

GROUP 9 - Formative Assessment Feedback

The overall look and format of the formative assessment is great and the project aims and background are described well.

With regard to the **distribution of genders**: be careful that the gender column contains 4 different types of entries: “F”, “M”, “O” and “N/A”. From the plot on page 1, it looks as if you have not plotted the missing values there, and you mentioned that you renamed the “N/A” values to “O”, but you should think about whether the “O” and “N/A” values actually mean the same. Be mindful that **“O” could also represent non-binary people**, and not necessarily just missing values.

The rest of the EDA is very well and clear, with meaningful plots. The only thing that could be enhanced is to **make the style of all plots the same**. I especially liked the plots on page 4 and 5!

On page 6, you mention that your final data set only contains 17,000 customers. This means that you have kept all entries with missing values for gender, income and age. It would have been interesting to know **how you plan to deal with these missing values** in the clustering process, e.g. if you want to impute them, and if so, how?

Finally, I like that you explain why you will focus the segmentation on demographics and the amount spent/profitability and link the analysis to a practical business context.

In general, there is a good balance between the plots and written explanation and overall I think this is a great formative assessment, and I enjoyed reading it. I look forward to seeing the final report!