Global Causes of Death

Project Overview

The primary objective of this project is to analyze global mortality data to uncover key trends across regions and time periods. By employing various data analysis and visualization techniques, this project aims to provide actionable insights into mortality patterns, offering a foundation for public health strategies. The analysis leverages Python, Excel, and Power BI to deliver comprehensive visualizations and interactive dashboards.

Tools & Technologies Used

- Python (Pandas, Matplotlib, Seaborn): For data cleaning, exploration, and advanced visualizations.

- Excel: For pivot table analysis and initial data inspection.

- Power BI: To develop a dynamic and interactive dashboard summarizing key insights.

Project Scope and Objectives

1. Data Acquisition & Preparation

The dataset is acquired from global health databases. The first step involves cleaning and processing the data to ensure consistency and reliability. This includes handling missing values, standardizing formats, and preparing the dataset for analysis.To create a structured dataset that is free from anomalies, enabling accurate analysis and meaningful insights.

Code Example: Data Cleaning

```
import pandas as pd
import os
import seaborn as sns
import matplotlib.pyplot as plt
import numpy as np
from sklearn.linear_model import
LinearRegression
from sklearn.model_selection import
train_test_split
from sklearn.metrics import mean_absolute_error

file_path =
"C:\\Users\\tarek\\Desktop\\hackathon\\cause_of
_deaths.csv"
df = pd.read_csv(file_path)
pd.set_option('display.max_columns', None)
print(df.columns)

#no columns to drop

#Check for missing values
missing_values=df.isnull().sum()
print(missing_values)

#dublicates
dublicates= df.duplicated()
print(dublicates.sum())
```

```
print(df.dtypes)
print(df.info())
print(df.head())
```

2. Exploratory Data Analysis (EDA)

The EDA phase aims to explore the data, identify the leading causes of death globally and regionally, and uncover preliminary trends. Insights from this phase will guide deeper analyses and visualizations.

Key Focus Areas:
- Distribution of causes of death by region and time period.
- Identification of any outliers or significant changes in mortality rates.

Code Example:

```
#EDA

total_deaths = df[['Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
```

```
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].sum().sort_values(ascending=False)
print(total_deaths)


variance= df[['Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
```

```
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].std().sort_values(ascending=False)

print(variance)

sum_deaths_per_country =
df.groupby('Country/Territory')[[
   'Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
```

```
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis'
]].sum().sort_values(by='Country/Territory' ,
ascending=False)

print(sum_deaths_per_country.head(5))
```

3. Visualization and insights :

This step focuses on evaluating long-term trends, such as identifying whether specific causes of death have increased or decreased over time.

Code Example

```
#visualaztion/insights
# 1. Total deaths per disease across all countries
# Insight: This will show the diseases that cause
the most deaths globally.
total_deaths = df[['Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
```

```
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].sum().sort_values(ascending=False)

plt.figure(figsize=(10, 6))
total_deaths.plot(kind='bar')
plt.title('Total Deaths by Disease Globally')
plt.xlabel('Diseases')
plt.ylabel('Total Deaths')
plt.show()

# 2. Variance of deaths per disease
# Insight: High variance might indicate significant
variability across countries or regions. because of
envoirnment, heat etc....
variance = df[['Meningitis',
```

```
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].std().sort_values(ascending=False)

plt.figure(figsize=(10, 6))
variance.plot(kind='bar', color='orange')
plt.title('Variance in Deaths by Disease')
plt.xlabel('Diseases')
plt.ylabel('Standard Deviation')
plt.show()
```

```
# 3. Total deaths per country
# Insight: This will show which countries have the
highest death toll across all diseases.
sum_deaths_per_country =
df.groupby('Country/Territory')[[
   'Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis'
]].sum()
```

```
top_10_countries =
sum_deaths_per_country.sum(axis=1).sort_values(ascen
ding=False).head(10)
plt.figure(figsize=(10, 6))
top_10_countries.plot(kind='bar', color='green')
plt.title('Top 10 Countries with Highest Deaths')
plt.xlabel('Countries')
plt.ylabel('Total Deaths')
plt.show()

# 4. Top 10 most common death causes globally
# Insight: The top causes of death globally can be
shown to see which diseases are most fatal.
top_10_death_causes = total_deaths.head(10)
plt.figure(figsize=(10, 6))
top_10_death_causes.plot(kind='bar', color='purple')
plt.title('Top 10 Most Common Death Causes
Globally')
plt.xlabel('Death Causes')
plt.ylabel('Total Deaths')
plt.show()

# 5. Pie chart of top 5 causes of death
# Insight: A pie chart helps understand the
proportion of deaths from major causes globally.
top_5_death_causes = total_deaths.head(5)
plt.figure(figsize=(8, 8))
top_5_death_causes.plot(kind='pie',
autopct='%1.1f%%', startangle=90)
plt.title('Top 5 Death Causes (Proportion
Globally)')
plt.show()
```

```
# 6. Pie chart of top 5 countries with most deaths
# Insight: Show the proportion of deaths across the
top 5 countries globally.
top_5_countries = top_10_countries.head(5)
plt.figure(figsize=(8, 8))
top_5_countries.plot(kind='pie', autopct='%1.1f%%',
startangle=90)
plt.title('Top 5 Countries by Total Deaths
(Proportion)')
plt.show()

# 7. Heatmap of death rates by disease and country
# Insight: Heatmaps will show patterns of death
across countries and diseases.
plt.figure(figsize=(12, 8))
sns.heatmap(sum_deaths_per_country[top_10_death_caus
es.index].head(10), annot=True, cmap='coolwarm')
plt.title('Heatmap of Top 10 Diseases by Country
(Top 10 Countries)')
plt.xlabel('Diseases')
plt.ylabel('Countries')
plt.show()

# 8. Stacked bar chart for top 5 death causes per
country
# Insight: This will show how the top causes of
death contribute to the overall death toll in the
top 5 countries.
top_5_countries_data =
sum_deaths_per_country.loc[top_5_countries.index,
top_5_death_causes.index]
```

```
top_5_countries_data.plot(kind='bar', stacked=True,
figsize=(12, 6))
plt.title('Top 5 Death Causes by Country')
plt.xlabel('Country')
plt.ylabel('Total Deaths')
plt.legend(loc='upper right')
plt.show()


# Group data by Year, summing the death counts for
each cause
deaths_by_year = df.groupby('Year')[['Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
```

```
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].sum()

# Visualizing trends for the top 5 causes of death
using scatter plots
top_5_causes = deaths_by_year[['Cardiovascular
Diseases', 'Neoplasms', 'Chronic Respiratory
Diseases',
                               'Lower Respiratory
Infections', 'Tuberculosis']]

plt.figure(figsize=(12, 6))
for cause in top_5_causes.columns:
    plt.scatter(top_5_causes.index,
top_5_causes[cause], label=cause)

plt.title('Trends of Top 5 Causes of Death Over
Time')
plt.xlabel('Year')
plt.ylabel('Number of Deaths')
plt.legend()
plt.grid(True)
plt.show()


# Calculate Year-over-Year percentage change
yoy_change = deaths_by_year.pct_change() * 100
```

```
# Visualizing percentage changes for top 5 causes of
death
plt.figure(figsize=(12, 6))
for cause in top_5_causes.columns:
    plt.plot(yoy_change.index, yoy_change[cause],
label=cause)

plt.title('Year-over-Year Percentage Change in Top 5
Causes of Death')
plt.xlabel('Year')
plt.ylabel('Percentage Change (%)')
plt.legend()
plt.grid(True)
plt.show()


# Sum of death causes by country
deaths_by_country =
df.groupby('Country/Territory')[['Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
```

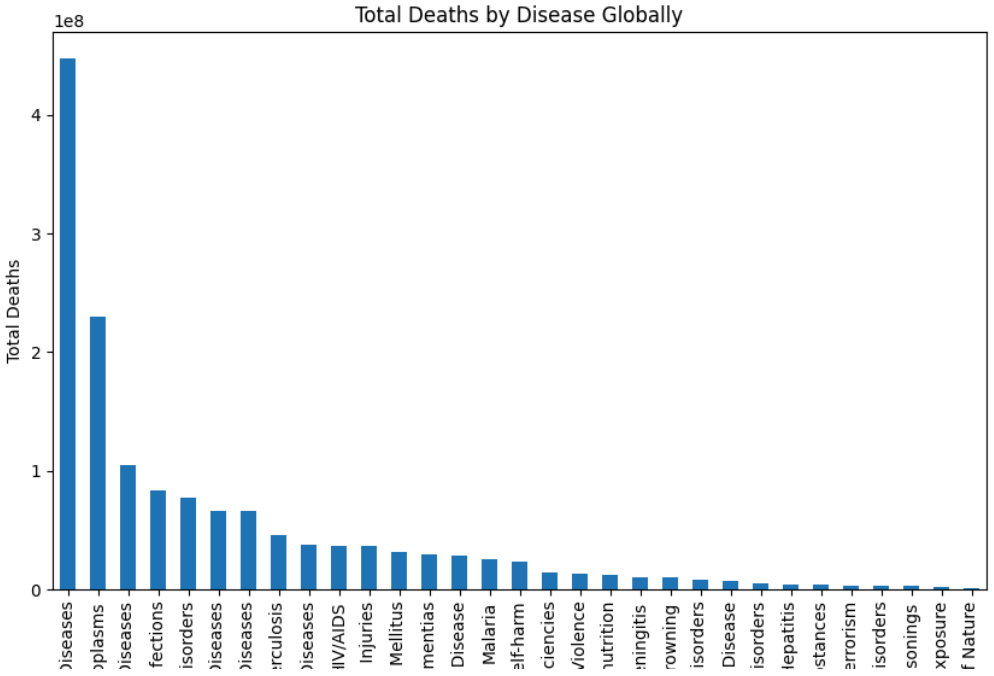
```
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].sum()

plt.figure(figsize=(12, 8))
sns.heatmap(deaths_by_country, cmap="YlGnBu",
linewidths=0.5)
plt.title('Heatmap of Death Causes by Country')
plt.xlabel('Death Cause')
plt.ylabel('Country')
plt.show()


# Heatmap for top 5 death causes by country
top_5_heatmap_data =
deaths_by_country[['Cardiovascular Diseases',
'Neoplasms',
                                        'Chronic
Respiratory Diseases', 'Lower Respiratory
Infections',

'Tuberculosis']]
```

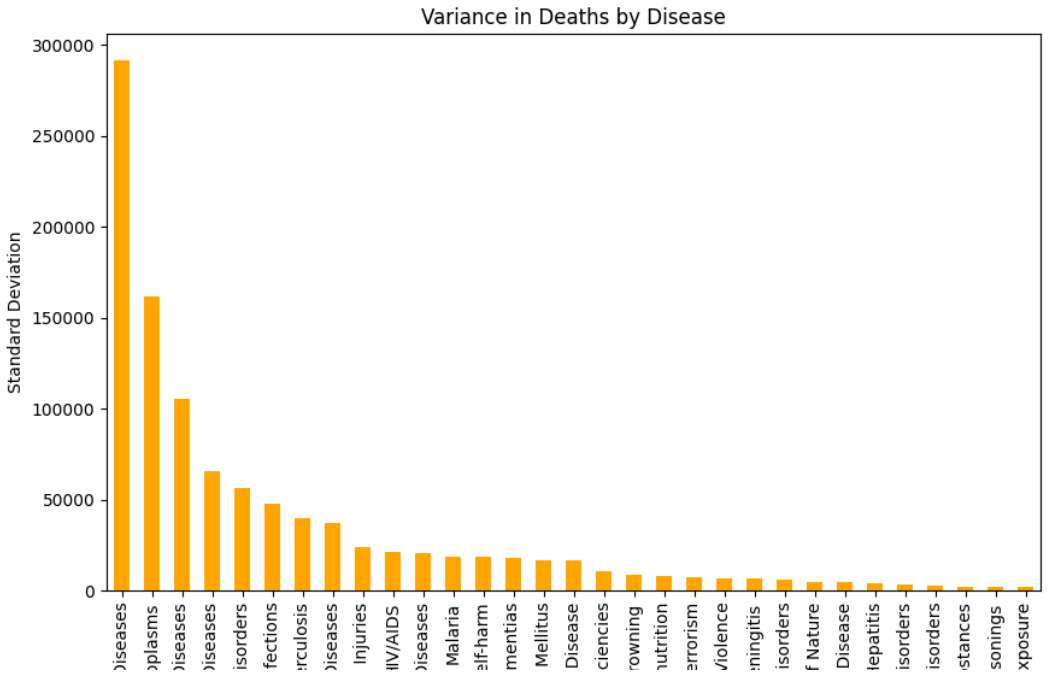
```
plt.figure(figsize=(12, 8))
sns.heatmap(top_5_heatmap_data, cmap="Reds",
linewidths=0.5)
plt.title('Heatmap of Top 5 Causes of Death by
Country')
plt.xlabel('Death Cause')
plt.ylabel('Country')
plt.show()


# Regional Comparisons
region_mapping = {
    'United States': 'North America',
    'Canada': 'North America',
    'Brazil': 'South America',
    'Germany': 'Europe',
    'India': 'Asia',
    'Nigeria': 'Africa',}

df['Region'] =
df['Country/Territory'].map(region_mapping)

regional_deaths =
df.groupby('Region')[['Meningitis',
       'Alzheimer Disease and Other Dementias',
'Parkinson Disease',
       'Nutritional Deficiencies', 'Malaria',
'Drowning',
       'Interpersonal Violence', 'Maternal
Disorders', 'HIV/AIDS',
       'Drug Use Disorders', 'Tuberculosis',
'Cardiovascular Diseases',
```

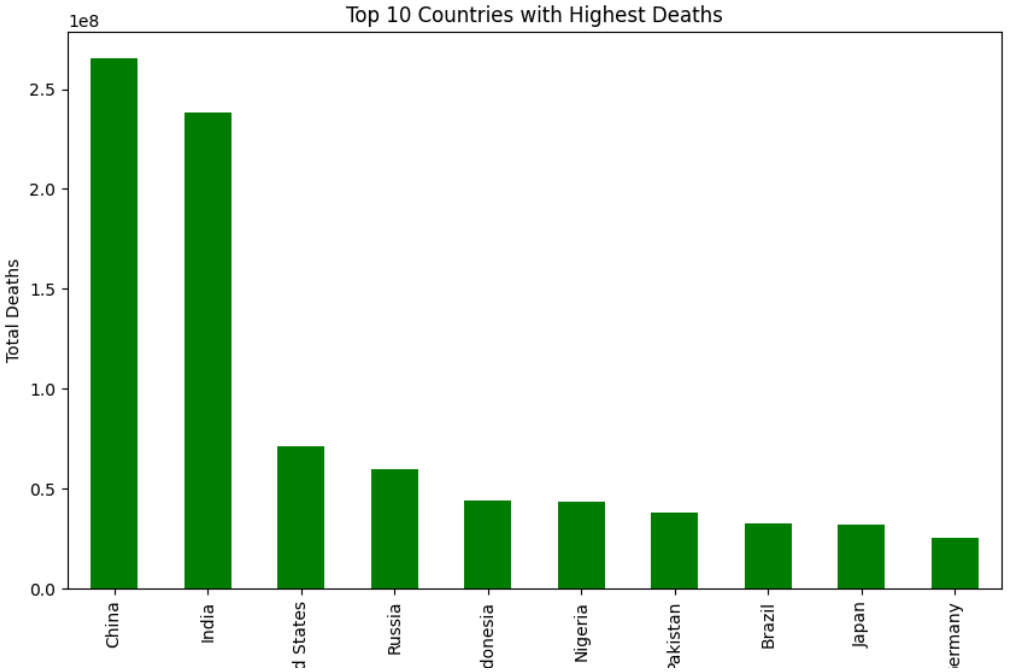
```
       'Lower Respiratory Infections', 'Neonatal
Disorders',
       'Alcohol Use Disorders', 'Self-harm',
'Exposure to Forces of Nature',
       'Diarrheal Diseases', 'Environmental Heat and
Cold Exposure',
       'Neoplasms', 'Conflict and Terrorism',
'Diabetes Mellitus',
       'Chronic Kidney Disease', 'Poisonings',
'Protein-Energy Malnutrition',
       'Road Injuries', 'Chronic Respiratory
Diseases',
       'Cirrhosis and Other Chronic Liver Diseases',
'Digestive Diseases',
       'Fire/ Heat/ and Hot Substances', 'Acute
Hepatitis']].sum()

print(regional_deaths)

top_5_death_causes = ['Cardiovascular Diseases',
'Neoplasms',
                      'Chronic Respiratory
Diseases', 'Lower Respiratory Infections',
                      'Tuberculosis']

plt.figure(figsize=(12, 8))
regional_deaths[top_5_death_causes].plot(kind='bar',
stacked=True)
plt.title('Regional Comparison of Top 5 Causes of
Death')
plt.xlabel('Region')
plt.ylabel('Total Deaths')
```

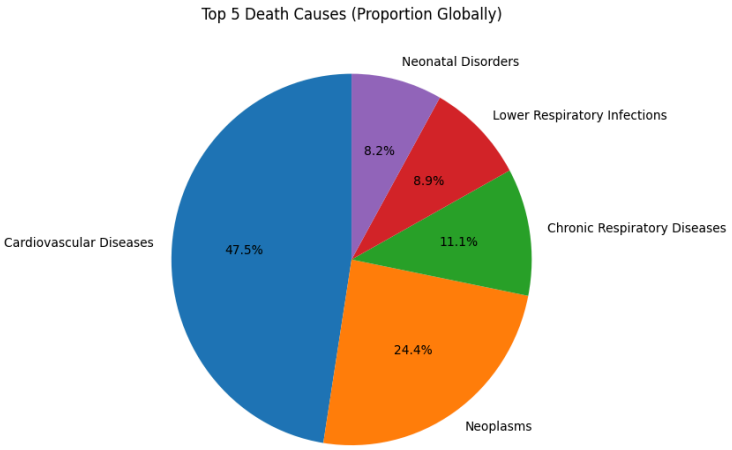
```
plt.legend(loc='upper right')
plt.show()

# specific health challenges unique to each region
plt.figure(figsize=(12, 8))
sns.heatmap(regional_deaths[top_5_death_causes],
cmap="Blues", linewidths=0.5)
plt.title('Heatmap of Top 5 Death Causes by Region')
plt.xlabel('Death Causes')
plt.ylabel('Region')
plt.show()
```

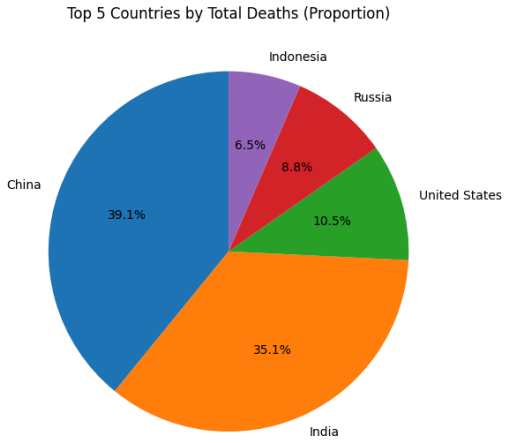
4.Visualization Examples:

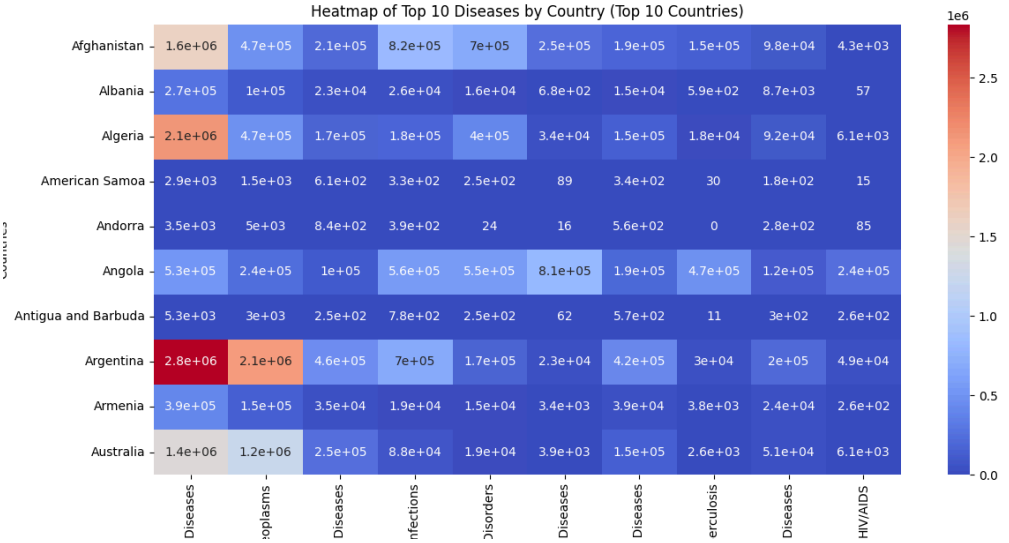


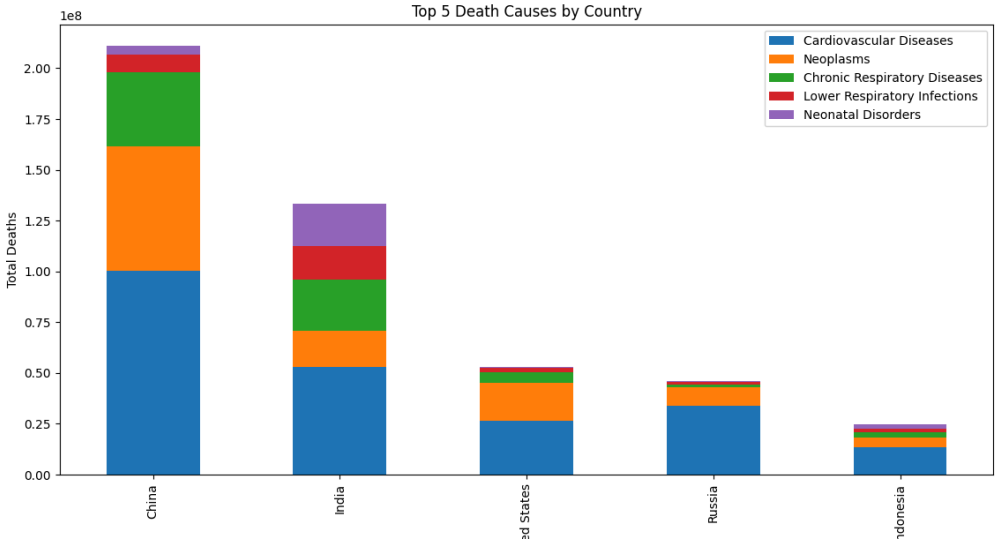


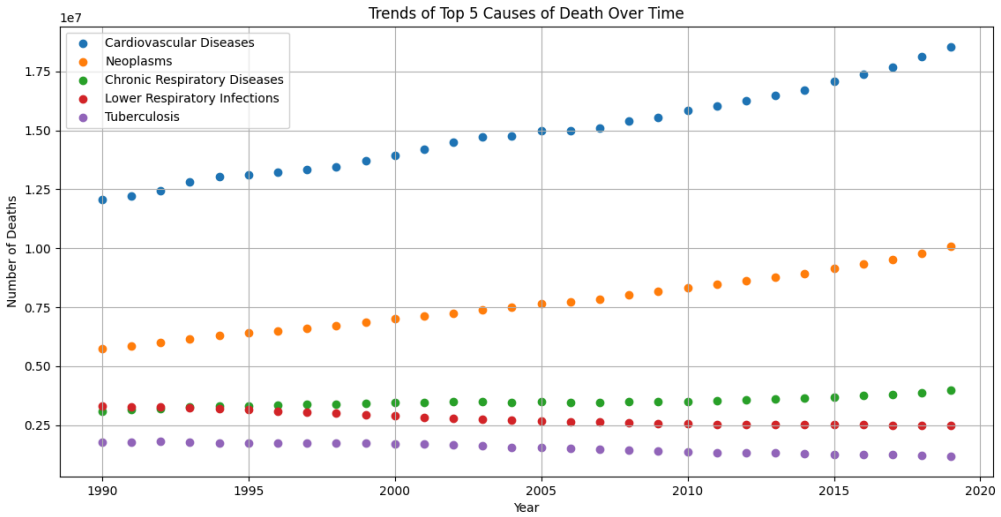

Top 10 Countries with Highest Deaths
1e8
Total Deaths
2.5
2.0
1.5
1.0
0.5
0.0
China
India
Russia
Nigeria
Brazil
Japan

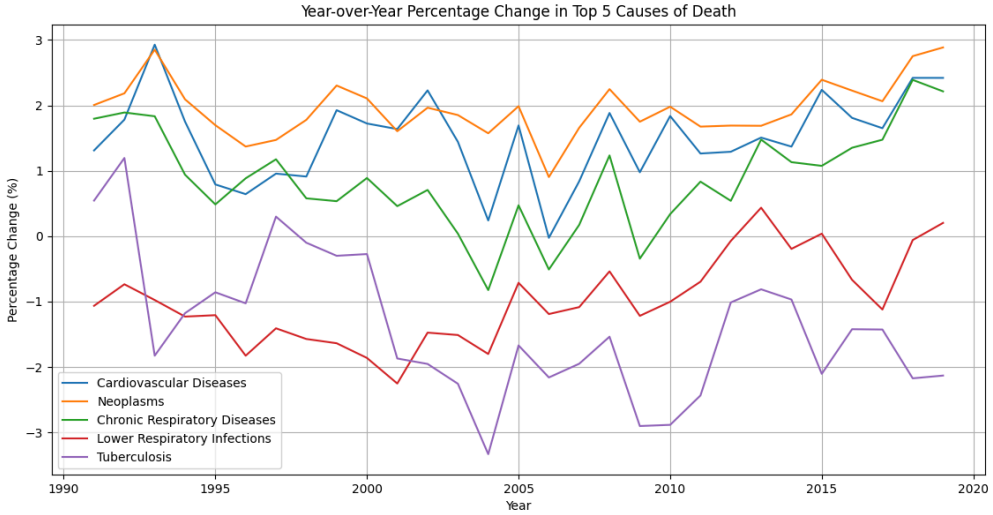

Top 5 Death Causes (Proportion Globally)
Neonatal Disorders
Lower Respiratory Infections
Chronic Respiratory Diseases
Cardiovascular Diseases
Neoplasms
8.2%
8.9%
11.1%
47.5%
24.4%

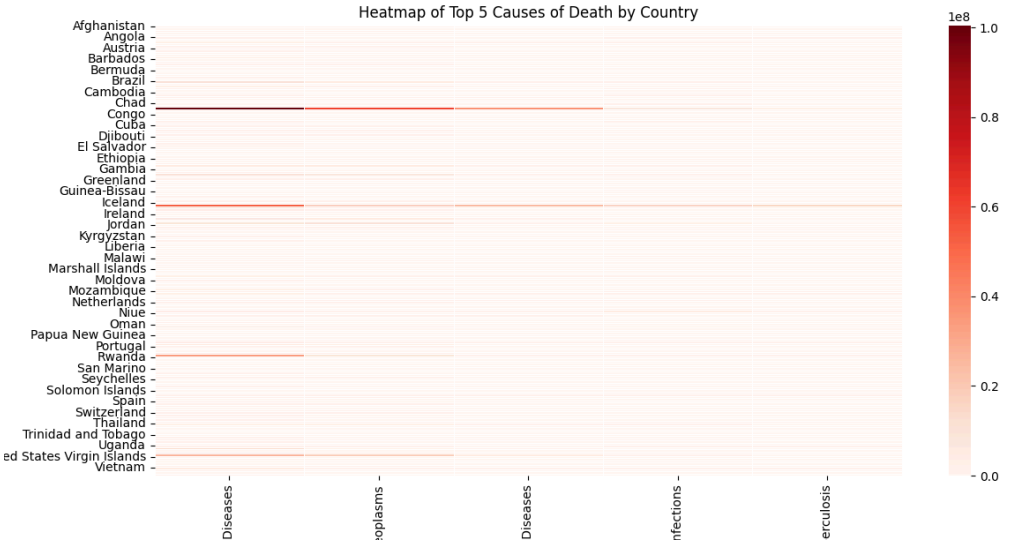

Top 5 Countries by Total Deaths (Proportion)
Indonesia
Russia
China
United States
India
6.5%
8.8%
39.1%
10.5%
35.1%

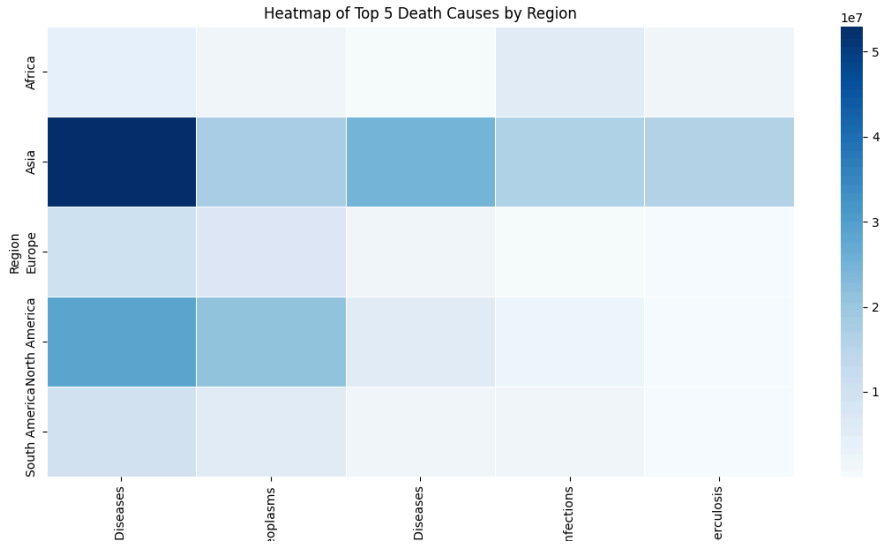

Heatmap of Top 10 Diseases by Country (Top 10 Countries)
Afghanistan 1.6e+06 4.7e+05 2.1e+05 8.2e+05 7e+05 2.5e+05 1.9e+05 1.5e+05 9.8e+04 4.3e+03
Albania 2.7e+05 1e+05 2.3e+04 2.6e+04 1.6e+04 6.8e+02 1.5e+04 5.9e+02 8.7e+03 57
Algeria 2.1e+06 4.7e+05 1.7e+05 1.8e+05 4e+05 3.4e+04 1.5e+05 1.8e+04 9.2e+04 6.1e+03
American Samoa 2.9e+03 1.5e+03 6.1e+02 3.3e+02 2.5e+02 89 3.4e+02 30 1.8e+02 15
Andorra 3.5e+03 5e+03 8.4e+02 3.9e+02 24 16 5.6e+02 0 2.8e+02 85
Angola 5.3e+05 2.4e+05 1e+05 5.6e+05 5.5e+05 8.1e+05 1.9e+05 4.7e+05 1.2e+05 2.4e+05
Antigua and Barbuda 5.3e+03 3e+03 2.5e+02 7.8e+02 2.5e+02 62 5.7e+02 11 3e+02 2.6e+02
Argentina 2.8e+06 2.1e+06 4.6e+05 7e+05 1.7e+05 2.3e+04 4.2e+05 3e+04 2e+05 4.9e+04
Armenia 3.9e+05 1.5e+05 3.5e+04 1.9e+04 1.5e+04 3.4e+03 3.9e+04 3.8e+03 2.4e+04 2.6e+02
Australia 1.4e+06 1.2e+06 2.5e+05 8.8e+04 1.9e+04 3.9e+03 1.5e+05 2.6e+03 5.1e+04 6.1e+03
Diseases
HIV/AIDS
1e6
2.5
2.0
1.5
1.0
0.5
0.0

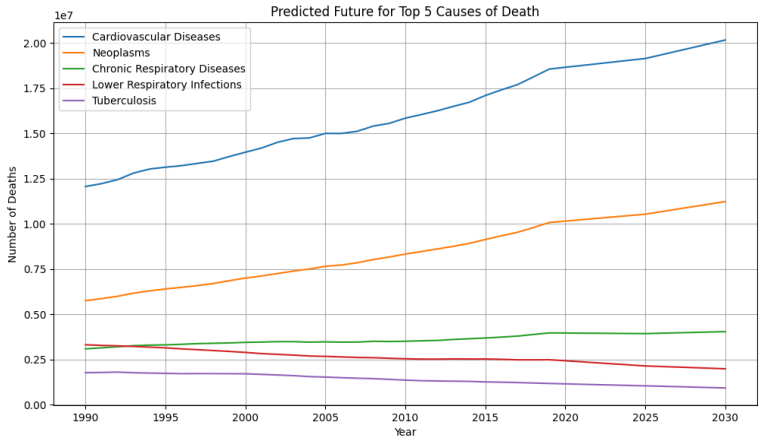

Top 5 Death Causes by Country
1e8
Total Deaths
2.00
1.75
1.50
1.25
1.00
0.75
0.50
0.25
0.00
Cardiovascular Diseases
Neoplasms
Chronic Respiratory Diseases
Lower Respiratory Infections
Neonatal Disorders
China
India
Russia

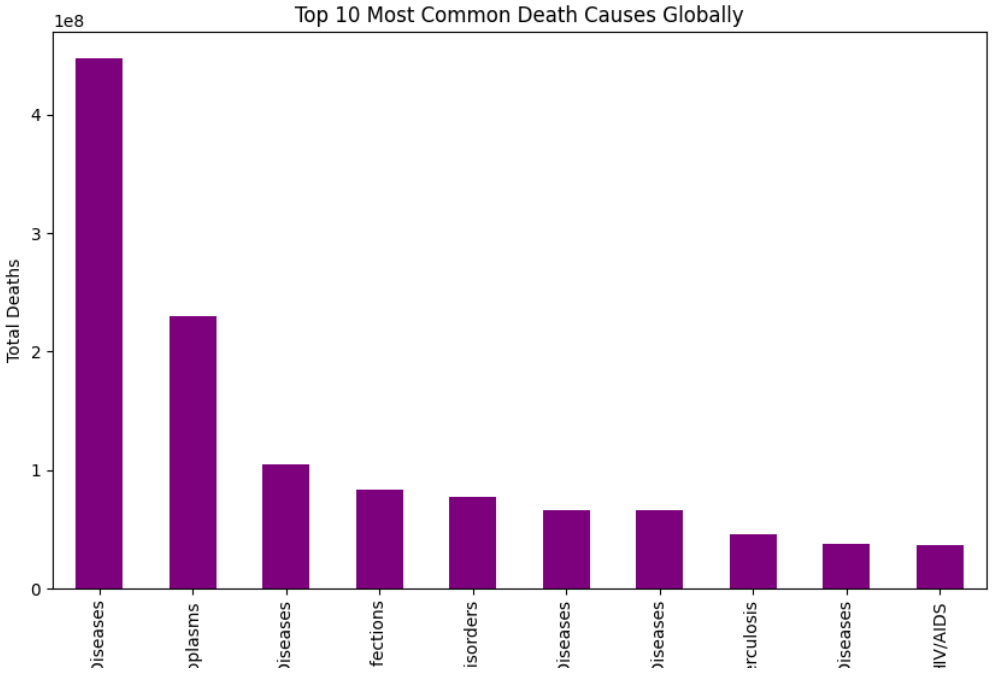

Trends of Top 5 Causes of Death Over Time
1e7
Cardiovascular Diseases
Neoplasms
Chronic Respiratory Diseases
Lower Respiratory Infections
Tuberculosis
Number of Deaths
Year


Year-over-Year Percentage Change in Top 5 Causes of Death
Cardiovascular Diseases
Neoplasms
Chronic Respiratory Diseases
Lower Respiratory Infections
Tuberculosis
Percentage Change (%)
Year


Heatmap of Top 5 Causes of Death by Country
1e8
Afghanistan
Angola
Austria
Barbados
Bermuda
Brazil
Cambodia
Chad
Congo
Cuba
Djibouti
El Salvador
Ethiopia
Gambia
Greenland
Guinea-Bissau
Iceland
Ireland
Jordan
Kyrgyzstan
Liberia
Malawi
Marshall Islands
Moldova
Mozambique
Netherlands
Niue
Oman
Papua New Guinea
Portugal
Rwanda
San Marino
Seychelles
Solomon Islands
Spain
Switzerland
Thailand
Trinidad and Tobago
Uganda
Vietnam


Heatmap of Top 5 Death Causes by Region
1e7
Region
Africa
Asia
Europe
North America
South America
Diseases
Diseases
erculosis


Predicted Future for Top 5 Causes of Death
1e7
Cardiovascular Diseases
Neoplasms
Chronic Respiratory Diseases
Lower Respiratory Infections
Tuberculosis
Number of Deaths
Year


Top 10 Most Common Death Causes Globally
1e8
Total Deaths
iseases
plasms
iseases
fections
isorders
iseases
iseases
rculosis
iseases
IV/AIDS

5. Predictive Modeling :

Base on the dataset we predicted which diseases would be dominant in the next 5 years starting from 2025

Code Example

```
#Predictive Analysis

top_5_causes = deaths_by_year[['Cardiovascular
Diseases', 'Neoplasms',
                               'Chronic Respiratory
Diseases',
                               'Lower Respiratory
Infections',
                               'Tuberculosis']]


X = np.array(top_5_causes.index).reshape(-1, 1)
y = top_5_causes
X_train, X_test, y_train, y_test =
train_test_split(X, y, test_size=0.2,
random_state=42)

models = {}
predictions = {}

for cause in top_5_causes.columns:
    model = LinearRegression()
    model.fit(X_train, y_train[cause])
    models[cause] = model
    predictions[cause] = model.predict(X_test)
```

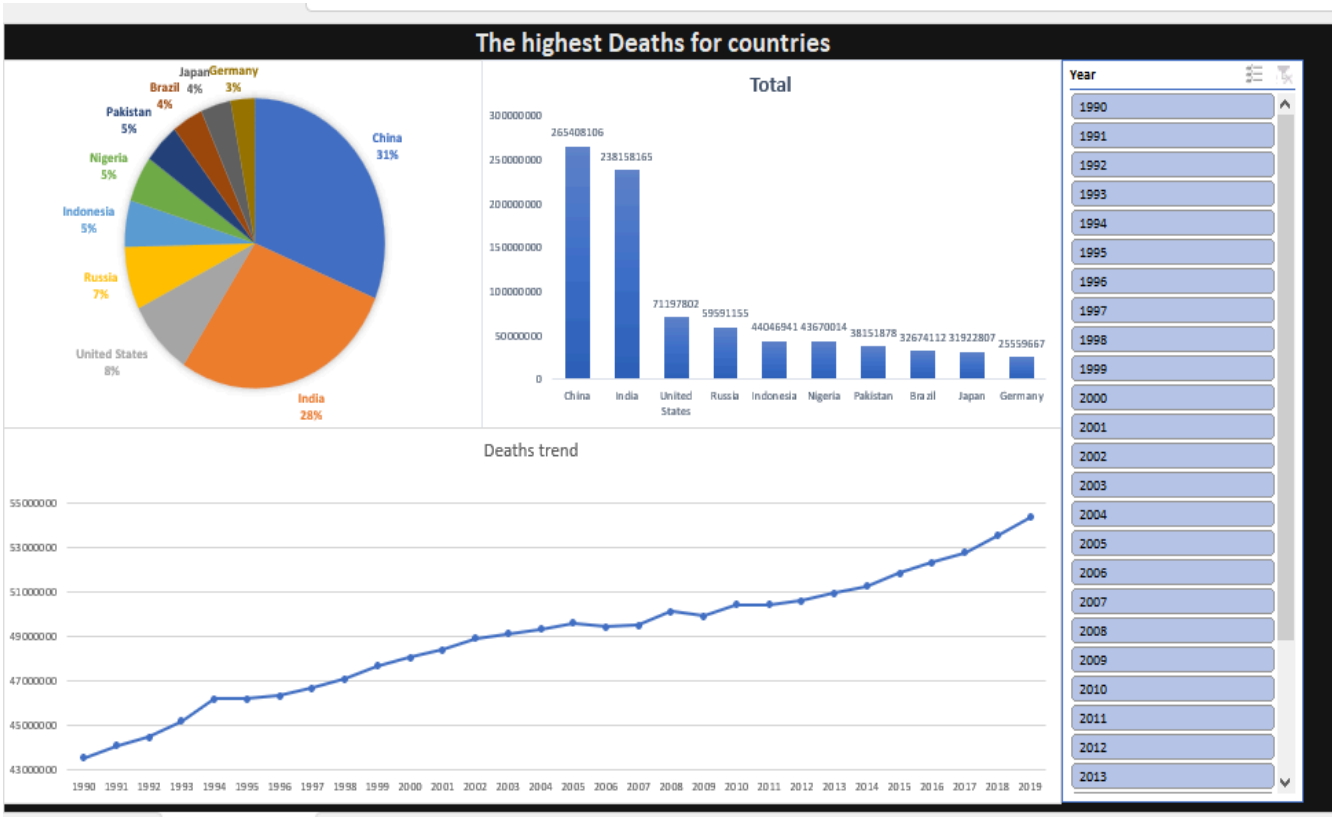
```
    error = mean_absolute_error(y_test[cause],
predictions[cause])
    print(f"Mean Absolute Error for {cause}:
{error:.2f}")

# deaths for the next 5 years
future_years = np.array([2025, 2026, 2027, 2028,
2029, 2030]).reshape(-1, 1)
future_predictions = {cause:
model.predict(future_years) for cause, model in
models.items()}

plt.figure(figsize=(12, 6))
for cause in top_5_causes.columns:
    plt.plot(np.append(X.flatten(),
future_years.flatten()),
             np.append(y[cause],
future_predictions[cause]), label=cause)

plt.title('Predicted Future for Top 5 Causes of
Death')
plt.xlabel('Year')


plt.ylabel('Number of Deaths')

plt.legend()
plt.grid(True)
plt.show()
```

6. Excel Pivot Analysis:

Excel is used to create pivot tables that summarize key findings from the dataset. This phase also includes the creation of basic visual charts

Objective:
To use Excel's capabilities for further data exploration, identifying key patterns using pivot tables and conditional formatting.

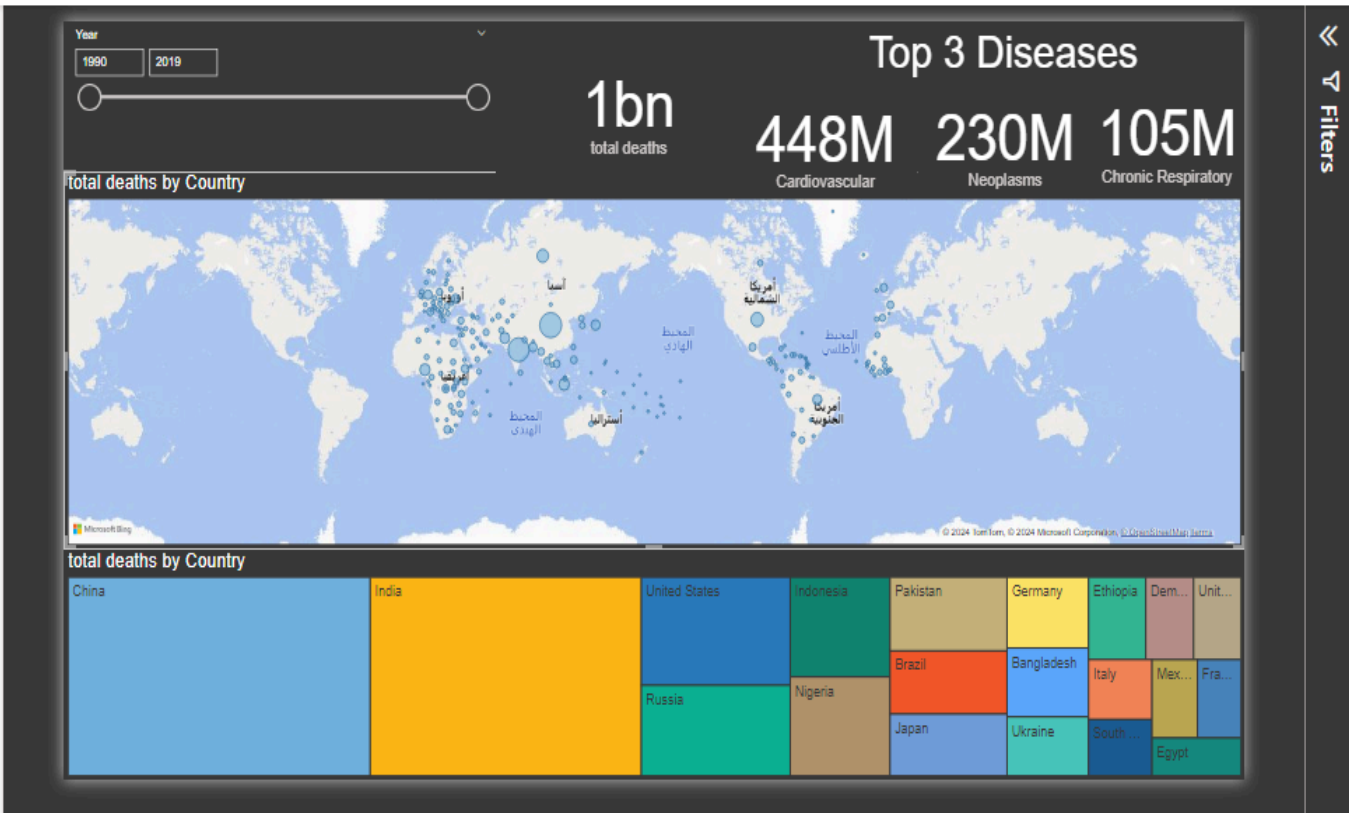
Visualization Example:

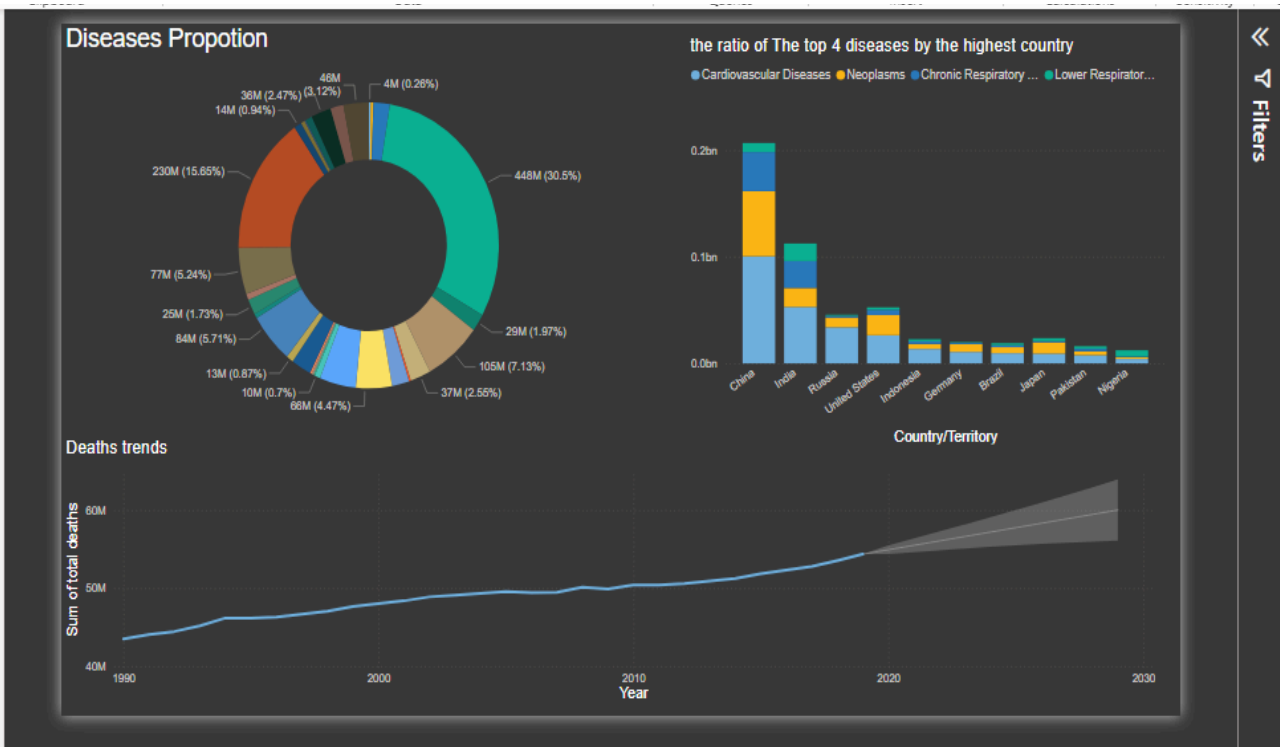


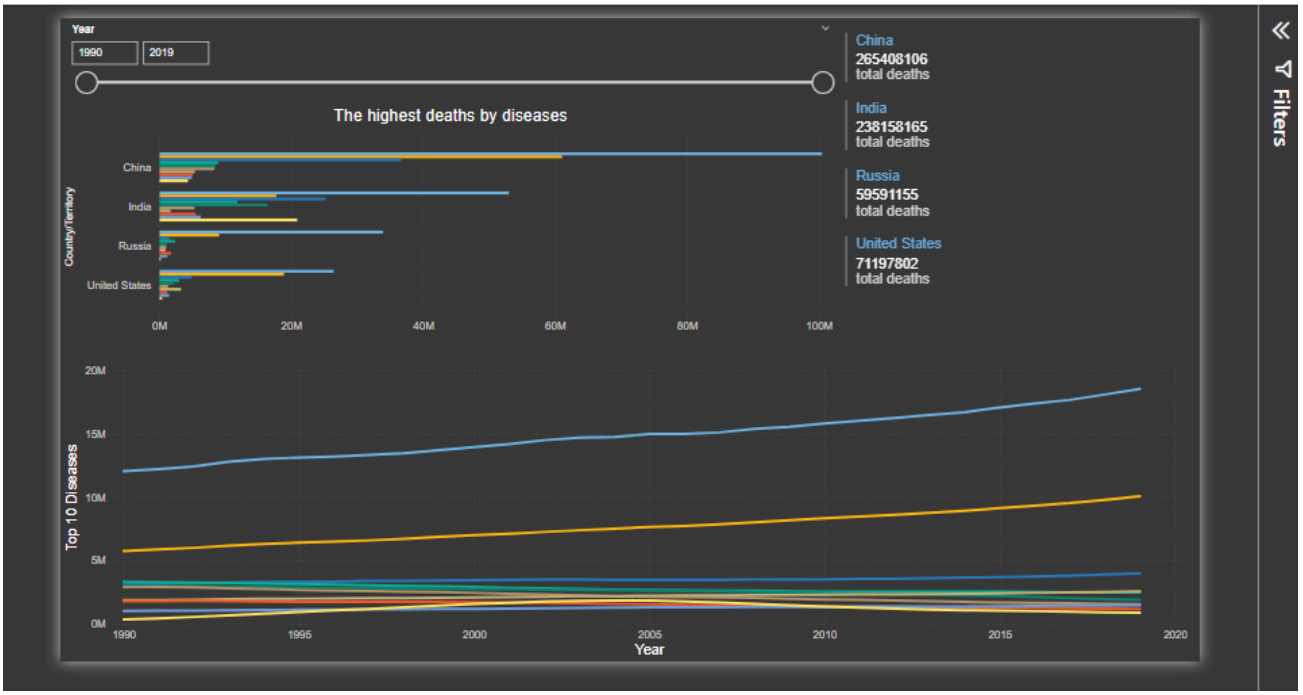
7. Power BI Dashboard :

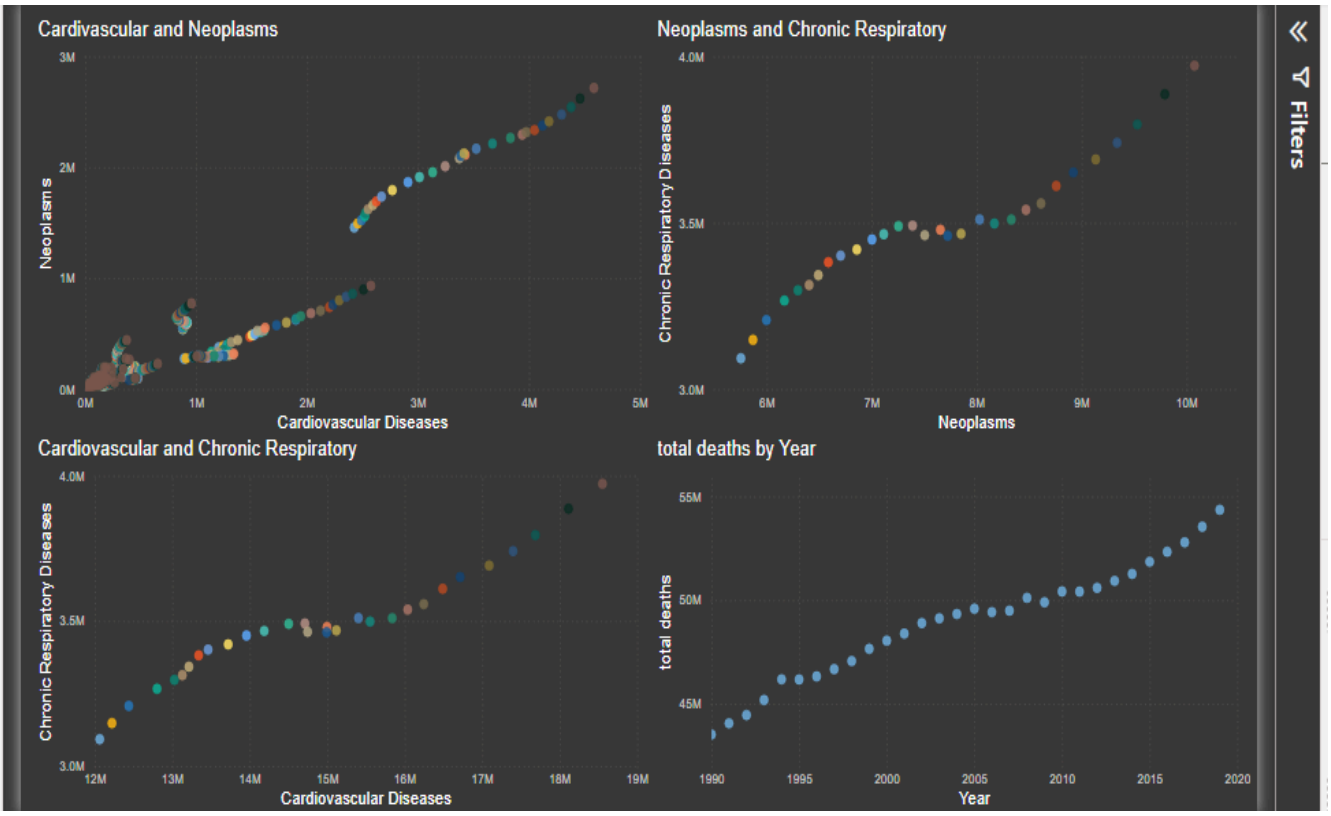
A comprehensive and interactive dashboard is developed using Power BI, allowing users to filter the data by various factors such as region, year, and cause of death. This dashboard offers a visual overview of global mortality trends with features that enhance user engagement and understanding.

Objective:
To create an intuitive and dynamic dashboard that allows for real-time interaction with the data, providing valuable insights at a glance.




Diseases Propotion
4M (0.26%)
448M (30.5%)
29M (1.97%)
105M (7.13%)
37M (2.56%)
66M (4.47%)
10M (0.7%)
13M (0.87%)
84M (5.71%)
25M (1.73%)
77M (5.24%)
230M (15.65%)
14M (0.94%)
36M (2.47%)
46M (3.12%)
the ratio of The top 4 diseases by the highest country
Cardiovascular Diseases
Neoplasms
Chronic Respiratory ...
Lower Respirator...
0.2bn
0.1bn
0.0bn
China
India
Russia
United States
Indonesia
Germany
Brazil
Japan
Pakistan
Nigeria
Country/Territory
Deaths trends
Sum of total deaths
60M
50M
40M
1990
2000
2010
2020
2030
Year
Filters


Year
1990
2019
The highest deaths by diseases
China
India
Russia
United States
Country/Territory
0M
20M
40M
60M
80M
100M
China
265408106
total deaths
India
238158165
total deaths
Russia
59591155
total deaths
United States
71197802
total deaths
20M
15M
10M
5M
0M
Top 10 Diseases
1990
1995
2000
2005
2010
2015
2020
Year
Filters



Conclusion:

This project has provided a comprehensive look into the global causes of death, uncovering trends and insights that can inform public health decisions. From detailed data cleaning to advanced visualizations, each phase of the project has contributed to a holistic understanding of mortality trends across different regions and time periods.