# Titanic EDA Task 5

## Overview

This project is part of the Data Analyst Internship Task 5, focusing on Exploratory Data Analysis (EDA) using the Titanic dataset.

We performed statistical and visual exploration to uncover patterns, trends, and anomalies in the data.

---

## Dataset

- \*\*Source\*\*: [Titanic Dataset](https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv)

- \*\*Description\*\*: Information about Titanic passengers including features like age, gender, class, fare, and survival status.

---

## Tools Used

- Python 3

- Pandas

- Matplotlib

- Seaborn

- Jupyter Notebook

---

## Steps Performed

1. \*\*Data Loading\*\*: Loaded the Titanic dataset.

2. \*\*Initial Inspection\*\*: Used `.info()`, `.describe()`, and `.isnull().sum()` to understand data structure and missing values.

3. \*\*Missing Value Handling\*\*:

- Filled missing 'Age' with median value.

- Filled missing 'Embarked' with mode.

- Dropped the 'Cabin' column due to too many missing values.

4. \*\*Univariate Analysis\*\*:

- Plotted distributions for Age and Survived.

5. \*\*Bivariate Analysis\*\*:

- Analyzed survival rate by Sex and Passenger Class.

6. \*\*Correlation Analysis\*\*:

- Created heatmap and pairplots for better feature understanding.

7. \*\*Observations\*\*:

- Females had higher survival rates.

- First-class passengers survived more often.

- Fare and survival rate were positively correlated.

---

## How to Run

- Download the `Titanic\_EDA\_Task5.ipynb` file.

- Open it in Jupyter Notebook, VS Code, or Google Colab.

- Run all cells to reproduce the results.

- Export to PDF if needed.

---

## Deliverables

- Jupyter Notebook: `Titanic\_EDA\_Task5.ipynb`

- (Optional) PDF Report (can be generated by saving the notebook as PDF)

Author

Barjinder Singh

Thank You