Amil Kazimoglu

2449223

CS 5803

Data Visualization Report

Introduction:

Data The dataset used in this research is titled “Auto Sales Data” and was obtained from https://www.kaggle.com/datasets/ddosad/auto-sales-data . It contains details of sales of transportation machines such as planes, trains, ships, and vintage cars. This project aims to demonstrate the power of data visualization in supporting decision-making and making predictions accordingly. As noted by Ingram (2024), applied sales analytics plays a crucial role in identifying market trends and shaping strategy, including making informed decisions and forecasts plays a crucial role in identifying market trends and shaping strategy like making decision and predictions.

Data Dictionary:

This is a Autosales data, which represents sales of some transportation machines. The data dictionary below shows the key original variables along with the new variables shown in bold and red:

Name	Description	Domain
ORDERNUMBER	Unique identifier for the order	Integer
QUANTITYORDERED	Number of items ordered	Integer
PRICEEACH	Price per item	Float
ORDERLINENUMBER	Line number within the order	Integer
SALES	Total sales value	Float
ORDERDATE	Date of the order	Date
DAYS_SINCE_LASTORDER	Days since customer's last order	Integer
STATUS	Current status of the order	Nominal
PRODUCTLINE	Category of the product	Nominal
MSRP	Manufacturer’s suggested retail price	Float
PRODUCTCODE	Unique product identifier	Nominal
CUSTOMERNAME	Name of the customer	Nominal
PHONE	Customer's phone number	Nominal
ADDRESSLINE1	Primary address line	Nominal
CITY	City of the customer	Nominal
POSTALCODE	Postal code of the address	Nominal
COUNTRY	Country of the customer	Nominal
CONTACTLASTNAME	Contact’s last name	Nominal
CONTACTFIRSTNAME	Contact’s first name	Nominal
DEALSIZE	Size category of the deal	Nominal (e.g., S/M/L)
ESTIMATED PROFIT	Total profit	Float
BAZAAR COST	Total cost of machine	Float
Profit per unit	Profit Per machine	Float

Persona And Questions:

Manager for a transportation company, who is trying to find out which machine to invest and where to invest for maximum profit.

Simple questions:

1 – Which Country has ordered most of the transportation machines?

2- Which transportation machine generates most profit per unit?

3 - What is the relationship between quantity sold and estimated profit across transportation machines?

Complex questions:

How profitable were the transportation machines sold in the USA in 2020 under large deal sizes, when measured against the non-profit zone?

Requirements:

1. To answer Q1, the user must perform a ranking task to identify which country has ordered the most transportation machines.

* A tree map is used to visualise total QuantityOrdered per Country, allowing users to quickly compare the volume of orders between countries based on the size of each block.

2. To answer Q2 the user must perform again ranking task to which transportation machine has the most profit per unit when it sold.

* Bar plot of average Profit per unit by ProductLine

3. To answer Q3 the user needs to see the correlation between how many machines has been sold (QuantityOrdered) and total profit from variety of transportation machines (ESTIMATED PROFIT).

* Scatterplot will be used, where each dot or point will be the type of transportation machine (ProductLine), plotted by x axis QuantityOrdered and y axis ESTIMATED PROFIT.

4. To answer Q4 the user needs to find out the profitability of transportation machines sold in specific country, specific deal size and a specific year, and needs to determine if that machines fall above or below the break-even threshold.

***** A scatterplot of QuantityOrdered vs Estimated Profit is used to visualise the relationship between volume sold and total profit.

***** The line represents the break-even point, to help the user to see profitable and unprofitable products.

***** Filters for Dealsize and year are included to narrow the analyses.

***** Tree Map can be used as a filter to choose the specific country**.**

***** Bar chart showing AVG (Profit per Unit) by ProductLine supports to find out which machine the invest to maximize the profit. Basically, it is helping for more performance.

Design:

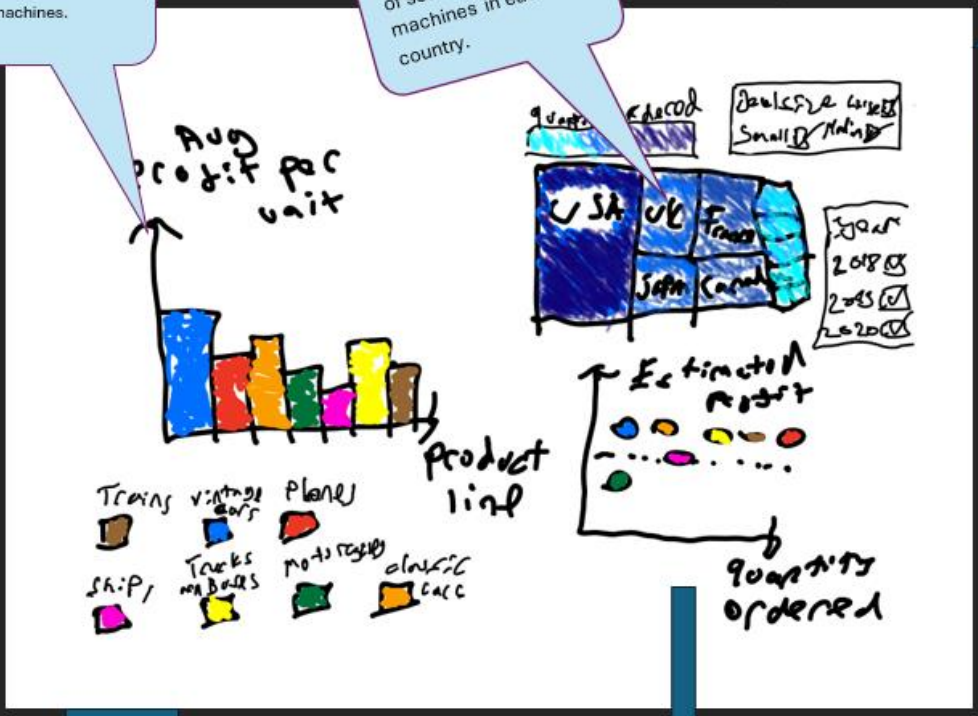
Paper landscape and Final Implementation:

Paper Landscape:

2. See the differences between profit per unit of all transportation machines.

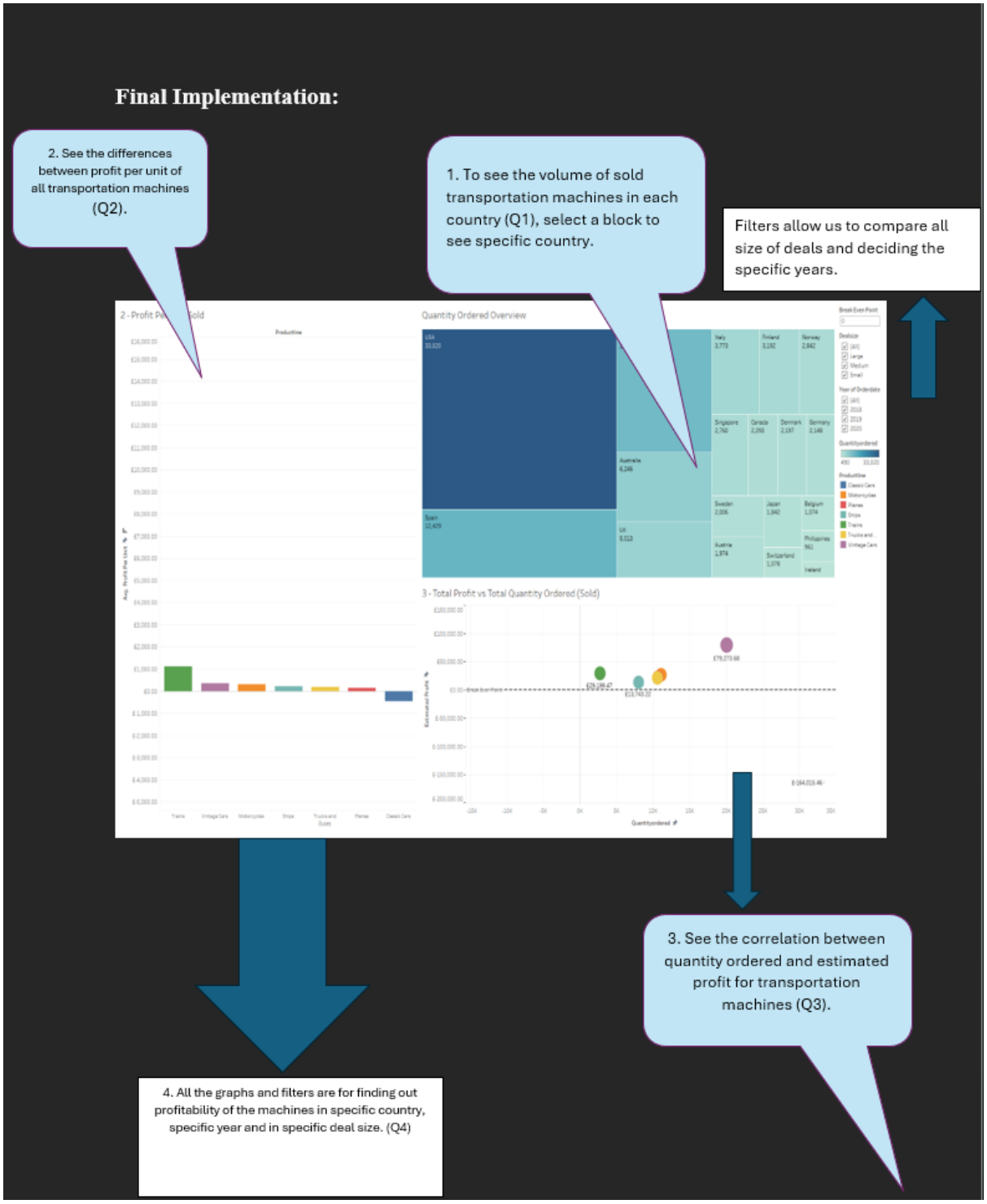
1. To see the volume of sold transportation machines in each country.

Filters allow us to compare all size of deals and deciding the specific years.



3. See the correlation between quantity orderd and estimated profit for transportation machines (Q3).

4. All the graphs and filters are for finding out profitability of the machines in specific country, specific year and in specific deal size. (Q4)



From the pictures above, the theoretical design and the real design are nearly the same. However, before reaching that point, there were two feedback sessions for my initial design — before the paper landscape and after the paper landscape. Before the paper landscape, the first design had two bar charts and one scatter plot, which was not sufficient for the given task. This problem occurred because of the approach. The approach that was used involved discovering the data and Tableau before developing the questions and persona, to create graphs and parameters to see who can use that data, what these graphs are giving to the user, and which graphs can be used in real-life problems. After this approach, the feedback changed one of the simple questions, rephrased the complex question with the stronger word "how," and changed one of the graphs to make the dashboard more interactive and useful for the filters that were being used. In the second feedback session, while showing the theoretical approach and implemented dashboard, everything was good and sufficient; only small changes were made, such as reducing axis ranges. These changes were actually very similar to what was learned

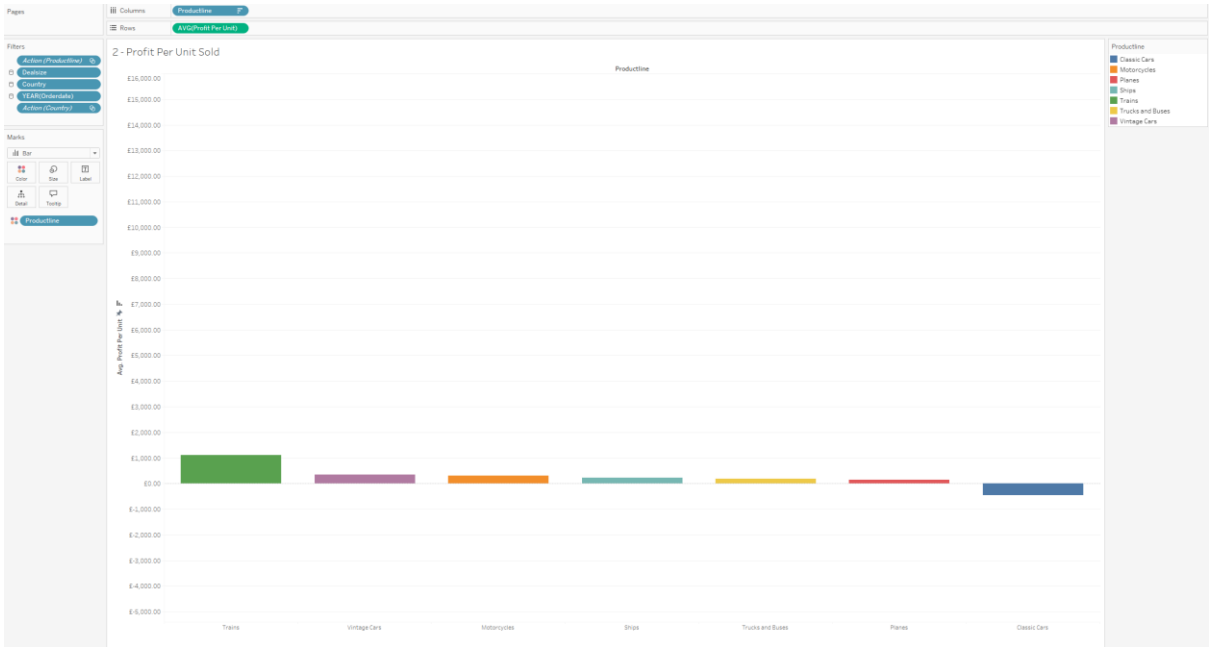
in the lecture, which is Few's (2013) principle that dashboard components must be user-centred and functionally grouped. With the graph change — where the bar chart was turned into a tree map — one of the filters, which was country, was removed and added to the graph to make it more user-centred, functional, and sufficient.

Implementation:

Sheet 1:

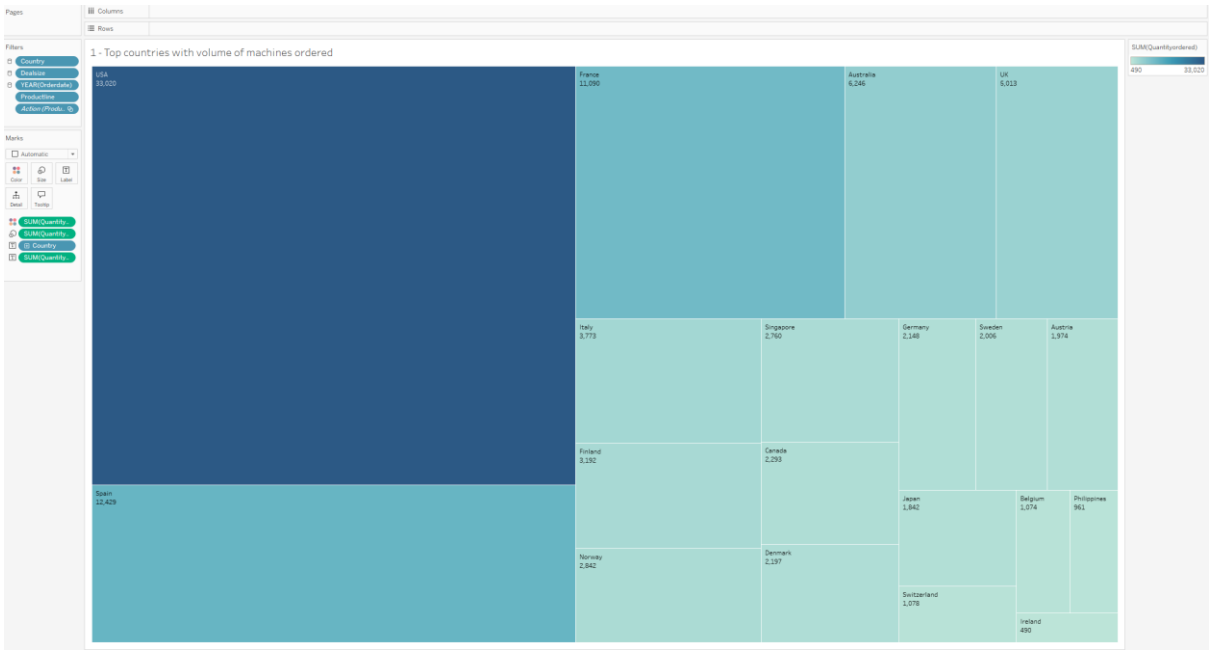
To understand the relationship between total profit and total quantity ordered of machines, created a scatter plot. To achieve this, firstly, need to create total profit parameter which is estimated profit in this figure. Then need to compare it with quantity ordered, to achieve this simply drop one of them in column and another to rows, then to find out which is which transportation machine simply label them with colour, then to understand the proportion of profit label them with size, lastly to understand which is profitable or not create a simple line parameter which is break-even point and simply give value of 0 to see profitable regions on the scatter plot. Usage of scatter plot in this question is very important because, according to Mackinlay (1986) ranking of graphical encodings, where position is the most important element here to see the correlation between x and y, thus usage of scatter plot is advised by Mackinlay (1986) in that situation.

Sheet 2:



To understand the relationship between total profit and total quantity ordered of machines, created a scatter plot. To achieve this, firstly, need to create total profit parameter, which is estimated profit in this figure. Then need to compare it with quantity ordered, to achieve this simply drop one of them in columns and another to rows, then to find out which is which transportation machine simply label them with colour, then to understand the proportion of profit label them with size, lastly to understand which is profitable or not, create a simple line parameter which is break-even point and simply give value of 0 to see profitable regions on the scatter plot. Usage of scatter plot in this question is very important because, according to Mackinlay (1986) ranking of graphical encodings, where position is the most important element here to see the correlation between x and y, thus usage of scatter plot is advised by Mackinlay (1986) in that situation.

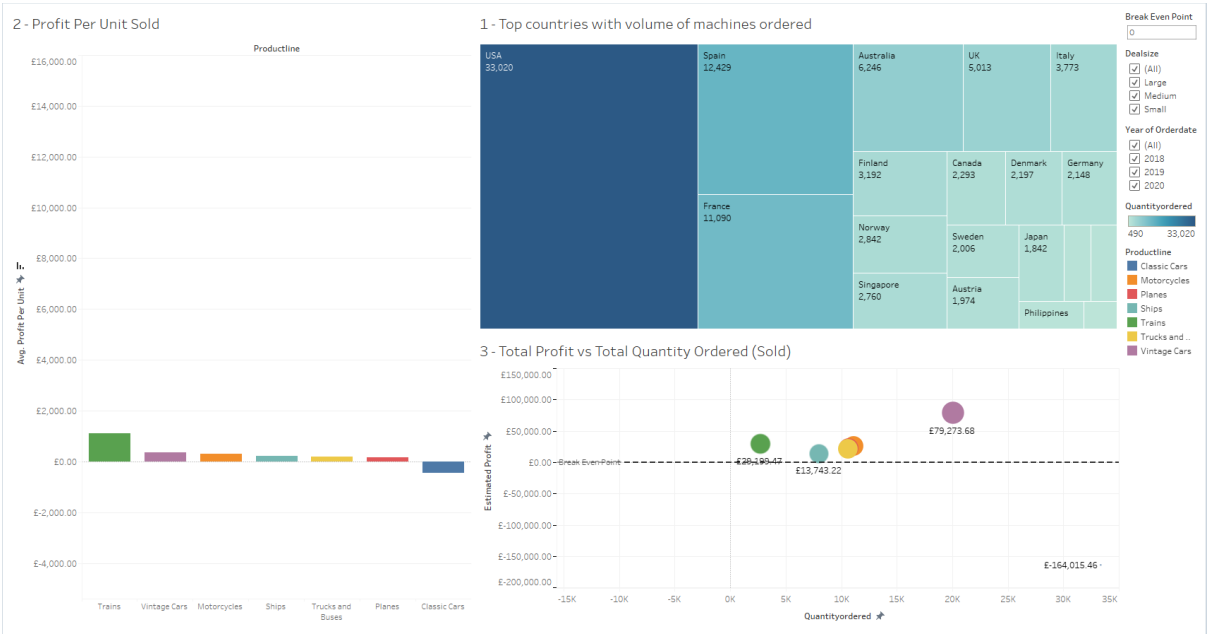
Sheet 3:



For seeing the difference in volume of total ordered transportation machines in different countries, created a tree map. Simply used colour encoding for different volumes of quantity ordered, darker blue to softer blue to show the difference between the orders. Also used size encoding as well to show difference of orders as well, to make user differentiate more easily. Other than that, showed the total orders at each square to show the

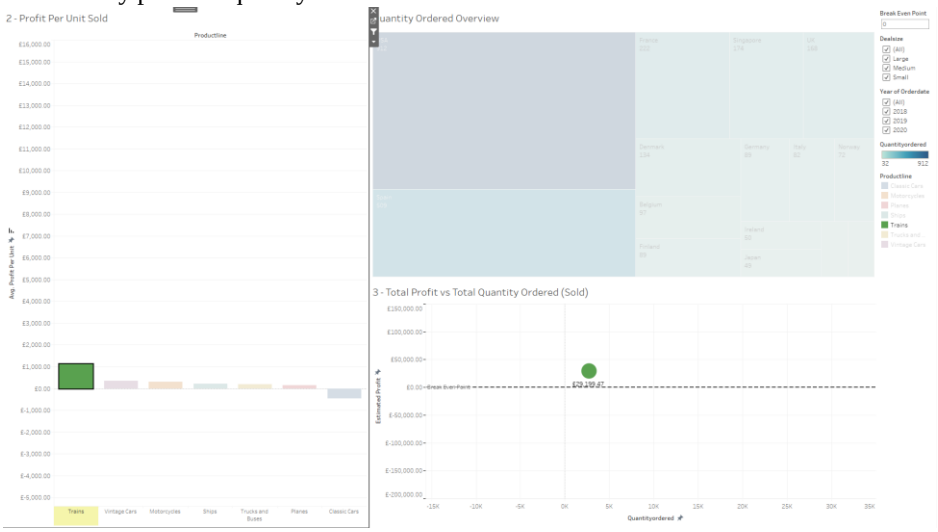
real deal to user, which makes sense to them to differentiate volumes with colour and size to see the relationship between real numbers and encoding. This idea also discussed in the lecture Week 19, according to Cleveland and McGill's work on graphical perception, noting that area is a weaker encoding channel compared to position and length (CS5703, Week 19 slides).

Dashboard:

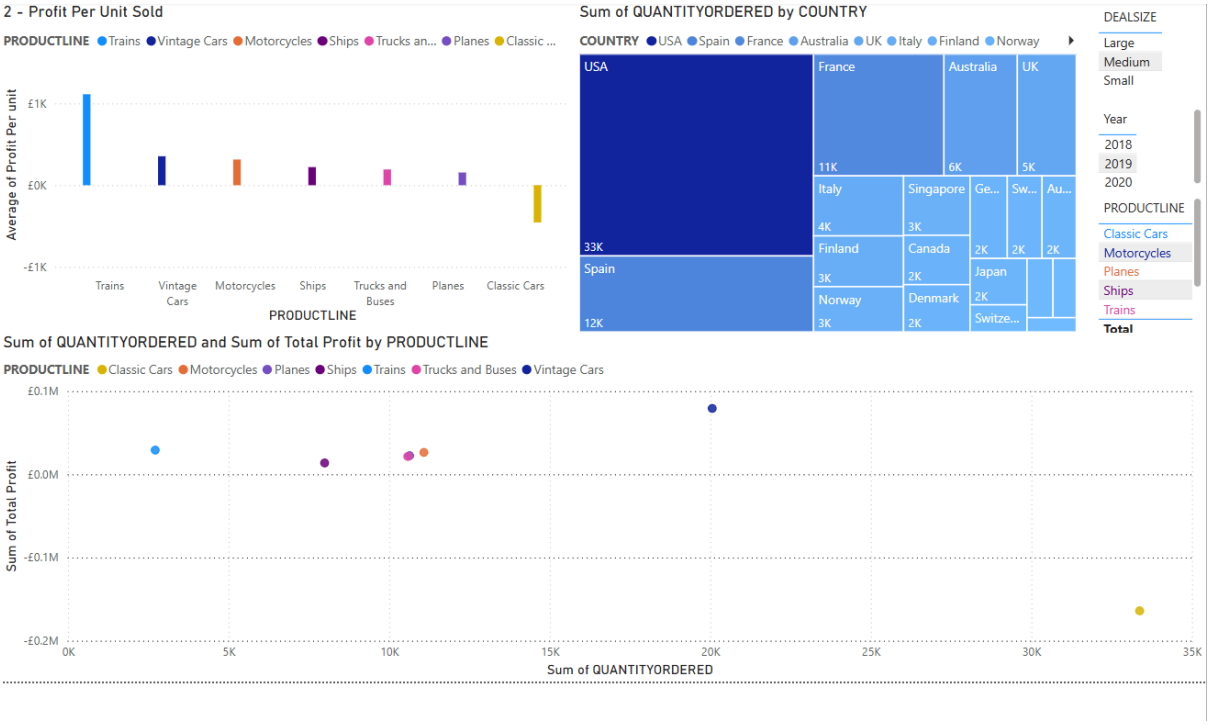


Lastly, to create a dashboard like that and to answer the complex question, merged 3 graphs in the dashboard screen, then added 2 filters like deal size and year of order date, used them to affect all the graphs and created an action which is an advanced Tableau tactic to make graphs more interactive and user-friendly, made all of them affect each other when clicking values in the graph and filters on the right side of them. This will give a huge advantage to the user to see the answer easily and interact with the graphs and filters by just clicking on them rather than a click 1 by 1 effect**.**

Additionally, a **Highlight Action** was implemented to further enhance user interaction. This feature allows users to click on a specific **ProductLine** in any graph, automatically highlighting related data points across all charts while dimming unrelated ones. This makes it easier for users to focus on relevant information and identify patterns quickly. This feature also 1 of the advanced tableau tactic.



Power BI vs Tableau:



In Power BI, there are some different steps compared to Tableau. First, it creates a column rather than a parameter like Tableau. Creating filters is also much more difficult than in Tableau, as you must manually drag the filter or value into the dashboard. Additionally, creating a parameter such as a break-even point is more complicated in Power BI; you must manually draw it, size it, and place it on the graph at the 0 value, whereas Tableau does this automatically within the graph.

While using Power BI, many tasks are handled manually, making it simpler but less dynamic. In a real-life scenario, if a manager needs to quickly create graphs and present them to a CEO or CFO within an hour, Power BI can be helpful due to its simplicity and manual approach. However, if the manager has a week to prepare and wants to deliver a more professional and interactive presentation, Tableau is much more beneficial. With its actions, filters, labels, functions, and parameters, Tableau offers a more complex but significantly more user-friendly experience, providing better visualisations for users to understand data and insights effectively.

Walkthrough:

Question 1 and Answer:

Question 1: Which country has ordered the most transportation machines?

To answer this question, a tree map was created to visualise all countries. First, it was necessary to sort the QuantityOrdered parameter in descending order. By observing the volume represented by the size, the colour intensity of each country, and the label showing the quantity ordered, it became easy to identify the country with the highest number of orders. The largest, darkest, and most prominent area on the tree map indicated the answer — USA, with 33,020 ordered transportation machines.

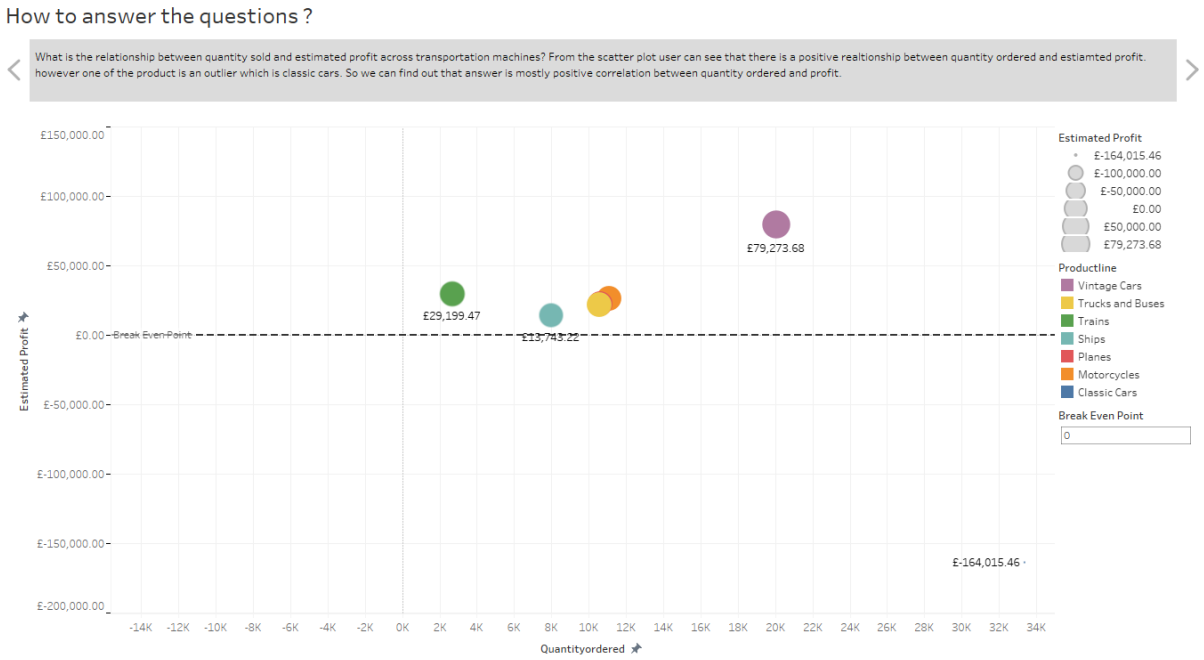
Question 2 and Answer:

Question 2: Which transportation machine generates the most profit per unit?

To answer this question, a bar chart was created to rank all transportation machines. The average profit per unit parameter was sorted in descending order. To differentiate the transportation machines, colour encoding was used. Additionally, the size of the bars helped to visualise the differences in profit, providing a clear ranking

for the user. By selecting the first and largest bar, the answer was identified as trains, which generate the most profit per unit.

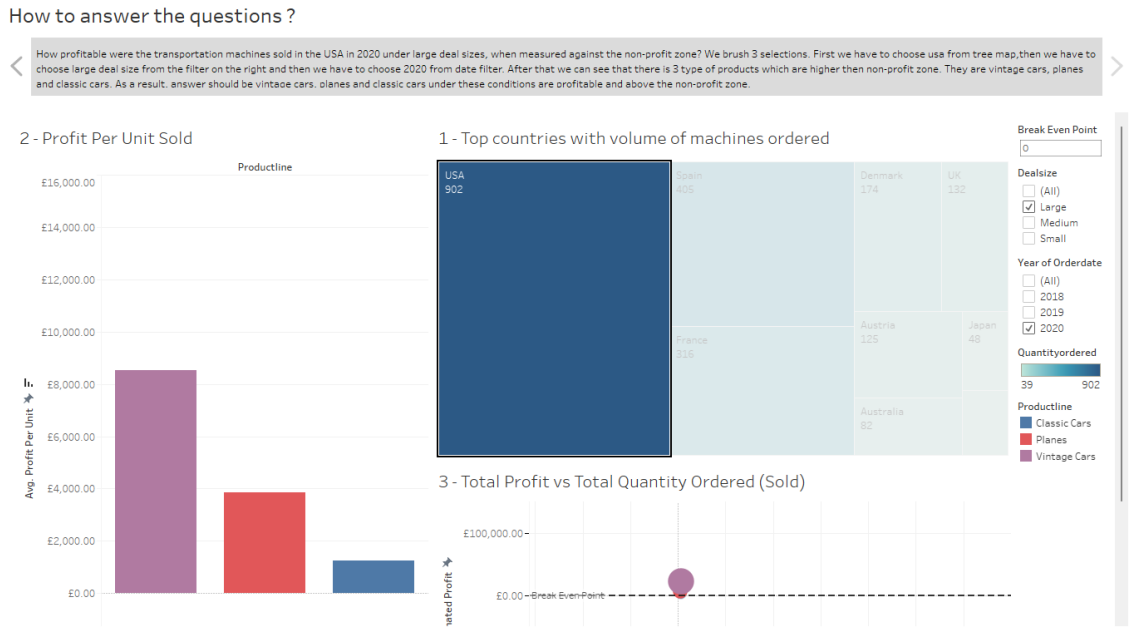
Question 3 and Answer:



Question 3: What is the relationship between quantity sold and estimated profit across transportation machines?

To find the answer to this question, a scatter plot was created, as it is suitable for identifying relationships or correlations. By observing the pattern of the circles, it was possible to see whether estimated profit increases or decreases in line with quantity ordered. If both variables increase or decrease together, it can be concluded that there is a positive relationship. However, one outlier can be observed — classic cars. Despite this single outlier, most transportation machines follow the same pattern, so the conclusion is that there is generally a positive relationship between quantity sold and estimated profit.

Question 4 and Answer:



Question 4: How profitable were the transportation machines sold in the USA in 2020 under large deal sizes, when measured against the non-profit zone?

This is the complex question. To find the answer, it was necessary to perform more than one interaction on the dashboard. First, the country needed to be selected from the tree map — USA — which triggered changes across the dashboard. Then, the date filter was used to select 2020. Lastly, the deal size filter was applied, selecting large. By observing the third graph, vintage cars, planes, and classic cars were above the break-even point, which represents the non-profit zone. Additionally, by reviewing the second graph to confirm the profitability per unit sold under these conditions, it can be concluded that vintage cars, planes, and classic cars are profitable in this scenario. That is the answer.

Discussion:

From the research and project, found that Tableau is a very powerful and useful data visualization tool for explaining data. I also learned that the graphs created in this project can be very useful in real-life problems to make assumptions — for example, using a scatter plot to compare the relationship between two values and make predictions about them for the future, draw conclusions, or take actions accordingly.

Other than that, I observed that even though Tableau is a powerful tool for data visualization, it has some limitations — such as dashboard sizing issues. This seems to be a compatibility issue between Windows and Tableau. Although my computer has a 1920×1080 resolution, Tableau created a scroll bar on my dashboard, which was unusual. My system indicated that my resolution was 1920×1080, but Tableau did not interpret it that way. To solve this problem, I used a higher-resolution monitor to test if the dashboard would fit correctly at 1920×1080, and it did.

However, Tableau also has many strengths. I discovered that while using Power BI, it has several functionalities that make the process easier and more automated, especially in filter creation, usage, and applying actions to build a powerful dashboard. The "Show Me" page, story points, and the ability to create stories are very powerful features in Tableau, which I came to appreciate even more while using Power BI.

Looking ahead, I aim to learn **Tableau Server**, as I assume many companies use it to publish and share dashboards internally. I believe gaining this knowledge will improve my professional readiness and allow me to showcase my work more effectively to potential employers.

Conclusion:

In conclusion, from this project, a person who is a transportation sales manager can decide what to invest in from the dashboards, stories, and sheets. This gives an idea of how Tableau can be used to solve real-life problems and provides huge prediction and decision-making power to the user.

To solve the questions, rather than first creating the persona or questions, playing with the data, exploring it, and using the software gave me a huge advantage in thinking about what to create and how to solve it. This shows how I approached this project from backwards to forwards, which suits the graphs and data because from past trends we can predict and decide what to do for real-life problems. Thanks to Tableau, this approach worked very well because interactive graphs, filters, labels, dashboards, and story points gave me a better idea of what parameters I could create from the data, which person could use this data, and what to investigate so that this data could help that person. That’s why this approach in Tableau helped me a lot to figure out everything.

From the problems created, I found out that the most profitable transportation machine per unit was trains, there is a relationship between quantity and profitability, the USA is the biggest bazaar for transportation machines, and lastly, in 2020, in the USA, large deal sizes gave a hint about what to invest in — vintage cars, planes, and classic cars — because they exceed the non-profit zone. From the solutions above, a sales manager can now predict and decide what to invest in for the future to gain the highest profit for his or her company, thanks to data visualization and math.

References:

Brunel University London (2025) CS5703: Encoding for Perception. [Lecture notes] Delivered by Dr Timothy Cribbin, Week 19.

Brunel University London (2025) CS5703: Interaction. [Lecture notes] Delivered by Dr Timothy Cribbin, Week 22.

Brunel University London (2025) CS5703: Presentation. [Lecture notes] Delivered by Dr Timothy Cribbin, Week 21.

Brunel University London (2025) CS5703: Representation. [Lecture notes] Delivered by Dr Timothy Cribbin, Week 20.

Cleveland, W.S. and McGill, R. (1984) 'Graphical perception: Theory, experimentation, and application to the development of graphical methods', Journal of the American Statistical Association, 79(387), pp.531–554.

Few, S. (2013) Information dashboard design: Displaying data for at-a-glance monitoring. 2nd edn. Burlingame, CA: Analytics Press.

Ingram, K.L. (2024) 'Applied sales predictive analytics for business development', Applied Business: Issues & Solutions, (1), pp.11–16.

Mackinlay, J. (1986) 'Automating the design of graphical presentations of relational information', ACM Transactions on Graphics, 5(2), pp.110–141.

Tableau Software (2023). Tableau Desktop Product Help. [online] Available at: https://www.tableau.com/support/help