COMP1800 DATA VISUALISATION

Tejesh Ramesh Bawa
001173097-8

Contents

1. Introduction to Data Visualisation

Data Visualisation is defined as the creation of graphical representation to show data. The main aim, as put forth by (Unwin, 2020), is to represent data and statistics visually in a way that allows interpretations to yield meaningful outcomes. (Islam and Jin, 2019) notes that organisations are leveraging data visualisations and analytical tools to make more effective inquires and make their decision-making process more efficient. Furthermore, human cognitive struggles to comprehend complex data in numbers and text compared to graphics. Usually, visual data allows decision makers to quickly identify patterns that are new and complex thereby making the decision-making process easier.

This report explores customer behaviour at each individual store. By utilizing Python libraries, a range of visualisations have been created to identify underlying trends in customer volume, seasonality and operational effectiveness that are present in Chrisco's store performance datasets. This analysis specifically categorizes store into high, medium and low customer volume groups allowing the company to allocate better resources and marketing strategies. The report further seeks to determine any anomalies among the stores like possible closure or opening of new stores through monitoring daily trends across the timeframe.

2. Visual Analysis and Insights (8 Visualisations)

2.1 Total Customers per Store (Segmented)

Figure 1 was chosen because it lays the groundwork for its ability to segment stores according on total customer volume, which is one of the primary goals in this report. By illustrating that total customer count for each store throughout the dataset, it established a crucial foundation for further analysis. Additionally, the plot helps directing the company's focus towards stores with high and medium customer volumes, while also identifying underperforming stores that may require attention. By making use of the colour code (green, orange and red) in conjunction with reference lines in segmentation enhances comprehension and allow instant identification of distribution of stores. Without this segmentation, interpreting the patterns of customer behaviour and store anomalies would lack the necessary clarity and context. This plot will help to structure for further analysis and directly support the store-by-store investigations that will follow, such as exploring the daily fluctuations, customer distributions in figure 2 and seasonality analysis in figure 3 and figure 4.

Figure 1 depicts a highly skewed distribution of customers across varies stores. Interestingly, four stores WPT, VVE, WHI and WBY stand out with an extremely large customer base and qualifying them as the high priority stores ranging from more than 2500,00 customers annually and are coloured green. This segmentation reveals a sharp drop off indicating that the customer activity is highest in these 4 stores. In the mid-sized store category, there are 8 stores which received between 100,000 and 250,000 customers annually are coloured orange. While half of the other stores are classified as low volume stores which have less than 100,000 which are coloured red.. This low overall customer numbers is a cause for concern because it showcases potential anomalies such as underutilized, disruption in stores or recently opened or closed stores.

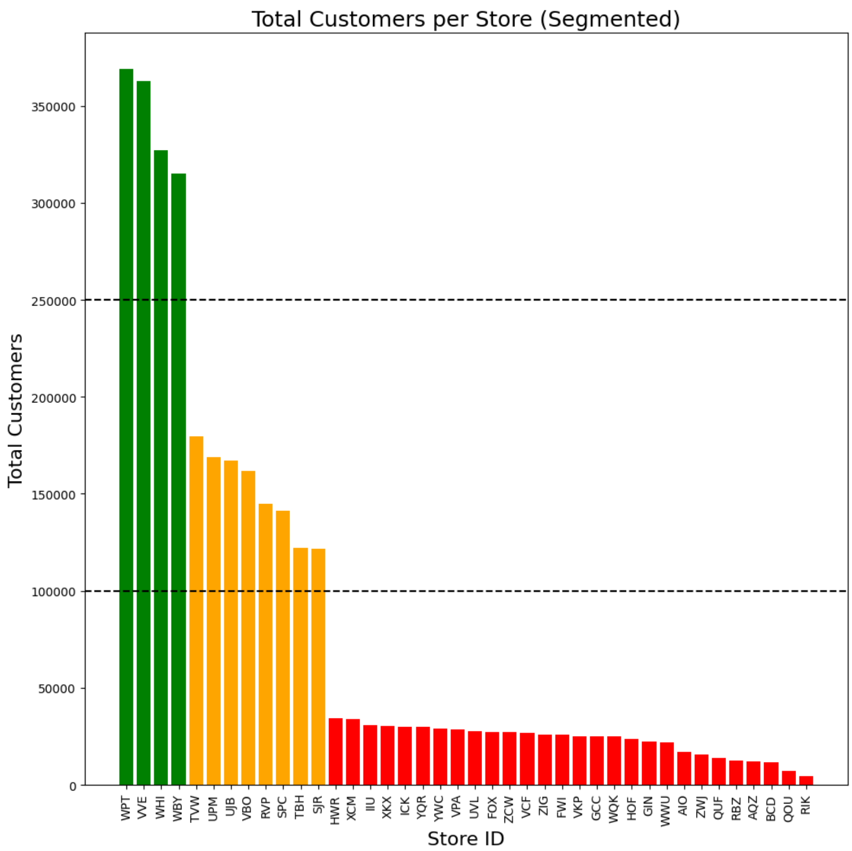

Figure 1 shows the Total Customers per Store (Segmented)

2.2 Distribution of Daily Customers per Store

Figure 2 makes it possible to examine the daily customer distribution at each store independently which offers valuable insights after segmentation. Unlike Figure 1 showing customer volumes as a whole, figure 2 visualises the fluctuations in customer volume on a daily basis for each store. It helps to distinguish between stores that have maintained a stable performance compared to those that have had displayed irregularities, skewed and minimal distributions. This is handy because it helps identify any outliers, unusual trends or detecting potential operational irregularities like sudden drops or low count depicting store closer or inaccuracies in the dataset. Furthermore, it also helps to understand and contextualise performance for example, 2 stores may have the same total customer volume but there is high possibility that it varies in operational realities and have different distribution.

Figure 2 reveals a high degree of variability in customer behaviour across different stores. High volume sores like WPT, VVE, WHI and WBY exhibit relatively symmetrical, unimodal distributions centred on high daily customer volumes. In contrast some low volume stores have skewed and compressed distribution. Furthermore, some of the lower volume stores have left skewed plots which are shown by sudden spikes in customer volume which show anomalies. This visualisation supports the justification for segmentation and provides a basis to investigate the trends in seasonality.

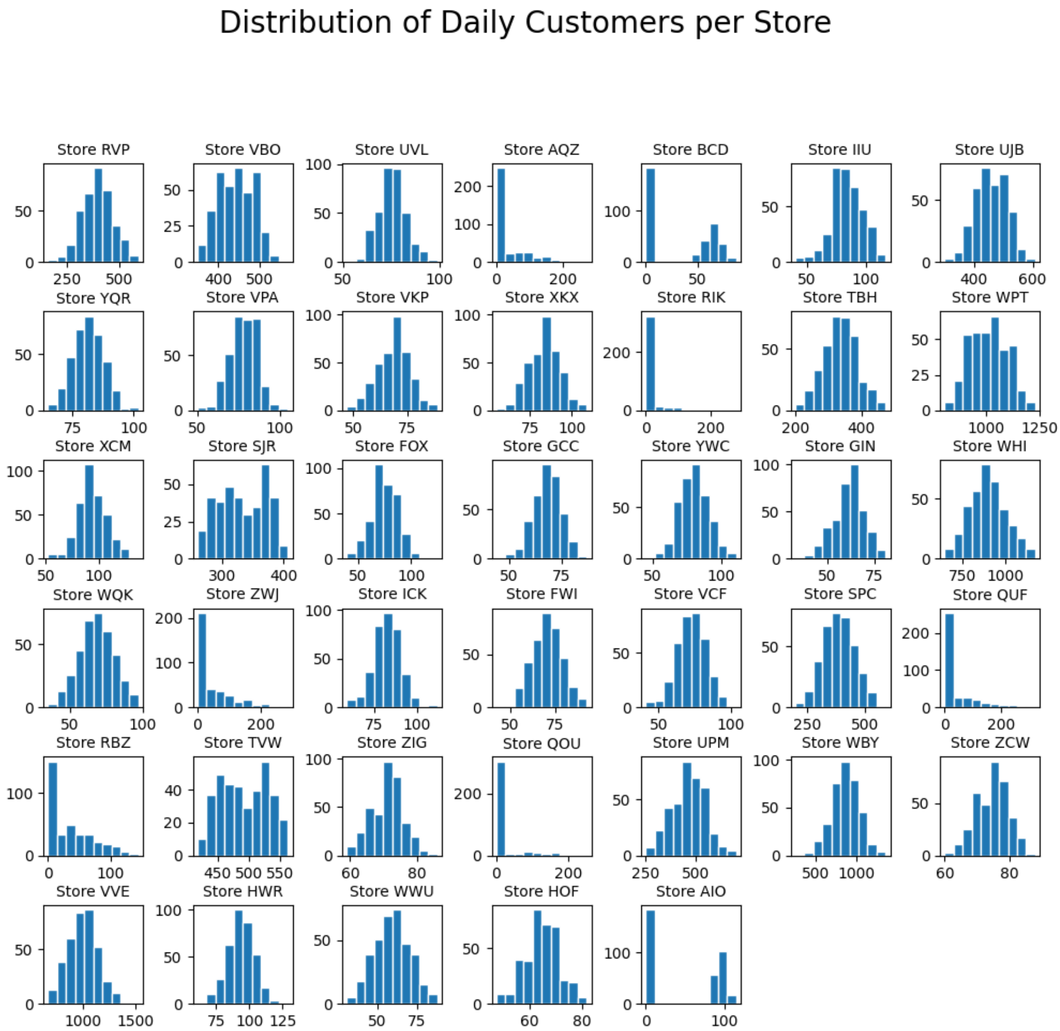

Figure 2 shows the Distribution of Daily Customers per Store

2.3 Daily Customer Trends for All Stores

Figure 3 was chosen to examine the fluctuations in customer volume throughout the day for each store, thereby helping to directly extend the daily patterns of daily customer distributions which was shown in Figure 2. While this previous plot only showed static representation of daily customer distributions, the current plot introduces a significant trend which reveals seasonality, variation, seasonal influences, and notable events like the opening or closing of a store. This step is very important as it helps identify anomalies, like steady growth or abrupt drops, and lays a groundwork for a comprehensive examination of high-volume stores in figure 4.

The chart consists of a grid made up of line plots, each depicting the trend daily customer trends for each individual stores. Several stores, for instance, WPT, VVE, WHI and WBY, show high numbers consistently across the period analysed, thereby supporting their classification in the earlier segmentation in figure 1 and figure 2. In contrast, other stores, like RIK, AIO, QOU and others, show flatlines or sudden endings, reflecting temporary or permanent closures. It is interesting to note that stores like RBZ and ZWJ have also displayed steep upward slopes, hinting at new opening activities sometime during mid-year. These findings help facilitate the identification of not just customer levels but also the dynamics that characterize the lifecycle at each store. Moving on to the figure 4, the high-volume stores will investigate week-over-week performance using smoothed averages.

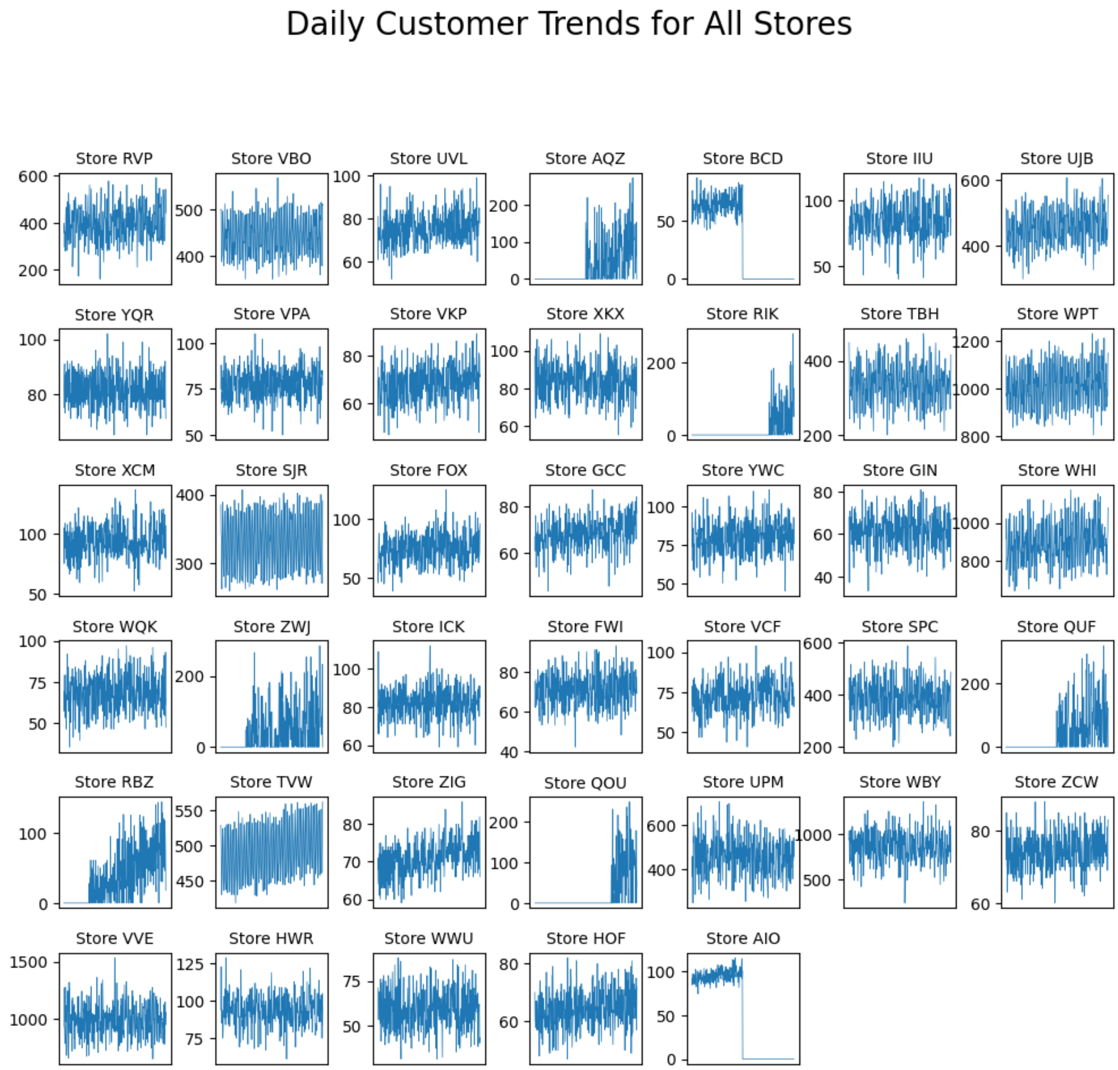

Figure 3 shows the Daily Customer Trends for All Stores

2.4 Daily Customers – High Volume Stores

Figure 4 provides a key improvement in shifting from a general, store-level examination to a more focused temporal analysis of the stores with the most significant impact. After identifying the top-performing stores (WPT, VVE, WHI, WBY), as shown in Figure 1 through segmentation, and the overall distribution and customer behavior across all stores, as shown in Figures 2 and 3, narrowing down the focus of high-volume stores for a more detailed temporal analysis. The hvPlot was chosen not only for its ability to provide an interactive and uniform comparison among several time series but also for its ability to visually aid the identification of inherent patterns, such as seasonality and anomalies. The use of color-coded and overlaid time series improves understanding of high-volume store behavior, which might otherwise be difficult to decipher when presented in the multi-line plot for all stores in Figure 3.

The figure provided shows a number of key trends in day-to-day customer visits to some leading retail stores. All four stores show considerable day-to-day variation, with distinctive peaks and troughs that are potentially associated with weekends, promotional offers, or some underlying factors in commerce. Interestingly, Store VVE appears to have the most extreme variability in customer visits, while WPT shows a more stable, but still irregular, trend. There is some initial evidence of short-term cyclical patterns—possibly weekly—although it is difficult to confirm such patterns because of the detailed level of the data.

This example clearly highlights the relevance of trend smoothing in enhancing insight into seasonal trends and patterns of a regular kind that could be lost behind volatility in raw figures. This result justifies moving to the following visualization, using a 7-day moving average for the same stores, which unearths stable trends and seasonal trends. Thus, this example works as a link phase working as a bridge between wide comparative analysis and complex pattern establishment, hence propagating the argument from segmentation to trends comprehension.

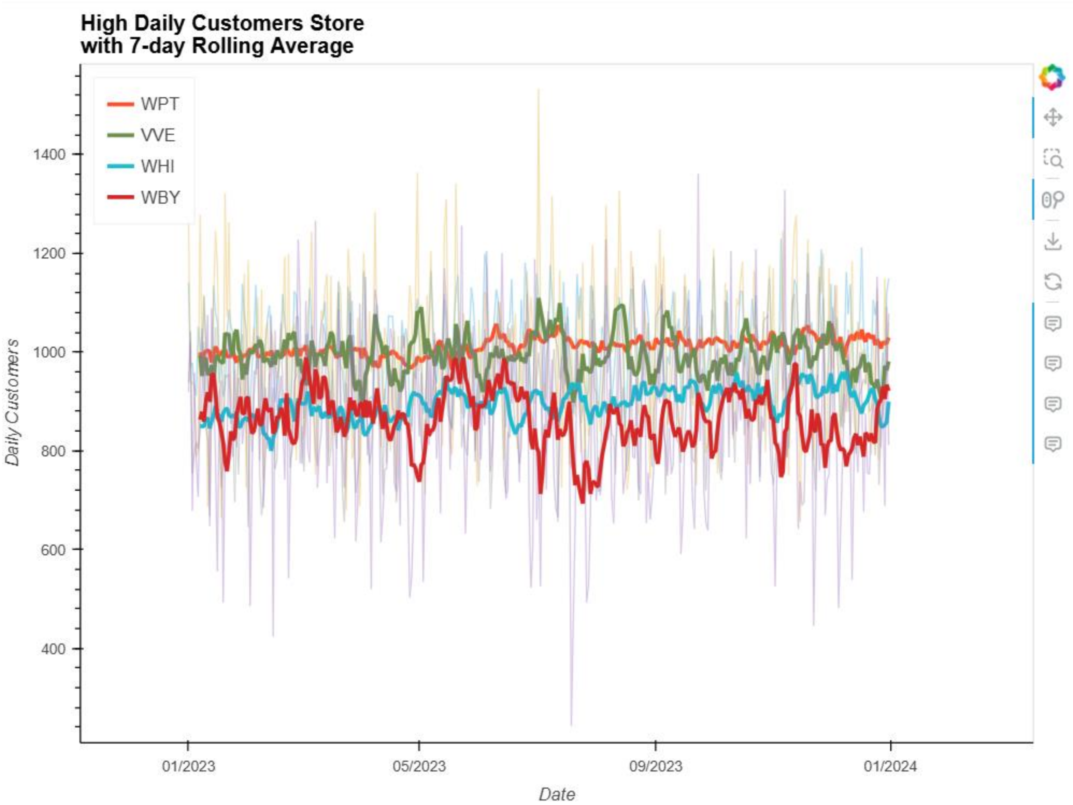

Figure 4 shows the Daily Customers – High Volume Stores

2.5 Correlation Heatmap – Store Summary Metrics

Figure 5 is used to analyse the reasons that contributes to the consistent appeal of some stores to larger crowds of customers, as seen in the figures above. This correlation matrix offers an initial overview of the interrelationships between variables like the size of a store, staff numbers, marketing expenditure, overhead costs, and overall customer counts.

The heatmap shows the correlation coefficients between essential store-level variables. Interestingly, 'Customers' demonstrated a very high positive correlation with 'Size' (0.96), 'Staff' (0.94), and 'Marketing' (0.96), suggesting that larger stores with more employees and greater marketing expenditure tends to have more customers. In contrast, 'Overheads' show a weak correlation with most variables, which implies that overhead costs have little to no effect on customer traffic. Also, Additionally, 'Customers per Staff' and 'Customers per sqm' show a negative correlation with 'Size' and 'Staff,' which indicates a potential trade-off between the operational scale and effectiveness. These insights are vital for assessing whether stores are underperforming or overperforming in terms of their available resources, which will be explored further in the following analysis.

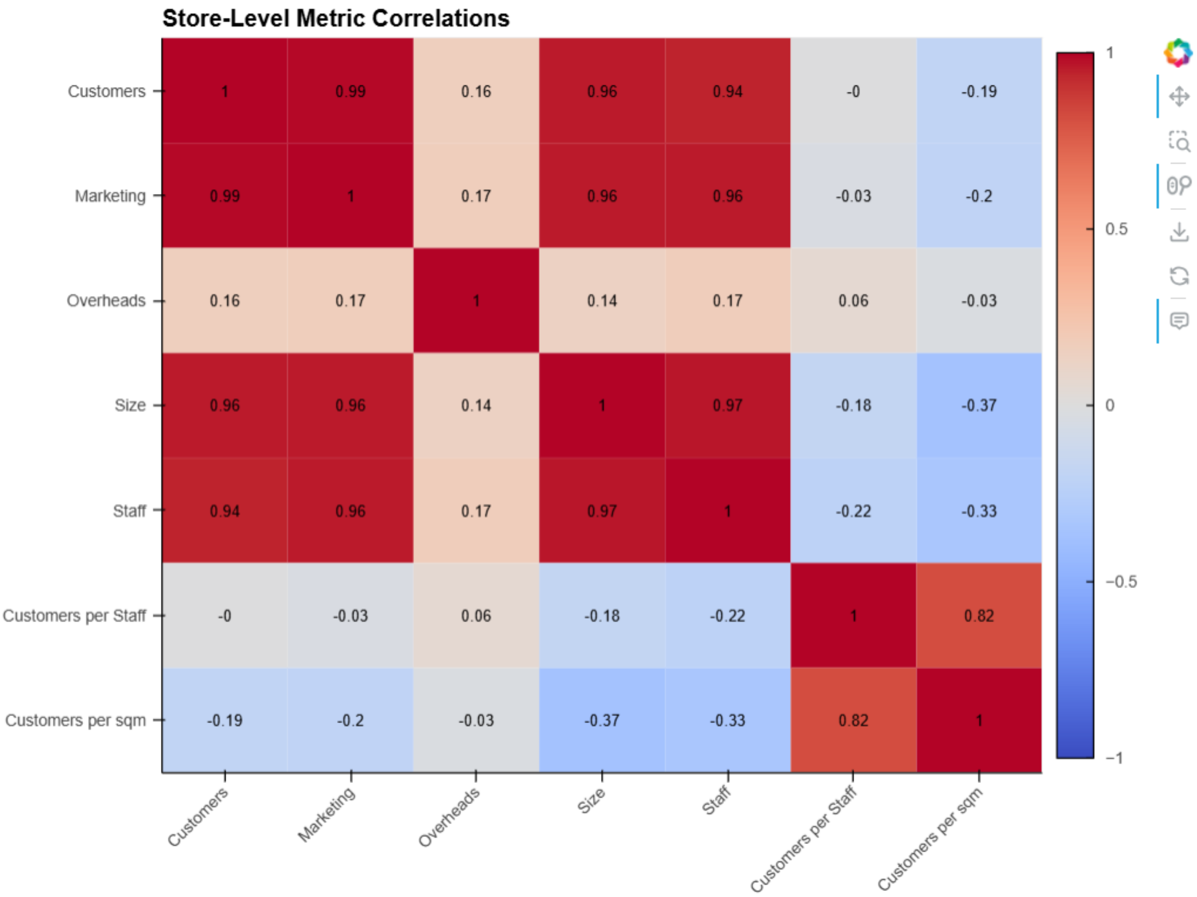

Figure 5 shows the Correlation Heatmap – Store Summary Metrics

2.6 Autocorrelation of Daily Customer Counts by Store

The autocorrelation plot serves as an important part of the analysis, that dives deeper than the volume and structure and explore the temporal patterns of customer behavior. After identifying store-level variables that affected the customer volume in figure 5, it becomes very important to examine the consistency of customer visits over time. This plot allows to assess whether stores display seasonality, periodic cycles, or random variations in customer patterns which are very information in operations planning.

The autocorrelation plots illustrate the extent of similarity between the current customer arrival at a store and its past records over a range of time intervals. It can be seen that several high-volume stores, like WPT, VVE, WHI and WBY , show progressively decreasing high autocorrelation with time, which suggests that there is seasonal patterns or regular weekly cycles. However, low-volume stores like RIK and AIO show sharp drops or sudden bursts, indicating randomness or irregularities, which can be due to the opening or closing of new stores. The trends observed point toward some stores operating within a predictable cycle, while others undergo external, short-term impacts or remain unstable.

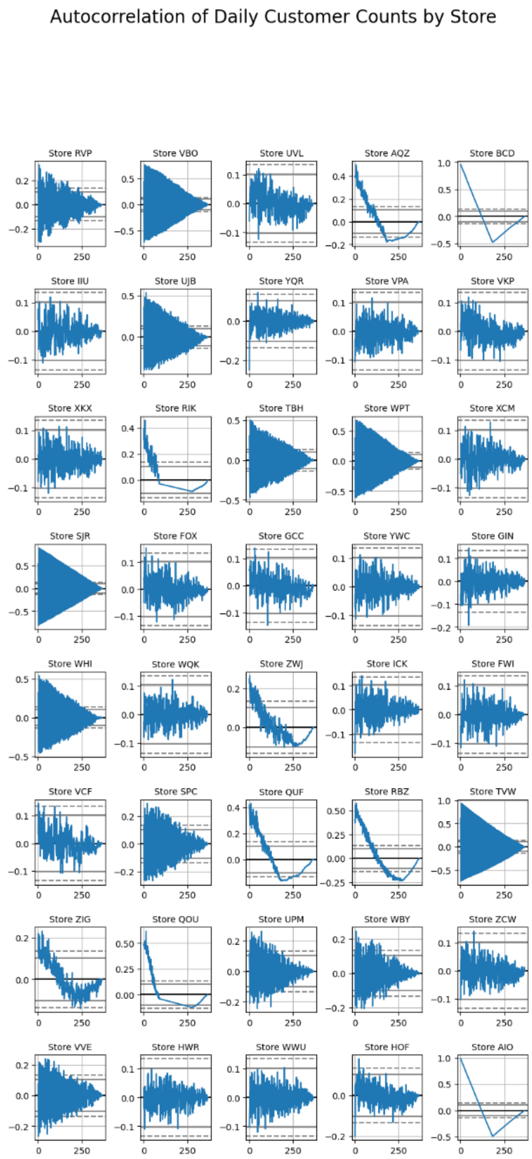

Figure 6 show the Autocorrelation of Daily Customer Counts by Store

2.7 Customers per Staff vs Customers per sqm (Segmented by Total Customers)

Following the autocorrelation analysis shown in figure 6. The plot now pays attention to store efficiency by assessing the two key metric Customers per Staff and Customers per square meter which are both found to be important in the correlation heatmap in figure 5. By categorising the stores by their total customer numbers into high, medium, and low categories, this scatter plot allows to compare between customer numbers but also assess the efficiency with which each category is associated. In addition, it will serve as a bridge to the final visual representation of figure 8, which uses radar charts to compare store profiles.

The scatter plot illustrates the distribution of stores in relation to two efficiency metrics which are customer per staff and customer per sqm. It is interesting that many low-volume and medium volume stores are shown in the upper right quadrant of the plot which indicates that there is more customer per staff and sqm, which is not optimal. High-volume stores which are shown in green tend to group around moderate to high efficiency levels, while medium-volume stores which are shown in orange have high variance and are all over the plot just like the low volume stores, indicating a lack of consistency in either staffing, space, or management approach. This visualisation is important in identifying where stores are over-resourced and under-utilized, therefore providing insights for strategic resource allocation.

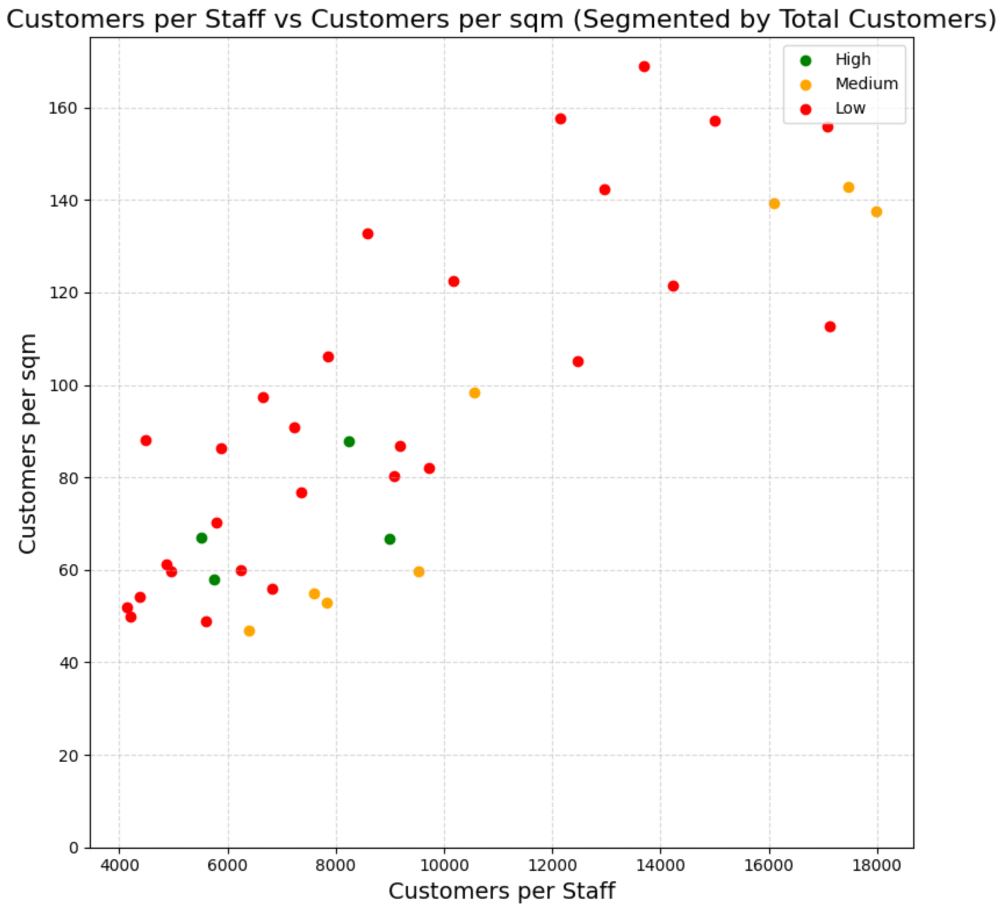

Figure 7 shows the Customers per Staff vs Customers per sqm (Segmented by Total Customers)

2.8 Radar Comparison of High-Volume Stores

The radar chart enhances the analytical process by integrating several key performance indicators into a single comparative visualization of the four high-volume stores mentioned (WPT, VVE, WHI, WBY stores). This radar chart is derived from the analysis shown in the figure 1, 5, and 7, which concentrated on the efficient metrics for all stores, therefore evaluating the unique strengths and weaknesses of each store. The radar chart is helpful in visualising the performance trade-offs across different dimensions like marketing, staff, and size, customer per staff and customer per sqm which helps to provide a management with a strategic overview.

The radar chart reveals significant operational disparities among the top four stores. WPT (blue) exhibits really high values for all main categories of investment (Marketing, Staff, Overheads, Size), but only shows low scores in customer per staff and customer per sqm which signals potential discrepancy in misalignment of allocation of potential funds. In contrast, WHI (red) also showed the same approach. VVE (green) stands out with its optimal balance between all variables despite paying relatively lower expenses towards marketing and overheads. WBY (cyan), even in terms of customer volume, significant trails in overhead costs and store size were noticed which may contribute to the varied outcomes as demonstrated. This plot encapsulates the implications of various strategies (high investment versus efficiency) in achieving similar customer volumes which can help the company to influence future discussions.

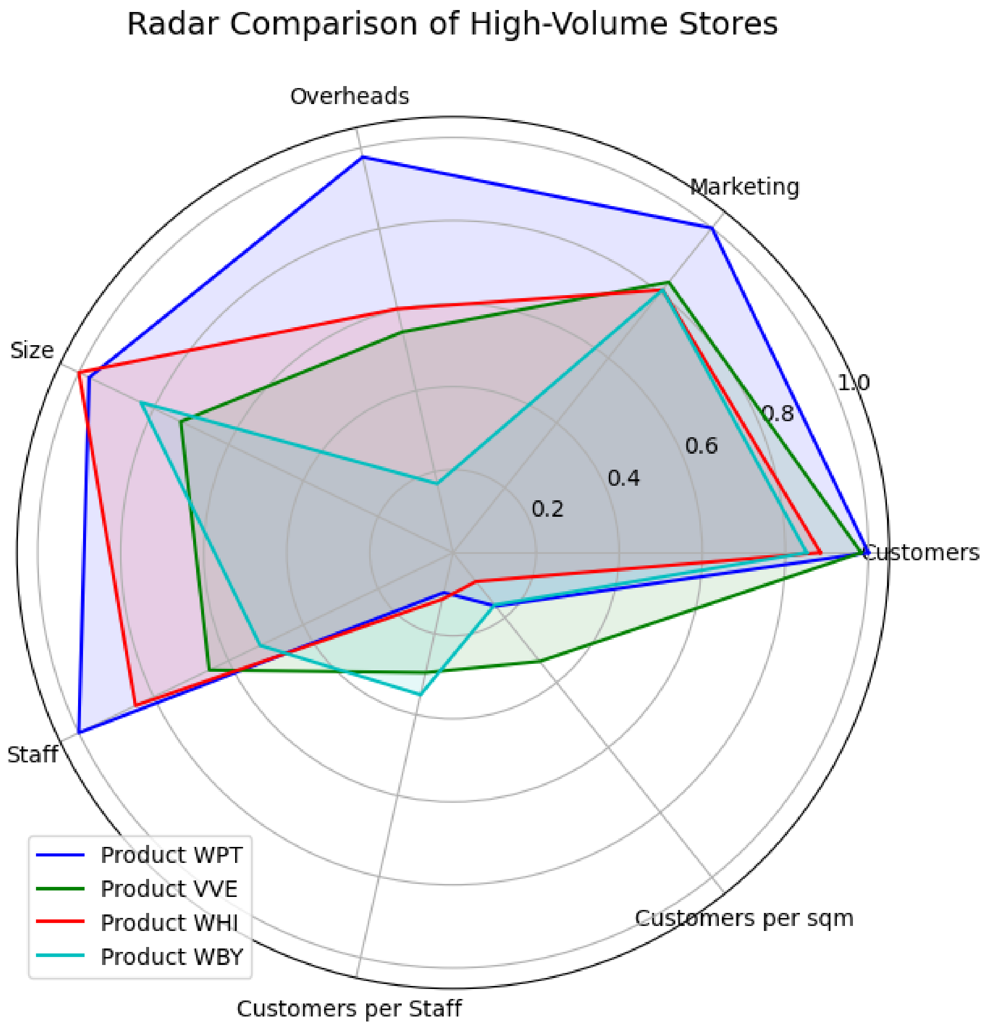

Figure 8 shows the Radar Comparison of High-Volume Stores

3. Critical Review

This coursework allowed me to apply the data visualization principles within a real retail dataset. Throughout the course, I was able to gain a deeper understanding of using visualisation techniques to investigate and communicate the insights I derived from these complex datasets. The assignments also encouraged me to pushed beyond basic plotting techniques, helping me to think and evaluate basic visualisation and understanding the different design elements and how they can contribute to the overall analysis.

One important lesson I learnt from this exercise was the importance of carefully structuring the visualisations to build a narrative that can be followed. Instead of just showing separate unrelated plots. I had made use of the best practices taught, like using colour segments, consistent labels and titles, and clear annotations. These approaches significantly enhanced my understanding of clarity and analytical impact of each figure.

Furthermore, I was able to understand how to combine statistical methods together with visualization tools. I noticed that when using the rolling average along with autocorrelation plot to detect seasonality which would otherwise be difficult to notice in a line plot. The radar and heatmap is another aspect that allowed for the multidimensional analysis which further broaden my understanding of how statistics and visualisation can work together.

By using different Python libraries including matplotlib, seaborn, and pandas which I regularly use, I also worked with hvplot for the first time. I understood that using it help the user understand the plots better as they can interact with it. Overall, this module helped enhance my capacity to produce visual analysis which are both precise and accurate which is very important in roles that are involved in using data to make decisions.

4. Conclusion

The analysis of ChrisCo's store dataset using visual techniques revealed many findings which could improve how the company understand to operate their stores through store performance and customer behavior. By using the eight carefully selected visualisation representation, the following insights were drawn:

- Customer segmentation based on volume effectively categorises the store into high, medium and low performing stores which is helpful to understand which stores need to be focused on, either to work on or to learn from.
- Store-level distributions revealed that there was daily variation in customer traffic, which includes outliers and anomalies.
- Time series trends of all stores allowed identify any potential stores that had opened or closed, as well as patterns of fluctuation that influenced the seasonal changes or local factors.
- The Rolling averages of the high-volume stores revealed consistent trend which help to understand the long-term customer engagement.
- Correlation metric helped to demonstrate all the strong positive relationships between total customers and store size, marketing expenditure, and staff, which suggested that these also influence the customer traffic.
- Autocorrelation graphs confirmed of repetitive patterns by the customers which then indicated seasonality.

- The scatter plot showing the customers per staff and sqm provided a way to evaluate productivity across stores.
- Radar chart provided an in-depth assessment of the high-volume stores to find out the strengths and weaknesses across the different metrics used in the heatmap.

These findings helped provide a insightful visual analysis of customer activity to lay a strong solid foundation for future data-driven decision-making.

5. References

Islam, M. and Jin, S. (2019) *An Overview of Data Visualization*,.

Unwin, A. (2020) Why is Data Visualization Important? What is Important in Data Visualization?, *Harvard Data Science Review*, MIT Press - Journals.