|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **CCT College Dublin**  **Assessment Cover Page**  *To be provided separately as a word doc for students to include with every submission*   |  |  | | --- | --- | | **Module Title:** | MSc in Data Analytics | | **Assessment Title:** | MSc DA CA2 | | **Lecturer Name:** | David Mc Quaid | | **Student Full Name:** | Stephen Kelly | | **Student Number:** | sba23305 | | **Assessment Due Date:** | 07/01/2024 | | **Date of Submission:** | 12/01/2024 |   **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. | |

Contents

[Abstract 3](#_Toc151146651)

[Programming 3](#_Toc151146652)

[Question 1 3](#_Toc151146653)

[Question 2 4](#_Toc151146654)

[Data Prep 5](#_Toc151146655)

[Question 1 5](#_Toc151146656)

[Question 2 6](#_Toc151146657)

[Question 3 6](#_Toc151146658)

[Question 4 9](#_Toc151146659)

[Statistics 9](#_Toc151146660)

[Question 1 9](#_Toc151146661)

[Question 2 11](#_Toc151146662)

[Question 3 12](#_Toc151146663)

[Question 4 12](#_Toc151146664)

[Machine learning 13](#_Toc151146665)

[Question 1 13](#_Toc151146666)

[Question 2 14](#_Toc151146667)

[Question 3 15](#_Toc151146668)

[Question 4 16](#_Toc151146669)

[Acknowledgements 17](#_Toc151146670)

[References 17](#_Toc151146671)

Table of Figures

[Figure 1: Box plot total population by year 7](#_Toc151147326)

[Figure 2: bar chart total population by gender for each year 8](#_Toc151147327)

[Figure 3: heat map for 2023 values 8](#_Toc151147328)

[Figure 4: scatter chart for each year by citizenship 9](#_Toc151147329)

[Figure 5: histogram displaying population value by frequency 11](#_Toc151147330)

[Figure 6: pie chart displaying total population for all years by gender 11](#_Toc151147331)

[Figure 7: bar chart displaying total population for all years by gender 12](#_Toc151147332)

[Figure 8: machine learning techniques by category 15](#_Toc151147333)

[Figure 9: results from different modelling techniques 16](#_Toc151147334)

[Figure 10: results between different confusion matrixes before and after adjustment 17](#_Toc151147335)

# Abstract

# Programming

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. [0-15]

The data was sourced using the link provided in the report. Additional data was sourced from the following site.

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.

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.

Modern Transport planning 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 Transport planning, using tufts principles, to showcase the information/evidence gathered following your Machine Learning Analysis. Detail the rationale for approach and visualisation choices made during development. Note you may not use Powerbi, rapidminer, tableau or other such tools to accomplish this (at this stage).

# Data Prep

## Question 1

# Statistics

## Question 1

# Machine learning

## Question 1

# Acknowledgements

Thank you to CCT for putting the effort into making the MSc fun and enjoyable to take part in. Apart from the material referenced all material came from either the lectures training material or my own understanding which I gained by attending class and studying the training material.

# References