A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Francisca Andrea Argandona Alvarado |
| *Student Number* | 2024247 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 2 Capstone Report |
| *Assessment Due Date* | 15th December 2024 |
| *Date of Submission* | 15th December 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

[Introduction 1](#_Toc180946812)

[Objectives 2](#_Toc180946813)

[Problem definition 3](#_Toc180946814)

[Scope 4](#_Toc180946815)

[Problem definition 4](#_Toc180946816)

[Data collection 4](#_Toc180946817)

[Discovery 4](#_Toc180946818)

[Data Preparation 4](#_Toc180946819)

[Model Planning and Building 4](#_Toc180946820)

[Results 5](#_Toc180946821)

[Timeline 5](#_Toc180946822)

[Data sources 6](#_Toc180946823)

[Ethics 7](#_Toc180946824)

[References 8](#_Toc180946825)

[GitHub Link 8](#_Toc180946826)

The report should be presented in a clear and concise manner and should demonstrate your ability to use a project management methodology.

Additionally, the report should provide a background to the business problem and its importance from a strategic viewpoint, an overview of the project's timeline, milestones achieved, and any challenges faced during the implementation phase.

It should highlight the key insights gained from analysing the data and present any significant trends or patterns observed. Finally, the report should address any limitations or constraints encountered during the project and propose potential solutions for future improvements.

Crisp DM

Timeline

Strategic Overview of the business problem

2. Project Plan

3. Business Understanding

4. Data Understanding

5. Data Preparation

6. Machine Learning Implementation

7. Findings & Conclusions

8. Any Future Recommendations

# Introduction

The economy of a country can be influenced by different sectors such as manufacturing, agriculture, construction, and tourism, among others.

One of the countries that has been positively affected by tourism is Iceland. It is a Nordic country located in the Atlantic Ocean. The variety of its landscapes, like geysers, volcanoes, hot springs, and glaciers, catch the attention of people all over the world, regardless of which season it is.

In the last decades, Iceland's economy has been diversifying into manufacturing and service industries, where tourism belongs. Although it was hugely negatively affected by the COVID-19 pandemic, the number of tourists increased even more after it.

As we can see in the following graphic that shows the “Percentage change in the number of international visitor arrivals to Iceland from 2010 to 2022”, the international tourism volume has experienced every year growth of 146% in 2022.

Tourism in Iceland was not a big topic to talk about 10 years ago. Due to the development that this area has grown up too fast, it can be important for many travel agencies to know the preferences of the tourists that visit the island, depending on how long they stay, how many people they travel with, how long in advanced they book their flights and accommodations, among other things, a personalised package can be arranged for them.

A graph with blue and white bars

Description automatically generated

# Objectives

In the beginning the main purpose of the following assignment was to predict how many tourists will visit Iceland per year in the next 5 years, nevertheless it will be changed to predict how likely the visitors from the most common countries that travel to the island, would visit it again, and what factors depends on their decision.

The reason of the objective changed is due to one of the variables to predict the future visitors is the price of the flights. That information depends on each country the tourists are from, the airlines and season. The timeline for this project is not enough to collect that information.

A diagram of data mining

Description automatically generated

# Methodology

For this project CRISP-DM will be implemented as methodology to solve the problem. The followings steps describe each stage of the project.

## Business Understanding

The landscapes, nature and culture of this country have caused an increase of the tourism industry getting attention from people all over the world. From the volcano hiking to northern light hunting, the variety of attractions this country has is diverse.

Due to the development of the tourism, there can be a saturation in the high seasons on some attractions, making people not getting a good experience. In the other hand, there are some patterns that people follow depending on the country they are visiting from. Finding this pattern depending on the nationality, length of stay, how long in advance they book the flights and organized the trip, travel companions, among others, is important to understand the preferences of the visitors. If they can get a good experience and they are likely to go back, that will be a positive result for the economy of the country in the future.

Due to the timeline for this project is short, I am going to analyse just the top 10 countries who visit Iceland in the last year of the collected data.

## Data Understanding

The data for this assignment was found in “Ferðamálastofa / Icelandic Tourist Board” website. They made a survey to all departure passengers who arrived at Keflavik International Airport, all year round. The data collected belongs to 2023, 2022, 2016, 2015, 2014 and 2013. The data are presented in excel and pdf. The top 10 countries in 2023 are: USA, United Kingdom, Germany, Poland, France, Italy, Spain, Canada, Netherlands and Denmark.

Due to the data is split in different files, the first step will be to join them in one file so it will be easier to understand.

## Data Preparation

## Modelling

## Evaluation

# Problem definition

Because tourism in Iceland is a new topic that has been more mentioned in the last ten years and it has increased exponentially, it may be a problem for the country in different aspects such as, budget, infrastructure, and services among others.

If the number of visitors continues increasing exponentially, the country needs to get more knowledge on how to approach this new area.

It is important to address this problem because it can avoid a saturation of sources in the facilities. For instance, the international airport could have a better flow on the landings and arrivals so that there will not be delays. Accommodations can hire the exact amount of employees in relation to get a profit.

# Scope

This project will analyse the data for the last 20 years regarding the amount of people who visited Iceland as tourists with the purpose of predicting the visitors in the next 5 years. The data must be collected from people who entered the island by aeroplane or ferry. People who have visited the country with a different purpose, for instance, work, will not be considered. Due to Iceland's recent development of tourism, the data collected will be no more than 20 years old.

Following a timeline and determined steps is crucial to reach the objective of the project in the two semesters given for it.

## Problem definition

Having a clear idea of the problem that must be solved is important for not doing extra work or collecting useless information.

## Data collection

The data has to be collected from reliable and updated sources.

## Discovery

Once we get the necessary data to reach our objective, it must be analysed to see any patterns it follows. EDA can be applied at this stage to make any complex data to a more understandable one.

## Data Preparation

The data must be cleaned to make an accurate prediction. Missing data must be eliminated or filled using different methods. Normalisation or PCA, can also be applied on this step if it is required.

## Model Planning and Building

On this step, the best model for the project must be selected. First, make clear if a supervised or unsupervised model will be needed.

Once the model is trained, make sure that it is not overfit or underfitted.

Jupyter Notebook will be used to build the model with Python language.

## Results

After the model is trained, communicate the results in a presentation at the end of the second semester. 

# Timeline

A chart with icons and text

Description automatically generated with medium confidence

# Data sources

The data that will be collected must include the number of tourists that have arrived in the country in the last 20 years, and it must be monthly or seasonally specified.

For this project, the data will be collected from different online sources. One of them will be Statistics Iceland. It is a website for official statistics in Iceland. They collect information from different aspects that affect the country, including tourism.

Another source will be [Ferðamálastofa Icelandic Tourist Board](https://www.ferdamalastofa.is/en), which is an independent authority under the Ministry of Culture and Business Affairs. One of the ways they collect the data is through surveys among domestics and foreign tourists.

# Ethics

It is important to analyse any ethical consideration that can affect the result of the project.

Personal information is not required for this project, except for nationality, which may be positive data to make sure it does not belong to just one group of people from one area.

Being aware of past events such as, economic crises, pandemics, or natural disasters is important to understand that it can affect the results.

The government and public and private companies will be positively affected by the result of the project so that they can get a better budget spend, but it is crucial to understand that the prediction of the model will not have 100% accuracy.

New international treaties or visas can affect the prediction of the model.

# References

Bjarki Bents­son, J. (2024). *Will tourist numbers rise this year?* [online] www.islandsbanki.is. Available at: <https://www.islandsbanki.is/en/news/will-tourist-numbers-rise-this-year>.

Dirección General de Comunicación, Diplomacia Pública y Redes de España (2024). *Islandia Islandia OFICINA DE INFORMACIÓN DIPLOMÁTICA FICHA PAÍS*. [online] Available at: <https://www.exteriores.gob.es/Documents/FichasPais/ISLANDIA_FICHA%20PAIS.pdf>.

Government of Iceland (n.d.). *Government of Iceland | Tourism in Iceland*. [online] www.government.is. Available at: <https://www.government.is/topics/business-and-industry/tourism-in-iceland/>.

López, A.M. (2024). *Change in international tourism in Iceland 2022*. [online] Statista. Available at: <https://www.statista.com/statistics/694121/tourism-industry-international-visitor-growth-iceland/>.

## GitHub Link

<https://github.com/CCT-Dublin/capstone-project-Pancha19.git>