**Model Summary : Abstract**

**Exploratory Data Analysis on Titanic Dataset**

**1. Introduction**

This project involves performing Exploratory Data Analysis (EDA) on the Titanic dataset. The goal is to uncover underlying patterns and insights by analyzing the data related to passengers on the Titanic. EDA helps in understanding the data better, preparing it for further modeling, and deriving meaningful conclusions about the factors that influenced passenger survival.

**2. Model Preprocessing**

**2.1)Analysis**

* Univariate Analysis: Examines individual variables to understand their distribution, central tendency, and spread.
* Multivariate Analysis: Explores relationships between multiple variables to uncover patterns and correlations within the dataset.

**2.2)Feature Engineering**

* Creating New Columns: Extracted useful information and generated synthetic features to enhance analysis.
* Modifying Existing Columns: Scaled, transformed, or encoded features to make them suitable for analysis and modeling.

**2.3)Handling Outliers**

* Detect Outliers: Used statistical methods and visualizations to identify data points significantly deviating from the expected range.
* Remove Outliers: Cautiously handled outliers by removing, transforming, or imputing them based on the context.

**3. Model Evaluation Results**

**Data Columns Classification**

* Categorical Columns: 'Survived', 'Pclass', 'Sex', 'SibSp', 'Parch', 'Embarked'
* Numerical Columns: 'Age', 'Fare', 'PassengerId'
* Mixed Columns: 'Name', 'Ticket', 'Cabin'

**Missing Values Handling**

* Cabin Column: Dropped due to more than 70% missing values.
* Age Column: Imputed using the mean value.
* Embarked Column: Imputed using the mode ('S').

**Data Type Modifications**

* Converted appropriate columns to categorical data types for better memory usage and analysis suitability.

Univariate Analysis Results

* Survived Column: Visualized survival distribution.
* Pclass Column: Identified the most crowded class.
* Sex Column: Analyzed gender distribution.
* SibSp Column: Explored the distribution of siblings/spouses.
* Parch Column: Analyzed the distribution of parents/children.
* Embarked Column: Examined the distribution of embarkation ports.

Age Column: Analyzed the age distribution and identified it follows a normal distribution for practical purposes.

**4. Conclusion**

* **Female Survival**: Chances of female survival were higher than male survival.
* **Pclass**: Traveling in lower classes (Pclass) was deadliest.
* **Age and Survival**: People aged 20 to 40 had a higher chance of dying.
* **Embarked Port:** Passengers embarking from Cherbourg ('C') had higher survival rates.
* **Family Size and Survival**: People traveling with small families had higher chances of surviving compared to those traveling alone or with large families.