Group 4 Formative Assessment Review

I think this is a very interesting summary of the report. I did notice a few things that I would like more clarification on:

- In the Exploratory Data Analysis, you mention the 13% of customers who have an age of 118. I did not understand what exactly you did with this customer segment. Did you decide to exclude them completely? The data does show that these customers have some transactions. If you decide to exclude them, you may need to explain why in some more detail.
- What is the purpose of the linear regression function? It is very interesting to see how the variables predict log(ATV), but how does that help you cluster the customers? Is log(ATV) going to be a variable you use in the clustering algorithm?
- I like the income range brackets and gender with the "average offer complete ratio". But what is "average offer complete ratio" and how is it calculated?
- In the Next Steps section, you mention you will apply clustering algorithms. Have you thought about which ones you will apply? K-means elbow method or will you explore several?
- The appendix shows you have a lot of other variables to play with! This is great because the rest of the report has a strong focus on age, income and gender. The derived variables seem very interesting, but it is not always clear what is meant. For example, what does the variable comp_rate_bogo_diff05_rew05_dur07 show you and how is it derived? Also, what is the difference between variables amount_transactions and transaction_freq?