**FAO Coffee Dataset (**2010-2022)

| **Column** | **Description** |
| --- | --- |
| Domain Code / Domain | The data domain, e.g., "Crops and livestock products" or "Food Balance Sheets" |
| Area Code (M49) / Area | Country code and name (e.g., 840 = USA) (190 countries) |
| Element Code / Element | **The metric type**, e.g., production, consumption, supply, imports, exports |
| Item Code (FBS) / Item | The item, e.g., "Coffee, green" |
| Year Code / Year | Year (2010-2022) |
| Unit | Measurement unit (e.g., tonnes, kg/capita/year) |
| Value | The actual measurement value |
| Flag / Flag Description | Metadata (e.g., estimated, imputed) |
| Note | Extra notes from FAO (optional) |

you have the following datasets:

* ✅ **NHANES**
  + DEMO (demographics: age, gender, income, education, urban/rural)
  + BMX (body metrics: BMI, waist, weight, height)
  + DR1TOT / DR1IFF (nutrient/food intake, including caffeine)

**✅ Key Columns Explained (from DEMO)**

| **Column** | **Description** |
| --- | --- |
| SEQN | Respondent ID (used to merge across files) |
| RIAGENDR | Gender (1 = Male, 2 = Female) |
| RIDAGEYR | Age in years |
| DMDEDUC2 | Education level |
| INDFMPIR | Income-to-poverty ratio (a proxy for socioeconomic status) |
| RIDRETH1 | Race/ethnicity |
| WTINTPRP, WTMECPRP | Sample weights (for statistical representativeness) |

Columns in BMX

| **Column** | **Description** |
| --- | --- |
| SEQN | Respondent ID (used for merging with other files) |
| BMXWT | Weight (kg) |
| BMXHT | Height (cm) |
| BMXBMI | Body Mass Index (kg/m²) |
| BMXWAIST | Waist circumference (cm) |
| BMXARMC | Arm circumference (cm) |
| BMXLEG | Upper leg length (cm) |
| BMXTRI | Triceps skinfold (mm) |
| BMXARML | Upper arm length (cm) |

As we’re focusing on coffee's relationship with health metrics, these are the key ones:

* **BMXBMI** → BMI, body fat indicator
* **BMXWAIST** → Waist circumference, abdominal fat
* **BMXWT** & **BMXHT** → Raw body size

**✅ Good KPI Questions (Key Performance Indicator Ideas)**

**🔵 Health & Consumption**

1. **How does BMI relate to caffeine (coffee) intake?**
   * KPI: Avg caffeine intake by BMI category (underweight, normal, overweight, obese)
2. **What is the average caffeine intake by age group and gender?**
   * KPI: Mean caffeine (mg/day) per age group × gender
3. **Is there a significant difference in caffeine intake between urban vs rural?**
   * KPI: Mean caffeine intake (mg) → Urban vs Rural
4. **What percentage of people drink more than 200mg of caffeine daily?**
   * KPI: % of population above recommended limit

**🟢 Socioeconomics & Preferences**

1. **Does income level or education correlate with coffee/caffeine consumption?**
   * KPI: Avg caffeine intake by income quartile or education level
2. **Are younger vs older people more likely to consume high-caffeine drinks?**
   * KPI: Age-wise caffeine brackets
3. **Do men or women consume more coffee on average?**
   * KPI: Mean caffeine intake (mg/day) by gender

**🌍 Cross-Country Comparison (NHANES + FAO)**

1. **How does individual caffeine intake in the US compare to national supply data (FAO)?**
   * KPI: Avg intake per capita (NHANES) vs FAO supply
2. **Which countries have the highest coffee consumption vs GDP per capita?**
   * KPI: Coffee consumption (kg/capita) vs GDP

**Detailed Deliverables (for your FAO + NHANES project):**

1. ✅ **Jupyter Notebook**:  
   A well-documented notebook that includes data loading, cleaning, analysis, visualizations, and results.
2. ✅ **Cleaned Datasets**:  
   Processed and merged FAO and NHANES datasets (e.g., in CSV format), ready for analysis.
3. ✅ **Data Visualizations**:  
   Graphs, charts, or maps showing trends in food consumption, caffeine intake, or nutritional comparisons.
4. ✅ **Statistical Analysis or Model**:  
   A simple machine learning model (e.g., regression/classification) or correlation analysis between food intake and health indicators.
5. ✅ **Final Report or Slide Deck**:  
   A PDF or presentation summarizing your methodology, key findings, insights, and conclusion.
6. ✅ **Code Repository** (optional):  
   GitHub repo containing code, README file, and instructions to reproduce the analysis.