Project Report

# GitHub URL

<https://github.com/andrewpquinn/-UCDPA_andrewpquinn->

# Abstract

In this project I analysed Ireland’s ranking according to it’s investment in supplementary pensions compared to other countries as well as how it ranks in terms of how well our old age state pension replaces pre-retirement income. To rank Irish investment in supplementary pensions I compared with OECD(Organisation for Economic Cooperation and Development) and non-OECD countries on two criteria, millions of US dollars invested(1) as well as as a percentage of GDP(2). To rank how well our old age state pension replaces pre-retirement income I used Gross pension replacement rate of average wage(3). The gross replacement rate is defined as gross pension entitlement divided by gross pre-retirement earnings. It is a measure of how effectively a pension system provides income during retirement to replace earnings, the main source of income prior to retirement.

Figures used were from the year 2020. This year was chosen as it was the most recent year where data was available for all 3 measures I was concerned with.

The features of the project are as follows:

1.Data

A real world dataset was used, source was OECD

2. Importing

-I imported from a number of OECD .csv files

-I retrieved data using an OECD API JSON

3.Preparation

-I created a number of data frames

-I was required to sort, index and group these data frames

-At different points I needed to both drop and add duplicates as well as replacing missing or extraneous values

-I merged a number of dataframes

4. Analysis

-I used conditional statements, looping and groupby functions and methods

-I created custom functions to suit my specific purposes and reused elements of this code for different value sets

-I used Numpy percentile functions to determine Ireland’s ranking on the above measures

-I created and used dictionaries and lists of dictionaries using from output from an AP

5. Visualisation

I created 3 charts using a mixture of Seaborn and Matplotlib

6. Insights

I derived five valuable insights from the analysis and justified my insights with reference to the charts or analysis

7. Machine Learning

I’d do a scatter plot with replacement rates and USD/%GDP on the respective axes, then suggest what replacement rate we ought to have given our investment in supplementary pensions or how much ought we to invest given our replacement rate. Regression.

Introduction

I chose this project as I work in the insurance industry and have a keen interest in pensions. I wanted to see where we stand with respect to other countries and whether there is plenty of room for improvement.

**Gross pension replacement rates**

The gross replacement rate is defined as gross pension entitlement divided by gross pre-retirement earnings. It measures how effectively a pension system provides a retirement income to replace earnings, the main source of income before retirement. This indicator is measured in percentage of pre-retirement earnings by gender.

OECD (2023), Gross pension replacement rates (indicator). doi: 10.1787/3d1afeb1-en (Accessed on 11 July 2023)

# Dataset

## <https://stats.oecd.org/index.aspx?queryid=69536#>

Datasets are from the Organisation of Economic Cooperation and Development’s data warehouse OECD.Stat, under the theme Social Protection and Well-being.

<https://www.oecd.org/daf/fin/private-pensions/Pension-Markets-in-Focus-2022-Statistical-Annex.xlsx>

API documentation as follows:

<https://data.oecd.org/api/sdmx-json-documentation/>

I made a number of CSV files from Pension Markets in Focus, (an OECD report).

I also pulled OECD data from a RESTful application programming interface (API) based on the SDMX-JSON standard.

What is SDMX-JSON format?

SDMX-JSON is a specific JSON-based format defined under the SDMX standard, which is used to describe and transmit statistical data.

https://stats.oecd.org/Index.aspx?DataSetCode=PAG

# Implementation Process

# I used the terminal to set up a local git repo for this project and linked to my remote git hub repo

# I imported a number of common packages and libraries in Jupyter Notebook to start with (added more later on)

# I downloaded an xls workbook from OECD with the title “Pension-Markets-in-Focus-2022-Statistical-Annex”

# I created a number of CSV files from this xls broken down by different categories

# In Apple Numbers I tidied up two of the csv files before exporting through python(Jupiter). I did this because there were extra rows that I wouldn’t need and to make it easier to create a dataframe

# I read the csv which looked at total assets in USD into a dataframe called “df” and had a look using .head , .tail and .shape functions.

# As data was included from 2001 to 2021 and I only wanted 2020 figures I removed columns using .drop and .iloc functions

# I also removed Rows where Notes, missing values, and regional aggregate data was included. This included use of .dropna funtion. I also ran a boolean function to indicate that only rows that did not include “..” as a value would be returned, this removed further missed values that were hiding as a string.

# I then repeated all of the steps but on a different csv from the OECD xls. This looked at total assets as percentage of GDP. I called this dataframe df1

# I merged these two dataframes to create a new dataframe called merged\_df

# I created a list which identified which of these countries were in the OECD, I then created 3 groups where 1 was OECD(excluding Ireland), one was non-OECD and the third group was Ireland

# After inspecting the data frame using dtypes I could see that the GDP and USD columns were not numeric, so I converted them using .replace and .astype functions

# I then sorted by USD using .sort\_values.

# I created a figure and axis and adjusted the figure size. Earlier versions of the plot were very difficult to read.

# I created a custom colour palette to make clear which countries were OECD, Non-OECD and Ireland.

# I set the x-axis to logarithmic as United States of America was too much of an outlier in an early attempt of the plot

# I created a bar chart using Seaborn

# I set labels and titles, adjusting the font size

# I set the font size of x and y ticks

# I adjusted spacing between the subplots and labels

# I used plt.show to demonstrate the first chart.

# As Ireland was hard to see and unclear I added legend patches and changed the colour palette. I also added percentile lines

# I calculated what percentile Ireland was in when it came to other countries in the set, using scipy.stats

# I repeated these steps 13-23(aside from 16 as it was not necessary this time to go logarithmic) for the %GDP column to create a chart of this measure.

# For my next dataset I wanted to use a different method of importing so I constructed an SDMX-JSON API URL using the OECD API documentation. This required a good deal of trial and error but eventually I had a JSON that I was able to read sufficiently to start building a dataframe.

# First I created a dictionary of observations, these represented the gross replacement rate for men and women in 53 countries but did not include the countries.

# Then I made this dictionary a dataframe called df\_observations

# Next I created a list of dictionaries from the dimensions data found again in the JSON, this contained 3 letter IDs for countries as well as names of countries in key:value format.

# Due to some of the punctuation in the list I was getting syntax errors so I converted into a string and back into a list, then created two different series using id and name to create a dataframe, this one called df\_countries

# I duplicated this data frame as I needed it to be the same size as df\_observations and line up with men and women figures, then I added a gender column and made every 2nd value men or women, knowing that this would correspond to df\_observations

# I reset the indexes to make the two data frames compatible and concatenated them

# I then grouped by country id and calculated the mean (men and women) Gross pension replacement rate. The last three columns were dropped, the new concatenated dataframe was called df\_new

# I then sorted by Gross pension replacement rate and created a graph, using a similar method and style to figures 1 & 2, this time showing Gross replacement rate ranking

(Describe your entire process in detail)

# Results

# Figure 1.

# Figure 2.

# Figure 3

# Insights

* Ireland is in the 71st percentile in terms of total assets invested in funded and private pension plans, as measured in millions of US dollars. See Figure 1
* 2.Ireland is in the 74th percentile in terms of total assets invested in funded and private pension plans, as measured as a percentage of GDP See Figure 2.
* Ireland is in the 10th percentile when it comes to the Gross Pension Replacement Rate by Country. See Figure 3
* Because the Irish Gross Replacement rate is so low, the 5th lowest of the 53 countries sampled in figure 3, there would appear to be a need to either a) improve state pension entitlements to more generous levels to reduce the likelihood of old age poverty/reduction of living standards or b) Increase incentives or otherwise encourage more investment in supplementary pensions amongst the Irish populace.
* Though at first blush being in the 71st and 74th percentiles for supplementary pension investment might seem impressive it’s important to see this in the context of our very low state pension replacement rates. Some of this may be due to larger pre-retirement incomes in Ireland but we can see for example that we are below United Kingdom in all 3 figures. UK is in some respects a comparable anglosphere economy and pension system, hopefully automatic enrolment might address this gap.

# References

(Include any references if required)

<https://www.oecd-ilibrary.org/sites/pens_outlook-2018-4-en/index.html?itemId=/content/component/pens_outlook-2018-4-en>

<https://www.oecd.org/daf/fin/private-pensions/Pension-Markets-in-Focus-2022-Statistical-Annex.xlsx>

OECD (2023), Gross pension replacement rates (indicator). doi: 10.1787/3d1afeb1-en (Accessed on 17 July 2023)