Project 2 Milestone 4 10 Questions with Answers

1. Why did you choose this subject for your project?
    a. I enjoyed my marketing analysis class during my undergrad when getting my Business Analytics Degree. I performed a similar problem utilizing K-Means that a really enjoyed and wanted to attempt something similar.
2. Are there any other visualizations that you could have used to improve your project?
    a. I attempted to utilize 3D clustering visuals; however, my laptop could not handle the output. I think an interactive visual of clustering would have attracted the human eye during my presentation.
3. What packages were used when coding your model?
    a. Sklearn.Cluster, Sklearn,processing Sklearn.confustion_matrix
4. Where did you acquire your dataset?
    a. The dataset was acquired from Kaggle website.
5. What visualization do you believe is your best from your project?
    a. It sounds simple but the Boxplots and Bar Graphs at the end of my analysis. This is because I wanted to focus on the interpretation of my results. This was requested on my last project, so I needed to try a different approach this time around.
6. What was your favorite insight from the EDA of part of your project?
    a. I always enjoy seeing the distribution of age when working on projects like this. So it was pretty cool to see how age when applying for credit is always younger and fades off as people become older.
7. Why did you only use KMeans for this project and not any other clustering methods?
    a. Mostly just time constraints. I attempted another method, but I could not get it too work properly in my notebook.

8. Which Cluster surprised you the most?
    a. Cluster 1. I thought there would be an even split of men and women within a cluster of higher credit. But overlooking my visualizations, it made sense because the dataset was skewed towards males.
9. What clustering method would you want to try next?
    a. Hierarchal clustering.
10. What do you believe is needed to make this a better project?
    a. I would want more data to work with. The skewness from the lack of females within the dataset I believe did not help when creating the clusters.