Data Analysis using Titanic Dataset

Load the Titanic dataset(https://www.kaggle.com/competitions/titanic/data)

Check the details of the features

Count null values and take the necessary measures to impute the missing values.

Check descriptive statistics of the dataset

Outlier Detection and Handling

Univariate Analysis: Exploring One Feature at a Time

Univariate analysis means looking at **one column (feature)** at a time to understand how its values are spread out.

Age

- What is the **age distribution** of passengers?
- Are there more **young people**, **adults**, or **elderly** on board?
- Do I see any **very young children or babies**?
- Is age spread out evenly or are there certain age groups with more passengers?

Fare

Understanding how much passengers paid for their tickets can tell us a lot about **social class, accessibility**, and even potential influence on **survival chances**.

What is the **fare distribution**?

- Are most ticket prices **low, medium, or high**?
- Are there any **extremely expensive tickets** (outliers)?
- Is the fare data **skewed** (most people paid low, but a few paid very high)?

Embarkation Analysis

Understand where passengers boarded the Titanic. This information is stored in the Embarked column, which includes three possible ports:

- **C**: Cherbourg
- **Q**: Queenstown
- **S**: Southampton

To explore this column

- How many passengers **embarked** from each port (C, Q, or S)?
- Which embarkation point was **most common**?
- Are there any **missing values** in this column?

Passenger Gender

Understand gender distribution.

- How many passengers were **male** and how many were **female**?
- Is there a **gender imbalance** on the ship?

Passenger Class (Pclass)

Explore passengerclass.

- How many passengers were in **1st**, **2nd**, and **3rd** class?
- Which class had the **highest number of passengers**?

Bivariate Analysis

explore how **two features relate to each other**. This is called **bivariate analysis**.

- Do passengers in higher classes pay more?
- Are younger passengers more likely to survive?
- Does where someone boarded the ship relate to their chance of survival?

Multivariate Analysis

- How do **Pclass**, **Age**, and **Fare** work together to affect survival?
- Are survival rates different depending on **Embarked location and Pclass**?

- Understand the **overall distribution** of this variable
- See if the dataset is **balanced or imbalanced**
- Explore which **other features (like sex, class, or embarkation)** may have influenced survival ## Countplot of Survival Distribution