📑 Findings Report: Titanic Survival Analysis

1. Missing Data

- **Age** had missing values → Filled using the **median age**.
- **Embarked** had missing entries → Filled with most frequent value **'S' (Southampton)**.
- **Cabin** had too many missing values → **Dropped** from the dataset.

---

2. Key Factors Influencing Survival

- **Gender:**
  - **Females** had a much higher survival rate compared to males.
- **Passenger Class (Pclass):**
  - **1st Class** passengers had the highest survival chances.
  - 3rd Class passengers had the lowest.
- **Family Size:**
  - Passengers traveling **with family** (siblings/spouses/parents/children) had **higher survival rates** compared to those traveling alone.
- **Embarkation Port (Embarked):**
  - Passengers who embarked from **Cherbourg (C)** had better survival rates.

---

3. Feature Engineering Insights

- **New features** like FamilySize and IsAlone were important.
  - Passengers **traveling alone** had **lower chances of survival**.
- Encoding of categorical variables (Sex, Embarked) improved model understanding.

---

4. Modeling and Results

Model	Accuracy	Key Notes
Logistic Regression	~81%	Good baseline model.
Decision Tree Classifier	~78%	Slightly overfit to training data.
Random Forest Classifier	~83%	**Best model**, handles feature interactions well.

- **Random Forest Classifier** had the **highest accuracy** (~83%) and is the recommended model for Titanic survival prediction.

---

5. Conclusion

- **Survival** strongly depended on:
    - **Gender** (Females had priority),
    - **Social Class** (Wealthier passengers survived more),
    - **Travel Group** (Families over solo travelers).
- **Random Forest** can be effectively used for predictive tasks on similar structured datasets.

🚀 **Bonus Recommendation:**

- Try **Hyperparameter Tuning** (Grid Search or Random Search) on the Random Forest to push accuracy even higher.
- Explore **feature importance plots** to visually confirm which features mattered the most.