

1: Title Slide

Title: Titanic Dataset - Data Cleaning & Exploratory Data Analysis (EDA)

Subtitle: Using the famous Kaggle Titanic dataset

2: Objective

Explore the Titanic dataset to understand key patterns and relationships, such as:

- * How age, gender, or class affect survival
- * Visualizing and interpreting survival trends

3: Step 1 – Import Libraries

```
python
```

```
import pandas as pd
```

```
import numpy as np
```

```
import seaborn as sns
```

```
import matplotlib.pyplot as plt
```

Dhruvik mori [ProdigyInfotech]

4: Step 2 – Load Dataset

python

```
df = pd.read_csv('titanic.csv')
```

5: Step 3 – Check Missing Values

python

```
df.isnull().sum()
```

Common missing columns:

- * Age

- * Cabin

- * Embarked

6: Step 3 – Handle Missing Values

python

```
df['Age'].fillna(df['Age'].median(), inplace=True)
```

```
df.drop('Cabin', axis=1, inplace=True)
```

```
df['Embarked'].fillna(df['Embarked'].mode()[0], inplace=True)
```

7: Step 4 – Exploratory Data Analysis (EDA)

We will now explore:

- * Overall survival

- * Impact of sex, class, age, fare on survival

8: Survival Count

python

```
sns.countplot(x='Survived', data=df)
```

```
plt.title('Survival Count')
```

Interpretation: 0 = Died, 1 = Survived

9: Survival by Gender

python

```
sns.countplot(x='Survived', hue='Sex', data=df)
```

```
plt.title('Survival by Gender')
```

Observation: Women had much higher survival rates.

10: Survival by Class

python

```
sns.countplot(x='Survived', hue='Pclass', data=df)
```

```
plt.title('Survival by Passenger Class')
```

Observation: 1st class had better survival odds.

11: Age Distribution

python

```
sns.histplot(df['Age'], bins=30)
```

```
plt.title('Age Distribution')
```

12: Survival by Age

python

```
sns.histplot(data=df, x='Age', hue='Survived', bins=30, kde=True)
```

```
plt.title('Survival by Age')
```

13: Fare vs Survival

python

```
sns.boxplot(x='Survived', y='Fare', data=df)
```

```
plt.title('Fare vs Survival')
```

14: Key Observations

- * Women survived more than men
- * 1st class passengers had better chances
- * Children had a higher survival rate
- * Higher fare = better chance of survival

Dhruvik mori [ProdigyInfotech]

15: Save Cleaned Dataset

python

```
df.to_csv('titanic_cleaned.csv', index=False)
```