Project Report

# GitHub URL

<https://github.com/BorisMoric/UCDPA_BorisMoric>

# Abstract

(Short overview of the entire project and features)

The aim of the project was to use Python and available read world data and generate valuable insights into question about factors that may lead to people’s decisions on having their kids at different times in their lives. The project used three data sets and was aiming to analyse and manipulate the data contained to enable better understanding of the contents and insights generation.

# Introduction

*(Explain why you chose this project use case)*

During November the family was watching Netflix documentary “Pepsi, Where’s My Jet?”. At the same time at UCD-CIDAB course we have just completed 3rd session when it was recommended to start thinking about the project ideas and related data sets. At one point my younger son asked a question “*Do people in rich countries have kids when they are older compared to poorer countries?*”. That sparked a short but lively discussion within the family as to *the reasons that may lead to a decision to have kids & when to have them in ones life*, which gave me the motivation to start looking for data sources that would provide some insights.

# Dataset

*(Provide a description of your dataset and source. Also justify why you chose this source)*

## Births registration

The first question to be answered was what public data exists about the births. By searching through the web I concluded that this info was collected and published by the governments of the countries for the country in question. Is there a public data set out there that combines the info from multiple countries. There is an interesting one I found on (UNICEF DATA, 2022). The overall file was 656Mb in size and was locally named “Percent of children\_UNICEF\_1.0\_all”. It contained 1,648,747 rows with 22 columns. The column names and short descriptions for each is contained within .iptnb file.

## Age of mothers at childbirth

From analysis of the birth registrations dataset (UNICEF DATA, 2022) it was concluded that it did not contain the information about the age of parents at the time of the first birth of their child so another source was found from (OECD) that seem to have that. The web site provides a number of datasets. The specific data set used was listed under ‘Fertility indicators’ section SF2.3. Generated file ‘*SF\_2\_3\_Age\_moth.xlsx*’ contained several sheets. Source data from xls was converted into csv as it was easier to work with csv in python. The XLS sheet ‘*Mean-age-first-birth*’ was converted to CSV file ‘*Mean-1st-birth.csv*’.

## Countries of the world

To get some insights into why people have babies at different times in their lives there was a need to add some more parameters into the mix like where in the world the countries are located, ie. are there any regional factors, if there is a correlation with wealth, countries population sizes and etc. “Countries of the world” dataset (Lasso, 2017) contains World fact sheet, with Information on population, region, area size, infant mortality and more. The data set was downloaded by using Keggle API (Kaggle, n.d.).

# Implementation Process

*(Describe your entire process in detail)*

To start learning about “*the reasons that may lead to a decision to have kids & when to have them in ones life*” I started to first look at what data sources are available on the subject of child births. Births registrations data from UNICEF was the starting point (UNICEF DATA, 2022). After reducing the data set to show the % of births I have decided to eliminate countries that have low % (<80%) of registered births not to influence the findings negatively. Resulting data set was reduced to 75 countries.

I now started to look into Age of mothers at childbirth (OECD) as second step with the intention to combine the two (inner merge on countries). After cleaning the dataset the result was 39 countries that have collected and reported the average age of the mother at the time of their first child births.

The [1st insight](#_Across_all_of) was based on data from one data source and was fairly obvious from the charts provided in the next chapter.

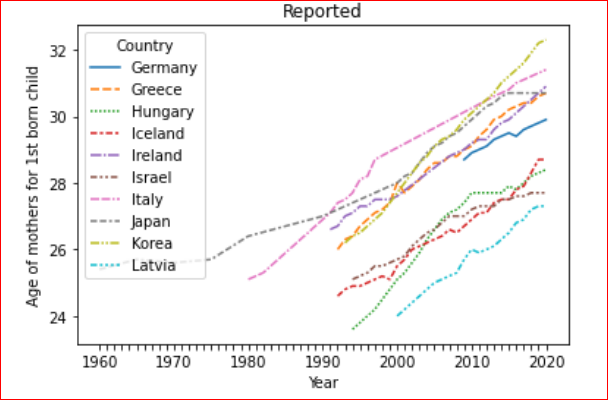
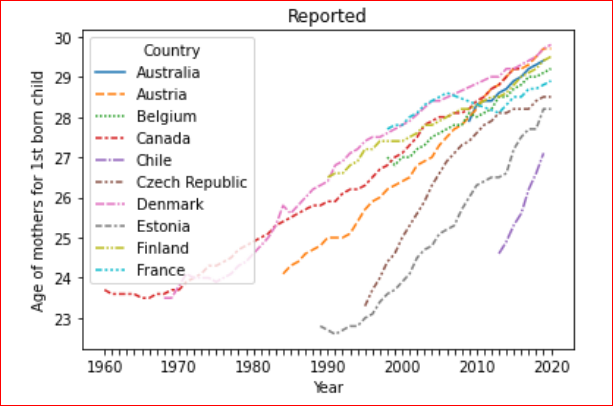
Planned merge of two datasets didn’t yield any overlap, apart from Turkey, which was surprising and resulted in [2nd insight](#_There_was_no) on the importance and difficulty in selecting the right data sources.

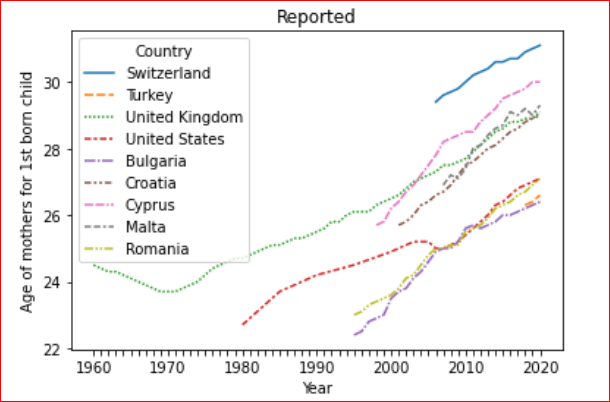
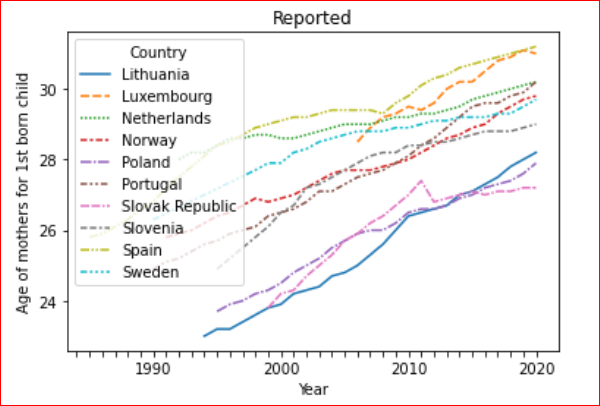
# Results

*(Include the charts and describe them)*

Charts below are derived from (OECD) report we looked at where 39 countries have reported the yearly average age of the mothers at first child birth in the span of last 60 years.

Our [1st insight](#_Across_all_of) can be derived from observing the trends in firstborns. Across all of the countries during last 60 years we can clearly see that people are deciding to have their first child later in life.





While merging the (UNICEF DATA, 2022) and (OECD) datasets with >80% condition of reported births it was noted that ([2nd insight](#_There_was_no)) there is no overlap – apart from Turkey – even when relaxing the “>80%” condition and looking at all of the 120 countries from (UNICEF DATA, 2022) which would indicate that either the datasets are incomplete or same data from the countries are not equally collected, analysed and published by both organisations - which in my opinion is less likely.

(UNICEF DATA, 2022)

['Afghanistan', 'Albania', 'Algeria', 'Angola', 'Argentina', 'Armenia', 'Azerbaijan', 'Bahrain', 'Bangladesh', 'Barbados', 'Belize', 'Benin', 'Bhutan', 'Botswana', 'Burkina Faso', 'Burundi', 'Cabo Verde', 'Cambodia', 'Cameroon', 'Central African Republic', 'Chad', 'Comoros', 'Congo', 'Cuba', "Côte d'Ivoire", "Democratic People's Republic of Korea", 'Democratic Republic of the Congo', 'Dominican Republic', 'Egypt', 'El Salvador', 'Equatorial Guinea', 'Eswatini', 'Ethiopia', 'Fiji', 'Gabon', 'Gambia', 'Georgia', 'Ghana', 'Guatemala', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti', 'Honduras', 'India', 'Indonesia', 'Iran (Islamic Republic of)', 'Iraq', 'Jamaica', 'Jordan', 'Kazakhstan', 'Kenya', 'Kiribati', 'Kosovo', 'Kyrgyzstan', "Lao People's Democratic Republic", 'Lebanon', 'Lesotho', 'Liberia', 'Madagascar', 'Malawi', 'Maldives', 'Mali', 'Marshall Islands', 'Mauritania', 'Mexico', 'Mongolia', 'Montenegro', 'Morocco', 'Mozambique', 'Myanmar', 'Namibia', 'Nauru', 'Nepal', 'Niger', 'North Macedonia', 'Pakistan', 'Panama', 'Papua New Guinea', 'Paraguay', 'Peru', 'Philippines', 'Republic of Moldova', 'Rwanda', 'Saint Lucia', 'Samoa', 'Sao Tome and Principe', 'Saudi Arabia', 'Senegal', 'Serbia', 'Sierra Leone', 'Singapore', 'Slovakia', 'Solomon Islands', 'Somalia', 'South Sudan', 'Sri Lanka', 'State of Palestine', 'Sudan', 'Suriname', 'Tajikistan', 'Thailand', 'Timor-Leste', 'Togo', 'Tonga', 'Trinidad and Tobago', 'Tunisia', 'Turkmenistan', 'Turks and Caicos Islands', 'Tuvalu', 'Türkiye', 'Uganda', 'Ukraine', 'United Republic of Tanzania', 'Uruguay', 'Vanuatu', 'Viet Nam', 'Yemen', 'Zambia', 'Zimbabwe']

(OECD)

['Australia', 'Austria', 'Belgium', 'Canada', 'Chile', 'Czech Republic', 'Denmark', 'Estonia', 'Finland', 'France', 'Germany', 'Greece', 'Hungary', 'Iceland', 'Ireland', 'Israel', 'Italy', 'Japan', 'Korea', 'Latvia', 'Lithuania', 'Luxembourg', 'Netherlands', 'Norway', 'Poland', 'Portugal', 'Slovak Republic', 'Slovenia', 'Spain', 'Sweden', 'Switzerland', 'Turkey', 'United Kingdom', 'United States', 'Bulgaria', 'Croatia', 'Cyprus', 'Malta', 'Romania']

# Insights

*(Point out at least 5 insights in bullet points)*

## Across all of the countries during last 60 years of reported data there is a clear trend of having a firstborn later in life.

## There was no overlap between countries with reported childbirths from (UNICEF DATA, 2022) dataset and Age of mothers at 1st child birth from (OECD) dataset (apart from Turkey).

# References

Kaggle. (n.d.). *Public API documentation | Kaggle*. Retrieved from https://www.kaggle.com/docs/api

Lasso, F. (2017). *Countries of the World | Kaggle.* Retrieved 12 30, 2022, from kaggle.com: https://www.kaggle.com/datasets/fernandol/countries-of-the-world

OECD. (n.d.). *OECD Family Database - OECD.* Retrieved Dec 30, 2022, from oecd.org: https://www.oecd.org/els/family/database.htm (Excel file was accessible through: https://www.oecd.org/els/soc/SF\_2\_3\_Age\_mothers\_childbirth.xlsx)

UNICEF DATA. (2022, May). *Birth registration data - UNICEF DATA data.unicef.org.* Retrieved Dec 30, 2022, from data.unicef.org.: https://data.unicef.org/resources/dataset/percentage-children-age-5-whose-births-registered-sex-place-residence-household-wealth-quintile/