**General Education Department**

**Project Fall 2024**

|  |  |  |  |
| --- | --- | --- | --- |
| **Course** | **Introduction to Programming - PROG101** | | |
| **Assessment Method** | Individual Project | | |
| **Date of Assessment** | 14/11/2024  24/11/2024 | **Duration / Deadline(s)** | 10 days |
| **Maximum Mark** | **100** | **Percentage of Final Grade** | **20%** |
| **Student ID** |  | **Student Name** |  |

|  |
| --- |
| **Instructions to Students**   1. **This assessment consists of 8 pages, including the cover page.** 2. **Copying from any source or collaborating with other students will result in a score of zero.** 3. **Each student must submit the project deliverables individually through the VLE (Virtual Learning Environment).** |

|  |
| --- |
| **Academic Honesty Statement**   * **Students are expected to uphold integrity by avoiding all forms of academic dishonesty, as outlined in HBMSU procedures and guidelines, as well as instructions from HBMSU personnel.** * **Any student found guilty of academic dishonesty may face disciplinary actions as specified in the Student and Academic Regulations.** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question (Section) No.** | **Deliverable Report (PDF), &**  **Python Code in PDF** | **Presentation in PPT** | **Total** | **Percentage** |
| **Marks Allocated** | **70** | **30** | **100** | **20%** |
| **Awarded Marks** |  |  |  |  |

# Project Objectives

The aim of this project is to provide the students with the opportunity to:

* Work with real-life datasets
* Demonstrate the competencies and skills needed to perform fundamental data analytics
* Write and present a brief executive report of findings and recommendations

# Project Description

You are required to work individual to analyze real time data using Pandas. In this project you will read the real-life dataset from national or international data sources on suitable domain. Set 10 objectives to analyze for the chosen data set using Python Pandas library. Write a well-structured report that contains executive summary and recommendations on your findings.

**Skills to be demonstrated:**

The selected dataset and derived data analytics challenge the student to demonstrate the following skills:

* Ability to read data from external files and store data in a Pandas Data Frame
* Analyze data by Sort data, Filter data, Group data
* Visualize data by appropriate plotting/charting

**Data Sources:**

Students are encouraged to select an appropriate dataset from any of the open data projects including:

* UAE Open Data Project: <https://bayanat.ae>
* USA Open Data Project: https://www.data.gov
* European Open Data Project: <http://data.europa.eu/euodp/en/home>
* Kaggle Data set: <https://www.kaggle.com/datasets>

The dimensions of an appropriate dataset are at least 100 rows by 4 “relevant” columns.

# Project Deliverables

### Project Report

Create and submit a well-structured MS Word Report including all Python code screen shots using the given template. The report must include the following sections: **[CLO4] [Deliverable 1] [50 marks]**

1. **Project introduction [4 marks]**

Describe the project aim and the chosen dataset in terms of source and columns.

1. **Analysis Questions:** **[10 marks]**

You need to define **10 questions** for the chosen dataset in order to analyze and visualize its data. Your 10 questions must have the constraints as shown below.

**Question1**: Data Analysis using statistical methods.

**Question2,3**: Data Analysis using conditional filtering with more than one condition.

**Question4**: Data Analysis using group the data by column(s)

**Question5**: Data Analysis using sorting the data in ascending/ descending order.

**Question6,7**: Data Analysis using combination of sorting, condition filter and/or grouping.

**Question8,9,10**: Visualize data using a chart with proper headings and legends.

These 10 questions **should be relevant**, by providing useful information that may help in making decisions.

Example of irrelevant questions:

* *How many columns? Show the 6 first rows*, …
* Questions related to reading data, cleaning data are NOT objectives.

Example ofrelevant questions:

* *How many students from Engineering Division got more than 89 marks?*
* *Display the number of students registered by each division using a pie chart.*

1. **Data Acquisition and Cleaning [10 marks]**

Provide the Python code to read the dataset, clean unnecessary data with clear explanation of why cleaning needed for selected data set and the screenshot of the code.

1. **Exploratory Analysis [28 marks]**

Provide the Python code that performs the data analytics for the analysis questions 1 to 7, the usage of conditions, sorting, grouping, and statistical commands, is mandatory. Provide Screenshot of coding with results for each objective with a clear explanation on the data analysis.

1. **Data Visualization [12 marks]**

Provide the Python code that uses different types of Python chart for the analysis questions 8 to 10, with suitable titles and colors to visualize the data for suitable analysis. Provide Screenshot of code and the result that shows the data visualization with a clear explanation on the code.

1. **Executive Summary [4 marks]**

Write an executive summary that provides a clear snapshot of the project and highlights all the key findings that is related to your analysis questions and give a clear recommendation with justifications.

1. **References [2 Marks]**

Include all the external references that you might have used. APA/MLA referencing style may be used.

### Presentation part

**Individual Presentation (30 Marks)**

Present your project’s key findings and recommendations with a well-structured presentation with good visualization to a live panel while explaining the process employed to perform your data analytics.

Your presentation content must have:

Title, Introduction, Main explanations on project, Individual contributions, Conclusion.

**See the Rubric for more details and evaluation**

# Project Marks Distribution

### Report-based ASSESSMENT [70%]

1. **Report in a Word / PDF forma**t using the given template including the Python code and screen shots of the output. [CLO4]

### Presentation-based ASSESSMENT [30%]

1. **Professional demonstration**: PPT Presentation. [CLO4]

# Rubric

## 1. Rubric for Project Report (Total: 70 Marks)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Section | Description | Max Marks | Poor | Good | Very Good | Excellent | Student Mark |
| **Project Introduction** | Clearly defines project aim, dataset source, relevance, and columns used. | 4 | (0-1 marks) Minimal or unclear aim and dataset description. | (2 marks) Basic aim and brief dataset description; partial relevance. | (3 marks) Clear aim, dataset details, and relevance mostly well-described. | (4 marks) Comprehensive and detailed aim, dataset, and relevance description. |  |
| **Analysis Questions** | Defines 10 logical, relevant questions aligned with analysis requirements. Each question is worth 1 mark. | 10 | (0-3 marks) Limited or irrelevant questions with minimal alignment. | (4-6 marks) Some relevant questions; partial alignment or clarity issues. | (7-8 marks) Mostly relevant questions, meeting analysis constraints. | (9-10 marks) Highly relevant, logical questions fully aligned with specified analysis constraints. |  |
| **Data Acquisition & Cleaning** | Data import, inspection, and cleaning with justification and clear results. | 10 | (0-3 marks) Incorrect or minimal cleaning; no justification. | (4-6 marks) Basic data loading with limited cleaning explanation. | (7-8 marks) Mostly correct cleaning with clear explanations. | (9-10 marks) Accurate, well-justified cleaning and documented code. |  |
| **Exploratory Analysis** | Uses Python to analyze data (sorting, filtering, grouping) for each question (7 questions at 4 marks each). | 28 | (0-10 marks) Incorrect code with minimal or unclear explanations. | (11-17 marks) Basic code but limited insights; partial explanations. | (18-24 marks) Mostly accurate code with relevant insights and clear explanations. | (25-28 marks) Accurate, well-documented code with insightful, relevant analysis and interpretations. |  |
| **Data Visualization** | Uses Python charts to visualize analysis questions 8-10 with clear titles, legends, and explanations (3 questions at 4 marks each). | 12 | (0-3 marks) Visuals lack clarity or correctness; minimal explanation. | (4-6 marks) Basic visuals with minor clarity issues, some missing elements. | (7-9 marks) Mostly clear visuals with appropriate titles and explanations. | (10-12 marks) High-quality, informative visuals with correct titles, legends, and well-explained relevance. |  |
| **Executive Summary & Recommendations** | Summarizes project findings, linking recommendations to data analysis. | 4 | (0-1 marks) Vague summary, lacks connection to insights. | (2 marks) Basic summary; limited recommendations with minimal link to analysis. | (3 marks) Clear summary and recommendations mostly based on findings. | (4 marks) Well-written summary with insightful, data-backed recommendations and clear connections to findings. |  |
| **References** | Proper citation of all external sources used in APA/MLA style. | 2 | (0 marks) No references or incorrect style. | (1 mark) Basic list with formatting errors. | (1.5 marks) Mostly correct references, minor formatting issues. | (2 marks) Correctly formatted references, free from errors. |  |
| **Total** |  | 70 |  |  |  |  |  |

## 2. Rubric for Presentation (Total: 30 Marks)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Section | Description | Max Marks | Poor | Good | Very Good | Excellent | Student Mark |
| **Content and Structure** | Focuses on what information is included and how well it is organized. The presentation should include Title, Introduction, Main Explanations, Key Findings, and Conclusion, in a logical flow that’s easy to follow. | 13 | (0-4 marks) Presentation lacks a clear structure. Significant sections are missing, or explanations are hard to understand. | (5-8 marks) Basic structure with some sections missing or with unclear explanations. Presentation needs better organization. | (9-11 marks) Mostly well-structured with clear explanations in each section; minor improvements needed. | (12-13 marks) Thoroughly organized presentation, covering all sections with clear, concise, and logically ordered explanations that are easy to follow. |  |
| **Clarity of Findings** | Focuses on how well key findings are explained and visualized. The presentation should clearly communicate project results using easy-to-understand language and supporting visuals. | 12 | (0-3 marks) Key findings are missing or confusing; visuals are either absent or do not support findings. | (4-6 marks) Findings are present but may lack clarity or have poor visuals that don’t enhance understanding. | (7-9 marks) Findings are mostly clear and are supported by relevant visuals, with minor clarity or quality issues. | (10-12 marks) Findings are well-presented, easy to understand, and supported by high-quality visuals that effectively enhance the explanations. |  |
| **Design and Readability** | Focuses on how visually appealing and readable the presentation is. Consistent formatting, readable fonts, and effective use of colors and visuals are key to making the presentation easy to look at and understand. | 5 | (0-1 marks) Presentation design is unprofessional, hard to read, or lacks formatting; visuals are unclear or inconsistent. | (2-3 marks) Basic design; some readability or formatting issues; visuals may lack clarity or consistency. | (4 marks) Mostly clear, organized design with readable fonts and consistent formatting; visuals support content well. | (5 marks) Excellent design with clear, professional formatting; fonts, colors, and visuals enhance readability and comprehension. |  |
| **Total** |  | 30 |  |  |  |  |  |

**Project Marks Distribution Summary**

|  |  |  |
| --- | --- | --- |
| **Assessment Component** | **Marks** | **Total Marks** |
| **Report-Based** - Project Report including Python code and screenshots | 70 |  |
| **Presentation-Based** - Professional presentation with findings and recommendations | 30 |  |
| **Overall Total** | **100** |  |