**Mini Project 1**

**Choose a Problem Statement of Your Liking**

**Problem Statement 1: To understand the factors contributing to heart disease.**

The following items have been provided to you:

1. Patients demographics and vital signals
2. Data Dictionary
3. Paper: “*Computational intelligence for heart disease diagnosis: A medical knowledge driven approach*” (Helps to understand the data)

**OR**

**Problem Statement 2: To understand the factors contributing to guest satisfaction.**

ReviewPro is an aggregator of hotels and restaurants. The raw data set contains information about all the hotels in Singapore. The respective tabs in the excel sheet contain:

1. Numeric score of reviews based on business center, cleanliness, decoration, entertainment, food & drink, location, reception, room, service, and value. An additional score (GRI) by ReviewPro is also shared.
2. Sentiment mentions related to topics such as room, breakfast, service etc. have been scraped off the internet and classified as positive and negative.
3. STB Rating: Star Category of a hotel. eg: Riverview Hotel is a 3-star hotel.

**Your tasks are as follows**-

1. Clean / transform / map / pre-process / slice and dice the data.
2. State assumptions if any.
3. Using your knowledge of visualisations and statistical analyses, apply appropriate techniques and tests (where applicable) to value-add and draw inferences from the data. Please place all your analysis in a python notebook with proper comments and report style explanations.
4. Highlight the key takeaways / recommendations from your “additional” analysis.
5. Prepare a presentation (10 slides only including Cover Page and Thank You). Include screenshots and visuals, as necessary. Your group will be invited to make a brief (10 minute) presentation.
6. Submit your codes and notebooks to Github.

**Hypothesis Testing:**

Hypothesis testing is the use of statistics to determine the probability that a given hypothesis is true. The usual process of hypothesis testing consists of four steps.

1. Formulate the null hypothesis H\_0 (commonly, that the observations are the result of pure chance) and the alternative hypothesis H\_a (commonly, that the observations show a real effect combined with a component of chance variation).

2. Identify a test statistic that can be used to assess the truth of the null hypothesis.

3. Compute the P-value, which is the probability that a test statistic at least as significant as the one observed would be obtained assuming that the null hypothesis were true. The smaller the P-value, the stronger the evidence against the null hypothesis.

4. Compare the p-value to an acceptable significance value alpha (sometimes called an alpha value). If p<=alpha, that the observed effect is statistically significant, the null hypothesis is ruled out, and the alternative hypothesis is valid.

**Sample PPT Template:**

1. Intro Page with Team Members
2. Business Objectives, Data Analytics Objectives
3. Data Description, Assumptions (If any)
4. Methodology
5. Findings and Results
6. Recommendations / Action Items
7. Thank You Slide
8. Appendix

**Grading:**

1. Assistant Trainers
2. Peer Evaluation

**Due:**

* Submission should be by 11pm Friday 5th March, 2021
* Presentation (1 person only) on Saturday 6th Mar 2021 (be ready to go by 9am; order to will be sent closer to the day)

**Evaluation Matrix:**

1. Business context, stakeholders and values

* Overall understanding of the business domain
* Explanation of the business context
* Formulation of the business question
* Understanding and engagement of stakeholders
* Estimation of the business value

1. Data exploration, analysis and visualisation (/1

* Data exploration showing the key entities and their business significance
* Using effective visualisation to communicate key aspects of the data

1. Delivery of the presentation, poise and audience engagement (/10)

* Ability to deliver a clear, concise and engaging presentation
* Ability to listen effectively and address questions
* Overall poise, confidence and rapport with the audience
* Keeping the time

1. Recommendations and Insights(/10)

* Delivery of key message
* Action Items / Strategies for business