# AI Data Preparation Challenge

## From Raw Data to AI-Ready Data

### Objective

Transform the Titanic dataset from raw, messy data into a **clean, structured format** that's ready for machine learning models.

*No models will be built in this challenge — focus purely on data preparation!*

### Dataset

We'll use the classic **Titanic dataset** from Kaggle:

https://www.kaggle.com/c/titanic/data

The file you need is `train.csv` — it contains passenger information and survival labels, with real-world complexities perfect for learning.

# Challenge Tasks

## 1. Load & First Look

- Load `train.csv` using pandas
- Display the first 10 rows
- Show dataset shape and column information (data types)

## 2. Basic Exploration

- Compute/Show descriptive statistics (mean, min, max)

- Show value counts for categorical columns
- Create at least one meaningful new column

## 3. Visual Analysis

- Make the necessary plots

## 4. Handle Missing Data

- Identify which columns have missing values and how many
- Apply appropriate imputation (Filling the missing data places)

## 5. Convert Categorical Data

## 6. Feature Engineering & Selection

- Drop unnecessary columns

# Deliverable

**One Jupyter Notebook** containing:

1. All code for the tasks above
2. Clear comments explaining each step
3. Visualizations
4. A **short written summary** (2-3 sentences) answering:

*"What key steps did you apply to make the Titanic data AI-ready?"*

**Remember:** Good data preparation is the foundation of successful AI projects. Take your time and understand why each step matters!