**CCT College Dublin**

**Assessment Cover Page**

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

|  |  |
| --- | --- |
| **Module Title:** | *Programming for DA*  *Statistics for Data Analytics*  *Machine Learning for Data Analysis*  *Data Preparation & Visualisation* |
| **Assessment Title:** |  |
| **Lecturer Name:** | *David Gonzalez*  *Marina Iantorno*  *Muhammad Iqbal*  *David McQuaid* |
| **Student Full Name:** | *Natalia Iolchin* |
| **Student Number:** | *sba23303* |
| **Assessment Due Date:** | *10/11/2023* |
| **Date of Submission:** |  |

**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. |

# Group ID - MSc in Data Analytics

Author: N. Iolchin

e-mail: [sba23303@student.cct.ie](mailto:sba23303@student.cct.ie)

Student ID: sba23303

1. **Introduction**

<https://www.gov.ie/en/publication/c1b0c9-national-broadband-plan/>

<https://nbi.ie/>

<https://esb.ie/media-centre-news/press-releases/article/2023/05/19/esb-networks-completes-process-for-development-of-private-mobile-network>

<https://eire.fandom.com/wiki/Regions_of_Republic_of_Ireland>

<https://aerfortel.com/maps-of-ireland/>

<https://data.cso.ie/product/pme>

<https://www.cso.ie/en/methods/informationnotefordatausersrevisiontotheirishnuts2andnuts3regions/>

<https://www.cso.ie/en/media/csoie/releasespublications/documents/ep/censuspreliminaryresults/2022/backgroundnotes/NUTS3_Region.xlsx>

**Development of Internet access in Ireland**

A decent internet connection nowadays is essential for many basic tasks. The objective of this work is to determine the number of people with internet access in Ireland and its regions, as well as the types of connectivity used.

To achieve this objective, data on internet access rates in Ireland and its regions will be gathered from the Central Statistics Office (CSO) website.

Data collection

While exploring CSO website I found that there are several sections that contain information about Internet Access: “People and Society” and “Census”

More deep exploration discovers that there are some data from 2009 – until 2016, some data are specifically related to only 2011, 2016 and 2022 years, another piece of data is related to 2020-2022 years or 2021-2022, and there is no any data for period from 2016 until 2020.

The recent data collected from the "People and Society" section for the years 2020 to 2022 is more varied than the data collected in previous years in terms of the classifications and data slices that it includes. However, the current work aims to assess the development of internet connectivity over a longer period of time, so as to include not only the current state of the industry and its future prospects, but also its historical development and the pace at which it has developed.

In order to provide the most comprehensive overview of the development of internet connections, data that maintains its structure over a longer period of time was used as the foundation. The geographical development of internet connections and the general classification of connection types were the primary factors considered.

A valuable addition to the base data would be to include information on the absence of an internet connection, as well as the number of people who did not specify their internet connection status at all.

To be able to assess number of people connected to the internet over the total number of estimated populations, an additional file from “People and Society” section was taken.

Three separate files were obtained from the CENSUS section for the years 2011, 2016, and 2022. These files contain data on the number of private households, types of internet connections and have nearly identical geographical structures.

To convert private household number into a people number and to have congeneric geographical data structure, additional research was performed. CSO website has a separate section with historical data of regions and counties classification and number of people in private households divided geographically for all relevant years.

For better understanding of the collected data, please check TABLE 1.

A screenshot of a computer

Description automatically generated

After data collection, there will be data discovery and exploration.

Next phase is data analysis and the results, that will include:

* The number and percentage of people with internet access in Ireland and its regions
* The types of connectivity used.
* Any trends, patterns or correlations found in the data.
* Conclusions.

The focus will be on persons using the internet in private households in Ireland, and other domains such as enterprises or educational institutions will not be included.

After CSO website research I found that there is no specific file (source) for the number or percentage of the people who have Internet Access. There will be necessary to find/calculate these numbers indirect from available data. Thus, was found sources with:

* Total number of people in the country and per region
* Number of private households and number of persons in private households
* Number of households with Internet access, including type of internet connection (3 separate files for 2011, 2016 and 2022 years)

The first observation shows that there is inconsistency in regions classification. Thus, we need a clear determination of Irish Regions.

To handle the region classification inconsistency, I’ve performed additional research of sources from the CSO website. As a result, a separate file with regional classification was created and the most aggregated classification will be used. Please use Table 1 as a reference.

|  |  |
| --- | --- |
| **Regions of**[**Republic of Ireland**](https://eire.fandom.com/wiki/Republic_of_Ireland) | |
| [**Border Region**](https://eire.fandom.com/wiki/Border_Region) | [Cavan](https://eire.fandom.com/wiki/County_Cavan) | [Donegal](https://eire.fandom.com/wiki/County_Donegal) | [Leitrim](https://eire.fandom.com/wiki/County_Leitrim) | [Louth](https://eire.fandom.com/wiki/County_Louth) | [Monaghan](https://eire.fandom.com/wiki/County_Monaghan) | [Sligo](https://eire.fandom.com/wiki/County_Sligo) |
| [**West Region**](https://eire.fandom.com/wiki/West_Region) | [Mayo](https://eire.fandom.com/wiki/County_Mayo) | [Roscommon](https://eire.fandom.com/wiki/County_Roscommon) | [Galway](https://eire.fandom.com/wiki/County_Galway) |
| [**Midlands Region**](https://eire.fandom.com/wiki/Midlands_Region) | [Laois](https://eire.fandom.com/wiki/County_Laois) | [Longford](https://eire.fandom.com/wiki/County_Longford) | [Offaly](https://eire.fandom.com/wiki/County_Offaly) | [Westmeath](https://eire.fandom.com/wiki/County_Westmeath) |
| [**Mid-East Region**](https://eire.fandom.com/wiki/Mid-East_Region) | [Kildare](https://eire.fandom.com/wiki/County_Kildare) | [Meath](https://eire.fandom.com/wiki/County_Meath) | [Wicklow](https://eire.fandom.com/wiki/County_Wicklow) |
| [**Dublin Region**](https://eire.fandom.com/wiki/Dublin_Region) | [Dublin](https://eire.fandom.com/wiki/County_Dublin) |
| [**South-East Region**](https://eire.fandom.com/wiki/South-East_Region) | [Carlow](https://eire.fandom.com/wiki/County_Carlow) | [Kilkenny](https://eire.fandom.com/wiki/County_Kilkenny) | [Tipperary](https://eire.fandom.com/wiki/County_Tipperary) | [Wexford](https://eire.fandom.com/wiki/County_Wexford) | [Waterford](https://eire.fandom.com/wiki/County_Waterford) |
| [**South-West Region**](https://eire.fandom.com/wiki/South-West_Region) | [Kerry](https://eire.fandom.com/wiki/County_Kerry) | [Cork](https://eire.fandom.com/wiki/County_Cork) |
| [**Mid-West Region**](https://eire.fandom.com/wiki/Mid-West_Region) | [Clare](https://eire.fandom.com/wiki/County_Clare) | [Tipperary](https://eire.fandom.com/wiki/County_Tipperary) | [Limerick](https://eire.fandom.com/wiki/County_Limerick) |

Detailed overview with observations of the data can be found in **this Jupiter Notebook**.

At this point we have all the data collected.

To be able to use our data with maximum efficiency there is a need to perform modification, cleaning, and other data manipulation.

Step by step data processing can be found in **this Jupiter Notebook**.

As a result, we have 1 file as a base for our next analysis. Table 2 reflect view of the final Data Frame:

A map of ireland with different colored areas

Description automatically generated

* **over 560,000 premises**
* **1.1 million people**
* **Over 65,000 farms**
* **44,000 non-farm businesses**
* **679 schools**
  1. **Objective**

Focus: Develop new products and services

* **Total amount of data used:** to predict the total amount of data that users will use based on their type of internet connection, frequency of usage, and demographic data.
* **Identification of trends in internet usage:** For example, information that certain demographics are using more data than others, can be used to develop new products and services, or to target marketing and advertising campaigns more effectively.
  1. **Scope**

Problems based on data about internet usage of households and enterprise, including:

* **Predicting future internet usage:** Machine learning algorithms can be used to predict future internet usage based on historical data. This can be useful for businesses to plan for future capacity needs, and for governments to plan for future infrastructure investments.
* **Identifying trends in internet usage:** Machine learning algorithms can be used to identify trends in internet usage, such as the increasing popularity of streaming video or the growing use of social media. This information can be used by businesses to develop new products and services, and by governments to develop new policies and regulations.
* **Segmenting customers:** Machine learning algorithms can be used to segment customers based on their internet usage patterns. This information can be used by businesses to target their marketing and advertising campaigns more effectively.
* **Developing new products and services:** Machine learning algorithms can be used to develop new products and services that are tailored to the needs of internet users.

1. Analyse if population is using internet with positive trend. (ICA77 2020)
2. Analyse if frequency of usage increases. ICA79 2017-2022
3. Analyse if estimated population will increase. Predict growth for 3-5 years per region. PEA04 2011-2023
4. Analyse age of people with max% of internet usage, that use internet more often. ICA108 2020-2022
5. Correlate this age with estimated population. (FORMULAS)
6. Analyse most disposable devices/recycling. SUST02 2022
7. Correlate mob and desktop/laptop devices Predict usage of different type of devices? (FORMULAS)
8. Correlate device usage (ages) with internet usage ages. (FORMULAS)
9. Analyse usage per region (ICA172 2020-2022)
10. Analyse type of internet connection per region Analyse trend (ICA174 2021-2022)
11. Correlate development of mob and fixed infrastructure. Predict usage of each type of infrastructure

**FUTURE RESEARCH**

Correlate Ireland trends with European?

Educational sector?

Educate population on how to use?

Develop TV streaming instead of current TV licence system?

Economic trend for Tech industry development?

Logistic vs e-commers?

1. **Materials and Methods**

Examples of machine learning algorithms that can be used to solve these problems:

* **Linear regression:** Linear regression can be used to predict future internet usage based on historical data.
* **Clustering:** Clustering can be used to identify trends in internet usage, such as the increasing popularity of streaming video or the growing use of social media.
* **Decision trees:** Decision trees can be used to segment customers based on their internet usage patterns.

We need to analyse trends in estimated population and internet usage to have the best picture for service and products delivery. The periods of time will be 3, 5 and 10 years.

Ireland’s official regions are.

* Border Region
* West Region
* Midlands Region
* Mid-East Region
* Dublin Region
* South-East Region
* South-West Region
* Mid-West Region

We will include one more region – State to have our stats for the whole country as well.

Future business strategy will depend on increase or decrease of population per specific region and age range, increase or decrease of internet usage per specific region and age range.

First let’s analyse estimated population of Ireland. For this we will use data from PEA04

1. Let’s analyse the tendency for population estimation in Ireland in the next 3, 5 and 10 years. For this we will use the dataset\_\_\_ and method\_\_\_.
2. Now let’s analyse same trend for each region\_\_\_.
3. To have full picture let’s analyse trends for selected age range.

Now, set’s see general trend of the internet usage for Ireland and it’s regions and create an estimation for the next 3, 5 and 10 years. For this we can use ICA79 file. (**Note:** this file was selected as it’s only one file that contains geolocation data with the maximum range of years. Thus, we can have more data for our future calculation and our predictions will be more accurate.)

From the data exploration we can see that age ranges are different, In the file with estimated population we have smaller ranges. Taking in consideration our focus on the internet usage and to have more accurate prediction there is a need to **adjust age ranges** from the dataset with estimated population to those from internet usage.

Here we can have a first stop for our first conclusion.

From the analyse of the estimated population and internet usage we can see positive trends. Our predictions just confirm those trends.

Our next part of analyze will be based on the data about people who use internet every day. For this we will use data from ICA108, that contain age ranges and from ICA180 with geolocation data.

1. **Results** 
   1. Programming for DA
   2. Statistics for Data Analytics
   3. Machine Learning for Data Analysis
   4. Data Preparation & Visualisation
2. **Discussion/Conclusions**

Identification of trends in internet usage. For example, information that certain demographics are using more data than others, can be used to develop new products and services, or to target marketing and advertising campaigns more effectively.

Overall, having data about type of internet connection, frequency of usage, and demographic data can be very useful for predicting internet usage and identifying trends. This information can be used to make better decisions about how to allocate resources and develop new products and services.