CCT College Dublin Continuous Assessment

Programme Title:	Statistical techniques for Data Analysis		
Cohort:	HDip in Science in Data Analytics for Business (FT)		
Module Title(s):	Statistical Techniques for Data Analysis		
Assignment Type:	Pairs	**Weighting(s)**:	50%
Assignment Title:	Dataset Analysis		
Lecturer(s):	Aldana Louzan		
Issue Date:	24/10/2022		
Submission Deadline Date:	18/12/2022		
Late Submission Penalty:	Late submissions will be accepted up to **5** calendar days after the deadline (23/12/2022). All late submissions are subject to a penalty of **10%** of the mark awarded. Submissions received more than 5 calendar days after the deadline above **will not** be accepted and a mark of 0% will be awarded.		
Method of Submission:	**Moodle**		
Instructions for Submission:	IPYNB file (Jupyter Notebook) & dataset		
Feedback Method:	**Results posted in Moodle gradebook**		
Feedback Date:			

Learning Outcomes:

Please note this is not the assessment task. The task to be completed is detailed on the next page.
This CA will assess student attainment of the following minimum intended learning outcomes:

1. MIMLO 1: Explore and evaluate datasets using descriptive statistical analyses. (Linked to PLO 1)
2. MIMLO 4: Use and understand current software tools and languages to produce result sets from existing data(e.g. Excel, R, Python). (Linked to PLO 4)

Attainment of the learning outcomes is the minimum requirement to achieve a Pass mark (40%). Higher marks are awarded where there is evidence of achievement beyond this, in accordance with QQI *Assessment and Standards, Revised 2013*, and summarised in the following table:

<table>
<tr><th rowspan="2">Percentage Range</th><th rowspan="2">CCT Performance Description</th><th colspan="2">QQI Description of Attainment</th></tr>
<tr><th>Level 6, 7 & 8 awards</th><th>Level 9 awards</th></tr>
<tr><td>90% +</td><td>Exceptional</td><td rowspan="3">Achievement includes that required for a Pass and in most respects is significantly and consistently beyond this</td><td rowspan="3">Achievement includes that required for a Pass and in most respects is significantly and consistently beyond this</td></tr>
<tr><td>80 – 89%</td><td>Outstanding</td></tr>
<tr><td>70 – 79%</td><td>Excellent</td></tr>
<tr><td>60 – 69%</td><td>Very Good</td><td>Achievement includes that required for a Pass and in many respects is significantly beyond this</td><td>Achievement includes that required for a Pass and in many respects is significantly beyond this</td></tr>
<tr><td>50 – 59%</td><td>Good</td><td>Achievement includes that required for a Pass and in some respects is significantly beyond this</td><td rowspan="2">Attains all the minimum intended programme learning outcomes</td></tr>
<tr><td>40 – 49%</td><td>Acceptable</td><td>Attains all the minimum intended programme learning outcomes</td></tr>
<tr><td>35 – 39%</td><td>Fail</td><td>Nearly (but not quite) attains the relevant minimum intended learning outcomes</td><td>Nearly (but not quite) attains the relevant minimum intended learning outcomes</td></tr>
<tr><td>0 – 34%</td><td>Fail</td><td>Does not attain some or all of the minimum intended learning outcomes</td><td>Does not attain some or all of the minimum intended learning outcomes</td></tr>
</table>

Please review the CCT Grade Descriptor available on the module Moodle page for a detailed description of the standard of work required for each grade band.

The grading system in CCT is the QQI percentage grading system and is in common use in higher education institutions in Ireland. The pass mark and thresholds for different grade bands may be different from what you have experience of in the higher education system in other countries. CCT grades must be considered in the context of the grading system in Irish higher education and not assumed to represent the same standard the percentage grade reflects when awarded in an international context.

Assessment Task

Students are advised to review and adhere to the submission requirements documented after the assessment task.

Assessment Overview:

This assignment consists of **3 sections**:

- **First section:** Dataset selection & Descriptive stats
- **Second section:** Hypothesis testing
- **Third section:** Correlation Analysis & Linear Regression

You **must** create **a notebook in Jupyter Notebooks (using Google Colab)** to complete the following tasks using your chosen dataset. Your dataset **must** contain quantitative and qualitative data, a minimum of 8 variables and around 10k entries. All your code must be commented.

First section: Dataset & Descriptive Stats (30 marks)

1. Pick a data set of your interest and upload it to your notebook. Add a brief about that data set, including the reason you picked this dataset and the data dictionary.
2. Using appropriate Python code show relevant information about the data set selected.
3. Pick **two different types** of variables (one qualitative and one quantitative) and describe them using appropriate statistical measures.
4. Using appropriate graphics, plot the two variables picked in 3 to get a better idea of its dispersion.
5. Provide a short description about the distribution of the variables picked in 3. Support your analysis with the descriptive statistics already calculated, and with the graphics that you plot in the previous item.
6. Do the variables picked in 3 have any missing values? Do they have any outliers? Justify your answer.

Second section: Hypothesis Testing (20 marks)

Choose **one variable** from your chosen dataset and perform a **Hypothesis Test.** All the steps **must** be supported by appropriate references, statistical concepts, and calculations.

You **must** interpret your results, provide your **own analysis and conclusion** based on your Hypothesis Test.

You will need to conduct **research** to find the parameters of the population you want to analyse.

Some examples:

- ✓ It is believed that the average age to finish High School is 17 years old. Based on this parameter, then you perform a HT to reject or not this hypothesis.
- ✓ It is believed that due to COVID and remote working, workers save 8.5hours on average, then you perform a HT to reject or not this hypothesis.

Third section:

Correlation Analysis (20 marks)

Carry out a **correlation analysis** between **2 variables**. Interpret your results and check if the **correlation implies causation**.

Provide a short **explanation and conclusion** based on your findings.

Linear Regression Model (20 marks)

Using the **same 2 variables** in the second section, build a **linear regression model** that allows you to predict information about those variables.

Interpret your results, and provide a short **explanation and conclusion** based on your findings.

Report (10 marks)

All your answers must be in form of a formal **report in an IPYNB file**, using **Markdown Language** to format your report.

The report must include an **introduction** to the selected topic and a **conclusion** at the end of the report that includes the insights found about the dataset selected

There are some sections that requires conducting some **research,** make sure to use appropriate external references.

All the sections must be answered using **appropriate statistical terms and calculation**.

You **must** add a **References list** at the end of your CA following the **Harvard Referencing Style**.

If you have used a **local dataset** to complete the third section, you **must submit** it as well.

Submission Requirements

All assessment submissions must meet the minimum requirements listed below. Failure to do so may have implications for the mark awarded.

All assessment submissions must:

- You must submit a IPYNB file
- You must use Markdown Language to format your CA, following a report style
- Be submitted by the deadline date specified or be subject to late submission penalties
- Be submitted via Moodle upload, (*faculty to specify submission method*)
- Use Harvard Referencing when citing third party material
- Be the student's own work.
- Include the CCT assessment cover page.

Additional Information

- Lecturers are not required to review draft assessment submissions. This may be offered at the lecturer's discretion.
- In accordance with CCT policy, feedback to learners may be provided in written, audio or video format and can be provided as individual learner feedback, small group feedback or whole class feedback.
- Results and feedback will only be issued when assessments have been marked and moderated / reviewed by a second examiner.
- Additional feedback may be provided as individual, small group or whole class feedback. Lecturers are not obliged to respond to email requests for additional feedback where this is not the specified process or to respond to further requests for feedback following the additional feedback.
- Following receipt of feedback, where a student believes there has been an error in the marks or feedback received, they should avail of the recheck and review process and should not attempt to get a revised mark / feedback by directly approaching the lecturer. Lecturers are not authorised to amend published marks outside of the recheck and review process or the Board of Examiners process.
- Students are advised that disagreement with an academic judgement is not grounds for review.
- For additional support with academic writing and referencing students are advised to contact the CCT Library Service or access the CCT Learning Space.
- For additional support with subject matter content students are advised to contact the CCT Student Mentoring Academy
- For additional support with IT subject content, students are advised to access the CCT Support Hub.