

# Titanic Data Analytics Challenge – Day 10 Final Documentation

## 1. Project Overview

This project is the final integration of the Python Data Analytics Coding Challenge using the Titanic dataset. The goal is to perform a complete end-to-end data analysis process that includes data cleaning, exploration, visualization, and insight generation. The Titanic dataset records information about passengers aboard the RMS Titanic, including their demographics, travel class, and survival status.

## 2. Workflow Steps

The project was executed following the below workflow:

1. **Data Import:** Loaded the Titanic dataset using pandas.
2. **Data Cleaning:** Handled missing values by replacing numeric ones with median and categorical with mode; removed duplicates.
3. **Exploration:** Used `head()`, `info()`, `describe()` to understand data structure and types.
4. **Visualization:** Created multiple plots using Matplotlib and Seaborn to explore trends, correlations, and distributions.
5. **Statistical Analysis:** Computed mean, median, and correlation between numeric features.
6. **Insights & Recommendations:** Derived key findings regarding survival trends and contributing factors.

## 3. Key Formulas and Functions

- `pandas.read_csv()` – Load dataset.
- `DataFrame.fillna()` – Handle missing values.
- `DataFrame.describe()` – Statistical summary.
- `groupby() + mean()` – Aggregation for insights.
- `matplotlib & seaborn` – Data visualization tools.
- `corr()` – Correlation matrix for numeric columns.

## 4. Visual Insights Summary

- The survival rate for `females` was significantly higher than for males.
- Passengers in `1st` class had a better chance of survival compared to lower classes.
- There is a positive correlation between `fare` and `survival`, implying wealthier passengers were more likely to survive.
- The `age distribution` revealed most passengers were between `20–40 years old`.
- The heatmap highlighted strong relationships between fare, passenger class, and survival.

## 5. Insights and Recommendations

- **Key Insights:**
  - 1. Higher-class passengers and women had better survival odds.
  - 2. There is a clear socioeconomic disparity visible in the data.
  - 3. Fare and class strongly influence survival probabilities.
- **Recommendations:**
  - Ensure equitable access to safety resources across all classes.
  - Maintain detailed demographic and class-based travel data for future planning.
  - Use similar analysis in transport safety studies to identify risk factors.

## 6. Learnings

This project reinforced essential data analytics concepts including:

- Efficient handling of missing and inconsistent data.
- Use of visualization to uncover hidden insights.
- Importance of statistical measures and correlations.
- Building a structured and reproducible data workflow in Python.