REPORT ON MINI – PROJECT

Name:	MALAVIKA .S
Batch Code:	TN_DA-FNB09
Email id:	malusounder1609@gmail.com

Project Title:	Global Coffee Intake and Health Analysis
Project Domain:	Healthcare
Submission date:	Jan 19,2026
Mentor Name:	JOSEPH DELMON. H
Dataset Link:	https://www.kaggle.com/datasets/uom190346a/global-coffee-health-dataset
Cleaned Dataset link:	https://github.com/malavikasounderrajan/Mini-Project-On-Excel-Power-BI/blob/main/Global%20Coffee%20Intake%20and%20Health%20Analysis.xlsx

GLOBAL COFFEE INTAKE AND HEALTH ANALYSIS

INTRODUCTION:

Coffee is one of the most widely consumed beverages across the world and plays a significant role in daily dietary habits. Its consumption has been linked to various health outcomes, making it an important subject for healthcare and nutritional analysis.

OBJECTIVE OF THE PROJECT:

This mini project, titled **Global Coffee Intake and Health Analysis**, aims to examine global coffee consumption patterns and their potential impact on health using data analysis and visualization techniques. The project utilizes datasets analysed through **Microsoft Excel and Power BI** to explore trends in coffee intake across different countries, age groups, and time periods, along with associated health indicators. By applying data cleaning, transformation, and analytical methods, meaningful insights are derived to understand correlations between coffee consumption and health-related factors.

The key objectives of this mini project are:

1. To analyse global coffee consumption patterns across different countries and regions.
2. To study variations in coffee intake based on demographic and lifestyle factors.
3. To examine the relationship between coffee consumption and selected health indicators.
4. To identify trends and patterns using data analysis techniques.
5. To develop interactive dashboards and visual reports using Power BI for effective insight presentation.

SCOPE OF THE PROJECT

The scope of this project includes:

- Analysis of historical and cross-sectional global coffee consumption data.
- Evaluation of selected health indicators related to coffee intake.

- Data cleaning, transformation, and modelling using analytical tools.
- Visualization of insights through charts, KPIs, and interactive dashboards.
- The project is limited to the available dataset and does not include clinical trials or medical diagnoses.

DATASET DESCRIPTION:

Column	**Type**	**Description**
ID	Integer	Unique record ID (1–10000)
Age	Integer	Age of participant (18–80 years)
Gender	Categorical	Male, Female, Other
Country	Categorical	Country of residence (20 countries)
Coffee_Intake	Float	Daily coffee consumption in cups (0–10)
Caffeine_mg	Float	Estimated daily caffeine intake in mg (1 cup ≈ 95 mg)
Sleep_Hours	Float	Average hours of sleep per night (3–10 hours)

Column	Type	Description
Sleep_Quality	Categorical	Poor, Fair, Good, Excellent (based on sleep hours)
BMI	Float	Body Mass Index (15–40)
Heart_Rate	Integer	Resting heart rate (50–110 bpm)
Stress_Level	Categorical	Low, Medium, High (based on sleep hours and lifestyle)
Physical_Activity_Hours	Float	Weekly physical activity (0–15 hours)
Health_Issues	Categorical	None, Mild, Moderate, Severe (based on age, BMI, and sleep)
Occupation	Categorical	Office, Healthcare, Student, Service, Other

Smoking	Boolean	0 = No, 1 = Yes
Alcohol_Consumption	Boolean	0 = No, 1 = Yes

METHODOLOGY:

The methodology adopted for this mini project follows a structured data analytics approach to analyse global coffee consumption patterns and their relationship with health indicators. The analysis was carried out using Microsoft Excel and Power BI.

1. Data Collection

The dataset was obtained from publicly available secondary sources related to global coffee consumption and healthcare statistics. The data includes information on coffee intake levels, demographic attributes, and selected health indicators across multiple countries and time periods.

2. Data Cleaning and Preprocessing

Column-Wise Data Cleaning & Formatting Process

a. ID:

Action Taken:

- Checked for **duplicates** and ensured each ID is **unique**
- No missing values found

Reason:

- Maintains **row-level uniqueness** and prevents double counting

b. Age:

Action Taken:

- Verified age values fall within a **realistic range (18–65)**
- Converted to **Whole Number**

- Removed invalid or blank entries.

Reason:

- Avoids outliers that could distort health and age-related analysis

c. Gender:

Action Taken:

- Standardized values to **Male / Female / Other**
- Corrected inconsistent spellings (e.g., “M”, “male” → “Male”)
- Converted to **Text (Categorical)**

Reason:

- Ensures accurate gender-wise aggregation and visuals

d. Country:

Action Taken:

- Standardized country names (e.g., “U.K” → “UK”)
- Removed leading/trailing spaces
- Set data type as **Text**

Reason:

- Prevents duplication during country-level analysis and mapping

e. Coffee Intake:

Action Taken:

- Checked for **negative or zero values**
- Converted to **Decimal Number**
- Rounded to **1 decimal place**

Reason:

- Represents realistic daily coffee consumption levels

f. Caffeine mg:

Action Taken:

- Converted to **Decimal Number**
- Removed unrealistic outliers
- Ensured consistency with Coffee Intake values

Reason:

- Critical for caffeine exposure and health impact analysis

g. Sleep Hours:

Action Taken:

- Validated range between **3–10 hours**
- Converted to **Decimal**
- Rounded to **1 decimal place**

Reason:

- Eliminates impossible sleep values and improves accuracy

h. Sleep_Quality:

Action Taken:

- Categorized into **Excellent / Good / Fair / Poor**
- Standardized text formatting
- Removed blank or invalid entries

Reason:

- Enables clean grouping for sleep-based insights

i. BMI:

Action Taken:

- Converted to **Decimal Number**
- Checked for unrealistic BMI values (<10 or >50)
- Rounded to **1 decimal place**

Reason:

- Prevents skewed BMI-health relationships

j. Heart Rate:

Action Taken:

- Converted to **Whole Number**
- Ensured values fall within **50–120 bpm**
- Removed abnormal outliers

Reason:

- Keeps physiological data realistic and reliable

k. Stress Level:

Action Taken:

- Standardized into **Low / Medium / High**
- Removed inconsistent labels
- Converted to **Text**

Reason:

- Enables accurate stress-wise behavioural analysis

l. Physical Activity Hours:

Action Taken:

- Converted to **Decimal**
- Validated values between **0–24 hours**
- Rounded to **1 decimal place**

Reason:

- Ensures logical activity tracking per individual

m. Health Issues:

Action Taken:

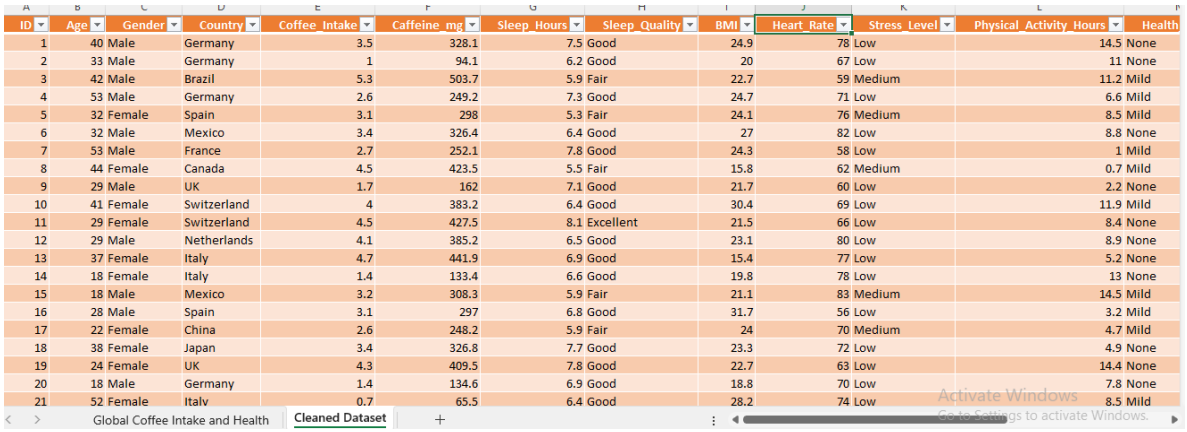
- Categorized into **None / Mild / Moderate / Severe**
- Cleaned spelling and casing
- Removed null values

Reason:

- Critical for severity-based health analysis

Final Formatting Applied

- Dataset converted to **Excel Table format**
- Consistent **column headers**
- Proper **data types assigned**
- Conditional formatting applied for better readability
- Ready for **Power BI import**



ID	Age	Gender	Country	Coffee_Intake	Caffeine_mg	Sleep_Hours	Sleep_Quality	BMI	Heart_Rate	Stress_Level	Physical_Activity_Hours	Health
1	40	Male	Germany	3.5	328.1	7.5	Good	24.9	78	Low	14.5	None
2	33	Male	Germany	1	94.1	6.2	Good	20	67	Low	11	None
3	42	Male	Brazil	5.3	503.7	5.9	Fair	22.7	59	Medium	11.2	Mild
4	53	Male	Germany	2.6	249.2	7.3	Good	24.7	71	Low	6.6	Mild
5	32	Female	Spain	3.1	298	5.3	Fair	24.1	76	Medium	8.5	Mild
6	32	Male	Mexico	3.4	326.4	6.4	Good	27	82	Low	8.8	None
7	53	Male	France	2.7	252.1	7.8	Good	24.3	58	Low	1	Mild
8	44	Female	Canada	4.5	423.5	5.5	Fair	15.8	62	Medium	0.7	Mild
9	29	Male	UK	1.7	162	7.1	Good	21.7	60	Low	2.2	None
10	41	Female	Switzerland	4	383.2	6.4	Good	30.4	69	Low	11.9	Mild
11	29	Female	Switzerland	4.5	427.5	8.1	Excellent	21.5	66	Low	8.4	None
12	29	Male	Netherlands	4.1	385.2	6.5	Good	23.1	80	Low	8.9	None
13	37	Female	Italy	4.7	441.9	6.9	Good	15.4	77	Low	5.2	None
14	18	Female	Italy	1.4	133.4	6.6	Good	19.8	78	Low	13	None
15	18	Male	Mexico	3.2	308.3	5.9	Fair	21.1	83	Medium	14.5	Mild
16	28	Male	Spain	3.1	297	6.8	Good	31.7	56	Low	3.2	Mild
17	22	Female	China	2.6	248.2	5.9	Fair	24	70	Medium	4.7	Mild
18	38	Female	Japan	3.4	326.8	7.7	Good	23.3	72	Low	4.9	None
19	24	Female	UK	4.3	409.5	7.8	Good	22.7	63	Low	14.4	None
20	18	Male	Germany	1.4	134.6	6.9	Good	18.8	70	Low	7.8	None
21	52	Female	Italy	0.7	65.5	6.4	Good	28.2	74	Low	8.5	Mild

Global Coffee Intake and Health | Cleaned Dataset

Fig: Cleaned Dataset

4. Data Analysis

Exploratory data analysis was conducted to identify trends, patterns, and correlations between coffee intake and health indicators. Key performance metrics and summaries were generated to evaluate:

- Global and regional coffee consumption trends
- Demographic variations in coffee intake
- Associations between coffee consumption and selected health outcomes

5. Data Visualization and Dashboard Development

- Interactive dashboards were developed using Power BI to present insights effectively. Visual elements such as charts, graphs, KPIs, and filters were used to enable dynamic analysis and easy interpretation of results.

6. Interpretation and Reporting

- The final step involved interpreting the analytical findings and documenting insights in a structured report. The results were summarized to highlight key observations, limitations, and potential areas for further study.

VISUALISATION IN POWER BI:

1. Card visuals:

High Overall Caffeine Exposure: The dataset records a total caffeine intake of 2.38 million mg, indicating substantial caffeine consumption across the population studied.

Strong Coffee Consumption Trend: With 25.09K total coffee intakes, coffee emerges as a major contributor to overall caffeine intake, reinforcing its role as a daily habit.

Balanced Sleep Duration Despite Caffeine: 66.36K total sleep hours suggest that, at an aggregate level, caffeine consumption does not drastically reduce sleep duration, though individual-level impacts may vary.

Active Lifestyle Indicator: 74.87K physical activity hours reflect a generally active population, which may help offset some negative health effects associated with high caffeine intake.

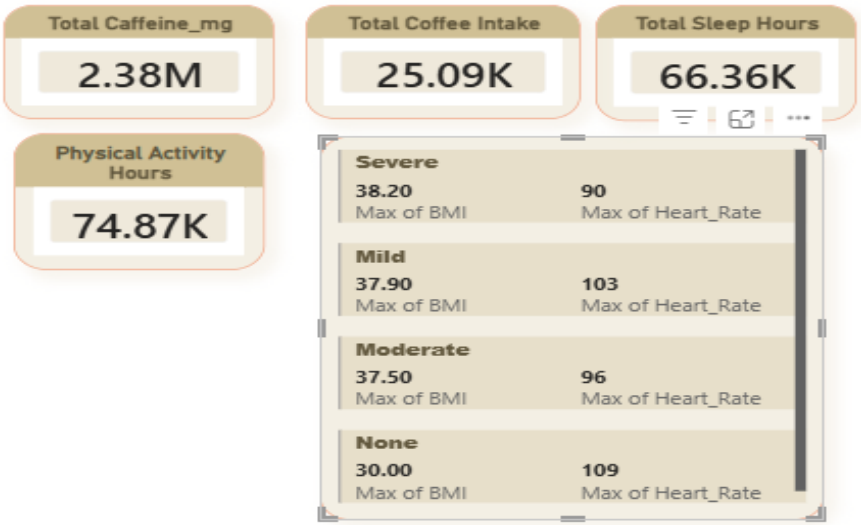


Fig: Cards Visuals

2. Line and Stacked column chart:

Age vs Health Severity

- Individuals with **Severe health issues have the highest average age (≈56.6 years)**, indicating that health severity increases with age.
- The **average age gradually decreases** from *Severe* → *Moderate* → *Mild* → *None*, showing a clear age-related health risk pattern.
- People with **no reported health issues are the youngest group (≈32.9 years)**.

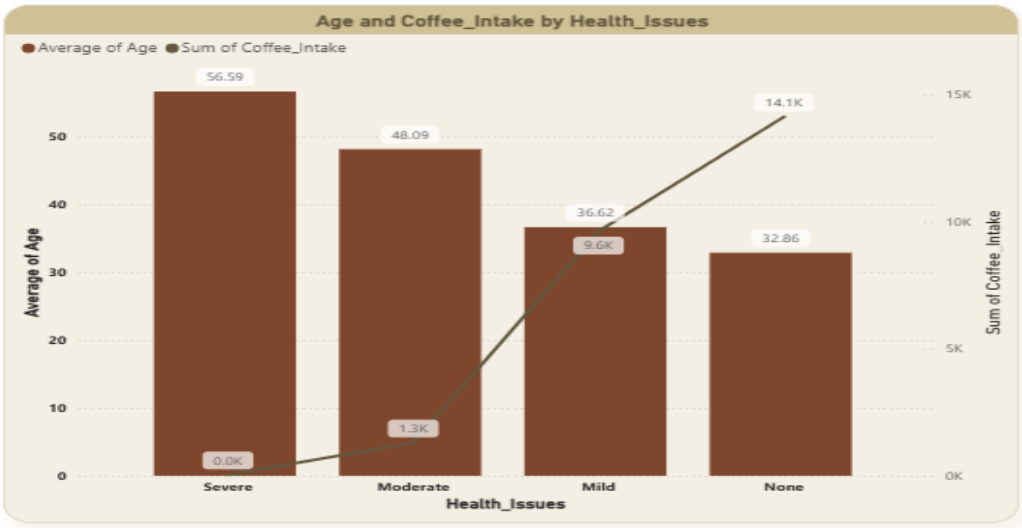


Fig: Line & Stacked Column Chart

3.Line Chart:

Highest Maximum Caffeine Levels

- Netherlands records the highest maximum caffeine intake (~780 mg), indicating very strong coffee or high single-day caffeine consumption.
- Australia and Belgium follow closely, both exceeding 740 mg, placing them among the top high-caffeine countries.

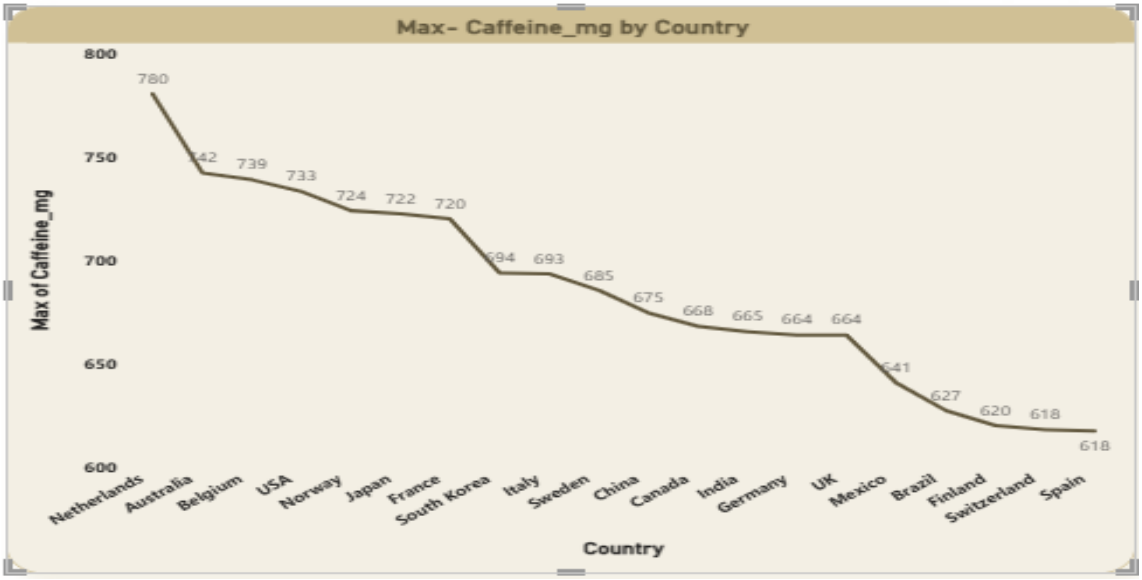


Fig: Line Chart

4. Area Chart:

Inverse Relationship Between Stress and Coffee Intake

- Low stress individuals consume the most coffee (~16.6K), indicating higher caffeine intake when stress levels are manageable.
- Coffee intake drops sharply to ~5.7K at medium stress, showing a significant behavioral shift.
- High stress individuals consume the least coffee (~2.9K), suggesting possible avoidance or reduced tolerance.

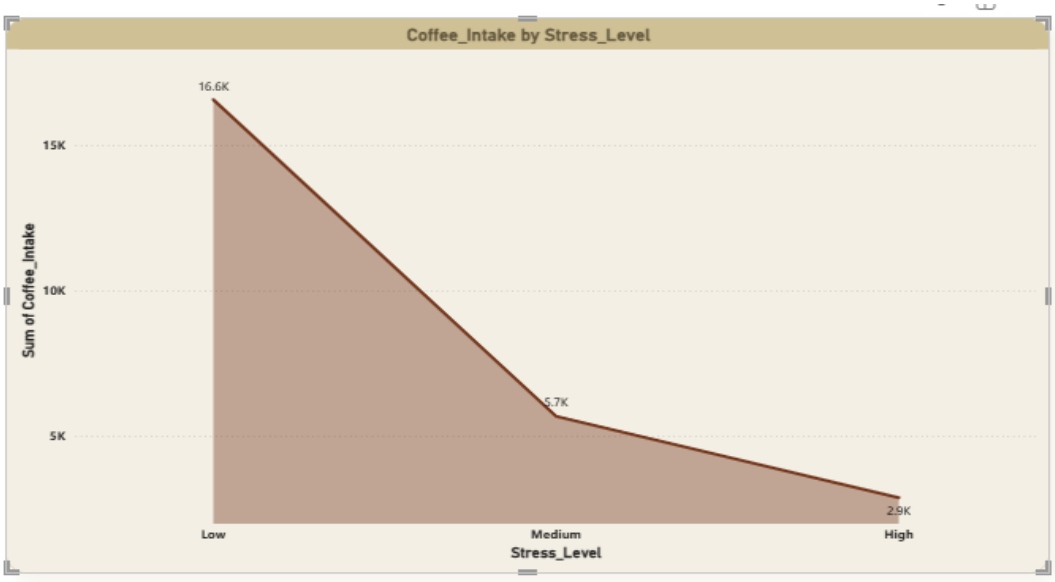


Fig: Area Chart

5.Bar Chart:

Female vs Male Consumption

- Females show the highest total coffee intake (~12.5K), slightly exceeding male consumption.
- Males closely follow at ~12.0K, indicating nearly equal coffee-drinking behavior across the two major gender groups.
- The small difference suggests that gender is not a strong differentiator in overall coffee consumption.

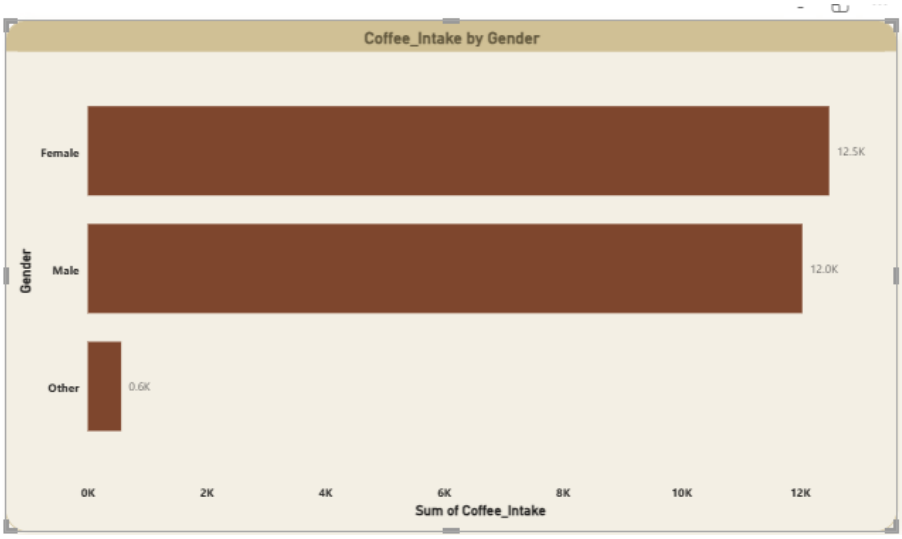


Fig: Bar Chart

6.Column Chart:

Coffee Intake Decreases with Health Severity

- Individuals with no health issues consume the highest amount of coffee (~14.1K).
- Coffee intake declines to ~9.6K for mild health issues, indicating early lifestyle moderation.
- A sharp drop is observed for moderate issues (~1.3K), suggesting stronger health-related restrictions.
- Severe health issues show almost zero coffee intake, highlighting strict avoidance or medical advice.

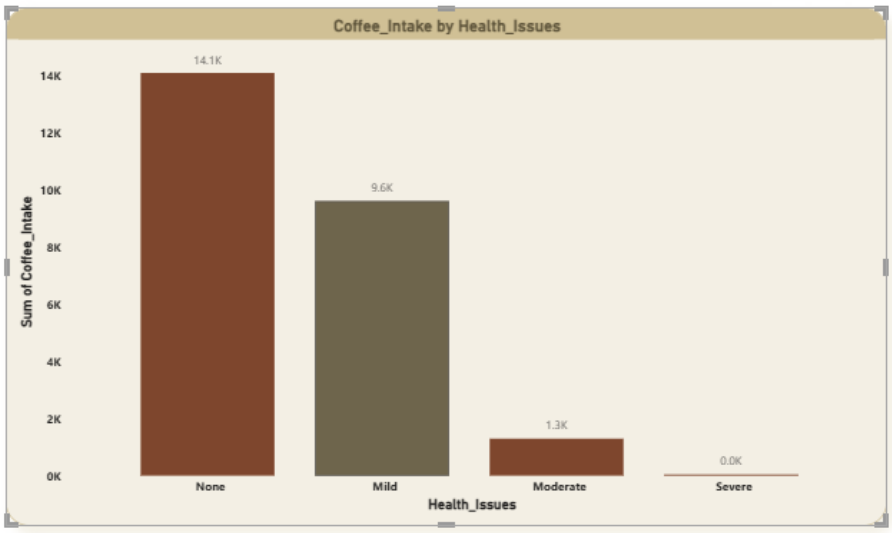


Fig: Column Chart

7.Pie Chart:

Balanced Coffee Consumption Across Occupations

- Coffee intake is evenly distributed across all occupation groups, with values ranging narrowly between ~4.9K and ~5.15K.
- No single occupation overwhelmingly dominates coffee consumption, indicating coffee is a universal habit across professions.

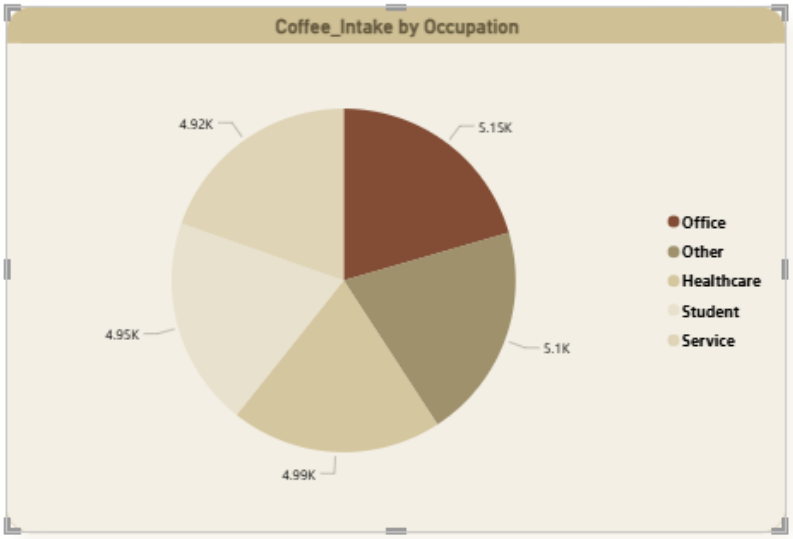


Fig: Pie Chart

8.Table:

Good Sleep Quality Stands Out

- Individuals with Good sleep quality show the highest coffee intake (~13.8K).
- They also record the highest physical activity hours (~42.3K), indicating an active lifestyle.
- This group has the highest total BMI value, likely reflecting a larger sample size, not necessarily poorer health.

Fair Sleep Quality

- Moderate coffee intake (~5.7K) and physical activity (~15.3K).
- BMI values are significantly lower than the *Good* category, suggesting reduced overall participation or sample count.

Poor Sleep Quality

- Individuals with Poor sleep consume less coffee (~2.86K) and engage in minimal physical activity (~7.33K).
- This indicates that poor sleep may be associated with lower energy levels and reduced activity, rather than higher caffeine use.

Excellent Sleep Quality

- Despite lower coffee intake (~2.77K), this group maintains relatively high physical activity (~9.97K).
- This suggests that good sleep may reduce dependence on caffeine while supporting active lifestyles.

Sleep_Quality	Sum of Coffee_Intake	Sum of Physical_Activity_Hours	Sum of BMI
Good	13,792.50	42,298.10	1,34,936.60
Fair	5,666.30	15,267.30	49,070.20
Poor	2,863.20	7,330.70	23,118.10
Excellent	2,770.30	9,974.30	32,743.70

Fig: Table

9.KPI:

Health Issues Significantly Exceed the Benchmark

- The current health issue count stands at 1.15K, which is well above the target goal of 448.
- This represents a +156.34% increase over the goal, signaling a substantial public-health concern.

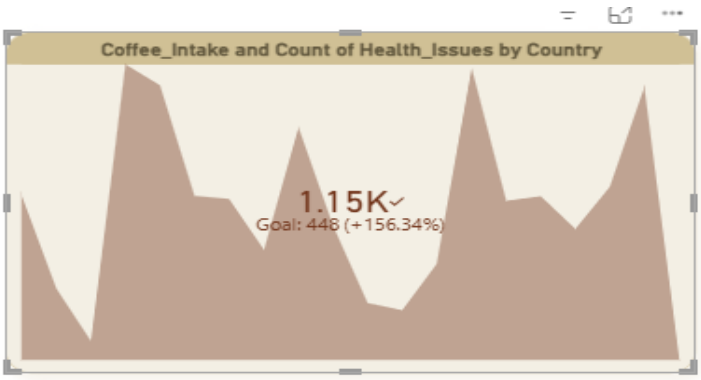


Fig: KPI

10.Scatter Plot:

Strong Positive Correlation

- The scatter points form a clear upward diagonal trend, indicating a strong positive relationship between alcohol consumption and smoking.
- As alcohol consumption increases, smoking frequency also rises, suggesting these habits often co-occur.

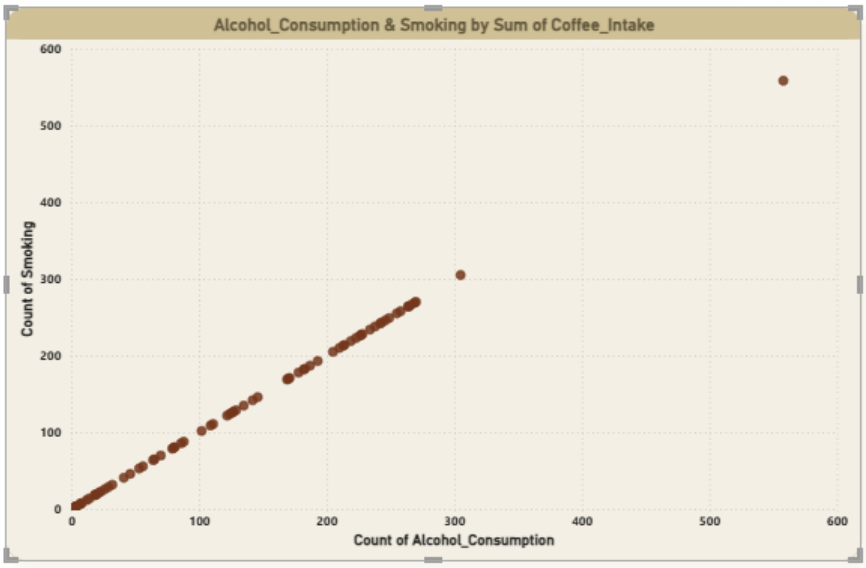


Fig: Scatter Plot

Final Conclusion:

Coffee consumption by itself does not directly indicate poor health outcomes. Instead, age, sleep quality, physical activity, stress, and associated habits like alcohol and smoking collectively shape health risks. Balanced lifestyles allow moderate-to-high coffee intake without adverse health effects.

Descriptive: Health issue counts vary widely by country.

Diagnostic: Elevated health issues may relate to combined effects of caffeine, lifestyle, and demographics.

Prescriptive: Country-level caffeine awareness and lifestyle programs are recommended.

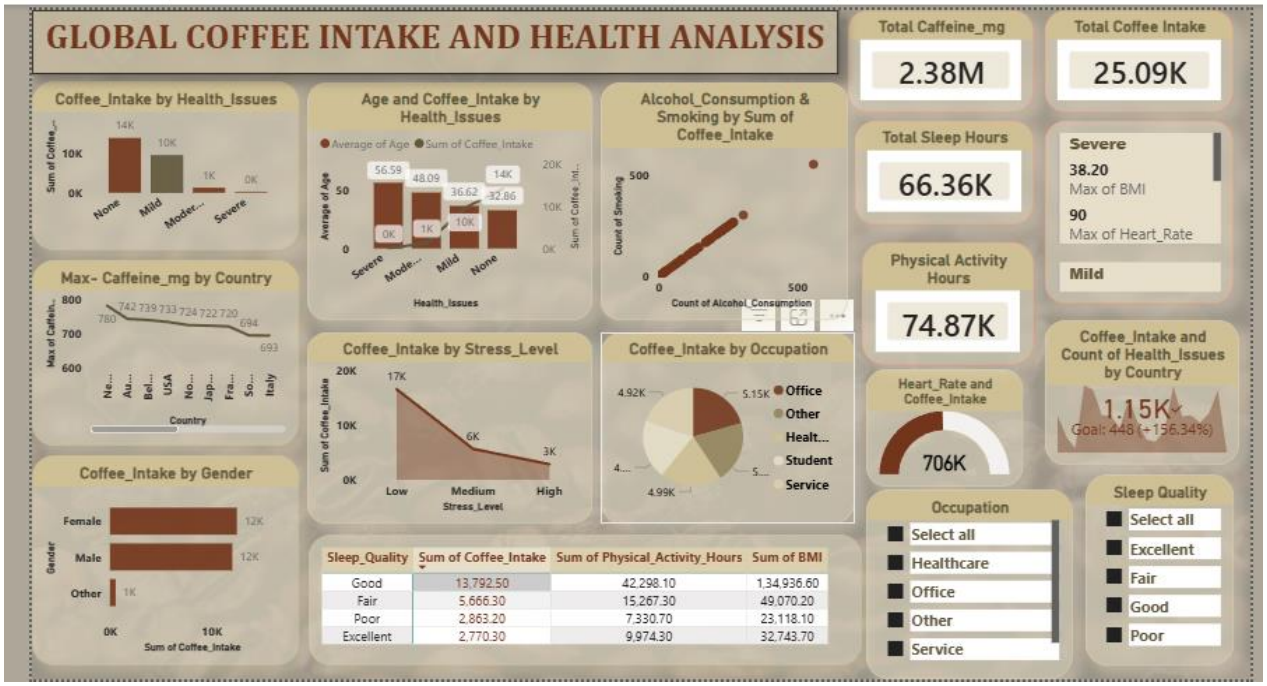


Fig: Final Dashboard

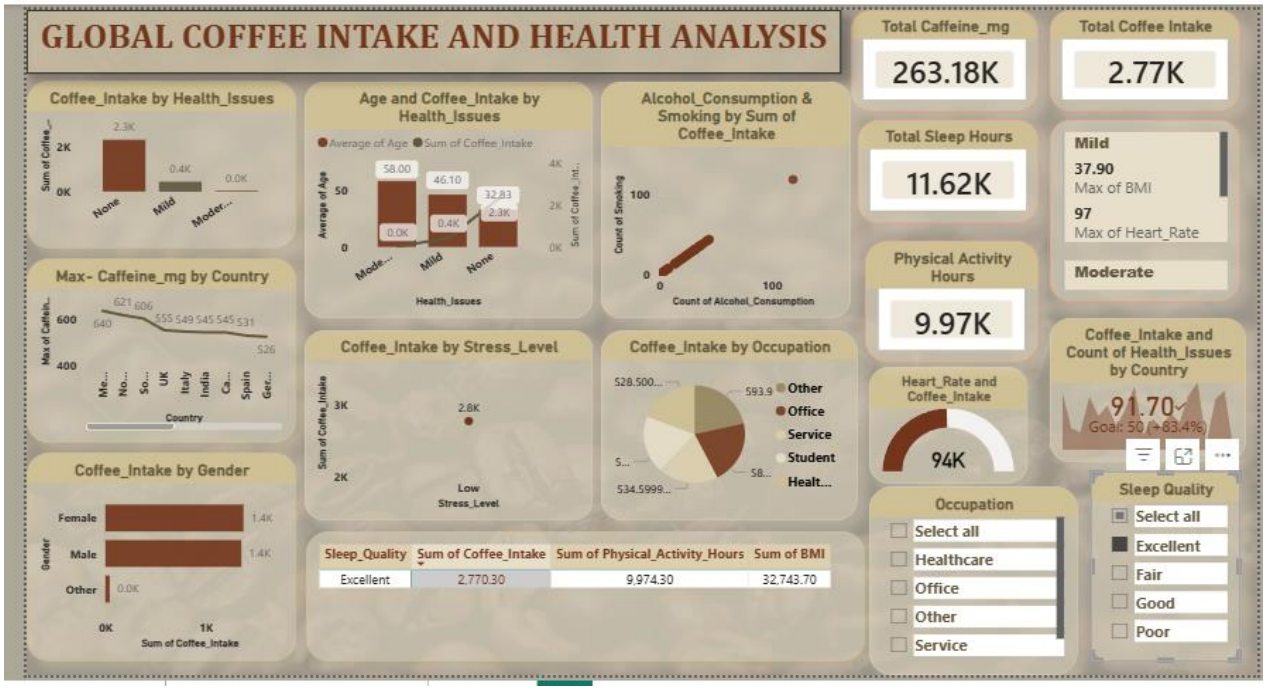


Fig: Excellent Category interactive Dashboard using Slicers

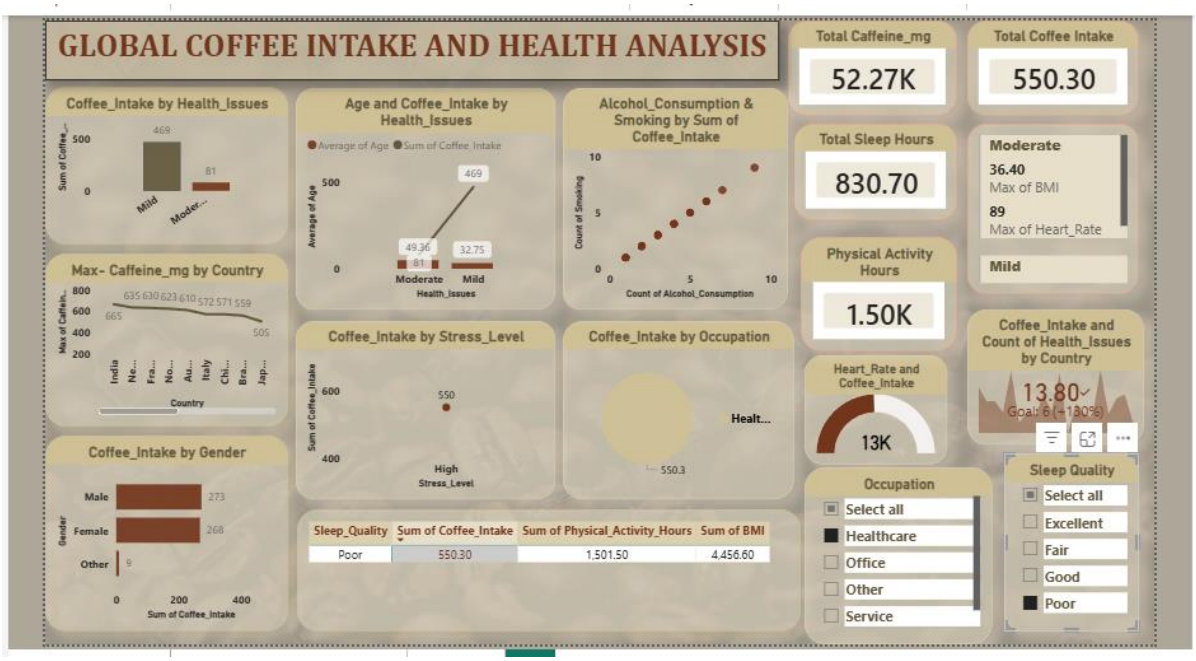


Fig: Healthcare -Poor Category using Slicers

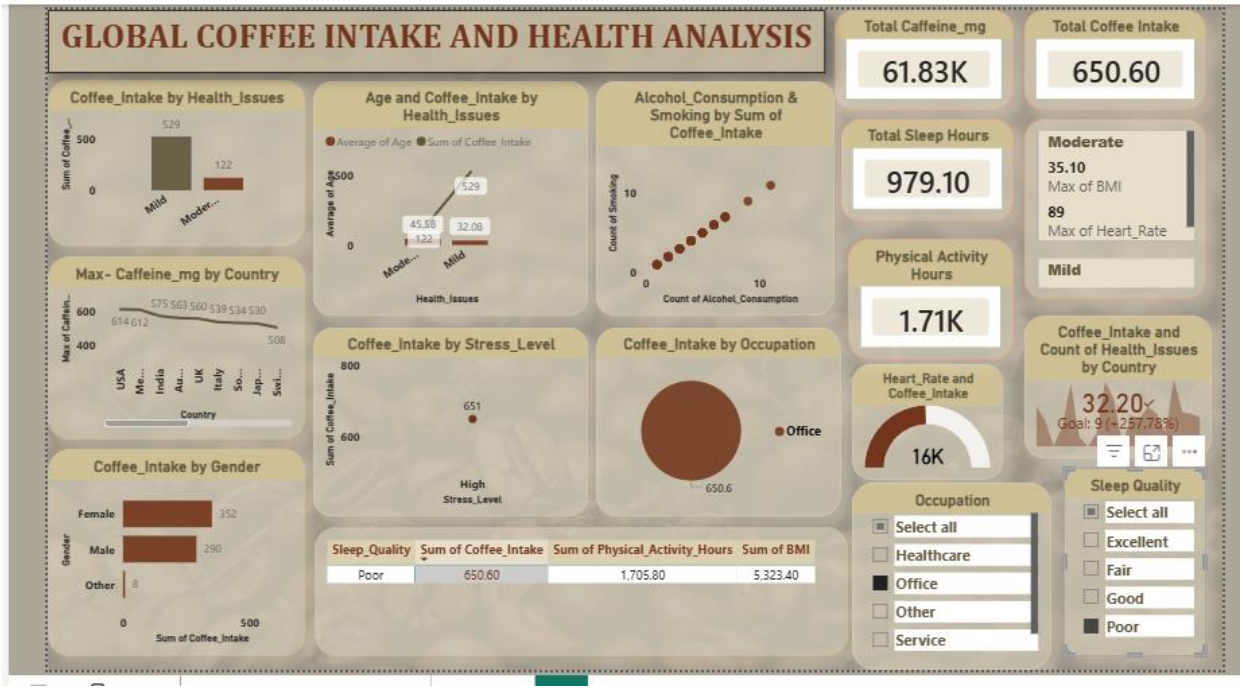


Fig : Office-Poor category using Slicers