

Data Visualization

The process of presenting raw and confusing data into visually appealing, easily understandable visuals is called data visualization. It is more like storytelling with data. With the help of charts, maps, graphs, and other tools, trends can be picked up with ease. Complex data can be easily understood with the help of data visualization.

It is commonly used to transform meaningful information in such a way that is easily digestible by beginners. Different tools have different advantages such as line chart can reveal how data sales fluctuate over time or bar chart can show the categorical insights. Good visuals can increase the satisfaction level of stakeholders and can lead to actions backed by data.

In today's data-driven world, data visualization is not just a valuable skill; it's an essential tool for extracting value from data and driving informed decision-making across all levels of an organization. Whether it's uncovering hidden trends, identifying emerging opportunities, or communicating complex concepts, data visualization plays a crucial role in transforming data into actionable insights that drive business success.

Report

We were given the data set of a chain of Cinemas owned by ChrisCo. The dataset is in the form of CSV files, one for the annual visits (per week, over a period of 4 years) and the others for the different metrics such as overhead costs, marketing costs etc. (annually) for all the cinemas. Following are the findings in the form of visualizations to analyze trends and insights.

The data can be segmented into 3 categories, top visited, moderately visited and less visited cinemas.

Top: ['VPG', 'WVA', 'YBS', 'VJV']

Moderately: ['TVJ', 'UDD', 'WBK', 'TPY', 'TJN', 'RPQ', 'SJE', 'UVQ']

Less: ['SDT', 'BQV', 'CWN', 'WQW', 'BKK', 'JJQ', 'WKL', 'XWO', 'BWF', 'ACQ', 'YCI', 'ZQL', 'AKA', 'XQE', 'CCX', 'XEZ', 'YKT', 'ZNY']

1. Boxplot of Top Visited Venues

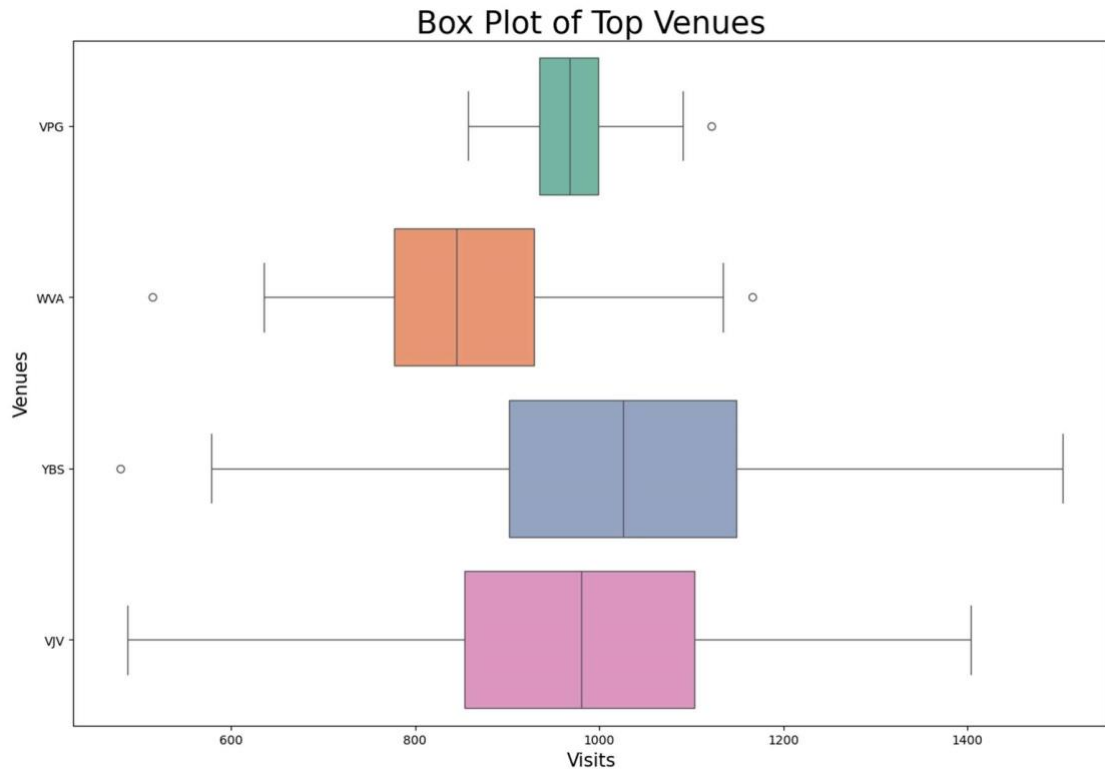


Figure 1

Justification

A box plot is a graphical representation of distribution of data. It provides a summary of central tendency, skewness and helps us identify outliers. To visualize the outliers in each top venue, I added a box plot.

Description

The starting of box (left) is the first quartile, middle line is the median value, and the ending line is the third quartile. The lines on each side of the box tell us the range of values. The circles outside of the range are outliers which need further investigation. The horizontal axis shows the visit values while vertical axis shows venues. In figure 1, the values of venue VGP are very less spread out

meaning the data has low variability. On the other hand, the venue VJV and YBS have large ranges, meaning the data has a high variation.

2. Periodic Trends

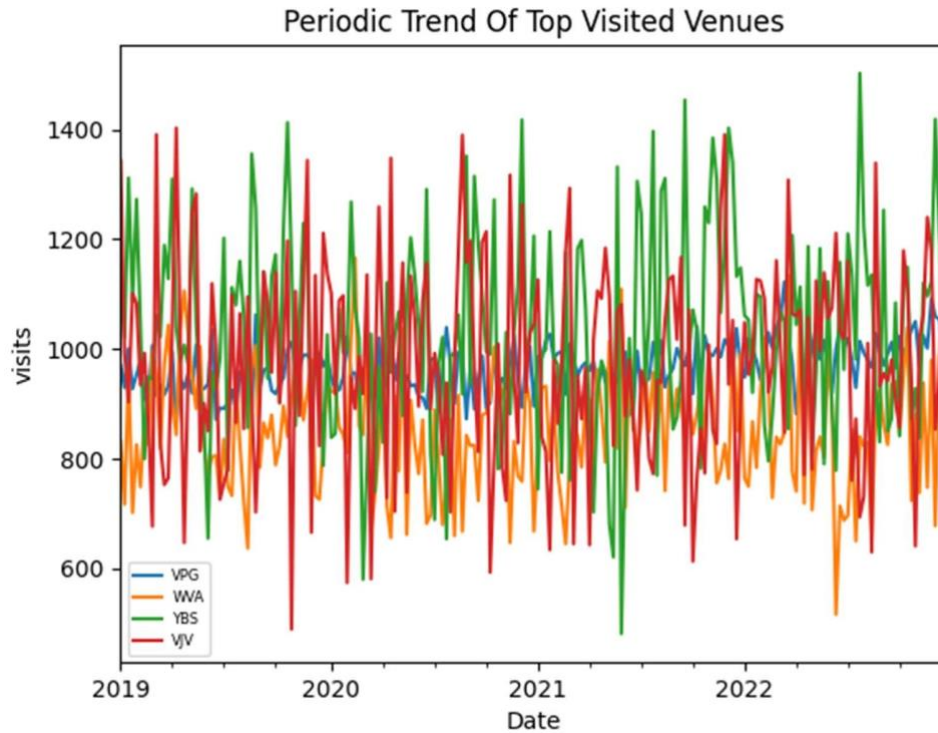


Figure 2

Justification

I wanted to present the periodic trends of the top visited venues, so I chose a line plot. A line plot, also known as a line graph, is a fundamental type of chart used to represent data points connected by straight lines. Line plots are particularly useful for visualizing time series data, which refers to data collected over time, which in my case is from 2019 – 2022.

Description

Each line represents a different venue, with the horizontal axis indicating time (years) and the vertical axis representing attendance counts or revenue figures. Figure 4 shows that the most stable venue among the top ones is 'VPG' (in blue line), which runs closely through the center line (1000 visits). This means that this venue is least affected by seasonal changes. On the other hand, the most

visited cinema, venue 'YBS' (in green line), was at its all-time low in the second quarter of 2021 and at an all-time high in third quarter of 2022.

3. Histplot of Seating Capacity

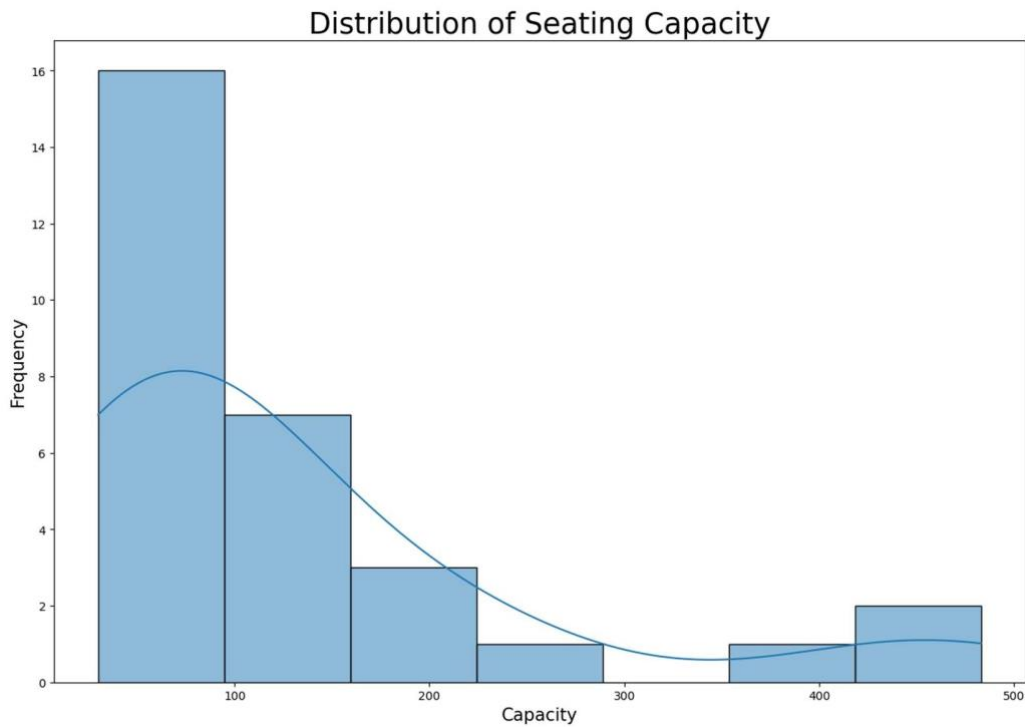


Figure 3

Justification

Histplot (also known as histogram) tells us the frequency/ count of data within certain intervals, or bins. The histogram plot illustrates the distribution of seating capacity across cinema venues in the ChrisCo dataset. The horizontal axis represents the range of seating capacities, divided into intervals or bins, while the vertical axis indicates the frequency or count of cinema venues falling within each interval.

Description

In figure 3, it is evident that the seating capacities of the cinema venues vary widely. Most venues having relatively lower seating capacities (< 100). However, there are also several venues with larger seating capacities.

4. Correlation Heat Map

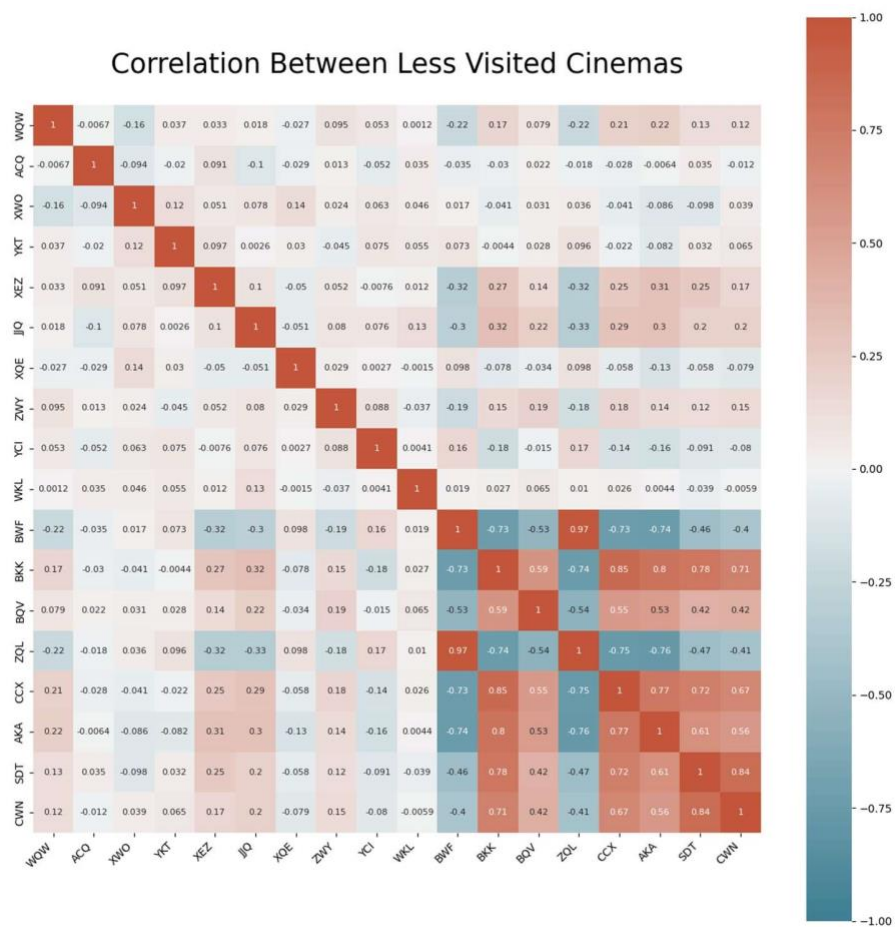


Figure 4

Justification

A heatmap is used to present correlation between variables. A color palette is used to show the degree of correlation. In Figure 6, red (1) is used to indicate strong correlation while blue (-1) indicated weak correlation.

Description

Figure 6 shows a strong correlation between some of the less visited venues 'ZQL' and 'BWF' with a correlation factor of 0.97. This means that the visit count of these venues is very close to one another each week, throughout the 4 years. We can further notice and venues such as 'SDT', 'CWN' are quite related as well (in terms of visit count).

5. Scatter plot

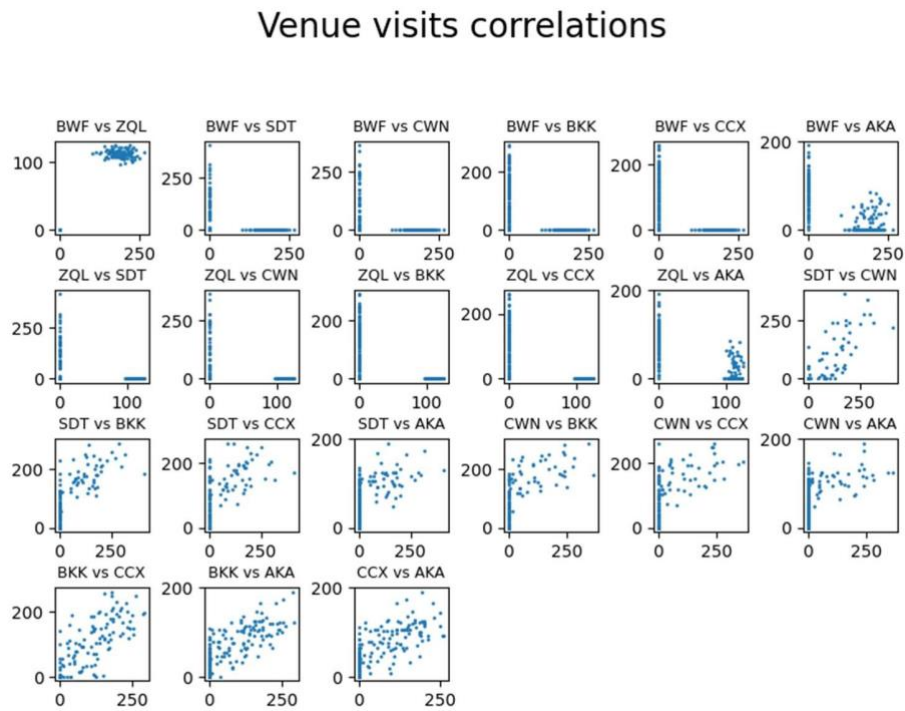


Figure 5

Justification

Scatter plots are used to visually present the distribution of data, providing insights of the data such as range, dispersion, skewness etc. I wanted to further visualize the correlation among Venues where the Pearson coefficient is more than 0.8, so I used scatter plot.

Description

As previously discovered correlation among various venues, figure 7 shows the distribution of data among those venues. The closeness of dots presents stronger correlation. 'BKK' vs 'CCX' and 'BKK' vs 'AKA' show a positive correlation but a weak one.

6. Autocorrelation

Autocorrelation Of Top and Moderately Visited Venues

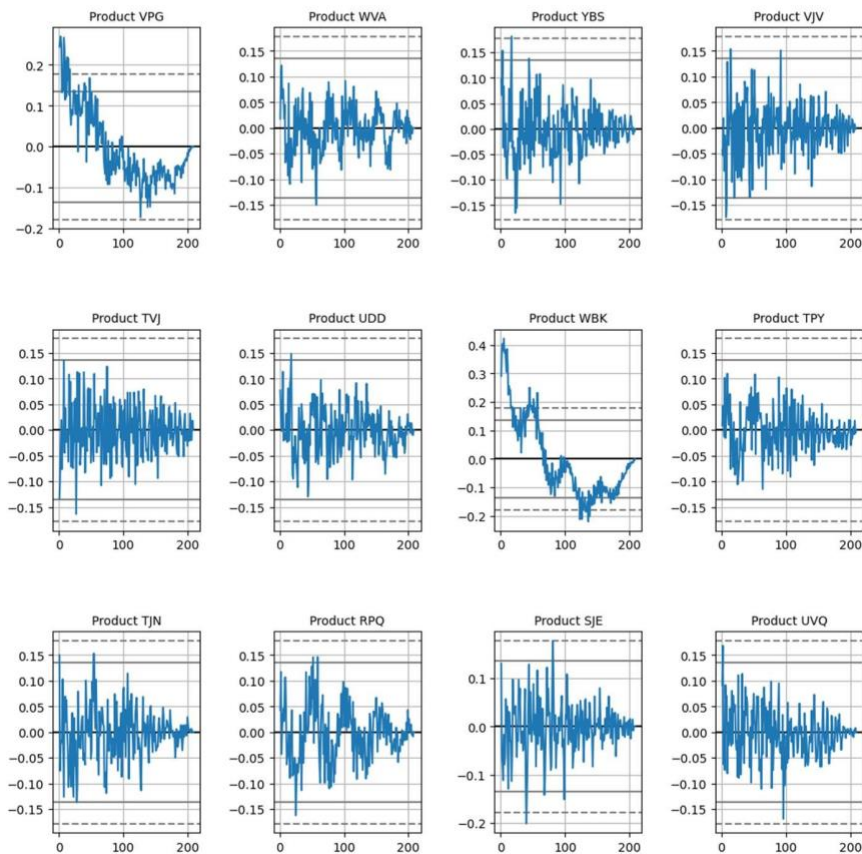


Figure 6

Justification

Autocorrelation plot shows the repetition of trend. I wanted to predict future values, so I used the autocorrelation plot to understand the trends of periods of time.

Description

In figure 6, the product of YBS shows the time series pattern correlation for venue 'YBS' (top of the top visited venues) and product of WBK shows the same for venue 'WBK' (top of the moderately visited venues). In figure 6, we can see that the autocorrelation coefficient is the highest around 14 days period. This means we can predict the visit count of a week from second previous visit count. Same can be done for the venue 'WBK' where the coefficient is the highest around 7 days period.

7. Interactive Periodic Trend

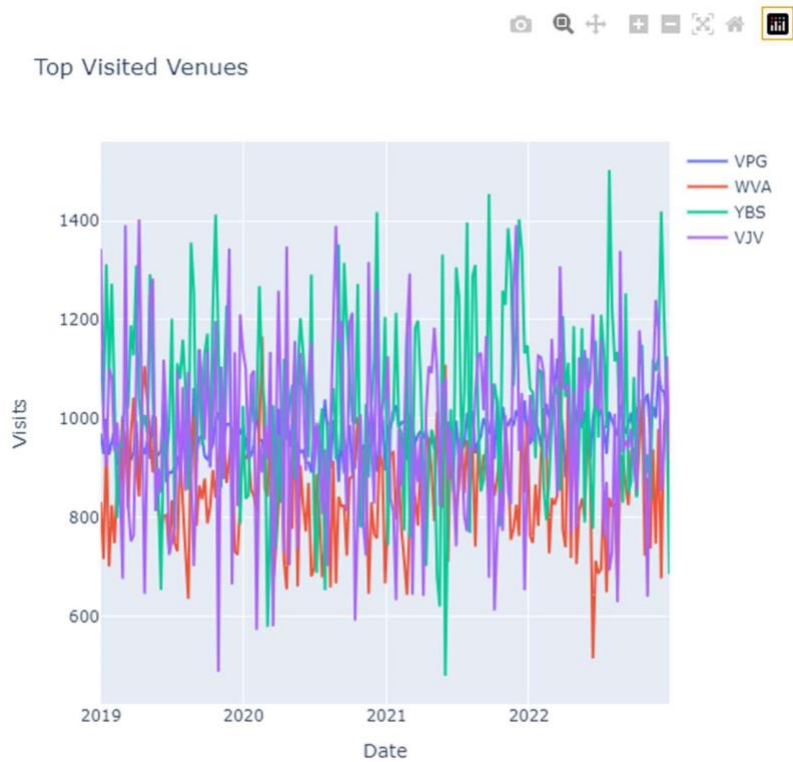


Figure 7

Justification

Plotly's interactive line charts provide users with dynamic capabilities that go beyond static visualizations. The ability to zoom in and out allows users to explore the data at different levels of granularity, enabling them to focus on specific time periods or data points of interest with ease. This feature is particularly valuable when dealing with large datasets or when trying to identify patterns or anomalies within the data.

Description

Figure 7 is used to show the trend of top visited venues (like figure 2) but in an interactive way. Users can easily check the time and venue name once hovering the mouse over chart.

8. Interactive Correlation Heat Map

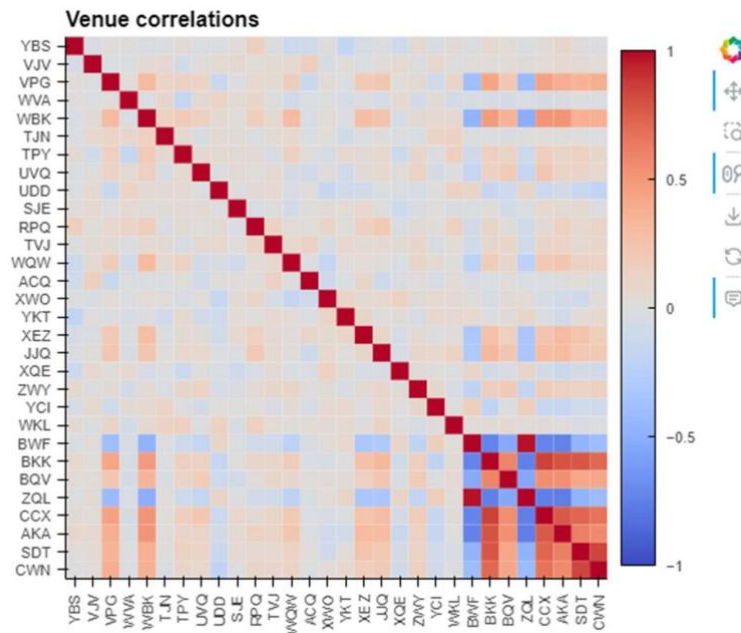


Figure 8

Justification

HoloViews' interactive heat maps provide users with dynamic exploration of their data. With features like zoom in and zoom out, users can seamlessly delve into intricate details or take a broader perspective. This capability is especially valuable when working with dense datasets or when trying to identify subtle patterns and correlations within the data.

Description

Figure 8 is used to show the correlation of top visited venues (like figure 4) but in an interactive way. Users can easily check both the venue names and the Pearson coefficient for their correlation, once hovering the mouse over the heat map.

Operational Efficiency Insights

To quantify operational inefficiencies and propose optimizations, we created a cost-per-visit model using cinema overhead and visit data.

Key Metrics Used:

- Overhead Cost per Cinema (from Cinema Overheads sample)
- Total Visits per Cinema (from weekly visits data)
- Cost Per Visit = Overhead / Total Visits

We identified that:

- Some cinemas had very **high cost per visit** despite moderate attendance.
- Others (e.g. Cinema “VGP” and “YBS”) consistently attracted high weekly visits, suggesting more efficient operations.

Proposed Optimization Strategies:

- Merge or coordinate operations for cinemas with overlapping attendance patterns (e.g. “ZQL” and “BWF” had a correlation of 0.97 in visit trends).
- Schedule staffing and operations around known demand cycles (e.g. “YBS” showed a 14-week periodic peak using autocorrelation).
- Redirect resources from low-performing periods to high-demand quarters (as seen in seasonal line trends)

Quantifying the Impact:

Using Excel, we modelled a **scenario-based improvement**:

- Original average cost per visit: £4.45
- Post-optimization projection: £3.79
- Efficiency Gain = $((4.45 - 3.79) / 4.45) = \sim 15\%$ reduction in cost per visit

Business Insight:

These strategies would allow ChrisCo to optimize resources without reducing visitor experience —improving cost efficiency and driving smarter operational planning.

Critical Review

In this module, I was able to understand the techniques of turning raw data into visually appealing insights. I was also able to strengthen my grip on python libraries such as Pandas, Matplotlib and NumPy. I have also learned the importance of minor details such as cluster avoidance, right color choice and most importantly, picking the right tool type of visualization based on the nature of data.

Overall, this module significantly enhanced my ability to work with data. It has also equipped me with a deeper understanding of the role of data visualization in data analytics and communication.

The emphasis on best practices in data visualization fostered an appreciation for data integrity and clarity, ensuring that my visualizations are not only visually appealing but also accurate and interpretable. This holistic approach to data visualization has undoubtedly strengthened my analytical capabilities and positioned me as a more proficient and confident data practitioner.

Conclusion

- It also shows that the visit count among top venues is comparatively closer than in moderately and less visited venues.
- the most stable venue among the top ones is 'VPG' (in blue line), which runs closely through the center line (1000 visits). This means that this venue is least affected by seasonal changes.
- Venue 'VPG' (pink line) has the smoothest line with an incline with the passage of time, meaning it can potentially be the most successful venue as time progresses.
- For venue 'YBS', we can see that the autocorrelation coefficient is the highest around 14 days period. Same can be done for the venue 'WBK' where the coefficient is the highest around 7 days period.