Covid19 Mortality and Association with Health System and Life Expectancy Analysis

**Health Analytics Project**

Contents

[1. purpose of this document 2](#_Toc75910000)

[2. inroduction & context 2](#_Toc75910001)

[Data Entry Applications 2](#_Toc75910002)

# purpose of this document

This purpose of this document is to explain the data process to create the COVID19 Analysis dashboard.

# 

# inroduction & context

## 

## Data Entry Applications

The Covid19 pandemic, is a global pandemic that hit almost all the countries around the world and left many dead where the total mortality reached more than 3.9 million. Why Covid19 hits some countries more than others. Although researchers started to find theories that explains why some countries had many cases and deaths compared to others, still researcher couldn’t identify all the confirmed reasons to respond to this question and better understand the virus.

# Data collection

The data used in this project was collected from different online sources, mainly the world health organization. Here are the data sets used and their sources.

|  |  |
| --- | --- |
| Covid19 deaths by country and date | worldmeter.info |
| Health system ranking | World health organization (WHO), Measuring Overall Health System Performance for 191 Countries |
| Life expectancy | World health organization (WHO), life expectancy(years) by country in 2019 |
| Population age distribution by country | Institute for Health Metrics and Evaluation (IHME) population forcasting |

# Data cleaning and processing

There was no major cleaning required on the extracted data sets, the data was complete and clean. Data processing techniques were applied to the data sets using python to prepare the data for visualization and analyses. Summary of the processing steps:

* Covid deaths data set was merged with life expectancy and health system ranking based on the country field into one data frame to visualize the association graphs.
* Data set were filtered and subset, not all information in the raw data was needed.
* Extracted new fields from the data set, for example period from date field to display the covid19 death by month, new age ranges were created to replace the existing ones for the analysis purposes.

Python was used for data processing, one python fille with comments that explain the steps is attached with the project documents.

# Streamlit dashboard

Covid19 mortality analysis dashboard is created using Streamlit – it is a tool used to create and share data application easily and fast.