**CCT College Dublin**

**Assessment Cover Page**

*To be provided separately as a word doc for students to include with every submission*

|  |  |
| --- | --- |
| **Module Title:** | Programming for DA  Statistics for Data Analytics  Machine Learning for Data Analysis  Data Preparation & Visualisation |
| **Assessment Title:** | MSC\_DA\_CA1 |
| **Lecturer Name:** | Sam Weiss/ David Gonzalez  Bharathi Chakravarthi/ Marina Iantorno  Muhammad Iqbal  David McQuaid |
| **Student Full Name:** | Aaron Bowman |
| **Student Number:** |  |
| **Assessment Due Date:** | 10/11/2023 |
| **Date of Submission:** |  |

**Declaration**

|  |
| --- |
| By submitting this assessment, I confirm that I have read the CCT policy on Academic Misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source. I declare it to be my own work and that all material from third parties has been appropriately referenced. I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution. |

# Abstract

*There a*

# Introduction

There are a variety of Income levels across the various Countries in the Republic of Ireland. Throughout the course of this analysis I aim to investigate the relationship between Income Levels without a County, said Counties population and the number of people working in major Industry groups within those group to determine which of these has the largest effect on predicting the income of households.

# Data Gathering and Exploration

## Data Gathering

All data for this research was gathered from the Central Statistics Office (CSO) website. All data contains numerous features. But most important, a year in which it was collected, and a County & State feature describing the geographic area that the data refers to.

Additional each of the of the datasets included at least one feature that I aimed to use for analysis.

## Exploratory Data Analysis

## Data Cleaning

## Data Visualisation

# Statistical Analysis

## Descriptive Statistics

## Distributions

## Normal Distributions

# Machine Learning

## Outline

## SEMMA Framework

## Regressions Analysis

## Clustering Analysis

## Analysis

In general, use a 12-point Times New Roman font, or other Roman font with serifs.

### Title and Authors

The title (Arial 14-point bold) runs across the full width of the page and is centred. Authors' names (Arial 12 point not-bold) and affiliations (Arial 12-point not-bold) are entered into the table at the top. We also recommend you add your postal address and e-mail address using the same style as for authors.

### Abstract and Keywords

Every submission should begin with an abstract of about 100 words in the normal text style but italicized. The abstract should be a concise statement of the problem, approach, findings, and conclusions of the work described. Keywords will be taken from your submission form and added when the publication is assembled.

### References and Citations

In the body of the text, cite the references like so (author, year).Use the standard Harvard format for listing references at the end of your report. You can refer to the reference list at the end of this document as an example.

### Page Numbers and Footnotes

Make sure you number pages unless explicitly asked not to. (For example, many journals request that page numbers are *not* used in submission for publication.) In general, avoid the use of footnotes. Stick to using references and citations.

## Sections

Breaking your report into sections can make it much easier to read. Main sections (Introduction, Materials and Methods, Results and Discussion/Conclusions should generally be in Arial 12-point bold title-case or follow specific instructions given with your real assignments.The use of screen shots or diagrams to enhance your explanations is encouraged.

### Subsections

Arial 12-point title-case.

#### Sub-subsection headings

Arial 12-point italic.

## Figures

Use Normal Times New Roman 12-point font for figure captions.

## Writing Style

* Use a simple writing style – if five words convey your meaning don’t use ten.
* Make use of headings and sub-headings to break the text and your key ideas into readable sections.
* Use past tense apart from descriptions about existing knowledge and results when you should use the present tense.
* Be specific about observations and quantities.
  + Avoid the use of adverbs (quickly, rapidly, slowly).
  + If a word you are using ends in 'ly' the chances are you can provide a measurement instead (over five minutes *instead of* slowly).
  + Be careful with adjectives which refer to quantities (few, lots, many).
  + Provide actual amounts (three Mayflies *instead of* a few Mayflies).
* Always back up your assertions with data or logical argument and references.
* Use a spell checker or dictionary before you print out or submit your final report.
* If you are unsure about grammar ask a friend or someone who is good at writing to read through your draft report.
* Do not use derogatory language and be careful and considerate if you criticise the work of others. Stick to the arguments and facts.

## Acknowledgements

The idea and parts of the formatting for this template came from the Association for Learning Technology (UK) ALT-C 2004 Research Paper Format Template.

## References

Knisely, K. 2005, *A Student Handbook for Writing in Biology*, 2nd Edition, Smauer Freeman.

Pechenik, J. A. 1997, *A Short Guide to Writing about Biology*, 3rd Edition, Addison-Wesley.

Silyn-Roberts, H. 1996, *A Practical Handbook for Science, Engineering and Technology Students,* Longman.

Harvard Style *UseIt,* Available at http://www.library.uq.edu.au/training/citation/harvard.html