**Titanic Machine Learning Model Evaluation Report**

**1. Data Overview**

The Titanic dataset was loaded into three Data Frames: df1 for training, df2 for testing, and df3 for gender submission. Initial exploratory data analysis was conducted to understand the dataset's structure.

Training Dataset Shape: (rows, columns): (891, 12)

Testing Dataset Shape: (rows, columns): (418, 11)

Gender Submission Dataset Shape: (rows, columns): (418, 2)

2. Data Exploration and Visualization

**2.1 Overview and statistics**

Summary statistics and data types were examined using the skim function and the info method.

**2.2 Data Visualization**

Statistical and visual analyzes were performed to understand relationships within the dataset.

A correlation plot was generated for numeric features.

Bar plots were used to visualize the relationship between 'Embarked', 'Sex', and 'Survived'.

Count plots displayed gender distribution and survival proportions.

**2.3 Additional analyses**

Additional analyzes included survival comparisons between genders, age distributions, and the impact of the number of parents/children ('Parch') on survival.

**3. Data Preprocessing and Label Encoding**

Categorical variables, 'Sex' and 'Embarked', were label-encoded to prepare the data for machine learning models.

**4. OLS Regression Model for Initial Analysis**

An Ordinary Least Squares (OLS) regression model was fitted to analyze the linear relationships between features and the target variable 'Survived'.

**5. Data Cleaning for Testing Dataset**

Preprocessing steps were applied to the testing dataset (df2) to handle missing values and prepare it for model evaluation.

**6. Support Vector Machine (SVM) Model Evaluation**

A Support Vector Machine (SVM) model was trained on the formatted training data, and its performance was evaluated on the testing data.

**7. Hyperparameter Tuning for SVM**

Grid search and cross-validation were employed to optimize the SVM model by tuning hyperparameters.

8. Logistic Regression Model Evaluation and Tuning

A Logistic Regression model was evaluated, and hyperparameter tuning was performed similarly to the SVM model.

**9. Final Model Building with PyCaret**

PyCaret was utilized to compare and create a final machine learning model.