



NGEE ANN
POLYTECHNIC

School of InfoComm Technology

Data Discovery & Visualisation

Specialist Diploma in Data Analytics

October 2022 Semester

ASSIGNMENT 2 (Individual Assignment)

Submission Deadline:

8 February 2023 (Wednesday), 2359 hrs

Student Name	:	<input type="text"/>
Student Number	:	<input type="text"/>
Video link	:	<input type="text"/>

Penalty for late submission:

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

NO submission will be accepted after **15 Feb 2023 (Wednesday), 2359 hrs.**

DATA VISUALISATION & STORY TELLING ASSIGNMENT

1. INTRODUCTION

The goal of this assignment is to perform data 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 DVST 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 tourism research team for Merlion Tour, a tour agency specializing in local Singapore tours. The team has collected data on Indonesian tourist visits to Singapore for past 2 years. The data is stored in the **Indonesian traveler data.xls** 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.

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 Merlion Tour 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.

5) Reflection

- How does the work on assignment 2 compare to assignment 1?
- What would I do differently next time, if given the chance?

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

5. PRESENTATION AND DATA STORY TELLING

Use your Jupyter notebook to share your findings in a live presentation, in a face to face class or online synchronous using MS Teams (depending on the prevailing COVID Situation).

The majority of the presentation should focus on the dashboards and the business questions it answers and your recommendation for interventions. The presentation should not exceed 10 minutes. Presentations which exceed the allotted time might be penalized.

If the presentation is done using Microsoft Teams, the session might be recorded, at the discretion of the tutor.

6. ASSIGNMENT DELIVERABLES

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

i. Assignment Report (in Jupyter Notebook format)

- Submit in "Assignment 2 Report Submission" folder in PoliteMail
- Make sure all outputs are generated and saved in your Jupyter Notebook

ii. Link to video recorded presentation (using **any tools** such as MS Team, PowerPoint, etc)

- Upload the video to YouTube as unlisted. **Make sure your Google account is registered and verified.**
- Copy and paste the YouTube link in "DDV Assignment 2 Report Cover".
- Below are guides on how to do recording and uploading Unlisted YouTube.

iii. Reference:

- **How to record Video with MS Team**
<https://youtu.be/hjfQ41LF0w>
- **How to Record Desktop Screen Using Zoom**
<https://youtu.be/P6cTbnUPwfY>
- **How to upload an unlisted video to Youtube**
<https://youtu.be/WkgOvUr5Alc>

7. ASSESSMENT CRITERIA

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

8. MARKING SCHEME GUIDE:

<p>Above Average A+ to B+</p>	<ul style="list-style-type: none"> • Excellent/Good list of excellent exploratory questions tailored for target audience. Excellent justification of target audience, and detailed data preparation steps. • In-depth analysis coupled with accurate application of analytical techniques, yielding valuable insights in answering all exploratory questions. • Excellent/Good mix of visualisations that are thoughtfully enhanced, coupled with correct and accurate interpretation of all visualisations and dashboards. • Excellent/Good design of all visualisations with perfect consistency of design elements across all visualisation and dashboards. • Demonstrated good learning analysis and critical thinking in reflection. • Presentation: <ul style="list-style-type: none"> ○ Skills: clear, lively, imaginative; great use of visual aids with lively animation. ○ Use of relevant and accurate evidence: key points supported with highly relevant and objective evidence, critically evaluate. ○ Detailed analysis of data insights, coupled with good data storytelling techniques for a suitable target audience. ○ Detailed sharing of data exploration, coupled with excellent use of relevant analytical techniques to showcase insights that are highly relevant to target audience.
<p>Average B to C</p>	<ul style="list-style-type: none"> • Largely correct level of analysis, with a suitable mix of analytical techniques applied to answer some exploratory questions. • Largely good variety of visualisations with accurate descriptions of visualisations and dashboards. • Largely correct visualisations with good effort in enhancing the visuals. • Use of dashboard that are populated with largely suitable visualisations. Some inconsistent design of visualisations with some/little naming of visuals and dashboards. • Some reflection with little critical thinking on learning and analysis. • Presentation: <ul style="list-style-type: none"> ○ Presentation Skills: conveys the meaning generally, but sometimes unclear; use of appropriate visual aids with some animation. ○ Minimal sharing on data insights and/or lack of suitable visualisations for data exploration.

<p>Below Average D+ to F</p>	<ul style="list-style-type: none"> • Minimal listing of exploratory questions and/or poor justification of target audience(s) and/or poor explanation of data preparation steps. • Minimal analysis using various analytical techniques or wrong interpretations. Incomplete and skimpy document. • Mixture of poor or incorrect visualisations, with little enhancements performed. • Mediocre/poor use and design of dashboards. Some inconsistent design of visualisations with some/little naming of visuals and dashboards. • Little or no reflection. • Presentation: <ul style="list-style-type: none"> ○ Presentation Skills: not always clear or easy to follow, clumsy, disjointed, dull. Basic use of visual aids. ○ Little to no sharing on data insights or wrong insights. ○ Use of relevant and accurate evidence: little or no evidence discussed; or irrelevant and inaccurate.
---	--