**Project Topic Area**

In this project, I aim to explore alcohol consumption and its effects on health, specifically heart disease. I researched multiple sources and identified three different datasets that will help me examine how alcohol consumption may relate to cardiovascular conditions and other health implications.

**Data Sources**

**Flat File:**

* **Dataset:** Heart Failure Clinical Records
* **Description:** This dataset contains 299 clinical records of patients who suffered from heart failure. It includes 13 health-related variables such as blood pressure, diabetes, smoking, age and more. While this dataset does not directly include alcohol consumption data, it still allows for an analysis of heart disease, which is commonly associated with excessive alcohol intake.
* **Source:** <https://www.kaggle.com/datasets/andrewmvd/heart-failure-clinical-data>
* **File Name:** heart\_failure\_clinical\_records\_dataset.csv

**API Source:**

* **Dataset:** WHO Global Health Observatory (GHO) - Alcohol Consumption & Disease Data
* **Description:** The WHO API contains 12,936 rows and 25 columns and provides up-to-date global health data on alcohol consumption and its impact on public health. It monitors alcohol consumption trends worldwide, its correlation with various health conditions, and policy responses from different countries. This dataset will allow me to analyze how alcohol-related health issues vary across regions.
* **Source:** WHO GHO OData API (<https://www.who.int/data/gho/info/gho-odata-api>)

**Website (Web Scraping):**

* **Dataset:** Wikipedia - List of countries by alcohol consumption per capita
* **Description:** The dataset includes 196 rows (4 columns) detailing total alcohol consumption per capita and 189 rows (9 columns) categorizing consumption by type (beer, wine, spirits). It measures alcohol intake in liters of pure ethanol per person annually. The dataset also categorizes alcohol consumption into three types: beer, wine, and spirits for each country.
* **Source:** Wikipedia (<https://en.wikipedia.org/wiki/List_of_countries_by_alcohol_consumption_per_capita>)

**Relationships between the datasets**

While these datasets do not share an explicit common key, meaningful connections can still be established through analytical comparisons:

1. **Heart Failure Clinical Data & WHO API Data:**

* The heart failure dataset provides individual-level patient health data, whereas the WHO API contains country-level statistics on alcohol consumption and its related heart outcomes.
* By analyzing cardiovascular risk factors (e.g., hypertension, diabetes, and smoking) in the heart failure dataset alongside WHO data on alcohol-related diseases, I can explore potential correlations between alcohol consumption and cardiovascular mortality trends.

1. **Wikipedia Alcohol Consumption & WHO API Data:**

* The Wikipedia dataset provides per capita alcohol consumption across various countries, while the WHO API details the health consequences of alcohol consumption in those same countries.
* This allows for an analysis of whether countries with higher alcohol consumption rates also exhibit higher alcohol-related mortality rates.

1. **Connecting all three datasets:**

* The heart failure dataset lacks country-specific information, which prevents direct merging with WHO API or Wikipedia dataset. However, I can still compare trends in alcohol consumption and cardiovascular disease prevalence from the WHO API and Wikipedia to infer possible relationships.
* For example, I can assess whether known risk factors in heart failure patients, such as hypertension or diabetes, align with alcohol-related health concerns reported in the WHO dataset.

**Project Approach and Challenges**

**Approach:**

* Extract and clean data from all three sources. This includes preprocessing tasks like handling missing values, standardizing country names across datasets, and ensuring numerical data consistency.
* Identify trends and correlations between alcohol consumption and country-level alcohol-related mortality rates.
* Apply statistical analysis techniques to explore potential relationships and patterns.

**Challenges:**

* **Data Merging:** The absence of a common key (e.g., country name) in the heart failure dataset requires indirect comparison rather than direct integration.
* **Variability in Data Formats:** The datasets originate from different sources, including medical records and country-level statistics, which requires careful standardization.
* **Correlation vs. Causation:** While alcohol is a well-known cardiovascular risk factor, providing direct causation requires careful interpretation and additional contextual analysis.

**Ethical Implications**

* **Privacy & Data Sensitivity:** Although the patient data is anonymized, handling health-related data requires ethical responsibility to avoid misuse or misinterpretation.
* **Responsible Data Interpretation:** Correlation does not imply causation. Careful statistical analysis and contextual understanding are needed to prevent misleading conclusions about alcohol’s direct impact on heart disease.
* **Public Health Relevance**: Raising awareness of alcohol’s role in cardiovascular disease could help promote better public health policies and interventions.

**Conclusion**

This project aims to identify correlations between alcohol consumption and heart disease while considering other health conditions linked to excessive alcohol intake. By combining patient-level medical records with country-specific alcohol consumption trends and disease data, I hope to explore meaningful patterns that could contribute to public health insights.