Titanic Project

Data Exploration & Visualization

1. **How many passengers survived and how many didn't? Plot a bar chart.**
2. **What percentage of males vs females survived? Use a countplot or pie chart.**
3. **Compare survival rate across passenger classes (`Pclass`). Which class had the highest survival rate?**
4. **Plot the age distribution. Then, compare the average age of survivors vs non-survivors.**
5. **Do passengers with family (`SibSp` or `Parch`) have higher survival rates?**
6. **What is the survival rate for passengers who paid more than the average fare?**
7. **Create a new column dividing passengers into "child" (Age < 16), "adult", and "senior" (Age > 60). Which group survived the most?**
8. **Check for missing values in the dataset. How did you handle them before modeling?**
9. **Create a heatmap showing correlation between numerical columns. What insights can you draw?**
10. **Plot survival counts for each Embarked location (`Embarked`). Which location had the highest survival rate?**

Modeling with Random Forest

11. **Train a Random Forest classifier and print the accuracy.**
12. **What happens to the accuracy when you change `n_estimators` (e.g., 10, 100, 500)?**
13. **Use `.predict()` to make predictions for the first 10 rows. Compare with actual survival values.**
14. **Plot a confusion matrix for the Random Forest model. Explain what the matrix tells you.**
15. **List the top 3 most important features according to the Random Forest model.**

16. **Use `cross_val_score` to evaluate model performance. What's the average accuracy across folds?**

17. **Try changing `max_depth` of the Random Forest. How does it affect accuracy and overfitting?**

18. **Compare Random Forest accuracy with a simple Decision Tree. Which performs better and why?**

Critical Thinking & Interpretation

19. **What 3 factors seem to affect survival the most based on your analysis and model?**

20. **If you were going to improve the model, what extra data (features) would you collect and why?**