

# Report on Titanic Dataset Analysis Code

The Titanic dataset analysis code utilizes several key functions from the pandas, matplotlib, and seaborn libraries to preprocess the data and visualize insights regarding passenger survival rates. Here's an explanation of the main functions used in the code:

## Key Functions Explained

### Pandas Functions

1. **pd.read\_csv():** This function is used to load data from a CSV file into a DataFrame. It allows for easy data manipulation and analysis.
2. **df.isnull().sum():** This method checks for missing values in the DataFrame. It returns the count of null values for each column, helping identify which columns require data cleaning.
3. **df['Age'].fillna(df['Age'].median(), inplace=True):** This function fills missing values in the 'Age' column with the median age. The `inplace=True` argument ensures that the changes are made directly to the original DataFrame.
4. **df.drop(columns=['Cabin'], inplace=True):** This function removes the 'Cabin' column from the DataFrame. It is used when a column has too many missing values, which could hinder analysis.
5. **df.dropna(subset=['Embarked'], inplace=True):** This method drops rows that have missing values in the 'Embarked' column. It ensures that only complete cases are analyzed for this categorical variable.
6. **df['Sex'] = df['Sex'].map({'male': 0, 'female': 1}):** This function converts the 'Sex' column from categorical text values ('male' and 'female') to numerical values (0 for male and 1 for female), facilitating analysis and modeling.
7. **pd.get\_dummies(df, columns=['Embarked'], drop\_first=True):** This function performs one-hot encoding on the 'Embarked' column, converting categorical variables into multiple binary columns. The `drop_first=True` argument avoids multicollinearity by dropping the first category.

## Matplotlib Functions

1. **plt.figure(figsize=(8, 5))**: This function creates a new figure for plotting with a specified size, allowing for better control over the appearance of the visualizations.
2. **plt.title()**: This method sets the title for the plot, providing context for the visualization.
3. **plt.xticks()**: This function customizes the x-axis tick labels, making them more interpretable for the audience.
4. **plt.ylabel()**: This method sets the label for the y-axis, clarifying what the values represent.
5. **plt.show()**: This function displays the plot on the screen, allowing users to visualize the results.

## Seaborn Functions

1. **sns.set(style='whitegrid')**: This function sets the aesthetic style of the plots to 'whitegrid', enhancing the visual appeal and readability of the graphs.
2. **sns.barplot()**: This function creates a bar plot, which is used to visualize the average survival rates based on categorical variables such as gender and passenger class. It computes the mean of the 'Survived' column for each category.

## Conclusion

These functions work together to preprocess the Titanic dataset, handle missing data, convert categorical variables, and visualize the survival rates across different demographics. The insights gained from these visualizations can inform further analysis and modeling efforts in understanding the factors that influenced survival on the Titanic.