

# Standard DS Template

Classification

John Doe

2023-01-01

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
<b>2</b>	<b>Problem Statement</b>	<b>2</b>
<b>3</b>	<b>Exploratory Data Analysis</b>	<b>2</b>
<b>4</b>	<b>Feature Engineering</b>	<b>4</b>
<b>5</b>	<b>Model Selection</b>	<b>4</b>
<b>6</b>	<b>Model Tuning</b>	<b>4</b>
<b>7</b>	<b>Model Validation and Testing</b>	<b>4</b>
<b>8</b>	<b>Results</b>	<b>4</b>
<b>9</b>	<b>Conclusion</b>	<b>4</b>

# 1 Introduction

Describe the dataset.

# 2 Problem Statement

Describe the problem. What are we trying to predict? Is there a baseline to measure against? Does prediction bring value? “So what?”

# 3 Exploratory Data Analysis

1. Profile the dataset.

Check for correct data types, nulls, uniqueness, and top value counts.

Table 1: Quality Check of All Fields

	Data Type	Mode	Mode % of total	Unique Count	Percent Null
Age	Float	24.0	4%	88	20%
Fare	Float	8.05	5%	248	0%
Survived	Integer	0	62%	2	0%
Pclass	Integer	3	55%	3	0%
SibSp	Integer	0	68%	7	0%
Parch	Integer	0	76%	7	0%
Name	String	Braund, Mr. Owen Harris	0%	891	0%
Sex	String	male	65%	2	0%
Ticket	String	347082	1%	681	0%
Cabin	String	B96 B98	2%	147	77%
Embarked	String	S	72%	3	0%

2) Compute descriptive statistics of numeric fields.

Table 2: Descriptive Statistics of Numeric Fields

	Min	Mean	Median	Max	Standard Dev	Kurtosis
Survived	0	0.4	0.0	1	0.5	-1.8
Pclass	1	2.3	3.0	3	0.8	-1.3
Age	0	29.7	28.0	80	14.5	0.2
SibSp	0	0.5	0.0	8	1.1	17.9
Parch	0	0.4	0.0	6	0.8	9.8
Fare	0	32.2	14.5	512	49.7	33.4

3) Explore dependent variable

Not necessary, as it is binary. Accomplished above.

4) Visualize independent variables.

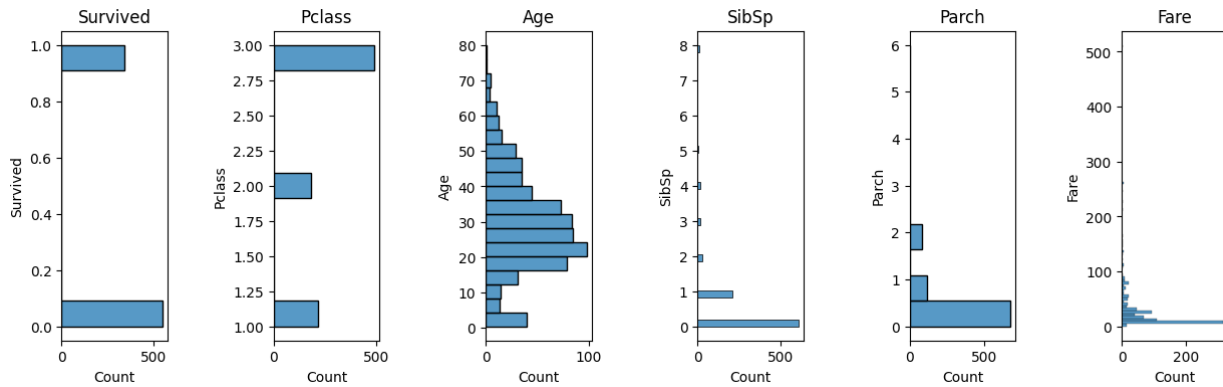


Figure 1: Distribution of Numeric Fields

5) Explore relationship independent variables have on dependent variable.

- Correlations
- Predictive Power Scores

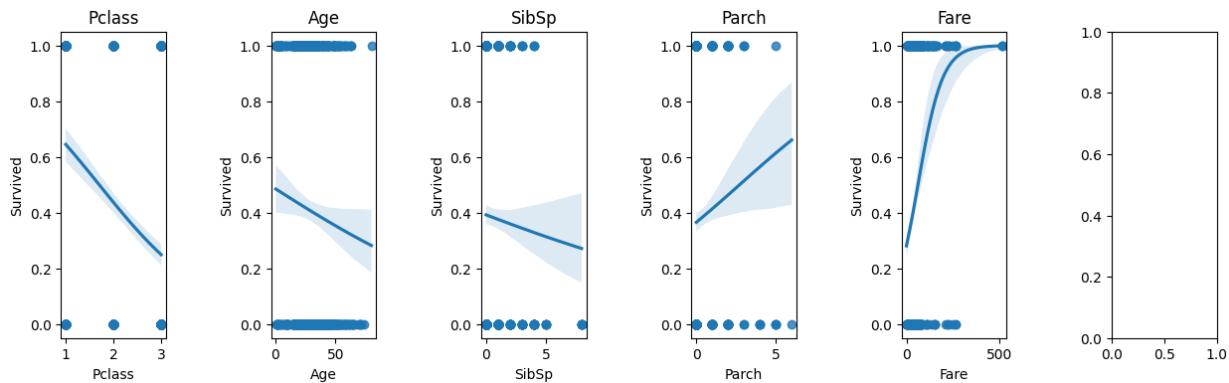


Figure 2: Relationship Between Numeric Fields and Target

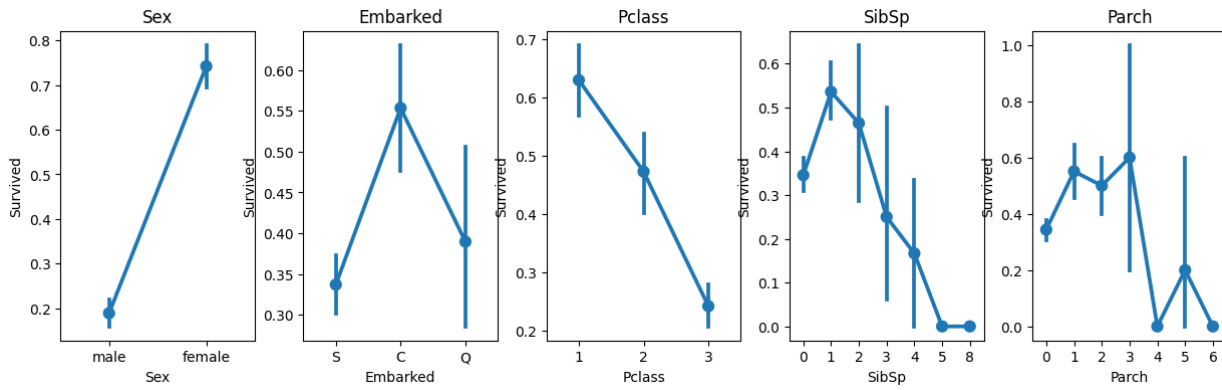


Figure 3: Relationship between Categorical Fields and Target

## 4 Feature Engineering

## 5 Model Selection

## 6 Model Tuning

## 7 Model Validation and Testing

## 8 Results

## 9 Conclusion

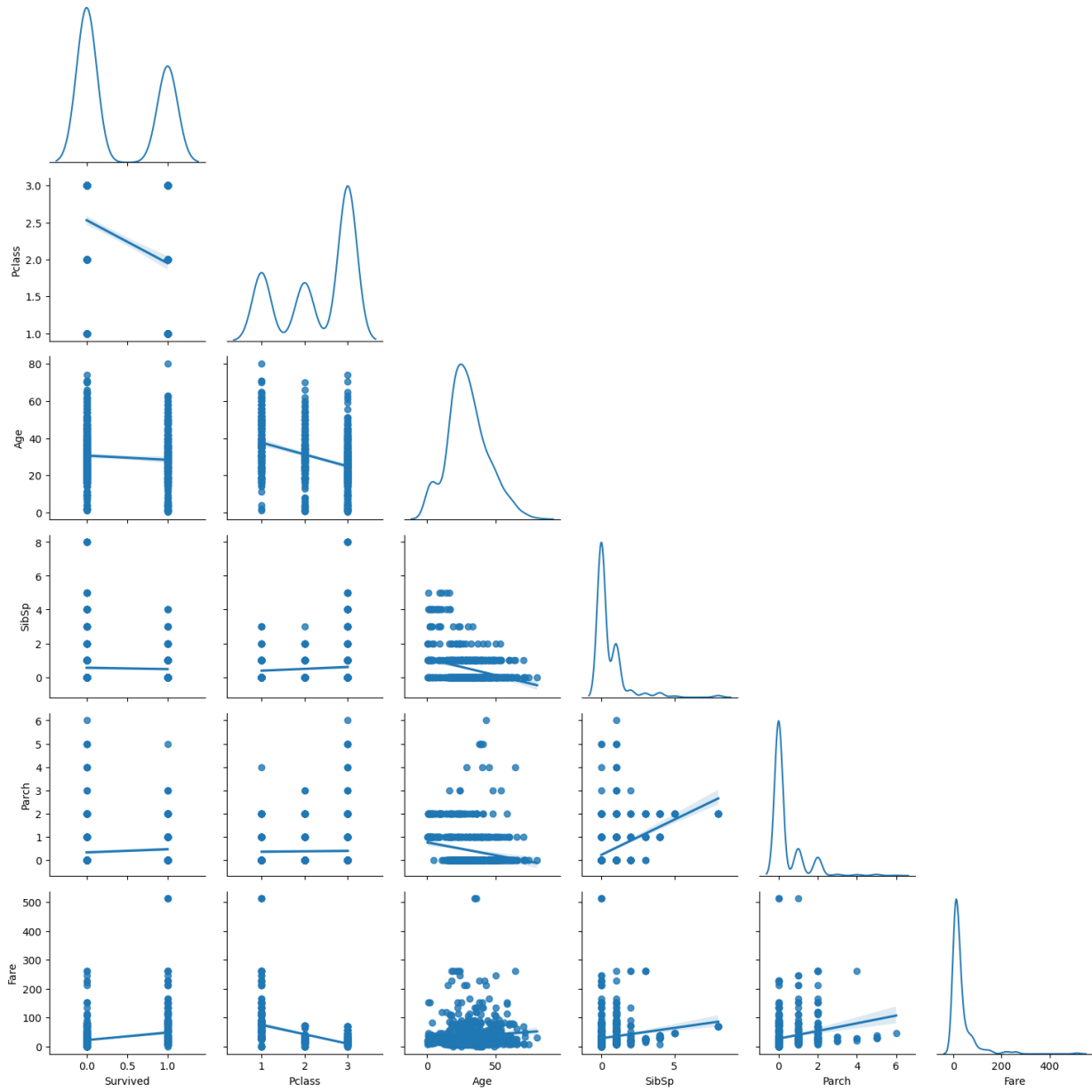


Figure 4: Relationships Between Numeric Fields