**Population in Ireland: Where are we heading to?**

(Word count\*: 1945)

*\* Excluding diagrams, code, references and titles.*

By

Leopoldo Rojo Romero (Student ID: 2023355)

Higher Diploma in Science in Data Analytics for Business

Strategic Thinking

James Garza

CCT College

Dublin, Ireland

**Table of contents**

[**Introduction** 2](#_Toc153730836)

[**Problem definition** 2](#_Toc153730837)

[**Scope** 3](#_Toc153730838)

[**Data source** 4](#_Toc153730839)

[**Ethical considerations** 4](#_Toc153730840)

[**Approach and data exploration** 5](#_Toc153730841)

[**Findings** 8](#_Toc153730842)

[**Machine Learning model** 11](#_Toc153730843)

[**Github** 12](#_Toc153730844)

[**References** 13](#_Toc153730845)

# **Introduction**

Data is a collection of facts. When data is summarised, it gives us information about the context within which it was collected, and with the use of analytical tools we can produce evidence of the information, and with the evidence we have today, we can attempt to predict future occurrences.

We would be producing a capstone project on an interesting and relevant subject ‘population-growth rates’. We would be conducting an in-depth analysis of population decline in Ireland as many of today’s largest economies are reported to be experiencing decline in population growth rate (World101, 2022).

(Honohan, 2021) reported Ireland as one of the largest GDP in Europe, as a result, its economic stability is imperative, if the European Union (EU) will retain its economic viability on the world stage.

The Organisation for Economic Co-operation and Development (OECD, 2023) and the United Nations (UN, 2023) reported that a country requires a fertility rate of 2.1 children per woman to ensure a stable population. For several decades, the EU are below this fertility rate and for most case, this decline has been below replacement level (UN, 2023). Many factors influence a countries population like birth rate, death rate and migration been key determining factors of population growth or decline (World101, 2022).

Through exploratory data analysis, we aim to find if there is any correlation between this 3 major factors and population decline / growth in Ireland. Our project aims to predict based on the decline or growth trend we find in our analysis when the birth rate in Ireland will not compensate for the rate of deaths.

**Objective**

* Observe through exploratory analysis of our data, the population growth/decline trend in Ireland from 1960 to 2022.
* Check correlation between the number of births, deaths, and marriages, been determining factors of population growth/decline and population growth/decline in Ireland.
* Find an optimal machine learning model that makes projections into the future regarding Ireland’s population.
* Analyse our results to make conclusions concerning Ireland’s population stability.

# **Problem definition**

Human capital is an important bed rock of any stable economy, a large population means more workers and customers which ultimately boosts a country’s GDP (World101, 2022).

As stated by (Wilmoth, Menozzi and Bassarsky, 2022) world population is expected to peak around 2100 at a level of almost 11 billion. (OECD and UN, 2023) both reported a country requires a birth rate of 2.1 children per woman to maintain a healthy population growth. However as shown in *fig.1,* Ireland has recorded a progressive decline in its birth rate since 1970 but were still within a healthy birth rate.

Gráfico

Descripción generada automáticamente

Figure : Fertility rate, children per woman 1970-2022 (OECD 2023)

Since 1991, Ireland has declined below a healthy population birth rate, and trend suggests that this decline may continue unless objective measures are implemented to prevent it.

We believe Ireland is facing a significant challenge in its birth rate which is a major driver of population growth or decline. This challenge could potentially impact Ireland’s economic and societal stability.

# **Scope**

In this capstone project we will analyse the current trend of the population in Ireland with public data collected by the Central Statistics Office (CSO) and the United Nations (UN) from 1960 to 2022. With this data we seek to explore the correlation between the number of births, deaths and marriages in Ireland to provide a better understanding of its population trend. With this data we aim to predict when the number of deaths will be greater than the number of births per year.

**Project phases:**

**Data preparation,** where we will perform an Exploratory Data Analysis (EDA) to get a better understanding of our dataset, this involves cleaning, sorting, describing and shaping the dataset and sourcing mora data if needed. Using a Jupyter Notebook and python language, this task will represent around 80% of the workload and time during project implementation due to the importance of high-quality data going into a machine learning model to be able to obtain reliable results.

**Machine Learning,** we will train multiple machine learning models and observe which model optimal for our dataset. Through metrics that measures regression models, we will ensure machine learning model makes accurate and reliable predictions. This task will represent around 15% of the workload and time during project implementation.

**Visualization,** we will get a better visual understanding of relationships and trends in our dataset using tools such as Seaborn and Matplotlib.pyplot which will help us draw our conclusions and provide visual evidence to stakeholders during presentation. This task will represent around 5% of the workload and time during the project.

# **Data source**

The dataset we intend to use for our independent features of this analysis is going to be gotten from the CSO Ireland office. The CSO is responsible for collection, compilation, extraction and dissemination of high-quality data used for analysis of economic and social important topics in Ireland.

Additionally, for our dependent feature we intend to use a dataset gathered by the United Nations. The United Nations (UN) is an international organization founded in 1945 and it remains the one place on Earth where all the world’s nations can gather together, discuss common problems, and find shared solutions that benefit all of humanity (United Nations, 2023a).

We believe they would be able to provide accurate and reliable data for our project, we got data on the quarterly birth rate, death rate, marriage rate, migration rate population trend of Ireland from year 1960 to 2022 by the CSO and the population per year of Ireland from the UN. With this data we seek to accomplish our project scope.

# **Ethical considerations**

We considered the socioeconomic importance of this project topic, as this can influence public action or inactions, for example people deciding to have more children because of fear of decline, but we established after consideration that we do not seek to make any conclusions on the population stability of Ireland but just to analyse and report our findings for capstone research purpose only.

Also, to prevent breach of anonymity of population from which the data was gathered, the CSO anonymised all the dataset.

We rang CSO office Dublin to advise us on permission we might require to use their data, On the phone CSO confirmed we do not require any permission to use data available on open source as shown in *fig.2* and [here](https://www.cso.ie/en/aboutus/whoweare/copyrightpolicy/%20) .

A diagram of a pyramid

Description automatically generated

Figure : Central statistics office on data usage

As well as the data from the CSO, the data from the UN is available in electronic format, for this data is important to mention the following disclaimer:

These documents do not imply the expression of any opinion whatsoever on the part of the Secretariat of the United Nations concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries (United Nations, 2022b).

# **Approach and data exploration**

At this first stage we will approach the case aiming to fit a Random Forest Regressor Machine Learning model, due to its multiple benefits like low risk of overfitting, easy to determine feature importance and not needing considerable pre-processing work, to name a few.

Let’s have a first look at our data:

Texto

Descripción generada automáticamente

Figure : Dataset 1, independent features.

Aplicación

Descripción generada automáticamente con confianza baja

Figure : Dataset 2, dependent feature (target).

As we can see, we will need to perform data cleaning, manipulation and feature engineering to be able to analyse our data.

First, we need to understand our features:

***Dataset 1***

Statistic Label

- Marriages Registered

- Marriage Rate Registered per 1000 Estimated Population

- Deaths Registered

- Death Rate Registered per 1000 Estimated Population

- Deaths of Infants under 1 Year Registered per 1000 Births

- Births Registered

- Birth Rate Registered per 1000 Estimated Population

- Opposite sex marriages

- Same sex marriages

- Civil partnerships

- Births registered that were first births

- Births registered outside marriage

- Opposite sex marriage rate

- Same sex marriage rate

- Civil partnership marriage rate

- Percentage of first births registered

- Percentage of births registered outside marriage

- Average age of mothers giving birth

Quarter

Years from 1960 (Q1) to 2023 (Q1); each year has 4 quarters (Q1, Q2, Q3 and Q4).

State

- State

UNIT

These represents the metric of the next feature (VALUE).

- Number

- %

VALUE

Values regarding the Statistic Label feature.

***Dataset 2***

Entity

Countries.

Code

Abbreviation of the countries.

Year

Years of the values.

Population - Sex: all - Age: all - Variant: estimates

Values of the population.

Population - Sex: all - Age: all - Variant: medium

Projections values of the population.

Subsequently, and knowing that we require to analyse values per year and that we are not analysing rates and groups of population (male, female or infants), we are in a position to perform our cleaning, manipulation and feature engineering to end up with the following dataset:

Tabla

Descripción generada automáticamente

Figure : Dataset after pre-processing the data.

# **Findings**

Population in Ireland throughout 1960 to 2022:

Gráfico

Descripción generada automáticamente

Figure : Population in Ireland from 1960 to 2022.

Births, deaths and marriages in Ireland throughout 1960 to 2022 (0 to 62):

Gráfico, Gráfico de líneas, Gráfico de dispersión

Descripción generada automáticamente

Figure : Independent features behaviour from 1960 to 2022.

Relationship between births and deaths in Ireland throughout 1960 to 2022 (0 to 62):

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Figure : Relationship between births and deaths.

Relationship between births and marriages in Ireland throughout 1960 to 2022 (0 to 62):

Gráfico, Gráfico de líneas, Gráfico de dispersión

Descripción generada automáticamente

Figure : Relationship between births and marriages.

Succeeding the observation of relationships between births, deaths and marriages, let’s have a closer look at births and deaths due to the declining and rising respectively.

Births in Ireland throughout 1960 to 2022 (0 to 62):

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Figure : Births in Ireland from 1960 to 2022.

As shown in *fig.10*, we can observe a clear pattern declining births in Ireland every 15 years and it seems that a rising pattern lasting 15 years its developing, therefore, we can expect 3 more years (from 2022) of further declining of births in Ireland reaching its lower value in 2025.

Deaths in Ireland throughout 1960 to 2022 (0 to 62):

Gráfico, Gráfico de dispersión

Descripción generada automáticamente

Figure : Deaths in Ireland from 1960 to 2022.

As shown in *fig.11*, we can observe that Ireland had a clear declining pattern from 1970 to 2005, however, since then, Ireland’s population deaths has been dramatically rising.

# **Machine Learning model**

As previously mentioned in our approach, we will use Random Forest Regressor from sklearn as our Machine Learning model for this first stage of the analysis.

To be able to obtain the best results possible for our dataset, we will use a tool called Grid Search CV from sklearn as well, this tool will help us to iterate multiple values of hyperparameters of our model searching for the best combination to get the highest R2 score attainable, being R2 score our metric to measure our model.

The R2 score, also known as the coefficient of determination, is a metric that measures how well a regression model fits the actual data, it gives us a value on a scale of 0 to 1, a value of 1 indicates that the model perfectly predicts values in the target field (our dependent variable) and a value of 0 indicates that the model has no predictive value (IBM, n.d.).

For the sake of science, we will train our model in 3 different proportions of our data to compare them, the results were the following:

r2\_score result (**training**)(90% training): 0.996

r2\_score result (**testing**)(90% training): 0.991

r2\_score result (**training**)(80% training): 0.996

r2\_score result (**testing**)(80% training): 0.994

r2\_score result (**training**)(70% training): 0.996

r2\_score result (**testing**)(70% training): 0.982

By only looking at our R2 scores, it would seem that our model with 80% of training data is the one that will perform the best in regards to predicting the population in Ireland, however, lets explore the features importance of our model:

90% training:

Year: 0.979

Births: 0.00826

Deaths: 0.00820

Marriages: 0.00378

80% training

Year: 0.954

Births: 0.00961

Deaths: 0.0320

Marriages: 0.00346

70% training

Year: 0.953

Births: 0.00851

Deaths: 0.0347

Marriages: 0.00345

We can observe that our ‘year’ feature is the main determining factor in our data, nevertheless, knowing the rapid rising of deaths in Ireland, we can understand why our model is taking ‘deaths’ feature as the second more important feature to predict the population in Ireland, in future years we can anticipate that this feature will gain more relevance.

Gráfico

Descripción generada automáticamente

Figure : Feature importance from our Random Forest Regressor model.

# **Github**

https://github.com/LeopoldoCCT/Strategic\_Thinking\_

# **References**

Central Statistics Office (n.d.). *Copyright Policy - CSO - Central Statistics Office*. [online] www.cso.ie. Available at: https://www.cso.ie/en/aboutus/whoweare/copyrightpolicy/ [Accessed 27 Oct. 2023].

Central Statistics Office, Ireland (2023). *Total Births, Deaths and Marriages Registered*. [online] Data.cso.ie. Available at: https://data.cso.ie/table/VSQ04 [Accessed 12 Oct. 2023].

Honohan, P. (2021). Is Ireland really the most prosperous country in Europe? *Central Bank of Ireland*, [online] 2021(1). Available at: https://www.centralbank.ie/docs/default-source/publications/economic-letters/vol-2021-no-1-is-ireland-really-the-most-prosperous-country-in-europe.pdf [Accessed 24 Oct. 2023].

IBM (n.d.). *R2*. [online] www.ibm.com. Available at: https://www.ibm.com/docs/en/cognos-analytics/11.1.0?topic=terms-r2 [Accessed 17 Dec. 2023].

IBM (n.d.). *What is Random Forest? | IBM*. [online] www.ibm.com. Available at: https://www.ibm.com/topics/random-forest [Accessed 16 Dec. 2023].

Organisation for Economic Co-operation and Development (2023). *Fertility rates - OECD data*. [online] OECD. Available at: https://data.oecd.org/pop/fertility-rates.htm [Accessed 13 Oct. 2023].

United Nations (2022a). *World Population Prospects - Population Division - United Nations*. [online] population.un.org. Available at: https://population.un.org/wpp/DefinitionOfProjectionScenarios/ [Accessed 11 Dec. 2023].

United Nations (2022b). *World Population Prospects - Population Division - United Nations*. [online] population.un.org. Available at: https://population.un.org/wpp/DefinitionOfProjectionScenarios/ [Accessed 16 Dec. 2023].

United Nations (2023a). *About Us*. [online] United Nations. Available at: https://www.un.org/en/about-us [Accessed 16 Dec. 2023].

United Nations (2023b). *Population*. [online] United Nations. Available at: https://www.un.org/en/global-issues/population [Accessed 24 Oct. 2023].

Wilmoth, J., Menozzi, C. and Bassarsky, L. (2022). *Why population growth matters for sustainable development POLICY BRIEF NO 130 Key messages*. [online] Available at: https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesa\_pd\_2022\_policy\_brief\_population\_growth.pdf [Accessed 27 Oct. 2023].

World101 from The Council on Foreign Relations (2022). *Global Population Growth Is Slowing Down*. [online] World101 from the Council on Foreign Relations. Available at: https://world101.cfr.org/global-era-issues/development/global-population-growth-slowing-down#:~:text=Studies%20differ%20about%20the%20exact [Accessed 24 Oct. 2023].