COMP1800 DATA VISUALIZATION

Richard Raja

001370307

MSc Data Science

Table Of Contents

1. Introduction

ChrisCo Cinemas, a successful and functional company, wants to manage a list of cinemas across the UK. Based on this requirement, the company has collected a vast amount of data through various means. All customer data has been compiled into a single dataset through its loyalty card scheme. This data spans a four-year period, and each cinema is identified with a unique code within the dataset, labeled as the "ID" column. A total of six datasets have been provided for this task. One dataset contains weekly visitor details, while the other five datasets contain information about the individual cinemas, such as age, capacity, marketing spend, and overhead costs. Using these six datasets, the entire visualization has been created using Python object-oriented programming.

2. Justification and Description

2.1 Discussion

A total of six data sets have been provided for this visualization process. After uploading all the given datasets that are in CSV form will be merged using the ID column. This uploading process has been done using the panda's library function of the dataset. Five data sets that contain the information about all individual cinemas have been merged based on the merge function concerning each ID of the cinemas.

2.2 Data Analysis

Data Frame 1: Customer_data

```
[16] # Transpose the cinema_visitors DataFrame to have dates as index
     customer_data = cinema_visitors
     # Display the customer data DataFrame
     print(customer_data.head())
```

```
         Date   YBS  BWF  WVA  SJE  XEZ  ZWY  XWO  UDD  YCI  ...  BQV  BKK  \
0  2019-01-01   990  182  830  340  170  156  173  368  135  ...    0    0
1  2019-01-08  1016  179  793  327  104  130  121  342  143  ...    0    0
2  2019-01-15   862  186  895  330  170  150  103  359  126  ...    0    0
3  2019-01-22  1078  184  738  332  160  180  125  352  134  ...    0    0
4  2019-01-29  1051  190  845  317  176  169  205  347  139  ...    0    0

   WQW  RPQ  XQE  YKT   VJV  TJN  ACQ   VPG
0  246  248  147  152  1004  530  168   841
1  220  325  145  169   717  431  177  1036
2  198  356  155  152  1092  548  172  1017
3  143  333  147  151  1064  505  166  1097
4  208  338  151  149  1161  488  173   963

[5 rows x 31 columns]
```

Data frame 1

(Source: Colab Notebook)

Based on the weekly visitor dataset this data frame has been created.

Data Frame 1: Summary_data

```
[17] # Create a summary DataFrame containing one row for each cinema
     summary_data = cinema_age.merge(cinema_capacity, on='Id')
     summary_data = summary_data.merge(cinema_marketing, on='Id')
     summary_data = summary_data.merge(cinema_overheads, on='Id')
     summary_data = summary_data.merge(cinema_spend, on='Id')

     # Rename columns for clarity
     summary_data = summary_data.rename(columns={
         'Avg age (yrs)': 'Avg_age_years',
         'Seating capacity': 'Seating_capacity',
         'Marketing (£000s)': 'Marketing_£000s',
         'Overheads (£000s)': 'Overheads_£000s',
         'Avg spend (£)': 'Avg_spend_£'
     })

     # Drop any duplicate columns from merging
     summary_data = summary_data.loc[:,~summary_data.columns.duplicated()]

     # Display the summary data DataFrame
     print(summary_data.head())
```

```
    Id  Avg_age_years  Seating_capacity  Marketing_£000s  Overheads_£000s  \
0  YBS             39               531               29               54
1  BWF             44                37                2               80
2  WVA             35               171               13               59
3  SJE             33               202                5               90
4  XEZ             34                63                4               46

   Avg_spend_£
0           24
1           21
2           13
3           13
4           15
```

Data frame 2

(Source: Colab Notebook)

The data margin process has been executed in this above figure where all five separate details of the cinemas have been merged into a single dataset that is summary data.

2.3 Visual Analysis

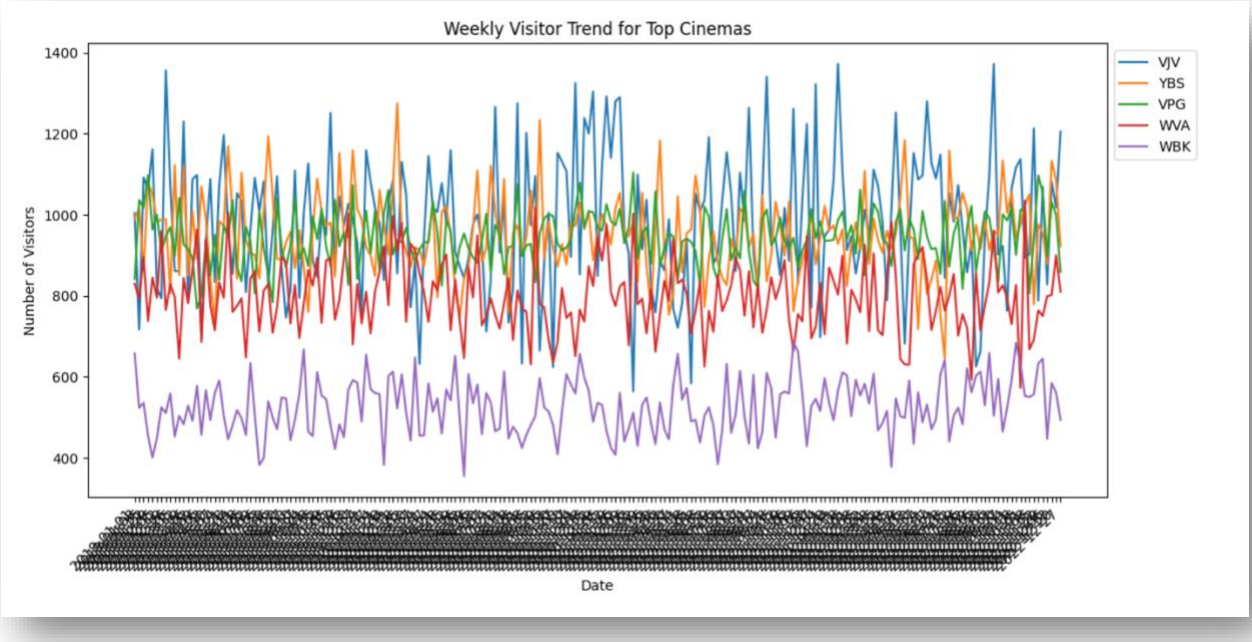

Figure 1: Visualization of Weekly Visitors Trends for Top cinemas

(Source: Colab Notebook)

This line plot reveals weekly visitor trends across all cinemas. We can identify cinemas with consistently higher traffic throughout the year, such as VJV, YBS, VPG, WVA, and WBK. These cinemas help to understand the factors influencing their visitor preference for a particular cinema.

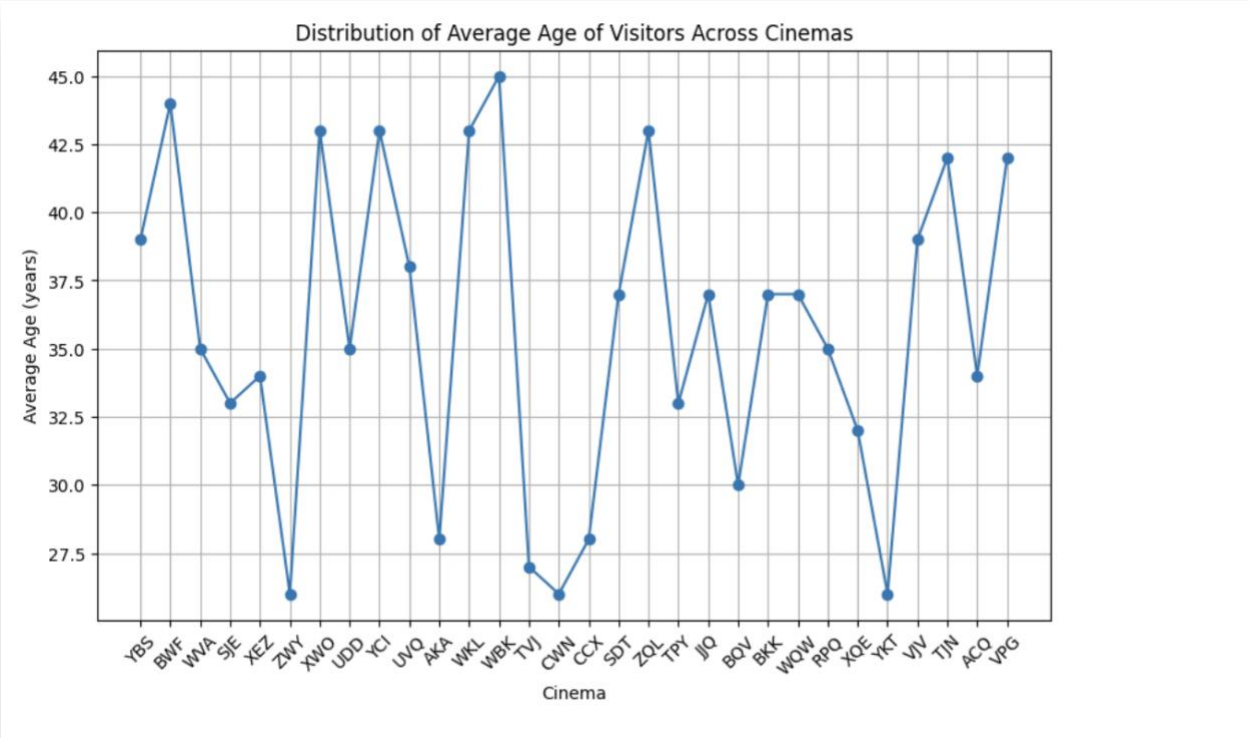

Figure 2: Visualization of Distribution of Age

(Source: Colab Notebook)

The histogram helps us understand the age distribution of cinema viewers. Analyzing this data can inform decisions about movie selection and marketing strategies to target specific age demographics.

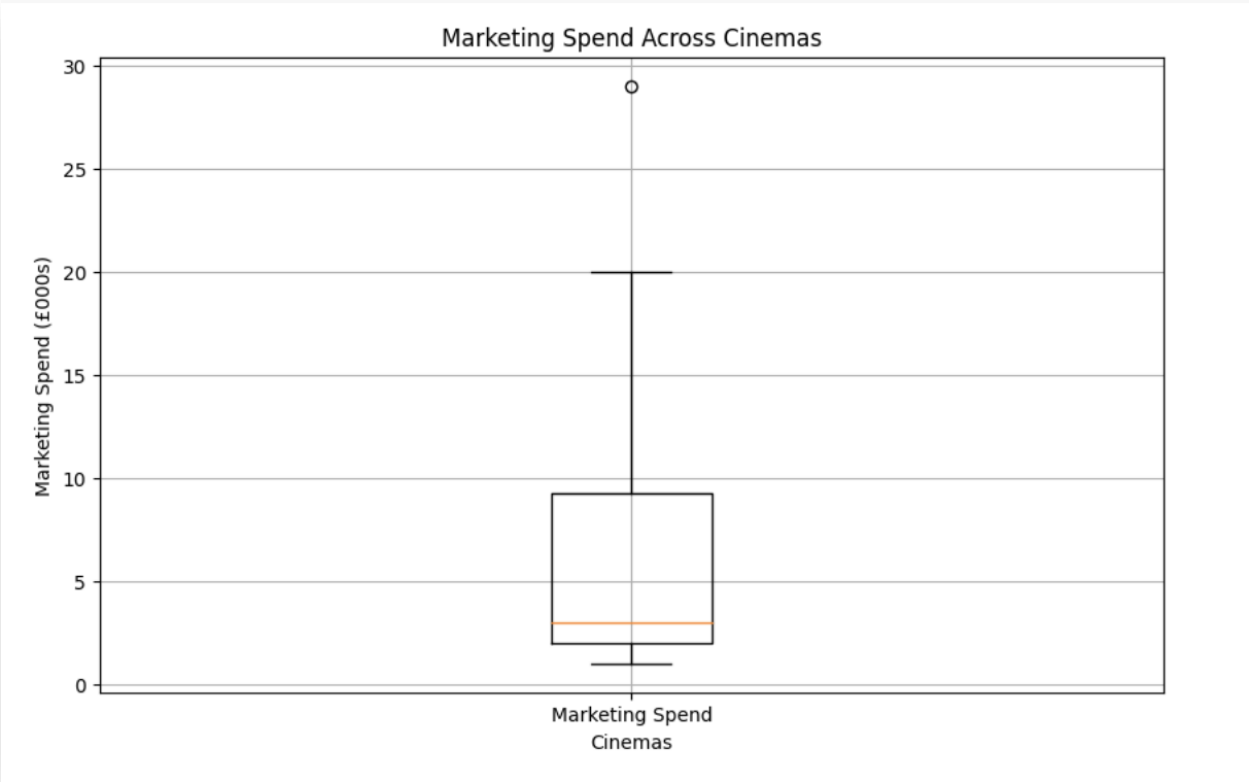

Figure 3: Visualization of Distribution of marketing span

(Source: Colab Notebook)

This histogram shows the distribution of marketing spending across cinemas. It reveals that some cinemas allocate significantly more resources for marketing compared to others. Here the marketing span of this organization gives a comparative view of the distribution and market changing across the cinemas.

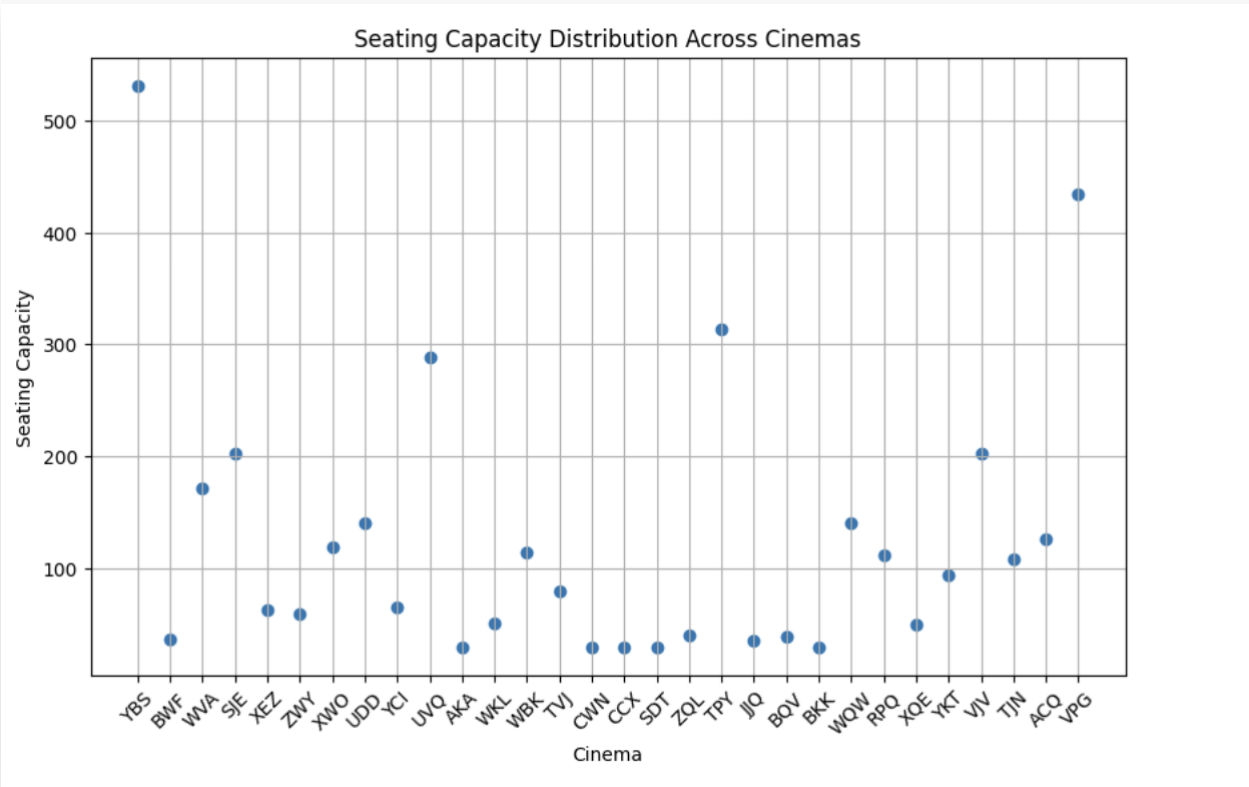

Figure 4: Visualization of Distribution of Seating Capacity

(Source: Colab Notebook)

This graph depicts the seating capacity for each cinema. YBS stands out with the highest capacity, exceeding 500 seats. This information can be considered when allocating resources or scheduling events for different cinemas.

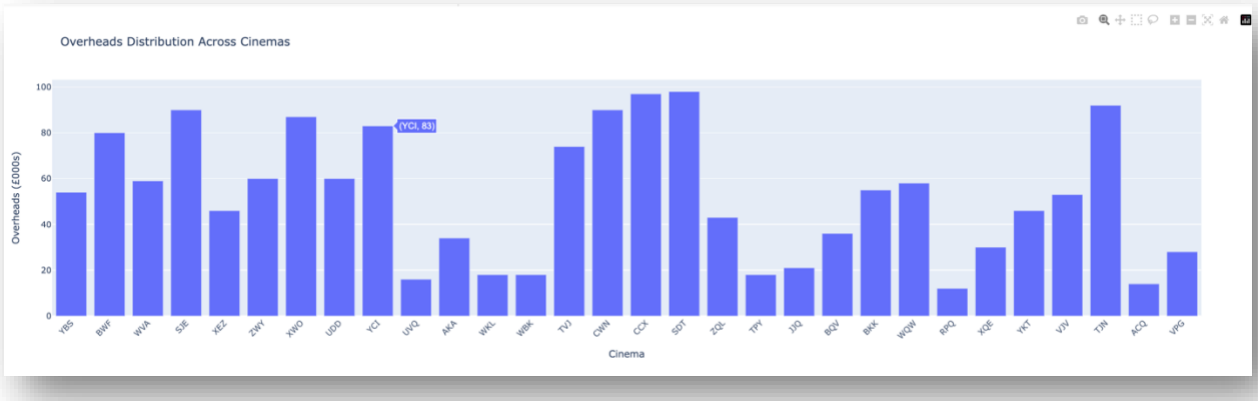

Figure 5: Visualization of Distribution of overhead

(Source: Colab Notebook)

The graph reveals the distribution of overhead costs for each cinema. SDT has the highest overhead cost, exceeding 90 pounds. Analyzing cost structures could help identify areas for potential cost optimization.

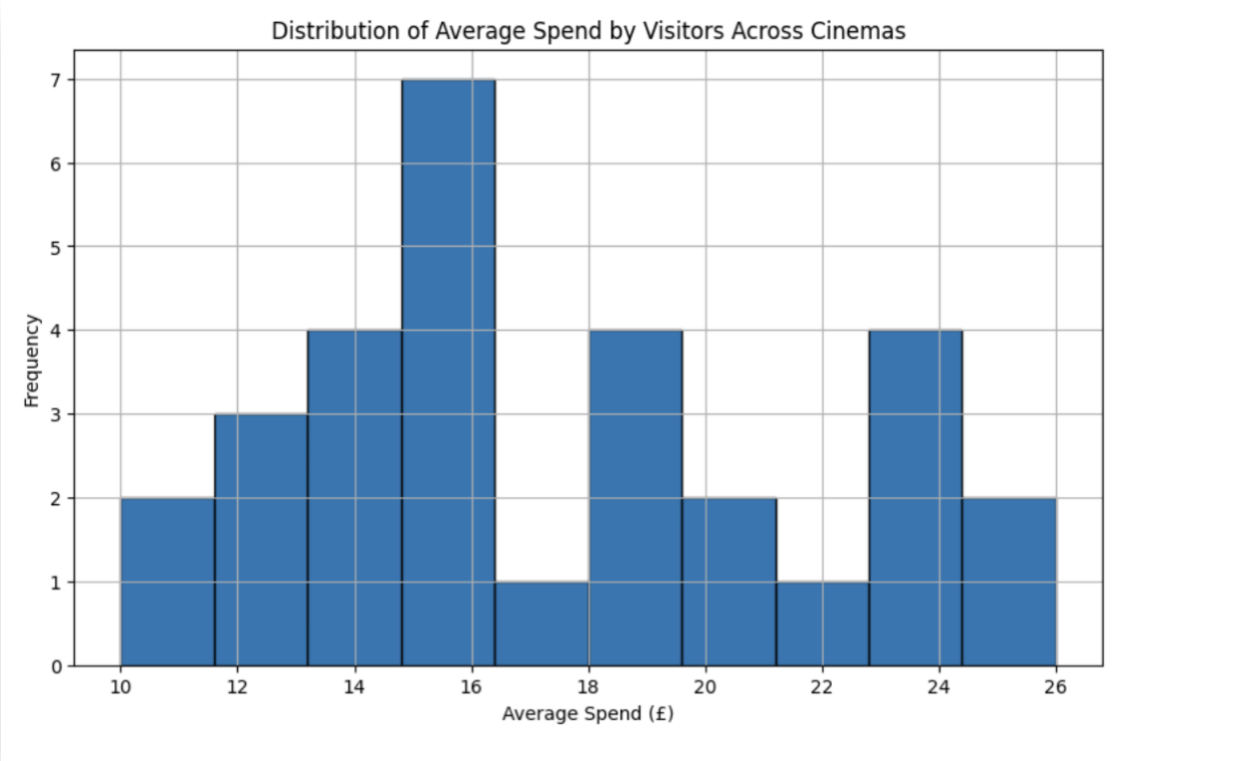

Figure 6: Visualization of Distribution of Average Spend on Cinemas

(Source: Colab Notebook)

This graph shows the average spending per visitor across cinemas. The maximum average spend observed is 16 pounds. Analyzing spending patterns can help tailor promotions or loyalty programs to encourage higher spending from customers.

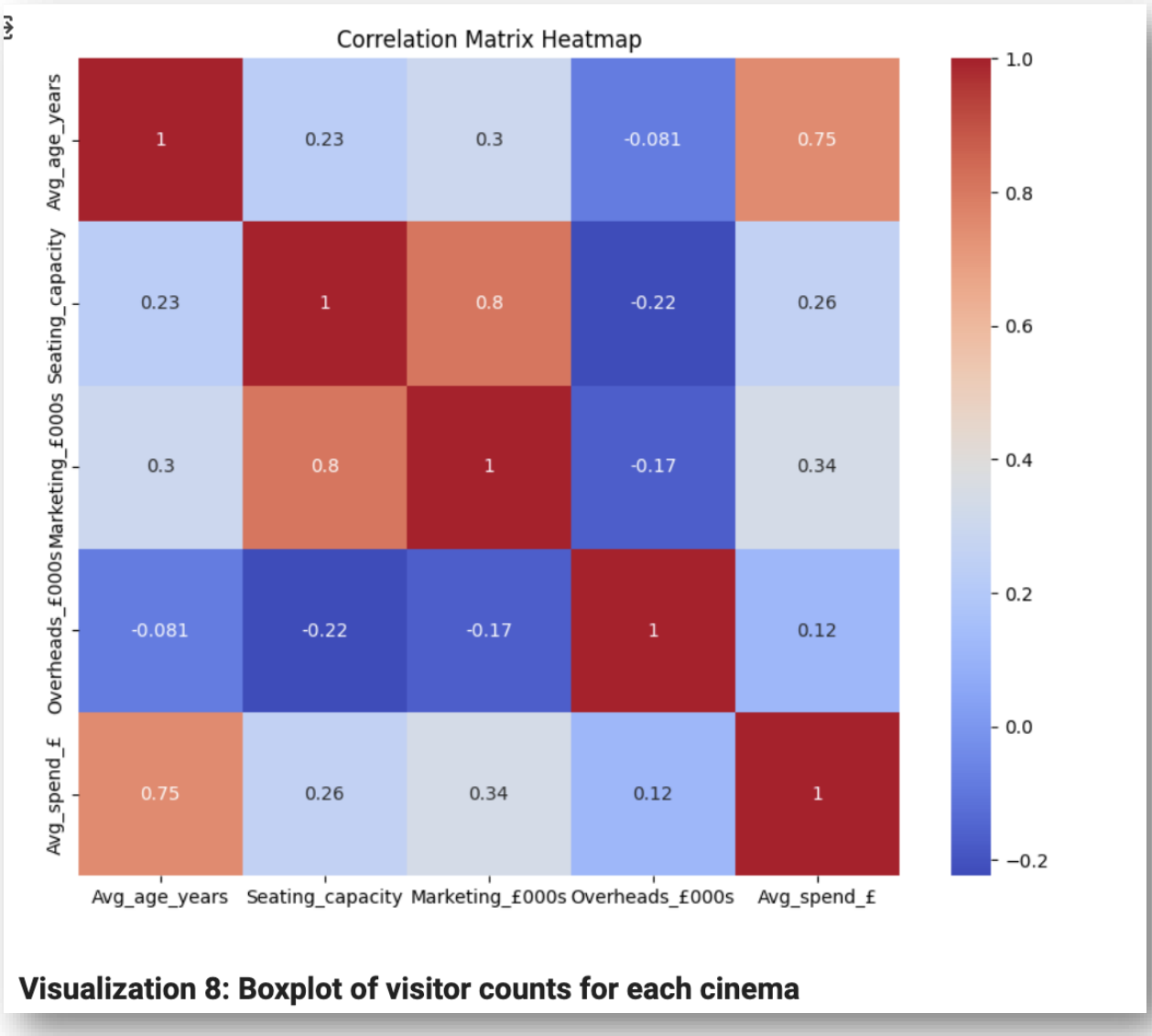

Figure 7: Visualization of Correlation Heat Map

(Source: Colab Notebook)

The correlation map visualizes relationships between different data points. The red squares indicate strong positive correlations, while light blue squares represent weak or negative correlations. This visualization helps identify potential connections between factors like marketing spend and average spending.

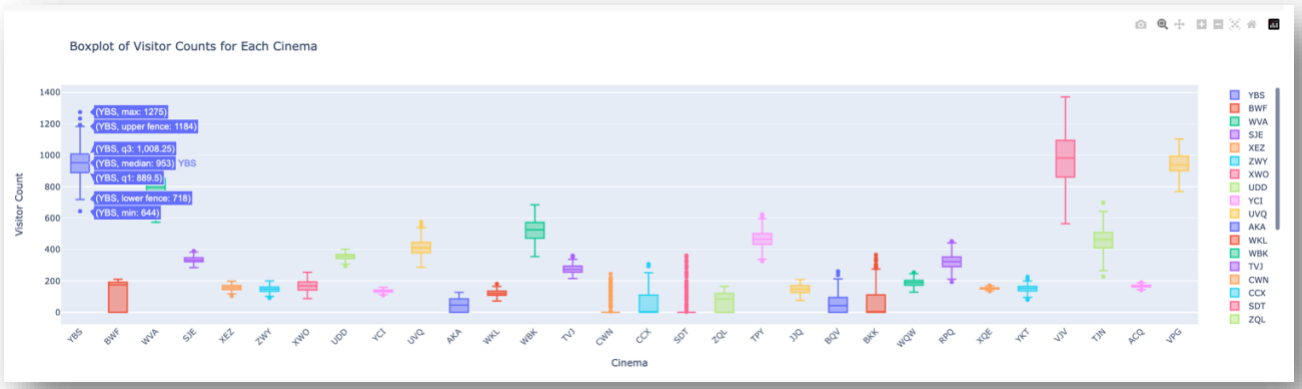

Figure 8: Visualization of Distribution of Boxplot for visitors count.

(Source: Colab Notebook)

This visualization is also justified because this plot displays the visitor count based on the individual cinemas.

The visualization reveals the distribution of visitor counts for each cinema. The spread and interquartile range highlight variations in visitor traffic across cinemas. This can help identify cinemas with consistently high, low, or variable visitor counts.

2.4 Description of the Visualization

The visualization helps to highlight the potential relationship within the data present in the cinema organization dataset. The correlation heatmap can help to understand the higher marketing spend with the increasing average spend by the visitors of the organization (Lau and Guo, 2020).

On the other hand, the trend of visitor line plots based on the cinema category can help to understand the interest of the customer for the particular movie. After that, the other plots like boxes, scattered, histograms, and others give a pictorial representation of the data to understand the relation within them.

3. Critical review

This assignment task can help to understand the hidden relationship within the data that are present in the dataset of ChrisCo Cinemas organization. All the data of this organization has been collected and stored in a different dataset such as marketing, age, overhead, spending and many more. Using various library functions of Python all these datasets are marge into one data set which is summary data. After margining the data, the visualization process has been executed using matplotlib and sklearn function of Python language (Caon *et al.*, 2021). Various visualizations such as distribution of age, marketing, spending, and overhead of all cinemas of the company have been executed in this assignment.

This assignment helped to understand the importance of critical analysis of data and also helped to interpret and extract the meaning of full details from the data using visualization. It is important to understand the preprocessing process before the execution of the visualization process (Yuan, 2023). This is the vital step in this type of analysis and visualization process to clean the dataset in a proper way to get a better result understanding. After analysis, it is time to understand the area of further development of the project. Here the further development of the project needs to incorporate more interactive visualization and also needs to explore advanced statistical analysis processes to understand the characteristics of the data in a better way.

4. Conclusions

In conclusion, the data visualization of ChrisCo Cinemas company data has been executed in this assignment. In this process, all real-time data of this company has been stored in different datasets such as age, spending, overheads, and many more. The visualization provides valuable insights into customer behavior, operational performance, and strategic opportunities. Based on the data set various comprehensive data visualizations such as weekly visitor counts, demographic information, and financial metrics have been executed here. The assignment highlights the average age of visitors, seating capacity, marketing spending, overheads, and average spending by visitors, enabling informed decision-making for the company. Th the part of critical review further improvement sector of this project is also mentioned to get an idea of further improvement of the project.

References

Cao, S., Zeng, Y., Yang, S. and Cao, S., 2021. Research on Python data visualization technology. In Journal of physics: Conference series (Vol. 1757, No. 1, p. 012122). IOP Publishing.

Lau, S. and Guo, P.J., 2020, November. Data Theater: A live programming environment for prototyping data-driven explorable explanations. In Workshop on Live Programming (LIVE).

Yuan, S., 2023. Design and Visualization of Python Web Scraping Based on Third-Party Libraries and Selenium Tools. Academic Journal of Computing & Information Science, 6(9), pp.25-31.