

Coursework Sample Report

The attached pages should give you a good idea of what your final coursework report might look like.

Note however, that there are some differences:

- The attached report is part of a larger coursework.
- The attached report deals with visitors to venues – your coursework deals with customers and stores.
- In the attached report, only 6 visualisations were required – in your coursework 8 are required.
- In the attached report there was no requirement to submit interactive visualisations – in your coursework there is.
- In the attached report there was no requirement to write an introduction, critical review or data conclusions – in your coursework there is.

Chris Walshaw

2 PART C: Information Visualisation

The data presented is too large to be explored as a whole, for this reason, it must be segmented, in this case, the venues have been segmented by the total amount of visitors they received in a year into four groups low, low-medium, medium-high and high. Figure 1, is a Bar Graph that shows the total visitors in the year 2019, each bar is placed above a venue code and is the height of the amount of visit. From Figure 1, it is clear that within this segmentation there is a clear group of four venues that have significantly more visitors throughout the year when compared to the other high visitor venues, these venues are, RDA, SPF, SJU and PXI. This could happen because of a wide range of possibilities, such as having popular items exclusively stocked in these venues as this would mean people would visit these venues to get the item. The rest of the venues may not have this demanded item and that may be why there is such a large difference between venues such as PXI and PDT that can be seen in Figure 1.

A Bar Graph see Figure 1, was included in this report because it is very simple and effective at showing one variable across a data set in a very clear way. The simplicity of a Bar Graph allows for a novice to be able to grasp what the visualisation is showing and to draw their conclusions from the graph. Additionally, it is very clear to see further sub-groups within the segment of data being visualised, these discovered sub-groups could then analysed further to find any trends within the groups.

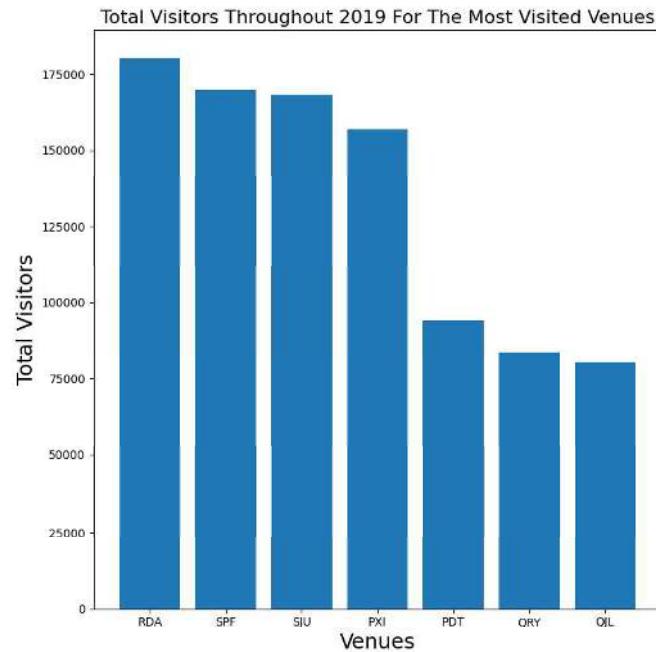


Figure 1: A Bar Chart Which Shows The Total Visitors In A Year Of The Most Visited Venues

Another part of the segmented data, low visitors, was visualised with a Line Plot. Figure 2 shows how the daily visitors change throughout the year 2019, where the colour of the line corresponds with the venue shown in the legend. The less visible line being the daily visitors for the venue, the solid line being a 14 day rolling average for the venue and the dashed line being the trend line of the venue's visitors. Figure 2 clearly shows that the majority of the lowly visited venues that seem to have opened partway throughout the year but are now steadily growing or closed during the year, with the only VRD and AXM having a consistently small amount of visitors throughout the whole year and show no signs of growth when looking at the trend lines. It can also be seen from Figure 2, that the venues have troughs and peaks at fairly consistent intervals throughout the year, this could be due to weekly habits of visitors, for example, that more visitors visit on the weekend rather than the start of the week.

Figure 2 was included, as it allows for a deeper insight into a venue's visitors for the year, as it allows for the visitor amount to be assessed over the time of a year allowing for patterns in visitors amounts to be seen. Such as consistent peaks and troughs or sudden decreases or increases to be seen that would be missed if it was displayed in other methods such as a Bar Graph. Additionally, Figure 2, with the trend lines included makes it clear to the viewer what way the data is growing and how fast. Which is useful for a line graph where finding the overall direction of the data shown can be difficult to grasp as they can get confusing due to the lines crossing.

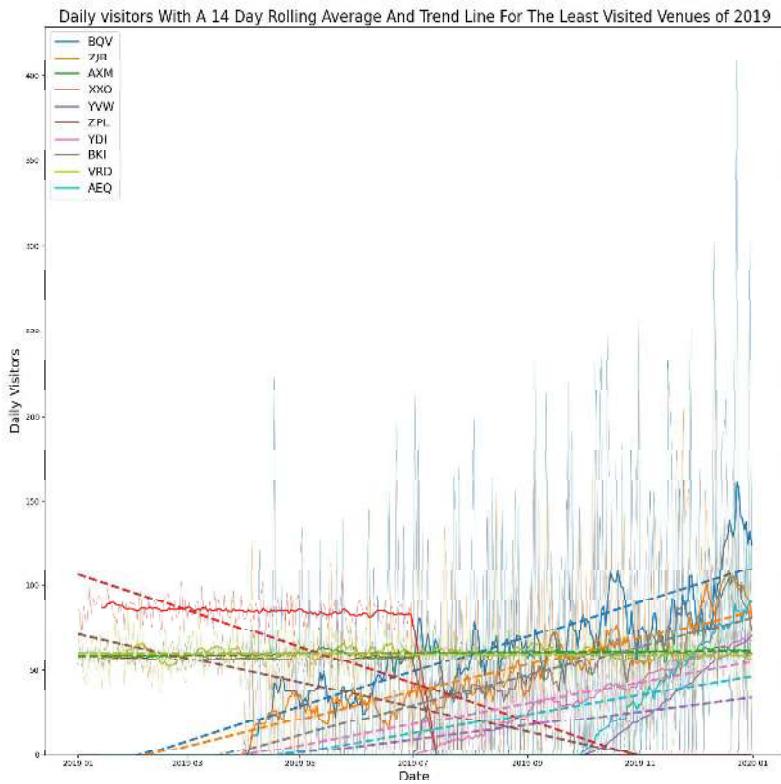


Figure 2: A Line Graph Showing The Venues With the Lowest Number Of Total Visitors Over The Year With the Rolling 14 Day Average And Trend Line For Each Venue

Figure 3 is a Heat Map, that shows the correlation between the six visible categories that make up the summary data. Correlation indicates the relationship between two or more categories. Heat Maps show how strongly correlated the categories on the x and y axes are through two means one is the colour of the square at the intersection of the categories. The next is the value indicated in the square where 1 is a perfect positive correlation where both of the variables increase together and -1 is a perfect negative correlation where one value decreases as the other increase. From Figure 3, it can be seen that there is not a lot of significant correlations throughout the summary data-frame apart from in two cases. Total Visitors and Average Distance displays a close to perfect correlation meaning that as the visitors rise so does the distance from the stores, this is unexpected but suggests that the venues that are further away are more desirable to visit. The next correlation that is worth exploring is the correlation between Average Age and Average Spend, this means that there is a trend that the older the average age of the visitor, the more the visitor spends at the venue.

A Heat Map was included see Figure 3 because it is more concise than looking at Scatter Graphs that compare each category as this would produce multiple graphs that will all need to be analysed by individually to see if there is a correlation between the two categories. Additionally, during the analysis, it can be difficult to see exactly how strong the correlation between the two categories is if it is weak. Utilising a Heat Map allows for the correlation to be shown mathematically between the values of -1 to 1. This removes any guesswork taken on the strength of the correlation, therefore, allowing for more informed conclusions to be drawn from the visualisation. Another reason that the Heat Map was included is that it is quite simple to understand, therefore people with little or no experience in graphs and visualisations will be able to understand the heat map with only a little explanation.

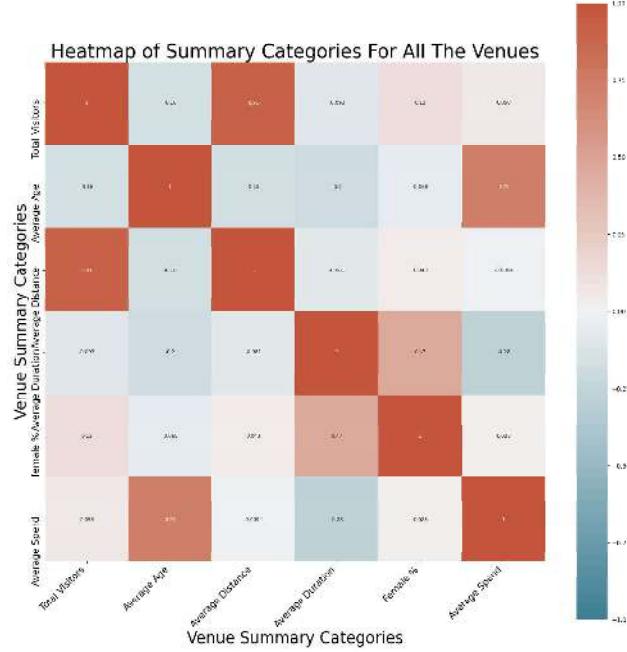


Figure 3: A Heat Map Showing The Correlation Between Venue Summary Categories

A Scatter Graph can be used to plot one category against another, showing where the values intersect for the venues with a small dot, Figure ?? also includes a trend line to show the overall direction the data is moving in this case diagonally to the right. Figure 4, shows the Scatter Plot of the Average Spend category against the Average Age category, confirming what was found in the Heat Map see Figure 3, that these two categories are quite strongly positively correlated due to the direction of the trend line and the relatively uniform placement of the dots. Figure 4, shows that as the Average Age of the visitor increases so does the Average Spend, while it is clear that the correlation is not perfect it is still strong enough to be significant. Additionally, from Figure 4 it can be seen that there is a small group of values grouped at the higher end of both categories that sits above the trend line, these points could indicate that the relationship in a larger data set may not be correlated as strongly as they are represented here, but within the dataset provided the two categories are clearly positively correlated. This correlation may exist as when you grow older your expenses become less meaning you have more disposable income or as someone gets older they could earn more money, in their career also meaning they have more disposable income.

Figure 4, was included in this report to explore further one of the correlations found in Figure 3. A Scatter Plot was selected for this purpose as it is easily understandable even for someone with no prior knowledge of visualisation techniques. Additionally, Scatter Plots also can show trends in the data through analysing the placement of the dots, it can also be used to identify clusters in the data that may be missed if using a different plot to visualise the information. Scatter Plots like Figure 4, also can be used to identify any extreme outliers in the data that may affect any conclusions that would be drawn from the data, with a scatter graph any outliers can be identified then can be dealt with accordingly, for example, excluding it and redoing the visualisation.

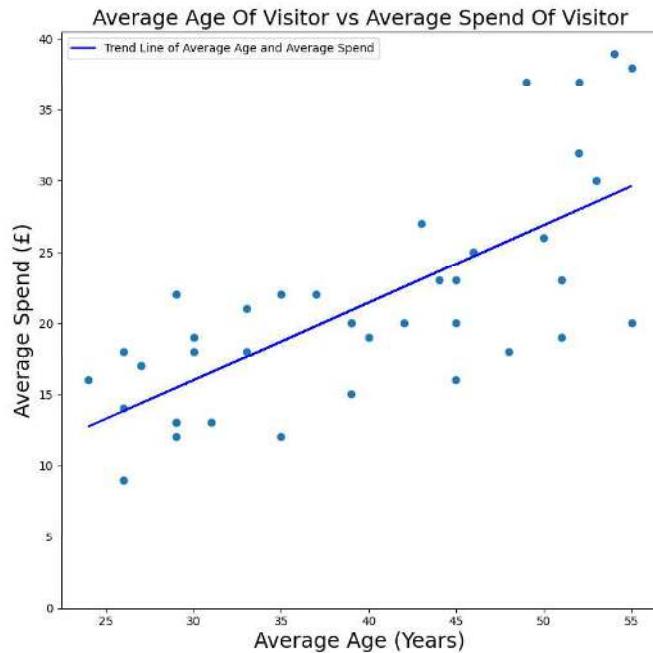


Figure 4: A Scatter Plot Of The Average Spend against the Average Age Labels Are Messed up and zero the data out

Figure 5, is a Radar Plot of the venues that have been categorised as having a large number of visitors throughout 2019. Figure 5, also supports the findings of Figure 3, in that as the Average Age increase so does the Average Spend. This can be seen in venue SPF and QJL and also supports the other correlation found of as the distance increases so does the visitors this can be seen in venues RDA and SJU. Figure 5, shows that there seems to be some amount of correlation between the Female % category and the Visit Duration. However, Figure 3, only indicates that there is a small correlation between these two categories over the whole data set. This suggests that the correlation is stronger for venues with a high amount of total visitors throughout the year, therefore this correlation would not be displayed as strongly within Figure 3, this could be because the venues shown in Figure 5, are aimed more towards female visitors so they therefore stay in the venue longer.

This Radar Plot see Figure 5, was included because it allows for the venues summary data to be compared wholly instead of just comparing one category against another through a method such as a Scatter Graph. This allows for additional relationships between the categories that may have been missed if only methods such as Heat Maps were employed. Additionally, radar plots are clearer and less cluttered than other plots that can achieve a similar goal, such as a comparative Bar Charts making it easier to notice trends and links within the visualised data which may have been missed if a graph such as this was included in the place of Figure 5.

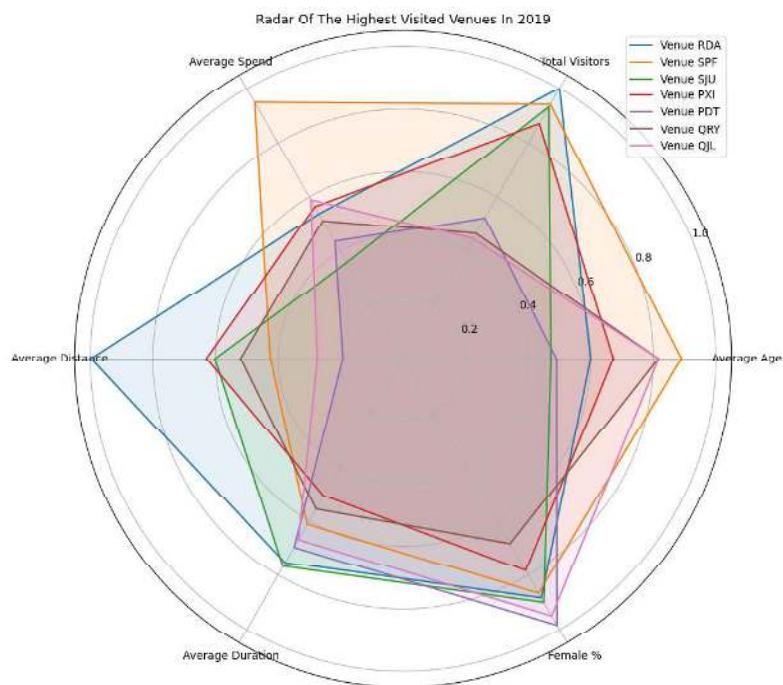


Figure 5: A Radar Plot Of the High Visited Venues Represented With Normalised Data

The Bubble Plot see Figure 6, is similar to a Scatter Plot such as Figure 4, as it will plot values from one category against another. However, Bubble Plots also inflate the size of the dot to show a different variable in the graph this allows for three categories to be compared. Figure 6, shows that the correlation discovered in the Radar Plot see Figure 5, is indeed a correlation and would be classified as a strong positive correlation, as the trend line shows that as the Visitor Female % increases so does the Average Duration of stay for the venue meaning that the two are connected and increase together. Figure 6, also shows that on average the most visited venues are visited by more females than males, this could mean that these venues are aimed more towards females meaning that more females visit these venues and stay for longer when they do.

Figure 6, was included in this report because it can be used to verify and confirm a potential correlation found in Figure 5, by plotting the categories that appear to correlate, on the axes then analyse the trend line these two categories make. While a Bubble Plot can be more confusing than a Scatter Plot that can achieve the same task such as Figure 4. The Bubble Plot Figure 6, was included this time to allow for a third category for each venue to also be evaluated and additional links and trends within the data to be discovered that may have been missed if a Scatter Graph was used.

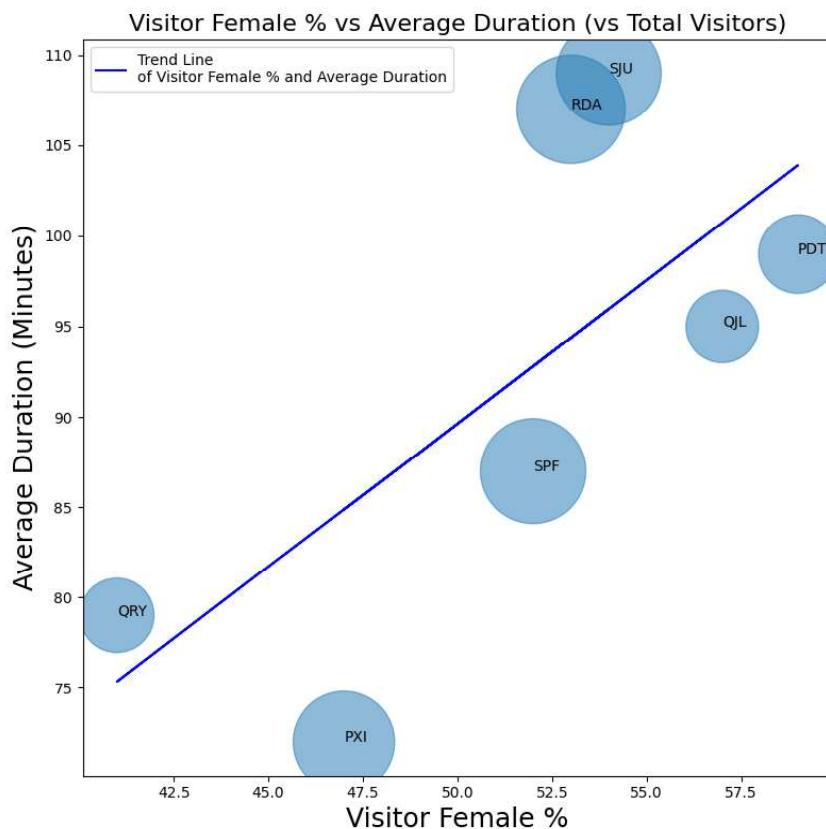


Figure 6: A Bubble Plot Of Visitor Female % and Average Duration, where the bubble size represent the total visitors for the High Visited Venues only