**CCT College Dublin**

**Assessment Cover Page**

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

|  |  |
| --- | --- |
| **Module Title:** | Data preparation and visualization, Machine Learning, Statistics and Programming |
| **Assessment Title:** | Irish and british transportation |
| **Lecturer Name:** | David McQuaid, Muhammad Iqbal, Taufique Ahmed, Sam Weiss |
| **Student Full Name:** | Fernando Tupa Leniz |
| **Student Number:** | 2023341 |
| **Assessment Due Date:** | 05/01/2024 |
| **Date of Submission:** | 04/01/2024 |

**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. |

**Introduction**

Transportation is a vital component of any thriving society, playing a pivotal role in fostering economic development, enhancing social connectivity, and promoting overall well-being. In the context of Ireland, understanding and optimizing the efficiency of transportation systems is crucial for the nation's growth. This research project delves into the intricacies of Irish transportation, with a specific focus on the analysis and prediction of bus types using the National Public Transport Access Nodes (NAPTAN) dataset.

The NAPTAN dataset serves as a comprehensive repository of information pertaining to public transportation nodes, offering insights into the network and infrastructure that facilitate the movement of people across Ireland. By leveraging this dataset, the research aims to employ machine learning models to predict the types of buses operating within the Irish transportation system. This predictive analysis involves the utilization of various models, including Naive Bayes and Logistic Regression, to discern patterns and relationships within the data.

To ensure the robustness and applicability of the developed models, a comparative analysis will be conducted by applying the same machine learning methodologies to both Irish and UK transportation datasets. By examining the similarities and differences between the two regions, this research seeks to identify factors that contribute to the variance in bus types and transportation dynamics. Such a comparative approach can enhance our understanding of the unique characteristics of Irish transportation and contribute to the development of targeted strategies for improvement.

Furthermore, beyond the quantitative analysis of transportation data, this research project also incorporates a qualitative dimension through sentiment analysis. By exploring public sentiments related to Irish transportation, the study aims to gauge the overall satisfaction and perception of the transportation system among the populace. This sentiment analysis will draw on textual data from various sources, such as social media, surveys, and public forums, providing a holistic view of the public's experiences and opinions regarding transportation in Ireland.

In conclusion, this research project aspires to contribute to the optimization of Irish transportation by combining NAPTAN analysis, machine learning prediction models, and sentiment analysis. Through a comprehensive examination of the transportation landscape, this study seeks to provide valuable insights that can inform policy-making, infrastructure development, and the enhancement of the overall transportation experience for the Irish populace.

**Data preparation and visualisation part:**

* **1)** 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]**

Obtaining the NaPTAN dataset for both Ireland and the UK presented a challenging yet crucial task for the research due to the requirement of acquiring the data in two different formats. This dual-format necessity added a layer of complexity to the data acquisition process, as compatibility issues and differing standards had to be addressed to ensure a seamless integration of information from both regions.

The challenge stemmed from the fact that the NaPTAN dataset for the UK was available under the Open Government Licence (OGL), as outlined in the official documentation provided by the National Archives at <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/> . This license, while fostering openness and collaboration, required careful adherence to its terms and conditions, including proper attribution and compliance with the specified usage guidelines.

Simultaneously, acquiring the NaPTAN dataset for Ireland posed a different challenge, as the dataset was made available under the Creative Commons Attribution 4.0 International License. This license, was found at <https://creativecommons.org/licenses/by/4.0> allowed for the sharing and adaptation of the data with the requirement of providing appropriate attribution to the data source.

Despite the challenges associated with navigating two distinct licensing frameworks, our research successfully obtained the necessary permissions to use the NaPTAN datasets for both Ireland and the UK. This achievement was vital in ensuring the ethical and legal use of the data, as well as facilitating a comprehensive and unified analysis that encompassed both regions.

The dual-licensing approach not only underscores the complexities of dealing with geographically diverse datasets but also highlights the commitment to upholding the principles of open data and responsible research practices. By navigating these challenges and securing the appropriate licenses, our research endeavors to contribute meaningfully to the understanding and enhancement of public transportation systems in both Ireland and the UK.

**Machine Learning part:**

**1)**

**2)**

**3)**

**4)**

**Statistics part:**

**1)**

**3)**

**4)**

**Programming part:**