**COMP1800 – Data Visualisation**

**2023 – 2024: Term 2**

**Data Visualisation Coursework**

**Nneka Ndubishi**

**ID: 001348699**

**MSc Data Science**

Contents

[Brief Introduction to Data Visualisation 2](#_Toc161416249)

[Visualisation 1: A Bar Chart illustrating the Total Number of Customer Visits per Cinema spanning the years from 2019 to 2022. 3](#_Toc161416250)

[Justification: 3](#_Toc161416251)

[Explanation: 3](#_Toc161416252)

[Visualisation 2: Four Radar Charts illustrating the Comparison of High Volume Cinemas based on the Total Weekly Visits, Average age, Seating capacity, Marketing, and Overheads Costs. 4](#_Toc161416253)

[Justification: 4](#_Toc161416254)

[Explanation: 4](#_Toc161416255)

[Visualisation 3: A Seasonal Decomposition Analysis of Weekly Visits for Cinema 'YBS' for 2022. 6](#_Toc161416256)

[Justification: 6](#_Toc161416257)

[Explanation: 6](#_Toc161416258)

[Visualisation 4: Radar Charts illustrating the comparison between High and Medium Volume Cinemas between the categories. 8](#_Toc161416259)

[Justification: 8](#_Toc161416260)

[Explanation: 8](#_Toc161416261)

[Visualisation 5: An Interactive Line Graph illustrating the Weekly Visits over a 28-day period for Medium and Low Volume Cinemas. 10](#_Toc161416262)

[Justification: 10](#_Toc161416263)

[Explanation: 10](#_Toc161416264)

[Visualisation 6: An Interactive Heatmap Correlation displaying metric comparisons among Medium Volume Cinemas. 11](#_Toc161416265)

[Justification: 11](#_Toc161416266)

[Explanation: 12](#_Toc161416267)

[Visualisation 7: Line Plot Explorations of Weekly Visits for all Low Volume Cinemas within the first two years. 13](#_Toc161416268)

[Justification: 13](#_Toc161416269)

[Explanation: 13](#_Toc161416270)

[Visualisation 8: A Heatmap Correlation illustrating metric comparisons among Low Volume Cinemas. 15](#_Toc161416271)

[Justification: 15](#_Toc161416272)

[Explanation: 15](#_Toc161416273)

[Critical review: 17](#_Toc161416274)

[Conclusion: 17](#_Toc161416275)

# **Brief Introduction to Data Visualisation**

In the contemporary era of data analysis, the ability to extract meaningful insights from vast datasets is essential. To accomplish this, both companies and individuals rely on an effective tool known as Data Visualisation.

Data Visualisation is the practice of presenting data in graphical format, this could be in the form of visual elements such as maps, graphs, and charts. It is a highly useful practice for companies and individuals across various industries as it transforms raw data into meaningful visual representations. Such visualisations aid individuals in comprehending complex information more easily which will assist them to gain valuable insights and informed decision-making.

One way to achieve this is by selecting the right kind of visualisation, as discussed earlier, this could take the form of a map, graph, or chart, and then applying a suitable design. This ensures that stakeholders or anyone viewing the visualisation can understand it very clearly.

For ChrisCo, the cinema chain operating across the UK, data visualisation will offer a lens through which the company can understand and analyse various aspects of its business operations. By tapping into the rich pool of customer data gathered through their loyal card program, ChrisCo can unearth insights that are crucial to well-informed decision-making and strategic planning.

# **Visualisation 1: A Bar Chart illustrating the Total Number of Customer Visits per Cinema spanning the years from 2019 to 2022.**

A graph of a number of customers

Description automatically generated with medium confidence

**Figure 1: Bar Chart**

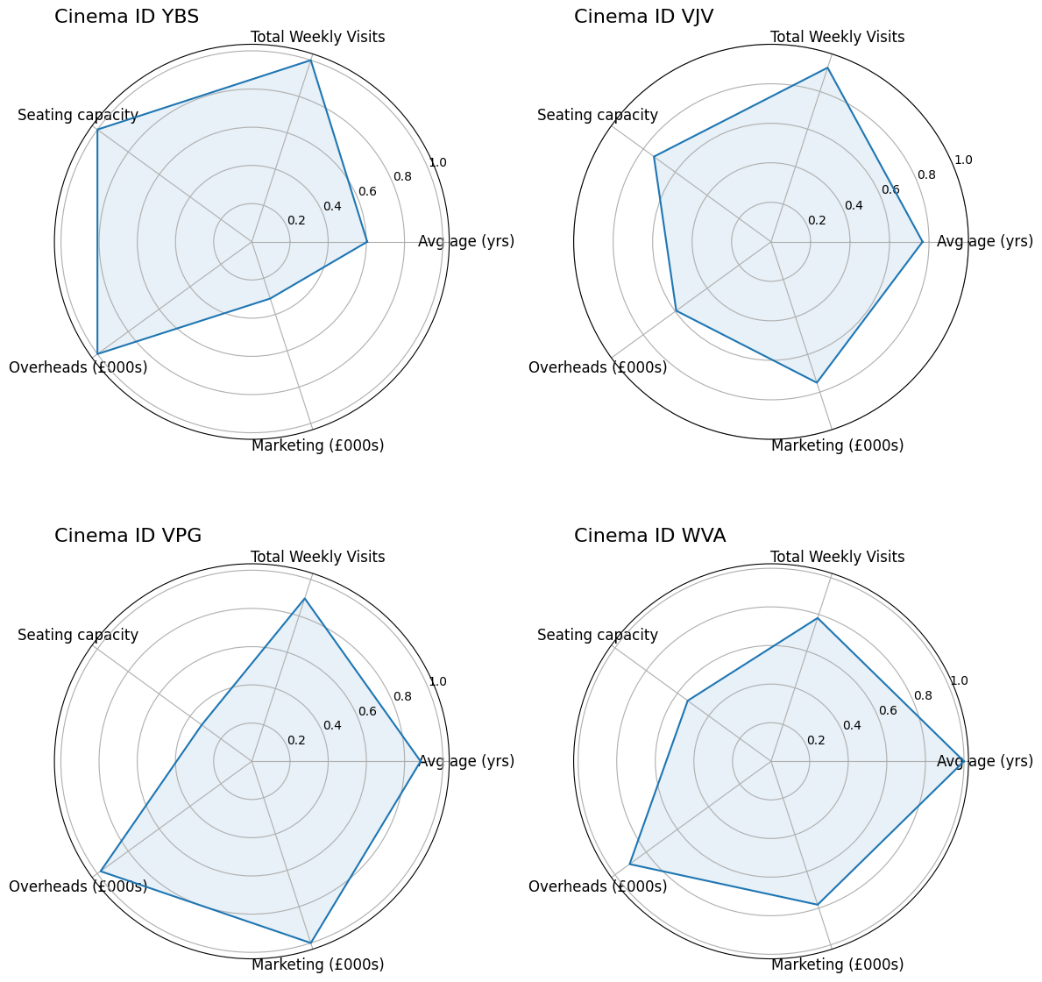
## Justification:

The bar chart in Figure 1 is chosen as the visualisation method to demonstrate the total customer visits per cinema spanning the years 2019 to 2022. This choice is based on its ability to effectively present the total visitor count for each cinema and to differentiate between high, medium, and low visit volumes. It provides a straightforward way to identify cinemas with high and low visitor numbers, which is crucial for ChrisCo’s reporting purposes. This presentation format enhances clarity and makes it easier for stakeholders to interpret the data effectively.

## Explanation:

Figure 1 categorises the cinemas into three distinct groups based on visitor volumes: high, medium, and low which are each distinguished by a unique colour scheme spanning the years 2019 to 2022. Cinemas attracting high volumes, displayed in green have accommodated over 150,000 visitors, while those in the medium volume range depicted in yellow have accumulated between 50,000 to 150,000 visitors. Low volume cinemas are in red and have accommodated fewer than 50,000 visitors. Cinemas such as YBS, VJV, VPG, and WBK emerge as the top performers in terms of visitor numbers. With these insights, ChrisCo can delve into factors influencing visitor volumes, such as overhead expenditure and seating capacity, to identify trends and plan accordingly.

# **Visualisation 2: Four Radar Charts illustrating the Comparison of High Volume Cinemas based on the Total Weekly Visits, Average age, Seating capacity, Marketing, and Overheads Costs.**



**Figure 2: Radar Charts**

## Justification:

The radar chart was selected as the most appropriate visualisation to illustrate the comparison of the four high volume visitor cinemas based on key attributes such as ‘Total Weekly Visits’, ‘Average age’, ‘Seating capacity’, ‘Marketing’, and ‘Overheads’ costs. Figure 1 highlighted these cinemas as the highest volume due to their cinema’s visitation activity. Figure 2’s radar chart has normalised values which involve scaling the values of each attribute commonly ranging between 0 to 1. Furthermore, the radar chart visually captures the strengths and weaknesses of each cinema as the shape that is formed by the lines offers insights into comparative advantages or areas for improvement. Each axis in the radar chart represents an attribute, in Figure 2’s case, the attributes are ‘Total Weekly Visits’, ‘Average age’, ‘Seating capacity’, ‘Marketing’, and ‘Overheads’ costs, this will facilitate an easy comparison in identifying the outliers or performers within this dataset.

## Explanation:

As for Cinema YBS, the normalised values begin with ‘Total Weekly Visits’ where it maintains a high value before gradually decreasing to ‘Average age’ and ‘Marketing’ which rises back to ‘Overheads’, forming a loop. This trend suggests a strong emphasis on visitor engagement with fluctuations indicating possible shifts in marketing strategies or customer demographics.

Cinema VJV, the second highest volume visitor cinema in Figure 1, exhibits consistent normalised values across all the attributes of the radar chart which suggests a balanced performance.

Furthermore, Cinema VPG in Figure 3 showcases a spike in ‘Marketing’ and ‘Overheads’ which is followed by a dip in ‘Seating Capacity’ indicating a potential area for improvement.  
Lastly, Cinema WVA displays a strong emphasis on ‘Average age’ and ‘Overheads’ with a sharp decrease in ‘Seating Capacity’ which could signify a focus on customer experience. Overall, these trends in the four radar charts provide the cinemas valuable insights into their performance area and priorities.

# A graph of a graph**Visualisation 3: A Seasonal Decomposition Analysis of Weekly Visits for Cinema 'YBS' for 2022.**

**Figure 3: Seasonal Decomposition Analysis**

## Justification:

A Seasonal Decomposition Analysis was chosen as the appropriate analysis to illustrate the Weekly Visits for the Cinema ‘YBS’ in 2022 in Figure 3. According to Figure 1, Cinema ‘YBS’ has the highest volume visitor count over the three years so this indicates its significance for this analysis. Focusing on the latest year's data will provide the most up-to-date insights which ensures an accurate analysis of the dataset. There are four components of the seasonal decomposition analysis as seen in Figure 3, they all serve a distinct purpose such as the ‘observed’ plot is the actual weekly visitation patterns; the trend reveals long-term movement; the seasonal component shows recurring patterns and the residual identifies random variability.

This visualisation in Figure 3 offers a unique insight into Cinema ‘YBS’s performance that other visualisations cannot provide. This is due to if the data was to become aggregated then this would defeat the purpose of analysing the monthly data and focusing on its trends.

## Explanation:

In Figure 3, the observed graph illustrates the visitation figures throughout the year, ranging from 1000 to 1200 visits. In this graph, the peaks are evident in January, while September experiences the lowest visits. The trend graph follows a similar trend to the observed graph, with August showing the highest values after a steep decline in July**.** In contrast, the seasonal graph maintains a consistent pattern throughout the year fluctuating between 0.98 and 1.02. Peaks do occur in the seasonal graph, around the beginning of each month whilst lows fall slightly below 0.98**.** Lastly, the residual shows random variations with August displaying the highest values, mid-September being the lowest and a similar dip in mid January suggesting external factors may have influenced. These insights help make better decision-making and can allow ChrisCo to see which months and seasons influence the cinema visitation activity.

# A diagram of a diagram of a diamond**Visualisation 4: Radar Charts illustrating the comparison between High and Medium Volume Cinemas between the categories.**

**Figure 4: Radar Charts**

## Justification:

Similarly to Figure 2, choosing the radar chart as a visualisation to compare two high volume cinemas and two medium volume cinemas across different categories seemed most appropriate. The radar charts will help visualise each cinemas performance in areas such as ‘Average age’, ‘Seating capacity’, ‘Marketing’ and ‘Overheads’. By showcasing both high and medium volume cinemas on the same chart, the stakeholders can see easily how they compare against each other in each category and makes it easier to identify strengths and weaknesses for each cinema and their category.

Another visualisation that could have served this purpose is a comparative bar chart, however, the radar chart offered distinct advantages against the comparative bar chart. The radar charts circular layout facilitates an interpretation of multi-dimensional data which creates better decision-making. Furthermore, the comparative bar chart does not necessarily provide a holistic view of performance whereas, the radar does as it allows for simultaneous comparisons across multiple attributes. Therefore, the most appropriate visualisation for the comparative analysis would be the radar charts, despite the bar chart’s potential utility.

## Explanation:

According to Figure 1, it highlights that Cinema ID ‘YBS’ and ‘VJV’ are the two highest high volume visitor cinemas and Cinema ID ‘WBK’ and ‘TPY’ are the two highestmedium volume cinemas. Figure 4’s radar chart can offer as a comparison between these cinemas across various key performance indicators. Furthermore, by comparing the medium volume cinemas with the high volume cinemas, they can identify areas for improvement and adjustments to enhance performance and potentially close the gap.

In Figure 4, the ‘Avg age (yrs)’ radar chart indicates a trend that cinemas with more customers tend to cater to the older audience, whilst those with a lower visitor volume attracts a younger demographic. This suggests a possible correlation between the average age of cinemagoers and the cinema’s visitor volume activity.

Furthermore, the ‘Seating capacity’ radar chart suggests lower values for the two High Volume Cinemas, VJV and YBS, implying smaller venues compared to the Medium Volume Cinemas, TPY and WBK. Whilst this might limit the High Volume Cinemas seating capacity, it doesn’t necessarily explain their high visitor figures as there could be other factors influencing their popularity.

The Medium Volume Cinema, TPY stands out in the ‘Marketing’ radar chart suggesting more efforts in their marketing compared to the other cinemas. Furthermore, analysing the ‘Overheads’ radar charts, the cinemas VJV and WBK score higher, indicating potentially higher overhead expenses compared to cinemas TYP and YBS.

Overall, through the use of radar charts, the stakeholders can gain valuable insights into cinema performance dynamics which will aid in strategic decision-making.

# **A screen shot of a graph Description automatically generatedVisualisation 5: An Interactive Line Graph illustrating the Weekly Visits over a 28-day period for Medium and Low Volume Cinemas.**

**Figure 5: Interactive Line Graph**

## Justification:

An interactive line graph is the most appropriate visualisation representation for viewing the weekly visits over a 28 day period for High and Medium Volume Cinemas. By selecting three Medium Volume Cinemas and three Low Volume Cinemas for the comparison, the line graph can avoid cluttering whilst providing the viewer with meaningful insights. Choosing to include three Medium and Low Volume Cinemas provides a balanced representation across different levels of attendance and this approach ensures clear insights into attendance patterns across the cinemas.

The interactive elements of the line graph has useful tools such as hover, zoom, pan, autoscale and reset axis which will enhance the viewer’s experience by allowing them to explore specific data points effortlessly. Figure 5’s screenshot of the interactive line graph shows that by hovering over the Low Volume Cinema “XWO” on June 22, 2021, one can identify easily that they had 236 visitors that week. Furthermore, the line graph’s ability to display seasonality within a one month period will enable the viewer to observe trends over time which includes peak and off-peak days.

## Explanation:

The interactive line graph in Figure 5 illustrations weekly visits over a 28 day period for 3 Medium Volume Cinemas and 3 Low Volume Cinemas, all 6 cinemas use different colours to help distinguish each cinema. Cinemas “WBK”, “TPY” and “UVQ” are Medium Volume cinemas which are classified in Figure 1. Whereas, Cinemas “WQW”, “ACQ” and “XWO” are low visitor volume cinemas.

All cinemas except for Cinema XVO experience a dip on Tuesday 15 June, 2021, this observation suggests that Tuesdays may not be favourable for cinema visits which highlights a potential trend in their attendance patterns.

Cinema XVO, a low volume cinema, experiences a peak in the first three weeks of June which further promotes investigation into the factors contributing to this surge in attendance. Medium Volume Cinema “UVQ” also experiences an increase in their customers with nearly 200 more within one week. However, the line graph doesn’t provide the exact reasons for these fluctuations but it will offer an insight into further analysis.

# **A screen shot of a screen Description automatically generatedVisualisation 6: An Interactive Heatmap Correlation displaying metric comparisons among Medium Volume Cinemas.**

**Figure 6: Interactive Heatmap Correlation**

## Justification:

The heatmap uses a numerical scale which ranges from -1.0 to 1.0, this range depicts the strength of the correlations among the various metrics. A correlation coefficient that is close to 1.0 indicates a strong positive correlation, this suggests to the viewers that as one metric increases, the other tends to increase also. Whereas, a coefficient correlation that is closer to -1.0 signifies a strong negative correlation which implies an inverse relationship between the metrics. The interactive features of the Heatmap will allow the viewer to easily analyse and explore the dataset. This functionality is facilitated by options such as “Zoom”, “Pan”, “Auto scale”, and “Reset axes” which will enable the viewers to utilise the interactivity of this visualisation. For instance, in Figure 4’s screenshot, I hovered over the comparison between Overheads and Marketing and it revealed a strong positive correlation of 0.9504.

Another type of comparison graph that could have been considered for this particular visualisation is a scatter plot graph. In this case, a scatter plot might have appeared relatively empty rather than cluttered due to the limited number of paired comparisons. Furthermore, the scatter plot graph could have still provided insights into individual relationships but the heatmap offers a more concise representation of the correlations simultaneously. The heatmap’s ability to show all relationships at once makes it a better choice for analysing the dataset compared to the scatter plot graph. Therefore, the Interactive Heatmap remains the best option for this analysis and visualisation.

## Explanation:

For this Interactive Heatmap in Figure 6, blue indicates a positive correlation, white indicates no correlation and orange/red indicates a negative correlation. The heatmap reveals for Medium Volume Cinemas that there is a strong positive correlation between “Overheads” and “Marketing” which indicates as marketing expenditures increase, overhead costs also tend to rise. On the other hand, a negative correlation also appears between “Average Spend” and “Average Age” which suggests as the average age of cinemagoers increases, their average spending decreases. In addition to this, the heatmap also indicates no significant correlation between “Marketing” and “Average Age”.

The colour-coded cells in the heatmap provides the viewers with visual clarity by using darker and bolder colours representing the stronger correlations and lighter, pale colours indicating weaker correlations. Overall, the heatmap has facilitated a deeper understanding of the dynamics within the Medium Volume Cinemas which allows viewers to discern patterns and trends that can impact decision making processes.

# **Visualisation 7: Line Plot Explorations of Weekly Visits for all Low Volume Cinemas within the first two years.**

A chart of blue lines

Description automatically generated with medium confidence

**Figure 7: Line Graph**

## Justification:

The line graph was chosen as the visualisation for the exploration of weekly visits for all Low Volume Cinemas within the first two years because of its ability to depict seasonal trends clearly. This visualisation allows for each comparison between the cinemas and reveals seasonality and anomalies over the two year period which is essential because it helps stakeholders identify potential areas for improvement.

Another type of visualisation that could have been used is the bar chart, this is due to the bar chart having the ability to present the total number of visits over the two year period. However, the issue with using the bar chart visualisation is that it cannot capture the seasonality that a line graph can offer so using the bar chart would be more limited as it may overlook important trends in the data.

Therefore, the line graph seems to be the most appropriate visualisation in Figure 7 as it explores the weekly visits for Low Volume Cinemas that can capture seasonal trends and will reveal anomalies.

## Explanation:

In Figure 7’s line graph exploring weekly visits for all Low Volume Cinemas within the first two years, Cinemas SDT, CWN, CCX, BKK and AKA exhibit static lines at 0 which suggests they may have been temporarily closed or not operating during this period. It is possible that the presence of numerous Low Volume Cinemas was influenced by the new openings or temporary closures of the cinemas within the first or second year of operation. Additionally, Cinemas ZQL and BWF show declines to 0 after the second year which indicates the possibility of temporary or permanent closures post the initial two year period. These insights show the importance of operational activity and longevity when interpreting visitation data as the fluctuations impact volume metrics.

# A diagram of a diagram Description automatically generated with medium confidence**Visualisation 8: A Heatmap Correlation illustrating metric comparisons among Low Volume Cinemas.**

**Figure 8: Heatmap Correlation**

## Justification:

The heatmap in Figure 8 illustrates how various metrics such as “Average Age”, “Seating Capacity”, “Marketing”, “Overheads” and “Average Spend” are correlated within the Low Volume Cinemas. Using a heatmap to analyse these metrics will provide the Low Volume Cinemas to uncover the correlations and patterns which may not be obvious with other visualisation tools. Another comparative visualisation that could have been used is a comparative bar chart, comparative bar charts can be useful for comparing individual metrics, however, they cannot show the correlations between them. However, heatmaps offer a clearer view of these relationships by using colours and a numerical scale which can highlight the strength of correlations between metrics. This visualisation approach is the most appropriate in this scenario as stakeholders can now easily identify which metrics move together enabling better decision-making.

## Explanation:

The colour scheme of Figure 8’s heatmap displays red representing a correlation of 1.0 which highlights a strong positive correlation, whereas, blue indicates a correlation of -1.0 which is a strong negative correlation. White reflects a correlation of 0 which translates to two metrics having no correlation between them.

Figure 8’s heatmap visualisation reveals several key insights and patterns about the data for the Low Volume Cinemas. Firstly, the strongest positive correlation is between the “Seating Capacity” and “Average Spend”, which is evident by the dark red tone in this heatmap graph. This positive correlation suggests that an increase in the cinema’s seating capacity corresponds to higher average spending per customer which hints at potential revenue optimisation through strategic seating arrangements. Furthermore, the category with the most negative correlations between the other metrics is “Overheads”, this is evident because there are 3 blue colours for the categories. This implies that when other factors change, such as Marketing or Seating Capacity, the overhead costs stay consistent. Another observation made from the heatmap is that there isn’t a correlation between Marketing and Seating Capacity, this suggests that the number of seats available does not influence any marketing strategies.

# **Critical review:**

My Data Visualisation coursework consisted of 8 visualisations that I have learned throughout the Data Visualisation module. The 8 visualisations consisted of a bar chart, two radar plots, a seasonal decomposition analysis, two heatmap charts and two line plot explorations. One of the heatmap charts and line plot explorations were created into interactive visualisations using Plotly, an interactive graphing library for Python, Hvplot and Bokeh are alternative interaction visualisations that I could have used for this report.

Moreover, producing all 8 visualisations for my report assisted with deepening my understanding of the importance of colour coding and labelling for visualisations. Not only do colour coding and labelling aid me in explaining trends and seasonality but it also assists viewers or stakeholders in discerning peaks and areas of poorer performance.

Furthermore, this Data Visualisation report has exhibited my grasp of the module by showcasing my application of theory to visualisations. For example, in Figure 5, I created a line graph exploration spanning 28 days, I chose this timeframe to highlight to the viewer what are the most popular days for cinema visits. Line graphs are effective for illustrating time series data so it enabled me to articulate the trends and seasonality. Additionally, each cinema in Figure 5 was colour coded which helps to simplify the interpretation of performance metrics.

# **Conclusion:**

* Many Low Volume Cinemas are recorded in Figure 1, the reason for this is explained in Figure 7.
* The openings or temporary closures of Cinemas SDT, CWN, CCX, BKK and AKA directly impacted the amount of Low Volume Cinemas.
* Figure 3 indicates dips in the observed graph in September, this could be due to factors like school holidays. The peaks in the seasonal graph can coincide with the beginning of new movie releases, attracting higher audience members.
* Despite being Medium Volume visitor cinemas, figure 4 shows that the Medium Volume Cinemas have a higher seating capacity than the High Volume Cinemas.
* The findings in Figure 4 indicate that the size of the cinemas does not correlate with the cinema activity.
* Majority of the Low and Medium Volume Cinemas in Figure 5 faced a dip in the visitation activity on the third Tuesday of the month, this could indicate that Tuesday’s are not a favourable day for visitors.
* Figure 6 shows that for Medium Volume Cinemas, there is no correlation between Average Spending and the average age of the cinemagoers.
* Figure 6 also displays that all other categories significantly affect seating capacity, as it shows the highest number of strong positive correlations among all other categories.
* In all Low Volume Cinemas, seating capacity and marketing are the only categories to exhibit no correlation (as seen in Figure 8).