

# **COMP1800 Coursework**

Author Name: Wahidul Alam Riyad

Submission Date: 08<sup>th</sup> March 2022

Authorisation: Module Leader

Distribution: All Students

**School of Computing and Mathematical Sciences**



## Table of Contents

<b>1. INTRODUCTION TO DATA VISUALISATION</b>	<b>- 4 -</b>
<b>2. DATA EXPLORATION FOR CHRISCO</b>	<b>- 4 -</b>
2.1 LINE PLOT VISUALISATION	5 -
2.2 BAR CHART VISUALISATION	6 -
2.3 SCATTER PLOT VISUALISATION	7 -
2.4 HEATMAP VISUALISATION	8 -
2.5 PEARSON COEFFICIENT SCATTER PLOT VISUALISATION	9 -
2.6 SEASONALITY GRAPH VISUALISATION	10 -
2.7 LINE PLOT INTERACTIVE VISUALISATION	11 -
2.8 HEATMAP INTERACTIVE VISUALISATION	12 -
<b>3.0 CRITICAL REVIEW</b>	<b>13 -</b>
<b>4.0 CONCLUSION</b>	<b>13 -</b>
<b>5.0 REFERENCES</b>	<b>14 -</b>

## Table of Figures

FIGURE 1 SHOWS THE LINE PLOT OF HIGH-VOLUME VENUES VISITS WITH 7-DAY ROLLING AVERAGE -----	5 -
FIGURE 2 SHOWS THE BAR CHART OF HIGH, MEDIUM AND LOW VOLUME VENUE VISITS-----	6 -
FIGURE 3 SHOWS THE SCATTER PLOT OF VENUE VISITS CORRELATION BETWEEN HIGH AND MEDIUM VENUE -----	7 -
FIGURE 4 SHOWS THE HEATMAP OF CORRELATION BETWEEN HIGH AND MEDIUM VOLUME VENUES -----	8 -
FIGURE 5 SHOWS THE PEARSON COEFFICIENT SCATTER PLOT OF POSITIVE CORRELATION VENUES -----	9 -
FIGURE 6 SHOWS THE SEASONALITY GRAPH OF HIGH AND MEDIUM VOLUME VENUE VISITS -----	10 -
FIGURE 7 SHOWS THE INTERACTIVE LINE PLOT OF HIGH-VOLUME VENUE-----	11 -
FIGURE 8 SHOWS THE INTERACTIVE HEATMAP OF ALL VENUE VISITS -----	12 -

## **1. Introduction to Data Visualisation**

According to (Venkatraman and Venkatraman, 2019) a graphical depiction of data is known as data visualisation. It visualises information as a picture or graphic to make it simpler to see patterns and comprehend complex ideas. Users may engage with the data by altering the settings to view more information and gain new insights using technology. We can reason and think more effectively about our data using good visualisations. By visualising the data, we can see the pattern right away. We might be able to detect a pattern if we look at numerical data in a table but recognising and conceptualising that trend will take a lot of effort on our behalf.

Businesses can observe enormous volumes of data in simple, unified ways and derive inferences from that information by employing graphical representations of business information (Eltweri, Faccia and Sawan, 2021). Additionally, because analysing data in a graphic format, firms can fix issues or respond to concerns more quickly. Using data visualisation to detect patterns in the company and market may provide companies with an advantage over the competition and, in turn, influence the bottom line. Outliers that impact product quality or customer attrition are easy to recognise and manage before they become significant concerns.

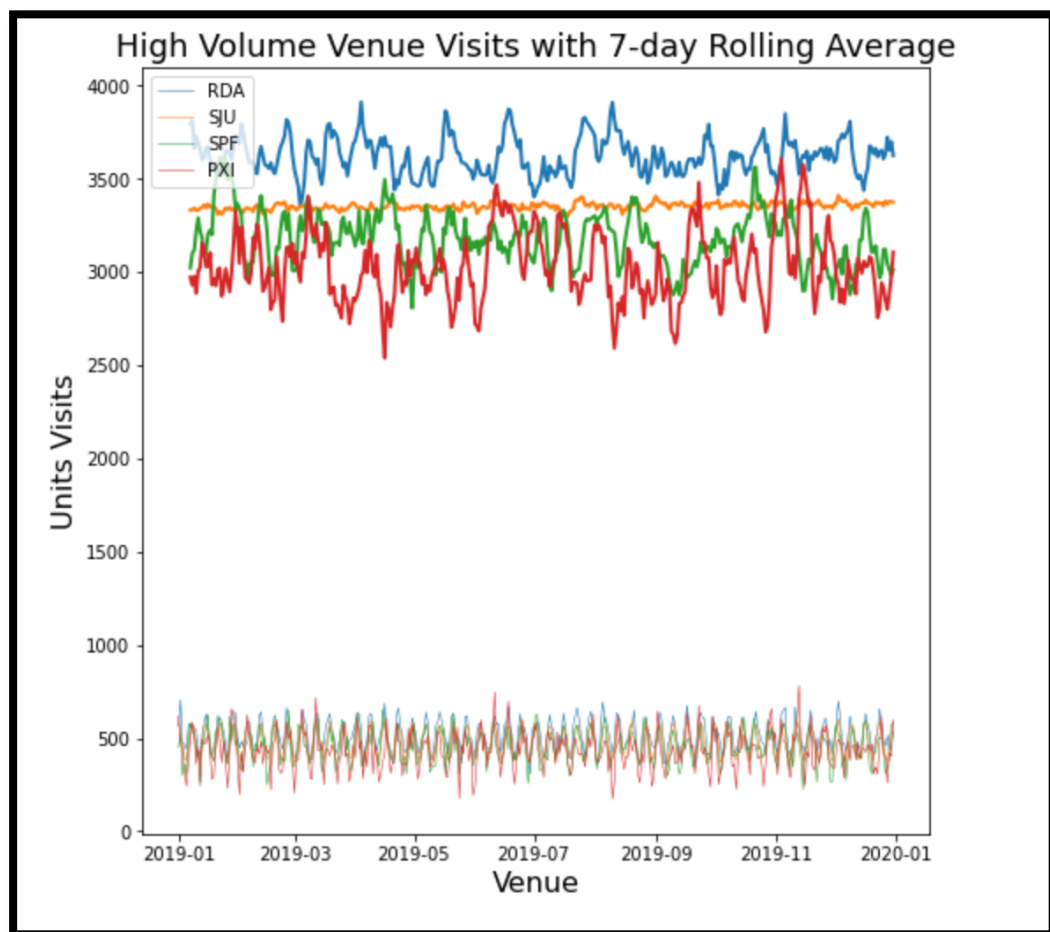
## **2. Data Exploration for ChrisCo**

According to the COMP1800 coursework, ChrisCo is a fictitious yet highly successful firm that manages several locations around the United Kingdom. ChrisCo gathers a great deal of information about individual customers who visit its venues via its loyalty card programme. However, this information has been aggregated/averaged to provide information about the company's 40 locations, each designated by a unique three-letter number.

## 2.1 Line Plot Visualisation

**Figure 1** represents the line plot of high-volume venues. It consists of four independent stores with a rolling average of client visits of seven days. Because we are working with vast amounts of data, we use a rolling average to minimise the build-up of different line plots of data and make the line plot more obvious. The y-axis indicates the number of visits, and the x-axis represents the time series in this line graph, which shows the frequency of client visits throughout that year.

From **Figure 1**, we can see that the Venue RDA, SJU, SPF and PXI are the high-volume venue visits with a 7-day rolling average. Venue RDA, SPF and PXI show fluctuations in customers over the year. However, Venue SJU remains constant with minimum customer fluctuations over the year. Overall, Venue RDA has the highest customer visits compared to other Venues.



*Figure 1 shows the line plot of High-Volume Venues Visits with 7-day Rolling Average*

## 2.2 Bar Chart Visualisation

A bar chart better represents the overall value of a variable. When working with significant amounts of data, this is incredibly handy. It quickly exposes the weight of several variables, allowing us to choose which variable is the most important by simply examining the bars. High-volume venue visits, medium-volume, and low volume venue visits are shown in **Figure 2**.

From **Figure 2**, we can see that RDA, SJU, SPF, and PXI are the four high-volume venues. These locations are within walking distance of one another and have had over 150,000 visitors. There are eight visits to venues with a medium volume. Close bars are present in the letters 'PDT', 'QRY', and 'QJL'. In contrast, close bars are present in the letters 'CWN', 'BEY', 'CQC', 'DKS', and 'AWF.' The lowest volume venue visit has three bar segments, with the final seven 'ZJB', 'BQV', 'BKI', 'ZPL', 'YDI', 'YVW', and 'AEQ' venues being new venues that debuted this year.

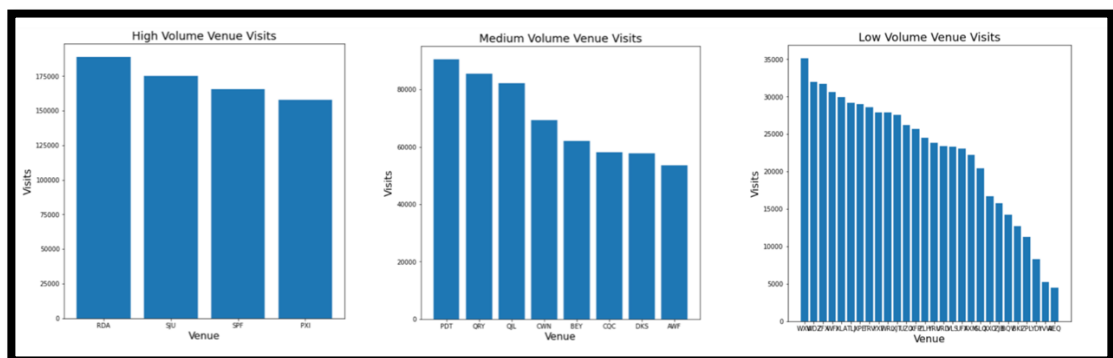


Figure 2 shows the bar chart of High, Medium and Low Volume Venue Visits

### 2.3 Scatter Plot Visualisation

Finding the correlation between two variables is the goal of a scatter plot. It allows us to see how closely two thoughts are connected or what influence one has on the other. But, more importantly, if we can forecast the link between these two factors, we will be in an excellent position to look forward and predict future behaviour. **Figure 3** shows a scatter plot of the correlation of different venue visits.

From **Figure 3**, we can see the correlation between high and medium venue visits. The closer the dots on a line, the stronger the correlation, whether positive or negative. However, it does not give a perfect view of the correlation. There is a strong positive correlation between 'RDA', 'SJU', 'PDT', 'QRY', 'CWN', 'BEY', 'CQC', and 'DKS'.

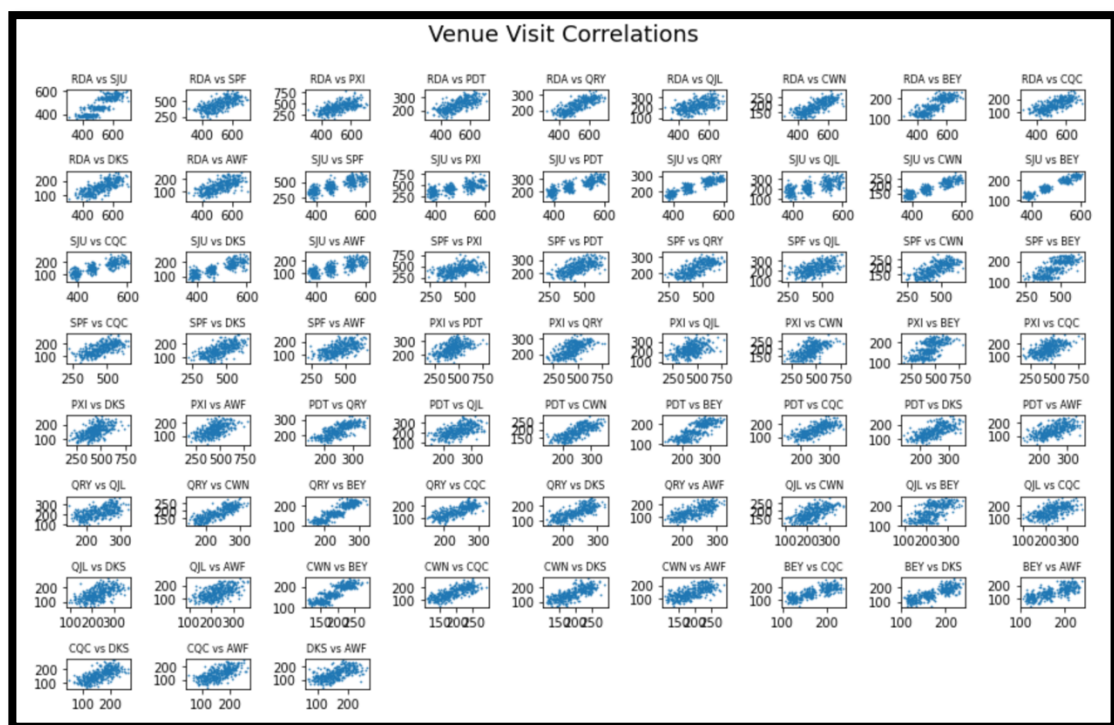


Figure 3 shows the scatter plot of Venue Visits Correlation between High and Medium Venue

## 2.4 Heatmap Visualisation

Heatmaps enable data analysis considerably faster and easier than studying data recorded in excel with infinite rows and columns by providing a visual depiction of numerical data in a ready-to-consume format. Overarching comparisons that would otherwise need data analysis and sorting days perform with a single glance at the resulting heatmap. **Figure 4** shows the heatmap of correlation between high volume and medium volume venues.

**Figure 4** shows that heatmaps provide a more accurate picture of variable relationships, unlike scattering plot graphs. Using a heatmap, we can observe more correlations and the strength of correlation between variables, which we could not notice using a scatter plot. The venues DKS and AWF, CQC and DKS, BEY and CQC, QWN and BEY, and others are substantially positive connected in the chart above. As a result, we will only create a new scatter plot to indicate positive correlations with a Pearson coefficient more significant than 0.8.

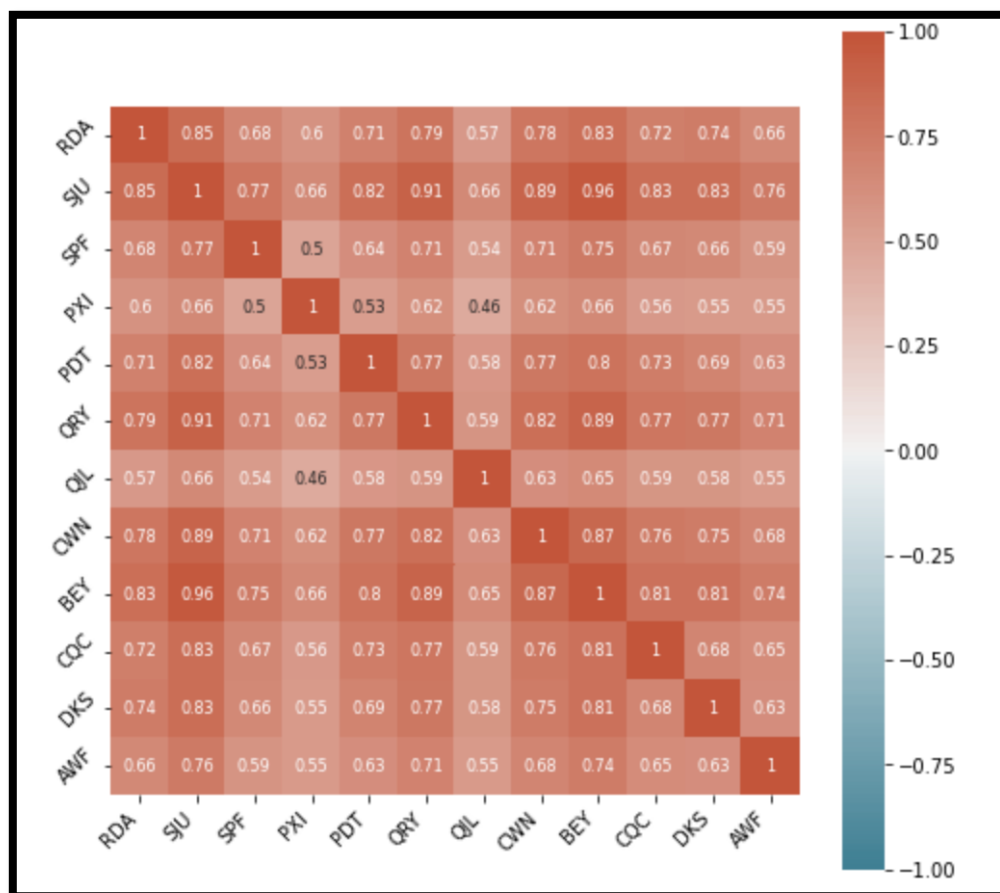


Figure 4 shows the heatmap of Correlation between High and Medium Volume Venues



## 2.5 Pearson Coefficient Scatter Plot Visualisation

Finding the correlation between two variables is the goal of a scatter plot. It allows us to see how closely two thoughts are connected or what influence one has on the other. But, more importantly, if we can forecast the link between these two factors, we will be in an excellent position to look forward and predict future behaviour. **Figure 5** shows a scatter plot of the correlation of different venue visits.

**Figure 5** illustrates a positive connection with a Pearson Coefficient greater than 0.8. The scatter plot includes the following venues: RDA, SJU, PDT, QRY, CWN, BEY, CQC, and DKS. RDA and QRY, as well as PDT and CWN, have a good relationship. On the other hand, other venues do not seem to have a significant association.

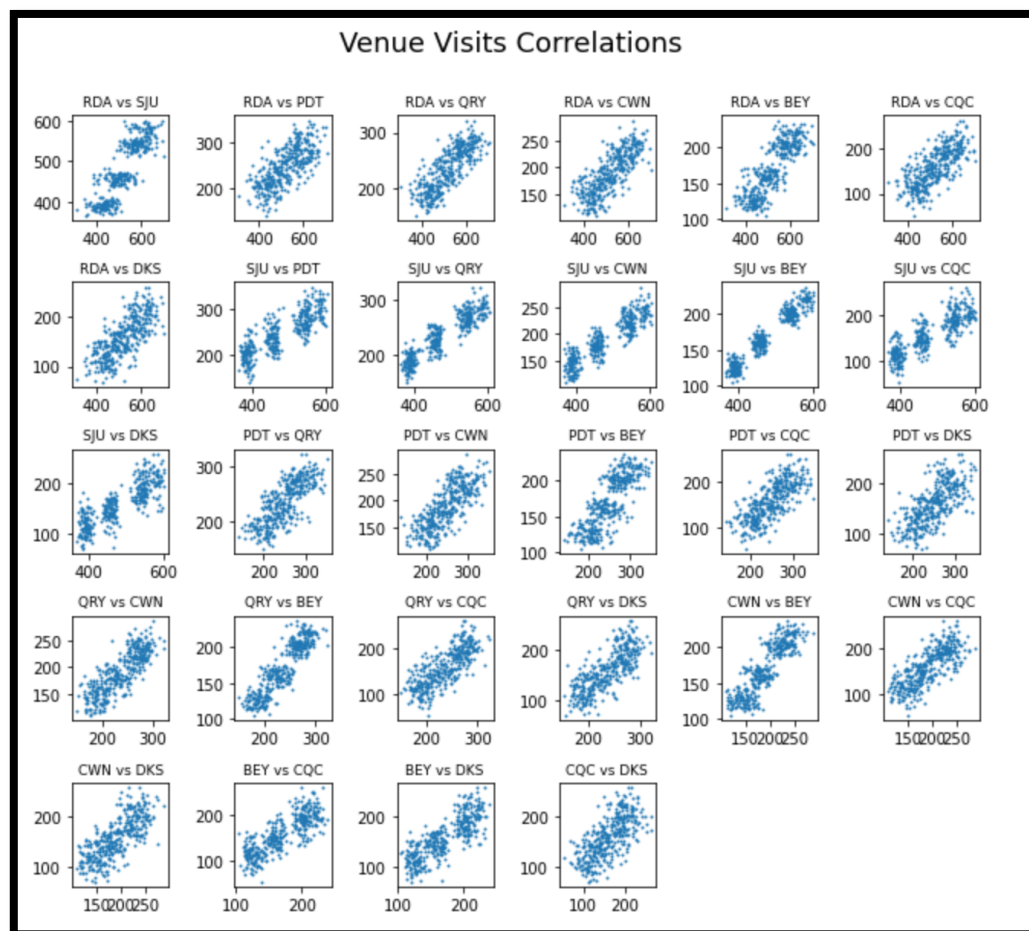
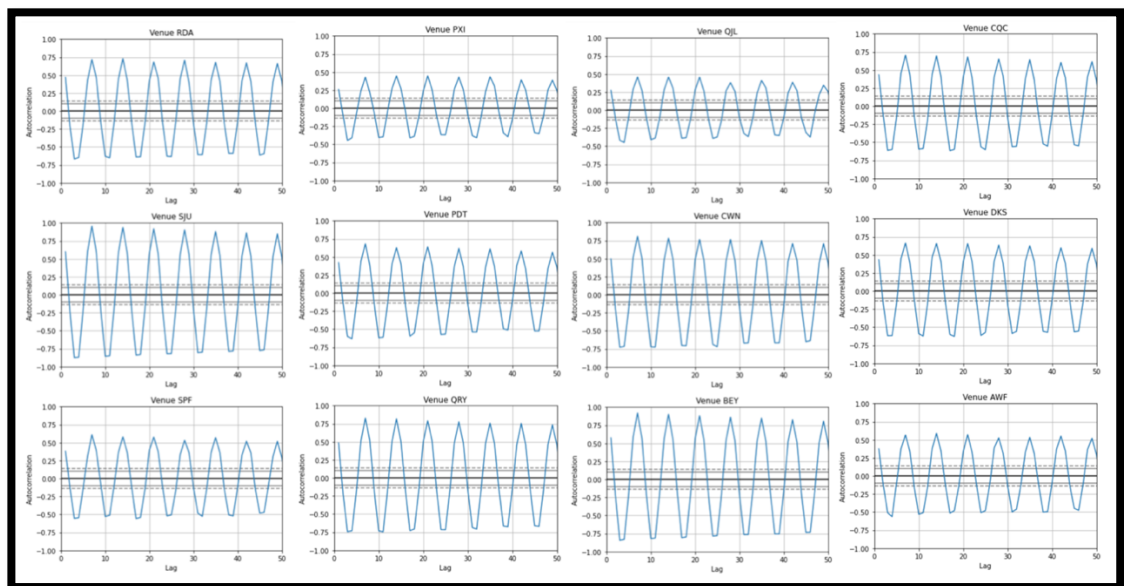


Figure 5 shows the Pearson Coefficient scatter plot of Positive Correlation Venues

## 2.6 Seasonality Graph Visualisation

Seasonality refers to the recurrence of data throughout time. Another critical feature of time series analysis is its ability to detect patterns in data. After eliminating the trend from the data, it is usually quantified by autocorrelation. The graph of weekly seasonality of high and medium volume venue visits is shown in **Figure 6**.

In **Figure 6**, the spikes that rise over the dashed lines are the subject of this illustration. Autocorrelation is evidenced when such tips are statistically significant. Notably, eight of the twelve locations have a weekly distribution, with a peak every seven days. Weekends and holidays, for example, might contribute to such seasonality. Furthermore, this implies that the venue's visitor figures are linked. To explain further, when the number of visitors increases, so do the lines on the graphs, and vice versa.



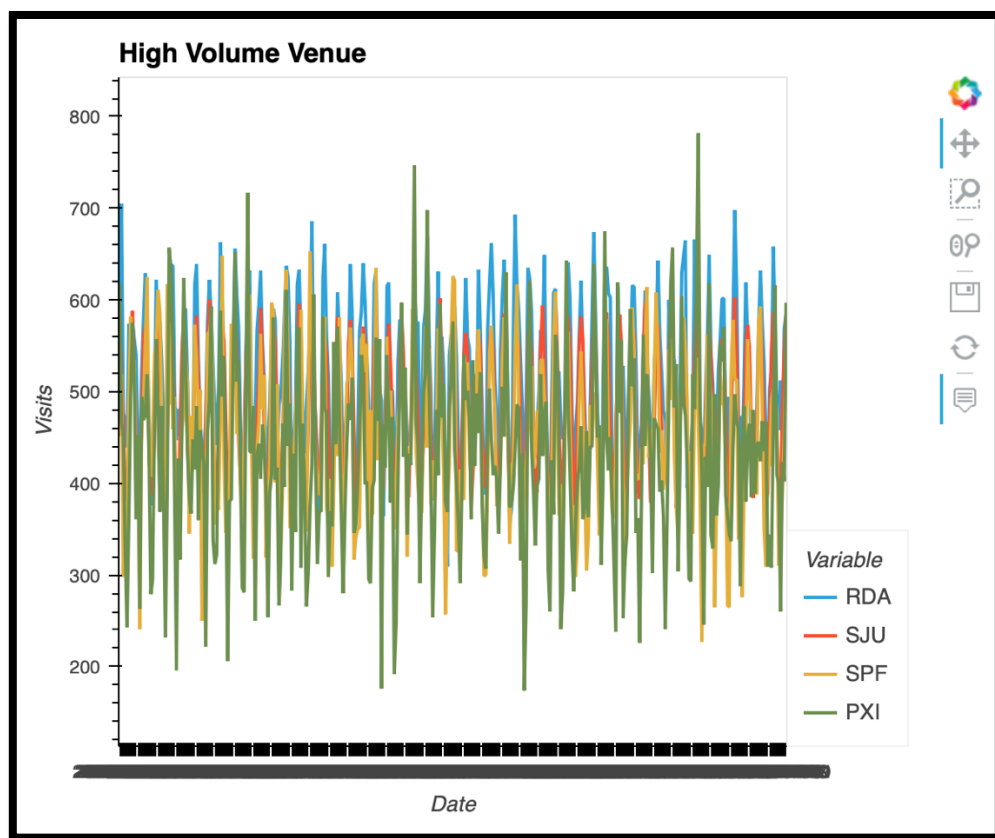
*Figure 6 shows the seasonality graph of High and Medium Volume Venue Visits*

## 2.7 Line Plot Interactive Visualisation

**Figure 7** represents the interactive line plot visualisation of a high-volume venue.

When users choose points or series in a chart, this enables us to create simple-looking visualisations that are rather detailed and a tool for users to explore our data to their heart's content. We may zoom in on a particularly intriguing variation and rapidly add more variables to describe the pattern we've discovered using interactive data visualisation. Furthermore, interactive visualisation will aid us in identifying findings that are not consistent with the theory.

Panning, box zoom, wheel zoom, save, reset, and hover are among the interactive capabilities shown in **Figure 7**. As seen in the graph, RDA has the most outstanding total score among the four locations. PXI, on the other hand, gets the most significant number of visits on certain days. The hover is used to determine which day it is. This indicates that PXI increased to 782 on November 13th, 2019. Even though PXI is the lowest, there was a day when it fell below its average. Use the hover to see which day it is. PXI plummeted to 174 on August 10th, 2019.



*Figure 7 shows the Interactive line plot of High-Volume Venue*

## 2.8 Heatmap Interactive Visualisation

**Figure 8** represents the interactive heatmap visualisation of all venue visits. When users choose points or series in a chart, this enables us to create simple-looking visualisations that are rather detailed and a tool for users to explore our data to their heart's content. We may zoom in on a particularly intriguing variation and rapidly add more variables to describe the pattern we've discovered using interactive data visualisation. Furthermore, interactive visualisation will aid us in identifying findings that are not consistent with the theory.

Panning, box zoom, wheel zoom, save, reset, and hover are among the interactive capabilities shown in **Figure 8**. Use the box zoom and approach to view the most robust correlations and values. This indicates a strong correlation between QRY vs SJU with a value of 0.9. There is also a strong correlation between BEY vs PDT, with a value of 0.79. Use the reset to return to its normal state.

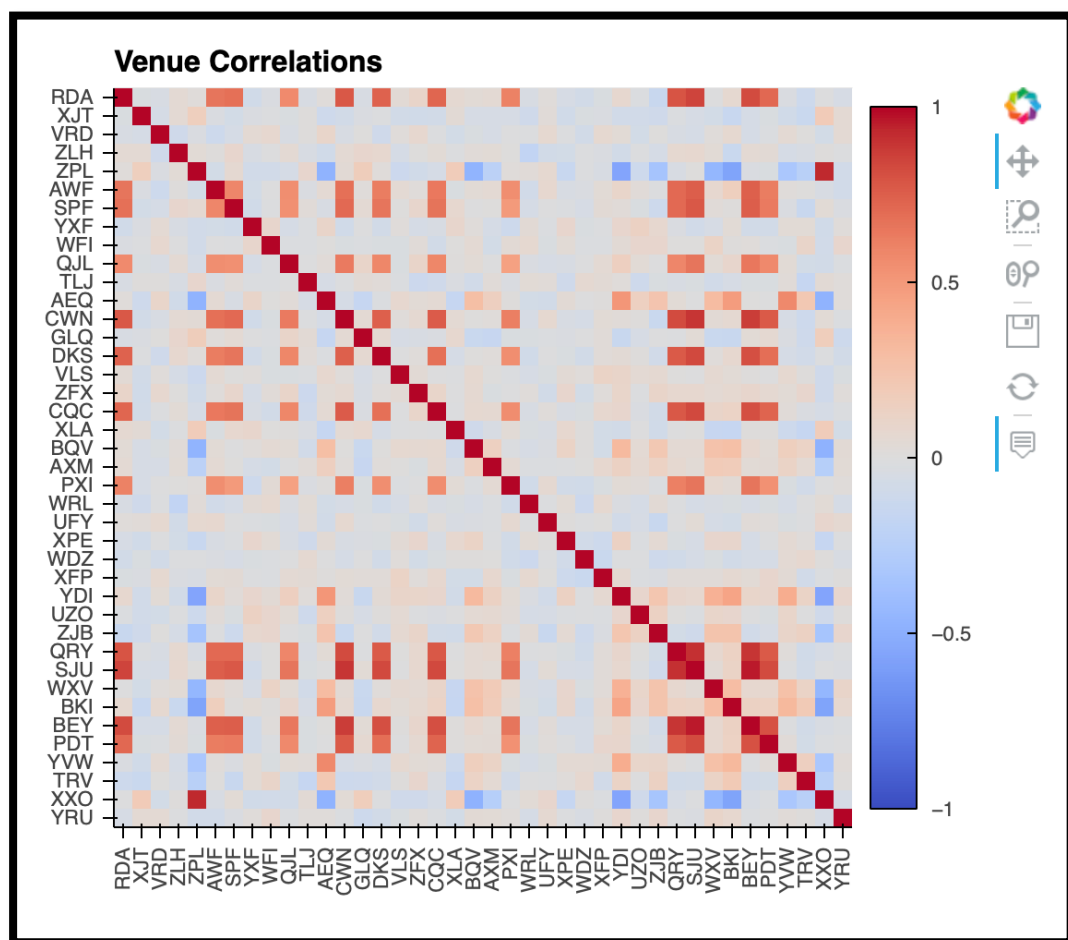


Figure 8 shows the Interactive heatmap of All Venue Visits

### 3.0 Critical Review

Starting with defining the goal, the data processing organisation wants to know what patterns and observations the data might provide, focusing on big and medium-sized venues. We can immediately begin applying the visualisations to extract essential information because the data displayed is already clean and fits the standards for data analysis. The code is repeatable and efficient because it prevents the repetition of undesired code by importing Python libraries for data science, such as Pandas and Hollo Views & Seaborn. The code incorporates loops to append multiple comparable graphs to maximise performance.

The necessary libraries and data sets are extracted at the start of the code blocks. Because the audience is not a visualisation specialist, it is critical to draw their attention to specific visual signals and patterns; the chosen graphs and charts are most effective and competent in conveying the vital information in the dataset of the shop. Furthermore, the representation supports labels and captions to explain how to understand the provided datasets. The notebook complies with all requirements since it allows for markdowns on using each block of code, which includes data output in texts and figures. Providing solid chart axis point lines, reducing clutter and disturbances, and using a variety of colours to help readers grasp data can help create such consistency and clarity.

### 4.0 Conclusion

In the end, the business concentrates on high and medium volume venues. Thus, this study presents significant findings and trends in those venues. The last seven 'ZJB', 'BQV', 'BKI', 'ZPL', 'YDI', 'YVW', and 'AEQ' venues are new venues that opened this year, with the lowest volume of venue visits having three-bar segments. There were no negative relationships between the characteristics provided. RDA has the highest median out of all four when it comes to high-volume venues. Furthermore, eight of the twelve exhibits weekly seasonality.

## 5.0 References

Eltweri, A., Faccia, A. and Sawan, N. (2021) The Importance of Big Data Visualisations for Auditors' Decisions, In *2021 5th International Conference on Software and e-Business (ICSEB)*, New York, NY, USA, ACM, pp. 7–12.

Venkatraman, R. and Venkatraman, S. (2019) Big Data Infrastructure, Data Visualisation and Challenges, In *Proceedings of the 3rd International Conference on Big Data and Internet of Things - BDIOT 2019*, New York, New York, USA, ACM Press, pp. 13–17.