DATA SCIENCE PROJECT:
Yellow Fever in Brazil

Introduction

This project is a personal initiative focused on learning.

In this project I used a dataset about cases of yellow fever in Brazil from 1994 to 2021, all the data is real, from the Ministry of Health of Brazil

Dataset: [Portal de Dados Abertos](#)

The objective of this project is to analyze the data to understand trends of yellow fever in Brazil. I used Python and Anaconda to clean and transform the data, the Jupyter notebook to document my process and Power BI to see the results.

Technologies used:

> **Python 3** - [Welcome to Python.org](#)
> **Anaconda** - [Anaconda | The World's Most Popular Data Science Platform](#)
> **Jupyter Notebook** - [Project Jupyter | Home](#)
> **Power BI** - [Power BI - Visualização de Dados | Microsoft Power Platform](#)

What is Yellow Fever?

Yellow fever is an infectious disease that, in severe cases, has a high fatality rate. It is transmitted by mosquitoes and has two transmission cycles: the urban cycle, where the main vector is the *Aedes aegypti* mosquito, and the sylvatic cycle, where the predominant transmitters are mosquitoes of the *Haemagogus* and *Sabethes genera*. This disease is of the resurgent type, meaning it can reappear periodically, with the highest incidence occurring during the summer.

Fortunately, there is a yellow fever vaccine. Since 2017, Brazil has implemented a policy of a single lifetime dose.

Exploratory Analysis and Data Cleaning and Transformation

I started loading the dataset in the Jupyter Notebook, looking for any problems, such as incorrect data types, null values and outliers.
I checked the age distribution using a histogram and the count of deaths with a bar chart.
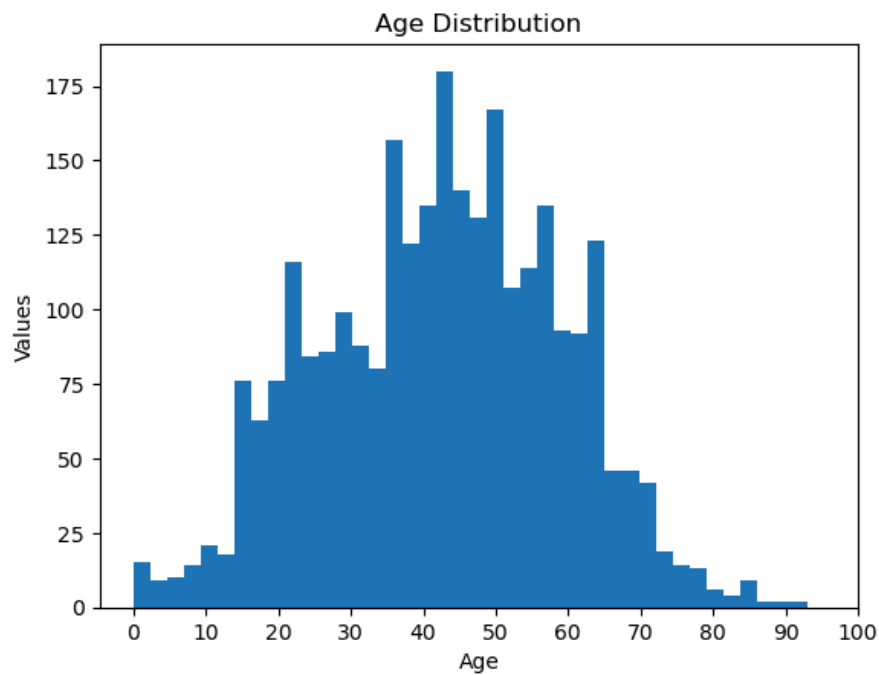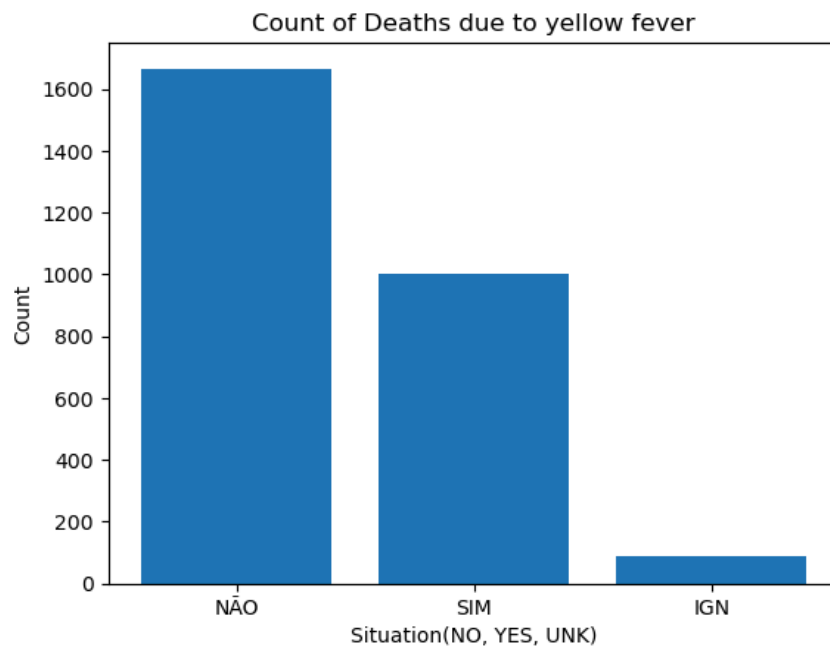
Figure 1 - Age Distribution

Figure 2 - Count of Deaths due to yellow fever

I found some null values using the functions "isnull().sum()", so i removed all null values from the columns 'SEXO', 'IDADE' and 'DT_IS', i tried to change the data type of 'DT_IS, 'but i found some strings and invalid days, so i removed the strings and replaced the invalid days.

I changed the 'DT_IS' type from object to date. It was necessary to create two new columns 'MONTHS', which contains the name of the month, and 'ESTADO' which contains the full name of the states; these columns are necessary for analyzing the number of deaths per month and by states.

Dashboard (Power BI)

it was created 2 pages for analysis, an overview about the situation of yellow fever in Brazil and an analysis of the states: