جامعة الحسين التقنية
Al Hussein Technical University

ASSIGNEMNT BRIEF

HTU Course No: 10204280 HTU Course Name: Principles of Data Science and Computing Systems	BTEC Unit No: BTEC Unit Name:

Assignment Brief Number: 2

Version: 1

جامعة الحسين التقنية
Al Hussein Technical University

Assessment Brief

Student Name/ID Number/Section	
HTU Course Number and Title	10204280 Principles of Data Science and Computing Systems
BTEC Course Number and Title	
Academic Year	Fall - 2022/2023
Assignment Author	Dr. Murad A. Yaghi
Unit Tutor	Dr. Murad A. Yaghi Dr. Heba Alawneh Bassam Kasasbeh
Assignment Title	Application of Data Science Life Cycle
Assignment Ref No.	2
Issue Date	08-01-2023
Formative Assessment Dates:	Every week until 25/01/2023
Submission Date	30/01/2023
IV Name & Date	Dr. Raneem Qaddoura 07/01/2023

Submission Format

You are expected to complete and submit your work according to the following guidelines:

Deliverables:

- A report in document (.docx) or pdf format.
- The project implementation in the form of python script (.py) or Jupyter notebook (.ipynb)
- The declaration form attached to assignment (filled and signed).

Formative Presentation:

- Weekly or biweekly presentations to present your progress

Oral:

- The oral assessment will be on Campus, with the instructor and according to the schedule specified by the instructor
- Attending the oral assessment at the determined date and time is mandatory.
- Any unanswered question in the oral assessment for any criteria results in losing that specific criterion, even if it was answered correctly in the report

Report and Code Guidelines:

- Your submissions should be in the form of a soft copy via the eLearning school system.
- Your work should be:
  - (a) Written in professional style format
  - (a) Include a report cover page: Tour name, Assignment Title, Organization name, date
  - (b) Your report must be supported with references using the Harvard reference system.
  - (c) If the percentage of quoted sources in your report is more than 15%, you shall fail the course.
  - (d) Any plagiarism, even if it is 1%, shall result in failing the course.

جامعة الحسين التقنية
Al Hussein Technical University

Unit Learning Outcomes

LO1: Demonstrate knowledge of key concepts, principles, tools and libraries and overarching themes in data science and the data science lifecycle under a computing system environment.

LO3: Acquire, cleanse, manipulate, preprocess, and analyze diverse datasets using the latest technologies, platforms, and techniques to gain deeper insights leading to innovative solutions.

LO4: Develop the ability to build and assess data-based models, and differentiate between learning models (supervised and unsupervised) and be able to communicate results efficiently

Assignment Brief and Guidance

Scenario:

You are a data scientist in a financial company's risk management department. The supervisor asked you to analyze the data related to a food trading company. The data collected by the company describes the various products that were sold last month at different stores within the country. This dataset contains information about the products, including product features and selling store features. The product features included product number, weight, retail prices, and others. The store features included the store size and location and others. You are required to clean, analyze, perform appropriate predictions, and present your results to customers who don't have any technical experience in this field.

Implementation Requirements:

- Given the provided dataset, develop two prediction models. The first prediction model uses Linear Regression to predict the total product sales. The second task is to use KNN (K-Nearest Neighbors) to predict the store from which the product was bought.
- Represent the performance of the prediction models using appropriate evaluation measures and visual representation.
- Use different optimization techniques to improve the performance of your prediction models.

Prepare a report that provides detailed insight into the following:

- Explain the main differences between supervised and unsupervised learning.
- Explain the main differences between regression and classification techniques.
- Discuss and explain the data science life cycle of your developed prediction model (any of them).
- List the data preprocessing techniques utilized for building prediction models.
- Discuss the use of each preprocessing technique utilized for building prediction models.
- Critically evaluate and explain the impact of each data processing and data cleaning technique on the performance of the machine learning prediction model and compare the results with and without applying these preprocessing techniques. Support your statements with performance evaluation and comparison.
- Investigate the computing systems used in cloud computing services (such as AWS and Azure), and compare them in terms of structure, GPU, CPU, type of storage devices used, and other related hardware aspects.

Learning Outcomes and Assessment Criteria		
Pass	Merit	Distinction
LO1 Demonstrate knowledge of key concepts, principles, tools and libraries and overarching themes in data science and the data science lifecycle under a computing system environment.		
P1: Illustrate the data science lifecycle and explain key concepts in data science and their applications in real life **P2**: Investigate the computing systems utilized in data science projects	**M1**: Assess the different libraries and frameworks used in data science projects	**D1**: Evaluate the different sources of data and their types
LO3: Acquire, cleanse, manipulate, preprocess, and analyze diverse datasets using the latest technologies, platforms, and techniques to gain deeper insights leading to innovative solutions		**D4**: Critically evaluate and explain the insights derived from analysis and communicate insights
P5: Illustrate the importance of data cleaning and data pre-processing.	**M3**: Assess the latest technologies and platforms to preprocess and analyze data.	
LO4: Develop the ability to build and assess data-based models, and differentiate between learning models (supervised and unsupervised) and be able to communicate results efficiently		**D5**: Critically evaluate and explain the insights derived from analysis and communicate insights derived from a data science project efficiently and clearly using appropriate tools and techniques
P6: Design a data science project and report results based on evaluation measures	**M4**: Differentiate between supervised and unsupervised learning and their applications.	

جامعة الحسين التقنية
Al Hussein Technical University

STUDENT ASSESSMENT
SUBMISSION ANDDECLARATION

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

Student name:		Assessor name:
Issue date: 08-01-2023	Submission date: 30-01-2023	Submitted on:
Program: **Computing**		
Course Name: Principles of Data Science and Computing Systems HTU Course Code: 10204280 BTEC UNIT:		
Assignment number and title: 2 / Application of Data Science Life Cycle		

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 thatyou 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 theconsequences of plagiarism. I understand that making a false declaration is a form of malpractice.

Student **Date**