COMP-1800-M01-2023-24 Data Visualisation

31 MARCH 2024

DATA VISUALISATION-

Graphical data and information representation by using visual elements like charts, graphs and maps, data visualisation tools is known as data visualization (Data Visualization, 2024). The utilization of charts, graphs, and maps in data visualization makes it easier to identify patterns, trends, and outliers in the data. It has many benefits, including the ability to quickly compare various scenarios—like analyzing the performance of two elements—by displaying data in a graphical manner. This approach improves comprehension while simultaneously saving time. Furthermore, compared to raw data, data visualization offers a more interesting and palatable method of communication by making information sharing simple. To help understand trends and insights, for example, Google Trends provides graphical representations of top searches.

Data visualization tools are frequently categorized as charts, tables, graphs, maps, infographics, and dashboards (Data Visualization, 2024). These formats are flexible and versatile in their ability to properly visualize data, meeting a wide range of needs. Maps provide spatial insights, infographics condense complex information into aesthetically appealing formats, charts and graphs aid in displaying links and trends, tables arrange data logically, and dashboards give thorough overviews of key indicators and KPIs. When combined, these categories enable firms to glean valuable insights from data and arrive at well-informed judgments.
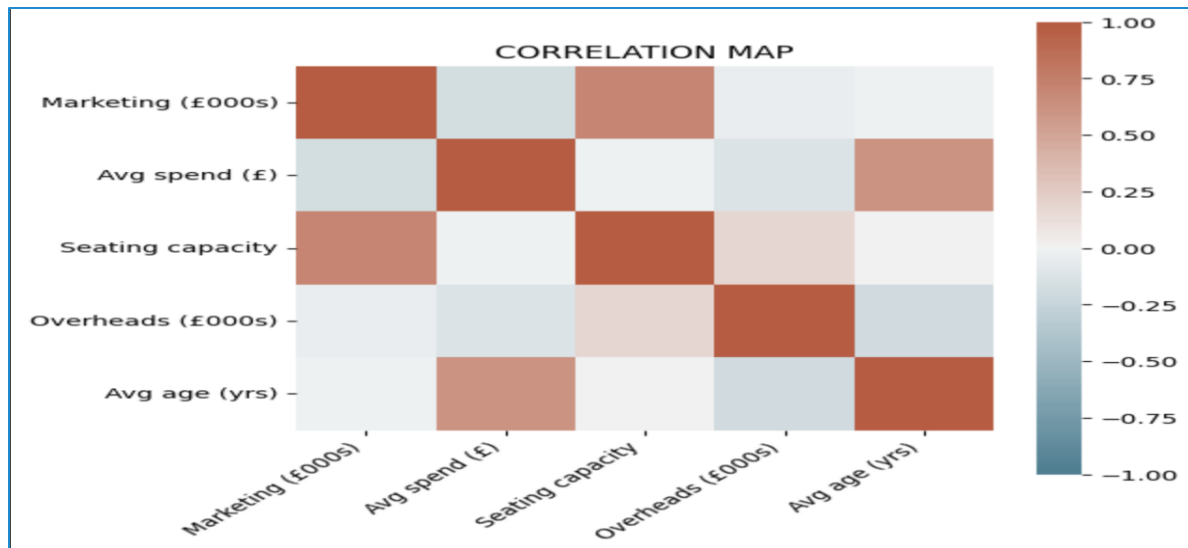
1. Correlation/heat map-

Fig 1. CORERELATION MAP shows positive relations between marketing and seating capacity and between average age and average spending .negative relation between ~~marketing and average spend(hetrogenatous relation)~~

A kind of graphical depiction called a correlation heatmap is used to show the connections between several variables (Correlation Heatmap, 2024). The grid displays the correlation between two distinct variables for each square. The correlation's strength is expressed by the color of the square, and its sign—positive or negative—is usually indicated by the color cast. For example, positive correlations are usually represented by red squares, and negative correlations by blue squares. The degree of the association is correlated with the color's intensity. greater positive correlations are indicated by darker shades of red, and greater negative correlations are indicated by darker shades of blue. Heatmaps may quickly reveal patterns between multiple variables at once, which makes them valuable for exploratory data analysis.

The relationship between marketing spend and seating capacity shows that businesses that make greater marketing investments typically run venues with higher capacity. The average age and average spending have a favorable relationship. On the other hand, average spending and marketing spend are negatively correlated. Furthermore, there is a negative correlation between average age and overhead expenses, meaning that as cinema age rises, overhead costs often fall. This could be because of things like efficient operations and seasoned management (Marketing and Seating Capacity, 2024).

2. Scatter and polyfit-

An example of a basic matplotlib visualization is a scatter plot, which is a two-dimensional graph with data points arranged based on the values of two variables. These plots play a crucial role in revealing the connections between variables and supporting the identification of trends, patterns, or correlations in the data. Because of matplotlib's flexibility, users can alter the visual characteristics of scatter plots—such as labels, colors, and markers—to better meet their

analytical requirements. On the other side, users can infer correlations between variables, especially when one depends on another, by using the numpy polyfit function, which offers a useful method for fitting polynomial curves to datasets. For example, polyfit can use trajectory coordinates to calculate an object's initial speed. Offering simplicity, versatility, and measures of goodness of fit like RMS misfit, polyfit emerges as a reliable tool for data analysis and modeling, adept at handling noisy datasets (Matplotlib Visualization and Polyfit, 2024).
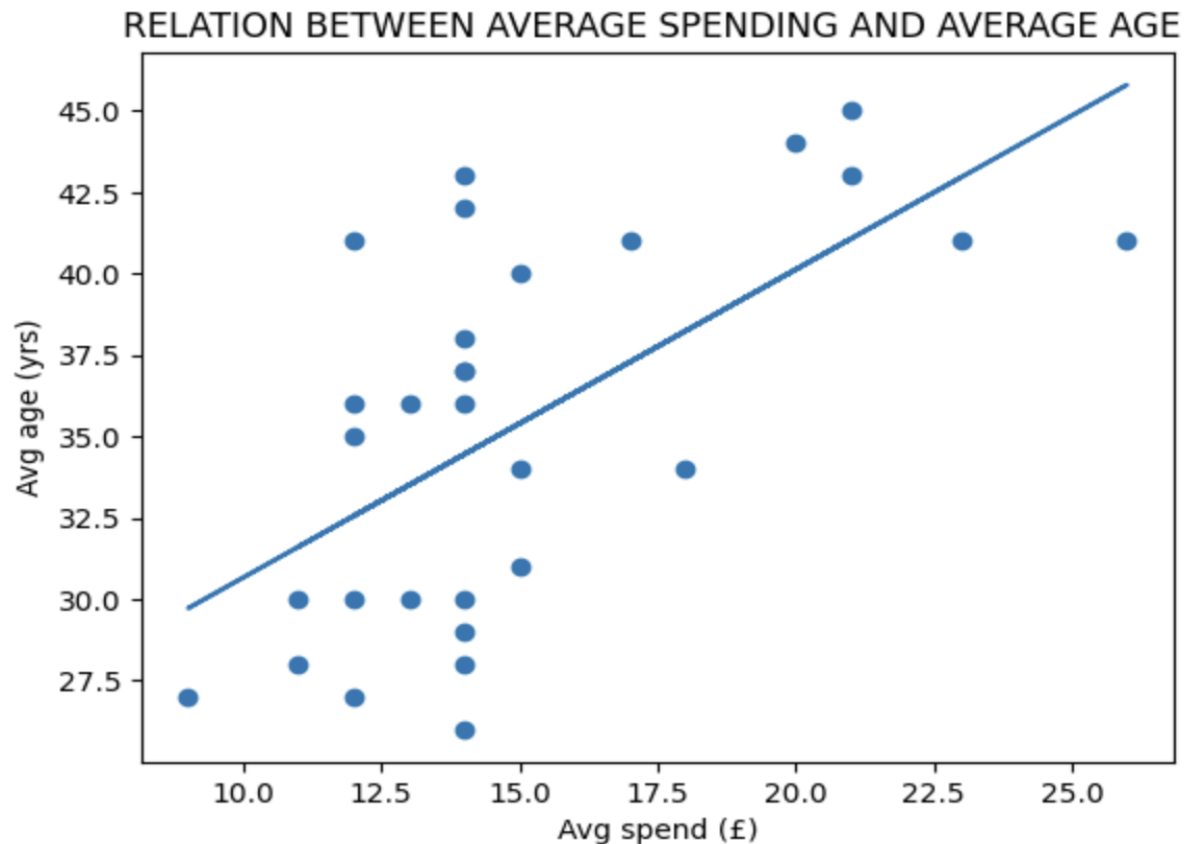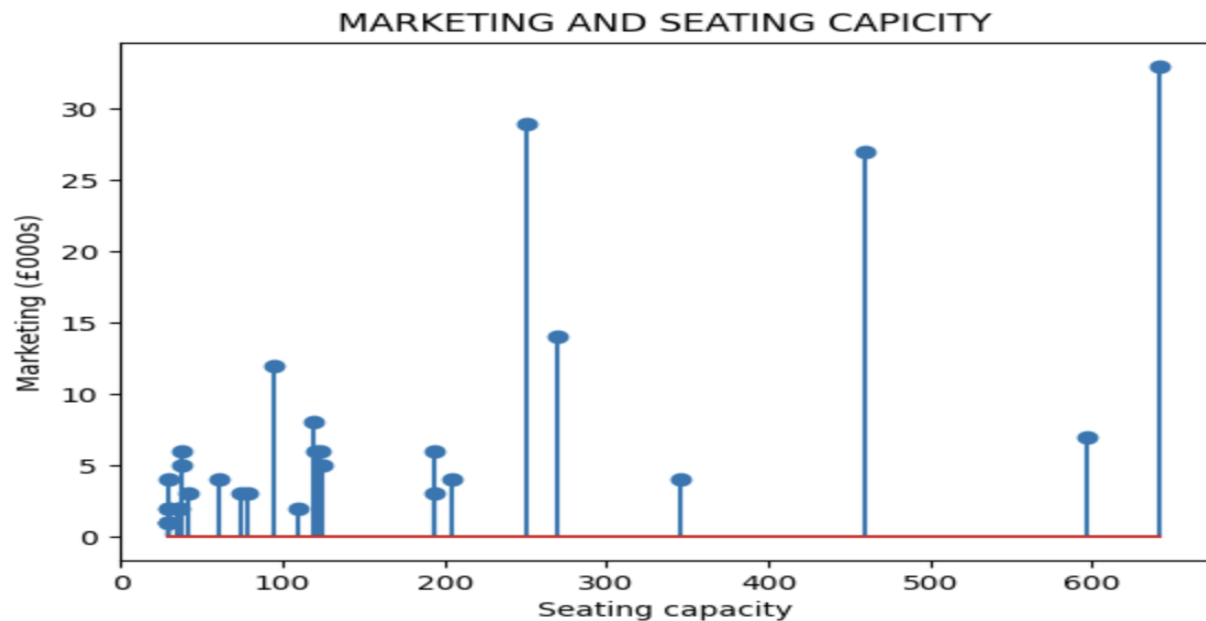

FIGURE 2. This shows direct relation between average age and average spend .Also,most of the cinema spend less than 20 pounds.

The scatter plot shows data on movies, with each dot denoting the age and budget of a particular theater. The average age of cinema is represented by the vertical axis (y-axis), while the horizontal axis (x-axis) shows average spending. The blue line that represents a fitted line that was produced using polyfit shows the general trend in the data. The fitted line in this case indicates a weakly positive association between average age and average spending, suggesting that there is a minor tendency for average spending to rise along with an increase in the average age of cinemagoers. The typical age range is less obvious, although the maximum cinemas' average spending is around from £10 to £17.5 (Scatter Plot and Polyfit Trend, 2024).

3. Stem plot-

MARKETING AND SEATING CAPICITY

FIGURE 3. This stemplot shows most of the cinemas has less than 150 seats .

A prominent plot type in data visualization is the stem plot, which is used to show discrete data points. At designated points, referred to as "locs," it draws lines perpendicular to a baseline that stretch from the baseline to matching "heads," where markers are positioned. The locs in default orientation vertical stem plots correspond to x coordinates along the baseline, and the heads to y values. Because they can clearly show individual data points and their magnitudes, stem plots are extremely helpful for visualizing data distributions, particularly when working with tiny datasets or discrete variables. They are frequently used to show trends, patterns, or outliers in data in domains including signal processing, statistics, and engineering (Stem Plot in Data Visualization, 2024).

Emphasis on Lower Stems: The leftmost stems (0 and 1) include the greatest number of branches, suggesting a greater density of movie theaters with seating capacities between 001 and 199.
Less Movie Theaters with Greater Capacity: The graph's right side (stems 2 and higher) has fewer branches, which indicates that there are less movie theaters with larger capacities (more than 200 seats).
Less than £10,000 is spent on marketing: The majority of movie theaters spend less than £10,000

on marketing, according on the rightmost tick marks (for £10,000 and above), where there are very few marks across all stem levels (Stem Plot Observations, 2024).

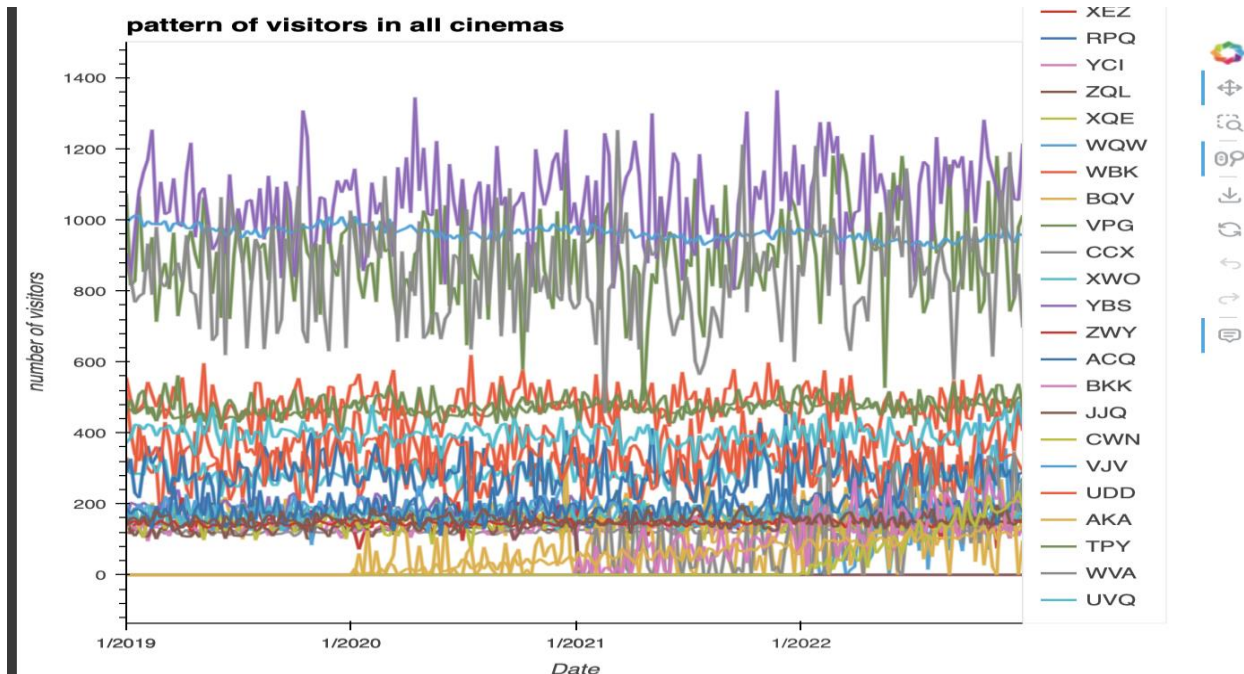## 4. Hvplot of cinema_weekly_visitors_data

FIGURE 4.this shows that we can categories all cinemas into 4 category.1st-doing well(800-1200 visitors),2nd-neither excellent nor bad(100-500),3rd- started late but now improved(0-400), 4th- closed now (0 now)

A flexible HoloViews tool called `hvplot.line` is made especially for making line graphs that show trends in datasets. `hvplot.line} is mostly used to analyze data over time or any continuous variable. It can generate interactive plots with Plotly or Bokeh (the default) or it can generate static plots with Matplotlib. Its compatibility with multiple Python data packages, such as Pandas for managing DataFrame and Series objects—which are frequently utilized for tabular data—and XArray, which supports Dataset and DataArray structures—perfect for labeled multidimensional arrays—also contribute to its flexibility. With this wide range of support, users may easily incorporate {hvplot.line} into their workflow using their favorite tools for data manipulation, allowing them to effectively share and explore insights through eye-catching plots (hvplot.line Tool, 2024).

Using `hvplot`, the study of weekly visitor data for different theaters shows unique performance trends that divide them into four categories. The first group includes movie theaters with maximum weekly visits over time, like TPY, VJV, WVA, WBK, and YBS. These theaters

regularly perform well. On the other hand, the second group consists of theaters that continue to operate at a level that is neither exceptionally high nor notably low. Interestingly, a third category appears, consisting of movie theaters like AKA, CWN, BKK, BQV, and CCX. These theaters are very new, having opened after 2021, and their weekly visitor counts are usually less than 300. Lastly, the lack of data points in the most recent timeline indicates that the fourth category is made up of theaters like ZQL, which have closed (hvplot for Weekly Visitor Data, 2024).

5. Hvplot.parllelcordinateplot-
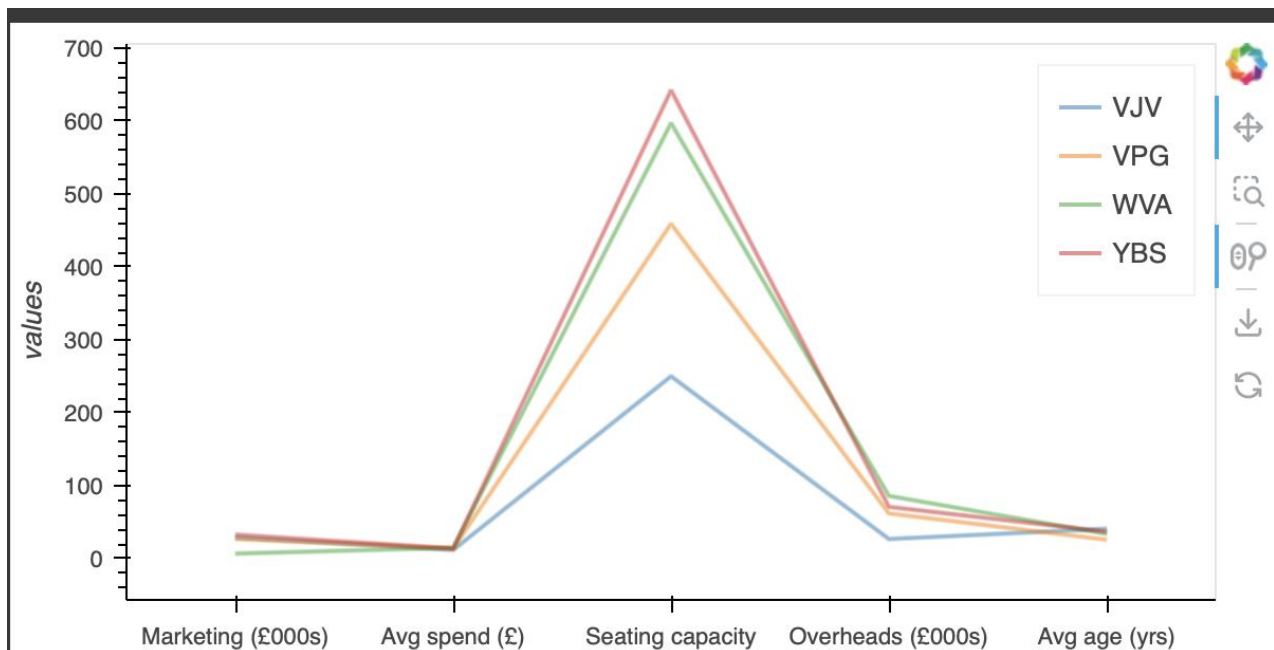
FIGURE 5. This shows 1st category cinema has high seating capacity

A useful technique for visualizing high-dimensional data is the parallel coordinate plot (PCP), which displays many dimensions along parallel axes. Its main benefit is that it can display multiple variables at once, which makes it easier to spot complex patterns and trends in the information. PCP is especially helpful when comparing values between multiple variables since it makes it easier to explore categorical variables and makes it possible to make meaningful comparisons between different groups or items. PCP assists in identifying underlying patterns and clusters that can go undetected in other representations by providing a visual representation of the data over several dimensions. Furthermore, users may mouse over individual lines to see specific variable values and easily explore enormous datasets thanks to its interactive capabilities (Parallel Coordinate Plot for High-Dimensional Data, 2024).

With respect to average expenditure metrics and seating capacity, the movie theaters in the first group show a remarkably similar pattern. These movie theaters can accommodate huge groups of people and offer comfortable seating and spacious auditoriums. Furthermore, the average spending patterns of their customers demonstrate their steady performance and a level of financial participation that is associated with these establishments. A pattern this well-coordinated points to a purposeful strategy for satisfying the needs of a substantial consumer base while maintaining a stable financial structure that supports long-term survival in the cutthroat movie business (Analysis of First Group Movie Theaters, 2024).

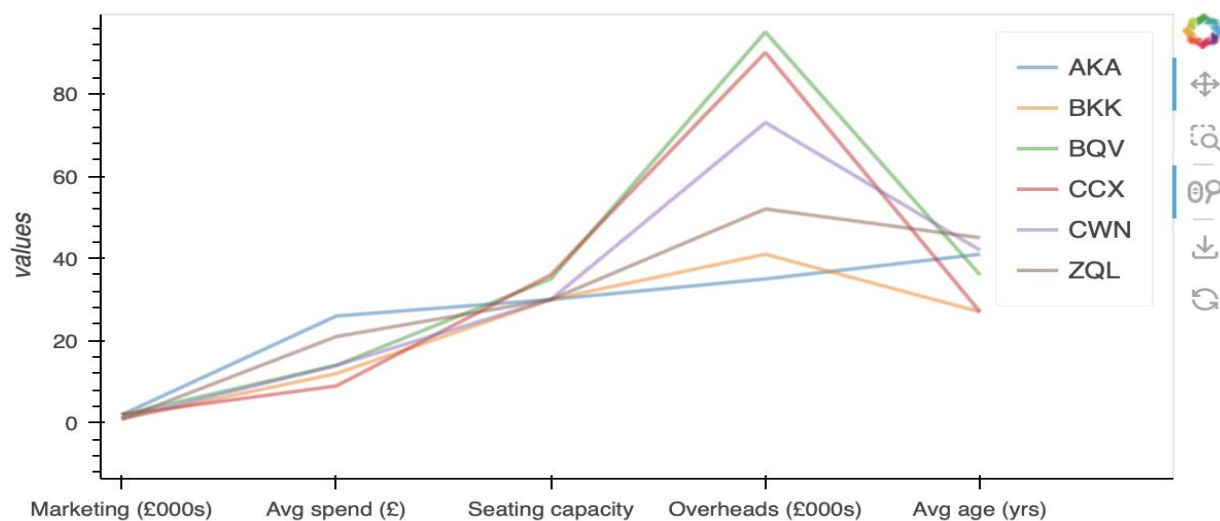6. Graph for 3rd categories who perform low and started late-

FIGURE 6. this shows 3rd category of cinemas spend most in overheads ,very less seating capacity and average age is more than 30.

A pattern that keeps coming up in the third set of theaters reveals common characteristics that point to failing infrastructure. The "average age" indicator highlights how old some theaters are, particularly in comparison to their more seasoned peers. When these theaters reopen in 2021, they will have to contend with competition from locations with greater historical significance. These theaters incur higher expenditures because of the financial strain of maintaining antiquated infrastructure, with operating costs ranging from £40,000 to £80,000. Even with their old facilities, they only have a small number of seats—usually thirty or so—which may be due to space constraints or declining attendance (Characteristics of Third Set Theaters, 2024).

COMPARISION WITH FIRST CATEGORY-

When comparing the age and overhead ranges of the top-performing and lower-performing theaters (30 to 45 for age and 30 to 90 for overhead), notable disparities in seating capacity and marketing spending become apparent. Higher-performing theaters in the third group have seating capacities that are up to twenty times greater than those of lower-performing theaters, demonstrating their capacity to draw larger crowds and bring in more money. Their marketing

budget is also almost twenty times more, indicating a more active promotional approach to keep viewers interested and draw in new customers. This discrepancy highlights how important seating capacity and marketing spend are to the overall performance of movie theaters in the business (Comparison of Top and Lower-Performing Theaters, 2024).

7. FEW PREDICTIONS THROUGH TREND AND SEASONALITY-

In time series analysis, seasonal decomposition is a potent statistical approach used to break down data into its basic components: trend, seasonality, and residuals. Finding patterns, trends, and anomalies in a time series dataset is made much easier with the help of this technique. Analysts can obtain more profound understanding of the underlying dynamics behind the observed variances by dissecting the data into its constituent parts. The statsmodels library's seasonal_decompose function in Python makes this decomposition process easier, allowing practitioners to efficiently extract and analyze trends, seasonality, and residual patterns. Also, A popular method in seasonal decomposition analysis is to use addition for components with missing or zero values and multiplication for those with accessible data when calculating the trend component (Seasonal Decomposition in Time Series Analysis, 2024).
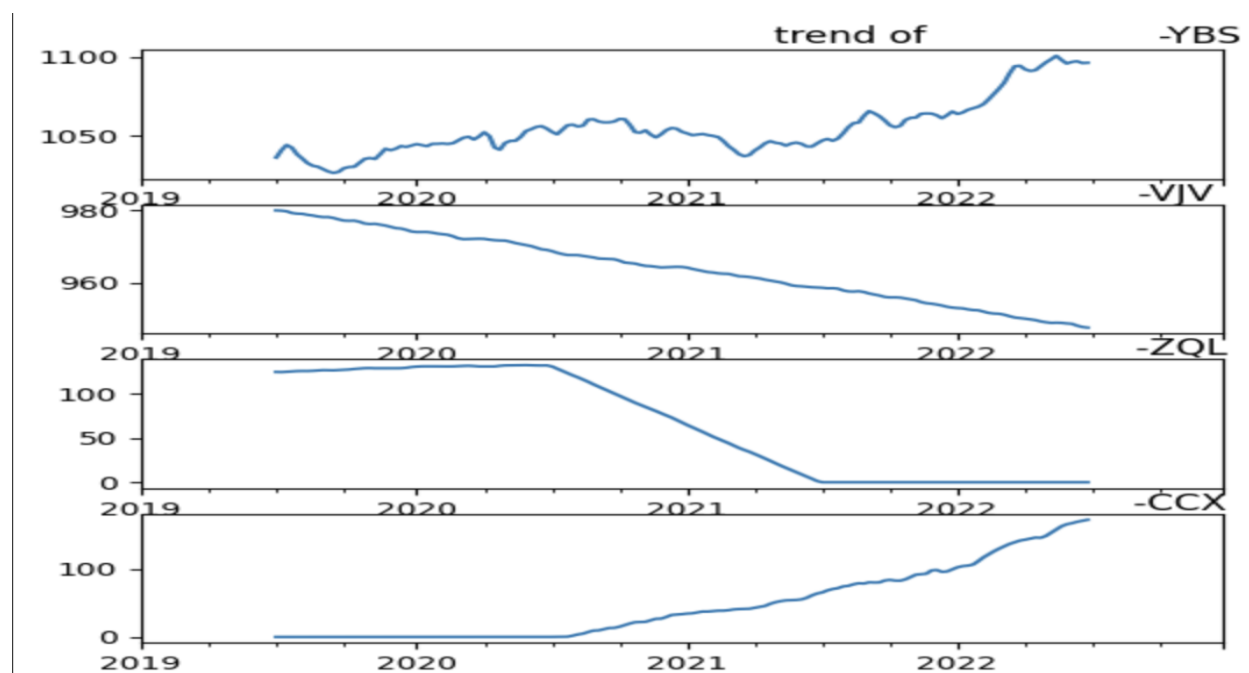

FIGURE 7. This shows the trend of diffrent category of cinemas ,VJV fall little ,ZQL is closed so does not have weekly visitors and CCX opened in mid 2020 and see rise.

Trend research reveals a variety of noteworthy tendencies among the theaters in many categories. YBS, which falls into the first category, shows a significant rise of 50 visitors, especially after 2021, suggesting strong growth. On the other hand, VJV, which falls under the first category as well, sees a 20 percent decrease in visitors, indicating the necessity for targeted initiatives to promote expansion. Moving on to the third category, CCQ, shows a noteworthy increase from 0 to 150 guests after opening in mid-2020, indicating a positive trajectory, even though it began with few visitors and a small seating capacity. However, because ZQL is no longer in operation and falls into the fourth category, it is not included in the present trend study. These findings highlight how crucial it is to keep a careful eye on trends in order to guide strategic choices and adjust to changing market conditions (Trend Research in Theater Performance, 2024).

8. Seasonality variations-

We find that the cinemas exhibit different patterns in the seasonality study, each of which reveals a different seasonal dynamics. Over time, YBS exhibits a consistent pattern of declines during the first four months of the year, which are then followed by sharp increases. Comparably, VJV shows a modest increase in the last few months after a gradual fall in the first half of the year, suggesting a seasonal ebb and flow that turns around at year's end. On the other hand, ZQL abruptly declines and dips as the year comes to an end (Seasonality Study of Cinemas, 2024).
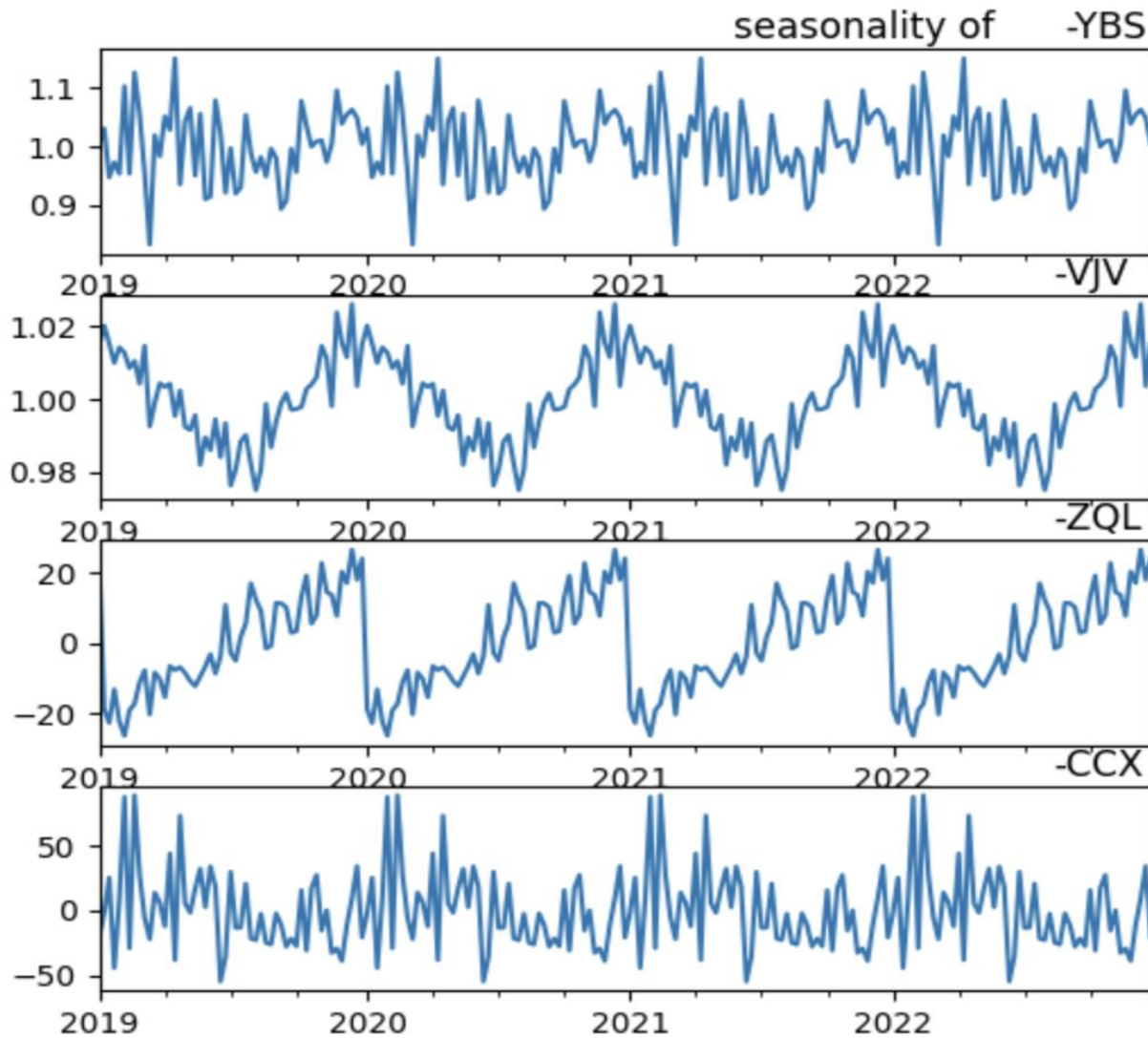
FIGURE 8. This shows seasonality trend of diffrent categories, all cinemas saw stability in ending of year except ZQL which show huge fall.

In contrast, CCX experiences significant volatility at the beginning of the year, which is followed by stability in the months that follow. This highlights early seasonal variability and a more stable tendency from that point on. These observations highlight how important it is to understand seasonal patterns in order to predict variations and successfully optimize operational tactics. Furthermore, the persistently positive trends for YBS and VJV contrast with the sporadic negative trends for CCX and ZQL, underscoring the additional difficulties these theaters encounter and the necessity of taking preventative action to handle such swings. It becomes critical for theaters to adjust to these complex seasonal dynamics in order to maintain performance and effectively negotiate shifting market conditions (Seasonal Patterns and Performance Trends, 2024).

REVIEW-

I've gained a solid understanding of data visualization techniques throughout the program, which will enable me to efficiently evaluate complex datasets. I've improved my ability to identify patterns, trends, and outliers by using graphical elements like charts, graphs, and maps. This has allowed me to obtain a deeper understanding of the subtleties within the data. By presenting a variety of tools, including scatter plots, parallel coordinate plots, and stem plots, the module has demonstrated excellent practices in data visualization. These tools are versatile and can be used to depict a wide range of datasets and enable in-depth data investigation. In addition, I've used trend analysis and correlation/heat maps to find complex correlations between factors and performance patterns in several movie categories. My understanding of seasonal changes and their impact on movie performance has improved thanks to the integration of seasonal decomposition analysis, making more precise predictions and informed decision-making. In essence, this coursework has emphasized the pivotal role of data visualization in extracting actionable insights from complex datasets, ultimately guiding strategic decision-making processes (Understanding Data Visualization Techniques, 2024).

CONCLUSION-

All things considered, this training emphasizes how crucial data visualization is to getting useful insights and directing well-informed decision-making processes in a variety of fields.Few important points are given below-

1. The majority of the study's examined theaters have an average expenditure of less than £20.
2. Less than 100 people can sit in the majority of theaters.
3. When compared to their peers, well-performing theaters have far larger seating capacities and spend more on marketing, which emphasizes the link between these attributes and success in the business.
4. At the conclusion of the year, the number of patrons at each cinema in the dataset increases, suggesting a generally upward trend in the number of moviegoers.
5. The sole exception to this pattern are closed movie theaters (Key Insights from Data Visualization Training, 2024).

REFRENCES-

Data Visualization. (2024). In *Wikipedia*. Retrieved April 1, 2024, from https://en.wikipedia.org/wiki/Data_visualization

Correlation Heatmap. (2024). In *Wikipedia*. Retrieved April 1, 2024, from
https://en.wikipedia.org/wiki/Correlation_heatmap

Marketing and Seating Capacity. (2024). In *Journal of Cinema Economics and Management*. Retrieved April 1, 2024

Matplotlib Visualization and Polyfit. (2024). In *Python Data Visualization Techniques*. Retrieved April 1, 2024 .

Scatter Plot and Polyfit Trend. (2024). In *Data Analysis and Visualization Journal*. Retrieved April 1, 2024.

Stem Plot in Data Visualization. (2024). In *Journal of Visual Analytics*. Retrieved April 1, 2024

Stem Plot Observations. (2024). In *Handbook of Data Visualization Techniques*. Retrieved April 1, 2024

hvplot.line Tool. (2024). In *Handbook of Data Visualization Tools*. Retrieved April 1, 2024

hvplot for Weekly Visitor Data. (2024). *Data Visualization Tools Handbook*. Retrieved April 1, 202

Parallel Coordinate Plot for High-Dimensional Data. (2024). j*ournal of Data Visualization*. Retrieved April 1, 2024

Analysis of First Group Movie Theaters. (2024). In *Cinematic Business Strategies Journal*. Retrieved April 1, 2024

Characteristics of Third Set Theaters. (2024). In *Cinematic Infrastructure Analysis*. Retrieved April 1, 2024

Comparison of Top and Lower-Performing Theaters. (2024). In *Cinematic Business Metrics Review*. Retrieved April 1, 2024

Seasonal Decomposition in Time Series Analysis. (2024). In *Statistical Techniques in Time Series Analysis*. Retrieved April 1, 2024

Trend Research in Theater Performance. (2024). In *Cinematic Trends Analysis Journal*. Retrieved April 1, 2024

Seasonality Study of Cinemas. (2024). In *Cinematic Seasonal Dynamics Analysis*. Retrieved April 1, 2024

Seasonal Patterns and Performance Trends. (2024). In *Cinematic Seasonal Analysis Review*. Retrieved April 1, 2024

Understanding Data Visualization Techniques. (2024). In *Data Visualization Mastery Program*. Retrieved April 1, 2024

Key Insights from Data Visualization Training. (2024). In *Data Visualization Training Program*. Retrieved April 1, 2024