

COMP-1800-M01- Data Visualisation

2024-25 Term 2

Task Report

Ganesh Ramakrishnan

001447349

MSc. Data Science

Table of Contents

1. Introduction.....	1
2. Findings Discussion	2
2.1. Store Segmentation by Customer Volume	2
2.2. Interactive Rolling Average Plot.....	3
2.3. Box Plot for Customer Variation in High Volume Stores.....	4
2.4. Correlation Heat Map for High Volume Store.	5
2.5. Correlation Heat Map for Medium Volume Store.	6
2.6. Interactive Grouped Bar Chart of Store Metrics.....	7
2.7. Combined Radar Plot: High and Medium Volume Stores	8
2.8. Daily Customer Trends in Very Low Volume Stores.....	9
3. Critical Review	10
4. Conclusion	11
Appendix A: Declaration of AI Use	12

Table of Figures

Figure 1 Total Customers for each Store (Year 2023)	2
Figure 2 High vs Medium Volume Interactive Rolling Avg Line Plot	3
Figure 3 High Volume Box Plot	4
Figure 4 High Volume Store Correlation Heat Map	5
Figure 5 Medium Volume Store Correlation Heat Map	6
Figure 6 Interactive Normalized Store Metric Bar Plot	7
Figure 7 High vs Medium Volume Store Metrics Combined Radar Plot	8
Figure 8 Very Low Volume Store Line Plot	9

1. Introduction

Data Visualization is the process of visualizing complex patterns of numbers which are either stored as information from an outside source or from within. To be able to draw an insight on the trends that this data creates and be able to portray them in the form of shapes that are both intuitive and accessible irrespective of the technical capacity of the next person. Visualization helps understand complex relations that exist in a group in a more meaningful manner.

Effective visualizations in a business setup lead to a more robust and high level decision making approach that yields a positive impact on the overall business through performance indicators, identifying anomalies and comparing different entities. It helps turn static information into a more visual representation which is helpful in the context mentioned earlier.

This coursework will focus on the overall footfall seen for over 40 stores in the business, see how they perform and to be able to draw insights from the data shared through visualization techniques learnt throughout the term. The goal is to be able to convey to the non technical shareholders what was noticed out of the data and prepare a narrative that will help them make critical business decisions to further improve.

2. Findings Discussion

2.1. Store Segmentation by Customer Volume

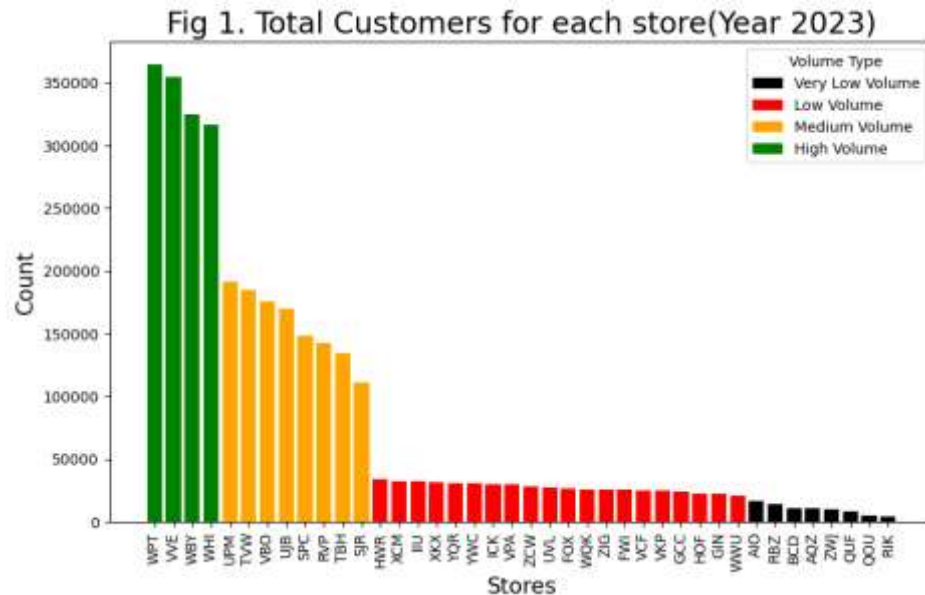


Figure 1 Total Customers for each Store (Year 2023)

We begin with Figure 1 where it shows us the total number of customers that have visited each of the 40 stores over the course of 2023, it shows a clear difference in customer traffic that then required a segmentation based on volumes such as high, medium, low and very low volume stores.

These stores were classified visually because of the stark difference in customer interaction between them in the bar chart. To which, each segment was then given a colour to help visualize them in the chart to enhance its readability and specify the distribution.

As seen in the chart, high and medium volume stores control the overall customer engagement for the business while the low and very low volume stores contribute significantly less. This helps us to maintain focus for the remainder of the analysis, on the high and medium volume stores as that is where majority of the business impact lies.

2.2. Interactive Rolling Average Plot



Figure 2 High vs Medium Volume Interactive Rolling Avg Line Plot

We now move onto the 2nd Figure, which shows an interactive rolling average plot between all the high and medium volume plots through out the year of 2023. Each line represents an individual store on which to smoothen the fluctuations, a 14-day rolling average has been applied to highlight the trends that each store goes through.

Interactive plots allow the users to be able to hover and zoom into particular dates to watch customer interaction trends without requiring to modify the code behind it multiple times to get the desired result.

Notably, we can see that the high volume store WHI goes through a period of seasonality with its regular rises and dips, right during spring, summer and later festive periods. This suggests the predictability of when the store appears to be busy and when not. Although not as much, the stores WPT and WBY appears to have rises and dips of its own during various parts of the year.

However, medium volume stores such as UPM, VBO, UJB and SPC appear to have a smoother line with an upward trajectory showcasing the fact that these stores see constant customer footfall irrespective of the time period throughout the year. These seasonality trends help us understand which of the stores require scheduling and targeted financial promotions to be able to boost its performance.

2.3. Box Plot for Customer Variation in High Volume Stores

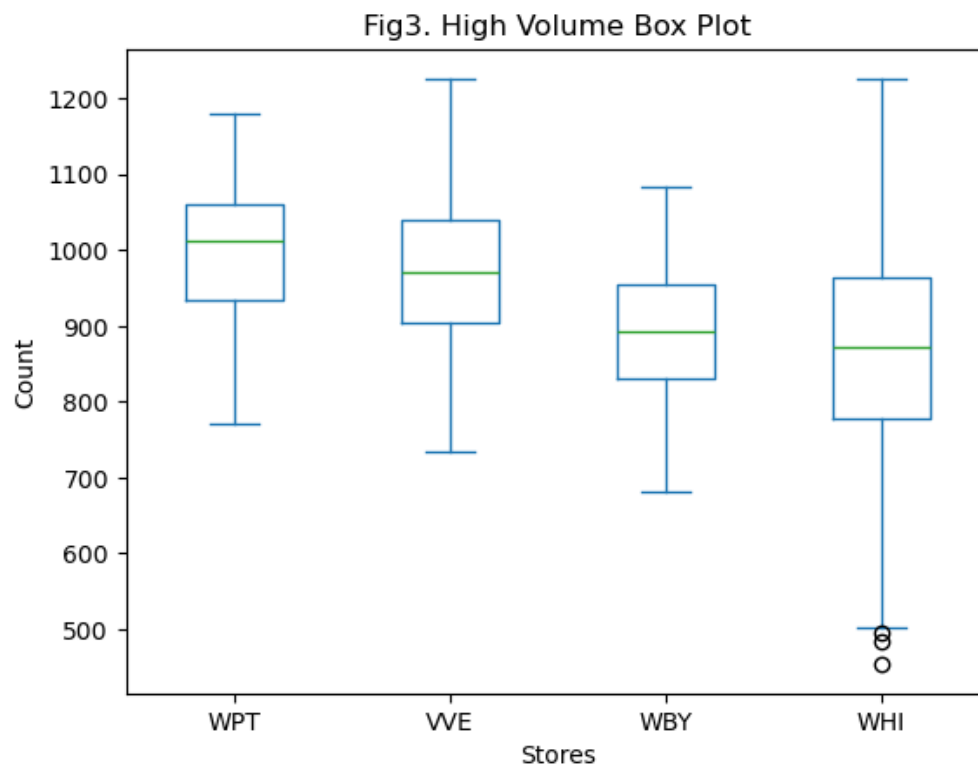


Figure 3 High Volume Box Plot

Figure 3 is a high volume box plot. This is used to add more information on the seasonal impact as seen and mentioned in figure 2. The box plot helps visualize the median for each of the stores. As discussed earlier, we can see that the stores WPT and WBY show consistent levels of customer inflow throughout the year and to a certain extent we can include store VVE in it as well.

Whereas, WHI which had shown us the varying rises and dips, can be seen to have a wider spread of customers coming to the store throughout the year. It also has multiple outliers supporting the observation made earlier across the seasons.

The figure was chosen to add context to the seasonal impact on the stores and help differentiate between the others to show which ones are high performing and which ones have differing outcomes which helps in forecasting and planning.

2.4. Correlation Heat Map for High Volume Store.

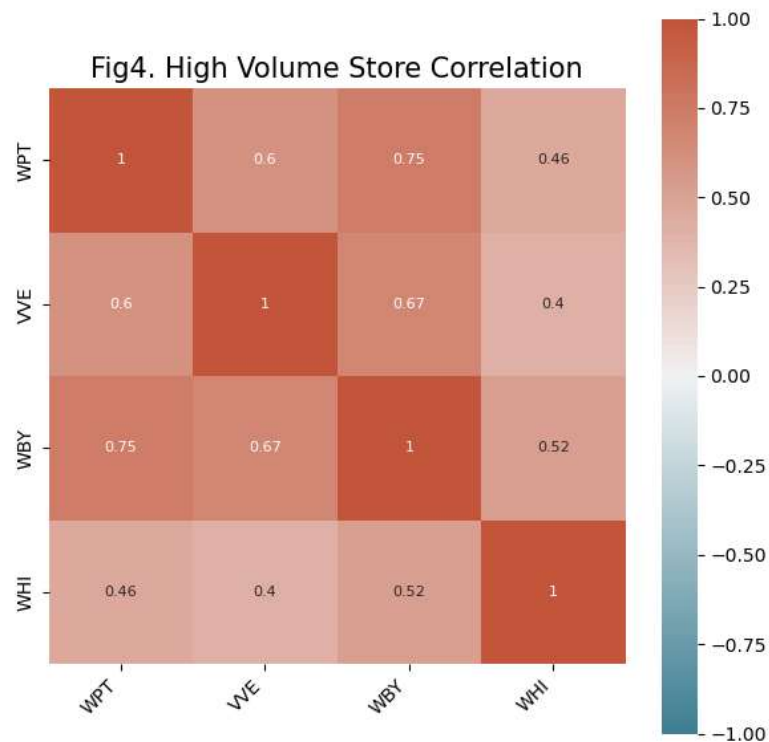


Figure 4 High Volume Store Correlation Heat Map

Figure 4 presents us with a correlation heat map between all the high volume stores. The map helps us understand the functionality and how relatable each store is between each other. The correlation coefficient ranges from -1 meaning they are not similar at all, i.e., opposite trends, 0 represents no relation and to 1, meaning its identical to each other.

We can see that the stores WPT and WBY show the highest correlation, of 0.75, with each other, it can be noted in the previous figure 2 as well that both the stores showed similar trends in peaks and troughs in the graph with respects to customer footfall.

On the other hand, while the correlation coefficient is still positive, the stores VVE and WHI, show the least similarity between each other. This could be due to seasonal conditions at the region or independent behaviour.

Overall, this figure highlights that the high volume stores operate mostly in sync with each other and if and when future strategies are developed to further improve them, it is likely that we will largely see a positive impact across all the stores.

2.5. Correlation Heat Map for Medium Volume Store.

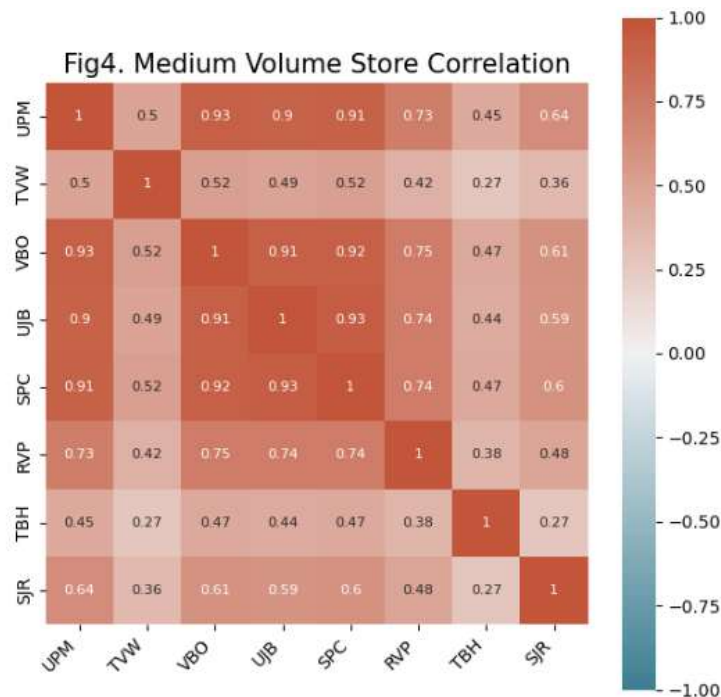


Figure 5 Medium Volume Store Correlation Heat Map

Figure 5 presents us with the Heat map between all the Medium Volume stores. Similar to the previous figure, it helps understand the correlation between each store on the customer footfall patterns throughout the year.

In contrast to the high volume stores, the Medium Volume stores display a wider range of correlations. This only indicates less of a uniformity in behaviour between each of the stores. For example, the stores UJB and VBO, UJB and SPC and VBO and SPC show a coefficient of 0.91, 0.93 and 0.92 respectively, showcasing strong similarities. However, the stores TVW and RVP, TBH and SPC and SJR and TBH show a coefficient of 0.42, 0.38 and 0.27 respectively.

This variation only suggests that majority of the Medium Volume stores operate on an independent level, this could likely be because of the locality or store specific conditions. As a result, developing a common strategy might prove difficult in this condition and would rather develop a more individual approach.

The takeaway is that some of the stores align closely with each other, while the group as a whole lacks consistency as seen with the high volume stores.

2.6. Interactive Grouped Bar Chart of Store Metrics

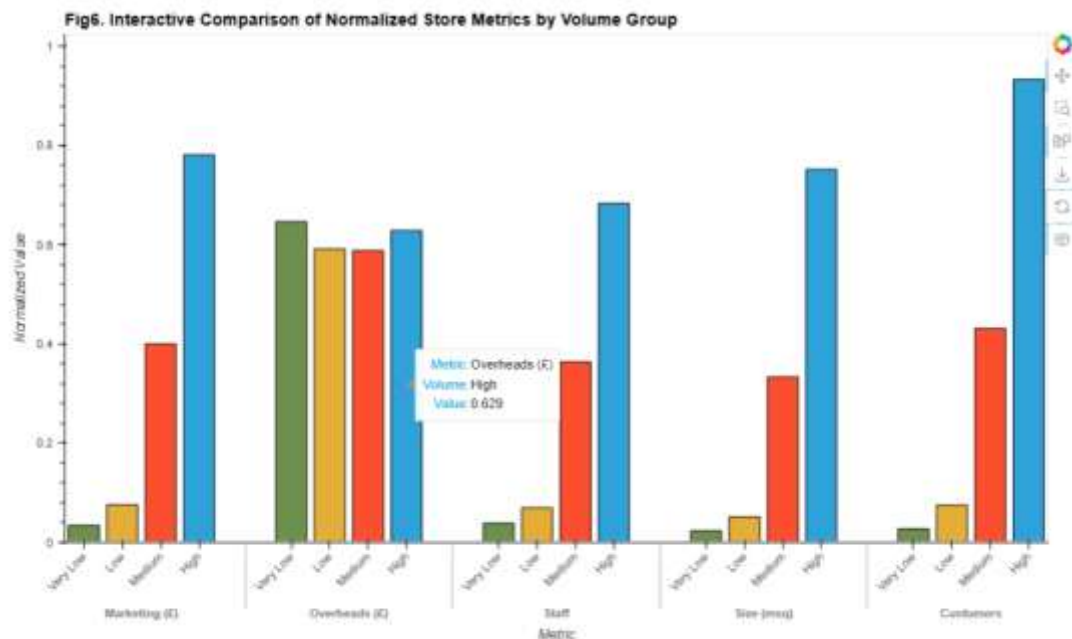


Figure 6 Interactive Normalized Store Metric Bar Plot

In Figure 6 we compare the overall store metrics from all the volume groups, for a change, the metrics was received as part of the 2nd dataframe and it holds details on the store's Marketing(£) spend, Overheads(£), the number of staff, size of the store (m²) and the total number of customers. As part of building this plot, the metric values were normalized and then grouped together into a single value in order to be able to better represent the data together in a single plot.

It is evident from the plot that high volume stores see balanced values across all metrics and see the most customers among the others, suggesting a profitable outcome. The medium volume stores however have a higher overhead expenditure, which does not make them as profitable.

The interesting part of this plot is that we can see low and very low volume stores have the lowest stats on almost all the metrics like in the size of the store, marketing, staff, even on the number of customers visiting them. But have the highest overhead spend amongst all the other groups and suggests where the resources are being allocated inefficiently.

Through this chart we are able to frame a narrative to the earlier segmentation which was formed in the first figure.

2.7. Combined Radar Plot: High and Medium Volume Stores

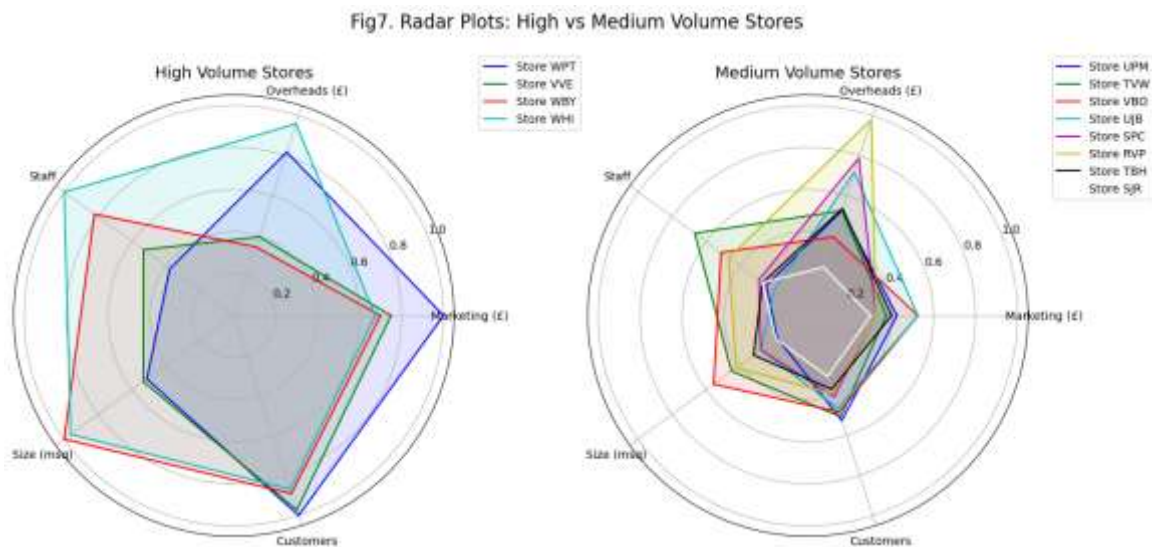


Figure 7 High vs Medium Volume Store Metrics Combined Radar Plot

We will continue to look into the high and medium volume stores and their metrics using figure 7 which shows a combined radar plot. The plot on the left shows the high volume store and the one on the right is the medium value store.

The combined radar plot allows all the stores to plot a graph of their own which is useful in making comparisons between each of them. The area the plot creates highlights where each store stands in the normalized scale.

The high volume store WHI, appears to cover the most area all the five metrics, whereas, WPT has the most money spent on marketing resulting in having the most number of customers visit the store, while having the smallest store and staff count amongst the stores. The stores in the high volume plot each have their own operational strength.

The medium volume stores on the other hand show how underperforming they are in comparison to the high volume ones, with stores like RVP, SPC and UJB having higher overspends compared to the customers that are entering their stores. This was seen in the previous figure as a whole, but here we can see which stores contribute to the overall score. Using this business can target improvements in the fields of reducing overhead expenditure and increase marketing to improve the customer footfall in the stores.

2.8. Daily Customer Trends in Very Low Volume Stores

Fig8. Very Low Volume Store Daily Customer Intake

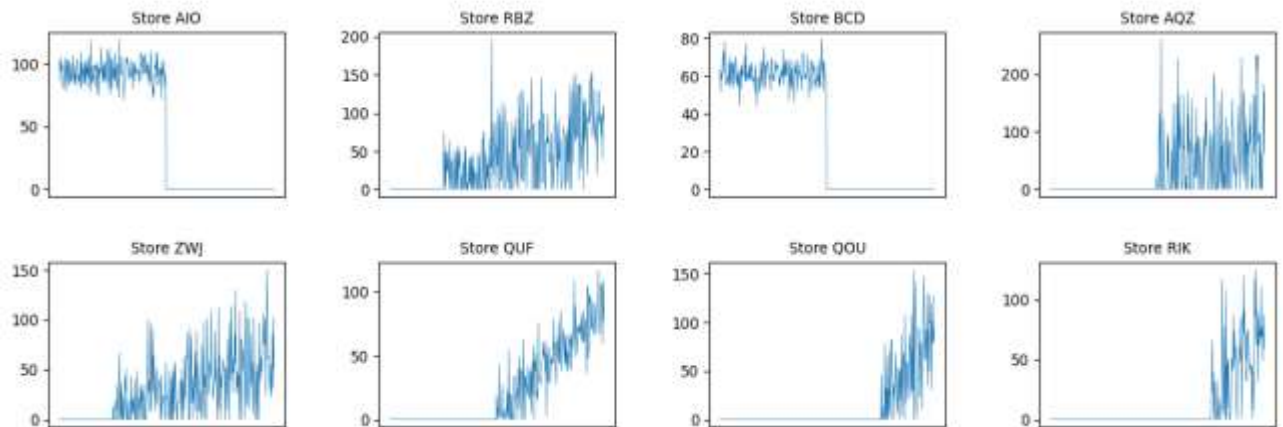


Figure 8 Very Low Volume Store Line Plot

Finally Figure 8 is dedicated to the very low volume stores using line plots. The line plot helps us see which stores were closed during the year of 2023 and the stores which were newly opened as well.

Stores AIO and BCD show immediate decline to no customers half way through the year, which likely means that it was closed. However, there are new stores opening at various points throughout the year like RBZ, AQZ, ZWJ, QUF, QDU and RIK.

This helps us validate the classification which was done earlier and gives us a perspective in the very low volume stores. It helps support the decision to focus business strategy on high and medium volume stores, while the low and very low volume stores are slowly closing or starting new.

3. Critical Review

Through this coursework, I learnt about the importance of data visualization as a data scientist and how it is really important to be able to communicate your findings in an insightful manner such that even a layman without any technical experience is able to understand and contribute to the potential business decisions taken. Each figure in this coursework was picked to be able to convey a narrative on the performance of each store by looking at trends, their heat maps on how the stores function either similarly or are unique and also to be able to highlight situations where business decisions are required.

I learnt about how without proper labelling and classification, visualizations like radar plots and box plots would be difficult to understand, whereas visualizations such as bar charts and line plots are relatively easier to understand but are really powerful, for example in the scenario of segmenting the stores. It would have been an interesting option to be able to make the interactive plots accessible in the form being able to modify in this report but it helped me understand the value of visuals are progressing and making it very flexible when compared with a static option.

Lastly, this coursework helps me in the aspect of being able to improve my narrator skills as this does require a lot of work, especially in a professional environment where things are moving really quick! So being able to plot readable and quickly identifiable visuals is the primary goal I wish to achieve through this file.

4. Conclusion

Through this coursework, we were able to identify 40 different stores and their operational abilities along with the customer inflow to each of the stores throughout a specific time period in 2023, with which we were able to classify these stores into high, medium, low and very low volume stores. The primary requirement as within the specifications was to focus on the high and medium volume stores and it was found that they were the key revenue drivers for the business. With box and rolling average plots we were able to see the seasonal impact that caused some of the stores to be volatile in nature, while the rest of the stores had seen constant influx of customers throughout the year irrespective of the season.

We utilized radar and grouped interactive bar plots to highlight the metrics of each of the stores to showcase their operational efficiency, which will provide the business with key insights as to how the stores are performing and where the money that was invested in these businesses are going, while through the line plot we could see some low volume stores which were shutting down half way through the year and also, it showed us majority of newly opened stores growing at a fast pace. The overall story supports the data driven business approach and resource planning that is required in order to be profitable regardless of the store tier.

Appendix A: Declaration of AI Use

I have used AI while undertaking my assignment in the following ways:

- To develop research questions on the topic – YES
- To create an outline of the topic – NO
- To explain concepts – YES
- To summarise the following articles/resources:
 1. The workings of HVPlots and how they can be configured
 2. The workings of BoxPlots
- In other ways, as described below:

To validate if the points or ideas I am conveying the essay are portrayed in the right tone and format.