INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA

FINAL ASSESSMENT – TEST 1
SEMESTER 2, 2022/2023 SESSION

KULLIYYAH OF INFORMATION AND COMMUNICATION TECHNOLOGY

Programme	: **BCS**	Level of Study	:	**Undergraduate**
Time	: **03.30 PM (17th May 2023) - 06.00 PM (19th May 2023)**	Date	:	**17th MAY 2023 – 19th MAY 2023**
Duration	: **47** Hours **30** Minutes			
		Section(s)	:	**1**
Course Code	: **CSCI 4341**			
Course Title	: **BIG DATA ANALYTICS**			

<u>INSTRUCTION(S) TO CANDIDATES</u>

This question paper consists of 4 printed pages including the cover page.

All the required details have been provided inside this booklet.

IMPORTANT:

DO NOT OPEN UNTIL YOU ARE ASKED TO DO SO

Any form of cheating or attempt to cheat is a serious offence which may lead to dismissal.

<u>APPROVED BY</u>

A. TEST QUESTION

In a typical data science environment, the first step involves working with raw data and making sense of it. However, different data scientists may draw various conclusions from the same dataset. It is your responsibility to determine the relevant business questions and transform them into data science inquiries. To accomplish this, thorough data exploration is necessary, often referred to as exploratory data analysis. This entails examining the data and its characteristics, utilizing statistics to extract features and test significant variables, employing data visualization to identify patterns and trends, and so on. The final and crucial step is interpretation, where the analytical outcomes should be understandable to a non-expert. The interpretations should address the data science questions formulated at the outset and lead to actionable insights.

Between exploration and interpretation, there exists an intermediate step: modeling the data. This step aids in creating prediction models and understanding certain grouping phenomena, among other things.

This take-home test provides you with a dataset/s pertaining to housing data containing information about tenure, housing costs, median income, fair rent, poverty levels etc. The **dataset** folder contains the following files:

1. housdata.csv
2. housdict.pdf

This is an **open-ended, exploratory and hands on test**. Your task is to explore the dataset, as per the guidelines below:

- You **must** use **Python** for the analysis.
- Methods are as per your choice & convenience.

You **MUST** prepare and submit your results, analysis & conclusions, etc. using a power point presentation along with the relevant codes or any other material. Documentation within the code is **very important.**

Since this is an open-ended test, you can choose to include any material in your presentation that you deem fit. Below are some of the sections that you might want to include:

- Overall Summary
- Background
- Data Science Questions
- Objectives
- Methodology
- Results
- Visualizations
- Any other discussion
- References (APA Format)

B. INSTRUCTIONS:

- Name & Matric Number must be included in all your submissions

- Submit your files by uploading them as following:

    o **iTaleem**: Only the power point presentation

    o **Google Classroom**: All files

- You must upload **actual** files to iTaleem & Google Classroom. Links to your files will **NOT** be accepted, like link to your Colab notebook, google slides, etc. are not acceptable.

- This is an **open ended**, **exploratory & hands on test**. You are free to refer to any online/offline resources as long as you don't plagiarize and refer to your sources.

- Referring to your sources is of utmost importance and not citing them will be deemed as plagiarism.

- Any form of plagiarism is highly unacceptable.

C. ASSESSMENT:

- **Test 1 bears 20 Marks (20% of your Total Course Evaluation).**

- **If you don't submit your "codes and/or other relevant material", your presentation will not be evaluated.**

- The marks are distributed as follows:

Presentation	15 Marks
Codes AND/OR Other Relevant Items	5 Marks

D. IMPORTANT DATES/DEADLINES:

STARTS: 17th May 2023, 03:30 PM

ENDS: 19th May 2023, 06:00 PM

END OF DOCUMENT