NGEE ANN POLYTECHNIC

School of InfoComm Technology

Data Visualisation

Diploma in Financial Informatics

October 2020 Semester

ASSIGNMENT 2
(Individual Assignment)

Submission Deadline:
11th February 2021 (Thursday), 12:00PM

Tutorial Group	**:**	**P01 / P02 / P03**
Student Name	**:**	
Student Number	**:**	

Penalty for late submission:

10% of the marks will be deducted every calendar day after the deadline.

NO submission will be accepted after 19th February (Friday), 10AM.

NGEE ANN POLYTECHNIC

DATA VISUALISATION ASSIGNMENT

1. INTRODUCTION

The goal of this assignment is to perform date preparation, exploration and visualisation on a given dataset. You will be involved in cleaning and preparing the data where necessary, exploring the data, and implementing your visualisations and dashboards **using Python programming in Jupyter Notebook**.

This assignment constitutes **40% of your overall DV module grade**. It is an individual assignment.

2. OBJECTIVES

a) Understand, prepare and clean a curated dataset programmatically.

b) Identify meaningful declarative and exploratory questions from the dataset.

c) Create a multitude of visualisations to effectively derive answers to the declarative and exploratory questions identified earlier.

d) Perform data story telling using dashboards and/or visualisations for target audience.

3. BACKGROUND

Imagine that you are part of the market research team for Cardio Good Fitness, a retail business specializing in the sales of treadmills. The team has collected data on individuals who purchased a treadmill at the Cardio Good Fitness retail stores for the past three months. The data is stored in the **cardio_good_fitness.csv** file. Through data preparation, exploration and visualisation, the market research team decides to investigate whether there are differences across the product lines with respect to customer characteristics.

The data consists of the following variables:

1. **Product:** Model of treadmill purchased (TM195, TM498, or TM798);
2. **Branch:** Branch location where purchase is made (North, South, East or West);
3. **Gender:** Gender of customer (Male or Female);
4. **Age:** Age of customer in years (integer values);
5. **Education:** Number of years of education customer had completed (integer values);
6. **MaritalStatus:** Marital status of customer (Single or Partnered);
7. **Usage:** Average number of times the customer plans to use the treadmill each week (integer values);

8. **Fitness:** Customer's self-rated fitness on a 1-to-5 scale (1: very unfit; 5: very fit);
9. **Income:** Customer's annual household income (integer values);
10. **Miles:** Average number of miles the customer expects to walk/run each week (integer values).

4. ASSIGNMENT REPORT RECOMMENDED STRUCTURE

You may use the following pointers to structure your report. You should include visuals, diagrams and/or tables where appropriate to enhance your report's readability.

1) Project Objectives

- Based on the Cardio Good Fitness scenario, come up with a list of primary business/research questions you are trying to answer with your visualisation(s).
- The questions may evolve over the course of the project. Update this part by adding or removing new business/research questions you consider necessary in the course of your analysis.

2) Data Preparation

- Describe the state of the data (e.g. is it clean, is it complete). Do you expect to do substantial data cleanup?
- How will you prepare your data so that it is suitable for exploration and analysis? Describe the steps involved in detail.

3) Visualisations

- Describe the process to perform univariate and multivariate analysis on your dataset.
- Using descriptive analytics techniques (i.e. statistical analysis, correlation analysis and basic visualisations etc), document your findings.
- Identify the core findings and insights that help to answer the business/research questions identified earlier. Create the visualisations using Python programming.
- What are the different visualisations you have considered? Justify the design decisions you made using visualisation and design principles.
- Describe each visualisation by highlighting the business/research questions it answers.
- Explain how to interpret the visualisations in order to answer the business/research questions (e.g. interactive elements).

4) Dashboard

- From your library of visualisations created earlier, identify suitable ones to create a dashboard. Recall that dashboards should incorporate only visualisations related to one main topic.
- Depending on your business/research questions, create multiple dashboards to aid you in performing your video recorded data story telling presentation subsequently.

Note: Include all the screenshots of your visualisations and dashboards in your Report.

5. PRESENTATION AND DATA STORY TELLING

Using your dashboards and visualisations, you are required to submit a **video recorded presentation** to perform data story telling and share your findings. The video recorded presentation **should not exceed 5 minutes**. Video recorded presentations which exceed the allotted time will be penalized.

You must record your video presentation using **Microsoft Teams**.

After completion of your video recorded presentation, you are required to **submit the link to your video** (from Microsoft Stream). Instructions to submit your video recorded presentation link are provided in the following section.

6. ASSIGNMENT DELIVERABLES

Upon completion of the assignment, you are required to submit the following deliverables:

i. Assignment Report (in Microsoft Word / PDF format)

- Submit in "**Assignment 2 Report Submission**" folder in MeL

ii. Jupyter Notebook (in .ipynb format)

- Submit in "**Jupyter Notebook Submission**" folder in MeL

iii. Link to video recorded presentation (using Microsoft Teams)

- Submit in "**Assignment 2 Video Presentation Submission**" Link in MeL **(Login using only your NP student account)**

All deliverables are due for submission by 11th February 2021 (Thursday), 12:00PM.

NGEE ANN POLYTECHNIC

7. ASSESSMENT CRITERIA

No.	Deliverable	Weightage
1.	Assignment Report	40%
2.	Jupyter Notebook	30%
3.	Video Presentation	30%
	TOTAL	100%