Data Visualisation

On

Cinema Company Sales

Name: Maitreyi Manish Kulkarni

MSc Big Data and Business Intelligence

Table of Contents

Table of Figures

Introduction

Data visualization is the graphical representation of information and data, utilizing visual elements such as charts, graphs, and maps to make data more readable. The main goal of data visualization is to communicate information visually by telling a compelling story with the help of graphs, charts, and maps. It's an essential phase in the data analysis process and a method or kind of art that can help all business sectors. A significant distinction exists between possessing vast quantities of data and truly knowing how to leverage it to obtain data-driven decisions and actions; data visualization serves to bridge this divide (Stevens, 2023). Data visualization offers a rapid and efficient means of communicating information. Additionally, by using this technique, organizations can estimate sales volumes, identify areas that require more attention or improvement, enhance stakeholder retention through data visualization, determine the best times and locations for particular product placement, and uncover the elements that influence consumer behaviour (Brush, 2023). There are various types of plots and charts that fall within the visualisation of data. The most common are bar graphs where the height of each bar represents the values of it. This way totals can be calculated and compared. When the relationship between two variables needs to be found out, a scatter plot can come in handy. Some other types include pie charts, bubble plot, clustering and so on. Each type of chart has its own usage and advantages.

The data provided is regarding a chain of cinemas across the UK. Information about the customers visiting the cinemas has been averaged over four years and combined to provide details on the company's cinemas, each of which has been assigned a distinct three-letter code.

Visualisation 1-

There can be various factors related to the analysis of cinemas, but it is imperative to first find out the high, medium, and low-volume cinemas. This is required for basic operational as well as promotional purposes. Segmenting the data into different sets will help in knowing where to allocate resources efficiently, which cinema needs staff adjustments, and what maximum revenue for medium and low volume ones. Cinema chains can compare their performance to that of their competitors by knowing the number of visitors to each of their locations. With the use of this data, plans to increase the chain's overall competitiveness can be developed. The cinemas that have a low volume of visitors or if they require improvements that can be found out as well. The cinemas that are performing well can provide insight into what is making the customers happy, and this could also help in future approaches.

Figure 1. describes a bar chart of all total weekly visits at every cinema in the given period of time. The height of each bar corresponds to the total number of customer visits to a particular cinema. Cinemas with taller bars indicate higher total visit counts, while shorter bars represent lower visit counts. For each cinema, the total visits are calculated by summing up all the values in the respective column. On the horizontal X-axis, the names of the cinemas are displayed, while the vertical Y-axis represents the total number of visits. By looking at this colour-coded bar plot, the distribution of total visits across different cinemas is evident. Bright pink and purple colours are assigned to the high and medium volume cinemas respectively and yellow colour is denoting the low-volume ones. This colour scheme helps visually distinguish between cinemas with different levels of visits volume.
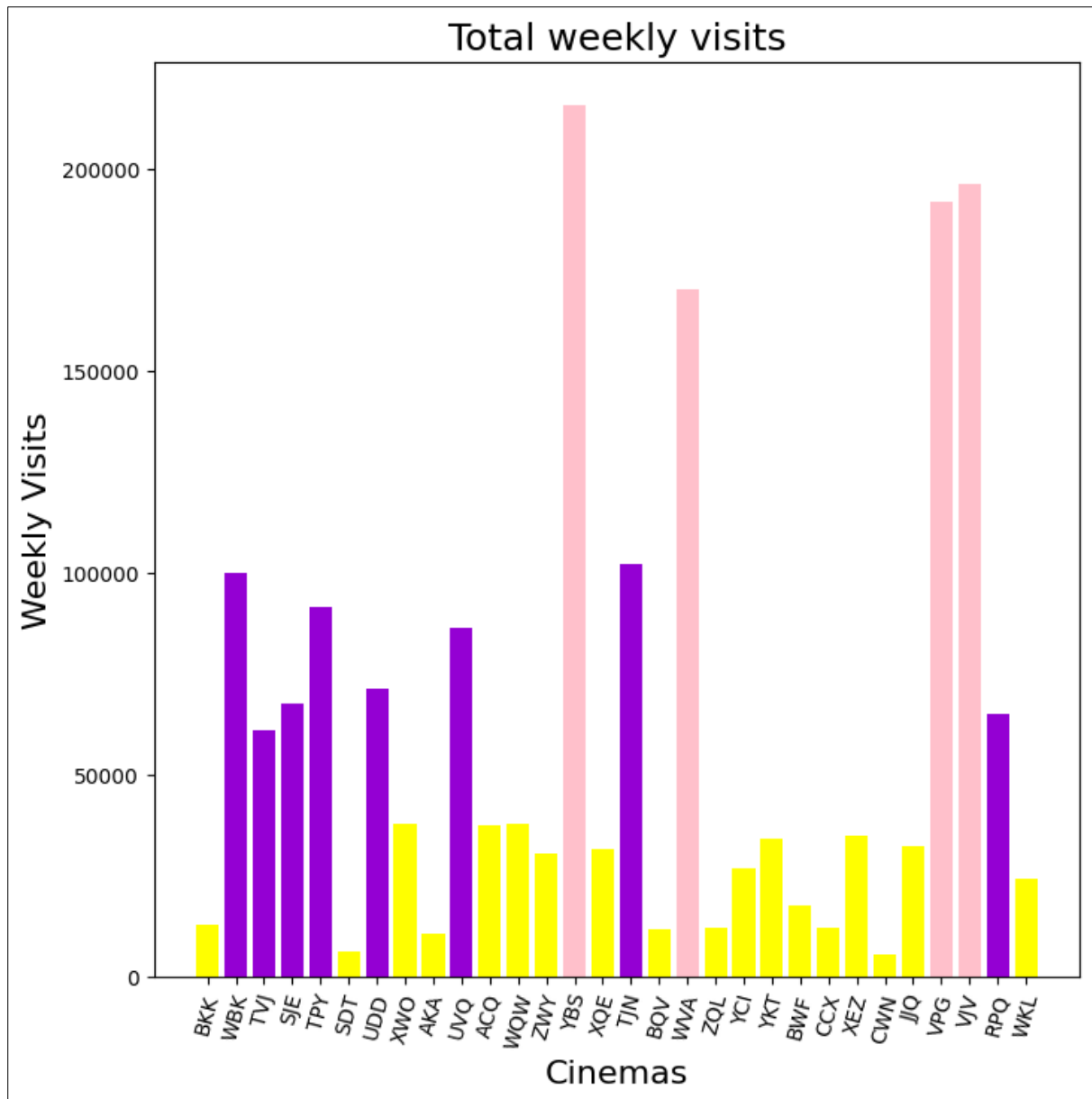
Figure 1. Bar Plot of Total Cinema Visits

Visualisation 2-

Once the data is segmented into three sets namely high, medium and low-volume cinemas further analysis can be performed on each of these sets. Some of these show interesting trends and to pinpoint it the lines must be plotted on a fixed number of points within a given period. Random ups and downs of day-to-day visits can obscure the data. Rolling average smoothens the data, which lessens the effect of noise or sudden variations. By reducing the impact of random variation, it aids in highlighting repetitive patterns in the data. In trend analysis, rolling averages are frequently utilized to find underlying trends or patterns in time series data. Additionally, the focus is on long-term averages hence the decisions are not limited to only the near future.

The following line plot, figure 2, shows all cinemas with medium volume, and it is used to calculate a rolling average with a window size of 14 days. This rolling average computes the mean of the previous 14 days' data for each day, effectively smoothing out short-term fluctuations and highlighting longer-term trends. The rolling average is denoted by a line of thicker width for clarity. Cinema 'TPY' shows a uniform weekly fluctuation. The visits rise and fall alternatively per week. WBK and TJN show drastic rises and drops. They rose and were at their peak in the time frame of 2021-01 and 2022-01 respectively. RPQ also fluctuates weekly, but it is not as uniform as TPY. The remaining cinemas UVQ, UDD, SJE, and TVJ are fairly constant with few highs and lows, but they are not that noticeable. The volume of their visits has not changed to a great extent.
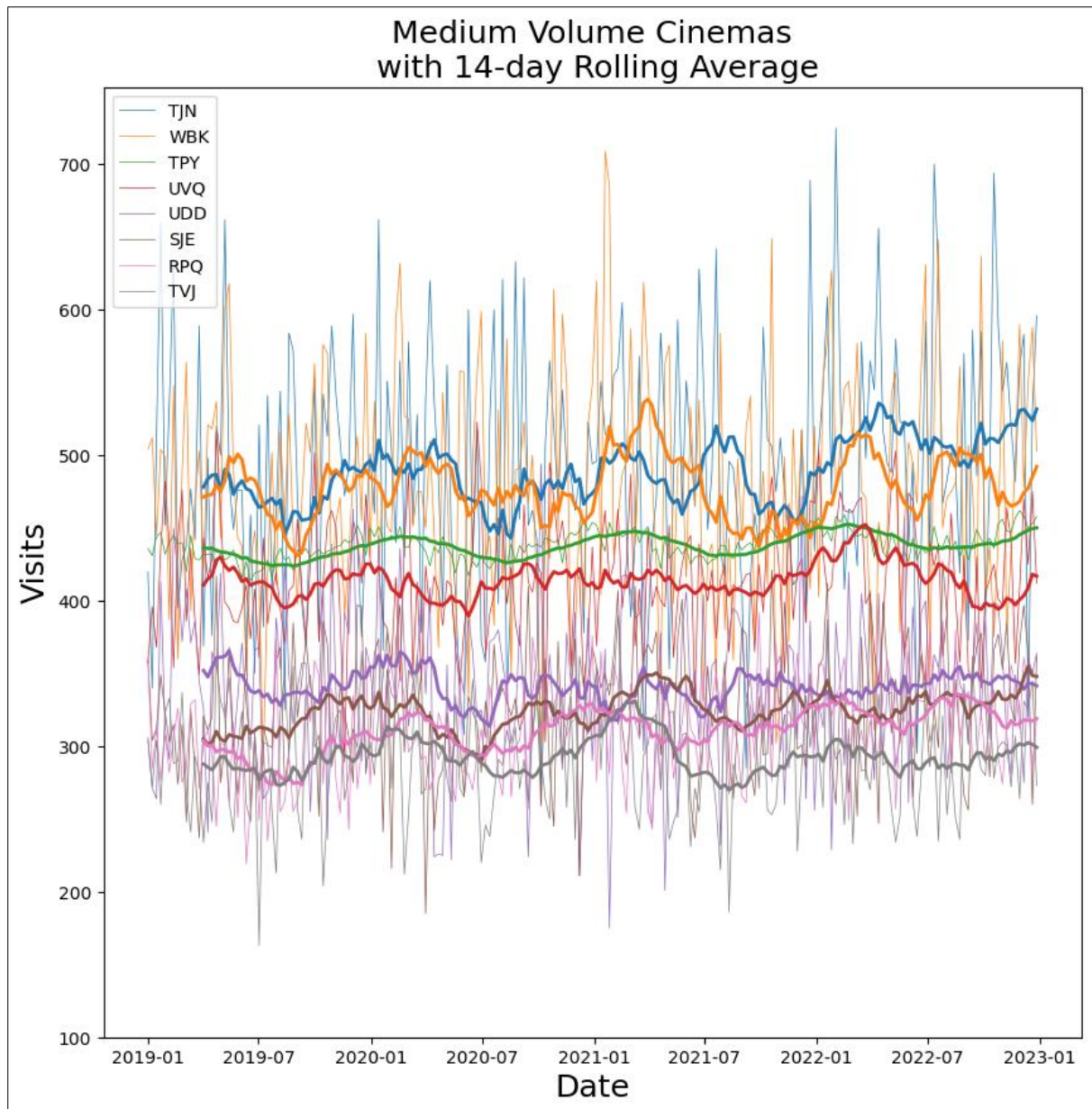
Figure 2. Line plot of medium volume cinemas with rolling averages

Visualisation 3-

A Line plot allows to visualize the trends in weekly visits to high volume cinemas over time. Most revenue is generated by the cinemas with high volume of visits. By observing the overall trend, we can identify patterns such as seasonality, upward or downward trends, and any significant fluctuations. Identifying anomalies, and effectively communicating customer visits data to stakeholders becomes easier with line plots, especially in high volume cinemas. It is a valuable tool for data analysis and decision-making in the cinema industry.

In the given pictographs, the number of visits at the high volume cinemas is plotted. Considering figure 1, YBS shows a steady increase in visits. There are a few dips in between but the overall number does not drop below a certain amount of around 600 even once. The highest number of weekly visits is at a value of 1475 in the year 2020 which is shown in figure 3.2 using an interaction tool 'hover'. In the same figure, a box zoom is displayed which helps to zoom specific blocks or boxes. Upon zooming it is observed that all 4 high volume cinemas have sharp peaks and lows. The footfall at these cinemas changes drastically and data is not found to be smooth. Cinema 'VPG' takes a drop in the middle of 2019, but gains popularity eventually and even surpasses VJV in 2021 and 2022. Overall, there are fluctuations present.
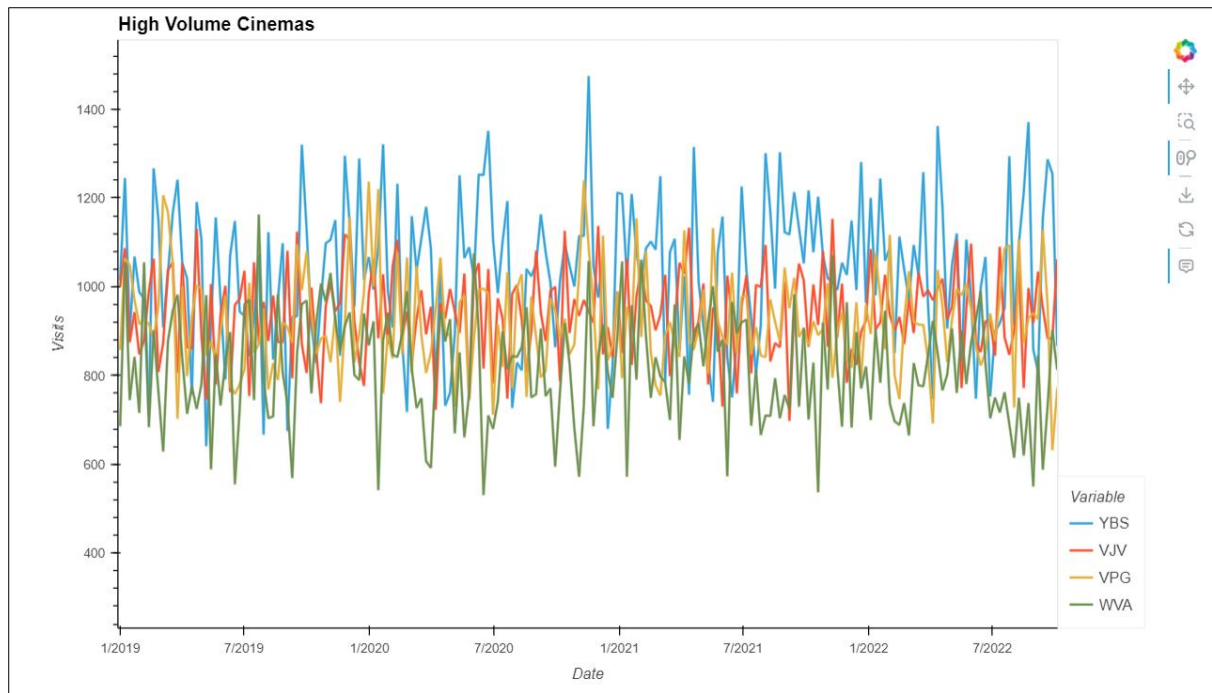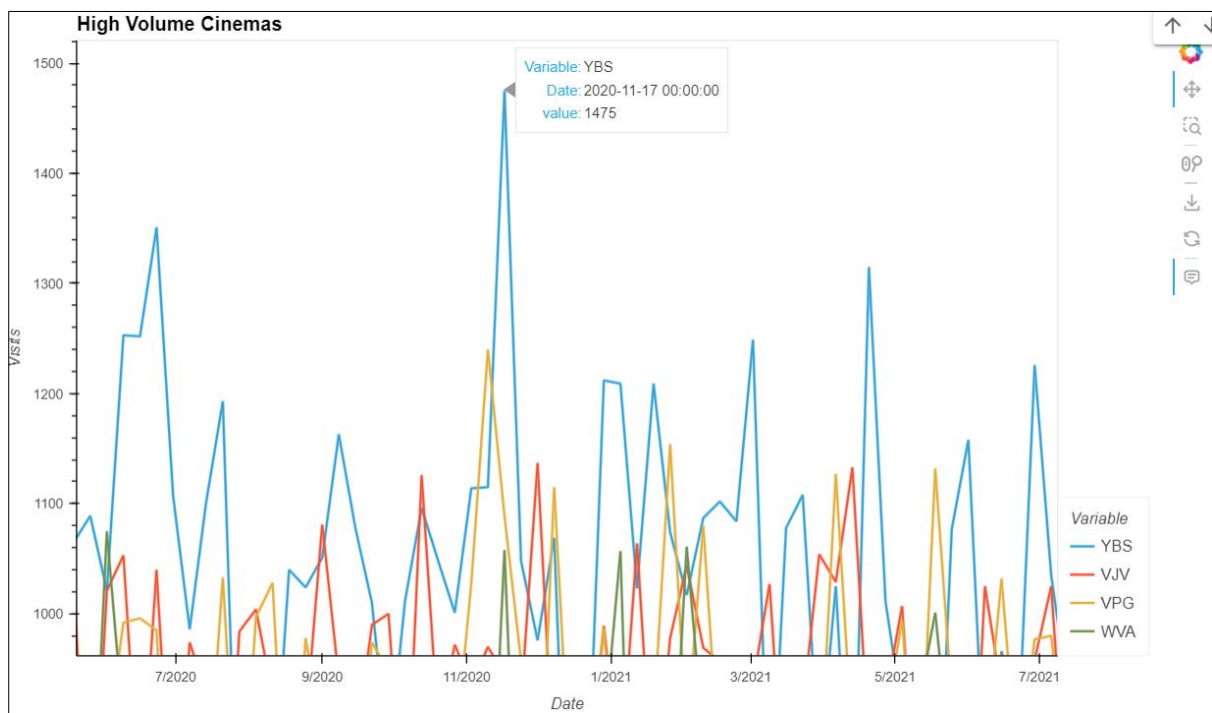
Figure 3.1. Line plot of high volume cinemas

Figure 3.2. Line plot showing zoom and hover

Visualisation 4-

As compared to the high and medium volume plots, the set of low volume visits data shows slower trends as it could take a lot of years to obtain quick and noticeable results. Nonetheless, over time, line plots enable us to identify any gradual increases or decreases in visitation, providing valuable information about the overall performance trail of these cinemas. Furthermore, abrupt increases or decreases in visits provide information about outside variables that affect going to the theatres, such as unique events or shifts in the local population. With this information at hand, management is better equipped to allocate resources, ensuring optimum staffing levels, inventory control, and focused marketing campaigns at low-volume cinemas.

Figure 4 illustrates the low volume cinemas with their trend lines plotted across the number of visits. Almost all of the low performing cinemas have a straight line indicating minor or no change in the years. The visits are capped and there seems to be no increase. Some of these show a rise quite later on. Before 2020, their visits are seen to be below 0 which means that they were newly opened. Post opening the visits have increased progressively. On the other hand, some cinemas have got shut down since their values are converging to 0 and falling even further. An example of that is ZQL and BWF. There are pointed peaks and lows which suggest sharp increase or decrease. This may be due to specific timelines or popular movies.
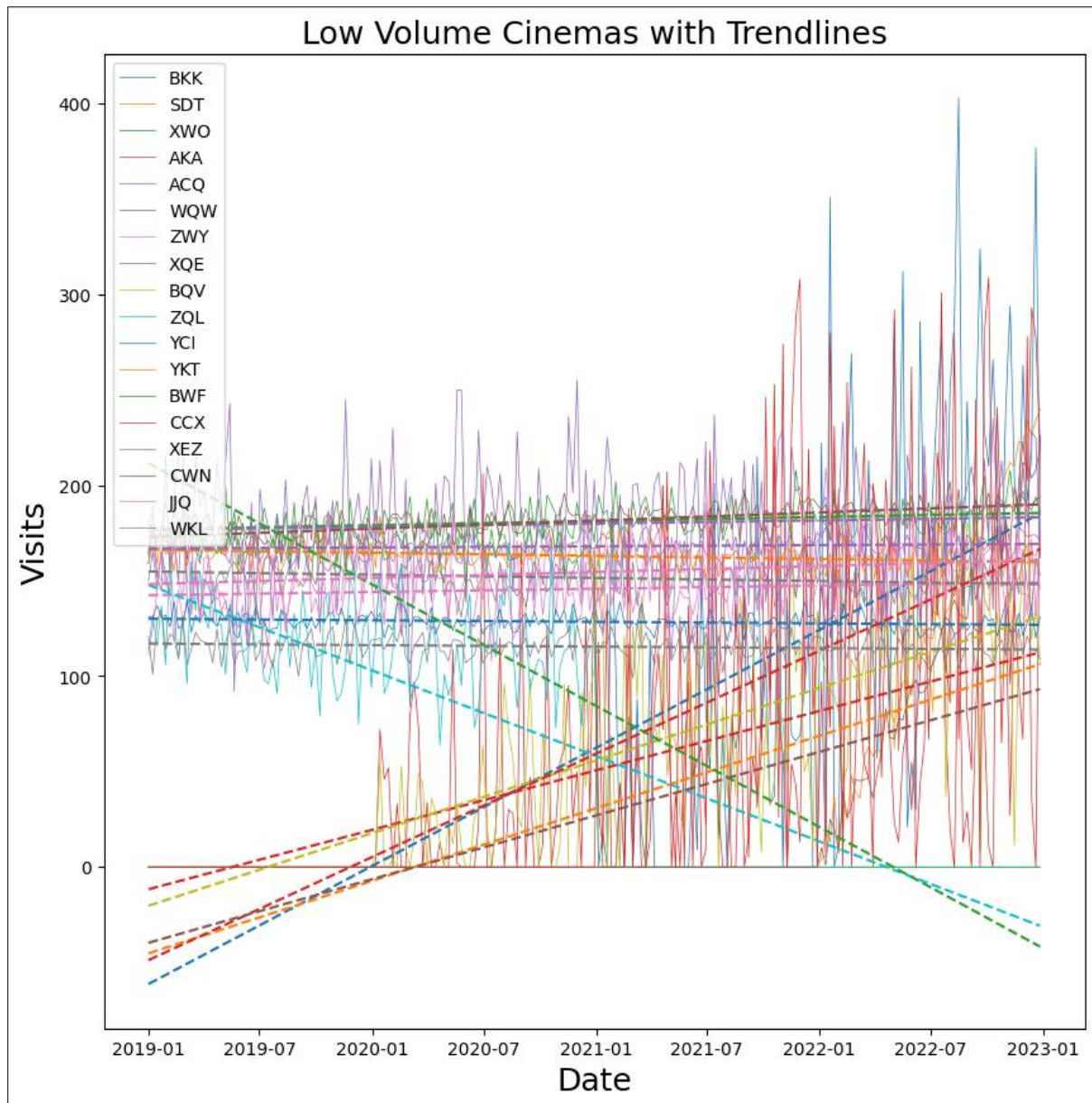
Figure 4. Line plot of low volume cinemas with trendline

Visualisation 5-

The Summary data is a compilation of all datasets concerning the marketing expenses, customers' spending, and the capacity of the cinema. This also includes total weekly visits to each cinema. These different aspects of the chain of the cinemas could most definitely be related to one another. A heatmap makes it simple to find patterns and correlations between variables by presenting the data in a grid format where each cell's colour intensity represents the magnitude of a certain variable. They can help identify trends and patterns instantly and any discrepancies can be spotted.

The values at the diagonal of the heatmap are always 1 because each parameter has perfect positive correlation with itself. According to the colour scheme it is easy to judge that three of the parameters namely visits, capacity and marketing have a strong positive correlation. In simple terms, it means that each of their values are dependent upon each other but they are not exactly the cause of one another. Their values are:

- ➢ Visits vs Capacity = 0.96
- ➢ Visits vs Marketing = 0.92
- ➢ Capacity vs Marketing = 0.96

It makes sense to assume that cinemas with additional seats will draw more visitors. It suggests that while other factors might also affect the number of visitors, expanding a cinema's seating capacity may result in a boost in attendance. More seats available mean more chances to sell tickets, which increases the number of visitors. Further, increased marketing expenditures typically result in a greater volume of visits. Successful marketing strategies can create interest increase awareness, and increase number of customers. The more the cinema spends on its local marketing the better is the chance for that cinema to be popular. Cinemas with larger seating capacities tend to allocate higher budgets to marketing efforts. This could be because larger cinemas have more seats to fill and therefore require more extensive marketing campaigns to attract visitors.

Another correlation is seen between age and the average spend although it is not as significant. The age of customers and the amount they spend at cinemas is correlated at 0.66 which is a moderately positive correlation. The greater the age of the customers, the more they are willing to spend.
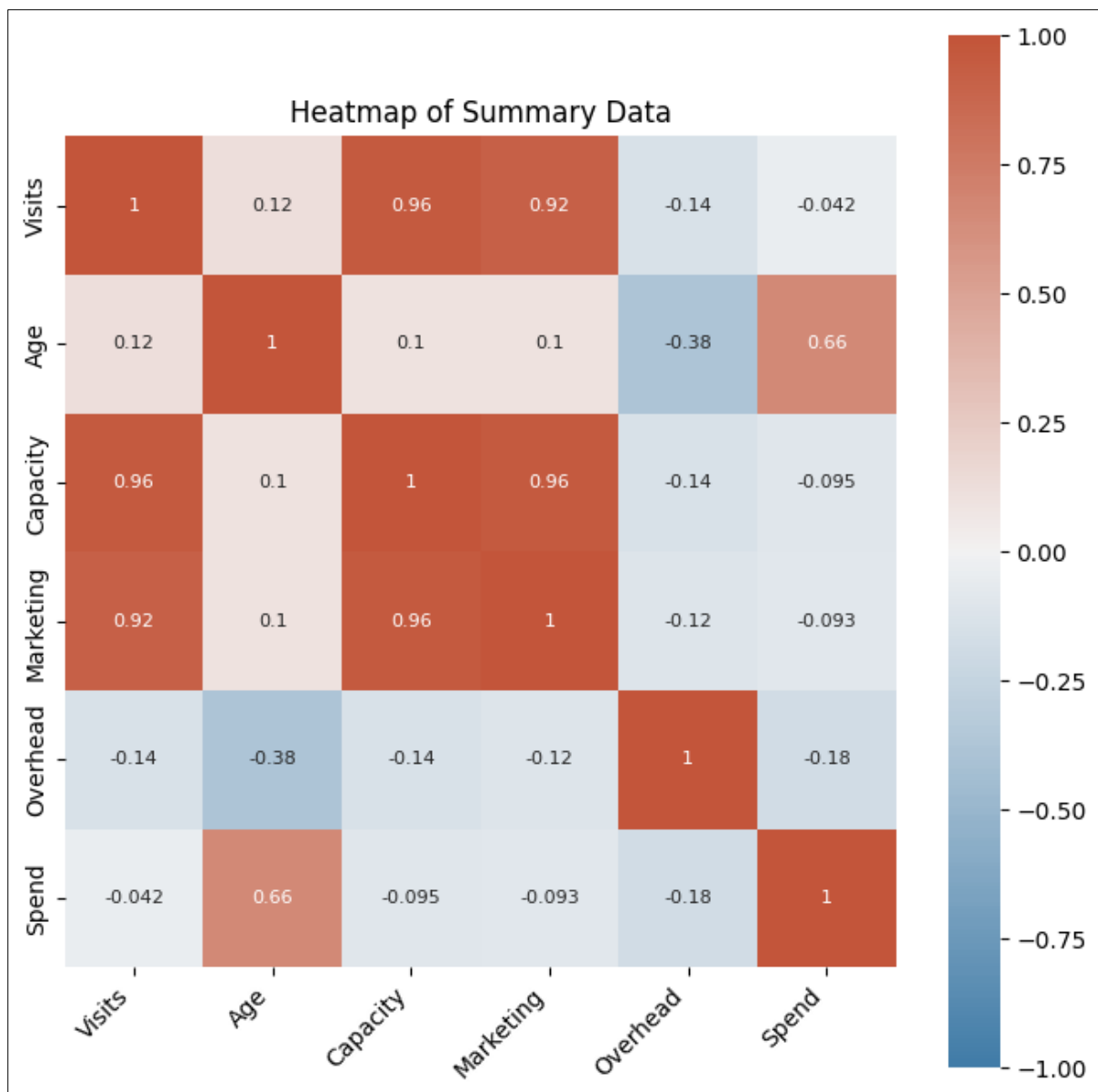
Figure 5. Heatmap of summary data

Visualisation 6-

Subplots make it feasible to compare the histograms of several busy cinemas side by side in a single chart. Histograms easier to read and efficient in presentation. Histograms are invaluable tools for distribution analysis as they provide a visual representation of the frequency distribution of a dataset. The spread or variability of the data can be assessed by examining the width and shape of the histogram. A wider histogram indicates greater variability in the data, while a narrower histogram suggests less variability. Additionally, the presence of outliers or extreme values can be visually identified in the tails of the histogram (FasterCapital, 2024).

In the figure 6, four subplots are displayed. They correspond to each of the cinemas in the high volume category. Cinemas YBS, VPG and VJV are seen as a bell-shaped curve. That is a symmetrical shape and it means that most observations are clustered around the average. They can be said to be approximately normal but slightly skewed. If the bins of the histogram are mostly towards the right side, it suggests that most cinemas had a high number of visits, and the data is skewed to the right. If the bins are mostly towards the left side, it suggests that most cinemas had a low number of visits, and the data is skewed to the left. By observing the shape of the histogram, analysts can assess the degree and direction of skewness in the data.
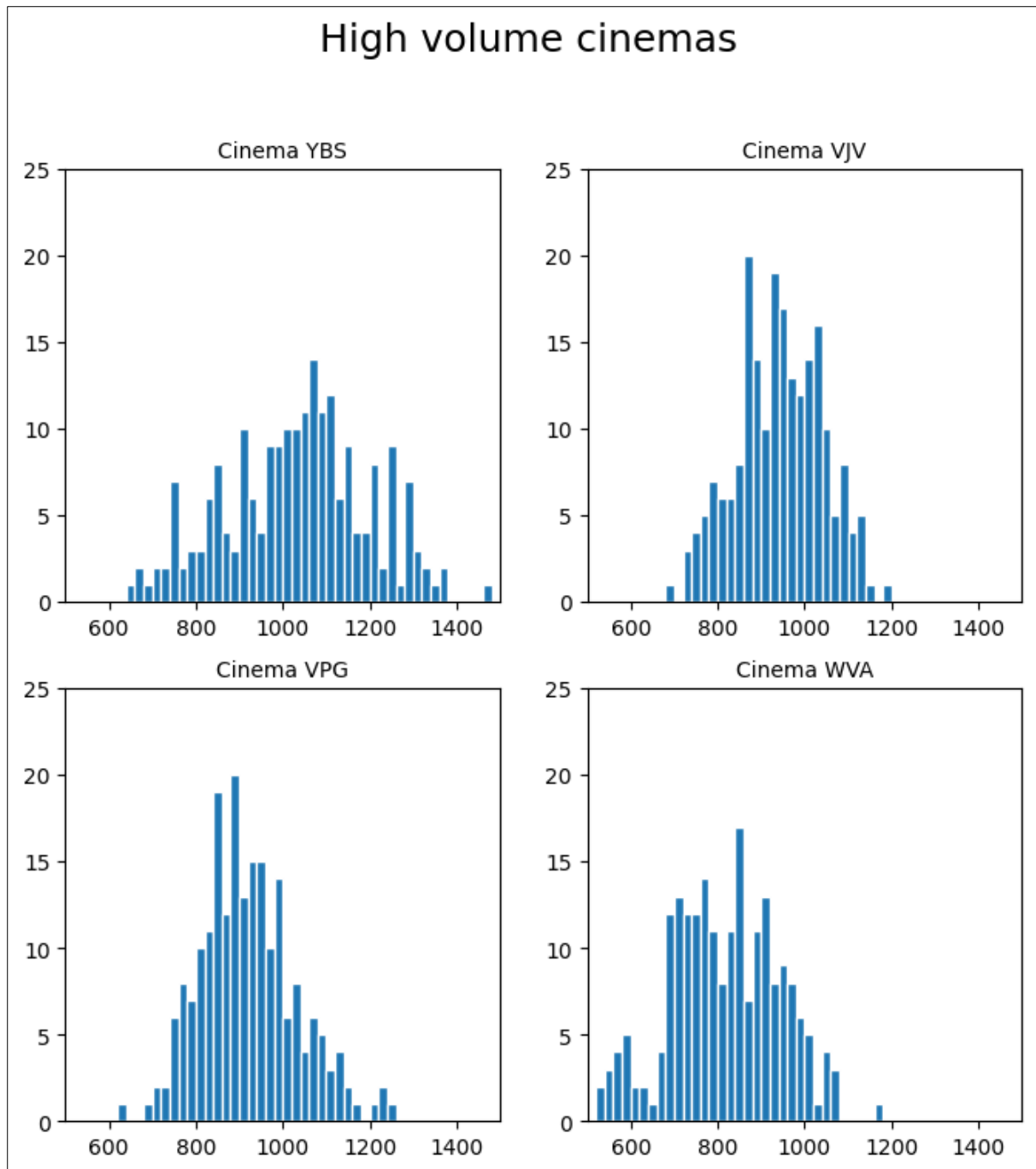
Figure 6. High volume cinemas subplots

Visualisation 7-

Scatter plots offer a visual way to confirm the relationships between variables that have been identified. They enable to examine the data points in person and see how well they match the predicted trends in light of the correlation coefficients. Scatter plots serve to illustrate the links between visitors, capacity, and marketing spend in the context of cinema operations by displaying the relevant variables against one another. These parameters show positively correlated scatter plot.

This scatter plot illustrates visits versus marketing expenditure , number of visits to a cinema and its seating capacity and last but not the least capacity versus marketing expenditure. This enables us to explore the impact of marketing efforts on cinema visitation , how cinema seating capacity relates to marketing investment is examined. As all three parameters are interrelated to one another, they have been taken in a single graph. Figure 7.2. shows that 'pan' tool is used and it is zoomed out causing the points to appear very close to one another. Also hover is used to check exact value of visits.
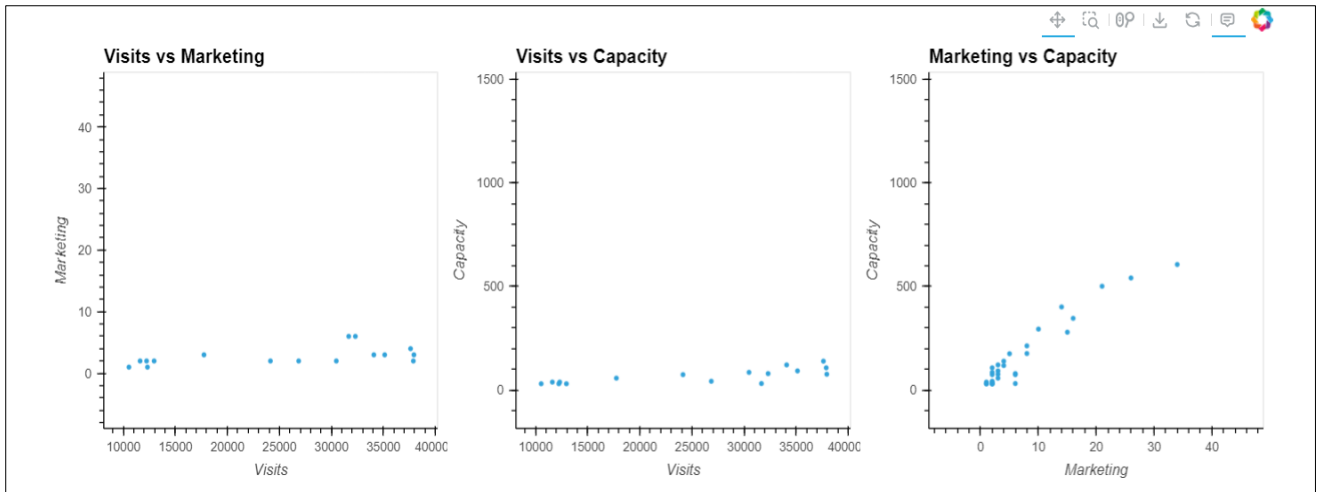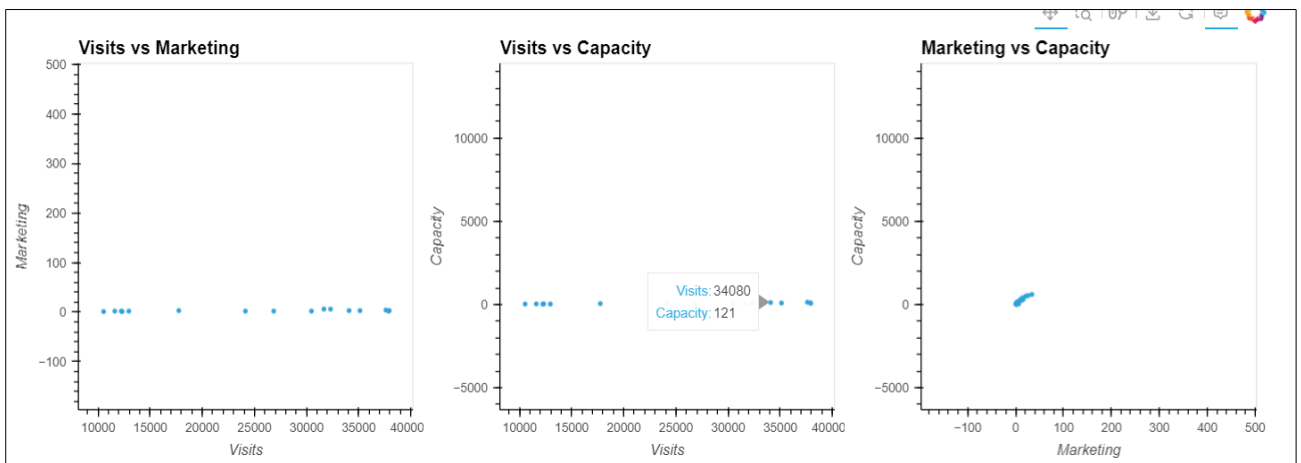
Figure 7.1. Summary data scatterplot

Figure 7.2. Summary data interactive scatterplot

Visualisation 8-

It is possible to find recurrent patterns or trends in weekly data by breaking them down. By decomposing weekly data, we can separate out the underlying trend from seasonal variations, noise, and irregularities in the data. Autocorrelation analysis complements decomposition by examining the persistence of patterns over time, verifying the presence of seasonal cycles, and assessing the effectiveness of decomposition models in capturing the underlying dynamics of the data. It measures the degree of correlation between a time series and a lagged version of itself at different time intervals, known as lags.

Cinema TPY depicts perfect seasonal fluctuations. It rises and dips alternatively at regular and predictable intervals. The data shows least variance as all data points are plotted almost on top of the line of best fit. This might be happening because of promotional offers or events taking place exactly every two weeks. The increase in visits is easy to be predicted as well.
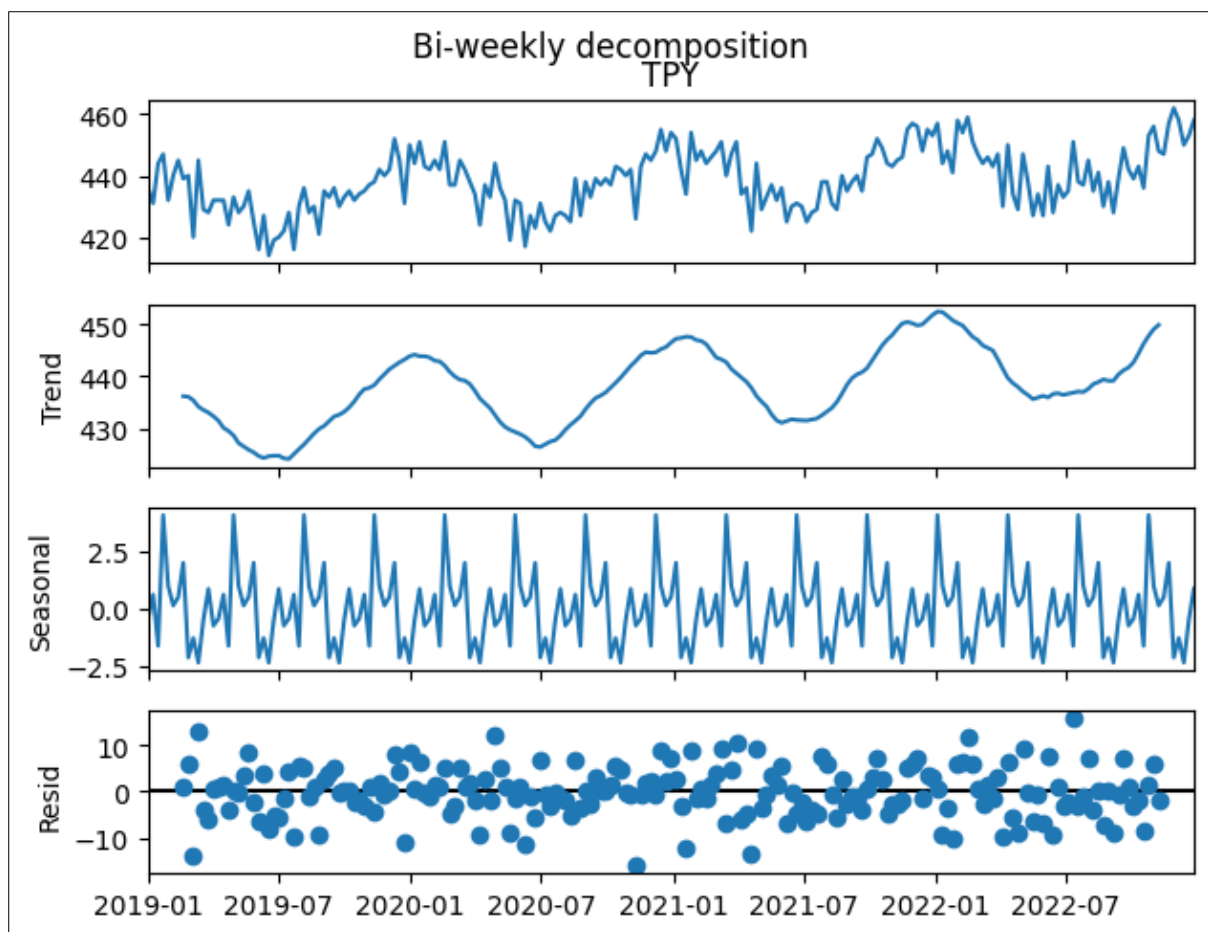
Figure 8. Bi-weekly decomposition

Critical Review

I used time series analysis throughout the coursework to investigate and comprehend a variety of datasets. This critical evaluation covers what I learned from the module, how I used it for the coursework, and examples of visualisation best practices. I have researched in depth about the segmentation of the data and its purpose. I have plotted variations of line plots for high, medium and low volume cinemas. The concepts like seasonality, rolling average and autocorrelation have been included. In the colab notebook I have tested and plotted many extra graphs as well. But, out of those, I have chosen the ones that display the most interesting results and effectively explained my findings. In this module I have learned the concepts of time series analysis, going over topics such as forecasting methods, autocorrelation, trend identification, and decomposition. I gained knowledge about the significance of pattern recognition, data preprocessing, and model selection for analysis. I used the skills I learned in the module to evaluate real-world datasets in the coursework, such as evaluating seasonal customer visits, establishing correlations between marketing of cinemas and its capacity. I used autocorrelation to find biweekly decomposition of period of 14 days to separate trend and seasonal components. Moreover, I have given sufficient information about high, medium, and low-volume cinemas. The data segmentation along with trendlines and expenditure data will help the client in further decision-making processes. I gained proficiency in utilizing programming tools such as Python and libraries like Pandas, NumPy, and Matplotlib for conducting analysis and generating visualizations. This learning experience has furnished me with valuable tools and capabilities that I can apply in future academic and professional endeavours.

Conclusions

➢ In this dataset there are four high volume cinemas which show steady increase throughout all years. They are 'YBS', 'VJV', 'VPG', and 'WVA'. They have a total sum of weekly visits above 150000. The medium volume cinemas are from the sum range between 50000 and 150000. They are 'TJN', 'WBK', 'TPY', 'UVQ', 'UDD', 'SJE', 'RPQ', and 'TVJ'. The low-volume cinemas are in the threshold even below 50000.

➢ In the moving average period of every 2 weeks uniform fluctuation is observed for cinema TPY. It shows seasonality and it seems to be predictable.

➢ The sharp peaks in the high volume line plot may be the result of a number of things, including the release of highly anticipated films, marketing, special occasions, holidays, or other outside influences. Cinema YBS has the topmost count of weekly visits.

➢ A large chunk of the data consists of low-volume cinemas and they do contribute to the overall marketing and overhead spend. Hence, it is crucial to know what went wrong or what seemed right. These are opened a bit later in year 2020- BKK, SDT, AKA, BQV, CCX, CWN. And they are performing exceptionally well. ZQL and BWF unfortunately have gone downhill. There is a steep decline in both those cinemas since the beginning of 2020. They have got shut down.

➢ The heatmap which is in statistical terms, a correlation matrix, suggests that cinemas looking to increase their visitor numbers might consider expanding their seating capacity. Since there is strong positive correlation between visits and capacity and marketing. However, other factors such as location, movie selection, and overall customer experience also play crucial roles in attracting visitors. It's important to note that while the correlation is strong, correlation does not imply causation. Simply increasing seating capacity may not guarantee an increase in visitation if other factors remain unchanged.

➢ The correlation also emphasizes how crucial intelligent marketing campaigns are in increasing moviegoer attendance. The cinemas might think about spending more money on marketing, particularly if their goal is to draw in more customers or publicize upcoming events or film releases.

➢ According to the grouped plot of high volume cinemas, there were about 20 days when the visits were between 800 to 1000.

➢ The high volume cinemas are almost normally distributed with slight skewness.

References

Brush, K. (2023). *data visualization*. Retrieved March 2024, from TechTarget:
     https://www.techtarget.com/searchbusinessanalytics/definition/data-visualization

FasterCapital. (2024, March 20). *Histograms: Visualizing Data Distribution through Absolute*
     Frequency. Retrieved March 2024, from FasterCapital:
     https://fastercapital.com/content/Histograms--Visualizing-Data-Distribution-through-
     Absolute-Frequency.html

Stevens, E. (2023, August 31). *What Is Data Visualization and Why Is It Important? A*
     Complete Introduction. Retrieved March 2024, from CareerFoundry:
     https://careerfoundry.com/en/blog/data-analytics/what-is-data-visualization/