# ****-:Task 01 Report:-****

**Data Visualization using Bar Chart and Histogram**  
**Data Science Project Report**

### ****Submitted By:****

Naima Shoukat

### ****Submitted To:****

### **Prodigy ImfoTech**

### ****Course:****

Data Science

### ****Date:****

08 August, 2025

### 📑 ****Table of Contents:****

1. Introduction
2. Problem Statement
3. Tools and Technologies
4. Methodology
   * Data Collection
   * Data Exploration
   * Data Cleaning
   * Visualization (Bar Chart & Histogram)
5. Results and Interpretation
6. Conclusion

## 1️⃣ Introduction:

Data visualization is one of the most powerful techniques in data analysis, allowing insights to be extracted at a glance. In this task, we explore and visualize the distribution of **categorical** and **continuous** variables — specifically, **Gender** and **Age** — using **bar charts** and **histograms** respectively. These visualizations are useful in summarizing patterns and characteristics of a dataset.

## 2️⃣ Problem Statement:

The objective of this task is to create **a bar chart or histogram** to visualize the distribution of variables such as **gender** (categorical) or **age** (continuous) in a population dataset. This analysis helps understand the demographic composition and trends within the dataset.

## 3️⃣ Tools and Technologies:

* **Programming Language:** Python
* **Libraries:** pandas, matplotlib, seaborn
* **Platform:** Jupyter Notebook
* **Visualization Techniques:**
  + Bar Chart (for Gender distribution)
  + Histogram (for Age distribution)
* **Data Source:** Titanic dataset (public dataset via GitHub)

## 4️⃣ Methodology:

### 🔹 4.1 Data Collection:

* Dataset Used: Titanic Dataset
* Source: GitHub Dataset Link
* Key Features:
  + Sex (categorical: male/female)
  + Age (continuous)

### 🔹 4.2 Data Exploration:

* Displayed first few rows using .head()
* Checked dataset info, column names, and missing values using .info()
* Summary statistics obtained via .describe()

### 🔹 4.3 Data Cleaning:

* Dropped rows with missing values in the Age column
* Retained columns: Sex, Age
* No encoding was necessary for Sex as it was directly usable for visualization

### 🔹 4.4 Visualization (Bar Chart & Histogram):

## 5️⃣ Results and Interpretation:

| **Visualization Type** | **Variable** | **Key Insight** |
| --- | --- | --- |
| Bar Chart | Sex | The dataset contains more male passengers than female passengers. |
| Histogram | Age | Majority of passengers were aged between 20 and 40 years. The KDE line shows the age distribution density. |

Both charts clearly illustrate the demographic distribution of the population, which is essential for any population-based analysis or model training.

## 6️⃣ Results and Conclusion:

✅ **Observations:**

* Clear gender imbalance (more males than females).
* Passengers were mostly young adults (20–40 age range).
* Useful for demographic modeling and further supervised learning tasks.

⚠️ **Limitations:**

* Dataset is not representative of the general population (limited to Titanic passengers).
* Missing values in Age column reduced total sample size.
* Further insights could be gained by segmenting data by class or fare categories.

## 🔄 Process Flowchart:

┌────────────────────────────┐

│ Load Titanic Dataset │

└────────────┬───────────────┘

↓

┌────────────────────────────┐

│ Explore Dataset (head/info)│

└────────────┬───────────────┘

↓

┌────────────────────────────┐

│ Clean Data (drop NA in Age)│

└────────────┬───────────────┘

↓

┌────────────────────────────┐

│ Bar Chart (Gender) │

└────────────┬───────────────┘

↓

┌────────────────────────────┐

│ Histogram (Age) │

└────────────┬───────────────┘

↓

┌────────────────────────────┐

│ Analyze & Interpret Charts │

└────────────────────────────┘