Data Science & Analytics

Louisville, Kentucky

2022-2023

#603893

Table of Contents

Introduction and Data Overview

Predicting Heart Disease Risk Drivers Using Machine Learning Algorithms and Global Insights on Cardiovascular Mortalities

Cardiovascular diseases (CVDs) have consistently been the leading cause of death across the globe. An estimated 18 million people die from CVDs every year, representing 32% of all global deaths. 85% of these CVD deaths are due to **heart attacks and strokes**. It is important to note that 38% of all premature deaths (under the age of 70 years), which were about 17 million in 2019, were due to CVDs (Hannah Ritchie et. al., 2019). The medical community believes that most cardiovascular diseases can be prevented by addressing behavioral risk factors such as tobacco use, unhealthy diet and obesity, physical inactivity, and harmful use of alcohol. It is important to detect early signals (or drivers) of cardiovascular disease to prevent deaths through effective management with counseling and medicines.

Cardiovascular diseases (CVD) refer to a group of diseases that involves the **heart or blood vessels**. CVD includes coronary artery diseases (CAD) such as angina and myocardial infarction (commonly known as a heart attack). Other CVDs include stroke, heart failure, hypertensive heart disease, rheumatic heart disease, cardiomyopathy, abnormal heart rhythms, congenital heart disease, valvular heart disease, carditis, aortic aneurysms, peripheral artery disease, thromboembolic disease, and venous thrombosis.

In a study done by Harvard professors, there are multiple categories to account for when estimating the cost of cardiovascular diseases, starting from screening primary prevention to acute hospital care and loss of productivity. Overall, the cost of treating CVDs could be as high as **US $20 trillion** from 2010 to 2030. It is estimated that 55% of this cost is attributable to healthcare, and the remaining 45% is primarily due to productivity loss from disability or premature death (Bloom et. al., 2011; RTI News Article, 2017).

Data Dictionary

Metadata for the dataset used to build ML models:

heart_statlog_cleveland_hungary_final (heart disease dataset): This dataset consists of 1190 records of patients from the US, UK, Switzerland, and Hungary. It has 11 features and 1 target variable (Manu Siddhartha, IEEE Dataport, 2020).

Attribute Name	Description	Datatype
Age	Patient's Age in years	Numeric
Sex	Patient's Gender - Male as 1 Female as 0	Binary
Chest Pain Type	Type of chest pain categorized into 1-typical, 2-typical angina, 3-non-anginal pain, 4-asymptomatic	Numeric
Resting BP	Level of blood pressure at resting mode in mm/HG	Numeric
Cholesterol	Serum cholesterol in mg/dl	Numeric
Fasting Blood Sugar	Blood sugar levels on fasting > 120 mg/dl represent 1 in case of true and 0 as false	Binary
Resting ECG	Result of an electrocardiogram while at rest – 0: Normal; 1: Abnormality in ST-T wave; 2: Left ventricular hypertrophy	Numeric
Max Heart Rate	Maximum heart rate achieved	Numeric
Exercise Angina	Angina induced by exercise - 0 depicts NO; 1 depicts Yes	Binary
Oldpeak	Exercise-induced ST-depression* in comparison with the state of rest	Numeric
ST Slope	ST segment* measured in terms of the slope during peak exercise - 0: Normal; 1: Upsloping; 2: Flat; 3: Down-sloping	Numeric
Target	Heart Risk - 1 indicates heart disease; 0 indicates normal	Binary

*ST-segment is the line between the “S” and the “T” on the readout of an EKG. If a person is in good health, the line appears at or close to the baseline level. A depressed or elevated ST segment indicates the presence of an underlying health condition.

Datasets from the UN and WHO were used for Global level insights at a country level:

Dataset	**Key Attributes**
Cardiovascular-disease-deaths-by-age	Cardiovascular Deaths
Cross-country-literacy-rates	Literacy Rates - share of the population older than 14 years that can read and write.
Human-development-index	Human Development Index: a long and healthy life, a good education, and a decent standard of living.
Population-since-1800	Percent WRT Population
Poverty-explorer	Share of population living in extreme poverty - Extreme poverty is defined as living below the International Poverty Line of $2.15 per day.
Share-of-adults-Alcohol	Consumption of alcohol is measured in liters of pure alcohol per person aged 15 or older
Share-of-adults-overweight	Share of adults that are overweight or obese
Share-of-adults-who-smoke	The share of men and women aged 15 and older who smoke any tobacco product on a daily or non-daily basis.
Share-population-female	Share of the female population

Purpose

The main objective of this study is to examine the global distribution of cardiovascular deaths and its relationship with behavioral and socio-economic factors and build machine learning models to predict CVD deaths and understand its drivers. Country-level data on CVDs and behavioral and socio-economic factors were extracted from various sites for this study. The key insights from this analysis can help WHO (World Health Organization) and national and state-level governments to develop data-driven plans to identify risk factors and proactively manage this disease to reduce deaths and its economic impact on society. The findings from this study will help develop cost-effective strategies to build preventive plans to reduce global deaths by focusing on the education and preventive healthcare of the top 20 nations that account for 70% of all CVD deaths.

Methods

A Machine Learning model has been created using a curated dataset that consists of 1190 patients from the United States, United Kingdom, Switzerland, and Hungary. The Global level insights have been created using United Nations and WHO data through datasets available in “Our World in Data"; the links to all the data files are in the References section. Data used for the global analysis consists of three decades, ranging from 1996 to 2019.

Exploratory Data Analysis

All the data files were extracted as .csv files, and analysis was done using Excel and Microsoft SQL Server. The first step was performing EDA on each dataset to understand the data in detail.
The analysis, performed in Excel, was done to understand the relationship between socioeconomic, demographic, and behavioral attributes with cardiovascular deaths.

Data Transformation

The “Final” dataset was created in SQL by combining key attributes from the above-mentioned datasets (from United Nations and World Health Organization) at a country level and yearly trend, ranging from 1990 to 2019.

Machine Learning Model to Predict Heart Risk

Multiple ML techniques were used to predict heart risk using the heart disease (mentioned in the Data Dictionary) dataset. Classification models - Decision Tree, Naïve Bayes, and Neural Network - were created using Visual Studio. The Decision Tree model was chosen due to its accuracy and explanatory power to explain the drivers of the model.

Results

Global Analysis & Insights on Cardiovascular Deaths

Data on CVD deaths were merged with socioeconomic attributes of individual countries, such as the percentage of the population with respect to the global population, the percentage of the population above 65 years old, the percentage of the population in extreme poverty, GDP per capita, percentage of the population that consumes Alcohol, percentage of the population that consumes Tobacco, Cardiovascular death, and Human Development Index. A derived variable - CVD Deaths percentage with respect to the population of that country was created.

CVD continues to be the leading cause of death worldwide, increasing from 28% in 1990 to 34% in 2019 (Figure 1). Deaths due to Cancer exhibited the second highest deaths and relative increase during this period. Alzheimer's and Diabetes showed significant positive relative change during this period while Diarrheal, Neonatal disorders, and Tuberculosis have seen a significant decline over this period.

Relative Change in the Distribution of Top Diseases in the Last 3 Decades (1990-2019)

Deaths by top Diseases	1990	2019	Relative Change (2019-1990)/1990
Cardiovascular Diseases	28%	34%	**23.1%**
Neoplasms (Cancer)	13%	19%	**40.2%**
Lower respiratory infections	8%	5%	**-39.9%**
Chronic respiratory diseases	7%	7%	2.8%
Neonatal disorders	7%	3%	**-49.9%**
Diarrheal diseases	7%	3%	**-57.6%**
Digestive diseases	4%	5%	10.4%
Tuberculosis	4%	2%	**-46.9%**
Diabetes mellitus	2%	3%	**87.6%**
Alzheimer	1%	3%	**131.7%**

Figure 1

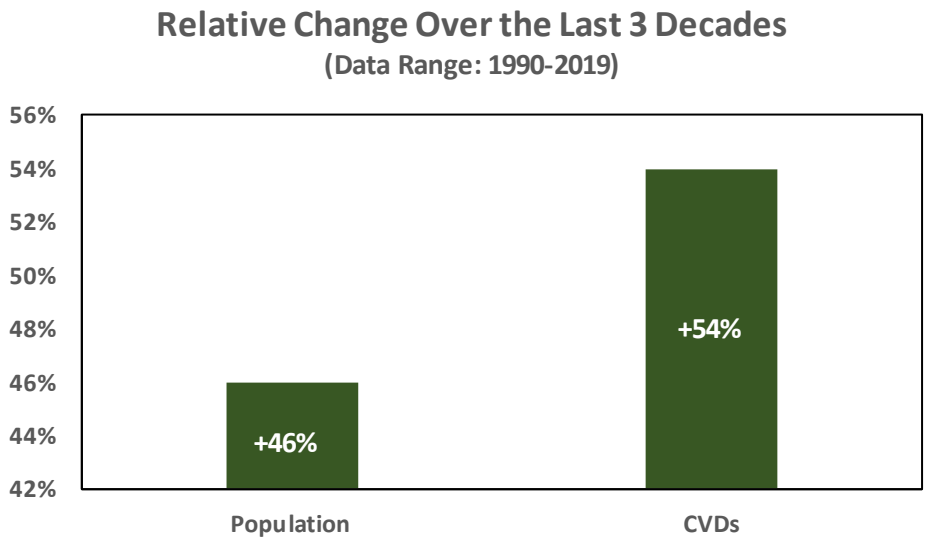


Figure 2

Figure 2 shows the relative change in global population (GP) [(GP2019-GP1990)/GP1990] and Cardiovascular Deaths (CVD) [(CVD2019-CVD1990)/CVD1990] over three decades from 1990 to 2019. The global population grew by 46% during this period while CVD deaths grew by 54%, outpacing population growth in the same period.

Distribution of CVDs by Age Groups in the last 3 Decades (1990-2019)

Distribution by Age Groups	% WRT All CVD Deaths	Change (2019-1990)
<5Yrs	0.1%	-0.5%
5-14Yrs	0.1%	-0.1%
15-49Yrs	6.7%	-1.5%
50-69Yrs	27.5%	-4.3%
>=70Yrs	65.6%	6.5%

Figure 3

Figure 3 shows the global distribution of CVD deaths by age group and its relative change over three decades [(2019-1990)/1990]. A significant increase in the percentage of deaths from older age groups is seen at the expense of a decline in other age groups.

Global Distribution of Change in CVDs WRT Population in the Last 3 Decades (1990-2019)

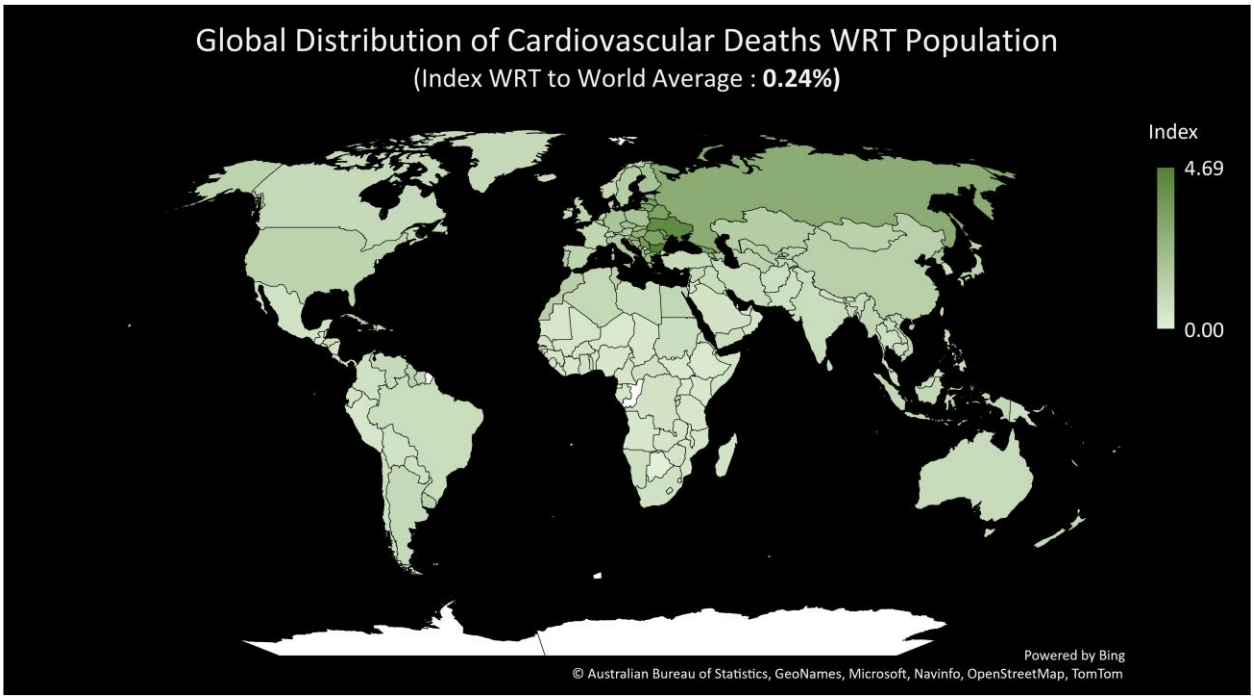


Figure 4

Figure 4 is indexed with the global average (total CVD Deaths WRT Global Population) to show which part of the globe is over-represented in CVD deaths. This chart indicates a higher index (higher CVD deaths WRT population) for the Siberia part of Russia, Eastern Europe, and Central Asia. The majority of the African continent, Australia, and South American countries indicate a lower index for CVD deaths.

Predicting Heart Risk Using Machine Learning Model & Understanding its Drivers

This section uses patient-level data on heart failure risk curated from five different data sets on a common set of 11 attributes from 4 different countries. These attributes cover demographic information about the patient and key risk attributes ranging from blood pressure to heart rate performance during exercise. This data was divided into training and testing data sets (70% training and 30% Testing). The training dataset was fed into Machine Learning Classification algorithms using Microsoft SQL Server 2020 and Visual Studio 2019. Multiple Machine Learning models were created to predict heart risk. All models showed relatively close classification accuracy and precision. Finally, the Decision Tree ML model was selected based on its explanatory power. Figure 5 below describes the methodology used for this work.

Methodology Used for Training and Testing ML Model to Predict Heart Failure Risk

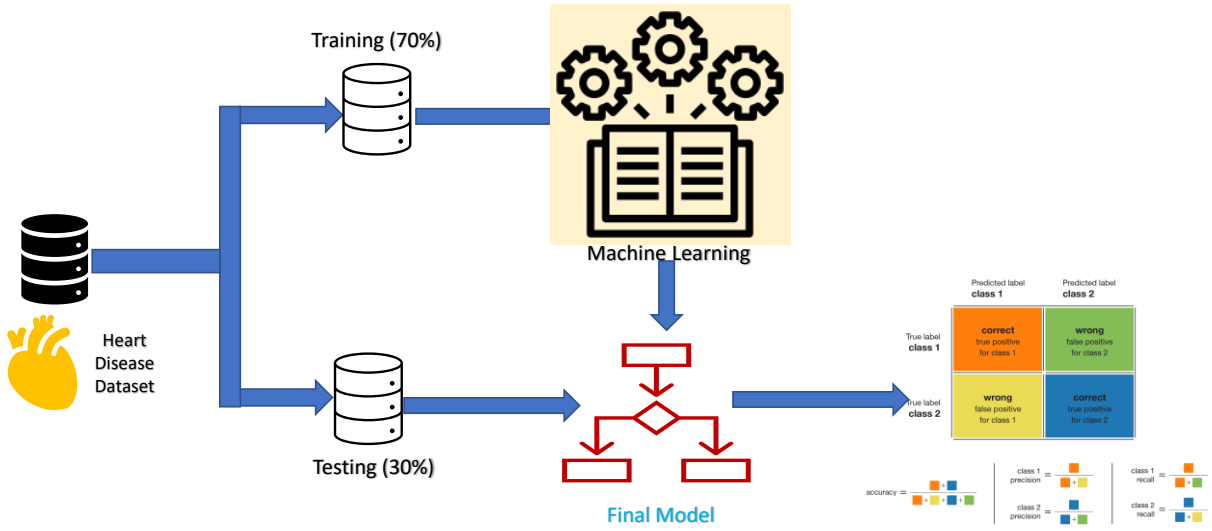


Figure 5

The classification accuracy of this decision tree model on the training data set is shown below in Figure 6. Overall, the accuracy of the model is 80%, which is reasonably good accuracy. The precision and Recall for predicting Heart Risk is also in the range of lower than 80%. These numbers exhibit the relative stability of the model.

Model Strength and Stability: Accuracy, Precision, and Recall of ML Model

Decision Tree ML Model	True Class	
Predicted	0 (Actual)	1 (Actual)
0	130	36
1	34	157

Precision for Target 1	Accuracy	Recall for Target 1
82%	80%	81%

Figure 6

Figure 7 below shows the dependency network of the decision tree model. This helps in understanding the importance of input attributes in predicting heart risk. ST Slope stands out as the most important attribute in predicting heart risk, followed by Chest Pain Type, Sex, and Max Heart Rate.

Dependency Network: Importance of Attributes

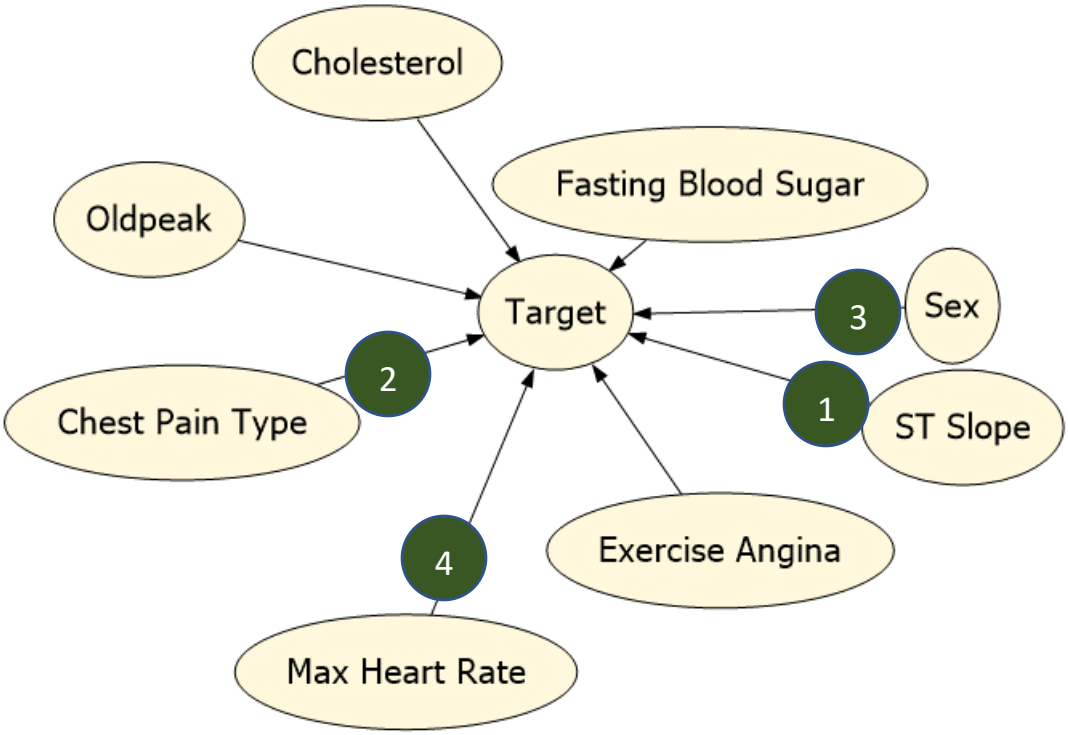


Figure 7

Attribute Definition and its Importance

Attribute Importance from Strongest to Least link	
ST Slope	Slope refers to the slope of each patient's stress test
Chest Pain Type	This refers to the Angina or Ischemic chest pain
Sex	Sex refers to the assigned sex of the patient at birth
Max Heart Rate	This refers to the maximum heart rate reached during a stress test. This is a hard variable because younger people have higher heart rates while older people have lower heart rates but are more likely to have heart problems
Cholesterol	This refers to cholesterol within the blood
Exercise Angina	Exercise Angina measures whether or not a patient had Angina (severe heart pain) during exercise
Fasting Blood Sugar	Fasting blood sugar measures a person's blood sugar and how it will likely change in the long term
Oldpeak	Oldpeak refers to whether or not the patient showed ST depression induced by exercise during the stress test

Figure 8

Figure 8 shows the detailed order of attribute importance in predicting heart risk.

Segmenting World Population Using Aggregated Attributes and Machine Learning Clustering Model

The change in Cardiovascular deaths by countries from 1990 to 2019 was examined to explore positive or negative trends in CVD over three decades. Figure 9 below shows the list of the top 15 countries that experienced a significant increase in CVDs in the last three decades. The majority of these countries belong to the eastern European region, where tobacco usage is on the higher side and the health development Index (HDI) is below the world's average. Earlier studies done on Tobacco (Ambuj Roy et.al., 2017, ResearchGate) and Overweight (Imre Csige et.al., 2018) are found to exhibit a significant correlation to CVDs, while a study on alcohol consumption (Jurgen Rehm et.al., 2016) did not establish a conclusive relationship with CVDs.

List of top 15 countries that experienced the highest increase in CVDs in the last three decades (1990-2019)

Top 15 Countries that Experienced Increase in Cardiovascular Deaths (CVD) WRT Population (2019-1990)		HDI	Alcohol	Overweight	Tobacco
Ukraine	**0.34%**	0.79	61.8	58.4	26.2
Bulgaria	**0.33%**	0.81	66.0	61.7	39.4
Moldova	**0.28%**	0.77	66.6	51.8	28.7
Bosnia and Herzegovina	**0.25%**	0.78	47.9	53.3	35.5
Montenegro	**0.25%**	0.84	54.0	59.4	31.8
Albania	**0.25%**	0.81	51.0	57.7	23.0
North Macedonia	**0.21%**	0.78	53.7	58.1	NA
Belarus	**0.21%**	0.82	73.6	59.4	30.9
Serbia	**0.20%**	0.81	60.0	57.1	40.1
Lithuania	**0.20%**	0.88	79.1	59.6	32.3
North Korea	**0.18%**	NA	40.3	32.4	17.8
Armenia	**0.18%**	0.78	26.0	54.4	25.8
Romania	**0.17%**	0.83	67.2	57.7	28.4
Dominican Republic	**0.13%**	0.77	43.0	61.2	10.9
Georgia	**0.13%**	0.81	35.1	54.2	31.7
Average of 15 Countries	0.22%	0.81	55.02	55.76	**28.75**
Global Average	0.01%	0.73	40.25	48.85	20.70

Figure 9

List of top 15 countries that experienced the highest decline in CVDs in the last three decades (1990-2019)

Top 15 Countries that Experienced decline in Cardiovascular Deaths (CVD) WRT Population (2019-1990)		HDI	Alcohol	Overweight	Tobacco
Portugal	**-0.10%**	0.87	69.2	57.5	25.3
Australia	**-0.10%**	0.94	79.4	64.5	14
Iceland	**-0.10%**	0.96	71.5	59.1	12.6
Switzerland	**-0.11%**	0.96	81.2	54.3	25.7
Belgium	**-0.11%**	0.94	76.5	59.5	23.9
Israel	**-0.12%**	0.92	55	64.3	21.6
Grenada	**-0.13%**	0.80	49.6	51.4	NA
Germany	**-0.13%**	0.95	79.4	56.8	22.5
Austria	**-0.15%**	0.92	78.1	54.3	27.1
Ireland	**-0.19%**	0.94	81.3	60.6	21.4
Sweden	**-0.19%**	0.95	73	56.4	24.8
United Kingdom	**-0.23%**	0.94	73.4	63.7	16.1
Luxembourg	**-0.23%**	0.93	91.8	58.7	21.6
Denmark	**-0.26%**	0.95	74.4	55.4	18.1
Norway	**-0.26%**	0.96	79	58.3	17.1
Average of 15 Countries	-0.16%	0.93	74.19	58.32	20.84
Global Average	0.01%	0.73	40.25	48.85	20.70

Figure 10

Figure 10 shows the list of the top 15 countries that experienced a significant decline in CVDs in the last three decades. The majority of these countries belong to the Western European region, where tobacco usage is on the much lower side and the health development Index is above the world's average.

List of top 10 countries accounting for 2/3rd of CVDs worldwide with change in CVDs in the last three decades (1990-2019)

Change in Cardiovascular Deaths (CVD) WRT Population for 10 Countries (2019-1990) Accounts for 63% of all Cardiovascular Deaths in the world		HDI	Alcohol	Overweight	Tobacco
China	**0.11%**	0.76	55.9	32.3	25.7
India	**0.05%**	0.65	38.8	19.7	28.1
Russia	**0.09%**	0.85	58.1	57.1	27.1
United States	**-0.07%**	0.93	71.7	67.9	23.4
Indonesia	**0.09%**	0.72	22.7	28.2	37.2
Ukraine	**0.34%**	0.79	61.8	58.4	26.2
Brazil	**0.01%**	0.77	40.3	56.5	13.2
Japan	**0.07%**	0.92	57	27.2	20.5
Germany	**-0.13%**	0.95	79.4	56.8	22.5
Pakistan	**0.003%**	0.55	1.2	28.4	20.8
Global Average	0.01%	0.73	40.25	48.85	20.70

Figure 11

Figure 11 shows the list of the top 10 countries which account for 63% of Cardiovascular Deaths worldwide. This chart indicates that 8 out of 10 countries experienced an increase in CVDs in the last three decades while only two countries, USA and Germany, experienced a decline over this period. These countries (USA and Germany) exhibit low tobacco usage compared to the world's average and the health development Index is above the world's average.

Country-specific data on the Human Development Index, Percentage of the Population consuming Alcohol, Percent of the population consuming Tobacco, and Percent of the population overweight are used to further generate segments using the Machine Learning Clustering algorithm using Microsoft SQL Server 2020. The three-segment solution was selected based on the sizes of the segments and homogeneity within segments and heterogeneity across segments.

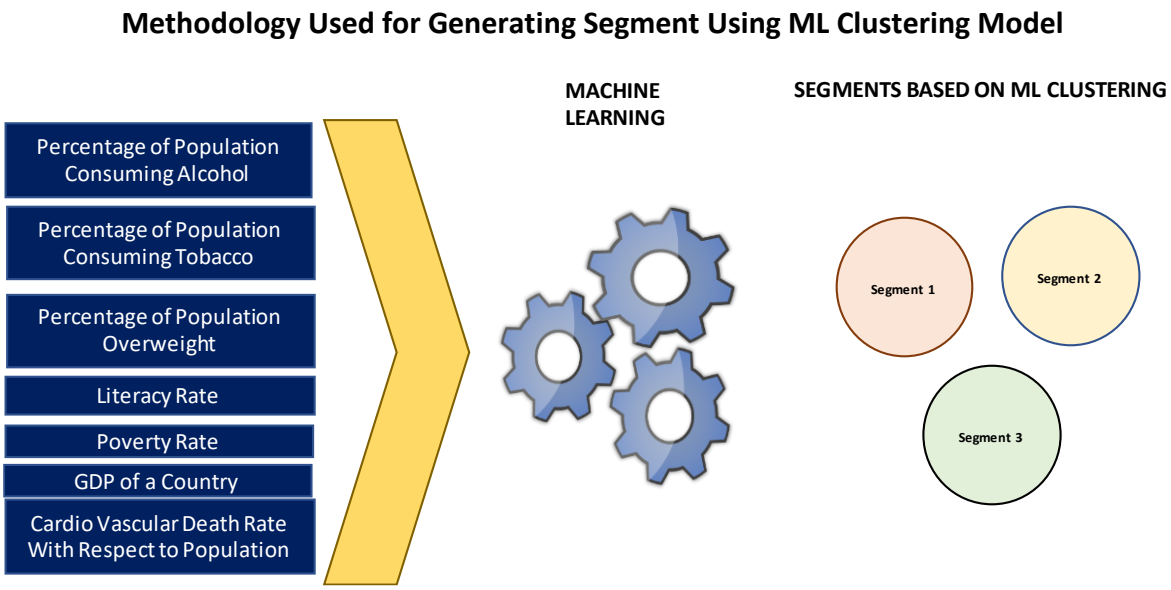


Figure 12

Figure 12 describes the methodology used for clustering country-specific data into meaningful groups.

Top Level Description and Profile of Segments

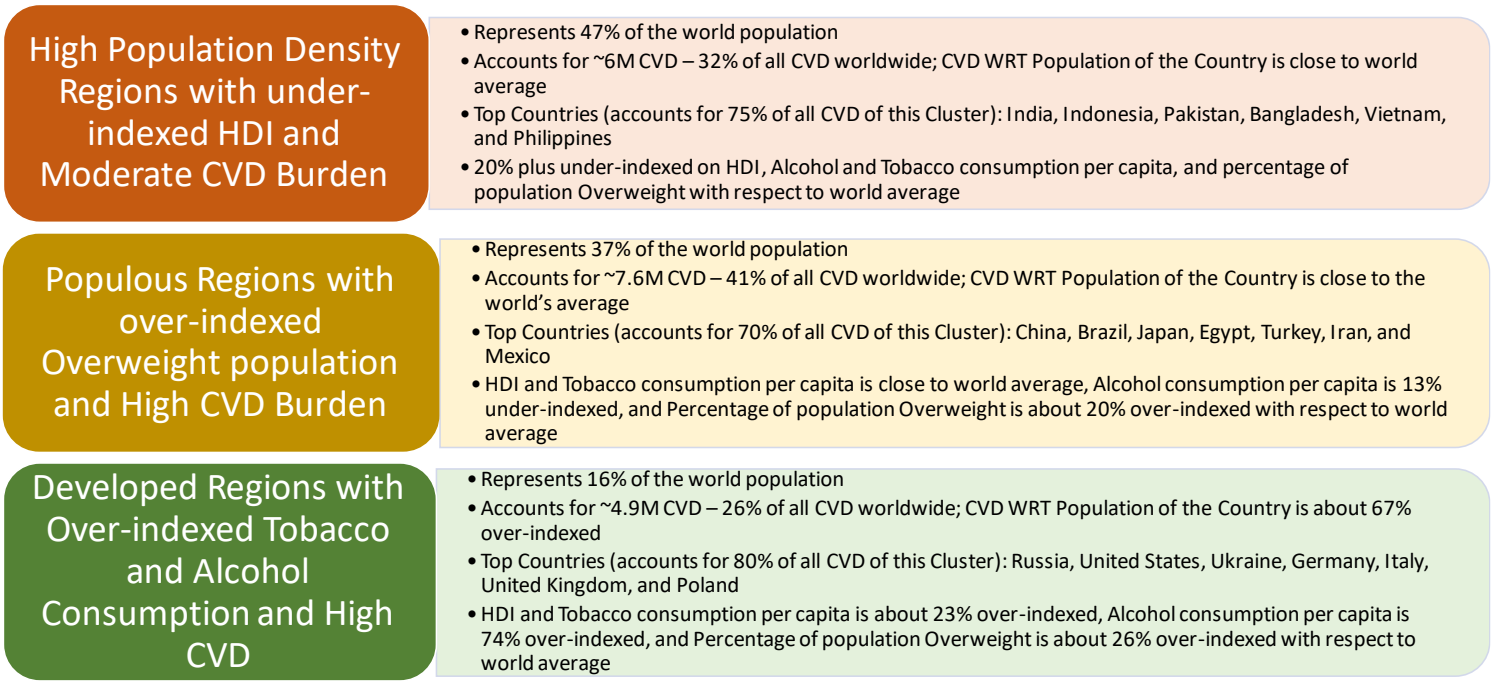


Figure 13

Figure 13 summarizes the key characteristics of each segment.

High Population Density Regions with under-indexed HDI and Moderate CVD Burden (Segment 1) are the biggest segment in size representing about 47% of the world population and accounting for 32% of all Cardiovascular Deaths.

- This segment is significantly under-indexed with respect to the world's average on all key dimensions- the percentage of the population consuming alcohol and tobacco, the percentage of the population overweight, and the Human Development Index.
- The top countries part of this segment are key Southeast Asian countries (India, Bangladesh, Pakistan, Indonesia, Vietnam, and the Philippines).
- This segment belongs to a high population density region of the world and is lower on the human development index.
- With economies growing and habit and work cultures changing, these countries pose a significantly higher risk of increasing CVD over the next few years if proper and relevant measures - such as promoting healthy foods and exercise routine, and investment in the healthcare system – are not undertaken by local Governments.

Populous Regions with an over-indexed Overweight population and High CVD Burden (Segment 2) represent about 37% of the world population and account for 41% of all Cardiovascular Deaths.

- This segment is at par with the world's average on Human Development Index and Tobacco consumption, under-indexed on Alcohol consumption, and about 20% over-indexed on the percentage of the population overweight.
- China, Brazil, Japan, Egypt, Turkey, Iran, and Mexico are the top countries in this segment and represents 80% of all CVDs in this segment.
- This region has a good opportunity to reduce CVDs by investing more in health infrastructure, preventive health care, and cardio health education.

Developed Regions with Over-indexed Tobacco and Alcohol Consumption and High CVD (Segment 3) represent only 16% of the world population and account for 26% of all Cardiovascular Deaths.

- This segment is about 67% over-indexed for Cardiovascular deaths WRT Population of the country compared to the world's average.
- This segment is 23% over-indexed on Human Development Index (HDI). This segment is also significantly over-indexed on all other key dimensions- the percentage of population consuming alcohol, tobacco, and overweight.
- The top countries part of this segment are Western developed nations or East European countries - Russia, United States, Ukraine, Germany, Italy, United Kingdom, and Poland, and represents 80% of all CVDs of this segment.
- The biggest challenge of this segment is the aged population and significantly higher consumption of Tobacco. Countries belonging to this segment can reduce CVD by focusing on educating their population on reducing tobacco consumption and focusing on eating healthy and exercising.

Conclusions

This study focuses on two key topics: (1) building a model using machine learning techniques to predict heart failure risk and its driver; (2) developing deeper global insights on cardiovascular deaths using 3 decades of data and examining the relationship between socioeconomic and demographic attributes with CVD deaths.

Key learnings from this study are as follows:

- Various ML models were built to predict heart risk. All models were able to predict patients with heart risk **with 80%+ classification accuracy**. Finally, the **decision tree model was selected** for this analysis because of its **explanatory power**. This ML model has a **precision of 82%** and **an 81% recall rate for heart-risk patient**s. The drivers of the model have been examined and have found that **ST Slope, Chest Pain Type, and Max Heart Rate** are the top drivers for predicting heart risk among patients.
- Over three decades, it was seen that the increase in CVD globally outpaced world population growth **(54% in CVDs vs 46% in the population).** The age group of 70+ years population accounts for **the majority of CVD** and the percentage of total cardiovascular deaths from this group has gone up in the last three decades. **Russia, Eastern European, and Central Asian** countries are over-indexed on CVD with respect to average CVD worldwide.
- Three decades of CVDs trend indicates that some of the **western countries with developed economies were able to reverse the CVDs trend** while many countries- such as India, China, and eastern European countries, saw an increase in CVDs. The trend in CVDs poses **a significant challenge as major populous countries are showing an increase in CVDs** over three decades.
- Machine Learning based cluster model using aggregated attributes such as the percentage of the population consuming alcohol and tobacco, and the percentage of the population overweight, with socioeconomic factors such as the Human Development Index, GDP, % of the population still below the poverty line, **helped segment the world's population into three key distinct segments**. These segments are homogeneous within and heterogeneous across and offer a great opportunity to develop segment base strategies to tackle Cardiovascular health across the globe.
- It was evident from the analysis that the percentage of the population using Tobacco was found to be positively correlated with a higher incidence of CVDs in that country, while higher HDI (Human Development Index) was found to be associated with countries with lower incidence of CVDs and a decline in CVDs over last 3 decades. This suggests that improving socio-economic factors could have a positive impact on cardiovascular health outcomes.

Next Steps

The insights generated from this study using patient level and aggregated data helped in establishing an underlying relationship between behavioral, and socioeconomic attributes and comorbidities with cardiovascular deaths. This study helps in identifying 3 key distinct segments representing different geographies and a very unique socio-economic profile that can guide WHO and individual countries to build strategies targeted to tackle cardiovascular disease. ML-based predictive modeling helped in identifying key drivers for heart risk. To further improve the accuracy of the ML model to predict heart risk failure with a 95%+ accuracy, efforts should be put in to gather patient-level information from diverse geographical regions on genetic pre-disposals, lifestyle, and dietary data on a longitudinal basis to enhance richness in attributes that feed into the ML model. Efforts should be made to develop Artificial Intelligence driven wearable devices that could help predict heart risk with very high accuracy well in advance to prevent deaths. This will be very helpful for developed nations (North America and Western Europe with aged populations) to prevent deaths due to cardiovascular episodes among the aged population.

To decrease the incidents of cardiovascular disease (CVD) deaths, it is crucial to implement more preventive measures. One effective measure is to incorporate health education about the hazards of sedentary lifestyles, tobacco and alcohol consumption, and unhealthy diets into the K-12 curriculum. Educating students about the long-term effects of these risk factors on their heart health, can encourage healthy behaviors early on and reduce the prevalence of CVD in the future.

These insights provide a solid foundation, but efforts to build richer and more extensive patient-level data representing a broader set of countries should be continued. Deeper studies on policies, programs, and preventive measurement practices adopted by top contributors to cardiovascular deaths worldwide should be carried out. Developed policies and practices focused on top drivers of heart risk and CVD will help reduce premature deaths because of CVDs and in turn mitigate the huge economic loss society bears due to this disease.

Digital Scientific Poster

Predicting Heart Disease Risk Drivers Using Machine Learning Algorithms and Global Insights on Cardiovascular Mortalities

#603893
Louisville, Kentucky

Introduction

Cardiovascular diseases (CVD) have consistently been the leading cause of death across the globe.

- Around 18 million people die from CVD yearly, representing 32% of all global deaths.
- 85% of these CVD deaths are due to **heart attacks and stroke** .

Most CVD can be prevented by addressing behavioral risk factors such as:

- tobacco use
- unhealthy diet and obesity
- physical inactivity
- harmful use of alcohol

CVD has a heavy economic burden on society. Cost of CVD is estimated in multiple categories - screening primary prevention, acute hospital care, loss of productivity. Cost of CVD could be as high as US$20 trillion from 2010 to 2030.

Purpose

The main objective of this study is to **examine global distribution of cardiovascular death** s and its relationship with behavioral and socio -economic factors and **build machine learning models to predict heart disease risk** and understand drivers of heart disease.

The key insights from this analysis can help WHO and national and state -level governments to develop data -driven plans to identify risk factors and proactively manage this disease to reduce deaths and its economic impact on society.

Methods

Machine Learning models have been created using a curated dataset consisting of 1190 patients from the United States, United Kingdom, Switzerland, and Hungary. Global level insights have been created using multiple datasets including:

- Worldwide Socioeconomic attribute and CVD deaths
- Three decades (1990 to 2019) of country level deaths by cause

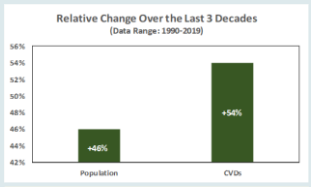
Exploratory Data Analysis is the first step performed on each dataset to understand the data in detail. **Statistical and ML techniques** are used to analyze this data to highlight underlying patterns and build models **(Decision Tree and Clustering)** to develop deeper and actionable insights.

Results (Global Distribution of Cardiovascular Deaths)

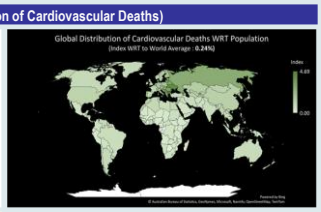
Deaths by top Diseases	1990	2019	Relative Change (2019-1990)/1990
Cardiovascular Diseases	28%	34%	**23.1%**
Neoplasms (Cancer)	13%	19%	**40.2%**
Lower respiratory infections	8%	5%	**-39.9%**
Chronic respiratory diseases	7%	7%	2.8%
Neonatal disorders	7%	3%	**-49.9%**
Diarrheal diseases	7%	3%	**-57.6%**
Digestive diseases	4%	5%	10.4%
Tuberculosis	4%	2%	**-46.9%**
Diabetes mellitus	2%	3%	**87.6%**
Alzheimer	1%	3%	**131.7%**

CVD continues to be the leading cause of deaths worldwide, increasing from 28% in 1990 to 34% in 2019.

Results (Global Distribution of Cardiovascular Deaths)



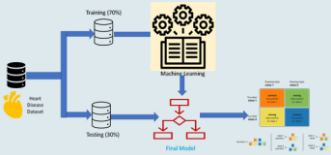
Relative change in CVD deaths from 1990 to 2019, grew by 54%, outpacing population growth in the same period.



This chart indicates higher index (higher CVD deaths WRT population) for Russia, eastern Europe and Central Asia.

Results (Predicting Heart Risk Using Decision Tree Model & Insights)

Methodology Used for Training and Testing ML Model to predict Heart Failure Risk



Classification Accuracy and Precision

Decision Tree ML Model	True Class	
Predicted	0 (Actual)	1 (Actual)
0	130	36
1	34	157

Precision for Target 1	Accuracy	Recall for Target 1
82%	80%	81%

Classification accuracy on testing data set is shown above. Precision and Recall numbers for predicting Heart Risk exhibit relative stability of the model.

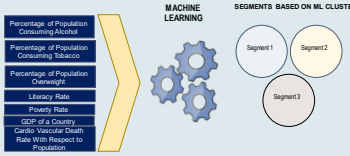
Attribute Importance from Strongest to Least link

ST Slope	Slope refers to the slope of each patient's stress test
Chest Pain Type	This refers to the Angina or Ischemic chest pain
Sex	Sex refers to the assigned sex of the patient at birth
Max Heart Rate	This refers to the maximum heart rate reached during a stress test. This is a hard variable because younger people have higher heart rates while older people have lower heart rates but are more likely to have heart problems
Cholesterol	This refers to cholesterol within the blood
Exercise Angina	Exercise Angina measures whether or not a patient had Angina (severe heart pain) during exercise
Fasting Blood Sugar	Fasting blood sugar measures a person's blood sugar and how it will likely change in the long term
Oldpeak	Oldpeak refers to whether or not the patient showed ST depression induced by exercise during the stress test

ST Slope is most important attribute in predicting heart risk followed by Chest Pain Type, Sex, and Max Heart Rate.

Results (Clustering and Key Segments)

Methodology Used for Generating Segment using ML Clustering Model



Machine Learning Based Cluster Model helped Segment World's Population into 3 Key Distinct Segments

- **High Population Density Regions with Under-Indexed HDI and Moderate CVD Burden (Segment 1)**
- **Populous Regions with Over-Indexed Overweight Population and High CVD Burden (Segment 2)**
- **Developed Regions with Over-Indexed Tobacco and Alcohol Consumption and High CVD (Segment 3)**

These segments are homogeneous within and heterogeneous across and offer great opportunity to develop segment base strategies to tackle Cardiovascular health across the globe.

Conclusion

Key learnings from this study are as follows:

- **ST Slope, Chest Pain Type, and Max Heart Rate** are top drivers for predicting heart risk among patients.
- CVD globally outpaced world population growth **(54% in CVDs vs 46% in population).** Age group of 70+ years population accounts for **majority of CVD** . Russia, Eastern European, and Central Asian countries are over -indexed on CVD.
- Some of the **western countries with developed economies were able to reverse CVD trend. Major populous countries show an increase in CVDs,** posing a significant challenge.
- **Three distinct segments** identified based on aggregated attributes - % of population consuming alcohol and tobacco, % of population overweight, other socioeconomic factors, allowing for **development of segment -based strategies to tackle CVD globally.**
- **% of population using Tobacco** was found to be **positively correlated** with a **higher incidence of CVDs** in a country.
- **Higher HDI** (Human Development Index) found to be **associated** with countries with a **lower incidence and decline in CVDs.**
- Analysis suggests that **improving socio-economic factors could have a positive impact on cardiovascular health outcomes.**

Next Steps

To improve the accuracy of ML model to 95% + accuracy, efforts should be put in to enhance richness in attributes by gathering patient level information from

- diverse geographical regions on genetic predisposal
- lifestyle and dietary data on longitudinal basis

To decrease the incidents of cardiovascular disease (CVD) deaths, it is crucial to implement more preventive measures on:

Social Factors:

- Develop AI driven wearable devices that could help predict heart risk with very high accuracy to prevent deaths. This will be very helpful for developed nations to prevent deaths due to CVD amongst aged population.
- Measures to incorporate health education into the K-12 curriculum on hazards of sedentary lifestyles, tobacco and alcohol consumption, and unhealthy diets should be undertaken. This can promote students to imbibe healthy behaviors early on and reduce the prevalence of CVD in future.
- Carry out deeper study on policies, programs and preventive measurement practices adopted by the countries that were able to reverse CVD death trends.

Economic Factors:

- WHO should encourage countries to adopt effective policies and practices focused on reducing burden of heart risk and CVD. This will help reduce premature deaths and in turn mitigate huge economic loss.

References

"About." *Our World in Data*, ourworldindata.org/about.

Bloom, D.E., Cafiero, E.T., Jané -Llopis, E., Abrahams -Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M. , Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. (2011). The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum.

Bradford, Alina. "What Is Blood Sugar?" *livescience.com*, 26 May 2018, www.livescience.com/62673 -what-is-blood -sugar.html.

"Cardiovascular Disease Costs Will Exceed $1 Trillion by 2035." *RTI*, 15 Feb. 2017, www.rti.org/news/cardiovascular -disease-costs-will -exceed-1-trillion -2035.

Cardiovascular Diseases | Division of Global Health Protection | Global Health | CDC www.cdc.gov/globalhealth/healthprotection/ncd/cardiovascular -diseases.html

Csige, Imre, et al. "The Impact of Obesity on the Cardiovascular System." *Journal of Diabetes Research*, vol. 2018, Hindawi Publishing Corporation, Nov. 2018, pp. 1 –12. https://doi.org/10.1155/2018/3407306.

References

"About." *Our World in Data*, ourworldindata.org/about.

Bloom, D.E., Cafiero, E.T., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L.R., Fathima, S., Feigl, A.B., Gaziano, T., Mowafi, M., Pandya, A., Prettner, K., Rosenberg, L., Seligman, B., Stein, A.Z., & Weinstein, C. (2011). The Global Economic Burden of Noncommunicable Diseases. Geneva: World Economic Forum.

Bradford, Alina. "What Is Blood Sugar?" *livescience.com*, 26 May 2018, www.livescience.com/62673-what-is-blood-sugar.html.

"Cardiovascular Disease Costs Will Exceed $1 Trillion by 2035." *RTI*, 15 Feb. 2017, www.rti.org/news/cardiovascular-disease-costs-will-exceed-1-trillion-2035.

Cardiovascular Diseases | Division of Global Health Protection | Global Health | CDC. www.cdc.gov/globalhealth/healthprotection/ncd/cardiovascular-diseases.html.

Csige, Imre, et al. "The Impact of Obesity on the Cardiovascular System." *Journal of Diabetes Research*, vol. 2018, Hindawi Publishing Corporation, Nov. 2018, pp. 1–12. https://doi.org/10.1155/2018/3407306.

"Heart Attack Symptoms, Risk Factors, and Recovery | cdc.gov." *Centers for Disease Control and Prevention*, 12 July 2022, www.cdc.gov/heartdisease/heart_attack.htm.

NCD Management-Screening, Diagnosis and Treatment. *HEARTS: Technical Package for Cardiovascular Disease Management in Primary Health Care: Risk-based CVD Management*. 13 July 2020, www.who.int/publications/i/item/hearts-technical-package.

Rehm, JÃ¼rgen, et al. "Modelling the Impact of Alcohol Consumption on Cardiovascular Disease Mortality for Comparative Risk Assessments: An Overview." *BMC Public*

Health, vol. 16, no. 1, BioMed Central, Apr. 2016, https://doi.org/10.1186/s12889-016-3026-9.

Ritchie, Hannah. "Causes of Death." *Our World in Data*, 14 Feb. 2018, ourworldindata.org/causes-of-death#cardiovascular-diseases.

Roy, Ambuj, et al. "Tobacco and Cardiovascular Disease: A Summary of Evidence." *The World Bank eBooks*, The World Bank, Nov. 2017, pp. 57–77. https://doi.org/10.1596/978-1-4648-0518-9_ch4.

Siddhartha, Manu. "Heart Disease Dataset (Comprehensive)." *IEEE DataPort*, 6 Nov. 2020, ieee-dataport.org/open-access/heart-disease-dataset-comprehensive.

UNdata | Explorer. data.un.org/Explorer.aspx.

VanMSFT. "Queries - SQL Server." *Microsoft Learn*, 28 Feb. 2023, learn.microsoft.com/en-us/sql/t-sql/queries/queries?view=sql-server-ver16.

Appendix

Data Analysis on Raw Data using Excel:

Country-Level Population:

Year	Country	Population
2019	Afghanistan	37769496
2019	Albania	2873883
2019	Algeria	42705372
2019	American Samoa	47341
2019	Andorra	76361
2019	Angola	32353592
2019	Anguilla	15408
2019	Antigua and Barbuda	92127
2019	Argentina	44745516
2019	Armenia	2820604
2019	Aruba	106458
2019	Australia	25357168
2019	Austria	8879939
2019	Azerbaijan	10232761
2019	Bahamas	404563
2019	Bahrain	1494195
2019	Bangladesh	165516224
2019	Barbados	280193
2019	Belarus	9673971
2019	Belgium	11510569
2019	Belize	389103
2019	Benin	12290443
2019	Bermuda	63833
2019	Bhutan	767467
2019	Bolivia	11777318
2019	Bonaire Sint Eustatius and Saba	25597
2019	Bosnia and Herzegovina	3360718
2019	Botswana	2499705
2019	Brazil	211782880
2019	British Virgin Islands	30631
2019	Brunei	438058
2019	Bulgaria	7052536
2019	Burkina Faso	20951638
2019	Burundi	11874838
2019	Cambodia	16207746
2019	Cameroon	25782342
2019	Canada	37522584
2019	Cape Verde	577037
2019	Cayman Islands	66152
2019	Central African Republic	5209329

Region	Total
Afghanistan	61994
Albania	12904
Algeria	97930
American Samoa	125
Andorra	168
Angola	25723
Anguilla	
Antigua and Barbuda	200
Argentina	101725
Armenia	13507
Australia	51615
Austria	32999
Azerbaijan	42138
Bahamas	867
Bahrain	1272
Bangladesh	324765
Barbados	905
Belarus	75338
Belgium	33627
Belize	464
Benin	11418
Bermuda	209
Bhutan	1208
Bolivia	15903
Bosnia and Herzegovina	18828
Botswana	3519
Brazil	397993
Brunei	508
Bulgaria	79119
Burkina Faso	22151
Burundi	10803
Cambodia	30313
Cameroon	26491
Canada	82644
Cape Verde	1073
Central African Republic	7210
Chad	13838
Chile	30114
China	4584273
Colombia	72629

Region		%WRT Population
Afghanistan	61994	0.16%
Albania	12904	0.45%
Algeria	97930	0.23%
American Samoa	125	0.26%
Andorra	168	0.22%
Angola	25723	0.08%
Anguilla		0.00%
Antigua and Barbuda	200	0.22%
Argentina	101725	0.23%
Armenia	13507	0.48%
Aruba		0.00%
Australia	51615	0.20%
Austria	32999	0.37%
Azerbaijan	42138	0.41%
Bahamas	867	0.21%
Bahrain	1272	0.09%
Bangladesh	324765	0.20%
Barbados	905	0.32%
Belarus	75338	0.78%
Belgium	33627	0.29%
Belize	464	0.12%
Benin	11418	0.09%
Bermuda	209	0.33%
Bhutan	1208	0.16%
Bolivia	15903	0.14%
Bonaire Sint Eustatius and Saba		0.00%
Bosnia and Herzegovina	18828	0.56%
Botswana	3519	0.14%
Brazil	397993	0.19%
British Virgin Islands		0.00%
Brunei	508	0.12%
Bulgaria	79119	1.12%
Burkina Faso	22151	0.11%
Burundi	10803	0.09%
Cambodia	30313	0.19%
Cameroon	26491	0.10%
Canada	82644	0.22%
Cape Verde	1073	0.19%
Cayman Islands		0.00%
Central African Republic	7210	0.14%

2019	Chad	16126868		Comoros	1143		Chad	13838	0.09%
2019	Chile	19039484		Congo	7096		Chile	30114	0.16%
2019	China	1421864064		Cook Islands	58		China	4584273	0.32%
2019	Colombia	50187404		Costa Rica	6665		Colombia	72629	0.14%
2019	Comoros	790993		Cote d'Ivoire	22377		Comoros	1143	0.14%
2019	Congo	5570736		Croatia	24720		Congo	7096	0.13%
2019	Cook Islands	17133		Cuba	39755		Cook Islands	58	0.34%
2019	Costa Rica	5084527		Cyprus	3112		Costa Rica	6665	0.13%
2019	Cote d'Ivoire	26147544		Czechia	51294		Cote d'Ivoire	22377	0.09%
2019	Croatia	4129749		Democratic Republic of Congo	89907		Croatia	24720	0.60%
2019	Cuba	11316699		Denmark	15331		Cuba	39755	0.35%
2019	Curacao	193119		Djibouti	1240		Curacao		0.00%
2019	Cyprus	1228840		Dominica	242		Cyprus	3112	0.25%
2019	Czechia	10536876		Dominican Republic	27139		Czechia	51294	0.49%
2019	Democratic Republic of Congo	89906896		Ecuador	21464		Democratic Republic of Congo	89907	0.10%
2019	Denmark	5795879		Egypt	263873		Denmark	15331	0.26%
2019	Djibouti	1073999		El Salvador	9694		Djibouti	1240	0.12%
2019	Dominica	71446		England	153285		Dominica	242	0.34%
2019	Dominican Republic	10881886		Equatorial Guinea	960		Dominican Republic	27139	0.25%
2019	Ecuador	17343742		Eritrea	6660		Ecuador	21464	0.12%
2019	Egypt	105618672		Estonia	8201		Egypt	263873	0.25%
2019	El Salvador	6280222		Eswatini	1518		El Salvador	9694	0.15%
2019	Equatorial Guinea	1553037		Ethiopia	76975		Equatorial Guinea	960	0.06%
2019	Eritrea	3498819		Fiji	2641		Eritrea	6660	0.19%
2019	Estonia	1327039		Finland	25384		Estonia	8201	0.62%
2019	Eswatini	1169620		France	166495		Eswatini	1518	0.13%
2019	Ethiopia	114120592		Gabon	2541		Ethiopia	76975	0.07%
2019	Faeroe Islands	51708		Gambia	2602		Faeroe Islands		0.00%
2019	Falkland Islands	3729		Georgia	29709		Falkland Islands		0.00%
2019	Fiji	918472		Germany	364285		Fiji	2641	0.29%
2019	Finland	5521539		Ghana	39650		Finland	25384	0.46%
2019	France	64399760		Greece	55921		France	166495	0.26%
2019	French Guiana	284088		Greenland	123		French Guiana		0.00%
2019	French Polynesia	299730		Grenada	263		French Polynesia		0.00%
2019	Gabon	2242788		Guam	474		Gabon	2541	0.11%
2019	Gambia	2508882		Guatemala	15882		Gambia	2602	0.10%
2019	Georgia	3770814		Guinea	14547		Georgia	29709	0.79%
2019	Germany	83148144		Guinea-Bissau	2214		Germany	364285	0.44%
2019	Ghana	31522294		Guyana	2430		Ghana	39650	0.13%
2019	Gibraltar	32715		Haiti	26469		Gibraltar		0.00%
2019	Greece	10574026		Honduras	15016		Greece	55921	0.53%
2019	French Guiana	284088		Greenland	123		French Guiana		0.00%
2019	French Polynesia	299730		Grenada	263		French Polynesia		0.00%
2019	Gabon	2242788		Guam	474		Gabon	2541	0.11%
2019	Gambia	2508882		Guatemala	15882		Gambia	2602	0.10%
2019	Georgia	3770814		Guinea	14547		Georgia	29709	0.79%
2019	Germany	83148144		Guinea-Bissau	2214		Germany	364285	0.44%
2019	Ghana	31522294		Guyana	2430		Ghana	39650	0.13%
2019	Gibraltar	32715		Haiti	26469		Gibraltar		0.00%
2019	Greece	10574026		Honduras	15016		Greece	55921	0.53%
2019	Greenland	55955		Hungary	61818		Greenland	123	0.22%
2019	Grenada	122737		Iceland	723		Grenada	263	0.21%
2019	Guadeloupe	395487		India	2574408		Guadeloupe	263	0.07%
2019	Guam	168632		Indonesia	651482		Guam	474	0.28%
2019	Guatemala	17106338		Iran	173601		Guatemala	15882	0.09%
2019	Guernsey	62391		Iraq	79953		Guernsey		0.00%
2019	Guinea	12877537		Ireland	10348		Guinea	14547	0.11%
2019	Guinea-Bissau	1970460		Israel	12394		Guinea-Bissau	2214	0.11%
2019	Guyana	798761		Italy	236507		Guyana	2430	0.30%
2019	Haiti	11160441		Jamaica	6456		Haiti	26469	0.24%
2019	Honduras	9958832		Japan	372482		Honduras	15016	0.15%
2019	Hong Kong	7496121		Jordan	11798		Hong Kong		0.00%
2019	Hungary	9771799		Kazakhstan	65391		Hungary	61818	0.63%
2019	Iceland	360787		Kenya	40612		Iceland	723	0.20%
2019	India	1383112064		Kiribati	348		India	2574408	0.19%
2019	Indonesia	269582880		Kuwait	4118		Indonesia	651482	0.24%
2019	Iran	86564208		Kyrgyzstan	17482		Iran	173601	0.20%
2019	Iraq	41563524		Laos	13523		Iraq	79953	0.19%
2019	Ireland	4896021		Latvia	15778		Ireland	10348	0.21%
2019	Isle of Man	83950		Lebanon	16329		Isle of Man		0.00%
2019	Israel	8607922		Lesotho	4027		Israel	12394	0.14%
2019	Italy	59727936		Liberia	4300		Italy	236507	0.40%
2019	Jamaica	2813776		Libya	12748		Jamaica	6456	0.23%
2019	Japan	125791680		Lithuania	21301		Japan	372482	0.30%
2019	Jersey	107055		Luxembourg	1332		Jersey		0.00%
2019	Jordan	10698682		Madagascar	35396		Jordan	11798	0.11%
2019	Kazakhstan	18754260		Malawi	15651		Kazakhstan	65391	0.35%
2019	Kenya	50951452		Malaysia	58825		Kenya	40612	0.08%

2019	Kiribati	124252		Maldives	522		Kiribati	348	0.28%
2019	Kosovo	1686856		Mali	19520		Kosovo		0.00%
2019	Kuwait	4441105		Malta	1490		Kuwait	4118	0.09%
2019	Kyrgyzstan	6323643		Marshall Islands	156		Kyrgyzstan	17482	0.28%
2019	Laos	7212058		Mauritania	3951		Laos	13523	0.19%
2019	Latvia	1916552		Mauritius	3309		Latvia	15778	0.82%
2019	Lebanon	5781912		Mexico	167536		Lebanon	16329	0.28%
2019	Lesotho	2225710		Moldova	23438		Lesotho	4027	0.18%
2019	Liberia	4985290		Monaco	161		Liberia	4300	0.09%
2019	Libya	6569085		Mongolia	9806		Libya	12748	0.19%
2019	Liechtenstein	38500		Montenegro	3942		Liechtenstein		0.00%
2019	Lithuania	2849083		Morocco	117034		Lithuania	21301	0.75%
2019	Luxembourg	619981		Mozambique	31693		Luxembourg	1332	0.21%
2019	Macao	663660		Myanmar	138139		Macao		0.00%
2019	Madagascar	27533134		Namibia	3748		Madagascar	35396	0.13%
2019	Malawi	18867340		Nauru	23		Malawi	15651	0.08%
2019	Malaysia	32804024		Nepal	46500		Malaysia	58825	0.18%
2019	Maldives	504518		Netherlands	42569		Maldives	522	0.10%
2019	Mali	20567430		New Zealand	11949		Mali	19520	0.09%
2019	Malta	503646		Nicaragua	8202		Malta	1490	0.30%
2019	Marshall Islands	44750		Niger	17201		Marshall Islands	156	0.35%
2019	Martinique	372254		Nigeria	163496		Martinique		0.00%
2019	Mauritania	4383853		Niue	9		Mauritania	3951	0.09%
2019	Mauritius	1296284		North Korea	102586		Mauritius	3309	0.26%
2019	Mayotte	294502		North Macedonia	14023		Mayotte		0.00%
2019	Mexico	125085312		Northern Ireland	5024		Mexico	167536	0.13%
2019	Moldova	3109496		Northern Mariana Islands	115		Moldova	23438	0.75%
2019	Monaco	37059		Norway	12724		Monaco	161	0.43%
2019	Mongolia	3232435		Oman	4998		Mongolia	9806	0.30%
2019	Montenegro	630403		Pakistan	341109		Montenegro	3942	0.63%
2019	Montserrat	4528		Palau	65		Montserrat		0.00%
2019	Morocco	36304408		Palestine	6709		Morocco	117034	0.32%
2019	Mozambique	30285602		Panama	5319		Mozambique	31693	0.10%
2019	Myanmar	53040212		Papua New Guinea	15540		Myanmar	138139	0.26%
2019	Namibia	2446648		Paraguay	9912		Namibia	3748	0.15%
2019	Nauru	12156		Peru	29215		Nauru	23	0.19%
2019	Nepal	28832500		Philippines	204312		Nepal	46500	0.16%

2019	Netherlands	17363260		Poland	174736		Netherlands	42569	0.25%
2019	New Caledonia	285374		Portugal	37354		New Caledonia		0.00%
2019	New Zealand	4959033		Puerto Rico	8823		New Zealand	11949	0.24%
2019	Nicaragua	6663919		Qatar	1121		Nicaragua	8202	0.12%
2019	Niger	23443400		Romania	150427		Niger	17201	0.07%
2019	Nigeria	203304496		Russia	1004932		Nigeria	163496	0.08%
2019	Niue	1943		Rwanda	11826		Niue	9	0.46%
2019	North Korea	25755442		Saint Kitts and Nevis	156		North Korea	102586	0.40%
2019	North Macedonia	2114175		Saint Lucia	445		North Macedonia	14023	0.66%
2019	Northern Mariana Islands	49880		Saint Vincent and the Grenadines	355		Northern Mariana Islands	115	0.23%
2019	Norway	5348285		Samoa	537		Norway	12724	0.24%
2019	Oman	4602769		San Marino	98		Oman	4998	0.11%
2019	Pakistan	223293280		Sao Tome and Principe	274		Pakistan	341109	0.15%
2019	Palau	17939		Saudi Arabia	45836		Palau	65	0.36%
2019	Palestine	4909775		Scotland	19285		Palestine	6709	0.14%
2019	Panama	4232538		Senegal	15982		Panama	5319	0.13%
2019	Papua New Guinea	9542492		Serbia	63768		Papua New Guinea	15540	0.16%
2019	Paraguay	6530029		Seychelles	241		Paraguay	9912	0.15%
2019	Peru	32824864		Sierra Leone	9009		Peru	29215	0.09%
2019	Philippines	110380800		Singapore	6809		Philippines	204312	0.19%
2019	Poland	38493600		Slovakia	26797		Poland	174736	0.45%
2019	Portugal	10289921		Slovenia	7797		Portugal	37354	0.36%
2019	Puerto Rico	3292888		Solomon Islands	2285		Puerto Rico	8823	0.27%
2019	Qatar	2807242		Somalia	18521		Qatar	1121	0.04%
2019	Romania	19524212		South Africa	82660		Romania	150427	0.77%
2019	Russia	145742288		South Korea	77480		Russia	1004932	0.69%
2019	Rwanda	12835029		South Sudan	7114		Rwanda	11826	0.09%
2019	Saint Barthlemy	10504		Spain	131493		Saint Barthlemy		0.00%
2019	Saint Helena	5470		Sri Lanka	46576		Saint Helena		0.00%
2019	Saint Kitts and Nevis	47739		Sudan	74371		Saint Kitts and Nevis	156	0.33%
2019	Saint Lucia	178594		Suriname	1422		Saint Lucia	445	0.25%
2019	Saint Martin (French part)	33150		Sweden	36086		Saint Martin (French part)		0.00%
2019	Saint Pierre and Miquelon	5952		Switzerland	23968		Saint Pierre and Miquelon		0.00%
2019	Saint Vincent and the Grenadines	104940		Syria	45681		Saint Vincent and the Grenadine	355	0.34%
2019	Samoa	211915		Taiwan	41061		Samoa	537	0.25%
2019	San Marino	34203		Tajikistan	21856		San Marino	98	0.29%
2019	Sao Tome and Principe	214612		Tanzania	54489		Sao Tome and Principe	274	0.13%

2019	Saudi Arabia	35827364		Thailand	115830		Saudi Arabia	45836	0.13%
2019	Senegal	16000783		Timor	2460		Senegal	15982	0.10%
2019	Serbia	7401056		Togo	8361		Serbia	63768	0.86%
2019	Seychelles	104386		Tokelau	4		Seychelles	241	0.23%
2019	Sierra Leone	8046830		Tonga	177		Sierra Leone	9009	0.11%
2019	Singapore	5866408		Trinidad and Tobago	4016		Singapore	6809	0.12%
2019	Sint Maarten (Dutch part)	43002		Tunisia	35008		Sint Maarten (Dutch part)		0.00%
2019	Slovakia	5453932		Turkey	176662		Slovakia	26797	0.49%
2019	Slovenia	2112905		Turkmenistan	18323		Slovenia	7797	0.37%
2019	Solomon Islands	675000		Tuvalu	42		Solomon Islands	2285	0.34%
2019	Somalia	15981312		Uganda	28150		Somalia	18521	0.12%
2019	South Africa	58087060		Ukraine	449376		South Africa	82660	0.14%
2019	South Korea	51803832		United Arab Emirates	8622		South Korea	77480	0.15%
2019	South Sudan	10447677		United Kingdom	188112		South Sudan	7114	0.07%
2019	Spain	47131372		United States	957455		Spain	131493	0.28%
2019	Sri Lanka	21649664		United States Virgin Islands	501		Sri Lanka	46576	0.22%
2019	Sudan	43232092		Uruguay	10004		Sudan	74371	0.17%
2019	Suriname	600305		Uzbekistan	107395		Suriname	1422	0.24%
2019	Sweden	10267922		Vanuatu	868		Sweden	36086	0.35%
2019	Switzerland	8575593		Venezuela	61510		Switzerland	23968	0.28%
2019	Syria	20098252		Vietnam	240161		Syria	45681	0.23%
2019	Taiwan	23777742		Wales	10520		Taiwan	41061	0.17%
2019	Tajikistan	9337008		Yemen	55753		Tajikistan	21856	0.23%
2019	Tanzania	59872584		Zambia	18157		Tanzania	54489	0.09%
2019	Thailand	71307768		Zimbabwe	17810		Thailand	115830	0.16%
2019	Timor	1280442					Timor	2460	0.19%
2019	Togo	8243093			**18739989**		Togo	8361	0.10%
2019	Tokelau	1775					Tokelau	4	0.23%
2019	Tonga	104965					Tonga	177	0.17%
2019	Trinidad and Tobago	1519960					Trinidad and Tobago	4016	0.26%
2019	Tunisia	12049319					Tunisia	35008	0.29%
2019	Turkey	83481688					Turkey	176662	0.21%
2019	Turkmenistan	6158424					Turkmenistan	18323	0.30%
2019	Turks and Caicos Islands	43104					Turks and Caicos Islands		0.00%
2019	Tuvalu	10983					Tuvalu	42	0.38%
2019	Uganda	42949076					Uganda	28150	0.07%
2019	Ukraine	44211100					Ukraine	449376	1.02%

2019	United Arab Emirates	9211660					United Arab Emirates	8622	0.09%
2019	United Kingdom	66778660					United Kingdom	188112	0.28%
2019	United States	334319680					United States	957455	0.29%
2019	United States Virgin Islands	100837					United States Virgin Islands	501	0.50%
2019	Uruguay	3428412					Uruguay	10004	0.29%
2019	Uzbekistan	32976946					Uzbekistan	107395	0.33%
2019	Vanuatu	304414					Vanuatu	868	0.29%
2019	Vatican	815					Vatican		0.00%
2019	Venezuela	28971686					Venezuela	61510	0.21%
2019	Vietnam	95776712					Vietnam	240161	0.25%
2019	Wallis and Futuna	11735					Wallis and Futuna		0.00%
2019	Yemen	31546694					Yemen	55753	0.18%
2019	Zambia	18380478					Zambia	18157	0.10%
2019	Zimbabwe	15354606					Zimbabwe	17810	0.12%

Cardio Deaths by County

Entity -2019	Cardio Deaths			WRT Avg Cardio Deaths	Country	Population
Afghanistan	61995	0.16%	0.49%	0.69	Afghanistan	37769496
Albania	12904	0.45%	0.45%	1.88	Albania	2873883
Algeria	97931	0.23%	0.23%	0.96	Algeria	42705372
American Samoa	125	0.26%	0.26%	1.10	American Samoa	47341
Andorra	169	0.22%	0.22%	0.92	Andorra	76361
Angola	25724	0.08%	0.08%	0.33	Angola	32353592
Antigua and Barbuda	200	0.22%	0.22%	0.91	Antigua and Barbuda	92127
Argentina	101725	0.23%	0.23%	0.95	Argentina	44745516
Armenia	13507	0.48%	0.48%	2.00	Armenia	2820604
Australia	51615	0.20%	0.20%	0.85	Australia	25357168
Austria	32998	0.37%	0.37%	1.55	Austria	8879939
Azerbaijan	42138	0.41%	0.41%	1.72	Azerbaijan	10232761
Bahamas	869	0.21%	0.21%	0.90	Bahamas	404563
Bahrain	1271	0.09%	0.09%	0.36	Bahrain	1494195
Bangladesh	324764	0.20%	0.20%	0.82	Bangladesh	165516224
Barbados	905	0.32%	0.32%	1.35	Barbados	280193
Belarus	75339	0.78%	0.78%	3.25	Belarus	9673971
Belgium	33627	0.29%	0.29%	1.22	Belgium	11510569
Belize	464	0.12%	0.12%	0.50	Belize	389103
Benin	11418	0.09%	0.09%	0.39	Benin	12290443
Bermuda	209	0.33%	0.33%	1.37	Bermuda	63833
Bhutan	1208	0.16%	0.16%	0.66	Bhutan	767467
Bolivia	15905	0.14%	0.14%	0.56	Bolivia	11777318
Bosnia and Herzegovina	18828	0.56%	0.56%	2.34	Bosnia and Herzegovina	3360718
Botswana	NULL	#VALUE!	#VALUE!	#VALUE!	Botswana	2499705
Brazil	397993	0.19%	0.19%	0.79	Brazil	211782880
Brunei	508	0.12%	0.12%	0.48	Brunei	438058
Bulgaria	79119	1.12%	1.12%	4.69	Bulgaria	7052536
Burkina Faso	22151	0.11%	0.11%	0.44	Burkina Faso	20951638
Burundi	10804	0.09%	0.09%	0.38	Burundi	11874838
Cambodia	30313	0.19%	0.19%	0.78	Cambodia	16207746
Cameroon	26491	0.10%	0.10%	0.43	Cameroon	25782342
Canada	82644	0.22%	0.22%	0.92	Canada	37522584
Cape Verde	1073	0.19%	0.19%	0.78	Cape Verde	577037
Central African Republic	7210	0.14%	0.14%	0.58	Central African Republic	5209329
Chad	13837	0.09%	0.09%	0.36	Chad	16126868
Chile	30115	0.16%	0.16%	0.66	Chile	19039484
China	4584273	0.32%	0.32%	1.35	China	1421864064
Colombia	72629	0.14%	0.14%	0.60	Colombia	50187404
Comoros	1143	0.14%	0.14%	0.60	Comoros	790993
Congo	7096	0.13%	0.13%	0.53	Congo	5570736
Cook Islands	58	0.34%	0.34%	1.41	Cook Islands	17133
Costa Rica	6664	0.13%	0.13%	0.55	Costa Rica	5084527
Cote d'Ivoire	22376	0.09%	0.09%	0.36	Cote d'Ivoire	26147544
Croatia	24721	0.60%	0.60%	2.50	Croatia	4129749
Cuba	39755	0.35%	0.35%	1.47	Cuba	11316699
Cyprus	3112	0.25%	0.25%	1.06	Cyprus	1228840
Czechia	51294	0.49%	0.49%	2.03	Czechia	10536876
Czechoslovakia	NULL	#VALUE!	#VALUE!	#VALUE!	Czechoslovakia	NULL
Democratic Republic of Congo	89906	0.10%	0.10%	0.42	Democratic Republic of Congo	89906896
Denmark	15330	0.26%	0.26%	1.11	Denmark	5795879
Djibouti	1240	0.12%	0.12%	0.48	Djibouti	1073999
Dominica	242	0.34%	0.34%	1.42	Dominica	71446
Dominican Republic	27140	0.25%	0.25%	1.04	Dominican Republic	10881886
Ecuador	21464	0.12%	0.12%	0.52	Ecuador	17343742
Egypt	263873	0.25%	0.25%	1.04	Egypt	105618672
El Salvador	9695	0.15%	0.15%	0.65	El Salvador	6280222
Equatorial Guinea	960	0.06%	0.06%	0.26	Equatorial Guinea	1553037
Eritrea	6660	0.19%	0.19%	0.80	Eritrea	3498819
Estonia	8201	0.62%	0.62%	2.58	Estonia	1327039
Eswatini	1519	0.13%	0.13%	0.54	Eswatini	1169620
Ethiopia	76975	0.07%	0.07%	0.28	Ethiopia	114120592
Fiji	2642	0.29%	0.29%	1.20	Fiji	918472
Finland	25383	0.46%	0.46%	1.92	Finland	5521539
France	166496	0.26%	0.26%	1.08	France	64399760
Gabon	2541	0.11%	0.11%	0.47	Gabon	2242788
Gambia	2603	0.10%	0.10%	0.43	Gambia	2508882
Georgia	29709	0.79%	0.79%	3.29	Georgia	3770814
Germany	364285	0.44%	0.44%	1.83	Germany	83148144
Ghana	39650	0.13%	0.13%	0.53	Ghana	31522294
Greece	55921	0.53%	0.53%	2.21	Greece	10574026
Greenland	123	0.22%	0.22%	0.92	Greenland	55955
Grenada	263	0.21%	0.21%	0.90	Grenada	122737
Guam	474	0.28%	0.28%	1.17	Guam	168632
Guatemala	15882	0.09%	0.09%	0.39	Guatemala	17106338
Guinea	14547	0.11%	0.11%	0.47	Guinea	12877537
Guinea-Bissau	2214	0.11%	0.11%	0.47	Guinea-Bissau	1970460
Guyana	2430	0.30%	0.30%	1.27	Guyana	798761
Haiti	26468	0.24%	0.24%	0.99	Haiti	11160441
Honduras	15015	0.15%	0.15%	0.63	Honduras	9958832
Hungary	61818	0.63%	0.63%	2.64	Hungary	9771799
Iceland	724	0.20%	0.20%	0.84	Iceland	360787
India	2574410	0.19%	0.19%	0.78	India	1383112064
Indonesia	651481	0.24%	0.24%	1.01	Indonesia	269582880
Iran	173601	0.20%	0.20%	0.84	Iran	86564208
Iraq	79952	0.19%	0.19%	0.80	Iraq	41563524

Ireland	10347	0.21%	0.21%	0.88	Ireland	4896021
Israel	12393	0.14%	0.14%	0.60	Israel	8607922
Italy	236507	0.40%	0.40%	1.65	Italy	59727936
Jamaica	6457	0.23%	0.23%	0.96	Jamaica	2813776
Japan	372483	0.30%	0.30%	1.24	Japan	125791680
Jordan	11798	0.11%	0.11%	0.46	Jordan	10698682
Kazakhstan	65391	0.35%	0.35%	1.46	Kazakhstan	18754260
Kenya	40611	0.08%	0.08%	0.33	Kenya	50951452
Kiribati	349	0.28%	0.28%	1.17	Kiribati	124252
Kosovo	NULL	#VALUE!	#VALUE!	#VALUE!	Kosovo	1686856
Kuwait	4119	0.09%	0.09%	0.39	Kuwait	4441105
Kyrgyzstan	17482	0.28%	0.28%	1.16	Kyrgyzstan	6323643
Laos	13523	0.19%	0.19%	0.78	Laos	7212058
Latvia	15779	0.82%	0.82%	3.44	Latvia	1916552
Lebanon	16329	0.28%	0.28%	1.18	Lebanon	5781912
Lesotho	4028	0.18%	0.18%	0.76	Lesotho	2225710
Liberia	4299	0.09%	0.09%	0.36	Liberia	4985290
Libya	12747	0.19%	0.19%	0.81	Libya	6569085
Lithuania	21301	0.75%	0.75%	3.12	Lithuania	2849083
Luxembourg	1332	0.21%	0.21%	0.90	Luxembourg	619981
Madagascar	35397	0.13%	0.13%	0.54	Madagascar	27533134
Malawi	15650	0.08%	0.08%	0.35	Malawi	18867340
Malaysia	58824	0.18%	0.18%	0.75	Malaysia	32804024
Maldives	522	0.10%	0.10%	0.43	Maldives	504518
Mali	19519	0.09%	0.09%	0.40	Mali	20567430
Malta	1491	0.30%	0.30%	1.24	Malta	503646
Marshall Islands	156	0.35%	0.35%	1.46	Marshall Islands	44750
Mauritania	3952	0.09%	0.09%	0.38	Mauritania	4383853
Mauritius	3309	0.26%	0.26%	1.07	Mauritius	1296284
Mexico	167535	0.13%	0.13%	0.56	Mexico	125085312
Micronesia (country)	331	#DIV/0!	#DIV/0!	#DIV/0!	Micronesia (country)	
Moldova	23438	0.75%	0.75%	3.15	Moldova	3109496
Monaco	161	0.43%	0.43%	1.82	Monaco	37059
Mongolia	9806	0.30%	0.30%	1.27	Mongolia	3232435
Montenegro	3943	0.63%	0.63%	2.61	Montenegro	630403
Morocco	117034	0.32%	0.32%	1.35	Morocco	36304408
Mozambique	31692	0.10%	0.10%	0.44	Mozambique	30285602
Myanmar	138139	0.26%	0.26%	1.09	Myanmar	53040212
Namibia	3747	0.15%	0.15%	0.64	Namibia	2446648
Nauru	23	0.19%	0.19%	0.79	Nauru	12156
Nepal	46501	0.16%	0.16%	0.67	Nepal	28832500
Netherlands	42569	0.25%	0.25%	1.02	Netherlands	17363260
New Zealand	11948	0.24%	0.24%	1.01	New Zealand	4959033
Nicaragua	8202	0.12%	0.12%	0.51	Nicaragua	6663919
Niger	17201	0.07%	0.07%	0.31	Niger	23443400
Nigeria	163496	0.08%	0.08%	0.34	Nigeria	203304496
Niue	8	0.41%	0.41%	1.72	Niue	1943
North Korea	102586	0.40%	0.40%	1.66	North Korea	25755442
North Macedonia	14023	0.66%	0.66%	2.77	North Macedonia	2114175
Northern Mariana Islands	116	0.23%	0.23%	0.97	Northern Mariana Islands	49880
Norway	12724	0.24%	0.24%	0.99	Norway	5348285
Oman	4998	0.11%	0.11%	0.45	Oman	4602769
Pakistan	341108	0.15%	0.15%	0.64	Pakistan	223293280
Palau	65	0.36%	0.36%	1.51	Palau	17939
Palestine	6710	0.14%	0.14%	0.57	Palestine	4909775
Panama	5318	0.13%	0.13%	0.53	Panama	4232538
Papua New Guinea	15540	0.16%	0.16%	0.68	Papua New Guinea	9542492
Paraguay	9912	0.15%	0.15%	0.63	Paraguay	6530029
Peru	29215	0.09%	0.09%	0.37	Peru	32824864
Philippines	204311	0.19%	0.19%	0.77	Philippines	110380800
Poland	174736	0.45%	0.45%	1.90	Poland	38493600
Portugal	37353	0.36%	0.36%	1.52	Portugal	10289921
Puerto Rico	8823	0.27%	0.27%	1.12	Puerto Rico	3292888
Qatar	1121	0.04%	0.04%	0.17	Qatar	2807242
Romania	150427	0.77%	0.77%	3.22	Romania	19524212
Russia	1004931	0.69%	0.69%	2.88	Russia	145742288
Rwanda	11826	0.09%	0.09%	0.39	Rwanda	12835029
Saint Kitts and Nevis	157	0.33%	0.33%	1.37	Saint Kitts and Nevis	47739
Saint Lucia	446	0.25%	0.25%	1.04	Saint Lucia	178594
Saint Vincent and the Grenadines	355	0.34%	0.34%	1.41	Saint Vincent and the Grenadines	104940
Samoa	539	0.25%	0.25%	1.06	Samoa	211915
San Marino	98	0.29%	0.29%	1.20	San Marino	34203
Sao Tome and Principe	273	0.13%	0.13%	0.53	Sao Tome and Principe	214612
Saudi Arabia	45837	0.13%	0.13%	0.53	Saudi Arabia	35827364
Senegal	15981	0.10%	0.10%	0.42	Senegal	16000783
Serbia	63769	0.86%	0.86%	3.60	Serbia	7401056
Seychelles	242	0.23%	0.23%	0.97	Seychelles	104386
Sierra Leone	9009	0.11%	0.11%	0.47	Sierra Leone	8046830
Singapore	6810	0.12%	0.12%	0.49	Singapore	5866408
Slovakia	26796	0.49%	0.49%	2.05	Slovakia	5453932
Slovenia	7797	0.37%	0.37%	1.54	Slovenia	2112905
Solomon Islands	2285	0.34%	0.34%	1.41	Solomon Islands	675000
Somalia	18521	0.12%	0.12%	0.48	Somalia	15981312
South Africa	82661	0.14%	0.14%	0.59	South Africa	58087060
South Korea	77481	0.15%	0.15%	0.63	South Korea	51803832
South Sudan	7114	0.07%	0.07%	0.28	South Sudan	10447677
Spain	131493	0.28%	0.28%	1.17	Spain	47131372
Sri Lanka	46577	0.22%	0.22%	0.90	Sri Lanka	21649664

Sudan	74371	0.17%	0.17%	0.72	Sudan	43232092
Suriname	1423	0.24%	0.24%	0.99	Suriname	600305
Sweden	36087	0.35%	0.35%	1.47	Sweden	10267922
Switzerland	23969	0.28%	0.28%	1.17	Switzerland	8575593
Syria	45682	0.23%	0.23%	0.95	Syria	20098252
Taiwan	41061	0.17%	0.17%	0.72	Taiwan	23777742
Tajikistan	21856	0.23%	0.23%	0.98	Tajikistan	9337008
Tanzania	54489	0.09%	0.09%	0.38	Tanzania	59872584
Thailand	115830	0.16%	0.16%	0.68	Thailand	71307768
Timor	2459	0.19%	0.19%	0.80	Timor	1280442
Togo	8361	0.10%	0.10%	0.42	Togo	8243093
Tokelau	4	0.23%	0.23%	0.94	Tokelau	1775
Tonga	178	0.17%	0.17%	0.71	Tonga	104965
Trinidad and Tobago	4016	0.26%	0.26%	1.10	Trinidad and Tobago	1519960
Tunisia	35007	0.29%	0.29%	1.21	Tunisia	12049319
Turkey	176663	0.21%	0.21%	0.88	Turkey	83481688
Turkmenistan	18323	0.30%	0.30%	1.24	Turkmenistan	6158424
Tuvalu	42	0.38%	0.38%	1.60	Tuvalu	10983
Uganda	28149	0.07%	0.07%	0.27	Uganda	42949076
Ukraine	449376	1.02%	1.02%	4.25	Ukraine	44211100
United Arab Emirates	8621	0.09%	0.09%	0.39	United Arab Emirates	9211660
United Kingdom	188113	0.28%	0.28%	1.18	United Kingdom	66778660
United States	957455	0.29%	0.29%	1.20	United States	334319680
United States Virgin Islands	501	0.50%	0.50%	2.08	United States Virgin Islands	100837
Uruguay	10003	0.29%	0.29%	1.22	Uruguay	3428412
Uzbekistan	107394	0.33%	0.33%	1.36	Uzbekistan	32976946
Vanuatu	869	0.29%	0.29%	1.19	Vanuatu	304414
Venezuela	61510	0.21%	0.21%	0.89	Venezuela	28971686
Vietnam	240161	0.25%	0.25%	1.05	Vietnam	95776712
Yemen	55752	0.18%	0.18%	0.74	Yemen	31546694
Zambia	18157	0.10%	0.10%	0.41	Zambia	18380478
Zimbabwe	17810	0.12%	0.12%	0.48	Zimbabwe	15354606

SQL Queries:

```
*/Selecting all rows from Heart Disease table*/
SELECT [ID]
      ,[age]
      ,[sex]
      ,[chest_pain_type]
      ,[resting_bp_s]
      ,[cholesterol]
      ,[fasting_blood_sugar]
      ,[resting_ecg]
      ,[max_heart_rate]
      ,[exercise_angina]
      ,[oldpeak]
      ,[ST_slope]
      ,[target]
  FROM [UN].[dbo].[heart_statlog_cleveland_hungary_final]

-----------------------------------------------------------------------------------
*/Altering the “Final Data” table*/

alter table [dbo].[Final_Data]
add Human_Development_Index float null,
 Alcohol float null,
 Overweight float null,
 Tobacco_Use float null,
```

```
 Percentage_of_female float null,
 Literacy_rates nvarchar(50) null,
 GDP float null,
 Poverty float null

-------------------------------------------------------------------------------
*/Updating the “Final Data” table to add defined attributes from the raw table*/

update [dbo].[Final_Data]
set [Tobacco_Use] =  a.[Tobacco_Use]
 from [dbo].[Final_Data] as b inner join [dbo].[WID_share-of-adults-who-smoke] as a on
a.Entity = b.Region  where a.[Year] = 2019


update [dbo].[Final_Data]
set [Overweight] =  a.[Overweight]
 from [dbo].[Final_Data] as b inner join [dbo].[WID_share-of-adults-overweight] as a on
a.Entity = b.Region  where a.[Year] = 2016


update [dbo].[Final_Data]
set  [Alcohol] =  a.[Alcohol]
 from [dbo].[Final_Data] as b inner join [dbo].[WID_share-of-adults-Alcohol] as a on
a.Entity = b.Region  where a.[Year] = 2016

update [dbo].[Final_Data]
set  [Percentage_of_female] =  a.[Percentage_of_female]
 from [dbo].[Final_Data] as b inner join [dbo].[WID_share-population-female] as a on
a.Entity = b.Region  where a.[Year] = 2019


update [dbo].[Final_Data]
set  [Literacy_rates] =  a.[Literacy_rates]
 from [dbo].[Final_Data] as b inner join [dbo].[WID_cross-country-literacy-rates] as a on
a.Entity = b.Region  where a.[Year] = 2015

update [dbo].[Final_Data]
set  [GDP] =  a.[Value]
 from [dbo].[Final_Data] as b inner join [dbo].[GDP] as a on a.[Region_Country_Area]  =
b.[Region] where a.[Year] = 2019 and a.[Series] ='GDP per capita (US dollars)'

update [dbo].[Final_Data]
set  [Poverty] =  a.[Share_of_population_below_2_15_a_day]
 from [dbo].[Final_Data] as b inner join [dbo].[WID_poverty-explorer] as a on a.[Entity]
= b.[Region] where a.[Year] = 2019


update [dbo].[Final_Data]
set  [Percent_WRT_Population] =  a.[Percent_WRT_Population]
 from [dbo].[Final_Data] as b inner join [dbo].[Final_Data_2] as a on a.[Region] =
b.[Region]
```

```
-------------------------------------------------------------------------------
*/Below few steps are related to extraction of data from “Final Data” table to perform
analysis on Excel and build an input data for ML modeling*/
```

```
SELECT [Region]
      ,[CardioVascular_Deaths]
      ,[Percent_WRT_Population]
      ,[Human_Development_Index]
      ,[Alcohol]
      ,[Overweight]
      ,[Tobacco_Use]
      ,[Percentage_of_female]
      ,[Literacy_rates]
      ,[GDP]
      ,[Poverty]
  FROM [UN].[dbo].[Final_Data]
  order by [Region]


  Select
Min([Human_Development_Index]),Max([Percent_WRT_Population]),avG([Percent_WRT_Population]
)
  from  [UN].[dbo].[Final_Data]
  Where [Region] IS NOT NULL


  Select
avg([Human_Development_Index]),avg([Alcohol]),avg([Overweight]),avg([Tobacco_Use]),avg([P
ercentage_of_female]),avg([GDP]),avg([Poverty]),count(*)
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] <= 0.15 and [Region] IS NOT NULL

  Select
avg([Human_Development_Index]),avg([Alcohol]),avg([Overweight]),avg([Tobacco_Use]),avg([P
ercentage_of_female]),avg([GDP]),avg([Poverty]),count(*)
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] > 0.15 and [Percent_WRT_Population] <= 0.22 and [Region]
IS NOT NULL

  Select
avg([Human_Development_Index]),avg([Alcohol]),avg([Overweight]),avg([Tobacco_Use]),avg([P
ercentage_of_female]),avg([GDP]),avg([Poverty]),count(*)
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] > 0.22 and [Percent_WRT_Population] <= 0.55 and [Region]
IS NOT NULL

  Select
avg([Human_Development_Index]),avg([Alcohol]),avg([Overweight]),avg([Tobacco_Use]),avg([P
ercentage_of_female]),avg([GDP]),avg([Poverty]),count(*)
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] > 0.55 and [Region] IS NOT NULL
  -----------
  Select
avg([Human_Development_Index]),avg([Alcohol]),avg([Overweight]),avg([Tobacco_Use]),avg([P
ercentage_of_female]),avg([GDP]),avg([Poverty]),count(*)
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] <= 0.2 and [Region] IS NOT NULL

  Select
avg([Human_Development_Index]),avg([Alcohol]),avg([Overweight]),avg([Tobacco_Use]),avg([P
ercentage_of_female]),avg([GDP]),avg([Poverty]),count(*)
  from  [UN].[dbo].[Final_Data]
```

```
  where [Percent_WRT_Population] > 0.2 and [Region] IS NOT NULL
  -----
   Select [Region]
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] <= 0.2 and [Region] IS NOT NULL

  Select [Region]
  from  [UN].[dbo].[Final_Data]
  where [Percent_WRT_Population] > 0.2 and [Region] IS NOT NULL
```

```
Select [Entity],[Population_historical_estimates]
from [dbo].[WID_population-since-1800]
where [Year] =2009 and [Entity] NOT IN (
'Anguilla','Aruba','Bonaire Sint Eustatius and Saba','British Virgin Islands','Cayman
Islands',
'Curacao','Faeroe Islands','Falkland Islands','French Guiana','French
Polynesia','Gibraltar','Guadeloupe',
'Guernsey','Hong Kong','Isle of
Man','Jersey','Liechtenstein','Macao','Martinique','Mayotte','Montserrat','New
Caledonia',
'Saint Barthlemy','Saint Helena','Saint Martin (French part)','Saint Pierre and
Miquelon','Sint Maarten (Dutch part)',
'Turks and Caicos Islands','Vatican','Wallis and
Futuna','Africa','Asia','Europe','European Union (27)','High-income countries',
'Lower-middle-income countries','Low-income countries','Netherlands Antilles', 'North
America','Oceania','Reunion',
'South America','Svalbard and Jan Mayen','Upper-middle-income countries','Western
Sahara','World')
Order by [Entity] ASC


Select [Entity],[Population_historical_estimates]
from [dbo].[WID_population-since-1800]
where [Year] =1990 and [Entity] IN ('Lower-middle-income countries','Low-income
countries','Upper-middle-income countries','High-income countries','World')
Order by [Entity] ASC


Select [Entity],[Population_historical_estimates]
from [dbo].[WID_population-since-1800]
where [Year] =2019 and [Entity] IN ('Lower-middle-income countries','Low-income
countries','Upper-middle-income countries','High-income countries','World')
Order by [Entity] ASC


Select
[Entity],[Deaths_Cardiovascular_diseases_Sex_Both_Age_Under_5_Number],[Deaths_Cardiovascu
lar_diseases_Sex_Both_Age_5_14_years_Number],
[Deaths_Cardiovascular_diseases_Sex_Both_Age_15_49_years_Number],[Deaths_Cardiovascular_d
iseases_Sex_Both_Age_50_69_years_Number],[Deaths_Cardiovascular_diseases_Sex_Both_Age_70_
years_Number]
from [dbo].[WID_cardiovascular-disease-deaths-by-age]
where [Year] =1990 and [Entity] Like '%World%'
Order by [Entity] ASC
```

```
Select
[Entity],[Deaths_Cardiovascular_diseases_Sex_Both_Age_Under_5_Number],[Deaths_Cardiovascu
lar_diseases_Sex_Both_Age_5_14_years_Number],
[Deaths_Cardiovascular_diseases_Sex_Both_Age_15_49_years_Number],[Deaths_Cardiovascular_d
iseases_Sex_Both_Age_50_69_years_Number],[Deaths_Cardiovascular_diseases_Sex_Both_Age_70_
years_Number]
from [dbo].[WID_cardiovascular-disease-deaths-by-age]
where [Year] =2019 and [Entity] Like '%World%'
Order by [Entity] ASC


Delete from [dbo].[Final_Data] where [Region] is NULL

Select * from [dbo].[Final_Data]

Alter table [dbo].[Final_Data]
Add Cardio_Deaths_Percent float Null

Alter table [dbo].[Final_Data]
Add CD_Bucket Varchar(256) Null

Update [dbo].[Final_Data]
set Cardio_Deaths_Percent = ([CardioVascular_Deaths]/18552134.0)

Update [dbo].[Final_Data]
set CD_Bucket =
CASE
WHEN Cardio_Deaths_Percent <= 0.0008 Then 'Very Low'
WHEN Cardio_Deaths_Percent > 0.0008 and Cardio_Deaths_Percent <= 0.002 Then 'Low'
WHEN Cardio_Deaths_Percent > 0.002 and Cardio_Deaths_Percent <= 0.004 Then 'Average'
WHEN Cardio_Deaths_Percent > 0.004 and Cardio_Deaths_Percent <= 0.009 Then 'High'
WHEN Cardio_Deaths_Percent > 0.009 Then ' Very High'
END
```

Student Copyright Checklist

STUDENT COPYRIGHT CHECKLIST *(for students to complete and advisors to verify)*

STUDENT: Answer question 1 below.

1) Does your solution to the competitive event integrate any type of music and/or sound? ☐ YES ☑ NO

If NO, go to question 2.

If YES, is the music and/or sound copyrighted? ☐ YES ☑ NO

If YES, move to question 1A. If NO, move to question 1B.

1A) Have you asked for author permission to use the music and/or sound in your solution and included that permission (letter/form) in your documentation? If YES, move to question 2. If NO, ask for permission and if permission is granted, include the permission in your documentation.

1B) Is the music/sound royalty free, or did you create the music/sound yourself? If YES, cite the royalty free music/sound OR your original music/sound properly in your documentation.

CHAPTER ADVISOR: Sign below regarding your student's answer(s) to the use of music/sound in his/her competitive event solution. Even if your student answers "NO" to question 1, please sign below noting that you have evaluated the competitive event solution and the student answered the question(s) accurately.

I, Lori Zebrack-Smith (chapter advisor), have checked my student's solution and confirm that any use of music/sound is done so with proper permission and is cited correctly in the student's documentation and/or the solution has been found to have no music/sound included.

STUDENT: Answer question 2 below.

2) Does your solution to the competitive event integrate any graphics/videos? ☑ YES ☐ NO

If NO, go to question 3.

If YES, is(are) the graphics/videos copyrighted, registered and/or trademarked? ☐ YES ☑ NO

If YES, move to question 2A. If NO, move to question 2B.

2A) Have you asked for author permission to use the graphics and/or videos in your solution and included a permission (letter/form) in your documentation for graphic/video used? If YES, move to question 3. If NO, ask for permission and if permission is granted, include the permission in your documentation.

2B) Is(are) the graphics/videos royalty free, or did you create your own graphic? If YES, cite the royalty free graphics/videos OR your own original graphics/videos properly in your documentation.

CHAPTER ADVISOR: Sign below regarding your student's answer(s) to the use of graphics/videos in his/her competitive event solution. Even if your student answers "NO" to question 2, please sign below noting that you have evaluated the competitive event solution and the student answered the question(s) accurately.

I, Lori Zebrack-Smith (chapter advisor), have checked my student's solution and confirm that the use of graphics/videos with proper permission and is cited correctly in the student's documentation and/or the solution has been found to have no graphics/videos included.

STUDENT: Answer question 3 below.

3) Does your solution to the competitive event use another's thoughts or research? ☑ YES ☐ NO

If NO, this is the end of the checklist.

If YES, have you properly cited other's thoughts or research in your documentation? ☐ YES ☐ NO

CHAPTER ADVISOR: Sign below regarding your student's answer(s) to having integrated any thoughts/research of others in his/her competitive event solution. Even if your student answers "NO" to question 3, please sign below noting that you have evaluated the competitive event solution and the student answered the question(s) accurately.

I, Lori Zebrack-Smith (chapter advisor), have checked my student's solution and confirm that the use of the thoughts/research of others is done so with proper permission and is cited correctly in the student's documentation and/or the solution has been found to have all original thought with no use of other's thoughts/research.

Student Name: Rishika Gautam

Chapter Advisor Signature: Lori Zebrack-Smith