

# Exploring and forecasting income disparities across Irish counties: A comprehensive socioeconomic research

MSc Research Project MSc in Data Analytics

Rohit Puranik Student ID: x22165967

School of Computing National College of Ireland

Supervisor:

Prof. Arjun Chikkankod

#### National College of Ireland



#### **MSc Project Submission Sheet**

School	of	Com	putina
501001	<b>U</b> I	COM	pating

Student Name:	Rohit Puranik							
Student ID:	X22165967							
Programme:	Data Analytics	Year:	2023-24					
Module:	MSc Research Project							
Supervisor:	Prof. Arjun Chikkankod							
Submission Due								
Date:	31 <sup>th</sup> Jan 2024							
Project Title:	Exploring and forecasting income disparities across Irish							
	counties: A comprehensive socioeconomic research							
	· · · · · · · · · · · · · · · · · · ·							
Word Count:	6000 <b>Page Count</b> 27							

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

<u>ALL</u> internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	Rohit Puranik
Date:	31 <sup>th</sup> Jan 2024

#### PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

# Exploring and forecasting income disparities across Irish counties: A comprehensive socioeconomic research

# Rohit Puranik x22165967

#### Abstract

Income inequality between urban and rural populations and among the various counties of Ireland constitutes a significant social and economic challenge for both society and policymakers. This research seeks to quantify and analyze these disparities within urban and rural counties of Ireland, utilizing two socioeconomic factors: Disposable Income per Person and Disposable Income per Person (excluding Rent). Drawing upon data from the Central Statistics Office of Ireland, this research aims to offer a comprehensive understanding of the extent and nature of income inequalities among the counties of Ireland.

Within the context of Ireland's historical and economic background, this research delves into the urban and rural classification of counties. The project underscores the global concern of income inequality and highlights a research gap specific to Ireland. To address this issue, visualization tools like Power BI and Python visualization libraries are employed. Additionally, forecasting time series machine learning models are utilized to predict future income disparities among the counties in the region. Throughout this research, there is also a focus on the impact of the Potato Famine (1845-1852).

To ascertain the significance of changes in income disparity, statistical tools such as the Lorenz Curve and Gini Coefficient are implemented. These tools are further validated through statistical tests such as t-tests and Wilcoxon rank-sum tests. The overarching goal of this research is to provide valuable insights for policymakers, contributing to the establishment of a more equitable society and fostering a sustainable Irish economy.

**Keywords**— Irish Economy, Income Disparity, Urban & Rural Counties, Forecasting Disparity, Potato Famine

# **1** Introduction

#### **1.1 Background and Motivation**

Income inequality is a major problem in society because of the damage it does to social dynamics and the economy. This disparity is especially noticeable between rural and urban areas, and it varies greatly throughout counties, making it difficult for politicians and the general public to address the issue. In order to foster a future that is more equal and sustainable, it is essential to address these inequities. This Research investigates income

disparity in Ireland in depth, first looking at how it varies across all counties and then differentiating between rural and urban regions.

'The Penal Laws' and the tragic Great Famine of the nineteenth century are just two of the defining moments in Ireland's troubled past. The island nation is located at the very point of the European landmass. A lengthy history of inequality has its roots in these events, which were marked by severe limitations and pervasive poverty. The Anglo-Irish War (1919–1921) and the subsequent creation of an independent Irish state were the climaxes of the nineteenth-century rise of Irish patriotism. After Ireland's 1973 accession to the European Economic Zone, the country's healthcare and infrastructure were among the best in the world, thanks to the EU's push for economic growth.

Significant changes in socioeconomic aspects have been spurred by recent obstacles, such as a high cost of living and the impact of the COVID-19 epidemic, despite Ireland's booming economy and active EU participation. For this reason, indicators like income inequality must be central to any comprehensive social assessment of the economy. Societal dynamics and economic stability are both impacted by income disparity. The increasing wealth gap between rural and urban areas, as well as across counties, is a serious social and economic problem for both society and policymakers.

If we want to build a future that is more just and sustainable, we must eliminate these inequities. The income gap in Ireland is the subject of this research, which starts by looking at the situation in every county before comparing and contrasting the wealth of rural and urban regions. Scott (2017) lays up a framework for classifying Ireland's counties as either urban or rural as per below.

#### **Classification of Rural and Urban Counties in Ireland:**

There is a high concentration of information technology occupations in the huge metropolitan region of Co. Dublin, which includes Ireland's capital city. A small number of software and information technology companies call the second-largest county in southern Ireland, the city of Cork. Co. Galway is an important western urban centre, a major city, and a university hub. Due to its lack of proximity to any seaports, the Mid-Ireland city of Limerick is considered landlocked. Due to its stunning history and beautiful shoreline, a historic city in County Waterford is well-known as a tourist attraction. The following counties are not part of the metropolitan areas but are considered rural or agricultural areas: Carlow, Cavan, Clare, Donegal, Kerry, Kildare, Kilkenny, Laois, Leitrim, Longford, Louth, Mayo, Wexford, Meath, Monaghan, Offaly, Roscommon, Sligo, Tipperary, Westmeath, and Wicklow.

Research has looked at a quantifiable criteria such as 'Disposable Income per Person (excluding Rent),' 'Disposable Income per Person for each county in Ireland, according to the Central Statistics Office. In order to answer the Research question, these data points will be used.

In addition to that this research also forecast Income & population for next 10 years till 2033 for all the counties of Ireland which would further help policy makers to foresee future.

## 1.2 Research Question

The primary objective of this research is to provide a nuanced understanding of the extent and nature of income disparities among the counties of Ireland and forecasting the same for next 10 years till 2033 therefore specific research questions are included:

RQ: Is income in-equality present between the various counties of Ireland and is income evenly distributed between the rural & urban population of Ireland considering the socio economic factors such as Disposable Income per person & Disposable Income per Person excluding rent and where this trend would go in next 10 years?

Sub RQ: Which time series forecasting machine learning model among LSTM, ARIMA Model Grid Search and Forecasting & ARIMA Model Forecasting with Fixed Parameters performs well for forecasting population for various counties of Ireland?

Sub RQ: What would be the population & disposable income per person for all the counties in Ireland for next 10 years till 2033?

Sub RQ: What is the impact of historical events, such as the Potato Famine (1845-1852) is influencing the present-day population forecasting models for Ireland?

## **1.3 Research Objective**

This study makes a contribution, to the existing socio literature by addressing a significant research gap related to income inequality in Ireland. By utilizing visualization tools like Power BI and Python visualization libraries, along with machine learning models for time series forecasting this research provides an understanding of income disparities both in the present and over the next decade. Additionally incorporating events such as the Potato Famine adds an aspect to the analysis enhancing our knowledge of its long term impact on population distribution.

The following sections of this report will explore Irelands economic background outline how counties are classified as urban or rural and elaborate on the methodologies used for data analysis. Detailed discussions will be provided on visualization tools and machine learning models with emphasis on predicting income and population trends for all counties in Ireland over the ten years till 2033. These predictions will play a role in identifying disparities between rural & metro counties. The findings will be validated through tools, like the Lorenz Curve, Gini Coefficient well as hypothesis testing using t tests and Wilcoxon rank sum tests.

In summary this research aims to provide policymakers with insights that can contribute to creating an equitable society and fostering sustainable economic development in Ireland. In addition to that in the below table are the objective which this research would achieve

Sr.	Description
1	Critical review of literature on the previous studies on Income disparity for various geography and how it was measured
2	Find the best Time series model among LSTM, ARIMA Model Grid Search and Forecasting & ARIMA Model Forecasting with Fixed Parameters for forecasting Income & Population of Ireland
3	Checking the impact of Potato famine on machine learning models for forecasting population of Ireland
4	Implementation of Time series model determined in #2 for forecasting Income for all 26 Counties of Ireland for next 10 years
5	Implementation of Time series model determined in #2 for forecasting Population for all 26 Counties of Ireland for next 10 years
6	Checking the income inequality among all the counties of Ireland starting 2000 till 2022 and where it will go by 2033
6.a	Considering Population of each county & Disposable income
6.b	Considering Population of each county & Disposable income excluding Rent
6.c	Without considering Population for the counties just by the Disposable income per person
6.d	Without considering Population for the counties just by the Disposable income excluding Rent
7	Checking the income inequality among rural and the urban population of Ireland starting 2000 till 2022 and where it will go by 2033
8	Statistically verifying if the change calculated in #6 & #7 which are observed from 2000 till 2022 and where it will go by 2033 are statically significant or not

Table (1) Research Objective

# 2 Related Work

Nowadays people, around the world are deeply concerned about income inequality. This has sparked a deal of interest among researchers who want to understand how factors like population, education levels and income disparities are interconnected in regions and ethnic groups. Several studies have been conducted to compare aspects.

Many studies had focused on the issues of growth and income inequalities in countries like Brazil, Russia, India and China (commonly known as the B.R.I.C. Countries). Moreover recent research has also been carried out to investigate and identify irregularities in income distribution within Ukraine.

What surprises many is that despite the attention given to income inequality in nations there seems to be a lack of research specifically addressing this topic in Ireland. The absence of investigation emphasizes the need for a study called "Income Disparities in the Urban and Rural Population of Ireland." This study holds importance given the economic impact of socio-economic factors. To bridge this gap our research paper will employ tools such as the Lorenz Curve, Gini Coefficient and ARIMA models. These tools will facilitate forecasting trends in disparity while analysing data points which are then proven by statistical tests.

Analytical tools serve as resources, for presenting information logically. These measures can offer insights, to government policymakers helping them gain an understanding of the income gaps between rural areas in Ireland. This knowledge will contribute to the development of an sustainable society as well as a thriving economy.

Extensive research has been conducted by scholars and experts in the field on income inequality. Numerous studies have been undertaken to comprehend the causes, consequences and potential solutions to this issue. The primary focus of these research endeavours is to shed light on the distribution of income among individuals and various groups across regions and countries.

Researchers from around the world have dedicated years to studying income inequality. They have conducted studies aiming to explore the link between education levels and disparities in income or expenditure. This research seeks to identify inequalities and income disparities within regions or ethnic groups. A notable study conducted by Das Gupta and Das (2014) focused on analysing and comparing B.R.I.C countries in terms of growth and income inequality. Their research examined factors such as government role and corruption, within these nations.

In a study conducted by Headey (2021) it was found that major developed economies, like the United Kingdom, Sweden and the United States have expressed concerns regarding income inequality (Alford, Reilly, & United States, 2012).

The Ukrainian economy has faced challenges due to war resulting in an economically unstable situation. Recent studies have also shown that the ongoing pandemic has further contributed to irregularities in income distribution among the population (Pashchenko and Zharikova 2021).

To accurately measure and address this issue researchers in Malaysia utilized the A ARDL testing approach (Goh et al., 2023). This approach incorporates the Gini coefficient to examine "Income Inequality, Income Growth and Government Redistribution in Malaysia; What Do We Know in the Long Run?".

However there is currently a lack of research conducted in Ireland to determine if there are any income disparities between rural households. Thus it is crucial to conduct a study on this topic during times of inflation. The objective is to identify any income disparities, within Irelands rural populations. Several scholars and institutions have conducted research, on income inequality in Ireland. Their studies aim to gain insights into the extent of income disparity and its impact on groups. These researchers have explored factors, including education, occupation, gender and regional differences to uncover the causes of income disparities in Ireland. By analysing data from surveys, censuses and other sources

Recently a comprehensive research study carried out by Maynooth University has focused on analysing the distribution of household incomes in Ireland. This study specifically examines the aspect of income inequality. Provides valuable insights into this crucial indicator. The research conducted by Walsh in 2023 sheds light on how social and economic changes have unfolded across regions of Ireland over the few decades. The findings indicate that there has been a decrease in income inequality between rural households. However it is important to note that this study relies on the parameter of 'median household income' from the CSO (Connolly, 2023) dataset.

Nevertheless it is worth mentioning that certain parameters such as 'Compensation and wages' 'Disposable Income per Person (excluding Rent)' and 'Disposable Income per Person' available, in the RAA02 dataset published by CSO (Connolly, 2023) were not considered during this analysis.

Furthermore there was no analysis conducted by Walsh (2023) utilizing the Lorenz Curve and Gini Coefficient. Was any qualitative analysis performed? Additionally the consideration of ARIMA (Autoregressive Integrated Moving Average) and SARIMA (Seasonal Autoregressive Integrated Moving Average) forecasting models was absent, in the study conducted by Goh et al. (2023).

This research paper aims to delve into factors specifically focusing on the calculation of the Lorenz Curve and Gini Coefficient. Moreover it will explore the utilization of ARIMA models to analyze income disparity between rural counties in Ireland.

In this topic we will be discussing the evaluation of income disparity using two measures. The Lorenz Curve and the Gini Coefficient. These measures are widely used to analyze and comprehend income distribution within a population. The Lorenz curve, which illustrates how income or wealth is distributed among a population held prominence in economics until the 1950s. It gained popularity as a concept frequently discussed within this field (Paglin, 1975 p. 598). As an approach utilizing the Gini coefficient has become increasingly popular, for assessing income inequality within populations or datasets.

The scale ranges from '0', to '1' where '0' signifies equality. It's like a world where everyone has an income. Conversely '1' represents inequality, where all the income is concentrated in one person or group. This is sometimes referred to as a scenario where a country's rich but with individuals (Muga, 2015). Gastwirth (1972 p. 306) argues that a significant portion of income inequality can be attributed to the Lorenz curve.

However due to its nature the Gini Coefficient has proven advantageous than the Lorenz Curve for calculating income inequality. It is widely recognized as the measure of inequality (Morgan, 1962 p. 270). Therefore we will utilize the Gini Coefficient to assess and compare income disparities between rural populations in counties across Ireland.

Analysing and understanding how the Gini coefficient varies across counties, in Ireland is an area of interest.

To examine the factors that contribute to inequality, in rural areas we can consider parameters. For instance a study titled "EDUCATION IN INCOME INEQUALITY IN BANGLADESH" (Mahmud and Akita 2018) utilizes Gini Coefficients to calculate income inequality for both rural populations in Bangladesh. This study confirms the presence of inequality in developing nations like Bangladesh as established by the American Sociological Review paper by Evans and Timberlake (1980). However when it comes to Ireland—a developed country with a standard of living advanced infrastructure and a strong economy it falls under a different context. As Ireland is a member of the European Union (EU) and the Organisation for Economic Co operation and Development (OECD) which consist of high income countries understanding income equality nuances between counties and rural counties becomes unfamiliar territory. Hence calculating the Gini Coefficient based on population as one of the weighted factors provides insights into inequality.

In this research paper our focus will be on analysing income disparities present between the counties of Ireland and to check is there any income disparity present between the urban & rural counties, in Ireland.We'll utilize a dataset obtained from CSO (Connolly, 2023). Employ statistical techniques, like the Gini Coefficient, ARIMA and forecasting models. Furthermore we'll utilize visualization tools to generate charts comparing the population and income trends of counties over time. We will also take into account the categorization of counties, into metro & rural regions.

# 3 Research Methodology

# **3.1** Determine the best performing model for forecasting Population for Irish geography

Following steps are performed in "Determining the best Forecasting Method.ipynb" script



Figure (1) Determining the best Forecasting Method.ipynb flow

#### 3.1.1 Libraries Import:

The script begins by importing necessary libraries such as NumPy, Pandas, Matplotlib, Seaborn, Plotly, SimpleImputer, MinMaxScaler, statsmodels, LabelEncoder, OneHotEncoder, and itertools.

#### 3.1.2 Read Population Dataset:

The script reads a population dataset from an Excel file using Pandas.

#### 3.1.3 Data Filtering:

Filters the population dataset based on criteria like sex and county.

#### 3.1.4 Grid Search ARIMA:

Defines a function (grid\_search\_ARIMA) for performing a grid search to find the best ARIMA parameters. It also utilizes the ARIMA model from statsmodels. Performs grid search for each county and stores the results in a DataFrame.

#### 3.1.5 Forecasting Using LSTM:

Imports necessary libraries for LSTM forecasting.

Defines functions for training an LSTM model and forecasting population for each county using LSTM.Performs training and forecasting for each county and stores the results in a DataFrame.

#### **3.1.6 ARIMA Forecast:**

Defines a function (train\_and\_forecast) to train an ARIMA model and forecast population for each county.Performs training and forecasting for each county and stores the results in a DataFrame.In this model p,d,q parameters are kept fixed as (1,1,1) and passed.

#### 3.1.7 Result Output:

The results for grid search ARIMA, LSTM forecasting, and ARIMA forecasting are saved in separate Excel files. Essentially MAE, RMSE & MAPE values for each county is stored for each type of model into .xlsx and used to determine the performance of the 3 models it

Drid_Search_ARIMA				ARIMA_Forecast					1 LSTM_Forecast											
CensusYear	County	VALUE	MAE	RMSE	MAPE	CensusYea	County	VALUE	MAE	RMSE	MAPE		CensusYear	County	Sex	UNIT	VALUE	MAE	RMSE	MAPE
2022	Carlow	52135.72825	9832.271751	9832.271751	15.86669208	2022	Carlow	52135.73	9832.272	9832.272	15.86669		2022-01-01 00:00:00	Carlow			48061.88	9089.546	9747.668	15.48244
2022	Dublin	1386616.835	71537.16489	71537.16489	4,906008891	2022 [	Dublin	1386617	71537.16	71537.16	4,906009		2022-01-01 00:00:00	Dublin			1394419	50283.08	52378.7	5.319465
2022	Kildare	222504	25270	25270	10.19881021	2022	Cildare	223524.7	24249.25	24249.25	9,786843		2022-01-01 00:00:00	Kildare			336535.1	73872.57	74901.48	33.17666
2022	Kilkenny	92873.26416	11286.73584	11286.73584	10.8359599	2022	Cilkenny	92873.26	11286.74	11286.74	10.83596		2022-01-01 00:00:00	Kilkenny			83429.34	14986.67	15605.27	14.93706
2022	Laois	76238.87736	15638.12264	15638.12264	17.02071535	2022 1	aois	76238.88	15638.12	15638.12	17.02072		2022-01-01 00:00:00	Laois			67849.87	16515.02	17431.21	19.03108
2022	Longford	37853.75669	8897.243308	8897.243308	19.03112941	2022 (	ongford	40866.74	5884.262	5884.262	12.58639		2022-01-01 00:00:00	Longford			37639.54	4644.971	5640.052	10.47857
2022	Louth	127105.615	12597.38495	12597.38495	9.017261586	2022 1	outh	125479.5	14223.53	14223.53	10.18127		2022-01-01 00:00:00	Louth			123720.9	7332.344	9633.958	5.758535
2022	Meath	195044	25782	25782	11.67525563	2022 1	Meath	190007	30818.99	30818.99	13.95623		2022-01-01 00:00:00	Meath			251112.3	32535.63	32693.7	17.58636
2022	Offaly	72149.87692	11000.12308	11000.12308	13.22925206	2022 (	Offaly	72149.88	11000.12	11000.12	13.22925		2022-01-01 00:00:00	Offaly			65441.07	12816.15	13292.22	16.06631
2022	Westmeath	81007.94697	15213.05303	15213.05303	15.81053308	2022	Nestmeat	80914.36	15306.64	15306.64	15.90779		2022-01-01 00:00:00	Westmeat	n i		73526.8	15616.32	16439.59	17.09752
2022	Wexford	139357.1793	24561.82073	24561.82073	14.98412065	2022 \	Nexford	139357.2	24561.82	24561.82	14.98412		2022-01-01 00:00:00	Wexford			123913.2	26994.52	28580.26	17.42879
2022	Wicklow	142425	13426	13426	8.61463834	2022 \	Nicklow	140471.1	15379.85	15379.85	9.868305		2022-01-01 00:00:00	Wicklow			161708.6	10052.42	10500.16	8.445665
2022	Clare	113576.8524	14361.1476	14361.1476	11.22508372	2022	Clare	113576.9	14361.15	14361.15	11.22508		2022-01-01 00:00:00	Clare			104879.8	15412.37	16347.68	12.57465
2022	Cork	532959.3922	51196.60776	51196.60776	8.764201302	2022 (	Cork	509568.7	74587.29	74587.29	12.76839		2022-01-01 00:00:00	Cork			475817.2	67710.65	74089.93	12.13228
2022	Kerry	144463.9129	11994.08714	11994.08714	7.666010773	2022	(erry	144324.9	12133.09	12133.09	7.754852		2022-01-01 00:00:00	Kerry			141115.7	8656.693	9881.19	5.683139
2022	Limerick	181530.9015	28005.09854	28005.09854	13.36529214	2022 1	imerick	181530.9	28005.1	28005.1	13.36529		2022-01-01 00:00:00	Limerick			171039.1	26162.86	27611.83	13.03531
2022	Tipperary	157195.2852	10699.7148	10699.7148	6.372860894	2022 1	Tipperary	157195.3	10699.71	10699.71	6.372861		2022-01-01 00:00:00	Tipperary			143863.7	17028.96	17741.64	10.45008
2022	Waterford	108864.0817	18498.91826	18498.91826	14.52456229	2022 \	Naterford	108864.1	18498.92	18498.92	14.52456		2022-01-01 00:00:00	Waterford			101905.7	16294.31	17557.91	13.47284
2022	Galway	239271.2157	38465.78434	38465.78434	13.8497155	2022	Salway	234780.2	42956.76	42956.76	15.4667		2022-01-01 00:00:00	Galway			218829.4	40041.38	42320.51	15.11622
2022	Leitrim	53720.226	18521.226	18521.226	52.61861417	2022 1	eitrim	34290.85	908.1493	908.1493	2.580043		2022-01-01 00:00:00	Leitrim			27998.94	4707.094	5027.127	14.07492
2022	Mayo	131848.7714	6121.228607	6121.228607	4.43663739	2022 1	Mayo	129447.8	8522.245	8522.245	6.176883		2022-01-01 00:00:00	Mayo			123385.9	9244.047	9985.902	6.88558
2022	Roscommon	88416.54623	18157.54623	18157.54623	25.84372996	2022	Roscomm	63279.73	6979.267	6979.267	9.933626		2022-01-01 00:00:00	Roscommo	n		59014.3	7011.398	7624.563	10.42016
2022	Sligo	67470.66754	2727.332457	2727.332457	3.88519966	2022 9	ligo	65147.35	5050.652	5050.652	7.194866		2022-01-01 00:00:00	Sligo			61163.84	5625.965	6120.451	8.292351
2022	Cavan	75816.78283	5887.217173	5887.217173	7.205543392	2022 (	Cavan	73093.51	8610.493	8610.493	10.53864		2022-01-01 00:00:00	Cavan			66535.84	9830.279	10580.02	12.58225
2022	Donegal	154704.7809	12379.21912	12379.21912	7.408979388	2022 (	Donegal	154704.8	12379.22	12379.22	7.408979		2022-01-01 00:00:00	Donegal			147820.9	13580.72	14171.24	8.32014
2022	Monaghan	62219.34357	3068.656432	3068.656432	4.700184463	2022 1	Monaghar	61665.38	3622.621	3622.621	5.548679		2022-01-01 00:00:00	Monaghan			55397.13	6544.79	6967.925	10.39274

Figure (2) Result Output

Comparing the performance of models based on cumulative MAE-Mean Absolute error it's clear that Forecasting with ARIMA Model Grid Search is yielding the lowest MAE therefore is considered for further analysis and used for forecasting Income & Population for till year 2033, with this the research completes the research objective 2 specified in the Table 1 - Research Objective.

Model	MAE				
ARIMA_Forecast	497,033				
LSTM_Forecast	522,591				
Grid_Search_ARIMA	485,126				
Table(2) Model Result					

### **3.2 Impact of Potato famine on the model performance while Forecasting Population for all the counties of Ireland till 2033**

Patato\_Femine\_Impact\_On\_Population\_Forecast.ipynb script which performs the below steps to determine the impact of potato famine



Figure(3) Patato\_Femine\_Impact\_On\_Population\_Forecast.ipynb Flow

#### 3.2.1 Dataset Selection

The script reads the population dataset

#### 3.2.2 Data Filtering

To see the impact of Potato femine the dataset is filtered to include only records where the CensusYear is greater than 1845, the Sex is 'Both sexes', and the County is not 'State'.

The 'CensusYear' column is converted to a datetime format as time series model will be applied on it.

#### 3.2.3 Grid Search ARIMA

The function grid\_search\_ARIMA is used to perform a grid search for the best ARIMA parameters.

#### **3.2.4** Forecasting Population

The script iterates through each county, performs grid search for ARIMA parameters, and then trains and forecasts using the best parameters.

The results, including forecasted values and metrics, are stored in a DataFrame (results\_df).

#### 3.2.5 Result Output

The forecasted DataFrame results\_df is displayed and saved to an Excel file Including\_Patato\_Femine\_Grid\_Search\_ARIMA\_v1.xlsx.All the steps from 1 to 5 are reexecuted by changing (df\_population['CensusYear'] > 1845) to (df\_population['CensusYear'] > 1852) & saving the output in Excluding\_Patato\_Femine\_Grid\_Search\_ARIMA\_v1.xlsx.

Ex e	cluding_Pata	to_Femine_(	Grid_Search_	ARIMA_v1			Including	g_Patato_Fer	nine_Grid_Se	arch_ARIM	A_v1
CensusYear	County	VALUE	MAE	RMSE	MAPE	CensusYear	County	VALUE	MAE	RMSE	MAPE
2022	Carlow	53441.25	8526.749	8526.749	13.75992	2022	Carlow	53441.25	8526.749	8526.749	13.75992
2022	Dublin	1387764	70389.63	70389.63	4.827311	2022	Dublin	1387764	70389.63	70389.63	4.827311
2022	Kildare	227074.7	20699.29	20699.29	8.354103	2022	Kildare	227074.7	20699.29	20699.29	8.354103
2022	Kilkenny	91578.74	12581.26	12581.26	12.07878	2022	Kilkenny	91578.74	12581.26	12581.26	12.07878
2022	Laois	79846.1	12030.9	12030.9	13.09458	2022	Laois	79846.1	12030.9	12030.9	13.09458
2022	Longford	39959.15	6791.854	6791.854	14.52772	2022	Longford	39959.15	6791.854	6791.854	14.52772
2022	Louth	128145.9	11557.11	11557.11	8.272631	2022	Louth	128145.9	11557.11	11557.11	8.272631
2022	Meath	195044	25782	25782	11.67526	2022	Meath	195044	25782	25782	11.67526
2022	Offaly	74089.49	9060.506	9060.506	10.89658	2022	Offaly	74089.49	9060.506	9060.506	10.89658
2022	Westmeat	85104.72	11116.28	11116.28	11.55286	2022	Westmeat	85104.72	11116.28	11116.28	11.55286
2022	Wexford	144844.4	19074.63	19074.63	11.63662	2022	Wexford	144844.4	19074.63	19074.63	11.63662
2022	Wicklow	142425	13426	13426	8.614638	2022	Wicklow	142425	13426	13426	8.614638
2022	Clare	112993.1	14944.91	14944.91	11.68137	2022	Clare	112993.1	14944.91	14944.91	11.68137
2022	Cork	527552.3	56603.7	56603.7	9.689826	2022	Cork	527552.3	56603.7	56603.7	9.689826
2022	Kerry	143844.2	12613.82	12613.82	8.06211	2022	Kerry	143844.2	12613.82	12613.82	8.06211
2022	Limerick	189509.5	20026.53	20026.53	9.557561	2022	Limerick	189509.5	20026.53	20026.53	9.557561
2022	Tipperary	154373.8	13521.21	13521.21	8.053371	2022	Tipperary	154373.8	13521.21	13521.21	8.053371
2022	Waterford	111759	15603.97	15603.97	12.25157	2022	Waterford	111759	15603.97	15603.97	12.25157
2022	Galway	249372.8	28364.18	28364.18	10.2126	2022	Galway	249372.8	28364.18	28364.18	10.2126
2022	Leitrim	32676.54	2522.456	2522.456	7.166272	2022	Leitrim	32676.54	2522.456	2522.456	7.166272
2022	Mayo	129505.7	8464.339	8464.339	6.134912	2022	Mayo	129505.7	8464.339	8464.339	6.134912
2022	Roscomm	64785.83	5473.165	5473.165	7.789984	2022	Roscomm	64785.83	5473.165	5473.165	7.789984
2022	Sligo	66867.28	3330.723	3330.723	4.744755	2022	Sligo	66867.28	3330.723	3330.723	4.744755
2022	Cavan	74781.02	6922.979	6922.979	8.473243	2022	Cavan	74781.02	6922.979	6922.979	8.473243
2022	Donegal	155542	11541.96	11541.96	6.907882	2022	Donegal	155542	11541.96	11541.96	6.907882
2022	Monaghar	60714 14	1572 959	4572 959	7 005664	2022	Monaghar	60714.14	4573.858	4573.858	7.005664

Figure(4) Model Output

As per the above results generated by cumulative MAE shown below in the table it's clear that excluding the years of potato famine decreases the mean absolute error which in turn increases the efficiency of model therefore these years are not considered for forecasting population in the research. With this the research completes the research objective 3 specified in the Table 1 - Research Objective

Туре	MAE
Excluding_Potato_Femine Years	374522
Including_Potato_Femine Years	425544
Table(2) Madel Decult	

Table(3) Model Result

# 3.3 Forecasting Population & Income for all the counties of Ireland till 2033

Following scripts Population\_Forecast.ipynb forecasts the population for each county using Grid Search ARIMA using below are the steps



Figure (5) Following scripts Population\_Forecast.ipynb flow

#### 3.3.1 Data Preprocessing

The dataset is filtered to include only records where the CensusYear is greater than 1970 so as to obtain more correct results removing outliers such as potato femine years and as Ireland joined european economic zone on 1973 but the talks of Ireland joinnning EU began in 1970 therefore used 1970 as the base year for forecasting . and where the Sex is 'Both sexes'. As time series model is required to be applied the 'CensusYear' column is converted to a datetime format.

#### 3.3.2 Grid Search ARIMA

The script includes a function (grid\_search\_ARIMA) for performing a grid search to find the best ARIMA parameters. It then defines a function (train\_and\_forecast) to train the ARIMA model and forecast population for each county.

#### 3.3.3 Forecasting(Up to 2033)

The script iterates through each county, performs grid search for ARIMA parameters, and then trains and forecasts using the best parameters. The results, including forecasted values and metrics, are stored in a DataFrame (results\_df).

#### 3.3.4 Result Output

The forecast is performed for each county for the next several years (up to 2033), and the results are saved to an Excel file (Population.xlsx). With this the research completes the research objective 5 specified in the Table 1 - Research Objective.

# **3.4** Forecasting Income for all the counties of Ireland till 2033 for 2 Socio economic indicators

Following the same process shown above for forecasting population, Income\_Forecast.ipynb script.



Figure (6) Income Forecasting flow

Essentially the scrit forecast the income for years till 2033 for two socio economic indicators "Disposable Income per Person" & "Disposable Income per Person (Excluding Rent)" into 2 seprate output files. With this the research completes the research objective 4 specified in the Table 1 - Research Objective

Disposable_Income_per_	Disposable_Income_per_Person								
Statistic Label	Year	County	UNIT	VALUE	Statistic Label	Year	County	UNIT	VALUE
Disposable Income per Person (excluding Rent)	2022	Co. Carlow	Euro	19503.38	Disposable Income per Person	2022	Co. Carlov	Euro	21285.59
Disposable Income per Person (excluding Rent)	2023	Co. Carlow	Euro	19786.97	Disposable Income per Person	2023	Co. Carlov	Euro	21428.04
Disposable Income per Person (excluding Rent)	2024	Co. Carlow	Euro	20141.02	Disposable Income per Person	2024	Co. Carlov	Euro	21620.19
Disposable Income per Person (excluding Rent)	2025	Co. Carlow	Euro	20585.19	Disposable Income per Person	2025	Co. Carlov	Euro	21880.79
Disposable Income per Person (excluding Rent)	2026	Co. Carlow	Euro	21142.13	Disposable Income per Person	2026	Co. Carlov	Euro	22236.05
Disposable Income per Person (excluding Rent)	2027	Co. Carlow	Euro	21833.27	Disposable Income per Person	2027	Co. Carlov	Euro	22721.89
Disposable Income per Person (excluding Rent)	2028	Co. Carlow	Euro	22670.85	Disposable Income per Person	2028	Co. Carlov	Euro	23384.7
Disposable Income per Person (excluding Rent)	2029	Co. Carlow	Euro	23654.29	Disposable Income per Person	2029	Co. Carlov	Euro	24277.07
Disposable Income per Person (excluding Rent)	2030	Co. Carlow	Euro	24781.55	Disposable Income per Person	2030	Co. Carlov	Euro	25447.2
Disposable Income per Person (excluding Rent)	2031	Co. Carlow	Euro	25426.53	Disposable Income per Person	2031	Co. Carlov	Euro	26933.17
Disposable Income per Person (excluding Rent)	2032	Co. Carlow	Euro	27196.89	Disposable Income per Person	2032	Co. Carlov	Euro	28783.17
Disposable Income per Person (excluding Rent)	2033	Co. Carlow	Euro	28826.26	Disposable Income per Person	2033	Co. Carlov	Euro	31042.36
Disposable Income per Person (excluding Rent)	2022	Co. Cavan	Euro	17608.68	Disposable Income per Person	2022	Co. Cavan	Euro	19255.58
Disposable Income per Person (excluding Rent)	2023	Co. Cavan	Euro	17783.17	Disposable Income per Person	2023	Co. Cavan	Euro	19361.05
Disposable Income per Person (excluding Rent)	2024	Co. Cavan	Euro	18002.13	Disposable Income per Person	2024	Co. Cavan	Euro	19505.76
Disposable Income per Person (excluding Rent)	2025	Co. Cavan	Euro	18279.05	Disposable Income per Person	2025	Co. Cavan	Euro	19704.71
Disposable Income per Person (excluding Rent)	2026	Co. Cavan	Euro	18632.32	Disposable Income per Person	2026	Co. Cavan	Euro	19978.69
Disposable Income per Person (excluding Rent)	2027	Co. Cavan	Euro	19072.58	Disposable Income per Person	2027	Co. Cavan	Euro	20356.23
Disposable Income per Person (excluding Rent)	2028	Co. Cavan	Euro	19637.63	Disposable Income per Person	2028	Co. Cavan	Euro	20875.1
Disposable Income per Person (excluding Rent)	2029	Co. Cavan	Euro	20363.74	Disposable Income per Person	2029	Co. Cavan	Euro	21581.92
Disposable Income per Person (excluding Rent)	2030	Co. Cavan	Euro	21291.96	Disposable Income per Person	2030	Co. Cavan	Euro	22526.83

Figure (7) Forecated Income

## 3.5 Calculating Income Inequality among the counties of Ireland

To analyze the income distribution across counties in Ireland and assess the Gini coefficient as a measure of income inequality 'Income Inequality Calculation.ipynb' script performs the below steps

#### 3.5.1 Data Collection and Preparation



Figure (8) Data collection flow

- Read historical dataset Income Dataset(RAA02) downloaded from CSO.ie,it contains disposable income information per person & disposable income information per person(excluding rent) for different counties and different years starting 2002 till 2020
- Read the Disposable\_Income\_per\_Person.xlsx & Disposable\_Income\_per\_Person\_excluding\_Rent.xlsx which has forecasted incomes from 2022 till 2033
- Concatenate the dataframe to get one single income dataset
- Performed the similar steps to determine population number for the counties
- Both the datasets have county & year common therefore have merged based on the 2 columns so as to get single dataframe which contains income & population for all the counties in one row from 2002 upto 2033

#### 3.5.2 Data Pre-processing and Transformation

- Created a new column 'MetroCounty' categorizing counties into metropolitan or nonmetropolitan based on a predefined list discussed above.
- Introduced new column 'Income' by multiplying 'Population' and 'Income\_Per\_Person' and saved the data into Forecasted\_Income\_Population.xlsx

Forecasted_Income_Po	pulation
----------------------	----------

Statistic Label	Year	County	UNIT	Income_Per_Person	MetroCounty	Population	Income
Disposable Income per Person	2002	Carlow	Euro	13406	no	46014	616863684
Disposable Income per Person	2002	Cavan	Euro	13244	no	56546	748895224
Disposable Income per Person	2002	Clare	Euro	14192	no	103277	1465707184
Disposable Income per Person	2002	Cork	Euro	14714	yes	447829	6589355906
Disposable Income per Person	2002	Donegal	Euro	12171	no	137575	1674425325
Disposable Income per Person	2002	Dublin	Euro	17005	yes	1122821	19093571105
Disposable Income per Person	2002	Galway	Euro	14121	yes	209077	2952376317
Disposable Income per Person	2002	Kerry	Euro	13019	no	132527	1725369013
Disposable Income per Person	2002	Kildare	Euro	15771	no	163944	2585560824
Disposable Income per Person	2002	Kilkenny	Euro	13510	no	80339	1085379890
Disposable Income per Person	2002	Laois	Euro	13576	no	58774	797915824
Disposable Income per Person	2002	Leitrim	Euro	13474	no	25799	347615726
Disposable Income per Person	2002	Limerick	Euro	14902	yes	175304	2612380208
Disposable Income per Person	2002	Longford	Euro	14400	no	31068	447379200
Disposable Income per Person	2002	Louth	Euro	14754	no	101821	1502267034
Disposable Income per Person	2002	Mayo	Euro	13637	no	117446	1601611102
Disposable Income per Person	2002	Meath	Euro	14614	no	134005	1958349070
Disposable Income per Person	2002	Monagha	r Euro	13441	no	52593	706902513
Disposable Income per Person	2002	Offaly	Euro	13653	no	63663	869190939
Disposable Income per Person	2002	Roscomm	Euro	13946	no	53774	749932204
Disposable Income per Person	2002	Sligo	Euro	13940	no	58200	811308000
Disposable Income per Person	2002	Tipperary	Euro	14273	no	140131	2000089763
Disposable Income per Person	2002	Waterford	Euro	14948	yes	101546	1517909608
Disposable Income per Person	2002	Westmea	Euro	14443	no	71858	1037845094
Disposable Income per Person	2002	Wexford	Euro	13647	no	116596	1591185612
	Figure	e (9) M	lerged	dataset			

#### 3.5.3 Gini Coefficient Computation

To calculate and quantify the income inequality among the counties of Ireland below methods are used



Figure (10) Gini Coefficient Calculation

#### 3.5.3.1 Method 1

Developed a method to calculate the Gini coefficient using the Lorenz curve. Plotted the Lorenz curve and visualized income distribution. Applied this method to historical and forecasted income data considering both the factors which are Disposable income & Disposable income excluding rent for calculating the Gini Coefficient.

#### 3.5.3.2 Method 2

Used an alternative method for Gini coefficient calculation using just Income data without considering the population. Applied this method to historical and forecasted income data. Considering both the factors which are Disposable income & Disposable income excluding rent for calculating the Gini Coefficient.

#### 3.5.4 Storing Gini Index results

Once the multiple Gini Coefficients are calculated they are stored in a relational database – PostGRE SQL table gini\_coef\_hist

_ Prop	perties SQL Statistics Dependencies D	ependents Processes	Database_DAP/postgres@PostgreSQL 15*
50	Database_DAP/postgres@PostgreSQL 15	~ 1	5
	🔁 🗸 🖍 🔨 🗶 No limit 🔹 🔳		
Que	ry Query History		
1	<pre>select * from gini_coef_hist order</pre>	by 1,2 desc	
Data	a Output Messages Notifications		
=+			
	parameter character varying (255)	year character varying (255)	gini_index character varying (255)
1	Disposable Income per Person (excluding Rent) M1	2033	0.1790206122477531
2	Disposable Income per Person (excluding Rent) M1	2027	0.0781478919644563
3	Disposable Income per Person (excluding Rent) M1	2022	0.064286308778304
4	Disposable Income per Person (excluding Rent) M1	2016	0.058273448140066506
5	Disposable Income per Person (excluding Rent) M1	2011	0.04102578620705131
6	Disposable Income per Person (excluding Rent) M1	2006	0.03454521452570873
7	Disposable Income per Person (excluding Rent) M1	2002	0.04462097407820323
8	Disposable Income per Person (excluding Rent) M2	2033	0.569098279257975
9	Disposable Income per Person (excluding Rent) M2	2027	0.5340052152943399
10	Disposable Income per Person (excluding Rent) M2	2022	0.5255124037541915
11	Disposable Income per Person (excluding Rent) M2	2016	0.5176149795854937
12	Disposable Income per Person (excluding Rent) M2	2011	0.49831450854583986
13	Disposable Income per Person (excluding Rent) M2	2006	0.49611405441093104

Figure (11) pgAdmin4

# **3.6** Calculating the significance of Income Inequality among the counties of Ireland

• 'Statistical Significance Test.ipynb' script is used to perform the statistical significance test as per the below flow chart



Figure (12) Gini Coefficient Statistical Test

- With alpha value of 0.05 TTEST & WILCOXON statistical tests are performed on the set of Gini Coefficients extracted from gini\_coef\_hist table present in PostgreSQL
- If the tests fail to reject the null hypothesis then there is no statistically significant difference in the Gini index between consecutive years.
- If the tests rejects the null hypothesis then there is statistically significant difference in the Gini index between consecutive years.
- TTEST & WILCOXON tests are performed for the gini index from the following years 2000 & 2033

#### 3.7 Income inequality between rural & metro counties of Ireland

• 'Statistical Significance Test.ipynb' script is again used to identify the income inequality for the metro and rural counties by performing the statistical TTEST & WILCOXON test as per the below flow chart



Figure (13) Income Disparity Check flow

• TTEST & WILCOXON tests are performed for the following years 2000,2020 & 2033

#### 3.8 Data visualization

- For creating final visualizations PowerBI is essentially used to read the data present in the RDBMS database table gini\_coef\_hist
- For visualization purpose below data modelling approach is used



Figure (14) Data Model

• Once the data is refreshed in the gini\_coef\_hist table the data in powerBI data model needs to be refreshed by refesh button



Figure (15) Refershing data source

• To differntiate between the forecasted value and the presnt & historic value conditional formatting on all the visuals is used where in the forecasted values are of lighter color



# 4 Design Specification

Figure 3 illustrates the Three Tier Architecture utilized in this study. The Data Layer, positioned at the base of the structure serves as the foundation, for data collection. This layer acquires data from two datasets; Estimates of Household Income (RAA02) and Population at Each Census (FY001) published by the Central Statistics Office Ireland portal in.xlsx format. Moving up from the data layer we provisioned the business logic layer. Here various machine learning models are employed to compare forecasting models for Income and Population. The outcomes of grid search ARIMA, LSTM forecasting and ARIMA forecasting are stored separately in Excel files. Notably Mean Absolute Error (MAE) Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) values are recorded for each county and model type. These values assist in identifying the time series machine learning model for obtaining values from 2022 to 2033. Additionally this layer is responsible for data processing tasks and performing business logic operations like GINI Index calculation as well as statistical tests such, as TTEST and WILCOXON to identify income inequality. The data visualization layer stores the result of the preceding two layers. In this layer the data stored in PostgreSQL by those two layers is. Used by PowerBI for visualization purposes.



Figure (17) Architecture

# **5** Implementation

The implementation phase is the stage of converting the proposed solution into a form. This section provides an overview of the outcomes, methodologies and tools used in the implementation process. While it doesn't include code listings or user manual descriptions the focus is, on the deliverables and the technologies utilized.

## 5.1 Produced Outcomes

#### 5.1.1 Population Forecast for Irish Geography

- Transformed Data- Filtered population datasets based on criteria like sex and county
- Model Outputs- Results obtained from ARIMA model, LSTM forecasting and grid search ARIMA are stored in Excel files
- Performance Metrics- Mean Absolute Error (MAE) Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) are used for model evaluation

### 5.1.2 Impact Analysis of Potato Famines Effect on Population Forecast

- Transformed Data- Filtered population datasets for impact analysis purposes
- Model Outputs- Results obtained from ARIMA model with and without Potato Famine years are stored in Excel files
- Performance Metrics- MAE, RMSE and MAPE calculated for each scenario
- 5.1.3 Population & Income Forecast for All Counties until 2033
  - Transformed Data- Datasets filtered to ensure results by removing outliers.
  - Model Outputs- Forecasted population values up, to 2033 are stored in Population.xlsx file.
  - Projected Income- The anticipated results, for the amount of money to each individual excluding rent are stored in output files.

#### 5.1.4 Calculating Income Inequality among Counties

- We combined datasets that include information about income and population for all counties from 2002 to 2033.
- Gini Coefficient Results- We calculated Gini coefficients using the Lorenz curve and an alternative method. The results were stored in a PostgreSQL table called "gini\_coef\_hist".

### 5.1.5 Statistical Significance Test for Gini Coefficient

- We extracted Gini coefficients from the PostgreSQL table for the years 2000 and 2033
- We then conducted TTEST & WILCOXON tests to determine the significance of differences in the Gini index

### 5.1.6 Income Inequality between Rural & Metro Counties

• We collected datasets for metro and counties.

• Then we performed TTEST & WILCOXON tests on these datasets for the years 2000, 2020 and 2033.

## 5.2 Methodologies Used

- To ensure accuracy and relevance of data we applied data filtering and pre-processing techniques.
- For population forecasting we utilized ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short Term Memory) models.
- To measure income inequality we used both the Lorenz curve and an alternative method to calculate Gini coefficients.
- Additionally we conducted TTEST & WILCOXON tests to assess significance.

### 5.3 Tools and Languages

- Our programming language, for data manipulation, analysis and modelling was Python used in Jupyter notebook development kit
- During the project we made use of libraries, like NumPy, Pandas, Matplotlib, Seaborn, statsmodels and others to handle data processing and analysis tasks.
- To manage the Gini results we utilized PostgreSQL for storing and retrieving them.
- For visualizing Gini coefficient, Income & Population trend we relied on PowerBI.

## 5.4 Summary

The implementation phase was successful in turning our proposed solution into outcomes. These outcomes include predicted population values, income forecasts Gini coefficient results with significance tests.

# **6** Evaluation

The evaluation section aims to provide an analysis of the research findings highlighting their relevance to the research questions and objectives. By utilizing tools and visual aids such, as graphs and charts this section offers an rigorous assessment of the experimental research outcomes.

## 6.1 Find the best Time series model among LSTM, ARIMA Model Grid Search and Forecasting & ARIMA Model Forecasting with Fixed Parameters for forecasting Income & Population of Ireland

Comparing the performance matrices for the below model by looking to MAE values it is determined that for forecasting ARIMA Model Grid Search is the best when compared to other 2 as it has the lowest MAE. Henceforth, the objectives 2 that was outlined in table 1 have been successfully accomplished

Model	MAE	
Fixed ARIMA	497,033	
LSTM_Forecast	522,591	
Grid Search ARIMA	485,126	
Table(3) Model Out	put	

#### 6.2 Impact of Potato Famine

To examine how the Potato Famine impacts the machine learning models capability for forecasting population of Ireland.

Туре	MAE
Excluding_Potato_Femine Years	374522
Including_Potato_Femine Years	425544
Table(4) Model Output	

Looking to the above table excluding the years of potato famine decreases the mean absolute error which in turn increases the efficiency. With this the research completes the research objective 3 specified in the Table 1 - Research Objective

## 6.3 Forecasting Population & Income for All Counties until 2033

PowerBI dashboard with all 26 counties on the left gives the Population & Income trend from 2000-2033 for each selected respectively, Also the light coloured bars are the forecasted years. With this the research completes the research objective 4 & 5 specified in the Table 1 - Research Objective



Eg. a. Income & Population for Carlow

Figure (18) Income & Population trend for Carlow



b. Income & Population trend for Dublin

Figure (19) Income & Population trend for Dublin

## 6.4 Calculating Income Inequality among Counties

a. Gini Coefficient which highlights the income disparities for the counties from 2000-2033 of Ireland is showcased in the below PowerBI visual



Figure (20) Gini Coefficient



b. Lorenz Curve: It indicates that the high amount of income is present with handful number of counties in Ireland

Figure (21) Lorenz Curve

Stastical Significance Test - To check if the change from 2000 to 2033 is significant or not

a. TTEST- There is no significant difference in the situation

b. WILCOXON- There is no significant difference in the situation

Results- It is clear that for income disparity decreased in 2008 due to the recession however it's likely to increase more by 2030 which means that more Income would lie with the handful of counties, With this the research completes the research objective 6,6.a,6.b,6.c & 6.d specified in the Table 1 - Research Objective

#### 6.5 Income Inequality between Rural and Metropolitan Counties

Income distribution among the counties of Ireland



Figure (23) County Wise Income



Figure (24) County Wise Income

After grouping Metro and Rural county income and for 2000, 2020 & 2033 and conducting TTEST & WILCOXON statistical test it was inferred that the difference between both Metro & Rural counties is significant for 2000(p=0.0095) & 2020(p=0.0176) however for the year 2033 when we considered the forecasted income & population figures the change is not significant. This means that as of today income disparity do exist between the urban & rural household but the extent would reduce by 2033. With this the research completes the research objective 7 & 8 specified in the Table 1 - Research Objective.

#### 6.6 Discussion

When examining the N experiments and case studies in detail we gain insights and considerations. It is important to have reflections, on the experiment design critique the findings and provide suggestions for improvement to advance research. Moreover by placing the findings within the context of research from the literature review we enhance our understanding. Lets delve into some discussion points;

#### 6.6.1 Time Series Model Selection-

The choice of using ARIMA Model Grid Search as the performing model raises questions about how sensitive the results to hyper parameter tuning.

Consideration- In research it would be beneficial to explore how robust this model is across other datasets and economic contexts to ensure its generalizability.

#### 6.6.2 Impact of Potato Famine

We should consider both the exclusion of Potato Famine years improving forecasting efficiency and its potential impact on factors. It's worth discussing whether it should be treated as an outlier.

Consideration- Conducting a sensitivity analysis regarding including or excluding events would contribute to a more comprehensive understanding of how responsive the model is.

# 7 Conclusion and Future Work

When it comes to studying the trends, in income and population in Ireland we have discovered that the ARIMA Model Grid Search is the tool for forecasting. It outperforms Fixed ARIMA and LSTM models. We found that excluding the years affected by the Potato Famine greatly improves the accuracy of our forecasts emphasizing how historical context plays a role in modelling. To present these trends effectively to policymakers and regional planners we have developed a PowerBI dashboard that provides a view of population and income trends for all 26 counties until 2033.

Our research predicts an increase in income inequality by 2033 with differences between the counties increases and our research also shows that since early 2000's there exist a significant disparity between Metro & Rural County of Ireland. But as the difference between the income levels of Metro & rural counties are not proven statistically significant which means that the rural counties would improve at much faster pace as compared to metro counties of Ireland by 2033.

This study has provided insights that can guide decision making processes. However there are limitations to consider such as assumptions made in long term forecasting and potential vulnerability to changes. For work it is important to enhance the versatility of models across diverse datasets and economic contexts to make them more applicable. Conducting an analysis of events with varying magnitudes and durations would provide us with a deeper understanding of their impact. Furthermore integrating research would offer insights into the socio economic factors influencing income disparities. Examining how our observed trends align with patterns would also provide a broader perspective, on our findings.

# References

Alford, R.D., Reilly, R. and United States (eds) (2012) *Income inequality: an alarming U.S. trend.* New York: Nova Science Publishers (Economic issues, problems and perspectives).
Connolly, M. (2023) *Home - CSO - Central Statistics Office, Estimates of Household Income.* CSO. Available at: https://www.cso.ie/en/index.html (Accessed: 30 July 2023).

Das Gupta, M. and Das, S.K. (2014) *Economic growth and income disparity in BRIC: theory and empirical evidence*. Singapore: World Scientific.

Dorfman, R. (1979) 'A Formula for the Gini Coefficient', *The Review of Economics and Statistics*, 61(1), pp. 146–149. Available at: https://doi.org/10.2307/1924845.

Evans, P.B. and Timberlake, M. (1980) 'Dependence, Inequality, and the Growth of the Tertiary: A Comparative Analysis of Less Developed Countries', *American Sociological Review*, 45(4), pp. 531–552. Available at: https://doi.org/10.2307/2095006.

Gastwirth, J.L. (1972) 'The Estimation of the Lorenz Curve and Gini Index', *The Review of Economics and Statistics*, 54(3), pp. 306–316. Available at: https://doi.org/10.2307/1937992.

Goh, S.K. *et al.* (2023) 'Income inequality, income growth and government redistribution in Malaysia: What do we know in the long run?', *Malaysian Journal of Economic Studies*, 60(1), pp. 69–87. Available at: https://doi.org/10.3316/informit.172790821694730.

Headey, B. (2021) *Housing policy in the developed economy: The United Kingdom, Sweden and the United States.* Routledge.

Mahmud, S.S. and Akita, T. (2018) 'Urban and Rural Dimensions of the Role of Education in Income Inequality in Bangladesh', *Review of Urban & Regional Development Studies*, 30(3), pp. 169–183. Available at: https://doi.org/10.1111/rurd.12089.

Miller, D.W. (1975) 'Irish Catholicism and the Great Famine', *Journal of Social History*, 9(1), pp. 81–98. Morgan, J. (1962) 'The Anatomy of Income Distribution', *The Review of Economics and Statistics*, 44(3), pp. 270–283. Available at: https://doi.org/10.2307/1926398.

Muga, F. (2015) *Rich Country, Poor People: the challenges of providing psychiatric services in the public and the private sectors in Papua New Guinea - Florence Muga, 2015.* Available at:

https://journals.sagepub.com/doi/full/10.1177/1039856215608293 (Accessed: 31 July 2023).

Paglin, M. (1975) 'The Measurement and Trend of Inequality: A Basic Revision', *The American Economic Review*, 65(4), pp. 598–609.

Pashchenko, O. and Zharikova, O. (2021) 'INEQUALITY IN THE HOUSEHOLD INCOME DISTRIBUTION IN UKRAINE', *Bioeconomics and Agrarian Business*, 12(3–4). Available at:

https://doi.org/10.31548/bioeconomy2021.03-04.013.

'RAA02.20230731183543.xlsx' (no date). Available at:

https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fws.cso.ie%2Fpublic%2Fapi.restful%2FP xStat.Data.Cube\_API.ReadDataset%2FRAA02%2FXLSX%2F2007%2Fen&wdOrigin=BROWSELINK (Accessed: 31 July 2023).

Scott, M. (2017) *Renewing Urban Communities: Environment, Citizenship and Sustainability in Ireland.* Routledge.

Walsh, J. (2023) *Income disparity between urban and rural dwellers reduces in recent years / Maynooth University, Maynooth University.* Available at: https://www.maynoothuniversity.ie/news-events/income-disparity-between-urban-and-rural-dwellers-reduces-recent-years (Accessed: 30 July 2023).