**CCT College Du CCT College Dublin**

**Assessment Cover Page**

|  |  |
| --- | --- |
| **Module Title:** | Data Preparation & Visualisation, Machine Learning for Data Analytics, Programming for Data Analytics, Statistics for Data Analytics |
| **Assessment Title:** | CA 2 |
| **Lecturer Name:** | David McQuaid, Muhammad Iqbal, Saw Weiss, Taufique Ahmed |
| **Student Full Name:** | Lylkend Silva Albuquerque |
| **Student Number:** | 2023290 |
| **Assessment Due Date:** | 05 /01/2024 |
| **Date of Submission:** | 05 /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.** |

# MSC\_DA\_CA2

Author: L. Albuquerque

e-mail: [2023290@student.cct.ie](mailto:2023290@student.cct.ie)

Student ID: 2023290

Dublin versus Hong Kong air traffic

## Acknowledgements

I would like to thank the following people who have helped me undertake this research:

My Professor David McQuaid, for his enthusiasm in all of our classes, for his support, encouragement and patience and of course to all the others professors who strive to put us on the path of knowledge and development such as:

Sam Weiss

Dr. Muhammad Iqbal

Taufique Ahmed

The CCT college, for input throughout this MSc programme.

All my colleagues which in their way had tried to give some contributions to data collection and some expertise was given in our class group.

To the Central Statistics Office (CSO) and DATA.GOV.HK is an online web interface provided by the Office of the Government of the Hong Kong Special Administrative Region, who concisely made available all the databases in which it was possible to develop this research.

My mother Maria do Socorro who almost every day left me a message of encouragement, one of which is always in his words "There is no glory without sacrifice", special thanks.

## Abstract

## *This research work synthesizes data up to 2023 from the Central Statistics Office (CSO) of Ireland and data.gov.hk of the Hong Kong Government, focusing on the air transport numbers between Ireland and Hong Kong. Intriguingly, the data reveals a lack of correlation between air transport metrics and broader economic indicators in the two regions.*

## *Despite the conventional belief that air transport figures are closely aligned with economic performance, the analysis of data from Ireland and Hong Kong indicates a decoupling of these variables. Passenger and cargo volumes, along with flight frequencies, exhibit trends that do not consistently mirror the economic growth patterns, trade volumes, or other financial indices in either region.*

## *In Ireland, the CSO data highlights fluctuations in air transport numbers, including passenger counts and cargo loads, which do not parallel the country’s economic growth trajectory. Similarly, data from Hong Kong’s data.gov.hk shows air transport metrics evolving independently of the region's economic fluctuations.*

## *This phenomenon raises questions about the factors influencing air transport trends in these regions. The abstract explores various potential influences, such as tourism patterns, governmental policies, technological advancements in air travel, and geopolitical events, which might contribute to this observed discrepancy.*

## *The findings challenge traditional assumptions about the relationship between air transport and economic health, suggesting a more complex interaction in the context of Ireland and Hong Kong. This analysis is crucial for policymakers, airline industry stakeholders, and economists who seek a deeper understanding of the dynamics between air transport and economic indicators.*

## *The data sources, CSO and data.gov.hk, ensure a high level of reliability in the findings, though the unconventional nature of the results underscores the need for further research in this area.*

**Introduction**

Embarking on a meticulous exploration, we delve into the intricacies of air traffic data encompassing the first nine months of 2023. Our focus converges on a comparative analysis, drawing insights from the dynamic realms of Dublin Airport and Hong Kong International Airport. These critical gateways serve as conduits for global connectivity, facilitating the movement of passengers and cargo with profound implications for regional and international economies.

The Dublin Airport data unravels a narrative of flights arriving and departing, revealing patterns that encapsulate the heartbeat of Ireland's air transport network during this period. From passenger numbers to cargo movements, we scrutinize the metrics that underpin the aviation landscape in the Irish context, seeking to unveil trends, challenges, and potential opportunities shaping the aviation sector.

Simultaneously, our gaze extends across continents to the bustling airspace of Hong Kong International Airport. Here, we navigate through data.gov.hk's comprehensive repository, dissecting the nuances of air traffic patterns unique to the vibrant metropolis. The soaring skyscrapers of Hong Kong bear witness to the constant ebb and flow of flights, each indicative of the region's economic pulse and global connectivity.

In this exploration, we aim not only to present a quantitative analysis of air traffic statistics but also to unravel the stories they tell about the economic dynamics, societal trends, and unforeseen factors shaping the aviation industry. By merging the Dublin Airport and Hong Kong International Airport data, we embark on a journey to illuminate the multifaceted facets of air traffic, offering a comprehensive understanding of the first nine months of 2023 in the realm of global aviation.

**Data Preparation and visualization**

* Data gathering

The procurement of raw data for this research project encompassed an exhaustive three-week quest to identify a dataset tailored to the specific nuances of Irish data. The positive aspect of this pursuit lay in the precision achieved, aligning the data closely with the research's Irish context. However, navigating licensing intricacies proved to be a formidable challenge, introducing delays and complexities. Licensing hurdles varied across geographies—UK datasets were relatively accessible but notably concise, Brazilian datasets, available only in Portuguese, demanded translation efforts, while Canadian datasets predominantly focused on accidents. The additional obstacle of dealing with non-English datasets, such as those from the United States, prolonged the conversion process. These experiences underscore the critical role of ethical considerations and legal compliance in data acquisition, shaping the course of the research and emphasizing the need for a meticulous and adaptable approach.

**Statistics for data analytics**

T-Test Results

T-statistic: 4.5869

P-value: 4.8949e-06

Interpretation: The t-test compares the means of the 'real\_number' variable between Ireland and Hong Kong. The p-value is very small, indicating a significant difference between the means. The difference in the means of 'real\_number' between Ireland and Hong Kong is statistically significant .

ANOVA Results:

F-statistic: 0.00176

P-value: 0.9665

Interpretation: The ANOVA assesses whether there are significant differences in the means of 'real\_number' across different groups (possibly months or directions). The p-value is high, indicating no significant differences. There is no evidence of a significant difference in the means of 'real\_number' among the groups in this case arrivals and departures.

Wilcoxon Results:

Statistic: -13.2923

P-value: 2.5645e-40

Interpretation: The Wilcoxon test assesses the difference in the distribution of 'real\_number' between Ireland and Hong Kong. The very low p-value indicates a significant difference. The distribution of 'real\_number' is significantly different between Ireland and Hong Kong.

Chi-Squared Results:

Chi-squared Statistic: 464.2253

P-value: 0.5145

Interpretation: The Chi-squared test assesses the association between two categorical variables (possibly 'Direction' and 'Month'). The high p-value suggests no significant association. There is no evidence of a significant association between the two categorical variables.

Tukey Post Hoc Results:

Interpretation: The Tukey post hoc test is applied after ANOVA to identify specific groups with significant differences. In this case, there's only one pair: 'Arrival' and 'Departure'. There is no significant difference in the means of 'real\_number' between 'Arrival' and 'Departure' groups.

Cramer's V Value:

Value: 0.7284

Interpretation: Cramer's V measures the strength of association between two categorical variables. A value close to 1 indicates a strong association. There is a strong association between the two categorical variables being analysed.

In summary, the t-test and Wilcoxon test suggest significant differences in the 'real\_number' variable between Ireland and Hong Kong. However, the ANOVA, Chi-squared test, and Tukey post hoc analysis indicate no significant differences in means or associations among groups. Cramer's V suggests a strong association between the analysed categorical variables.

## References

Barone A. 2023. Binomial Distribution: Definition, Formula, Analysis, and Example. Investopedia [Internet]. [accessed 2023 Nov 12]. https://www.investopedia.com/terms/b/binomialdistribution.asp

IBM Documentation. 2021. [accessed 2023 Nov 10]. https://www.ibm.com/docs/sr/spss-modeler/saas?topic=dm-crisp-help-overview

KDD Process in Data Mining. 2018. GeeksforGeeks [Internet]. [accessed 2023 Nov 10]. https://www.geeksforgeeks.org/kdd-process-in-data-mining/

Kumar M. 2022. Project Management in Data Science using SEMMA. Accredian [Internet]. [accessed 2023 Nov 10]. https://medium.com/international-school-of-ai-data-science/project-management-in-data-science-using-semma-a51ddf99cbd9

What is Exploratory Data Analysis? What is exploratory data analysis? [Internet]. [accessed 2023 Nov 7]. https://www.ibm.com/topics/exploratory-data-analysis