**Candidate Name: Mohamad Aznor Bin Mohamed Yusof**

Please note the following:

1. Both questions use the same dataset
2. All source code must be provided with comments along with instructions on how to execute the code.
3. R is preferred, but you are free to consider other tools. Regardless of the tool you use, source code / workings must be provided and clearly documented
4. Two datasets are provided **spenddata.csv** and **testdata.csv.** You are free to decide how best to use them.
5. You do not need to augment this dataset with any external data. However, if you choose to do so, you must document the reasons.

**Question 1**

You are given a set of survey data which captures spend amount among other data points. However, some of the spend amount (**totshopping.rep**) is missing.

Build a model that will help predict the amount spent by a visitor based on the set of data provided.

* 1. Please explain the choice of metric used
  2. What are the assumptions you made when building this model?
  3. What were the approaches you considered? Please explain the reason for the technique / approach used as well as the pros and cons.
  4. Please explain under what conditions will the model you choose **be not appropriate**
  5. How confident are you of the model’s robustness and how would you explain the model’s performance?
  6. Why is your model performing well / not well?
  7. Was any feature engineering required? If yes, what were they. If no, why?

**Question 2**

Some of the respondents have been tagged as belonging to group 1 – 6. However, due to a data calculation issue, some of the respondents have had their groups (**pov6**) missing.

Build a model that will classify these respondents back into one of the 6 groups.

1. Please explain the choice of metric / evaluation criterion used
2. What are the assumptions you made when building this model?
3. What were the approaches you considered? Please explain the reason for the technique / approach used as well as the pros and cons.
4. Please explain under what conditions will the model you choose **be not appropriate**
5. How confident are you of the model’s robustness and how would you explain the model’s performance?
6. Why is your model performing well / not well?
7. Was any feature engineering required? If yes, what were they. If no, why?

**Question 3**

Attached is a mock survey dataset for visitors from a particular country.

1. List the areas you would look into when it comes to data preparation
2. Highlight any data idiosyncrasies that you might have observed
3. Put together a short powerpoint presentation to tell a story about the key insights and observations, taking into consideration dataset limitations.

Do include your supporting scripts/output from R and or Microsoft Excel.