percent of total”.

Drag “market” to the filters. In that filter card click on edit filter, choose “Top” -> click on “By field”, then choose Top 3 by “sales sum”. The reason for choosing the sales sum is, it makes the char more meaning when compared to the average value. As the global filter is already added, click on the filter to view the sales volume for various year ranges from 2011 to 2014. In this graph, the label shows the various field related to the visualisation. The colour cards for the segment consumer, corporate, and home office is shown at the filter area of the Tableau worksheet shown Fig 3.

The top 3 market according to the 2014 year were APAC, LATAM, and EU. The visualisation and percent of sales varies with respect to the chosen year.

A screenshot of a computer

Description automatically generated

Figure 4: Sales volume of Top 3 country

## 3.3 DISTRIBUTION OF SALES VS SUB\_CATEGORIES AMONG TOP 3 MARKETS

The Sales performance of Top 3 markets is implemented using Stacked bar chart. The various subcategories involved phones, storage, Accessories, Copiers, Bookcases, Machines, Chairs, Appliances, Art, Supplies, Binders, Paper, Tables, Furnishings, Envelopes, Labels, Fasteners etc., Drag the Market and sub\_category to the “column” shelf and “sum of sales” to the “Row” shelf.

Drag and drop the segment to the “colour”, “sum of profit” to the “Detail” mark card, “sum of sales” to the “Label”. Follow the same procedure for “market” filter and additionally drop the “segment” to the “filter” card.

This chart clearly shows the distribution of sales by product categories among Top 3 markets is implemented. The segment filter at the right side will shows the “colour” code of the each customer segment. By clicking each colour segment, the sales volume of each segment along with the profit for top 3 markets is shown. Fig 5.

A screenshot of a graph

Description automatically generated

Figure 5: Sales Vs Sub\_categories among Top 3 markets

The Year filter in the filter area shows the changes across the Top 3 market in the selected year with respect to sales volume. Using this chart, we can interpret the various sales volume and profit across different subcategories among various segments (CS5703 week 20, 2024)

A screenshot of a computer

Description automatically generated

Figure 6: Sales Vs Sub\_categories among Top 3 markets across Customer segment.

## 3.4 AVERAGE SHIPPING COST ACROSS DIFFERENT REGION BY SHIP MODE:

To implement this, Tree map is used to visualize how the shipping cost varies across different region by ship mode (CS5703 Lab 4, 2024)

**A screenshot of a computer

Description automatically generated**

Figure 7: Average Shipping Cost across different Region

To generate a tree map, drag and drop the Average shipping cost to the “Size” of the mark card and Ship mode to the “colour” of the mark card. Region, Year, and Average shipping cost should be added to the “Label” mark card. Click on the show me option and choose the “tree map”. By generating the Tree map, the average cost along with the region visualised by “color” card is shown.

By clicking the year filter, you can see the difference in the Average shipping cost with respect to different ship mode among various years.

## 3.5 SECONDARY IMPLEMENTATION USING POWER BI:

For the secondary implementation, Power BI is used. PowerBi uses the DAX formula language for analytics techniques.

The initial step involves load the data to the Power Bi. Click on the “Get data” option from the menu bar and choose the file type which is to be uploaded. Then choose the file from the device or add the website link for the dataset which need to be uploaded.

By clicking the report view, the creation of the dashboard is initiated.

**3.5.1 SALES PROFIT ACROSS DIFFERENT COUNTRY**

To create a choropleth, map the following steps follows:

**A screenshot of a computer

Description automatically generated**

Figure 8: Choropleth Map in Tableau

For creating a geographical map in the Power Bi, click on the visualisation tab on the right side. Choose the option “Map” then from the dataset choose “country” to get the geographic location. Then click on the “sales” to visualize the sales across different countries. The filter automatically appears on the filter area on the dashboard. To make changes in the geographic location, “colour” has to modify manually, and the rules have to be set to get the colour changes according to the range of sales profit.

**3.5.2 SALES VOLUME OF TOP 3 MARKETS:**

Packed bubble chart not available among the visualization charts. By clicking on the three dots displayed on the visualization chart, click on get more visuals, then it will redirect to the page which contains available package for tableau. By clicking on add the package is added to the power Bi worksheet.

**A screenshot of a computer

Description automatically generated**

Figure 9: Packed Bubble chart

To view the sales volume across top 3 markets, add the segment, year and market to the category part in the visualization pane. bubble chart is created by changing the year filter and it will vary according to the year chosen,

## 3.5.3 DISTRIBUTION OF SALES VS SUB\_CATEGORIES AMONG TOP 3 MARKETS

To Implement this chart, click on the Stacked column chart from the visualization, then add the axis to create a stacked column chart. Click on “sub\_category” and add to the x-axis along with the market. Add the “sum of sales” to the y-axis. By adding the filter, for market, segment, sub\_category, and year, the stacked bar chart will be visualized as shown below fig 10.

**A screenshot of a computer

Description automatically generated**

Figure 10: Stacked bar chart for Sales Vs Subcategory

## 3.5.4 AVERAGE SHIPPING COST ACROSS DIFFERENT REGION BY SHIP MODE:

Tree map is used to visualize this graph using power Bi. Add the ship\_mode, year and region to the “category” shelf, region on the “detail” card. Add the average of shipping\_cost to the value shelf.

A screenshot of a computer

Description automatically generated

Figure 11: Tree map for Average shipping cost across different region

A screenshot of a computer

Description automatically generated

Figure 12: Interactive Dashboard Power Bi

* Colour intensity and saturation must be taken in consideration while implementing the visualisation (CS5703 Lecture 2, 2024)**.** High-contrast colour may affect the visualisation. While implementing the choropleth map in Power Bi, the colour hue is a bit challenging factor. In the Tableau, it automatically fetches the colour range if we use the measures to the map and the colour range in the tableau is soft when compared to power bi.
* In tableau, during the implementation of stacked bar, we can implement two categories on the same access at the same graph . It makes the graph meaningful. But in the power Bi it is hard to implement the same thing. During the implementation, the two categories were added on the same axis which might confuse the users while visualizing the dashboard.
* Packed bubble chart is a default visualization chart in the Tableau. On the other hand, in Power Bi, the chart need to be add from the resources. Adding filter is also a complex thing in PowerBi.
* Tree map is more similar process in both Tableau and Power Bi. To create a calculation field, DAX formula language is used in Power Bi whereas in Tableau it is a simple calculation according to the variables which we have.

# 4. WALK THROUGH

The objective of the dashboard is to give a complete view of the supermarkets sales ,profit to extend their business to the next level and how much revenue they got at the end of the year to improve their profit too. The first question to be answered is the sum of profit across the country with respect to year. By answering the question the Tableau story is created fig 12

* The annotated screenshot shows that the store gains a profit of 8$ from the state Chelyabinsk of Russia.
* Range value filter is chosen for the profit so that the user can easily access the minimum and maximum range of profit.
* Global filter is added to the Year to get a complete view of the transformation carried out in each sheet.
* By sliding the profit value, a clear view of the profit distribution over country can be viewed.

**A screenshot of a computer

Description automatically generated**

Figure 13: Tableau Story on Choropleth Map

In the Fig 13, the story sheet shows the solution for the sales volume of Top 3 market, categorized by customer segments and analysed over various years and the percentage of sales. The justification are as follows:

* In the annotated screenshot it is clearly shown that the label marks the Market, Customer segment, Year and percentage of sales over the year.
* By changing the filter to another year, a clear view on how the sales volume varies across different segment.

A screenshot of a computer screen

Description automatically generated

Figure 14: Tableau Story 2 on sales volume

By answering the 3rd question the Fig 14 Tableau story 3 on sales performance of Top 3 market is created.

A screenshot of a graph

Description automatically generated

Figure 15: Tableau story 15 on sales performance and its percentage

* The Label in the Fig 14 shows the value of Market, Segment, Sub \_category, profit, and sales across the top 3 markets by year. Label is added instead of annotation, the reason behind that is while doing the annotation, it affects the next chart created by changing the filter. And the dropdown filter is added to know the clear view of trend across each year by segments and markets.
* To answer the question how the sales performance contributes the overall profitability, the profit depends on the number of sales, and the quantity of each items sold. The bar chart shows the trend according to the sales rate.

The average shipping cost across different region various with respect to the year and ship mode. The below Fig 15 shows

A screenshot of a computer

Description automatically generated

Figure 16Tableau story 4 Average shipping cost by region

* In the Fig 15, the label shows the values needed for the question. The average shipping cost and how it varies according to the shipping mode.
* The label shows the region, the shipping mode, the year in which it takes place and the average shipping cost.

# 5. CONCLUSION AND DISCUSSION

The main objective of creating this dashboard is to answer the questions needed for the persona. This dashboard helps the supply chain manager to interact with the various aspects needed to improve the organisation profit and sales across the world. Interactive dashboard helps the user to understand the sales profit, sales performance along with the shipping cost across the world, how every worksheet interacts with other with a single filter. This dashboard will create a best user experience. Using tableau, the overall aim is achieved when compared to Power Bi.

The choropleth map part is much easier in the Power Bi when compared to tableau. It automatically fetches the location on behalf of the geographical variables we have. But in Tableau, If the location is not assigned properly, it will show null or unknown values and packages need to install for some geographic location. Other experience with Tableau is much easier and accessible when compared to Power Bi.

Future implementation includes working with the live dataset for the data analysis process.

## REFERENCE:

2. Dahl, D.W., Chattopadhyay, A. and Gorn, G.J. (2001) ‘The importance of visualisation in concept design’, *Design Studies*, 22(1), pp. 5–26. Available at: <https://doi.org/10.1016/s0142-694x(99)00028-9>.

3. Tim Cribbin, (2024) ‘CS5703 week 19 Teaching Material, Lab 3:Representation, Availabe at <https://brightspace.brunel.ac.uk/d2l/le/lessons/43662/topics/1515364>

4. Tim Cribbin, (2024) ‘CS5703 week 20 Teaching Material, Lab 4: Presentation’, Availabe at <https://brightspace.brunel.ac.uk/d2l/le/lessons/43662/topics/1520289>.

5. Tim Cribbin, (2024) ‘CS5703 week 20 Teaching Material, Lecture 2: Encoding for perception, Availabe at <https://brightspace.brunel.ac.uk/d2l/le/lessons/43662/topics/1508004>.

1. https://www.kaggle.com/datasets/laibaanwer/superstore-sales-dataset [↑](#footnote-ref-1)