

Page **1** of **6**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ASSIGNEMNT BRIEF   |  |  | | --- | --- | | **HTU Course No:** 10204281  **HTU Course Name:** Data Science Programming | **BTEC Unit No:**    **BTEC Unit Name:** | | | |
| Assignment Brief Number: 1 | |  |
| Version: 1 |  | |

**Assessment Brief**

|  |  |
| --- | --- |
| **Student Name/ID Number/Section** |  |
| **HTU Course Number and Title** | **10204281 Data Science Programming** |
| **Academic Year** | Fall 2022 |
| **Assignment Author** | Raneem Qaddoura |
| **Unit Tutor** | Raneem Qaddoura  Sinan Kamal  Bassam Alkasasbeh |
| **Assignment Title** | **Data preprocessing, modeling, and reporting** |
| **Assignment Ref No.** | 1 |
| **Issue Date** | 11/12/2022 |
| **Formative Assessment Dates:** | Every week until 12/1/2023 |
| **Submission Date** | 1/2/2023 |
| **IV Name & Date** | Murad Yaghi 7/12/2022 |

|  |
| --- |
| **Submission Format** |
| The assignment should be submitted to the university’s eLearning system within the deadline specified above from the link: <https://elearning.htu.edu.jo>. The assignment is in the form of:   * A technical document: it must follow the below guidelines, and it should contain the sections described in the assignment brief. In your documents, you should make use of headings, paragraphs, and subsections as appropriate. The expected word limit is 2000-5000 words, although you will not be penalized for exceeding the total word limit, do your best to be within the word limit. Your report should be:   + In a form of **WORD document** soft copy submitted to the university’s eLearning system.   + Written in a formal business style using single spacing and font size 12, of times roman.   + Must be supported with research and referenced using the Harvard referencing system. * The source code file (ipynb). * Signed declaration Form. |

|  |
| --- |
| **Unit Learning Outcomes** |
| **LO1** Understanding the basic concepts of a common programming language and data structures  **LO2** Understanding the common libraries used in the field of data science. How to load data and prepare data for analysis, data cleaning/preprocessing.  **LO3** Introduction to modeling and model evaluation.  **LO4** Produce a technical document to report analysis results and communicate effectively through visualization. |
| **Assignment Brief and Guidance** |
| **Scenario:**  You have a position as Junior Data Scientist for a telecom company with systems that the IT department implements. These systems have generated a massive amount of data, and the company wants to extend its work to include data science applications.  The company wants to know the possibilities of applying data science to its applications and wants to start with the complaints data provided in this [dataset](https://hatuniversity.sharepoint.com/:f:/s/DSclassespreparation/EsQsRoHFhfFIsuIEqMJ9zEsB0YPTYUFhTaNu-U3QF2jZ4w). The company also wants to evaluate your ability to do the work. You are assigned to give a proof of concept for the needed application by generating several models and evaluating the different models. You then need to communicate your work to the management by producing a comprehensive technical document to show some basic technical concepts and the results obtained from work.  The dataset contains the complaints that are received by the complaint system. The features of the dataset include the offer name, customer type, open date, close date, etc. The labels are the COMPLAINT\_TYPE, CASE, and PRODUCT. You can choose any of the labels to generate the model as a proof of concept for the company.  **Part 1:**   1. Write a program by installing and working with a popular programming language and tool used in data science. 2. The program should demonstrate your ability to prepare and load data for analysis 3. The program should show the different generated models using the appropriate libraries. 4. Make use, in your program, of the built-in data types, data structures, conditional programming, and looping. 5. Evaluate the models using different evaluation measures. 6. Support the results obtained with appropriate charts to visualize them.   **Part 2:**   1. Produce a comprehensive technical document that compares and contrasts the different types of data structures used in data science, and explain, with examples, the common libraries used in data science. The document should outline the different plotting and visualization libraries used in data science. 2. The document should include an explanation of the experiments conducted and the results obtained with appropriate charts to visualize them. The document should compare the different models applied using different evaluation measures. 3. The document should evaluate the choice of data structures used in the implementation of your program, the selection of the appropriate libraries and data preprocessing steps, and the effectiveness of different models applied. 4. You should communicate an analysis of the results and make valid, justified recommendations in the document. |

|  |  |  |
| --- | --- | --- |
| **Learning Outcomes and Assessment Criteria** | | |
| **Pass** | **Merit** | **Distinction** |
| **LO1 Understanding the basic concepts of a common programming language and data structures.** | | **D1** Evaluate the choice of data structures used in the implementation of a program. |
| **P1** Able to install and work with a popular programming language and tool used in data science.  **P2** Write programs with built-in data types, data structures, conditional programming, and looping. | **M1** Compare and contrast the different types of data structures used in data science. |  |
| **LO2 Understanding the common libraries used in the field of data science. How to load data and prepare data for analysis, data cleaning/preprocessing.** | | **D2** Critically evaluate the effectiveness of different models applied. |
| **P3** Explain, with examples, the common libraries used in data science.  **P4** Demonstrate an ability to prepare and load data for analysis. | **M2** Evaluate the selection of appropriate libraries and data preprocessing steps. |
| **LO3 Introduction to modeling and model evaluation.** | |
| **P5** Write a program that generates different models using the appropriate libraries. | **M3** Compare the different models applied using different evaluation measures. |
| **LO4 Produce a technical document to report analysis results and communicate effectively through visualization.** | | **D3** Communicate an analysis of the results and make valid, justified recommendations. |
| **P6** Outline the different plotting and visualization libraries used in data science. | **M4** Produce a comprehensive report which includes an explanation of the experiments conducted and the results obtained with appropriate charts to visualize them. |

# STUDENT ASSESSMENT SUBMISSION AND DECLARATION

When submitting evidence for assessment, each student must sign a declaration confirming that the work is their own.

|  |  |  |  |
| --- | --- | --- | --- |
| **Student name:** | | **Assessor name:** | |
| **Issue date:** 11/12/2022 | **Submission date:** 1/2/2023 | | **Submitted on:** |
| **Program:** Computing | | | |
| **Course Name:** Data Science Programming  **HTU Course Code:** 10204281 **BTEC UNIT:** | | | |
| **Assignment number and title:** Data preprocessing, modeling, and reporting | | | |

# Plagiarism

Plagiarism is a particular form of cheating. Plagiarism must be avoided at all costs and students who break the rules, however innocently, may be penalized. It is your responsibility to ensure that you understand correct referencing practices. As a university level student, you are expected to use appropriate references throughout and keep carefully detailed notes of all your sources of materials for material you have used in your work, including any material downloaded from the Internet. Please consult the relevant unit lecturer or your course tutor if you need any further advice.

**Student declaration**

I certify that the assignment submission is entirely my own work and I fully understand the consequences of plagiarism. I understand that making a false declaration is a form of malpractice.

**Student signature:**

**Date:**