Project Report

# GitHub URL

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

# Abstract

(Short overview of the entire project and features)

# 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 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 America have kids when they are older compared to other 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 outthere 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.

## Age of mothers at childbirth

From analysis of the birth registrations dataset (UNICEF DATA, 2022) it was clear that it did not have the information about the age of parents at the time of the first birth of their child so I found another source in this case from (OECD) that seem to have what I was looking for. 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 and as pandas provide better support for working with CSV files then XLS the 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, then countries population sizes and if there is a correlation with wealth etc. [a bit unclear. Just from ie onwards]

<https://www.kaggle.com/datasets/fernandol/countries-of-the-world>

About Dataset

Context - World fact sheet, fun to link with other datasets.

Content - Information on population, region, area size, infant mortality and more.

Inspiration - When making visualisations related to countries, sometimes it is interesting to group them by attributes such as region, or weigh their importance by population, GDP or other variables.

# 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 identifying the rows containing the

Also it is important to eliminate countries that have poor practice of registering the births not to influence the findings negatively.

*I’ve converted source data from xls into csv as it was easier to work with csv in python.*

The data set had 15487 rows and 22 columns. This data set needed to be analyzed in order to reduce the scope to countries where the % of children whose births were registered is higher than 80%. After some closer data examination and filtering the data set was reduced to 75 countries (out of 142) that reported child births in more than 80%.

Once the birth registrations dataset was analysed it was clear that it did not have the information about the age of parents to answer the fundamental question, which resulted in a need to search for more specific information. OECD seem to have what was needed and “Age of mothers at childbirth” data set was imported in the notebook next.

# Results

(Include the charts and describe them)

# Insights

(Point out at least 5 insights in bullet points)

# References

(Include any references if required)

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/