# Project Proposal: Titanic Survival Prediction

## 1. Project Title

Analyzing and Predicting Survival on the Titanic

## 2. Background & Motivation

The sinking of the RMS Titanic in 1912 is one of the most tragic maritime disasters in history. Out of over 2,200 passengers and crew, only about 710 survived. The Titanic dataset provides passenger details such as age, gender, ticket class, and survival status, making it an excellent dataset to practice data cleaning, exploratory analysis, and predictive modeling.

## 3. Objectives

- Explore and clean the Titanic dataset for analysis.  
- Perform Exploratory Data Analysis (EDA) and identify patterns in survival.  
- Determine the relationship between passenger attributes (age, gender, class, etc.) and survival.  
- Build a machine learning model that predicts passenger survival.  
- Document insights and make the project reproducible for others.

## 4. Research Questions

1. What proportion of passengers survived overall?  
2. Did gender influence survival rates?  
3. How did passenger class (Pclass) affect survival?  
4. Were children more likely to survive than adults?  
5. Did traveling with family impact survival chances?  
6. Did ticket fare or embarkation port influence survival?  
7. Which features are most important for predicting survival?

## 5. Methodology

1. Data Collection – Use Titanic dataset from Kaggle.  
2. Data Cleaning – Handle missing values and convert categorical variables.  
3. Exploratory Data Analysis (EDA) – Visualize survival patterns.  
4. Feature Engineering – Create new features like FamilySize, IsAlone, Title.  
5. Modeling – Logistic Regression, Decision Tree, Random Forest.  
6. Documentation – Summarize results and share on GitHub.

## 6. Expected Outcomes

- Identification of key factors that influenced survival.  
- Cleaned and structured dataset ready for analysis.  
- Visualizations explaining survival patterns.  
- Predictive model with ~75–80% accuracy.  
- GitHub repository with notebooks, scripts, and documentation.

## 7. Tools & Technologies

- Programming: Python  
- Libraries: pandas, numpy, matplotlib, seaborn, scikit-learn  
- Environment: Jupyter Notebook  
- Version Control: GitHub

## 8. Timeline

Week 1: Data cleaning, EDA, feature engineering.  
Week 2: Modeling, evaluation, documentation, GitHub upload.