

COMP-1800 Data Visualisation

Name: Chun Hin Attic Lee

Student ID: 001183418

Submission date: 8/4/2022

Introduction

Data visualisation is representing the data that is visualise to graphical elements for the readers, it allows readers to understanding the trend, pattern, and the relationships of the information very quickly and easily. illustration is easier to interpret by people as it does not require any training and it acts as a universal language among people. For example, a relationship between different variation can be observed through a scatter plot, a line graph can represent how the product performs over time, and a histogram can visualise the distribution of the data over certain period. Data visualisation is an excellent way allows user to interact and have a clear idea with the data, hence produce new insight in their business. This data visualisation report is to investigate the data of individual customers visiting venues of ChrisCo using its loyalty card scheme.

Discussion

1.

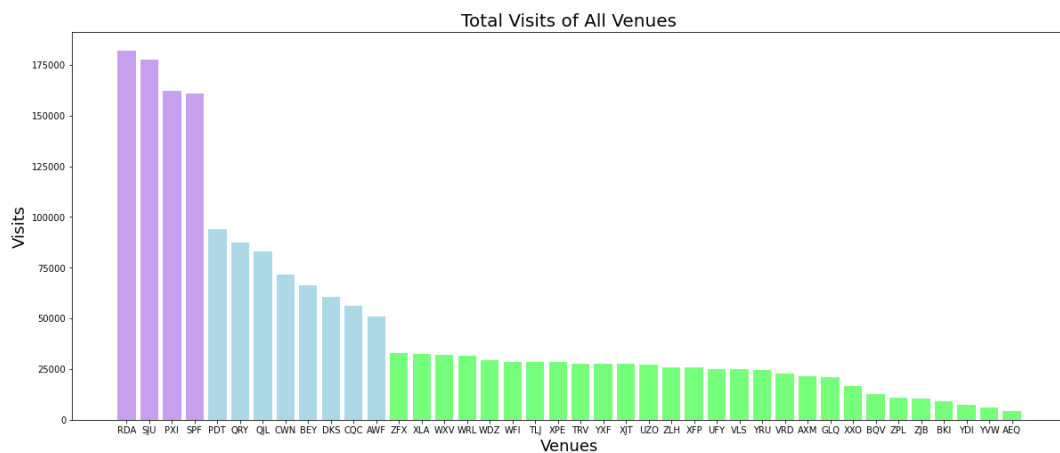


Figure 1 Bar chart of Total Visits of All Venues

Figure 1 shows the number of total visitors of all the venues, which gives an idea about the total volume of each venue and which venues are the most popular among all venues. Also, it helps us to categorise the venues into different category for future analysis.

The bar charts shows that there are 3 group of venue can be segmented into 3 different categories based of their total volume of visits, which is 'High volume', 'Medium volume', and 'Low volume'. The top 4 highest total visitor venues with purple bars are 'High volume' which includes 'RDA', 'SJU', 'PXI', 'SPF'. The total visitor between 15000 to 50000 are 'Medium volume' venue which is in light blue bars. It includes 'PDT', 'QRY', 'QJL', 'CWN', 'BEY', 'DKS', 'CQC', 'AWF'. Finally, 'Low volume' refers to the venues that the number of total visitors is below 50000, which is in light green bars.

2.

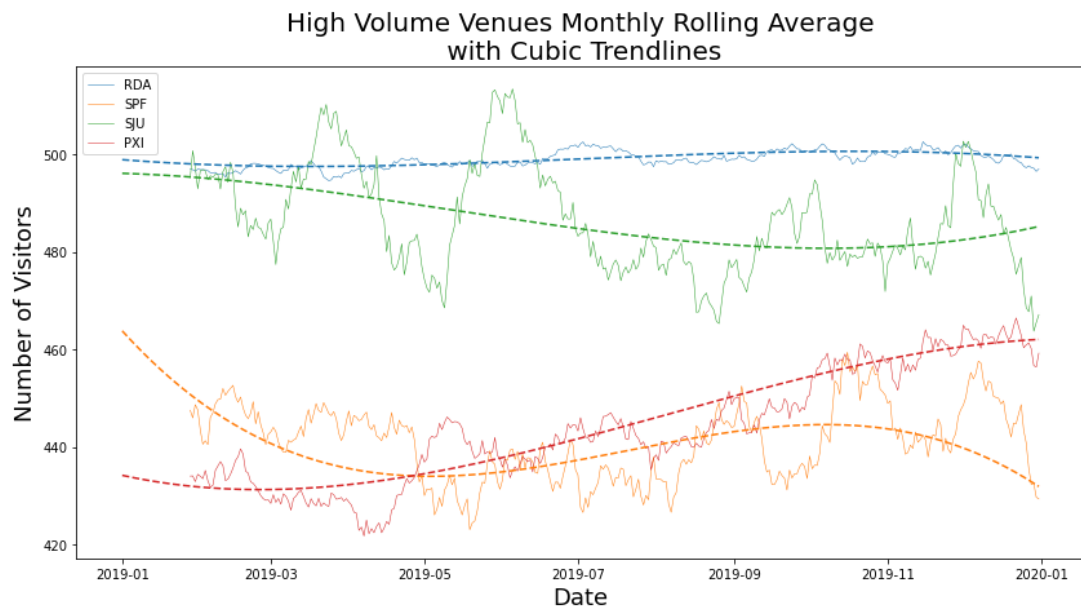


Figure 2 High Volume Venues Number of Visitors Over Time

Figure 2 shows a time-series plot of the monthly average of high volume venue with trend line. This graph provides visualised data of the number of visitors of 4 high volume venues. The behaviour of the venues can be observed through this graph and a business decision can be made to boost a certain period of time in the future if needed.

Figure 2 shows the trend of each high volume venues. 'RDA' is the steadiest venue and is the highest number of visitor venue, which might because it is saturated, demand of this venue is high than its capacity: people prefer 'RDA' more than other venue and it always busy. 'SJU' is the second high number of visitors venue, but the trend of it is going down. Also, a few peaks of 'SJU' can be observed: the number of visitors of 'SJU' soared in April, June, and December. It might because of easter, summer holiday, and Christmas holiday that customer visited 'SJU' for gifts. The trend of 'SPF' started higher than 'PXI', but it was taken over by 'PXI' in May and the increasing trend of 'PXI' is the strongest among the other high volume venues.

Although the number of visitors increased afterward, the growth of ‘PXI’ is still superior.

3.

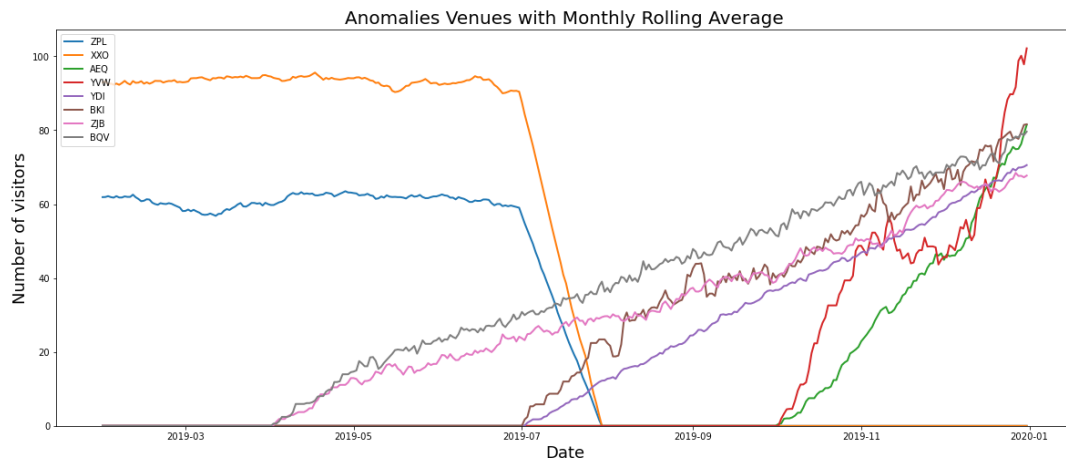


Figure 3 Line Graph of 8 Least Visitor Venues

Figure 3 shows the 8 least total visitor venues daily over the year, which provides the observation to the anomalies venues in ChrisCo.

Figure 3 indicates that ‘ZPL’, ‘XXO’, ‘AEQ’, ‘YVW’, ‘YDI’, ‘BKI’, ‘ZJB’, ‘BQV’ are anomalies venue since their number of visitors dropped to zero during the year, or their number not increasing at the beginning but only during the year.

From this line plot, which indicates ‘XXO’ and ‘ZPL’ were closed during the year in August. BQV and ZJB opened in April. BKI and YDI opened in July, and AEQ and YVW opened in October. ‘AEQ’ and ‘YEW’ are the latest venues, but the total volume of visitors is similar to others ‘Low volume’ venues, which means they have higher daily visitors than other ‘Low volume’ venues. Also, near the end of the year, ‘AEQ’ and ‘YEW’ has a spike on daily visitors. It might because of the venues has some shop or facility that people would like to visit during the Christmas holiday.

4.

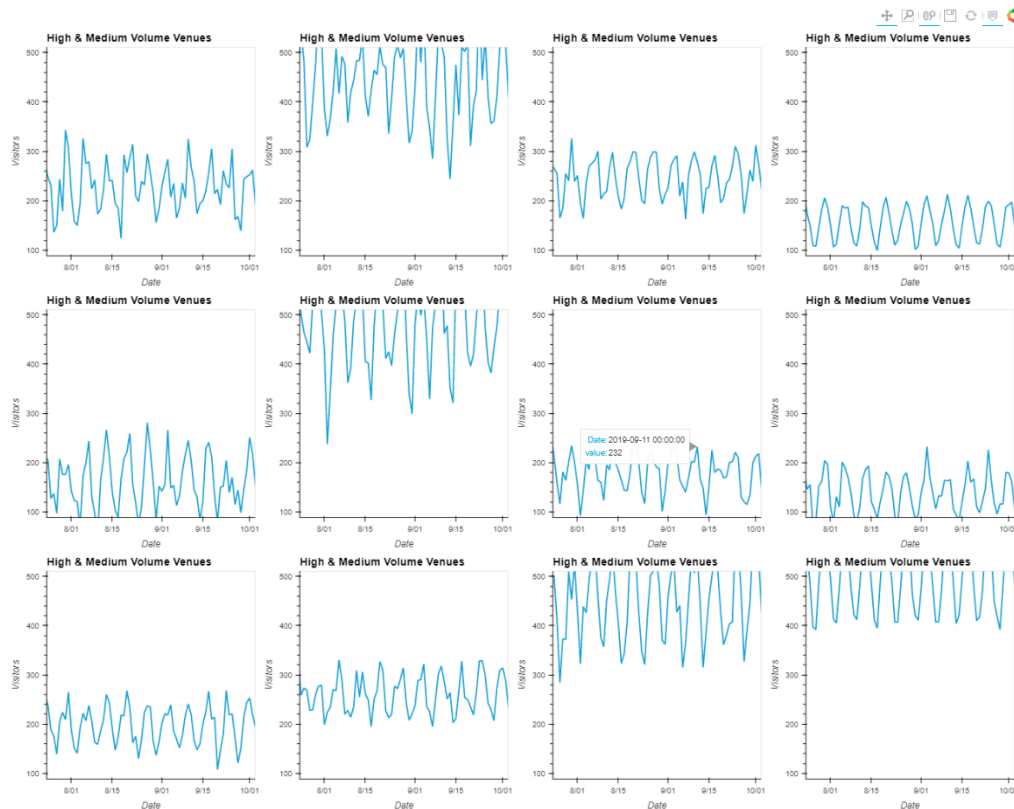


Figure 4 Line plot of High and Medium Volume Venues for seasonality

Figure 4 is a time-series line plot of high and medium venues. It allows user to observe seasonality through interaction. User can zoom in to each subplot by using the tools in top-right corner, hovering to every point of the line to observe the exact date and value of each point, allows more understanding of the data.

In Figure 4, every venue in high and medium volume has the same pattern. By using the interaction function of the visualisation, the pattern is obvious as there is a new peak every 7-days. The data has a weekly seasonality. Seasonality means the characteristic of the time-series data that it has a pattern that repeat regularly. In our case, every venue reached their peak of number of visitors every 7 day, and it achieves another peak after 7 days, which indicates the weekly seasonality of the data.

5.

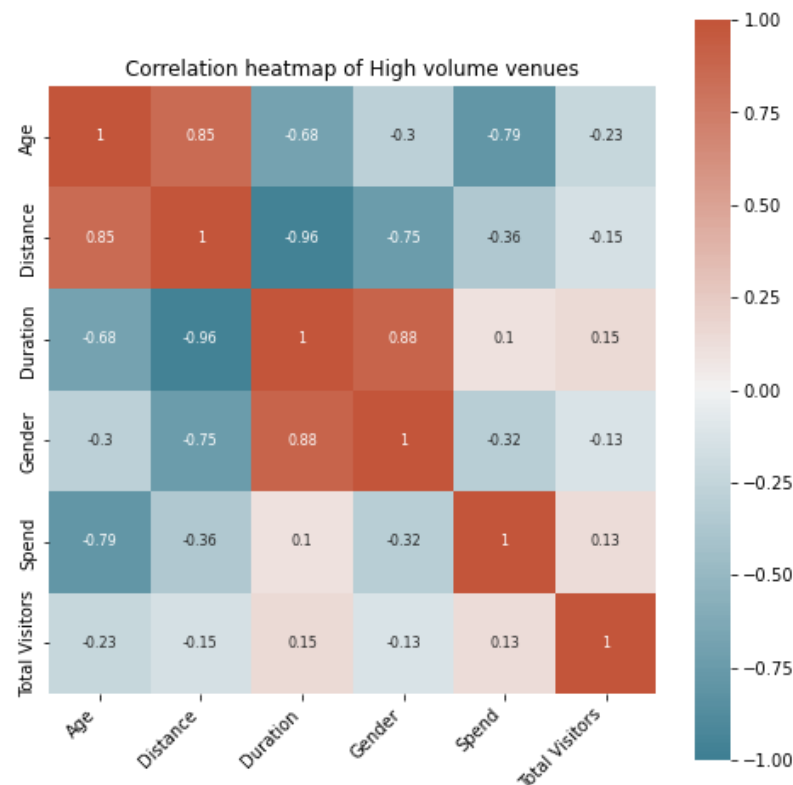


Figure 5 Correlation Heatmap of High Volume Venues

Figure 5 shows the correlation between different elements in High Volume venues.

This heatmap illustrates that relationship between different categories of the venues, it displays some trend that might related to two or more variables.

In Figure 5, darker orange indicates stronger positive correlation, and darker blue indicates stronger negative correlation. In high volume venues, Gender and Duration has the strongest positive correlation, which is 0.88, while Distance and Age also has a high positive correlation, which is 0.85.

Strong positive correlation means both variables (Gender vs Duration, Distance vs Age) tend to go in same direction. In High volume venues, more female customers, the average duration they spend at the venue would be high; and older people tend to visit further and travel longer to the venue.

However, there are some negative correlations in High volume venues. Duration and Distance have strong negative correlation, at -0.96. Spend and Age has the second strongest negative correlation, at -0.79.

Negative correlation means one variable increases as the other decreases. In High volume venues, the Duration increases as Distance decreases, and vice versa; Spend increases as Age decrease, and vice versa.

6.

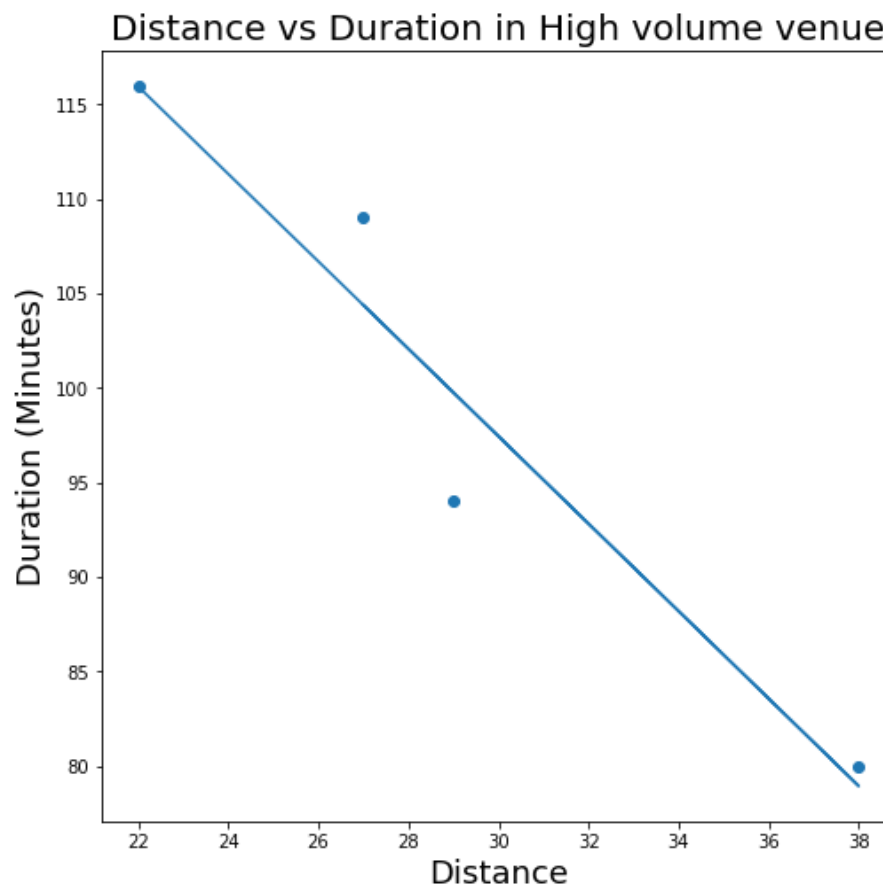


Figure 6 Relationship between Distance and Duration in High volume venue

Figure 6 is a scatter plot visualising the relationship between distance and duration in high volume venue. X axis is Distance and Y axis is Duration, a relationship can be observed between these two variables. As acknowledging the correlation of variables, an impactful business decision and prediction can be made based on the pattern observed. The trend lines describe the pattern of the data.

In Figure 6, there is a strong negative correlation as the trend line goes from top left to bottom right. Which also means as Distance or Duration increases, the other variable decreases. The lower the distance of visitors travel; they tend to stay longer in the venue. It might be because they give more effort of travelling longer distance, therefore

they might tend to spend more time in the venue as a reward for the effort to travel to the venue.

7.

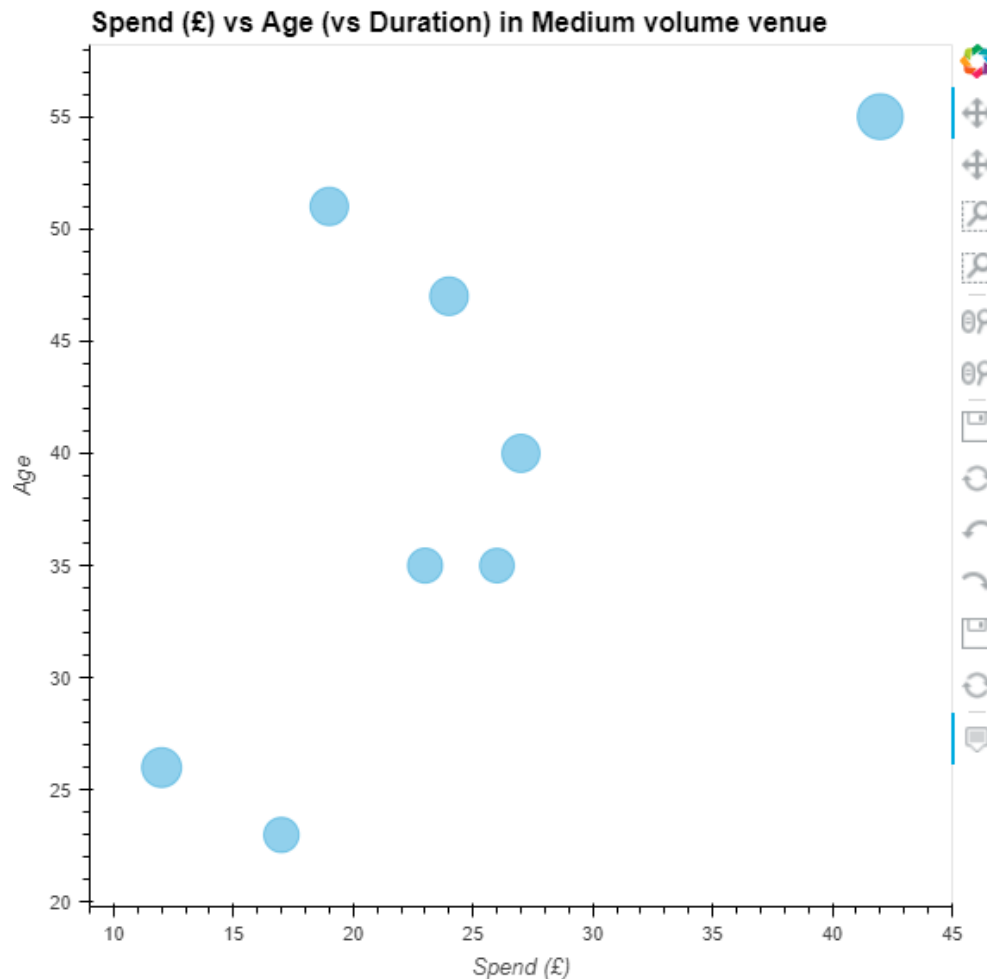


Figure 7 Relationship between spend, age, and duration in medium volume venue

Figure 7 shows the relationship between Spend and Age and Duration in medium volume venues. X axis is the average spend of the visitor, Y axis is the average age of the visitor, and the bubble size represents the average duration of the visitor stay in the venue. Figure 7 shows the correlation between these three variables and allows more efficient business decision to be made.

Figure 7 shows that there is a positive correlation of the data as the X axis and Y axis increases simultaneously. As the age of the visitors increases, they tend to spend more in the venues, it might be because they are more capable to spend more as people usually have higher income when they grow older. Another reason for that is the target audience the shop in the venues are usually older people, therefore older people would be more likely to spend more in the venues. Hence, since they tend to spend more in the venue, most likely they will stay longer in the venues than the younger people. Most of the bubbles lay on within this trend, however, there are an outlier that does not follow the trend: Venue AWF, the bubble located in the top left of the graph, does not fit the trend line, as the age increases, the spend of that venue does not increase at the same rate. The bubble size of data also increases as the spend and age increases, but not so significant.

8.

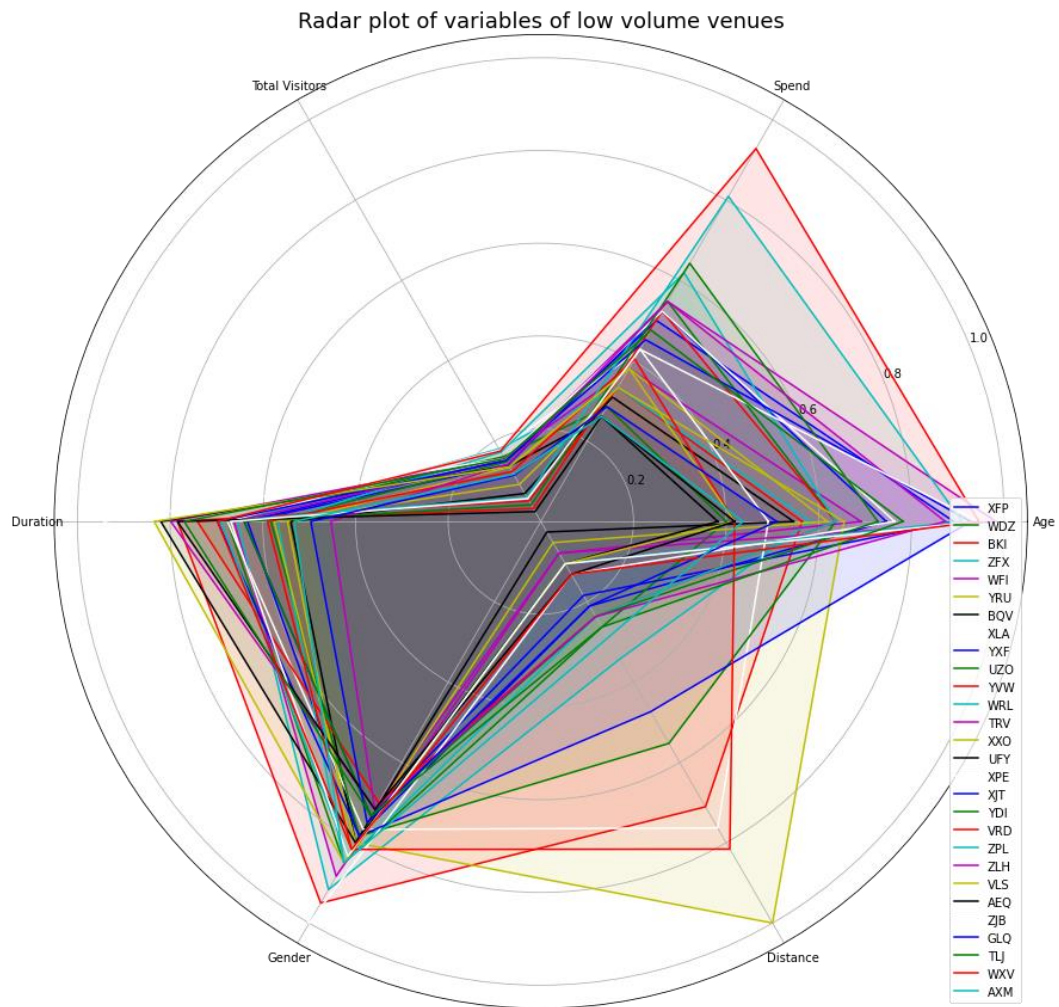


Figure 8 Radar plot of variables of low volume venues

Figure 8 shows the radar plot of variables of low volume venues. It illustrates a comparison and a summary of the variables in low volume venues. Provides a better insight for what the attribute is lacking and what might want to minimise.

Figure 8 shows that the total visitors of all low volume venues are low. Some of the venues has very high value of average distance, which means the location might be the factor that the number of visitors is low. In contrast, some venues have low value

of distance but also low number of visitors, that might because of those venues are not attractive at all, people do not want to go there even it is very close to them.

Moreover, the female proportion is relatively high in low volume venues. It might because the target audience of the venues are mainly female customers. Although the total visitor of low volume is low, some venues have high average spend in it. It might because of the venues lack of some variety, their target audience is too narrow, they will most likely spend money in the venues, but it is only a small amount of people, and they only go to the venue when they need to buy those products.

Critical Review

This coursework applies covers most of the topics in this module, including time, dimension, interaction, etc. It requires me to apply all the learned topic into this piece of work with careful selecting which graph is the most suitable for the topic/ situation. From the beginning, I must understand which venues is in which categories in order to carry out the next steps. It was great that the data is obvious that some are high volume venues, and some are medium and low volume venues after sorting total visitors column and visualise it using a bar chart. For the anomalies venues, a line plot with rolling average made all the work simple as it clearly indicates the venue that drop to zero during the year or starts during the year with the monthly rolling average. Also, the seasonality is obvious through the interaction tool to observe the weekly peak of the time-series data. However, the correlation did not work the way I expected. The correlation did not give an obvious or clear insight of the data. Although it has some strong correlation, those does not have an obvious reason behind it. Hence, it also makes the scatter plot/ bubble more difficult to explain. Nonetheless, this coursework provides a great opportunity to deal with a task that might be quite common in real world – some meaningless data and we must explore it in our own way. It is very different from the lab exercise, which help us to learn data visualisation with some instruction. But the

coursework is much like a real task in business world that require us to apply what we learn in the module into the work, which is a practical and worthwhile practice.

Summary

In this data visualisation, there are some findings of the data:

1. RDA is the most popular venue and SJU's number of visitor soars during holidays.
'PXI' has a strongest increasing trend among the others high volume venues
2. ZPL, XXO, AEQ, YVW, YDI, BKI, ZJB, BQV are anomalies venues.
Venue XXO and ZPL closed during the year, BQV and ZJB opened in April. BKI and YDI opened in July. AEQ and YVW opened in October.
3. In high volume venues, Gender and Duration has strongest positive correlation,
Distance and Age has the second strongest positive correlation. Distance and duration have strongest negative correlation; Spend and age have second strongest negative correlation.
4. Medium venues have many positive correlation but all of these are not strong.
5. Low volume venues have low total number of visitor but some of them has high average spend per customers and some have high percentage of female customers.