

S3 MCA (2024-26 Batch)

M24CA1L306 Data Science Lab

Lab Exercise #03

Date of Submission: 12 Sept 2025

- Download the dataset **titanic.csv**
- Using Python, visualize the various characteristics of the dataset.
- Use Google Colab for implementation.
- **Define your own conditions to learn the characteristics**
- **After each visualization write your findings**
- EDA Questions for Students
 1. **What percentage of passengers survived?**
 - Plot survival counts using bar charts.
 2. **Did gender influence survival chances?**
 - Compare survival counts between males and females.
 3. **How did passenger class (Pclass) affect survival?**
 - Use stacked barplots or grouped survival rates by class.
 4. **Was age a factor in survival?**
 - Plot survival by age distribution (histogram or KDE plots).
 5. **Which age group had the highest survival rate? (children, youth, adults, elderly)**
 - Create age bins and compare survival proportions.
 6. **Did family size (SibSp + Parch) affect survival chances?**
 - Create a new feature family_size and compare survival across groups.
 7. **Did the embarkation port influence survival?**
 - Compare survival rates among passengers from ports C, Q, and S.
 8. **Did fare have any relationship with survival?**
 - Compare survival against fare distributions using boxplots.
 9. **Which combination of factors best explains survival? (Gender + Class + Age)**
 - Use grouped barplots or FacetGrid plots.
 10. **Were there passengers with missing data, and how might this affect analysis?**
 - Inspect null values in dataset and discuss missingness.

- Use the following types of charts for visualization:
 1. Histogram
 2. Quartile (Box) plot
 3. Distribution chart
 4. Scatterplot
 5. Scatter multiple
 6. Scatter matrix
 7. Bubble chart
 8. Density chart
 9. Parallel chart
 10. Deviation chart
 11. Andrews curves