# Analysis of Social Housing, World Bank Statistics and Population Projections in Ireland.

Revanth Vijay Kumar
Masters in Data Analytics
National College Of Ireland
Dublin, Ireland
x21218374@student.ncirl.ie

Tharunendra Varma
Masters in Data Analytics
National College Of Ireland
Dublin, Ireland
x21194815@student.ncirl.ie

Madhuri Dhatrak
Masters in Data Analytics
National College Of Ireland
Dublin, Ireland
x21208808@student.ncirl.ie

***Abstract*— In relation to the social and economic context in the EU, this paper compares the population and social housing construction in Ireland. This study seeks to comprehend the current economic and population situation considering recent increases in rental costs and a lack of housing in Ireland. In terms of GDP and population, Ireland comes in ninth place in Europe. Ireland's unemployment rate in 2021 is 6.3%. There will be 5 million people living in Ireland in 2021, and 315,00 of them will be unemployed. Since 2016, the construction of social housing has produced 16,745 units.**

***Keywords—Ireland, Population, Housing***

## I. INTRODUCTION

This study exhibits the use of data science tools, with real world data, for fact finding, exploration and risk mitigation. The data selected for this is focused on socio economic indicators that pertain to population, education, unemployment, and housing. The motivation to select the area of study is to use official statistics published by central agencies to understand the current situation and future scenario of accommodation in Ireland with a focus on those who are dependent on the state for accommodation. Financial and social conditions are among the primary factors leading to homelessness [1]. Using the world bank data, this paper aims to capture and study the various indicators available for social and financial indexes and understand Ireland's position in comparison to other European countries. Data published by the Ireland government is used to understand the situation in Ireland. The status update on the social housing construction initiative by the is studied to know where the country stands in its readiness to accommodate people who are dependent on the state. Furthermore, scenario wise estimates of population by the government are used as a third dataset for a detailed understanding of the Irish population in the times to come.

Using this information and leveraging the data base tools MongoDB and Postgres, this study builds and accessible database for analyzing this information. Using this infrastructure, the paper aims at answering the following questions

*1) Ireland in comparison with other EU menber nations in terms of in population, gross domestic product(GDP) and unemployment*

*2) Which age groups in Ireland have the largest population projections*

*3) Identify the current state of social housing projects in Ireland*

The current situation makes the relevance of these questions extremely important from a social welfare perspective. It also provides insights for future planning and be used as a guide for policy evaluation and targeted action. The charts and visualizations are developed to provide a an easy to under representation of the above questions

## II. RELATED WORK

In the paper [2], the authors exhibit the use of various data mining techniques to identify association in public health metrics and economic indicators using the World Bank Statistics. Rapid miner, a proprietary tool for data mining and analysis is used for this research. The study exhibits the use of correlation functions and plot charts to show the association between the different metrics. This study however does not offer as a framework with the use of said licensed tool and has a central focus on medical statistics. It is used as a starting point to expand the use of World Bank data to other aspects of social welfare like accommodation. The paper also drops the country data and aggregates the available data to get an understanding of the global situation which will generally stand true for countries with statistics closer to the global average and be inaccurate for smaller countries.

This premise of health and economic interaction is taken a step further in paper [3]. It is different however in its scope as the study if focused on data from Finland and Japan exclusively. The understanding and evaluation of health is equated to life expectancy. The specific selection of the above countries is based on their similarities as countries with relatively small landmass and higher earning capacity than the global average. The authors offer a model with a unilateral goal of predicting life expectancy but does not shed light on other aspects that quantify the quality of life.

A study on the impact of economic indicators on housing is performed in paper [4]. This analysis is performed on data from China where the authors perform an impact analysis on housing prices. This paper highlights how real estate pricing is highly dependent on economic fundamentals. It also and identifies as earning capacity, availability and cost of construction as primary factors impacting demand. The difference in the government and housing policies and structure of China in comparison with most democratic countries impede the direct application of these conclusions to other markets. Also, a price point analysis makes affordability rather than availability as the central theme of this analysis.

The authors of paper [5] offer a statistical simulation of the housing market. It elaborately covers various scenarios such as people who must to stay versus people who stay by choice and calculate the variation in housing demand based on different levels of amenities, proximity to the city, occupation. The result is a framework that offers an estimated distribution of empty house for a location with certain amenities. This hypothesis however is not tested in a real-world scenario. To

do so, a database system is required that can be implemented in a real world.

While the above papers discuss applications of data technologies for housing and economy, the paper [1] offers a model that uses similar indicators to estimate the susceptibility of long-term homelessness in an individual. The study is conducted in America and the final model proposed uses age, employment, race, education, and physical disabilities as the significant factors for prediction. The results show that this study is 13.64% better than random selection at its predictions. This approach can be emulated with other sample groups to identify high risk individuals.

A unique approach to accommodation planning is offered in [6] which advocates the mixing of high-income and low-income groups for the development of public rental housing. The concepts of this study can be experimented with in a social housing setup. The study suggests that a heterogenous mix of income groups facilitates the development of the are and also the low-income groups. This could possibly be used in social housing interspersed with rental housing to prevent long term homelessness and encourage capable individual to move to rented accommodation.

With age being a significant factor in most of the above studies, paper [7] is studied to understand the prediction of the population of different age based. This facilitates efficient planning of resources. This study only concentrates on China which has very different population, growth, and migration trends in comparison with North America or Europe. While the model might not perform as expected when directly applied to these areas, the principles could be used, and new models be generated to the above regions. The indicators identified can be used as a starting point to know if the factors used for prediction age group population in China are the same for countries in Europe

## III. METHODOLOGY

### A. Datasets

#### *a) Social Housing Construction status*

This dataset is collected from the public access government data portal data.gov.ie. It raw json has 14604 rows and 7 columns. The report is published in Q3 2021. It contains the status of the social housing construction projects across the Republic of Ireland. The data details the projects starting from Q1 2016 up until Q3 2021. 26 counties and 5 cities comprise of the total area divisions in the dataset. The data has information about the funding package, welfare scheme under which the construction is done, the number of units being built, the approving body for the specific construction and the stages of completion with dates.

#### *b) World Bank Statistics*

This dataset is compiled from the World Bank Open Data Catalog API. The open-source python package WBGAPI [8] was used to create this dataset. This package provides the functionality to interact with the data API using a python client. The 18 indicators selected for building this dataset are described in Table I. The curated dataset has 30132 rows and 4 columns, with annual data starting from 1960 to 2021.

TABLE I. WORLD BANK INDICATORS

| ***Indicator*** | ***Description*** |
|---|---|
| SP.POP.TOTL | Total Population |
| SP.POP.TOTL.FE.IN | Female Population |
| SP.POP.TOTL.MA.IN | Male Population |
| SP.DYN.CBRT.IN | Birth Rate |
| SP.DYN.CDRT.IN | Death Rate |
| SE.COM.DURS | Compulsory Education Dur. |
| SL.IND.EMPL.ZS | Employment in Industry(%) |
| SL.AGR.EMPL.ZS | Employment in Agriculture(%) |
| SL.AGR.EMPL.FE.ZS | Female Employment in Agriculture(%) |
| SL.IND.EMPL.FE.ZS | Female Employment in Industry(%) |
| SL.UEM.TOTL.ZS | Unemployment(%) |
| NY.GDP.MKTP.CD | GDP in USD |
| NY.ADJ.NNTY.PC.KD.ZG | National Income per Capita |
| NY.GSR.NFCY.CD | Net income from Abroad |
| NV.AGR.TOTL.CD | Agriculture value added(in USD) |
| EG.USE.ELEC.KH.PC | Electric Power Consumption(kWH per capita) |
| EG.FEC.RNEW.ZS | Renewable Energy Consumption (%) |
| EG.USE.COMM.FO.ZS | Fossil Fuel Consumption (%) |

#### *c) Economic and Social Research Institue(ERSI) Population Projection*

This dataset is collected from the public access government data portal data.gov.ie. The raw json file has 12,524 rows and 24 columns. The range of the projection is 20 years starting from 2020 up until 2040. The dataset has age and area wise projections for 31 local authority areas with an age span from 0-100. The projections are made based on 4 scenarios

- Baseline projections made based on current data that quotes net inward immigration at 15,000 per annum
- 50:50 projections that assume net immigration at the same level as the baseline but with an additional assumption that the population distribution would be 50:50 between rural and urban areas
- High Migration scenario projections with net inward immigration assumed to be 30,000 per annum
- Low Immigration scenario projections with net inward immigration assumer to be 5,000 per annum.

### B. Data Processing

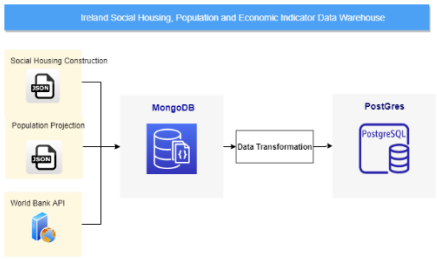



Fig. 1. Data Process

This project applies Extract, Transform, Load (ETL) methodology. Fig. 1 describes the process flow used to implement the methodology.

*a) Extract*

In the extract phase data is collected programmatically using python notebooks. The data for social housing construction and population projection is available in json files shared by the data.gov.ie website. These files are read using the json and pandas python libraries. The world bank data catalog has an application program interface (API) available to access its data. Open-source python library web API is used to retrieve this data. As the collected data is in different formats from multiple sources a semi structured data base is used as a landing server where the collected data is stored. MongoDB is used to create the landing point for the different data. It is selected as it provides a NoSQL database to store different types of data and files or varying sizes. Collections are created for population projection, social housing construction and EU economic indicators respectively. It provides an easy to access environment to query the unprocessed data

*b) Transform*

In the transform stage the collected data is processed into a relational database structure. The collections created in MongoDB in the extract stage are read programmatically and the following steps are performed in this stage

- Data cleaning is performed on the collected data. Null values obtained from the world bank data are removed. Data from the gov.ie portal contained encoding errors in the names due to presence of non-English alphabets. These were removed to prevent errors while matching the names column as string. Duplicates are removed from the datasets and the collections are transformed into a row column structure using pandas' data frames.
- Data structuring is performed to convert the data. The social housing construction data is converted from wide format with each to long column by adding stage and date columns that identify the date when a particular stage was completed for a project. Similar processing is performed for population projection. As it is in hierarchical structure where the top-down breakdown is for the four scenarios (Baseline, 50:50, Low Migration and High Migration) which have all the locations which in turn have the population projections for each age from 1 to hundred. This data is flattened by adding scenario, location and age columns for the projection values.
- Calculated columns are created to generate the date column by combining the quarter and year columns. The total population for an area us calculated by adding the projections in all age groups in that area. Stage wise completion date column is calculated for the social housing constructions dataset.
- Data formatting is performed to convert the data into correct datatypes before loading. The age column in the population projection dataset is converted to integer data type. All the date columns are cast into datetime data type

*c) Load*

In this stage the formatted data is loaded in postgres SQL database tables.

## IV. Results and Evaluation

The created database is used to perform analytics on the available data.

The correlation matrix in Fig 2 highlights the correlation between the different macroeconomic indicators. A high correlation is observed between the total population index with education, agricultural value addition, power usage, fossil fuel consumption, usage of electricity and national overseas income. These are understandable as the usage of resources and domestic and foreign produce is driven by the population.

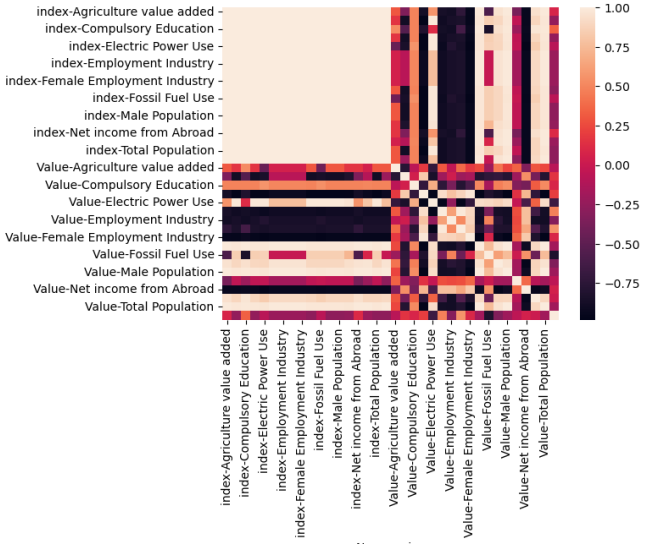



Fig. 2. Economic Indicators Correlation Matrix

The population trend in Ireland starting from 2009 has been visualized in Fig. 3. The national population of Ireland in 2009 was slightly higher than 4.5 million. In the next 10 years, by 2019 the population had risen to slightly higer than 4.9 million. It can be observed that the population continued to rise during the Covid 19 pandemic years and now stands at 5 million. The poplution growth rate can be calculauted at aroun 10 percent per annum. Inspite of the steady rise in numbers, Ireland falls in the countries with the lowest population when compared with other European nations

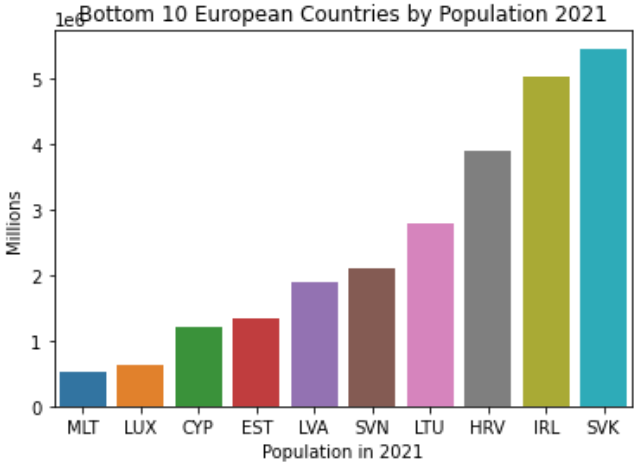



Fig. 3. Countries with lowest population

A bar chart showing Ireland in comparison with other European nations. It selects the countries with the lowest population and with a current population of approximately 5 million people, Ireland ranks 9$^{th}$ in the list of countires with

th lease poplation. With about half a million people, Malta comes at the bottom as the country with the least population.

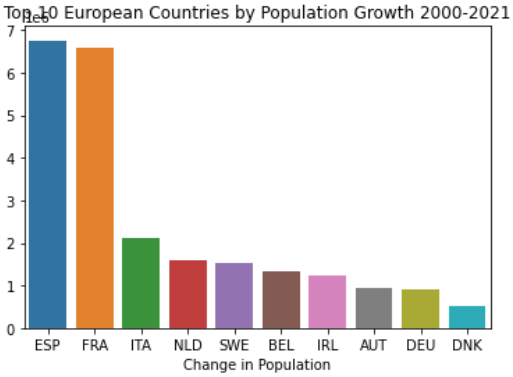



Fig 4. European Countries by Population

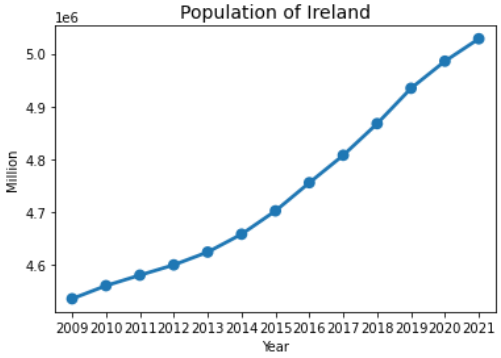



Fig 5. Population Of Ireland

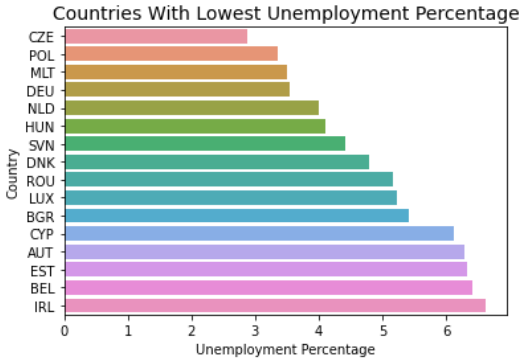



Fig 6. Countries with lowest Unemployment Percentage

The gross domestic product of the member nations is visualized in that selects the top 10 countries with the highest GDP in Europe. While being the 9th smallest country by population, Ireland ranks 9th in the GDP list. In the year 2021, the GDP of Ireland was around 500 billion USD. This is comparatively higher when compared with other European countries nations with a similar population. For example, Slovakia has a population of 5.1 million which is close to Ireland but a GDP of 100 billion USD which is about 1/5th of Ireland's GDP.

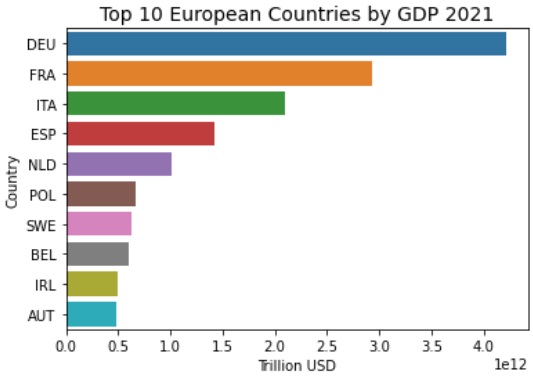



Figure 7. European Countries by GDP

Ireland features on the list of European nations with lowest unemployment at the 16th position. With an unemployment percentage of 6.3 it falls in the middle range when compared to other European countries. With an unemployment percentage of 3.2 Czech Republic ranks as the country with the lowest unemployment and Greece tops the list with 14.2 percent.

The baseline, 50:50 city, high migration, and low migration scenarios for population projections for the year 2040, shown in Figure 9.

It is observed that the different scenarios have similar trends, but the magnitude varies based on immigration. In the baseline prediction it can be observed that more than half the local areas have a population of less than 4000. This is also observed in the high migration scenario. It is found that for the high population areas, the baseline prediction stands slightly higher than the 10,000 marks whereas the high immigration scenario is around 12,000. From the data it is predicted that the total population of Ireland in the year 2040 would be 5.6 million (baseline), 5.5 million in the case of low immigration and 6 million in the case of high.

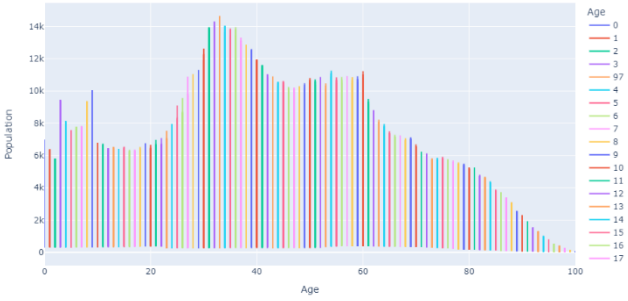



Fig.9 Ireland Population projection for Local Authority

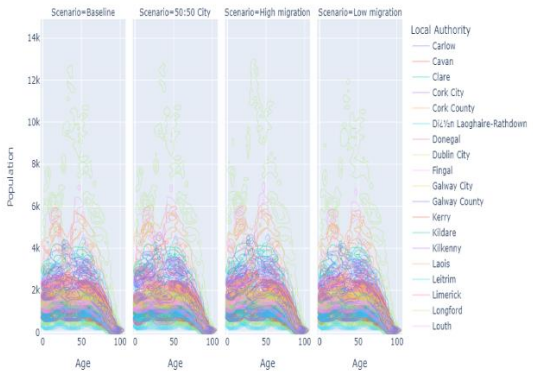
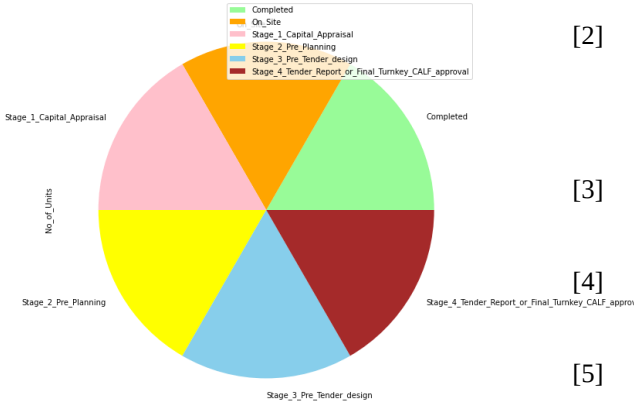
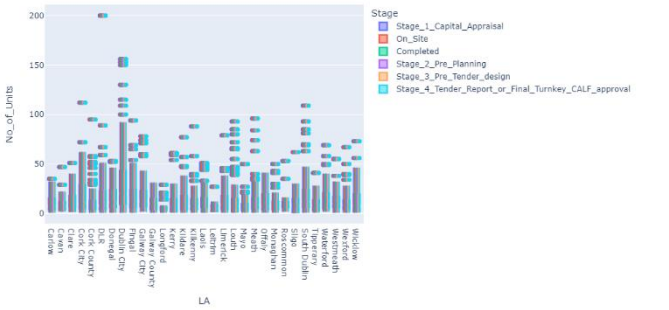



Fig 9. Ireland population projection

Fig. 10 is a pie chart showing the current stage wise distribution of the social housing construction projects. The report has information of projects starting from the first quarter of 2016. Construction is being done in 31 areas with the highest number of units in Dublin City at 4358 units and the lowest at Leitrim at 130 units. The projects are divided in 3 stages, Completed, On Site and Tender report or Final Turnkey (CALF approved). As per the Q3 2021 report, 16,745 unites have been completed, 9028 units are on site and 1053 units are Tender Report or Final Turnkey Stage




## V. Conclusion and Future Works

The data studied in this project highlights Ireland's growing economy and population. A relational database system allows the analysis of the various metrics and generate easy to interpret visualizations. With a growing economy and a steady population increase rate of 10 percent, Ireland will see a rise in its population in the coming decades. An unemployment rate of 6.3 for a population of 5.5 million as per baseline predictions suggests that around 315,000 people in Ireland will not have a job and might be vulnerable to homelessness. The current state of the social housing construction shows that 62 percent of projects currently undertaken have been completed. But with a total construction of 26,826 of which 16,745 have been completed, the large number of unemployed people might strain the system

Further study could be conducted on this topic to better understand the causes of unemployment and a better estimation of the demand for social housing construction. The data sources could be periodically refreshed to have a comparative update on the situation as it unfolds. The database created in this project could be used to generate datasets for machine learning algorithms for the prediction of housing demand, population, and homelessness

## References


[1] S. Purao, M. Garfield, X. Gu, and P. Bhetwal, "Predicting the slide to long-term homelessness: Model and validation," in *Proceedings - 21st IEEE Conference on Business Informatics, CBI 2019*, Jul. 2019, vol. 1, pp. 31–40. doi: 10.1109/CBI.2019.00011.

[2] H. Arif and J. Ali, "Using Data Mining Techniques on World Bank Statistics," in *Proceedings on 2018 International Conference on Advances in Computing and Communication Engineering, ICACCE 2018*, Aug. 2018, pp. 439–444. doi: 10.1109/ICACCE.2018.8441744.

[3] N. Kerdprasop, K. Kerdprasop, and P. Chuaybamroong, "Ensemble Modeling Method to Predict Life Expectancy of Population in High-Income Countries: Japan and Finland."

[4] H. Liu and Y. Shen, "Housing Prices and General Economic Conditions: An Analysis of Chinese New Dwelling Market," vol. 10, no. 3, 2005.

[5] H. Wang and C. J. Chang, "Simulation of housing market dynamics: Amenity distribution and housing vacancy," in *Proceedings of the 2013 Winter Simulation Conference - Simulation: Making Decisions in a Complex World, WSC 2013*, 2013, pp. 1673–1684. doi: 10.1109/WSC.2013.6721549.

[6] Y. Xu and H. Ying, "The planning mode of adjusting the proportion of public rental housing - Based on the concept of 'mixed-incoming housing,'" in *2011 International Conference on Electric Technology and Civil Engineering, ICETCE 2011 - Proceedings*, 2011, pp. 1708–1711. doi: 10.1109/ICETCE.2011.5776152.

[7] X. Yang, B. Chen, C. Chen, and Y. Wang, "Dynamic model of population prediction based on Aging Chains," in *Proceedings of the 28th Chinese Control and Decision Conference, CCDC 2016*, Aug. 2016, pp. 216–219. doi: 10.1109/CCDC.2016.7530983.

[8] Tim Herzog, "WBGAPI," *https://pypi.org/project/wbgapi/*, Jul. 22, 2022.