

COMP 1800  
Data Visualisation Logbook.

Student ID - 001002629

9<sup>th</sup> April 2021

## Contents

<b>1</b>	<b>ChrisCo Conclusions</b>	<b>3</b>
<b>2</b>	<b>Critical Analysis</b>	<b>3</b>

## 1 ChrisCo Conclusions

While exploring data that has been provided for ChrisCo 14 day rolling averages for medium volume pages are trending in an upward direction, except for 010 which saw a downward trend over the twelve month period. Correlations between medium volume pages is highest between pages 155 and 156, as a heat map with annotations shows a 0.83 correlation between these two pages, where closer to 1.0 is seen as a higher correlation.

Distribution for hits per day on high volume pages shows that page 001 has a smaller distribution of hits per day when compared to page 015, this can be seen in a box plot where the tails for page 001 are shorter in length. There is also a single outlier for page 001 on the low end of this distribution and 015 has two outliers, one for a low number of hits per day and one for a high number of hits per day. Viewing page hits for these pages in a histogram to see the frequency, show that page 001 has a much tighter distribution than 015 with just over 1000 visits per day 70 times while page 015 had just over 1300 visits per day over 40 times over the period recorded.

Auto-correlation plots for pages 010, 155 and 156 have clear peaks when zoomed for more detail, this indicates that there is a seasonality to these pages of around every 7 days, however when viewing page 010 revenue data, this is much lower than page 015 when seen in a radar plot. From the medium volume pages 003 has very low viewing time but very high revenue.

To cluster these pages, elbow and silhouette methods are used to estimate the number of clusters that exist within this data, where the elbow method shows 4 distinct clusters with a silhouette maxima of 1.

## 2 Critical Analysis

Building the visualisations that have been presented in lectures has been fun, especially the interactive graphs which is something that can be extremely useful but hard to figure out how to implement properly without proper guidance. All the of visualisations have their own title and labelled  $x$  and  $y$  axis to help readers of these graphs understand what is being presented. Not overloading the visualisations with too much information is important to extract meaningful information, while small things like legends that are ordered help with this extraction of information too.

Overall as a Data Science student, it is a critical skill of either creating these visualisations or to interpret data is being presented, learning these skills set out in this module will be of great use moving forward and put to good use during the Masters Projects during the summer and beyond.