A picture containing icon

Description automatically generated

**CCT College Dublin Continuous Assessment**

|  |  |  |  |
| --- | --- | --- | --- |
| **Programme Title:** | MSc in Data Analytics | | |
| **Delivery Mode:** | FT/SB+ | | |
| **Cohort Details:** | *MSc in Data Analytics (Sept24 start) Stage 1 Sem 1* | | |
| **Module Title(s)**: | *Data Prep & Vis*  *Programme schedules are all published on the* [*CCT IQR Provider Profile*](https://irq.ie/providers/cct-college-dublin?id=fec9ea7a-ace4-42c7-9fd5-7fccb6f0a53a&ref=%257B%2522search%2522:%2522cct%2522%257D) | | |
| **Assignment Type:** | *Individual* | **Weighting(s):** | *Data Prep & Vis 100%*  *Capped at 40%* |
| **Assignment Title:** | *MSC\_DA\_CA2* | | |
| **Lecturer(s)**: | *David McQuaid* | | |
| **Issue Date:** | April 5th, 2025 | | |
| **Submission Deadline Date:** | May 5th, 2025 | | |
| **Late Submission Penalty:** | Late submissions will be accepted up to **5** calendar days after the deadline. 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**  **Use the submission link on the Data Visualisation and Preparation Module page** | | |
| **Instructions for Submission:** | *Please do not ZIP your files. ALL files must be uploaded individually (to a maximum of 20 files)*  *Expected files : Written report (word document only, NO PDF’s) ,Code files (Jupyter notebook (.ipynb) ONLY, NO PYTHON FILES), Data Files*  *Note that the maximum number of Jupyter Notebooks is 4* | | |
| **Feedback Method:** | **Results posted in Moodle gradebook** | | |
| **Feedback Date:** | *After exam board Jun 2025* | | |

### Assessment Outline

### Description of Assessment Task

## Note: This is an academic exercise and not a hypothetical report and the most important aspect of this report is evaluating and rationalizing your decisions in the domain of Data Analytics NOT the problem domain.

## All Project files MUST be uploaded into MOODLE, this is your responsibility, if any files are not uploaded to MOODLE, even if they are available on GITHUB, they will NOT BE GRADED.

## Criteria of Analysis (ALL EXPERIMENTAL WORK MUST BE CARRIED OUT USING PYTHON IN JUPYTER NOTEBOOK)

***Scenario***

*“Like it or not, every construction company—and solutions provider—is now also in the data business. How well we help our customers transform that data into intelligence that drives better decisions to deliver projects more efficiently and more sustainably, with higher quality, lower costs and fewer risks is what defines the next frontier of construction management. Data is the key to improving the bottom line as well as protecting it. Our ability to break down data silos and transform raw data into action and intelligence is the crux to solving most challenges that rear their head in our industry. Solve the data problem and everything else falls into place.”—Jon Fingland, General Manager, Collaboration Solutions, Trimble*

You have been tasked with analysing Ireland's Construction data and comparing the Irish Construction sector with other countries worldwide. Your Research could include export, import, trade imbalance, house production, material stock, labour/skill pool, etc. (or any other relevant topic EXCEPT Climate change) with Ireland as your base line.

**Note:**

**While topical, Construction impact on Climate Change SHOULD NOT be chosen as an area of research for this assessment.**

You must source appropriate data sets from any available repository to inform your research (all datasets MUST be referenced, and the relevant licence/permissions detailed).

***Criteria of Analysis***

It is Required that you use GitHub Classroom as your version control repository etc with regular commits of code and report versions. You may be called to a Viva to defend your work.

Please find the GitHub Classroom link below:

<https://classroom.github.com/a/4odn-Ai9>

**Data Preparation & Visualisation Tasks**

**You Must Conduct an analysis of your Data based on the problem domain and data you have chosen following which you must compile a report (2000 words +/- 10 % (not including code, code comments, titles, references, or citations)** **) based on the following criteria**

* Discuss in detail the process of acquiring your raw data, detailing the positive and/or negative aspects of your research and acquisition. This should include the relevance and implications of any and all licensing/permissions associated with the data (This will require research outside of class material). **[0-15]**
* Exploratory Data Analysis helps to identify patterns, inconsistencies, anomalies, missing data, and other attributes and issues in data sets so problems can be addressed. Evaluate your raw data and detail, in depth, the various attributes and issues that you find. Your evaluation should reference evidence to support your chosen methodology and use visualizations to illustrate your findings.**[0-25]**
* Taking into consideration the tasks required in the machine learning section, use appropriate data cleaning, engineering, extraction and/or other techniques to structure and enrich your data. Rationalize your decisions and implementation, including evidence of how your process has addressed the problems identified in the EDA (Exploratory Data Analysis) stage and how your structured data will assist in the analysis stage. This should include visualizations to illustrate your work and evidence to support your methodology.**[0-30**]
* Modern construction has a great dependence on technology and relies upon visualizations to communicate information, this includes web based, mobile based and many other digital transmission formats. Develop an interactive dashboard tailored to modern construction professionals, using tufts principles, to showcase the information/evidence gathered following your Analysis. Detail the rationale for approach and visualisation choices made during development making reference to Tufts Principles. **Note you may not use Powerbi, RapidMiner, tableau or other such tools to accomplish this (at this stage).[0-30]**

**Total Mark = 15+25+30+30=100:(100%)**

All assessment submissions must meet the following minimum requirements:

* Be submitted in the format outlined in the assignment summary table.
* 2000 (+/- 10%) words in report (not including code, code comments, titles, references, or citations)
* Report submission MUST be a word document only (No PDF’s!).
* Code in a Jupyter Notebook file (.ipynb) only but may be referenced in the word document.
* Be submitted by the deadline date specified or be subject to late submission penalties.
* Be submitted via Moodle upload
* Use [Harvard Referencing](http://40.115.124.2/sp/subjects/guide.php?subject=harvardref) when citing third party material.
* Be the student’s own work.
* Include the CCT assessment cover page.

## Learning Outcomes:

This assessment addresses the following module learning outcomes for this module:

**Data Preparation & Visualisation**

1. Programmatically Implement graphical methods to identify issues within a data set (missing, out of range, dirty data)(linked to PLO 3, PLO 5)

2. Propose, design, develop, and implement an interactive data visualisation solution, for a given data set and potential audience, detailing the rationale for approach and visualisation choices made during development for a given use case, data characteristics and multiple transmission media (linked to PLO 2, PLO 5)

3. Perform a critical analysis of a data set to optimise the data for a given problem space. Document the rationale behind the decisions to peers and stakeholders.(linked to PLO 5, PLO 6)

4. Collaboratively perform a critical analysis of a data set to optimise the data for a given problem space. Document the rationale behind the group’s decisions to peers and stakeholders.(linked to PLO 5, PLO 6)

5. Engineer new features selection in data with the goal of improving the performance of machine learning models. (linked to PLO 2, PLO 4)Statement of Acceptable Use of Artificial Intelligence

|  |
| --- |
| **Acceptable and Unacceptable Use of AI**  *This statement is useful when you are allowing the use of AI tools for certain purposes, but not for others.* |
| * The use of generative AI tools (e.g. ChatGPT, Dall-e, etc.) is permitted in this assignment for the following activities:   + Brainstorming and refining your ideas;   + Fine tuning your research questions;   + Finding information on your topic;   + Drafting an outline to organise your thoughts; and   + Checking grammar and style. * The use of generative AI tools is not permitted in this course for the following activities:   + Impersonating you in classroom context   + Generating code for your assignment   + Writing a draft of a writing assignment   + Writing entire sentences, code, paragraphs or papers to complete class assignments. * You are responsible for the information you submit based on an AI query. Your use of AI tools must be properly documented and cited. * Any assignment that is found to have used generative AI tools in an unauthorised way will be subject to college disciplinary procedures as outlined in the QA Manual. * When in doubt about permitted usage, please ask for clarification. |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criteria** | ***Criteria 1***  *Discuss the process of acquiring raw data, detailing its positive/negative aspects and licensing/permissions implications..* | ***Criteria 2***  *Evaluate raw data, identify attributes/issues, support methodology with evidence, and use visualizations to illustrate findings.* | ***Criteria 3***  *Address EDA issues using appropriate techniques, rationalize decisions, and demonstrate how structured data aids analysis. Include visualizations and evidence.* | ***Criteria 4***  *Develop an interactive dashboard aligned with Tuft’s principles, clearly explain visualization choices, and address modern farmers' needs.* |
| **Weighting per criteria** | **15 marks** | **25marks** | **30 marks** | **30 marks** |
| *Excellent (+70%)* | Comprehensive and insightful discussion of data acquisition. Explains all positives and negatives, relevance, and licensing/permissions with evidence of thorough external research. | Thorough evaluation of raw data with well-documented attributes/issues. Strong rationale for methodology supported by evidence, using clear and insightful visualizations. | Excellent application of cleaning/engineering techniques. Decisions are well-justified and clearly linked to EDA findings. Strong use of visualizations and evidence to demonstrate process and outcomes. | Fully functional, interactive dashboard tailored to farmers. Choices are expertly justified using Tuft’s principles, with clear and effective visualizations. |
| *Very Good (60 - 69%)* | Clear and detailed discussion of data acquisition. Covers most aspects, including positives, negatives, and licensing implications, supported by moderate research. | Detailed evaluation of raw data, identifying most attributes/issues. Good methodological rationale, supported by relevant visualizations. | Solid application of techniques with clear rationale. Most decisions are linked to EDA findings and supported by appropriate visualizations. | Well-designed dashboard with good interactivity and relevance. Most choices are justified using Tuft’s principles, and visualizations are clear and effective. |
| *Good (50 - 59%)* | Adequate discussion of data acquisition with some gaps in detail or research. Covers positives, negatives, and licensing but lacks depth or clarity in some areas. | Adequate evaluation of raw data, addressing some attributes/issues. Methodology referenced but with limited supporting evidence or visualizations. | Adequate cleaning/engineering with partial justification. Some linkage to EDA findings but may lack clarity or depth. Visualizations are present but could be more insightful. | Adequate dashboard with some interactivity. Justifications reference Tuft’s principles but lack depth. Visualizations are functional but could be improved. |
| *Acceptable (40 - 49%)* | Basic discussion of data acquisition with minimal detail. Mentions positives or negatives superficially, with limited or no reference to licensing implications. | Basic evaluation of raw data with few identified attributes/issues. Minimal or unclear rationale for methodology, with weak or missing visualizations. | Limited application of techniques with minimal justification. Weak or unclear connection to EDA findings. Visualizations may be poorly executed or missing. | Basic dashboard with limited interactivity or relevance. Minimal justification or unclear application of Tuft’s principles. Visualizations are underwhelming. |
| *Fail (> 39%)* | Minimal or incomplete discussion. Little or no reference to positives, negatives, or licensing, with no external research evident. | Minimal or incomplete evaluation of raw data. Lacks clear methodology, supporting evidence, or visualizations. | Minimal or incomplete application of techniques. No clear rationale or connection to EDA. Lacks meaningful visualizations. | Incomplete or poorly designed dashboard. No justification of visualization choices or reference to Tuft’s principles. |

## Grading Criteria Data Preparation & Visualisation

**The Irish Grading System**

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 experienced 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.

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 and review the marking criteria outlined in this assignment brief for a breakdown of the marking criteria for this specific assignment.

**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.
* For additional support with subject matter content students are advised to contact the [CCT Student Mentoring Academy](https://moodle.cct.ie/course/view.php?id=827)
* For additional support with IT subject content, students are advised to access the [CCT Support Hub](https://moodle.cct.ie/course/view.php?id=1861).