# **Assignment**

# **On**

Global Store Data Analysis

*MSC Artificial Intelligence*

*Brunel University London*

CS5803

Module Code:

Data Visualisation

Module Title:

Visualisation Design Task

Assignment Title:

2021/2022

Academic Year:

# 

Submitted by:

Sookchand Harripersad

Student ID: 2121403

Table of Contents

[**Assignment** 1](https://d.docs.live.net/ada748aa0acc4af6/Documents/Brunel%20MSC%20AI/Term_2/Data%20Visualization/Course%20Work/Data%20Visualization/Data%20Visualization%20Coursework/Visualisation%20Design%20Task.docx#_Toc98833075)

[**On** 1](https://d.docs.live.net/ada748aa0acc4af6/Documents/Brunel%20MSC%20AI/Term_2/Data%20Visualization/Course%20Work/Data%20Visualization/Data%20Visualization%20Coursework/Visualisation%20Design%20Task.docx#_Toc98833076)

[1](#_Toc98833077)

[Introduction 3](#_Toc98833078)

[Persona Name: John Smith 3](#_Toc98833079)

[Data Dictionary 3](#_Toc98833080)

[Requirements 3](#_Toc98833081)

[Design 4](#_Toc98833082)

[Paper Landscape (design) 4](#_Toc98833083)

[Dictionary 5](#_Toc98833084)

[Revised Landscape 5](#_Toc98833085)

[Digital Prototype 6](#_Toc98833086)

[Implementation 7](#_Toc98833087)

[Percentage of revenue per region (fig. 1) 7](#_Toc98833088)

[Segment by Month (Fig 2) 7](#_Toc98833089)

[Regional quarterly profit (fig. 3) 7](#_Toc98833090)

[Discount vs Profit vs Sales by Product(Fig 4) 7](#_Toc98833091)

[State vs Profit (fig. 5) 7](#_Toc98833092)

[Average Measures Sheet (fig6) 7](#_Toc98833093)

[Dashboard (Fig 7) 8](#_Toc98833094)

[Walkthrough 8](#_Toc98833095)

[Tableau Walkthrough 8](#_Toc98833096)

[Tableau to Power BI implementation 13](#_Toc98833097)

[Reflection 19](#_Toc98833098)

[References 20](#_Toc98833099)

# Introduction

## Persona Name: John Smith

John’s parents ensured he had lots of exposure to the business. He has A focused Business Management graduate with a 2:2 degree, specialising in innovative methods to efficiently manage a business of any size and scope – from managing your own start-up business to working in a multinational organisation. Domain knowledge he ranked seven, technical knowledge seven, experience seven, and computer for business use five. These figures are ranked from one through ten and ten being the highest.

John was asked by his father to go to the USA and turn around the company. Knowing that I was completing my MSc Artificial Intelligence, he then called me up and handed me a data set containing the company« Global Store» sales and ask me to carry out a business analysis focusing on the US region. This solution would be for Johns’ personal use and not be as a presentation piece. I can tailor the solution to make use of things like tooltips and build in a lot of underlying information he can use.

The Global Superstore dataset is a collection of data on global retail sales, including information on products, customers, and sales. The dataset is sourced from the Global Superstore website (Kaggle Website: <https://www.kaggle.com/gauravtopre/global-superstore-dataset>), a leading retailer in the global market. The dimension of the original dataset is 24 x 51291 and the dimensions of the one I am proposing to use is 12 x 9750.

### Data Dictionary

Name US Global Superstore Dataset.

1. Order Date: The order date for each product – Description (Date Time)
2. Segment – The products are categorized into three separate business blocks (String).
3. City: The name of the city – Description (String).
4. State: The name of the state – Description (String).
5. Region: The name of the region – Description (String).
6. Country: The name of the country – Description (String).
7. Category: The category of the product – Description (String).
8. Sub-Category: The sub-category of the product – Description (String).
9. Sales: The total sales of the product – Description (Decimal).
10. Quantity: The number of units of the product sold – Description (Number)
11. Discount: The discount applied to the product – Description (Decimal).
12. Profit: The profit generated from the sale of the product – Description (Percent value).

Data description![Table

Description automatically generated]()

### Requirements

Johns Questions are:

1. Which Segment had the smallest revenue percent, how did it perform over time, 2014 which quarter had the highest negative value, and which products contributed to it? For this question, we have all the variables required to answer and these are - Segment (String), have Sales (decimal), order date (Year), and sub-categories (String).
2. For Idaho, Montana, Wyoming, North Dakota, South Dakota, Nebraska, Minnesota, and Iowa who has the highest and lowest profit, what product contributes to the success of the state with the higher profit, and how did it perform over time? - For this question, we have all the variables required to answer and these are – State (String), we have Profit (decimal), order date (Date).

I know that John has Tableau and all the requirements set out above can be met. I shall move forward with the project using Tableau. I can also utilize the tooltips on Tableau to add a further layer of information should he choose to drill down. The outcome of the project is to present an infographic solution that addresses the two concerns given below.

# Design

## Paper Landscape (design)

A picture containing diagram

Description automatically generated

## Dictionary

Shape, arrow

Description automatically generated

Looking at the paper landscape I realize that Johns’s question was not being answered. In short, I think John wants to know who is making money and who is not? Also, the dataset that is reported in the paper landscape is too large as it is giving trouble to load and move between visuals so I will reduce it some more. The data dictionary has the final version.

## Revised Landscape

A picture containing text, whiteboard

Description automatically generated

## Digital Prototype

![Graphical user interface

Description automatically generated]()

Itten’s qualitatively grounded contrast model formalized the concepts of warm and cool colors and postulated that tints (light colors) represent the brighter and better aspects of life, while shades (dark colors) represent the darker, sad, and negative forces (Bartram, Patra, and Stone, 2017). Losses or negative profit is bad and the color that equates to this is red, while green suggests things are good and making a profit is good. Hence Red to green will be my colour palette as to would add to what I am trying to communicate via diagrams and numbers. To make the diagram pop having red to green colour, I would use a black background and white font. Finally, the use of colour would be consistent where possible in all sheets.

Both Fig. 1 and Fig. 2 will be connected to a variable called Segment. The pie chart on each state will also hold sales by segment so it too shall share a colour with Fig. 1 and 2. Fig. 3, Fig. 4, and Fig 5 hold the profit variable, hence, the color scheme for all 3 will be red-green divergent.

Since he is asking me, I can assume that getting the company involved in any way is not a good move. Another constraint I might be facing is the limited analytical functionality in Tableau as compared to Rstudio. Since I am not going to present this solution in person to John, I need a tool that can automate the analysis have use involvement and be easy to walk through. The solution should allow John to get answers to his questions visually while looking at the various view without moving from one page to another.  
  
I think the Tableau report is well suited for the task. Tableau report requires data to be input into the software to generate graphs and charts.  
  
I know that John has Tableau and all the requirements set out above can be met. I shall move forward with the project using Tableau. I can also utilize the tooltips on Tableau to add a further layer of information should he choose to drill down. The outcome of the project is to present an infographic solution that addresses the two concerns given below. These requirements seem a bit vague ad what is necessary to have on the infographic solution still requires some fleshing out.

# Implementation

## Percentage of revenue per region (fig. 1)

To Answer question one, I need to be able to select the segment with the smallest revenue percent. I will use a pie chart for this first operation.

1. Region🡪 Column 🡪 sales 🡪 rows.
2. Show me🡪 Pie chart 🡪View 🡪 menu bar 🡪 Region 🡪Labels 🡪 Sum sales 🡪 Labels.
3. Sum sale 🡪 drop-down menu 🡪 quick table calculation 🡪 choose percent total.
4. On data pane 🡪 Calculated field 🡪 label axis “Zero axis”🡪 0
5. Drag zero axis unto row and repeat it.
6. On the second tab created “sum zero axis” 🡪 choose the dual axis.
7. Zero axis 🡪 show header
8. Format 🡪 Black. Rename pane “Percent revenue per region.” Format as per usual, instruction above.

## Segment by Month (Fig 2)

Fig1 is connected to Fig.2 and I will have to report how the segment performs over time.

1. Rows 🡪 Product margin %. Column 🡪 Order Date. Click on “+” to add another Order date.
2. Show me 🡪 Area chart 🡪 Segment 🡪 Colour
3. Drilling down further, he would want to know if these losses are seasonal or something else. This led to the solution for question nine “regional quarterly profit.” That information is encoded in figure four.

## Regional quarterly profit (fig. 3)

Fig.1 will be connected to fig. 2. I will have to search on the visual for 2014 for the highest negative profit

1. Rename 🡪 Regional Profit 🡪 Regional quarterly profit.
2. Background colour 🡪 black 🡪 formatting as before.
3. Region 🡪 Rows. Order 🡪 Column 🡪 Column click on the plus sign on order date.
4. Format 🡪 Profit 🡪 Sheet.
5. Sum 🡪 avg 🡪 Marks –Colour 🡪 red to green.

## Discount vs Profit vs Sales by Product(Fig 4)

1. Column 🡪 Sub-category. Profit 🡪 Rows 🡪 Sum 🡪 Average.
2. Discount 🡪 Size 🡪 Average Profit 🡪 Colour 🡪 red – Green divergent. Sakes 🡪 Labels.

## State vs Profit (fig. 5)

The steps I took to implement average measures for requirements one, two, and three (map chart, average, and profit).

1. State 🡪 Rows. Map layers 🡪 uncheck all. Profit 🡪 Color 🡪 red-green divergent.
2. Rows 🡪 Latitude 🡪 Ctrl Hold 🡪 the second latitude on Rows.
3. Sales 🡪 Size. --> Circle 🡪 Blue. Label 🡪 Marks. Rows 🡪 Latitude 🡪 Dual Axis.
4. Remove sales 🡪 marks 🡪 Category 🡪 Color 🡪 Circle 🡪 Pie 🡪 Sale 🡪 Angle 🡪 Average Sale

## Average Measures Sheet (fig6)

The steps I took to implement average measures for requirements. The solution should accurately describe what is being shown. Hence, the first page is the title of the page:

1. Open the Tableau 🡪 text file 🡪 click on sheet 1 and rename “Average Measures”.
2. Click on Format (formatting for all pages) 🡪 shading 🡪 black.
3. Click on the top left-hand corner 🡪 format ..Calibri, bold, 15 and white font(for all page all formatting follow this generally)
4. Click on Format 🡪 font 🡪 white.
5. Click on fields 🡪 measure values 🡪 font (15) 🡪 alignment (center) 🡪 currency (custom $).
6. Click on-axis 🡪 font (15).
7. Click on measures name 🡪 alignment(center) and repeat eh same for measures value.
8. Click on border 🡪 choose None for all variables.
9. On measure values 🡪 click on each drop-down arrow of each variable and change to average.
10. All requirements have been met.

## Dashboard (Fig 7)

After I finished designing the dashboard on the computer, it was just a matter of placing the bin in the correct place then dragging and dropping each page on the relevant bin.

1. Click on add dashboard 🡪 check show title 🡪 insert the name of the page and format put the font in white 🡪format entire page to black 🡪 drag and drop 3 horizontal bin one at a time so that page is separated into four 🡪 fit dimensions of the bins according to the design 🡪 drag and drop two horizontal bin one at a time on the second row. This should produce five bins. The first is for the title.
2. Second drag and drop average measure under the title” Global Store Fata Analysis.” 🡪 Drag and drop state vs profit in the third-row first column 🡪 drag and drop percentage vs Region next to state vs profit. Click on drop 🡪 click o drop-down arrow on filters 🡪 chose float and place them on the relative diagram.
3. Finally drag and drop regional profit unto the last row 🡪 click on enter view 🡪 fit entire view. Save project.

# Walkthrough

## Tableau Walkthrough

In Question one, it requires you to the segment with the smallest revenue and for that, we have Fig. 1

Fig.1

![Graphical user interface, application

Description automatically generated]()

After selecting the smallest segment, you would need to see how it is performing over time and Fig. 2 visualize this.

Fig. 2

![A picture containing background pattern

Description automatically generated]()

Regarding the identification of the highest negative quarter in 2014, you need Fig 3 visual to do this.

![Graphical user interface, text

Description automatically generated]()

To tell which product was most responsible for the negative profit you only need to look at Fig. 4. Discount is encoded by size, profit by colour, and sales in figures and product by name.

Fig. 4

![Chart, treemap chart

Description automatically generated]()

We need to encode the state into a map. This would allow us to choose the states relevant to our question and see any relationships with the rest of the visual. Fig. 5 was developed with this in mind.

Fig. 5

![Map

Description automatically generated]()

The title announces what this report entails and so we have Fig. 6.

Fig. 6

Text

Description automatically generated

Fig. 7

![Graphical user interface

Description automatically generated]()

Let’s see how the visualization work on answering Johns’ question:

1. Which Segment had the smallest revenue percent, how did it perform over time, 2014 which quarter had the highest negative value, and which products contributed to it?

* On the chart of Percent of Revenue by Segment click on the smallest revenue segment (Fig. 8)🡪 then look at the Regional quarterly profit chart and click on the highest negative quarter for 2014 (Fig. 9)🡪 answer for the product is given in Discount vs profit vs sales by-products.

Fig.8

![Graphical user interface, chart

Description automatically generated]()

Fig. 9

![Graphical user interface

Description automatically generated]()

1. For Idaho, Montana, Wyoming, North Dakota, South Dakota, Nebraska, Minnesota, and Iowa who has the highest and lowest profit, what product contributes to the success of the state with the higher profit, and how did it perform over time?

* First brush over the area to capture only the countries on the “Average profit by state. Sales by segment visual 🡪 color code can answer highest and lowest with tooltips providing further assistance 🡪 click on the pie chart on the highest profit state 🡪 check Discount vs profit vs sales by product visual for answer🡪 check segment by month for performance over time.

Fig. 10![Graphical user interface, application

Description automatically generated]()

Fog. 11

![Graphical user interface, application, website

Description automatically generated]()

## Tableau to Power BI implementation

**Steps:**

* First load the Superstore dataset into PowerBI

A screenshot of a computer

Description automatically generated

* Next Click on Load Button so that the Data gets loaded to PowerBI. Now you can start creating the visuals.

Graphical user interface, application

Description automatically generated

* Click on View and change the View to the Dark Colour.

Graphical user interface, application

Description automatically generated

* After this Add the Text Box and Write Down the Title of Dashboard inside that.

Graphical user interface

Description automatically generated

Graphical user interface, application

Description automatically generated

* Now Add the first Sheet Donut Chart for Category wise Revenue. Drag the Segment to Legend and Sales to the Values.

Graphical user interface, application

Description automatically generated

* Now add the Map Chart from the set of visuals available. Add the State to Location, Segment to Legend and Sales to Value. Also Click on this Sales and choose Average instead of Sum.

A screenshot of a computer

Description automatically generated

* Now add the Tree Maps from the Set of Visuals available in the right side. Add Sub-Category to Group, Discount to Values and Profit & Sales to the Tooltips. Select the level of aggregation as Average for Profit and Discount.

A screenshot of a computer

Description automatically generated

* Now Go to the Data Tab. Create a New Measure to Find the Margin by Formula

**Product Gross Margin = SUM(Sheet1[Profit])/CALCULATE(SUM(Sheet1[Sales]),ALL(Sheet1))**

Table

Description automatically generated

* Now Add the Area Chart from right side. Add the Order Date to X-Axis, Segment to Legend and the Product Gross Margin to the Values.

Chart

Description automatically generated

Now the last visual is adding the Matrix from the right side. After adding Matrix drag Regions to Rows, Order Date to Columns and Profit to the Values. The aggregation of profit must be changed to Average instead of Sum by just clicking this

Graphical user interface, chart

Description automatically generated

Now the Dashboard is Fully Ready and Could be Saved by clicking Ctrl+S

# Reflection

The strengths for a data analysis done on Tableau would include the ability to easily create graphs and charts, as well as the ability to filter data. The weaknesses for a data analysis done on Tableau would include the lack of flexibility in terms of data manipulation and the lack of ability to run complex statistical analyses. The strengths of a data analysis done on Rstudio would include the ability to manipulate data and to run complex statistical analyses. The weaknesses for a data analysis done on Rstudio would include the lack of flexibility in terms of creating graphs and charts, as well as the lack of ability to filter data. I have gained respect for both technologies and see the need to use them where they best fit and wherever I can.

There is no one-size-fits-all answer to this question, as the best data visualization tool for a given organization depends on that organization's specific needs and preferences. However, some general advantages and disadvantages of Tableau vs Power BI can be identified. Tableau is generally considered to be more user-friendly and intuitive than Power BI, making it a good choice for organizations whose employees are not highly technical. Tableau is also better suited for visualizing data in real-time, which can be important for some organizations. However, Power BI is generally more powerful and flexible than Tableau, making it a better choice for organizations with more complex data visualization needs.

The ability to easily create graphs and charts, as well as the ability to filter data, are among the strengths of a Tableau data analysis. The shortcomings of a Tableau data analysis would be a lack of flexibility in data manipulation and the inability to run complex statistical analyses. The ability to manipulate data and run complex statistical analyses would be strengths for a data analysis done in Rstudio. The shortcomings of a Rstudio data analysis would be a lack of flexibility in terms of creating graphs and charts, as well as a lack of ability to filter data. I've grown to appreciate both technologies and see the value in implementing them.

Regarding Power BI versus Tableau, there is no one-size-fits-all answer to this question because the best data visualization tool for a given organization is determined by the organization's unique needs and preferences. However, there are some general advantages and disadvantages of Tableau vs Power BI. Tableau is generally thought to be more user-friendly and intuitive than Power BI, making it a good choice for organizations with non-technical employees. Tableau is also better suited for real-time data visualization, which is important for some organizations. Power BI, on the other hand, is more powerful and flexible than Tableau in general, making it a better choice for organizations with more complex data visualization needs.

The interest in text visualization and visual text analytics has been increasing for the last ten years(Kucher and Kerren, 2015). My goal coming out of this module is to improve my skills in text visualization as I hope to become an NLP Researcher.

# References

1. Analytics with Nags (2021) *Tableau to Power BI Migration - Power BI Project Dashboard (5/30)*. Available at: https://www.youtube.com/watch?v=0eAPLogu0iI (Accessed: 13 March 2022).
2. Bartram, L., Patra, A. and Stone, M. (2017) ‘Affective Color in Visualization’, in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI ’17), pp. 1364–1374. doi:10.1145/3025453.3026041.
3. Chabot, C. (2009) ‘Demystifying Visual Analytics’, *IEEE Computer Graphics and Applications*, 29(2), pp. 84–87. doi:10.1109/MCG.2009.23.
4. Keim, D.A., Mansmann, F. and Thomas, J. (2010) ‘Visual analytics: how much visualization and how much analytics?’, *ACM SIGKDD Explorations Newsletter*, 11(2), pp. 5–8. doi:10.1145/1809400.1809403.
5. Kucher, K. and Kerren, A. (2015) ‘Text visualization techniques: Taxonomy, visual survey, and community insights’, in *2015 IEEE Pacific Visualization Symposium (PacificVis)*. *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 117–121. doi:10.1109/PACIFICVIS.2015.7156366.
6. *YouTube* (no date). Available at: https://accounts.google.com/signin/v2/identifier?service=youtube&uilel=3&passive=true&continue=https%3A%2F%2Fwww.youtube.com%2Fsignin%3Faction\_handle\_signin%3Dtrue%26app%3Ddesktop%26hl%3Den%26next%3Dhttps%253A%252F%252Fwww.youtube.com%252Fwatch%253Fv%253D0eAPLogu0iI&hl=en&gae=cb-23983171&flowName=GlifWebSignIn&flowEntry=ServiceLogin (Accessed: 13 March 2022).