Candidate Name:

Please note the following:

1. Both questions use the provided datasets
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. Refer to **mock survey data 3.1.xls** for Question 1 and 2.
5. You are free to augment this dataset with publicly available data to sharpen your insights / recommendations. However, if you choose to do so, you must document the reasons and how you are using the data to augment your insights / recommendations.
6. You can refer to the sheet “dictionary” for a description of the fields to aid in your analysis.

**Background**

The marketing team is looking to refresh their marketing strategy and reached out to the Data Science team for help. A survey was commissioned to collect visitor data over a period of 12 months to understand their spending behaviours and preferences. You are now in-charge of **analysing this data and presenting visitor behaviour insights** to help the **inform the marketing team on their strategy to increase visitorship and spend** in Singapore.

**Question 1**

With reference to the mock survey dataset (**mock survey data 3.xlsx**),

1. Describe the **analytical** **approach you will take and data fields** **you would look into** when it comes to doing exploratory data analysis.
2. Highlight the **data idiosyncrasies / issues** you found in this dataset and **how would you address** the issues identified.
3. What are the considerations that you will take when analysing survey data.

Use this data set to **illustrate and explain** **your approach and considerations**. Please include your supporting **scripts/output from R or an alternative tool** of your choice.

**Question 2**

Using the mock survey data (**mock survey data 3.1.xlsx**), answer the following questions.

1. You are given a set of survey data which captures spend amounts among other data points. The Marketing team is curious to understand **the different segments of visitors** who are coming to Singapore. You are given the task to analyse the survey data so that the marketing team can formulate **data driven strategies** to draw in more visitors to Singapore.

Using the set of data provided, answer **ALL** of the following questions:

* 1. How many different segment of visitors can be identified from the survey data
  2. Please explain the choice of metric you used
  3. What are the key assumptions you have taken to identify the different segments?
  4. What were the other approaches you considered? Please explain the reason for the technique / approach used as well as the pros and cons.
  5. Why is your analytical approach performing well / not well?
  6. Was any feature engineering required? If yes, what were they. If no, why?

1. Based on the analysis done in (1), answer the following business questions:
   1. **Describe and profile** the segment tourists that you have identified
   2. What can we learn about our visitors from the survey data as well as the visitor segments that will help the Marketing team better market Singapore as a tourist destination them?
   3. Is there a **strong or weak** **relationship between travel companions and choice of hotel**? Please provide an explanation on how you arrived at your conclusion.
2. With your **findings from (Questions 1, 2.1 and 2.2)**, **prepare a concise PowerPoint presentation to pitch** your strategy to the Marketing Team on how will your insights help STB attract visitors to spend in Singapore. *Hints:*
   * *You are encouraged to explore the data* ***beyond*** *the questions asked in (a) – (c)*
   * *You* ***should use your analysis results from question 1*** *to support your recommendations*
   * *Use appropriate charts / visuals to communicate your story*
   * *Be careful with the chart-ink ratio*