**High Impact Skills Development Program**

**in Artificial Intelligence, Data Science, and Blockchain**

**Module 3: Data Visualization**

Instructor: Dr. Arham Muslim

Assistant Professor, SEECS, NUST

# NAME: MUHAMMAD SAQLAIN

# ROLL NO: GIL\_DSAI\_255

# SECTION: 05

# ****Exploratory Data Analysis Report****

# ****DATASET:** Healthcare**

## ****Introduction****

This report outlines the Exploratory Data Analysis (EDA) performed on a healthcare dataset using Tableau. The dataset includes patient-related information across various hospitals, providing insights into factors such as patient demographics, medical conditions, medications, and hospital admissions. The key attributes in this dataset include:

* Admission Type
* Age
* Medical Condition
* Medication
* Hospital
* Insurance Provider
* Gender
* Date of Admission
* Discharge Date

The objective of this analysis is to explore and uncover patterns in the data that will inform the development of machine learning models in the future. Specifically, the EDA seeks to understand demographic trends, correlations between medical conditions and treatments, and hospital performance metrics. These insights will help in feature engineering for predictive models, such as predicting patient outcomes based on treatments, and optimizing hospital resource allocation.

## ****Visualization Process****

To extract meaningful insights from the data, various visualizations were created using Tableau. Each visualization was chosen with a specific analysis goal in mind.

### 1. ****Age Distribution (Histogram)****

* **Objective**: To understand the distribution of patient ages and identify any key demographic trends.
* **Chart Type**: Histogram
* **Steps**: A histogram was created by first converting the Age field into bins, grouping ages into 10-year intervals. This allowed us to see the distribution of patients across different age ranges.
* **Rationale**: A histogram is ideal for visualizing continuous variables like age, and the bins help reveal clusters or trends within the data.

### 2. ****Admission Type (Pie Chart)****

* **Objective**: To assess the distribution of different types of admissions, such as emergency, scheduled, and walk-in.
* **Chart Type**: Pie Chart
* **Steps**: A pie chart was used to visualize the percentage breakdown of each admission type. The Admission Type field was dragged to the **Color** shelf to differentiate categories.
* **Rationale**: Pie charts are effective for showing proportions, making it easy to compare the frequency of each admission type.

### 3. ****Blood Type Distribution (Bar Chart)****

* **Objective**: To analyze the distribution of different blood types among patients.
* **Chart Type**: Bar Chart
* **Steps**: A bar chart was used to count the occurrences of each blood type by dragging Blood Type to the **Rows** shelf and using the **Count** function.
* **Rationale**: A bar chart provides a clear view of categorical distribution, making it easy to compare the different blood types.

### 4. ****Correlation Between Medication and Medical Condition (Heat Map)****

* **Objective**: To explore the relationship between medical conditions and the medications prescribed.
* **Chart Type**: Heat Map
* **Steps**: A heat map was created by dragging Medical Condition to the **Rows** shelf and Medication to the **Columns** shelf. The number of patients receiving each medication for a specific condition was used as the measure, and color intensity was used to reflect the frequency.
* **Rationale**: Heat maps are highly effective for showing correlations between two categorical variables, allowing us to quickly identify patterns in treatment methods.

### 5. ****Hospital Statistics (Bar Chart)****

* **Objective**:
* To compare hospital performance based on the number of patients treated.
* **Chart Type**: Bar Chart
* **Steps**: A bar chart was created by dragging the Hospital field to the **Rows** shelf and using the **Count** function to determine how many patients were treated at each hospital.
* **Rationale**: Bar charts are well-suited for comparing categories, making them ideal for visualizing hospital performance.

## ****Decision-Making Justification****

The design of the dashboard was focused on clarity, interactivity, and ease of use. Each chart type was chosen to maximize understanding of the dataset:

* **Histogram**:
* This chart was used to capture the overall distribution of patient age, a key demographic variable that affects medical outcomes.
* **Pie Chart**:
* The pie chart for admission types provides a quick glance at how patients enter hospitals, offering insights that could be relevant for resource planning and patient flow optimization.
* **Heat Map**:
* The heat map was instrumental in finding patterns between medications and medical conditions. The use of color gradients helps in highlighting the most common treatment combinations.
* **Bar Charts**:
* These were chosen for categorical comparisons, such as hospital performance and blood type distribution, ensuring clear visibility of differences.

Interactive filters for gender, hospital, and medical condition were added to provide users with the ability to drill down into specific segments of the data, enhancing exploration capabilities.

## ****Challenges and Solutions****

Several challenges were encountered during the EDA process:

**1:Age Field Binning**:

Initially, Tableau did not recognize the Age field as a numerical value, preventing the creation of bins for the histogram. This was resolved by converting the Age field from a dimension to a measure, allowing appropriate age bins to be generated.

**2:Handling Missing Data**:

Some records had missing values for key fields such as Medication or Medical Condition. To ensure the integrity of the visualizations, filters were applied to exclude incomplete records from certain analyses without affecting the overall dashboard's quality

3:**Correlating Medications and Conditions**:

Visualizing the relationship between medications and medical conditions was complex due to the categorical nature of both fields. Using a heat map allowed for an intuitive and visually appealing solution to this challenge.

## ****Conclusion****

This EDA provided key insights into patient demographics, hospital performance, and treatment patterns:

* **Age Distribution**: Patients aged 40-60 were the most commonly admitted group, suggesting this age range could be a critical demographic for certain medical conditions.
* **Admission Types**: Emergency admissions accounted for the majority, highlighting the importance of resource allocation for unplanned hospital visits.
* **Medication and Medical Condition Correlation**: The heat map revealed strong correlations between certain conditions and specific medications, which can be used to build predictive models for treatment effectiveness.

These insights will be crucial for developing machine learning models in the future. For example, age and medical condition could be strong predictors of patient outcomes, and the patterns in medication use could inform models aimed at optimizing treatment plans.

The interactive dashboard, published to Tableau Public, will serve as a valuable tool for healthcare professionals and data analysts, allowing them to explore the data and generate hypotheses for further analysis.

### ****Tableau Public Link:****

**https://public.tableau.com/views/InteractiveHealthcareDashboard/InteractiveHealthcareDashboard?:language=en-US&publish=yes&:sid=&:redirect=auth&:display\_count=n&:origin=viz\_share\_link**