**Life Expectancy: A Data-Driven Exploration and Prediction**

**Milestone: Data Exploration and Visualization**

Group 13

Sai Varun Kumar Namburi

Prajwal Srinivas

[namburi.sai@northeastern.edu](mailto:namburi.sai@northeastern.edu)

[srinivas.pra@northeastern.edu](mailto:srinivas.pra@northeastern.edu)

**Percentage of Effort Contributed by Student1: 50%**

**Percentage of Effort Contributed by Student2: 50%**

**Signature of Student 1: Sai Varun**

**Signature of Student 2: Prajwal**

**Date of Submission: 5th Mar 2023**

**Data Source:** This dataset is taken from kaggle.com

<https://www.kaggle.com/code/philbowman212/life-expectancy-exploratory-data-analysis/data>

**Data Description:**

The World Health Organization's Global Health Observatory (GHO) maintains records of the health status and related factors of all countries. The data concerning life expectancy and health factors for 193 countries were obtained from the WHO's Global Health Observatory data repository. It was noted that over the past 15 years, there has been significant progress in the health sector, leading to a significant improvement in human mortality rates, particularly in developing nations compared to the last 30 years. In this project, data from the years 2000 to 2015 for 193 countries were selected for further analysis. In this dataset, we have 22 columns as below

(Country, Year, Status, Life expectancy, Adult Mortality, infant deaths, Alcohol, percentage expenditure, Hepatitis B, Measles, BMI, under-five deaths, Polio, Total expenditure, Diphtheria, HIV/AIDS, GDP, Population, thinness 1-19 years, thinness 5-9 years, Income composition of resources, Schooling)

**Data Exploration:**

Observation

* Life Expectancy has a negative correlation with adult mortality
* Life Expectancy has a strong correlation with Schooling and Income composition of resources
* The strong correlation between thinness\_1to19\_years and thinness\_5to9\_years
* There is a non-negligible correlation between Life Expectancy and BMI and body diseases

Chart, scatter chart

Description automatically generated

Dealing with missing data:

Text

Description automatically generated

Percentage of Null Values for each column:

Text

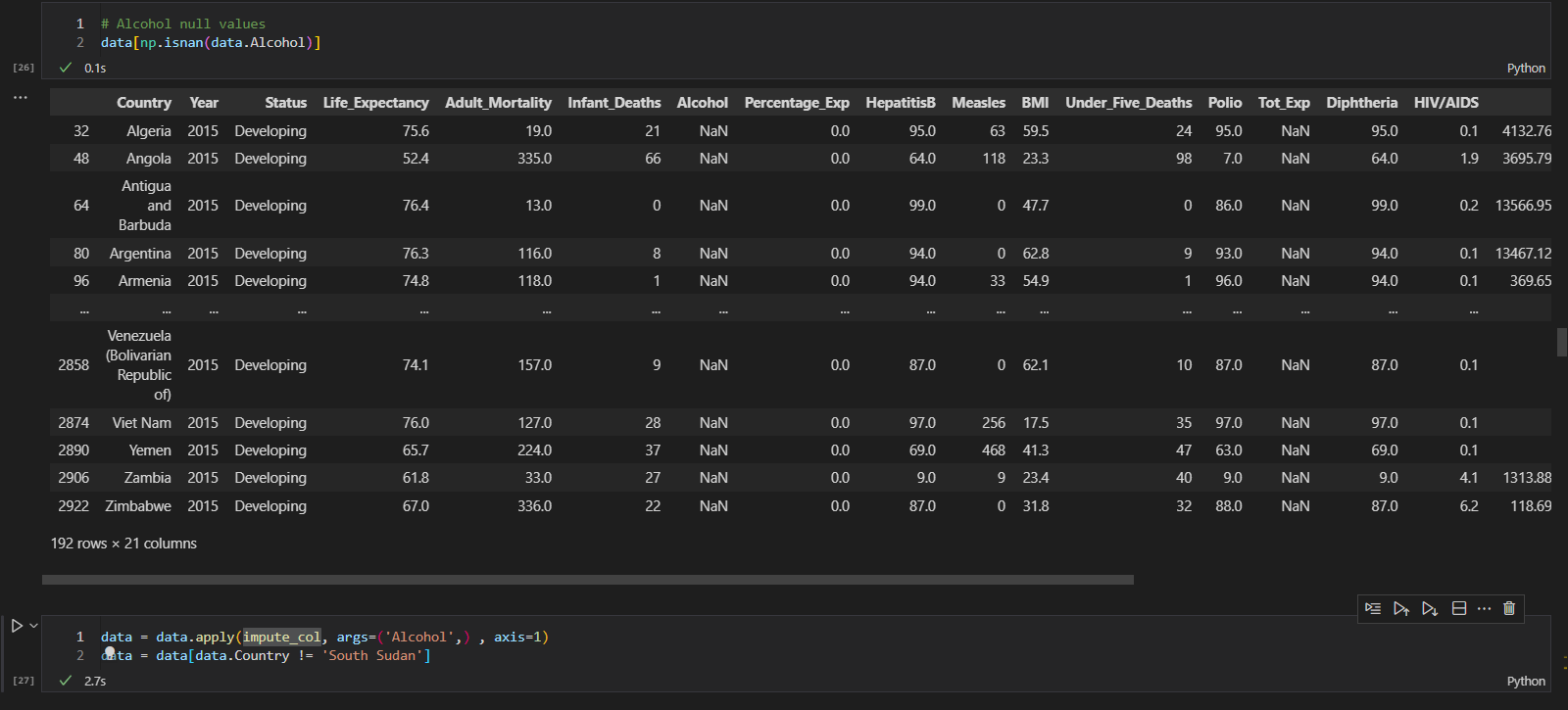
Description automatically generated

Treat null values using interpolation:

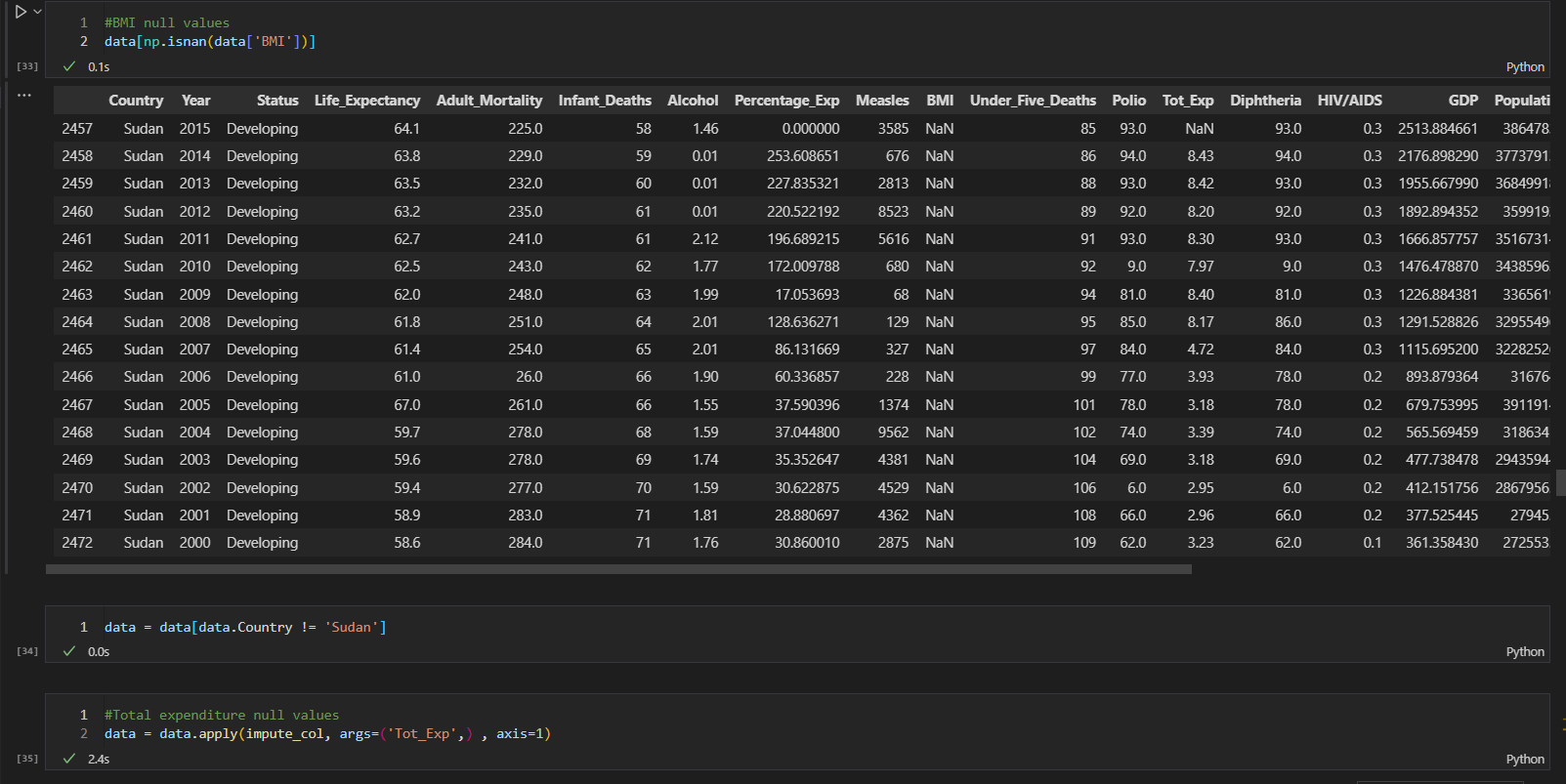
Text

Description automatically generated

Alcohol null values and imputing it:



BMI null values::



Impute the total expenditure null values

Population null values and imputing it with mean

Graphical user interface, text, application

Description automatically generated

Removing the unnecessary columns and imputing it with mean

Text

Description automatically generated

Above we can see that now we don’t have any null values

**Data Visualizations:**

Distribution of the response variables

Chart, histogram

Description automatically generated

Grouping the data based on status and avg life expectancy

**Chart

Description automatically generated**

Boxplot grouped by country life expectancy

**Chart, box and whisker chart

Description automatically generated**

Plotting the graph for the below four countries and grouping it with mean

**Graphical user interface

Description automatically generated**

Plotting a bar graph where we group data by each year

**A picture containing chart

Description automatically generated**

Finding a covariance and plotting a heatmap among variables

**Graphical user interface, chart

Description automatically generated**

**Outliers:**

Text

Description automatically generated

Chart, box and whisker chart

Description automatically generated

Chart, box and whisker chart

Description automatically generated