**🚢 Titanic ML Project Pipeline (Step by Step)**

**1. Exploratory Data Analysis (EDA)**

* Inspect dataset: head(), info(), describe()
* Check for missing values & class balance
* Visualize survival rate across:
  + Gender (Sex)
  + Ticket class (Pclass)
  + Age groups (binning ages)
  + Embarked location (Embarked)

**2. Data Preprocessing**

* Handle missing values:
  + Age: fill with median or predictive model
  + Embarked: fill with mode
  + Cabin: drop or extract first letter
* Encode categorical features:
  + Sex, Embarked, Pclass → one-hot encoding
* Feature engineering:
  + Extract Title from Name
  + Create FamilySize (SibSp + Parch + 1)
  + Bin Age and Fare into categories

**3. Model Building**

* Define features X and target y
* Split into train/test (using provided train.csv internally with train\_test\_split)
* Try different models:
  + Logistic Regression (baseline)
  + RandomForest
  + GradientBoosting / XGBoost (stronger)
* Use cross-validation for robustness

**4. Model Evaluation**

* Compare models on:
  + Accuracy
  + Precision, Recall, F1-score
  + Confusion matrix
* Feature importance (for tree-based models)

**5. Hyperparameter Tuning (Bonus)**

* Use GridSearchCV or RandomizedSearchCV
* Optimize RandomForest / XGBoost parameters
* Compare tuned vs baseline models

**6. Final Prediction**

* Train best model on full training data
* Predict on test.csv
* Create submission.csv for Kaggle

!!!! Perfect 😎 — let’s make the barplot **with percentages annotated** on top of each bar:

import seaborn as sns

import matplotlib.pyplot as plt

plt.figure(figsize=(6,4))

ax = sns.countplot(data=train, x="Survived", palette="Set2")

# Annotate percentages on top of bars

total = len(train)

for p in ax.patches:

height = p.get\_height()

percentage = f'{100 \* height / total:.1f}%'

ax.annotate(percentage, (p.get\_x() + p.get\_width() /