# sba21725 - MSc in Data Analytics

Author: G. Baudo

e-mail: [sba21725@student.cct.ie](mailto:sba21725@student.cct.ie)

Student ID: sba21725

Github Repo: <https://github.com/sba21725/MSC_DA_CA1>

# Abstract

This Report aim at applying the knowledge and skills learnt at CCT College. Technique

# Introduction

Tourism is an important business sector for the economy of a country. A study by (Stryzhak, 2019) demontrates *a positive relationship between the development of the tourist sector and the level of economic welfare of the country*.

Machine Learning in Tourism (Egger, 2022). Data Analytics in Industry 4.0 (Duan, 2021).

Related Concepts

Current Interdisciplinary Research between Data Analytics and Industry 4.0 (lit rev)

# Method

## Research approach

This research aim at solving the following problems:

1. Support the author learning process. The timeframe given to the project is strictly defined and limited.
2. Identify the problem. A mathematics scientist engaged in a piece of research can attach his problem straight away because a scientific framework already exists. In other words the problem is well defined (Popper, 2005).
3. Apply tools and techniques from statistics, machine learning, data preparation and visualization, and programming.
4. Project Report.

The research design, or research template by consequences of the intertwined subjects is to be an ensamble of Describe, Explain, Correlate, and Predict (Walliman, 2021).

Interdisciplinary approach in Data Analytics aiming at discovering patterns in data and making predictions using the tools and techniques from four different subjects: Data Preparation and Visualization, Statistics, Programming, and Machine Learning.

Although those subjects are different by nature, some characteristics are shared between them. For example praparing and visualizing a dataset is a prerequisite for statistical analysis and machine learning algorithms.

A Human Computer cognitive learning process will be employed. {four way of thinking book}

As a consequence the main focus of this research is applying the tools and techniques of the subejcts mentioned above to the domain of Tourism and Travel in Ireland. Because the main objective is applying as many techniques as possible to the domain, it may appear to be explored too broadly and seemingly not uniform.

The literature review is omissis cause a limited space and a short time frame.

## Programming approach

The programming approch is mainly iterative delivered by a Jupyter Notebook and the present report. Along with the iterative paradigm, an Object Oriented Programming will be employed to define a class that can be instanciated and used for transformation pipelines.

A Function Programming Paradign will be preferred to pipeline, manipulate, and transform datasets.

A class hence OOP will be used to create a class for a pipeline

Data are collected from cso web site using the REST API made available by the platform. This technique will be consumed only to retrieve the datasets.

For each dataset a dataframe will be created using a convenience function which returns a Pandas DataFrame. This allows the use of a single variable name “df” to indicate the dataframe being considered. Identify the original dataframe from the context in which it is being called.

# Tourism in Ireland

Tourism expenditure in Ireland during 20xx to 20xx

# Results

## Overseas Trips to Ireland by Non-Residents in 2009

Expenditure from turism coming from, see tab. 1: Overseas Trips to Ireland by Non-Residents in 2009.

Table 1 – Overseas Trips to Ireland by Non-Residents in 2009

|  |  |
| --- | --- |
| **Area of Residence** | **Number** |
| Germany | 812,000 |
| France | 577,000 |
| Italy | 392,000 |
| United States and Canada | 2,412,000 |
| Great Britain (includes England, Scotland, Wales) | 3,788,000 |
| Australia and New Zealand | 225,000 |
| Other Europe (14) | 2,129,000 |
| Other countries (18) | 473,000 |

From this statistic, Area of Residence Expenditure Overseas Trips to Ireland

A graph with different colored bars

Description automatically generated

## Correlation between Number of Trips and Average Number of Nights

A graph showing different colored dots

Description automatically generated

Data Sources

In order to carry on the research the following institution will be browsed.

<https://ec.europa.eu/eurostat>

<https://www.cso.ie>

## Tourism Balance between Travel to and from Ireland

The balance is positive for Ireland

# Discussion

Finding relevant dataset to analyze is an activity per se. Once the raw data is available the next step is

# Conclusion e Future Research Direction

Unsupervised machine learning, Artificial Neural Networks, and Deep Learning over more sofisticated visualization skills and statistical techniques. By skill is meant the ability to quickly program a visualization function in a Jupyter Note.

# Bibliography

Duan, L. a. D. X. L., 2021. Data analytics in industry 4.0: A survey. *Information Systems Frontiers,* pp. 1-17.

Egger, R., 2022. Machine Learning in Tourism: A Brief Overview: Generation of Knowledge from Experience. *Applied data science in tourism: Interdisciplinary approaches, methodologies, and applications,* pp. 85-107.

Popper, K., 2005. *The Logic of Scientific Discovery.* New York: Routledge.

Stryzhak, O., 2019. Assessment of the relationship between the tourism sector development and other sectors of economy. *Economics of Development,* Volume 2, pp. 10-18.

Walliman, N., 2021. *Research methods: The basics.* London: Routledge.

# Annex