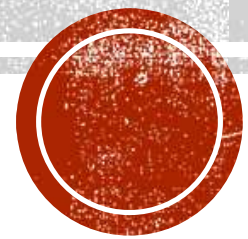
EXPLORATORY DATA ANALYSIS ON TITANIC DATASET

DATA CLEANING, VISUALIZATION, INSIGHT EXTRACTION



Presented by:
VATHALUR.PRASAD

OBJECTIVE:

- Perform Exploratory Data Analysis (EDA).
- Extract meaningful insights using visual and statistical exploration.
- Tools used: Python (Pandas, Matplotlib, Seaborn)

TOOLS & LIBRARIES USED:

- **Pandas** — Data loading, manipulation
- **Numpy** — Numerical operations
- **Matplotlib, Seaborn** — Visualization

DATASET OVERVIEW:

- Dataset: Titanic Test Dataset
- Loaded using Pandas(read_csv)
- Inspected using .info(),.describe(),.value_counts()

DATA CLEANING:

- Dropped the **Cabin** column (many missing values)
- Handled missing values:
- Filled **Age** with median
- Filled **Fare** with median

EXPLORATORY DATA ANALYSIS (EDA):

- Displayed dataset (head)
- Checked missing data
- Checked missing data

VISUALIZATIONS:

- Correlation Heatmap (sns.heatmap())
- Pairplot if applicable (sns.pairplot())
- Histograms, Boxplots (if you add them)

INSIGHTS & OBSERVATIONS:

- Findings from EDA
- Correlations observed
- Trends and anomalies detected

INTERVIEW QUESTIONS (PREPARATION):

- Importance of EDA
- Plots to check correlation (heatmap, pairplot)
- Handling skewed data
- Detecting multicollinearity
- Difference between heatmap and pairplot

CONCLUSION:

- Dataset cleaned and visualized successfully.
- Ready for further modeling or prediction.