

Data Visualisation

ChrisCo Visualisation Report

Akeem Babalola

Introduction

This report is a compilation of tasks completed for a Data Visualisation module that encourages the exploration of data to visualise datasets linked to ChrisCo, a fictional company that manages a chain of cinemas across the UK. 'Data visualisation is the graphical representation of information and data. In the world of big data, data visualisation tools and technologies are essential for analysing massive amounts of information and making data-driven decisions' (Tableau, 2024). The graphics seen throughout this report have been visualised using the programming language, Python and respective libraries.

This written piece delves into the investigation of datasets by comparing them accordingly to highlight characteristics that inform the condition of cinemas, and allow for conclusions to be drawn for effective business decisions. ChrisCo are mostly interested in the analysis of cinemas, particularly those of high and medium customer volumes as they are likely focused on aspects of the business with significant gains.

The task entailed the exploration of six datasets: the weekly number of customer visits for each cinema over a period of time, the average age of visitors at each cinema, the seating capacity for each cinema, the average annual spend on local marketing for each cinema, average annual overheads for each cinema and the average spend by visitor at each cinema, all in which were compiled into two data frames containing customer and summary data.

Bar Chart - Summary of Customer Visits

Figure 1 is a barchart for all cinemas within the customer dataset and displays each of their total weekly visitors over a time period of approximately 4 years. The data graphically suggests the difference between high ['YBS', 'VJV', 'VPG', 'WVA'], medium ['WBK', 'TPY', 'TJN', 'UVQ', 'UDD', 'SJE', 'RPQ', 'TVJ'], low volume cinemas ['ACQ', 'WQW', 'XWO', 'YKT', 'XEZ', 'JJQ', 'XQE', 'ZWY', 'YCI', 'WKL'] and very low volume cinemas ('BWF', 'ZQL', 'BKK', 'CCX', 'BQV', 'AKA', 'SDT', 'CWN') which will be referenced upon throughout the report. The assignment of these categories are based on respective thresholds of customer visits per cinema being, 125000 and above (high), between 50000 and 125000 (medium), between 25000 and 50000 (low), and below 25000 (very low).

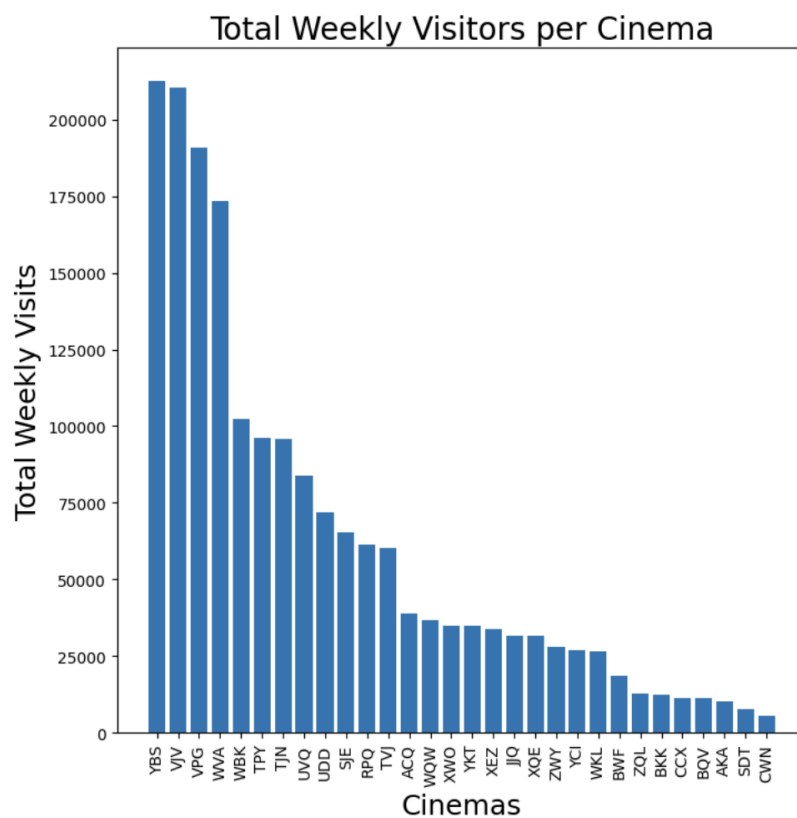


Figure 1: Bar Chart of Total Weekly Visitors per Cinema

Line Chart Analysis - Very Low Volume Customer Visits

The line chart in Figure 2 is a representation of the low volume cinemas and the trajectory of their weekly customer visits over a period of approximately 4 years, including trendlines that showcase the correlation between the variables. The screenshot suggests that between the period of 2019 and early 2020, the cinemas ['BKK', 'CCX', 'BQV', 'AKA', 'SDT', 'CWN'] received 0 number of weekly customer visits throughout which implies, these very low volume cinemas may be newly established as of 2020 and informs the relative lack of significance for analysis from ChrisCo. This also emphasises the analysis of high and medium volume cinemas to understand the optimal strategies to apply across cinemas within the portfolio.

Figure 2 also showcases an abrupt collapse in weekly customer visits for very low volume cinema's 'BWF and ZQL' towards the beginning of 2021. The weekly customers fell to 0 from the first week of 2021 and remained stagnant, which indicates the closure of cinema's 'BWF and ZQL' possibly as a result of an internal or geographical emergency that disabled customers from visiting. The trend lines for the cinemas previously mentioned differ from the remaining very low volume portfolio as they possess negative correlations that suggest a reduction in value as time goes on, which speaks to the eventual discontinuation of both cinemas.

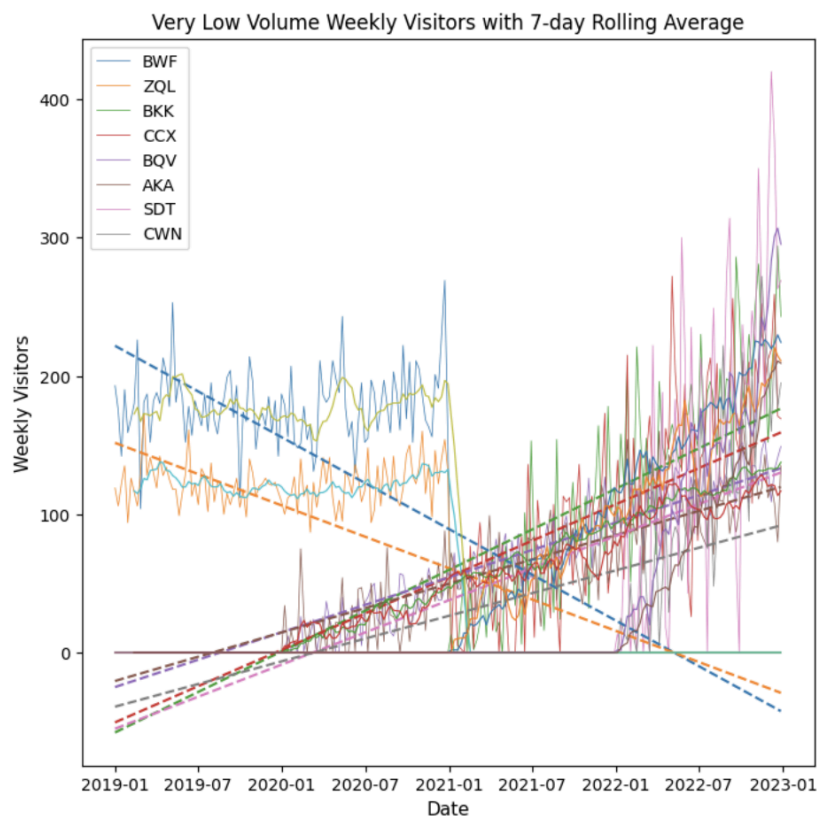


Figure 2: Line Chart of Very Low Volume Cinemas

Heatmap - Correlation between Cinemas Variables

The visualisation in Figure 3 presents a heatmap of variables relating to each cinema within the summary data, represented as numerical data linked to colours that represent different degrees of correlation. As red is associated with positive correlation, we can infer a strong correlation between the total cinema visitors and profit which generally justifies the ranking of high to low volume cinemas; the more weekly visitors a cinema receives, the greater the likelihood of converting visitors to customers which in turn increases profits.

Incidentally, there is a moderately positive correlation between a seating capacity and the annual marketing spend of a cinema. The larger the seating capacity, the greater the *potential* for a larger ‘consumer’ base which suggests the need to allocate a higher marketing budget in targeting the relevant audience size.

This also highlights the importance of brand awareness as cinemas of bigger capacity have connotations of an immersive grand experience which is attractive information for a consumer. ‘Companies can adopt a boutique approach which suggests a smaller sized business that sells luxury products or services and brands within a niche’ (Shopify, 2024). Boutique-like cinemas may allocate big budgets towards marketing to boost their image and deliver a strong impression that indirectly influences the perception of capacity. This can be inferred from the weak negative correlation between seating capacity and total number of cinema visitors as it suggests that seating capacity isn’t a significant variable in influencing a consistent number of visitors over a period of time and emphasises the need for marketing to attract customers and establish a brand.

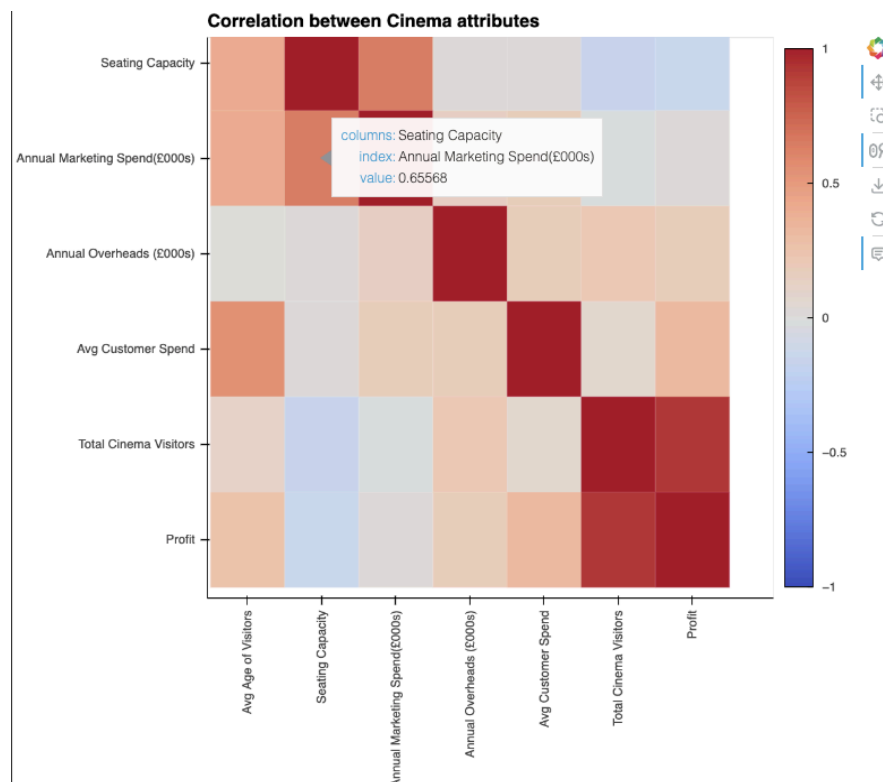


Figure 3: Heat Map of Cinema Variables - Summary Data

Radar Plot - Comparison of Cinema Variables

Figure 4 is a display of the four cinemas with the highest average of weekly visitors. The radar plot suggests a commonality between the selected, being that the average age of visitors is relatively high which coincides with their nature of high volume weekly customers as the dataset suggests that middle-aged individuals are more inclined to attend and spend at cinemas.

Cinema VJV provides a great representation of the correlation between cinemas and the variables of maintaining high volume weekly customers. VJV spends £87,000 on overheads annually which is a huge budget relative to some of the other cinemas selected. Considering the seating capacity is 184 which is relatively low in comparison to other cinemas under ChrisCo, this suggests that the nature of the cinema is boutique and luxury as the amount spent on overheads could possibly cater to aspects such as insurance and utilities that contribute to the sustenance of the cinema.

The standard of maintenance at Cinema VJV can help to establish esteem, however, the plot suggests that they require an additional campaign to reach their desired target audience. VJV spent £14,000 on annual marketing which is proportionally larger than the other cinemas with a high volume of weekly visitors. The cinema attracts the highest average age of visitors, aged 45 as seen in the dataset, and the average customer spend is £24. This implies that VJV set the price of their goods and services higher in order to make up the cost of business and lengthen their profit margin. In Figure 6, the data point for £24 is amongst the outliers in the data which reinforces the uniqueness and appeal of cinema VJV to customers.

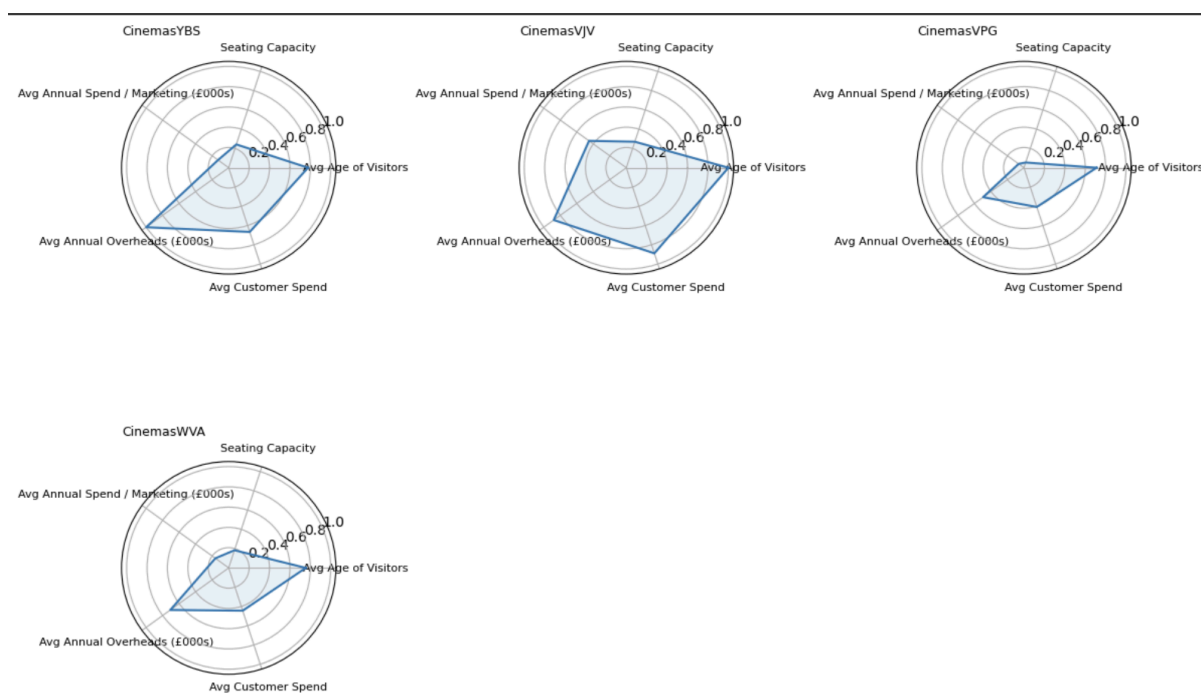


Figure 4: Radar Plot of High Volume Cinemas

The cinemas ['YBS', 'VJV'] have comparatively high average annual overheads which indicates that the *selected* invest heavily in operations and management of their businesses (*i.e. rent, utilities*), which is exclusive of the products and services they provide. This could correlate to their volume of weekly customers as it shows the cinemas are of high maintenance which is a factor that proves attractive to customers in establishing a favoured brand reputation.

As overheads are relatively high, the cinemas with a high volume of weekly visitors would generally need to increase the price of their goods and services offered in order to better their profit, which is evident through the 'Avg Customer Spend' displayed in the radar plot as this commensurate with the 'Avg Annual Overheads.' Whilst this can be considered disadvantageous as customers generally prefer lower prices, there is a correlation between the main market audience for the *selected* cinemas (31 - 36 year olds) due to their nature of prioritising additional factors such as comfort and quality within their experience.

Bar Chart - Profit Summary

Figure 5 is a bar chart that consists of the selected high and medium volume cinemas to display their respective profits. The methodology behind this begins with the total number of visitors for each of these cinemas multiplied by the average customer spend, to form the *assumed* revenue of each cinema. I have then combined the annual marketing spend and overheads to equate to the total costs for each cinema. Please note that **the total costs** are an assumption as there may be additional costs that haven't been accounted for in the dataset (*i.e. employee wages, inventory and supplies, taxes, technology, etc*).

The graph suggests that high volume cinemas are typically more profitable relative to others and that profit is generally consistent across cinemas according to their volume size. The cinemas VJV (£4,646,608 in profit over 4 years) and YBS (£3,416,024 in profit) noticeably have the highest profits which indicates their success within the ChrisCo portfolio, but additional factors such as conversion rates of visitors to 'customers', or unaccounted costs could vary the profit margin per cinema which may alter the consistency of the dataset.

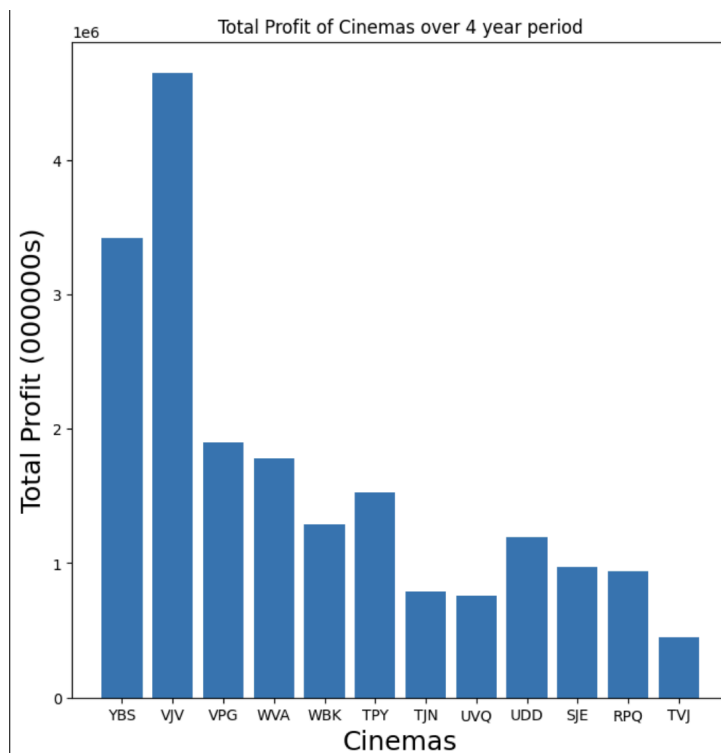


Figure 5: Bar Chart of Cinemas Total Profit (High and Medium)

Scatter Plot - Average Age vs Average Spend

The plot in Figure 6 represents the relationship between summary data in regard to a visitors average age and their average spend; the data suggests that there is a moderate positive correlation between the average age of visitors and average spend per cinema. The main demographic with the most consistent spend within the data set of cinemas are between 31 and 36 years of age which suggests that these individuals or groups are likely to have more disposable income that can be spent on cinemas. This also acts as an indicator for cinema companies in regard to marketing as this is their ideal target audience.

There are outliers that exist within the data geared towards a high average spend outside the typical 'arrangement' of values, this can be considered 'extreme.' Though they aren't significant enough to skew the dataset but are still an important factor to note as individuals approximately of age 40 may have tendencies to spend a higher average within cinemas.

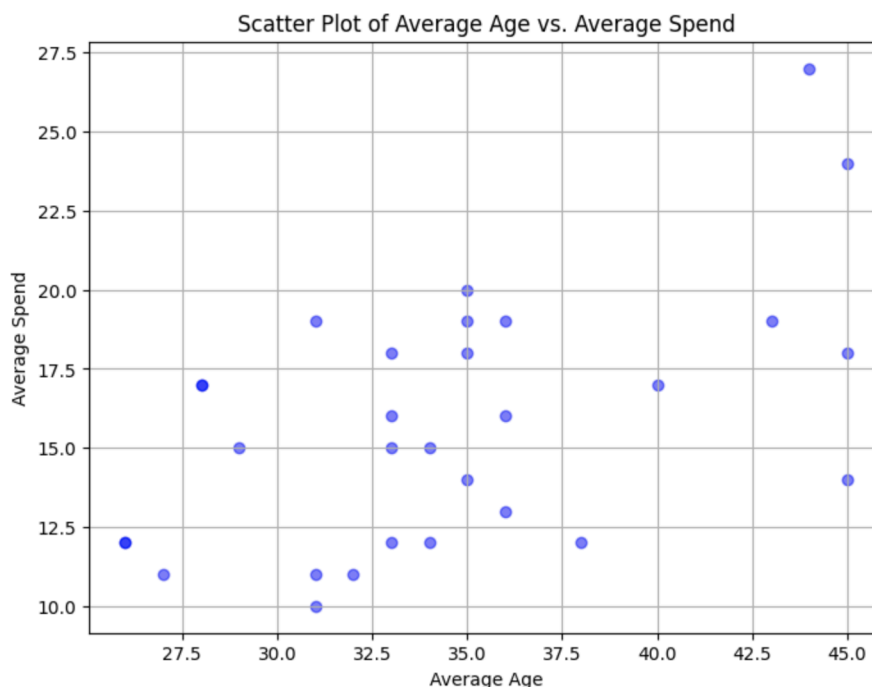


Figure 6: Scatter Plot of Average Age & Spend

BubblePlot - Marketing vs Total Weekly Visitors (vs Profit)

Figure 7 is a graphical representation of an interactive scatter plot that represents profit by bubble size and depicts the relationship between marketing and average customer spend. The plot generally suggests a weak positive correlation between annual marketing spend and the average customer spend which means that the more cinemas spend on marketing, the more visitors they can attract to spend. However, there are other factors that can also provide a competitive advantage such as the experience of cinemas, incentives or levels of accommodation that encourage visitors to spend. This is evident for Cinema YBS, the second most profitable cinema as the annual marketing spend (£4,000) is amongst other cinemas of significantly less profitability.

The bubble with the largest size belongs to Cinema VJV and reiterates the fact that they achieved the highest profits of the cinemas. At the same time, the cinema spends a relatively high amount on marketing (£14,000) and has a higher average customer spend (£24) which suggests they may be an exception to the rule and that the amount the cinema spends on marketing positively influences the connection with consumers' willingness to spend. This also indicates that VJV are optimising their marketing budget to achieve a greater return on investment (ROI) through their realised profits.

Cinema YKT can be considered an outlier within the plot as they spend the most on marketing (£31,000) and have a decent average customer spend (£19) yet receive relatively low levels of profit. This highlights the cinema's inability to market effectively to a pool of their desired target audience and potentially the high cost associated with acquiring their current visitor/customer base.

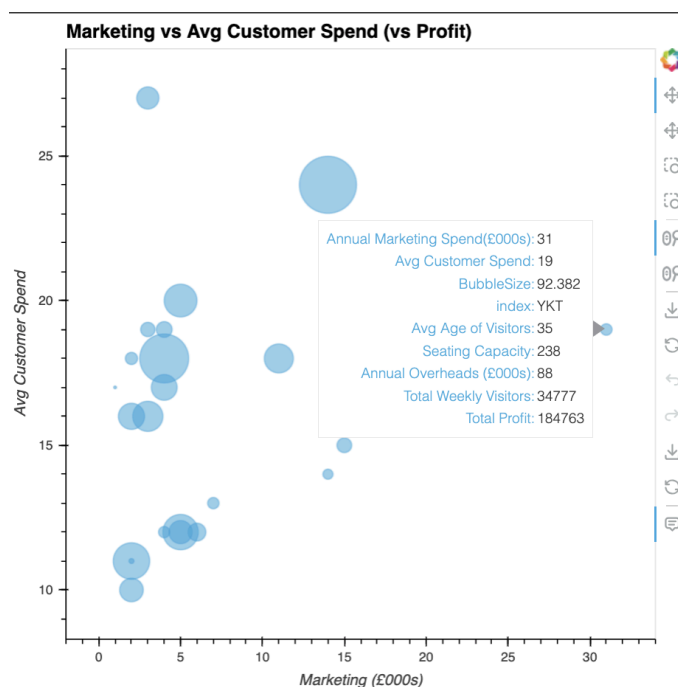


Figure 7: Bubble Plot - Marketing vs Total Weekly Visitors (vs Profit)

Autocorrelation

The autocorrelation in Figure 8 showcases the correlation between a time-series and the data points at different lag points. As the dataset used is of a weekly data, this suggests each lag represents the difference of 7 days. The lines shown in Figure 8 do not fall below the threshold of 0 which indicates a positive correlation reinforcing that as time goes on, the more weekly visitors Cinema VJV has which reinforces the positive growth trajectory, hence its consideration as a high volume cinema.

A secondary peak for VJV shown below at approximately lag 8 (week 8) is an indicator that there is seasonality and that 56 days (lag of 8) can be considered a season. This means that a particular week within this highlighted period is likely to have experienced a peak in weekly visitors that corresponds with similar seasonal cycles across the 4 years. For instance, this could be related to traction gained from the use of Cinema VJV's annual marketing budget, or potentially changes in weather that cause visitors to partake in indoor activities. In addition, industry events of significance could also result in a peak in weekly visitors such as the lead up to film festivals such as Cannes Film Festival which may encourage increased viewership in anticipation of yearly nominated films.

The line in Figure 8 presents spikes that tend to fluctuate between every other lag which indicates a weekly trend of fluctuation in the volume of visitors. As cinema's generally aren't an everyday activity, this suggests that visitors (particularly between age 31 and 36) may be more inclined to visit on the weekends or during 'free time' as a result of time off work which means the dips may be caused by the relative lack of visitors throughout the weekdays.

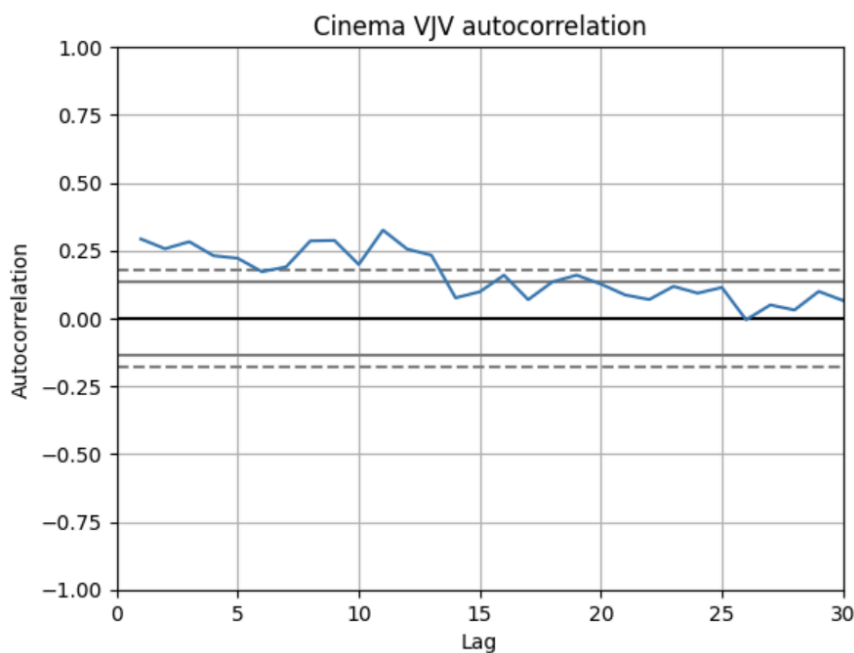


Figure 8: Seasonality of Weekly Visitors (Cinema VJV)

Critical Review

The data visualisation module has facilitated the understanding of techniques and principles used to explore data through graphical representations to analyse trends and patterns within datasets. This module provided insight into the utilisation of Python libraries (Pandas, Matplotlib, NumPy), various forms of visualisations and ways they differ in suitability as well as the theoretical context of data analytics and statistics, all of which enables comprehension and allows for the application of the acquired, to complete this coursework.

The company, ChrisCo, provided a range of data for analysis that enabled insights to be drawn through exploration and highlighted significant attributes of the cinema portfolio. The demonstration of best practices can be seen throughout the coursework in methods such as: the definition of the story in which the company data presents through grouping data files into two dataframes for assessment, the implementation of clear labels on their appropriate axis, the comparison of data points to showcase relationships between attributes, and use of interactive plots that enable a dynamic interaction and deeper understanding of data points.

It could be argued that the data exhibits sampling bias as the time series covers a period of 4 years as opposed to a longer time frame that could allow for a fairer categorisation of cinemas and analysis of overall performance. Also, high volume is associated with the number of weekly visitors but could have several connotations such as conversion to customers. This would allow for a more accurate analysis of cinema performance and debunk the assumptions made surrounding profits. Nevertheless, the data provided enables insight into the characteristics of cinemas and the variables that influence their individual and overall growth.

While the visualisations were well executed throughout the coursework, there are some improvements that can be added to reinforce confidence in methodology. The use of advanced techniques such as dashboards could have allowed for the inclusion of more graphical insights due to the tools capability to deploy more than one visualisation. The inclusion of more references could also support the credibility of inferred insights to demonstrate an understanding of theory.

All in all, the methods learnt during this module allowed for a fundamental understanding of visualisation concepts, the application of relevant tools and techniques, and the use of cognitive skills that will prove effective throughout a career in data and beyond.

Data conclusions

From the company data explored throughout this coursework, we can conclude that:

- The more weekly visitors a cinema has, the higher likelihood of customers which in turn leads to increased revenue. Cinemas VJV and YBS are noticeably the most profitable cinemas within the portfolio.
- Cinemas with a higher seating capacity are expected to spend more on marketing to reach a larger customer base whereas smaller venues can focus on appealing to a niche demographic of big spenders to be more profitable.
- Visitors between the ages 31 and 36 are the ideal demographic that visit the cinema, but visitors above age 40 tend to spend the most. Most cinemas are likely to target this audience in marketing due to their inclination to spend; there is more emphasis in knowing and reaching your target audience than the amount spent on marketing.
- Cinemas can gain a competitive advantage by capitalising on attributes other than marketing such as the cinema experience or providing better incentives, to increase profits.
- Cinema VJV has the strongest seasonality of the portfolio and experiences a peak of weekly visitors within a season of 56 days each year.
- Generally, visitors are more likely to attend the cinema on weekends as opposed to weekdays.
- The more visitors a cinema has, the higher the likelihood of customer conversion.
- The growth of weekly visitors can be influenced by external factors outside business operations such as environmental, economical and social trends.

References

Tableau (2024) Data visualisation beginner's guide: a definition, examples and learning resources. Available at: <https://www.tableau.com/en-gb/learn/articles/data-visualization> (Accessed: 31th March 2024)

Keenan, M. (2023) How to start an online boutique business in 2024. Available at: <https://www.shopify.com/uk/blog/start-online-boutique> (Accessed: 31th March 2024)