**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 McQuinn  David Gonzalez  Marina Iantorno  Muhammad Iqbal |
| **Student Full Name:** | Ana Carla Ferreira de Oliveira |
| **Student Number:** | sba23438 |
| **Assessment Due Date:** | 05th January 2024 |
| **Date of Submission:** | 05th January 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. |

**CA2 | MSC in Data Analytics**

***1º S E M E S T E R***

*Ireland’s Aviation Trends in Focus Alongside European Counterparts*

***https://github.com/Oliveiranac/CA2\_MScDataAnalytics.git***

**S t u d e n t:**

|  |  |
| --- | --- |
| **sba43** | Ana Carla Ferreira de Oliveira |



**L e c t u r e r s:**

|  |  |
| --- | --- |
| **David McQuaid** | Data Preparation & Visualization |
| **David Gonzalez** | Python Programming |
| **Marina Iantorno** | Statistics |
| **Muhammed Iqbal** | Machine Learning |

**J A N U A R Y 2 0 2 4**

Contents

[1. INTRODUCTION 5](#_Toc155125047)

[1.1 Objectives 5](#_Toc155125048)

[1.2 Research Questions 5](#_Toc155125049)

[1.3 Significance of the study 6](#_Toc155125050)

[2. DATA COLLECTION AND PREPARATION 6](#_Toc155125051)

[2.1 DATA 1 6](#_Toc155125052)

[2.1.1 Data Source 6](#_Toc155125053)

[2.1.2 Data Description 7](#_Toc155125054)

[3. MODELLING 7](#_Toc155125055)

[4. CONCLUSION 7](#_Toc155125056)

[5. REFERENCES 7](#_Toc155125057)

List of Figures

# **INTRODUCTION**

The project analyzes airport transport in Europe by comparing Ireland's air transport with three more European countries that have a population proportional to Ireland. The aim is to gain insights into the trends and relationships between key indicators among the countries. The datasets under consideration cover passengers, freight, commercial flights, and the conversation passengers exhibit during the travel. The data used in the analysis spans from 2020 to 2023.

The research involves a multistage process, including data collection, pre-processing, visualization, statistical analysis, machine learning, and sentimental analysis. Python programming language is utilized in the analysis of the data. The significance of the project is to shed light on industry stakeholders, policymakers, and the general public on useful insights that can be implemented to their advantage.

## Objectives

The main goal is to comprehensively analyze Ireland's air transport in comparison with other European countries from 2020 to 2023. The study focuses on passengers, freight, and commercial flights.

## Research Questions

In this project, several research questions are taken into consideration to help in gaining insights into the data. These are:

1. Are there differences in mean values for airports in Ireland?

Hypothesis 1: Null hypothesis(H0): there is no significant difference in the mean values of VALUE across different Airports in Ireland. Alternative hypothesis(H1): there is a significant difference in the mean values of VALUE across different Airports in Ireland.

1. Is there a difference between distribution of mean VALUE between Ireland and Denmark?

Hypothesis 2: H0: There is no significant difference in the distributions of VALUE between Ireland and Denmark. H1: There is a significant difference in the distributions of VALUE between Ireland and Denmark.

1. Does the Year and VALUE variables have correlation between each other?

Hypothesis 3: H0: There is no correlation between VALUE and Year (correlation coefficient equals 0). H1: There is a significant correlation between VALUE and Year (correlation coefficient is not equal to 0).

1. Is there a difference in distribution across different months of the year?

Hypothesis 4: H0: There is no significant difference in the distribution of VALUE across different months. H1: There is a significant difference in the distribution of VALUE across different months.

1. Are mean values of scheduled flights different from unscheduled flights?

Hypothesis 5: H0: There is no significant difference in the mean values of VALUE between Scheduled and Unscheduled flights. H1: There is a significant difference in the mean values of VALUE between Scheduled and Unscheduled flights.

1. Is there a correlation between flight type and direction?

Hypothesis 6: H0: There is no correlation between direction and flight type. H1: Flight\_Type and Direction have a significant correlation.

1. Does the distribution of VALUE for scheduled flights alter according on arrival and departure?

Hypothesis 7: H0: For scheduled flights, there is no discernible change in the VALUE distribution between the points of arrival and departure. H1: For scheduled flights, there is no discernible difference in the distribution of VALUE between the points of arrival and departure.

## Significance of the study

This initiative is important because it has the potential to benefit many stakeholders, such as members of the public, legislators, and business experts, by offering insightful and useful information.

# **DATA COLLECTION AND PREPARATION**

This section offers a thorough look at the data that was used to analyze air travel in Ireland and other European nations. The study makes use of two data sets. These are the following:

## DATA 1

This dataset provides a thorough overview of Ireland's aviation trends, enabling in-depth examination and research of passenger, freight, and commercial flight patterns throughout time.

### 2.1.1 Data Source

Data.gov is the source of Ireland's passenger, freight, and commercial flight data. The Pandas package in Python can be used to access the dataset, which offers insightful statistics about air travel.

### 2.1.2 Data Description

With 29,160 entries and 9 columns, the collection contains data on a number of topics. The following are the columns and the corresponding data types for them:

Statistical\_Label: Divided into 'Freight,' 'Passengers,' and 'Commercial Flights.'

Month: The data's recorded month.

Airports in Ireland: A variety of airports located around Ireland, such as 'Cork,' 'Dublin,' 'Kerry,' 'Knock,' and 'Shannon.'

Country: The nations taking part in the transit.

Direction: 'All directions,' 'Arrival,' or 'Departure.'

Flight\_Type: Divided into three categories: scheduled, unscheduled, and all flights.

UNIT: The value column's measurement unit.

VALUE: The number corresponding to the particular category.

Year: The year the data was first recorded.

Float values that indicate different metrics pertaining to passengers, freight, and commercial flights are found in the 'VALUE' column.

# **MODELLING**

# **CONCLUSION**

# **REFERENCES**