|  |
| --- |
| **EBIS3033 Programming for Business Applications**  **2024-2025, Semester 2** |

**GENERAL GUIDELINES:**

**Team Formation Deadline:** by Week 2 (**5 members** per group)

**Group Assignment Deadline:** Week 13 (the detailed deadline time shall be released on iSpace)

**Group Presentation:** Week12-13

**Group Assignment Instructions:**

**Comprehensive Python Project - Building a Data Analysis Tool**

In this group assignment, you will utilize the concepts and skills learned throughout the course to develop a comprehensive data analysis tool that demonstrates your understanding of these concepts and your ability to apply them in a practical setting.

* Your team is tasked with downloading a comprehensive dataset pertaining to a specific firm, store, or industry from either Tianchi (https://tianchi.aliyun.com/dataset) or Kaggle (https://www.kaggle.com/datasets).
* After the data collection, randomly sample **500** observations out of the dataset.
* Upon acquiring the data, you are required to conduct an in-depth analysis aimed at understanding its historical performance. This involves selecting and utilizing appropriate **Key Performance Indicators (KPIs)** that you deem fit for evaluating the entity’s performance (e.g., conversion rate, Customer Lifetime Value, Customer Satisfaction Score).
* Furthermore, in calculating these KPIs, you must leverage **Python** programming skills acquired through your coursework. This includes utilizing libraries such as Pandas for data manipulation and preprocessing, NumPy for numerical computations, and Matplotlib or Seaborn for data visualization. Specifically, you should demonstrate proficiency in tasks such as handling missing values, converting data types, and performing necessary aggregations or transformations to derive the selected KPIs with self-defined functions. Moreover, you are expected to write clean, efficient, and well-documented **code**, adhering to best practices in data analysis and ensuring reproducibility of your results.

**Formatting:**

* The report should be in **15-18 pages**, including tables and figures, excluding the cover page and table of contents; *avoid using too many screenshots of codes and outputs, show and explain main content only.*
* **Spacing:** Double Spacing
* **Font:** Times New Roman; Size 12
* **Margin:** 2.54 cm per Side
* **Submit a compressed file (e.g., Group?.zip file) containing:**
  + 1. **a Group?.ipynb file: i**nclude the Python code used for data acquisition, transformation, exploration, and analysis, ensure the code is well-commented.
    2. **Data file (csv or txt file)**
    3. the report, a **Group?.pdf file**.
* All files must be written in **English** only.

**Content Template:**

**Cover Page (1 page):**

* **Project Title,**
* **Details of the Group**
  + - Students' IDs
    - Names
    - Group Number
    - Course name
    - Section Number

**Table of Contents (1 page)**

1. **Project Overview:**o Provide a brief overview of the project, including the objective and expected outcomes.  
   o Explain how the knowledge gained from the course lectures will be applied in the project.
2. **Data Acquisition and Transformation:**o Describe the source of the data and how it was collected.  
   o Explain how Python was used to read and sample the data.  
   o Detail the steps taken to transform the data into a suitable format for analysis.
3. **File Handling:**o Explain how Python was used for file handling during the project.  
   o Discuss any challenges encountered during this process and how they were resolved.
4. **KPIs Calculation, Dataframe Manipulation, and Results Interpretation:**o Present KPIs analysis and interpret your calculation process; Elaborate on the methods used to manipulate the data to calculate KPIs.  
   o Explain the reasoning behind these methods and how they help in achieving the project's objectives.
5. **Data Visualization:**o Describe how data visualization was used to identify patterns and trends in the data and KPIs.  
   o Present the visualizations and explain what they represent.
6. **Error Handling and Debugging:**o Describe any errors encountered during the coding process and how they were resolved.  
   o Explain how Python's debugging tools were used to identify and fix these issues.
7. **Conclusion:**o Summarize the key findings from the data analysis and how they meet the project objectives.  
   o Discuss the implications of these findings and suggest areas for further investigation.

**Group Presentation Instructions:**

* **Formatting: 20mins** for presentation & **5-10mins** for Q&A
* **Content** is decided by your group to address the key points of your team project. For example: dataset explanation, KPI definition and calculation logic and visualization outputs.
  + **All** group members should present.
* Print a **hard copy** of your slides (2-4 slides per one side of a page) and submit it to me before presenting your presentation.
* **Structure tips**: outline and clear titles, avoid typos in slides.