

TITANIC SURVIVAL ANALYSIS

1. Project Overview

- **Objective:** To identify the key factors that influenced survival rates on the Titanic and provide actionable insights through a multi-tool analytical pipeline.
- **Problem Statement:** Moving from raw, messy data to a structured database and visual dashboard to tell the story of "Women and Children First."

2. Technical Stack

- **Programming:** Python (Pandas, NumPy)
- **Database:** PostgreSQL (SQLAlchemy for ETL)
- **Visualization:** Matplotlib, Seaborn, Power BI
- **Environment:** Jupyter Notebook, pgAdmin4

3. The Analytical Pipeline

1. **ETL (Extract, Transform, Load):** * Extracted Titles from names to handle missing Age values accurately.
 - Simplified complex columns (Cabin → Has_Cabin).
 - Engineered Family_Size to determine the impact of social structures on survival.
2. **Database Management:**
 - Automated the upload of cleaned data to PostgreSQL.
 - Executed complex SQL queries to bucket passengers into logic-based categories (Fare buckets, Age stages).
3. **Visualization:**
 - Developed a high-fidelity Power BI dashboard for dynamic filtering by Class and Gender.

4. Top 3 Key Insights

- **The Gender Gap:** Single women had a survival rate of **74%**, while single men dropped to **18.89%**, validating the "**Women and Children First**" protocol.
- **Socio-Economic Priority:** 1st Class passengers had a **62.96%** survival rate, nearly triple that of 3rd Class passengers (**24.24%**).
- **The "Alone" Penalty:** Passengers traveling with family survived at a rate of **50.56%**, compared to only **30.35%** for those traveling alone.

5. Final Conclusion

Survival on the Titanic was not random; it was a function of social class, gender, and family support systems. This project demonstrates the ability to handle the full lifecycle of data—from cleaning and engineering to SQL querying and executive dashboarding.