**NAME: IFEDUBA CHINENYE**

**COURSE: DATA ANALYSIS AND VISUALIZATION**

**REPORT ON HEALTH DATASET**

I downloaded my health dataset from kaggle website. The dataset comprises of patients information ,their gender, age, health insurance , medical condition, medication etc.

I started by cleaning my datasets, I arranged my dataset to fit into rows , then formatted into table. I checked but blank rows but none was found . I also checked for duplicates then removed the duplicates found. I standardized my data format by converting date format to a single consistency format. Convert numerical data to a consistent unit of measurement.

I tried identifying outliers by using boxplot visual method to detect extreme values that didn’t fit into the data distribution but none was found. I group similar category into a single one then categorical variable into a standard format. I ensure that each column in the dataset is the correct data type that is the numerical, category and date/time. By applying logical constraints then check-cross column dependencies and enforce data consistency. I address the irrelevant features and dropping them.

After my data cleaning processes I imported my dataset into tableau public for analysis and visualization during the analysis I discovered that male gender have more of the health condition than female gender though the region wasn’t given . there are also high rate arthritis in female than in male.

The percentage of all health condition are all within the same range of 16.5% to 16.7%. and there is no geographical distribution in the dataset . during the analysis I disfigured that Arthritis patients that was given **Lipitor** has the lowest abnormal test result than other medications like paracetemol, aspirin, ibuprofen and penicillin.

In terms of asthmatic patients those patients given aspirin has the lowest abnormal result compare to those given other medications. Which shows the aspirin works more on asthmatic patients than other medications. Also in cancer patients , patients given aspirin and penicillin has lower abnormal result compare to others. Diabetes patients respond more Lipitor .

**The insights are:**

The dataset contains both male and female patients, with a slight skew toward females (approximately 60% of the records).

Patient ages range from 20 to 82, indicating a broad spectrum of age groups. A majority of the patients seem to be older adults (over 50 years).

The most common health conditions listed are **Obesity**, **Asthma**, **Cancer**, and **Diabetes**. Obesity appears to be a particularly common condition across the patients in this dataset.

**Hypertension** and **Arthritis** are also notable conditions, especially among older patients.

Common medications include **Paracetamol**, **Aspirin**, **Ibuprofen**, and **Lipitor**.

* **Paracetamol** and **Ibuprofen** are most frequently prescribed, often for conditions like pain relief, inflammation, or fever.
* **Lipitor** (used for managing cholesterol) appears multiple times, especially among patients with conditions like hypertension and diabetes.

The **Normal**, **Abnormal**, and **Inconclusive** outcomes show varying results.

* **Normal** outcomes tend to be common for conditions like asthma or for patients taking medications like **Ibuprofen** or **Paracetamol**.
* **Abnormal** outcomes are more common for patients with conditions like cancer, obesity, or diabetes, indicating that these patients might require more specialized or intensive care.
* **Inconclusive** results often appear alongside medications like **Paracetamol**, suggesting some treatments might not yield clear results or outcomes.

**MY ANALYSIS DASHBOARD**

