Project Report

# GitHub URL

(insert URL here)

# Abstract

(Short overview of the entire project and features)

# Introduction

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

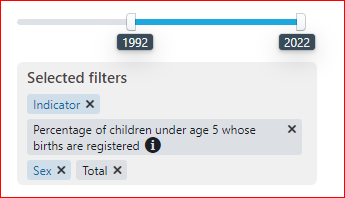
During November the family was watching (the)“Pepsi, Where’s My Jet?” documentary on Netflix. At the CIDAB course we have just completed (the)3rd session which was about -the- data frames, data import as well as filtering, data cleaning and grouping. At that stage I was yet to select my data set and during the documentary at the back of my mind was the question, what would be a good data set to use? Then 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

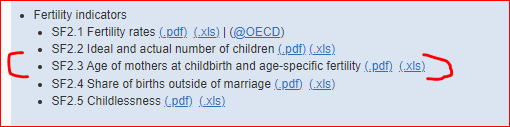
The first question we needed to answer is what public data exists about the births. There is a dataset per country on birth registrations on UNICEF web: <https://data.unicef.org/resources/dataset/percentage-children-age-5-whose-births-registered-sex-place-residence-household-wealth-quintile/> . The overall file was 656Mb in size and to narrow down dataset below is the filter applied to produce the dataset used in the project and generated CSV file is ‘*fusion\_GLOBAL\_DATAFLOW\_UNICEF\_1.0\_.PT\_CHLD\_Y0T4\_REG+DM.csv*’. This file was locally renamed to ‘fusion\_GLOBAL\_UNICEF.csv’ as Git was throwing the ‘Filename too long’ error when tried to add it to repository.



## Age of mothers at childbirth

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 at: <https://www.oecd.org/els/family/database.htm> .

(The) Above linked web page provides (a) number of datasets and the specific dataset I used in my project is indicated below under the ‘Fertility indicators’ section SF2.3. Generated file is ‘*SF\_2\_3\_Age\_mothers\_childbirth.xlsx*’ from which sheet ‘*Mean-age-birth*’ was copied and converted to CSV (ie. file ‘*SF\_2\_3\_Age\_mothers\_childbirth.csv*’) so can be imported.



## 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 childbirths.

By searching through the web I quickly concluded that data about the childbirths was generally collected and made public by the countries on the national level hence looking for data sets at the national levels would be too limiting and NGO’s that operate worldwide would potentially be a good place to look for data aggregates. Also it is important to eliminate countries that have poor practice of registering the births not to influence the findings negatively. Births registrations data from UNICEF was the starting point (***Citation needed***).

*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)