

Report for Titanic Dataset

created on 2020-06-11 22:12

Content Table

1 Training Result

1.1 Training Result

- .. 1.1.1 Metric Scores
- .. 1.1.2 Confusion Matrix
- .. 1.1.3 Confidence Distribution
- .. 1.1.4 Reliability Diagram

2 Feature Importance Analysis

2.1 Feature Importance Ranking

3 Model-Agnostic Explainer

3.1 Result

4 Data Statistics Analysis

4.1 Simple Data Statistic

- .. 4.1.1 Data Class (Label) Distribution
- .. 4.1.2 Data Field Attribute
- .. 4.1.3 Data Missing Value Check
- .. 4.1.4 Categorical Field Distribution
- .. 4.1.5 Numerical Field Distribution
- .. 4.1.6 Text Field Distribution

Report for Titanic Dataset

created on 2020-06-11 22:12

1 Training Result

This section summarized the training performance

1.1 Training Result

1.1.1 Metric Scores

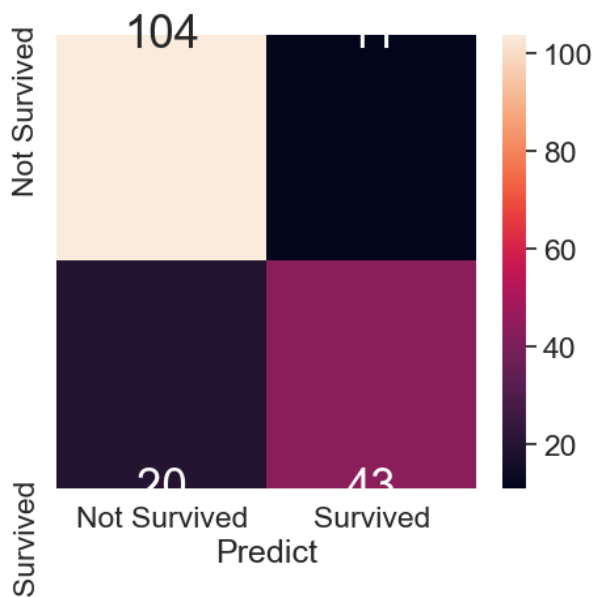
This section shows binary-class evaluation metric.

Overall Result

Split/Round	Testing
Precision	0.7963
Recall	0.6825
F1	0.7350
Accuracy	0.8258
Auc	0.4787

1.1.2 Confusion Matrix

This section shows confusion matrix result.

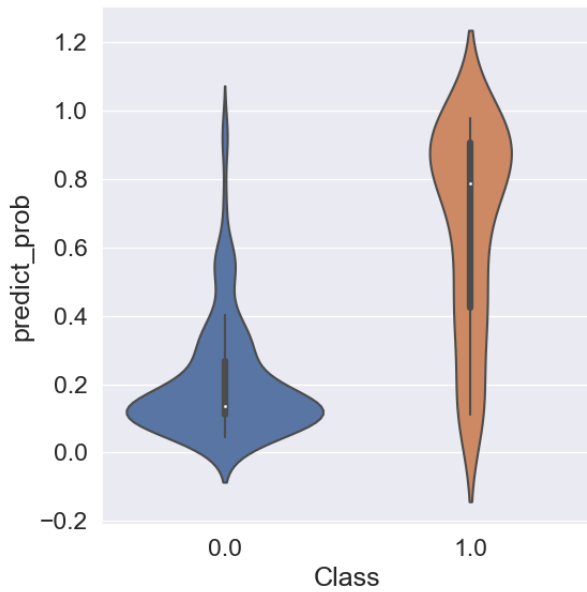


1.1.3 Confidence Distribution

This section shows binary-class confidence distribution result.

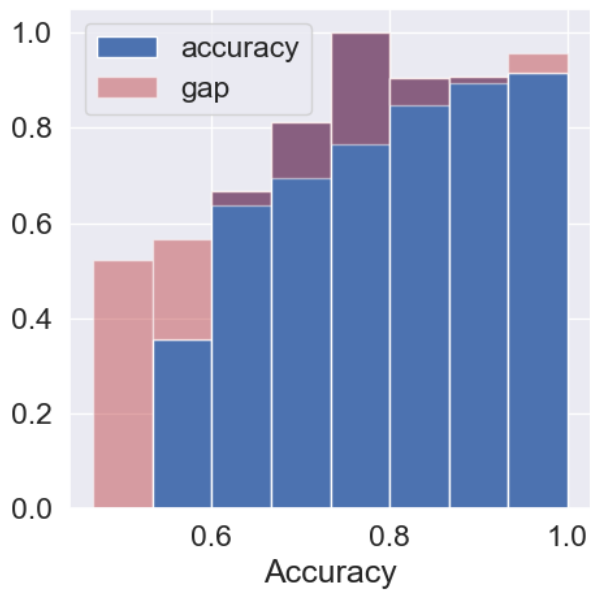
Report for Titanic Dataset

created on 2020-06-11 22:12



1.1.4 Reliability Diagram

This section shows binary-class reliability diagram result.



Report for Titanic Dataset

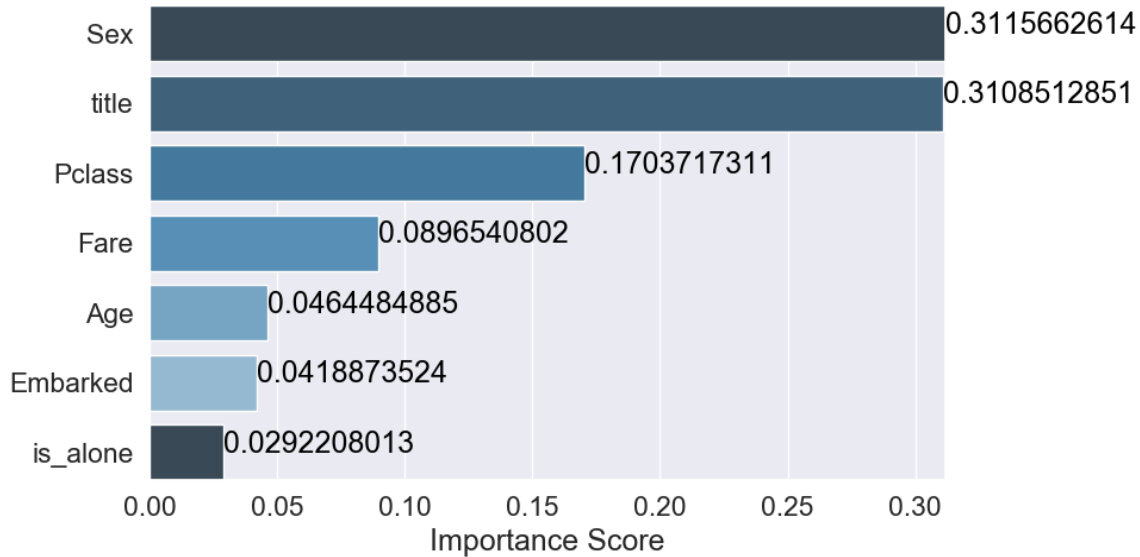
created on 2020-06-11 22:12

2 Feature Importance Analysis

This section provides the analysis on feature

2.1 Feature Importance Ranking

The figure below shows importance ranking for all features from the trained model.



The features which have an importance score larger than 0.005 are listed in the below table.

Feature	Importance
Sex	0.3115662614
title	0.3108512851
Pclass	0.1703717311
Fare	0.0896540802
Age	0.0464484885
Embarked	0.0418873524
is_alone	0.0292208013

Report for Titanic Dataset

created on 2020-06-11 22:12

3 Model-Agnostic Explainer

This section provides a model-agnostic explainer

3.1 Result

The local explainer is generated as `explainer.pkl`

Explainer Configuration

- Domain: **tabular**
- Algorithm: **lime**
- Training data shape: **(712, 7)**
- Number of features in explanations: **5**

Explanation Samples

Example 0:

Class Survived - Confidence: 0.32009357142266565

- Pclass=1: **0.126**
- Fare=higher than 31: **0.058**
- Embarked=S: **-0.049**
- title=Mr: **-0.239**
- Sex=male: **-0.260**

Class Not Survived - Confidence: 0.6799064285773343

- Sex=male: **0.260**
- title=Mr: **0.239**
- Embarked=S: **0.049**
- Fare=higher than 31: **-0.058**
- Pclass=1: **-0.126**

Example 1:

Class Not Survived - Confidence: 0.4364647404698348

- Pclass=3: **0.139**
- Embarked=S: **0.052**
- Fare=lower than 7.91: **0.045**
- title=Miss: **-0.135**
- Sex=female: **-0.245**

Class Survived - Confidence: 0.5635352595301651

- Sex=female: **0.245**
- title=Miss: **0.135**
- Fare=lower than 7.91: **-0.045**
- Embarked=S: **-0.052**
- Pclass=3: **-0.139**

Report for Titanic Dataset

created on 2020-06-11 22:12

4 Data Statistics Analysis

This section provides the analysis on data

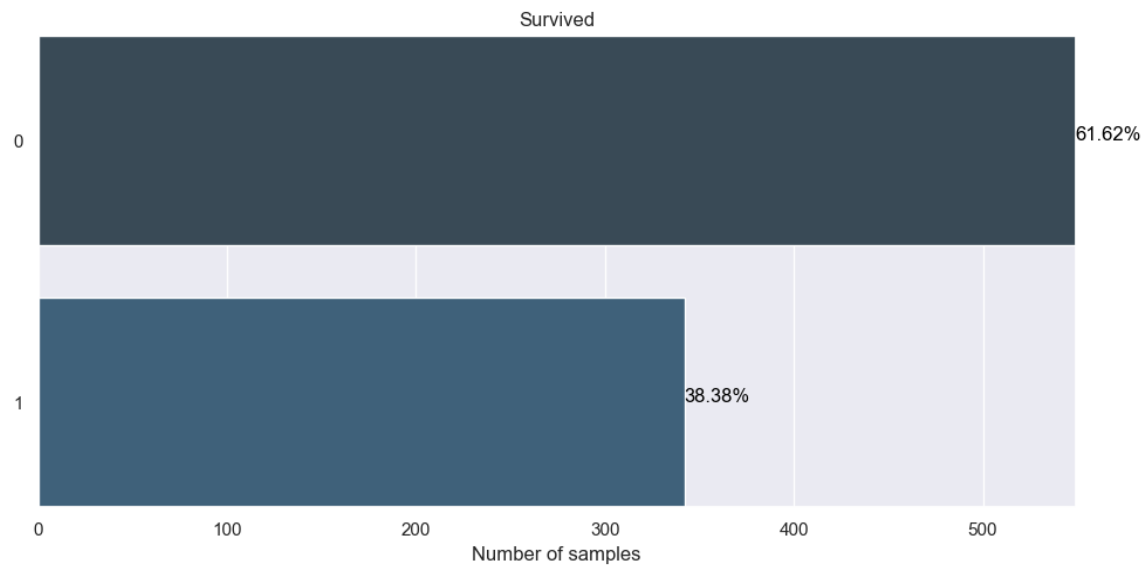
4.1 Simple Data Statistic

4.1.1 Data Class (Label) Distribution

This section shows the distribution of dataset on the label/split.

	Count	Distribution
Survived	891	See below graphs

Distribution for **Survived**



4.1.2 Data Field Attribute

This section shows the attributes of fields in the dataset.

Field Name	Structured	Type	Used
PassengerId	attribute	key	False
Survived	attribute	label	True
Pclass	attribute	categorical	True
Name	attribute	text	True
Sex	attribute	categorical	True
Age	attribute	numerical	True
SibSp	attribute	categorical	True
Parch	attribute	categorical	True
Ticket	attribute	key	False
Fare	attribute	numerical	True
Cabin	attribute	key	False
Embarked	attribute	categorical	True

4.1.3 Data Missing Value Check

This section shows the percentage of sample data with missing values.

Report for Titanic Dataset

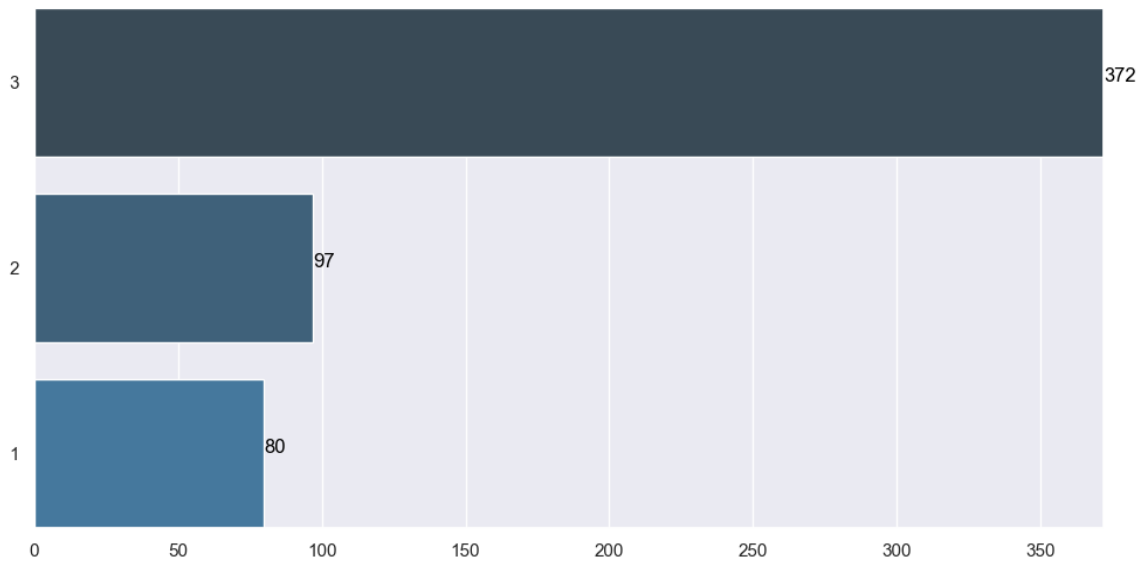
created on 2020-06-11 22:12

Feature	Count	Percentage
Name	0 / 891	0.00%
Sex	0 / 891	0.00%
Fare	0 / 891	0.00%
Embarked	2 / 891	0.00%
SibSp	0 / 891	0.00%
Age	177 / 891	0.20%
Pclass	0 / 891	0.00%
Parch	0 / 891	0.00%

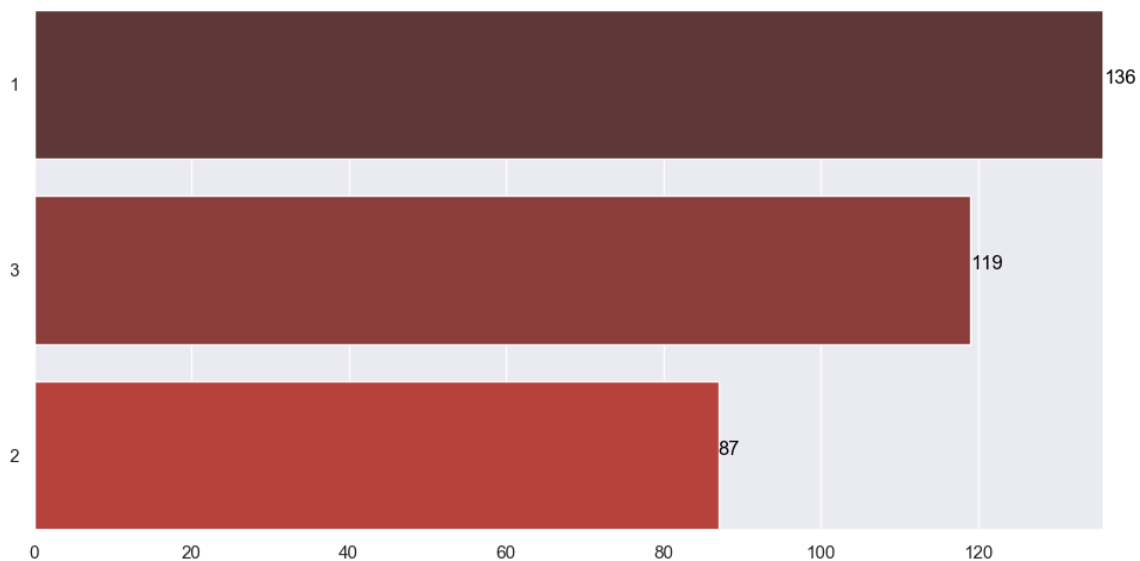
4.1.4 Categorical Field Distribution

Distribution for Pclass

- For 0 samples



- For 1 samples

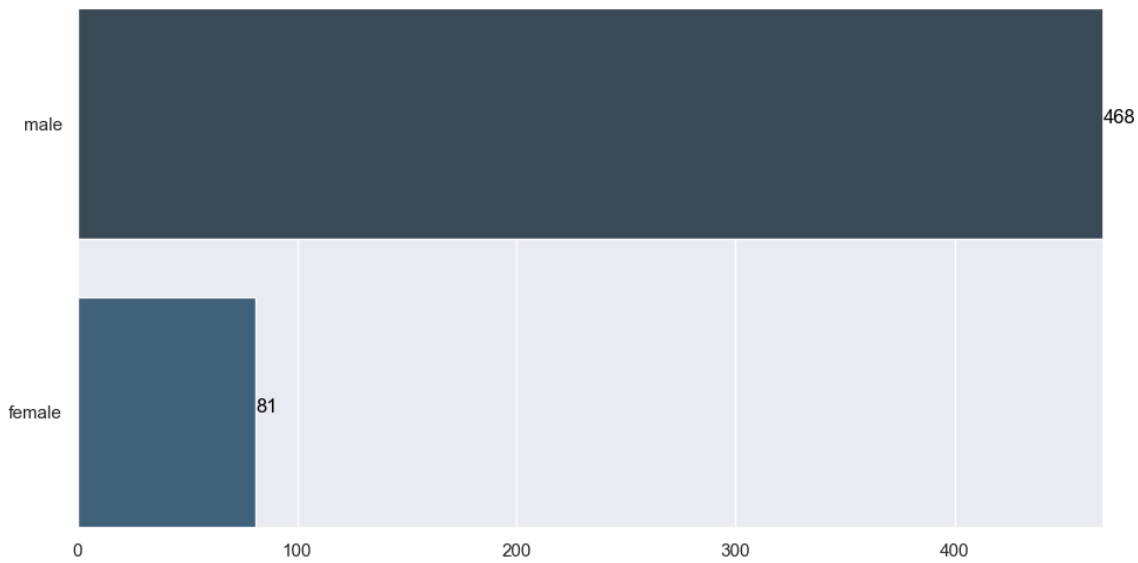


Distribution for Sex

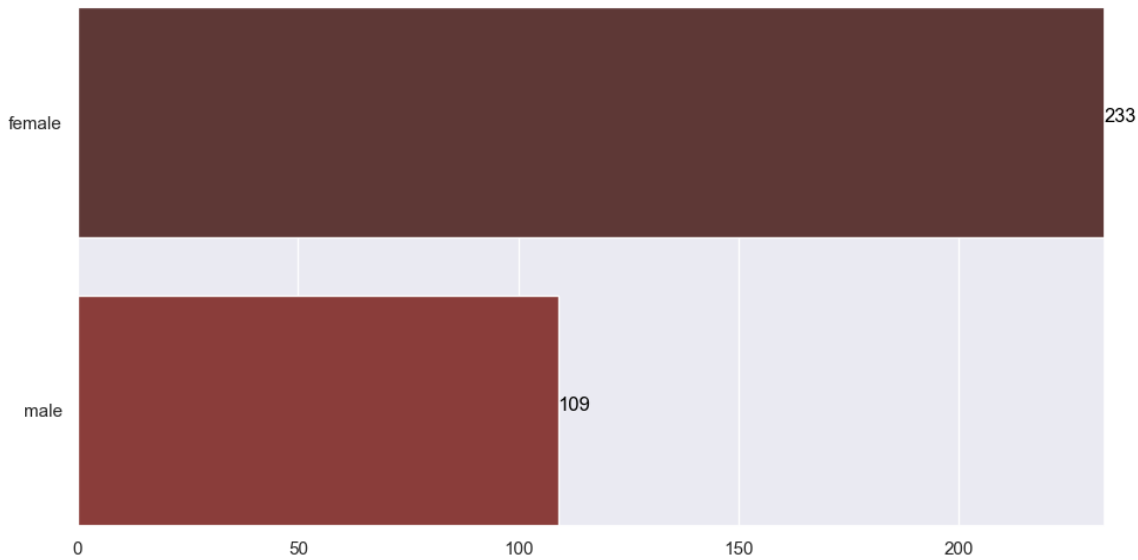
Report for Titanic Dataset

created on 2020-06-11 22:12

- For 0 samples



- For 1 samples

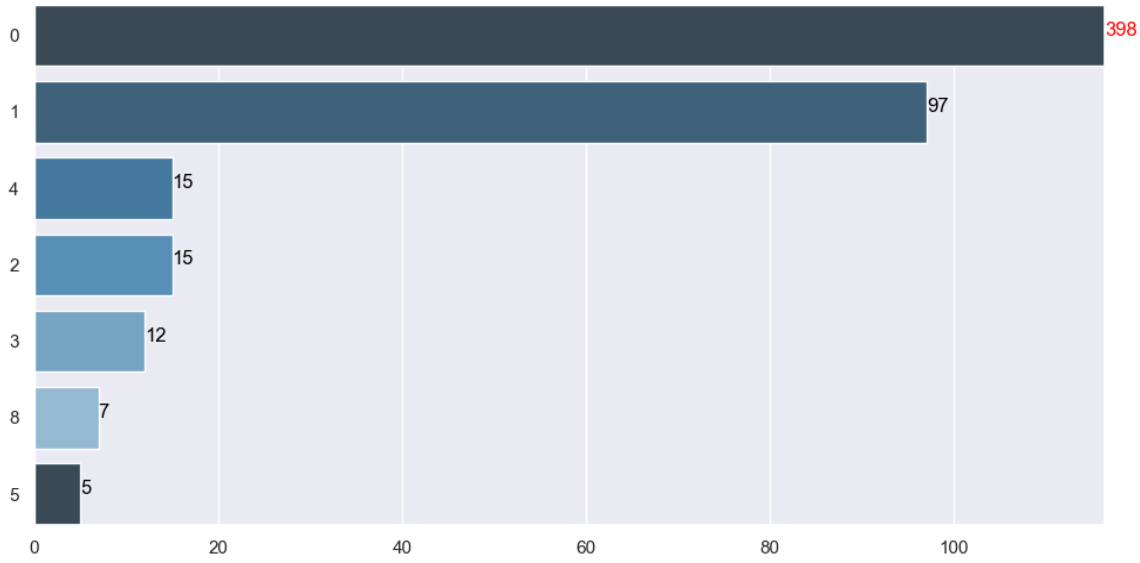


Distribution for SibSp

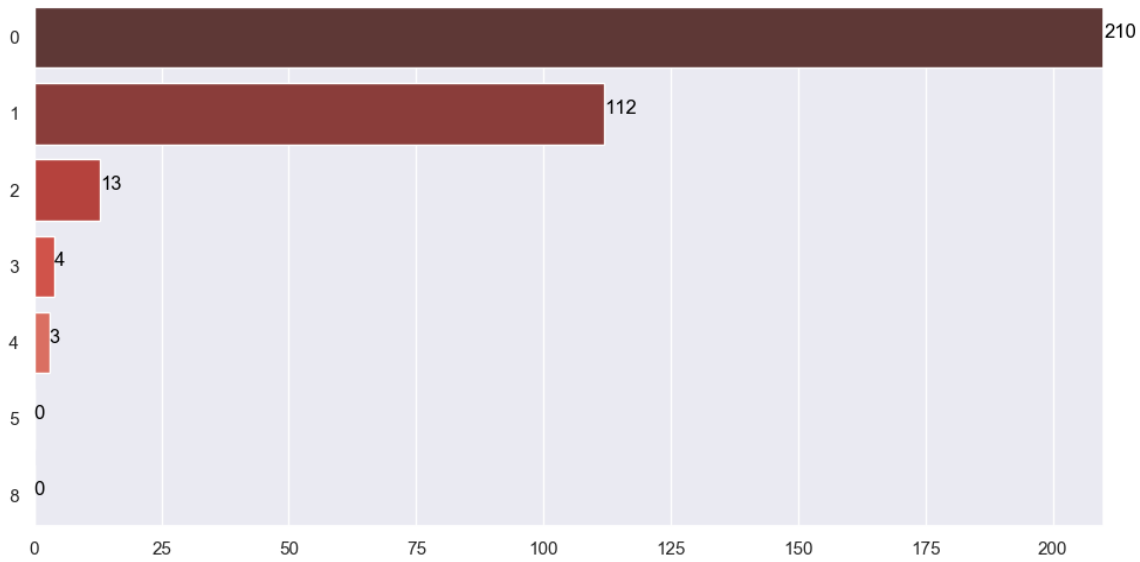
Report for Titanic Dataset

created on 2020-06-11 22:12

- For 0 samples



- For 1 samples

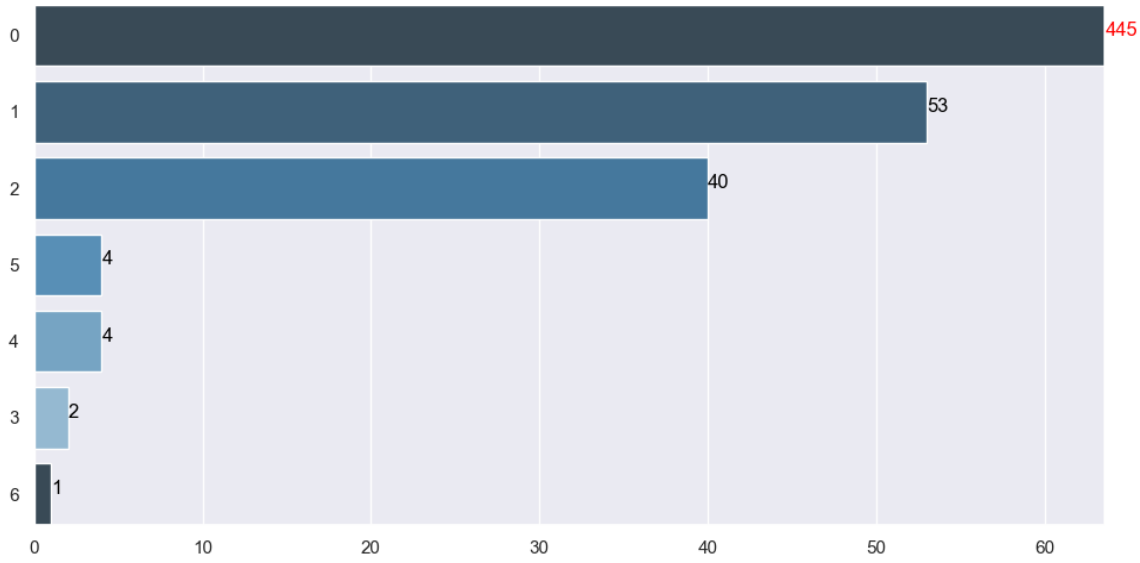


Distribution for Parch

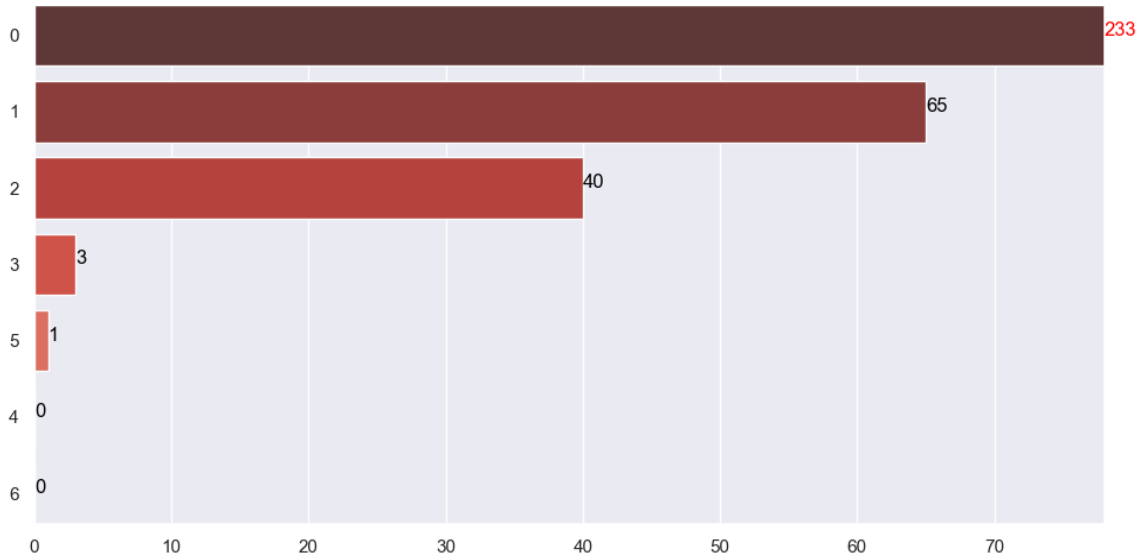
Report for Titanic Dataset

created on 2020-06-11 22:12

- For 0 samples



- For 1 samples

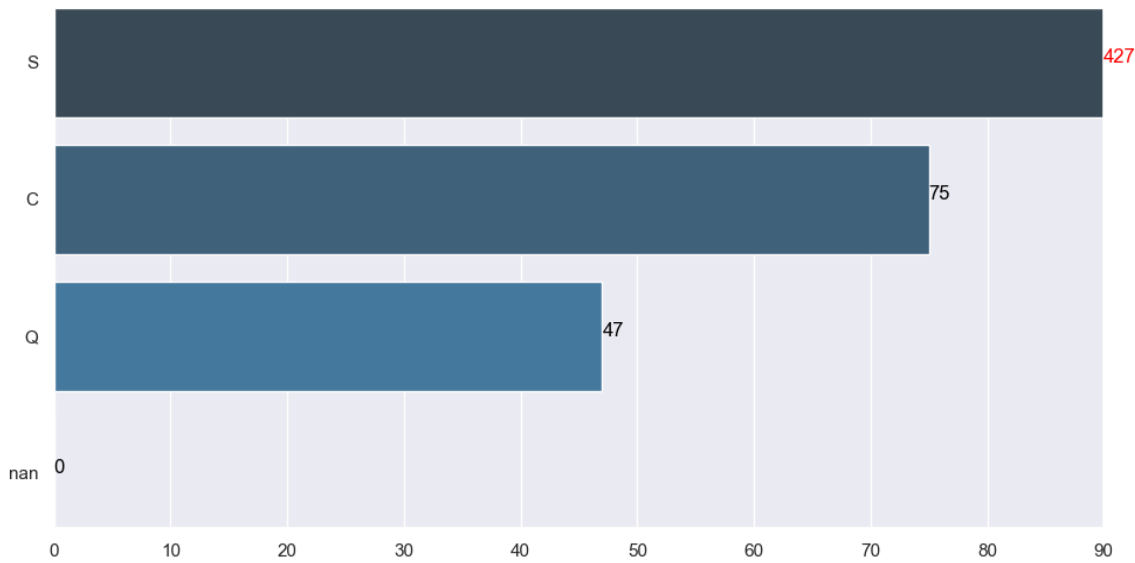


Distribution for Embarked

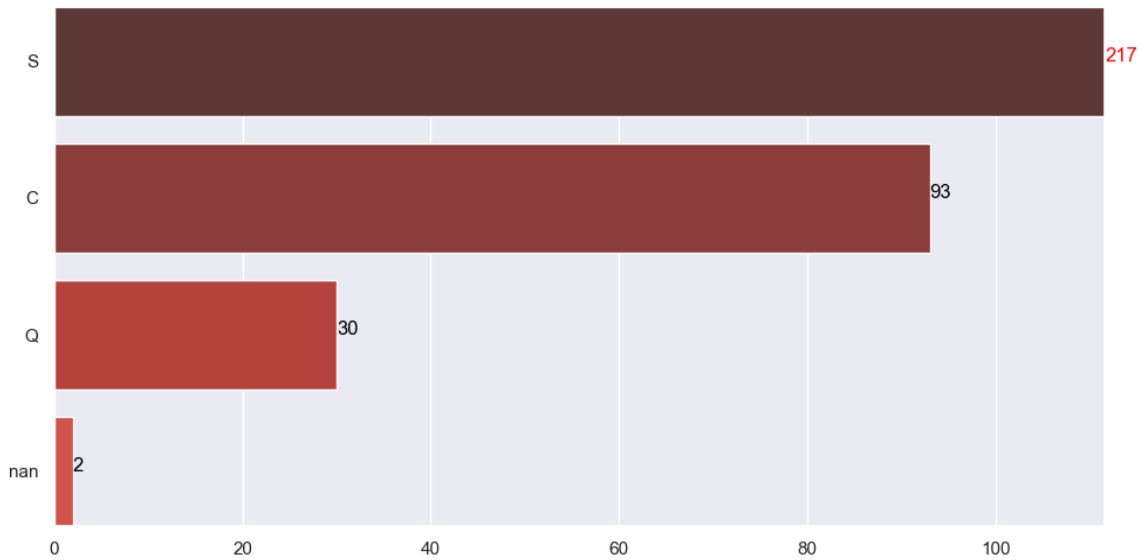
Report for Titanic Dataset

created on 2020-06-11 22:12

- For 0 samples



- For 1 samples

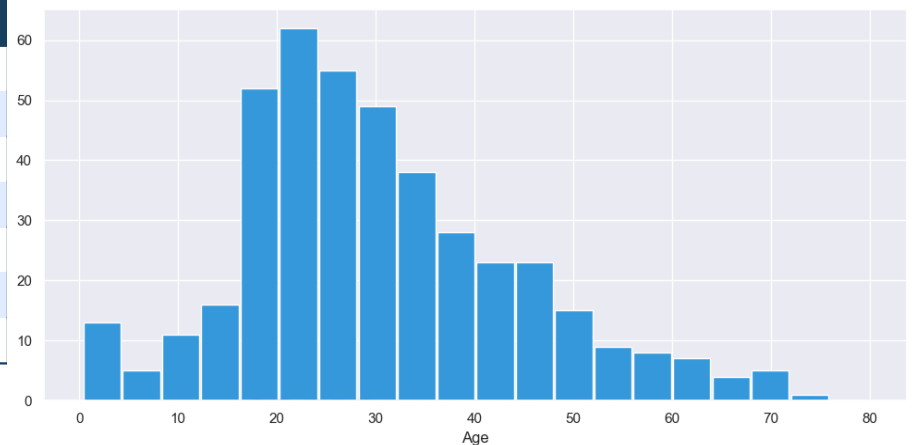


4.1.5 Numerical Field Distribution

Distribution for Age

- Distribution for 0

Statistical Field	Value
Total valid count	424
Min	1
Max	74
Mean	30
Median	28
Standard deviation	14
NAN count	125

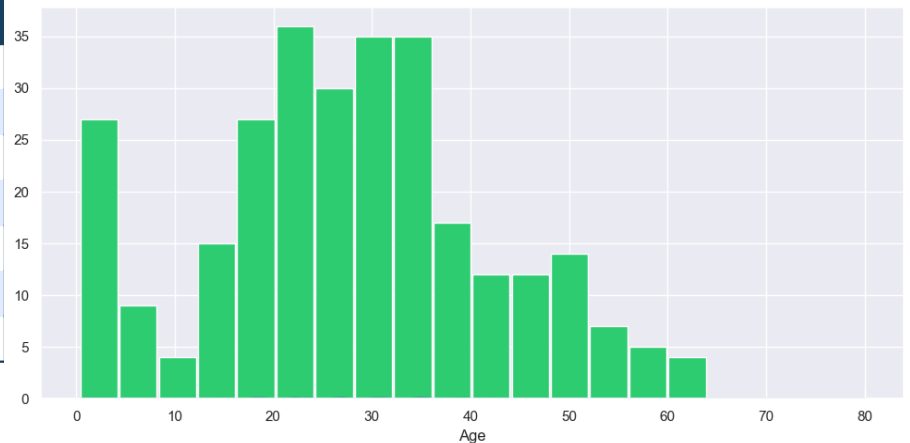


Report for Titanic Dataset

created on 2020-06-11 22:12

- Distribution for 1

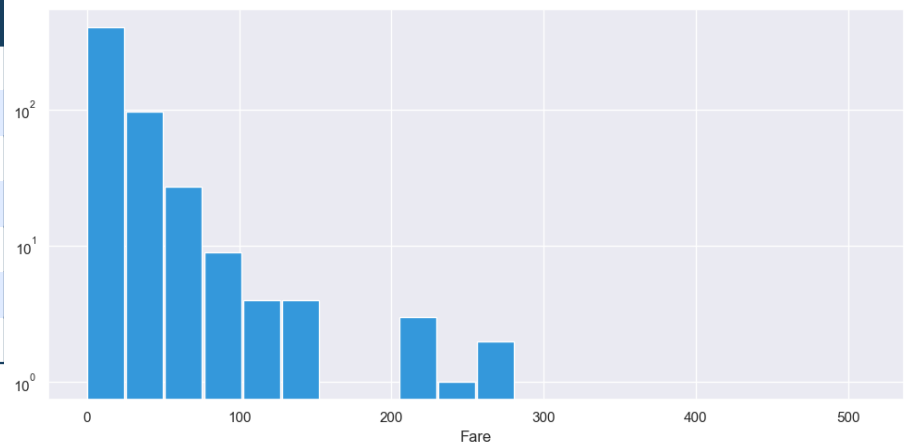
Statistical Field	Value
Total valid count	290
Min	0
Max	80
Mean	28
Median	28
Standard deviation	14
NAN count	52



Distribution for Fare

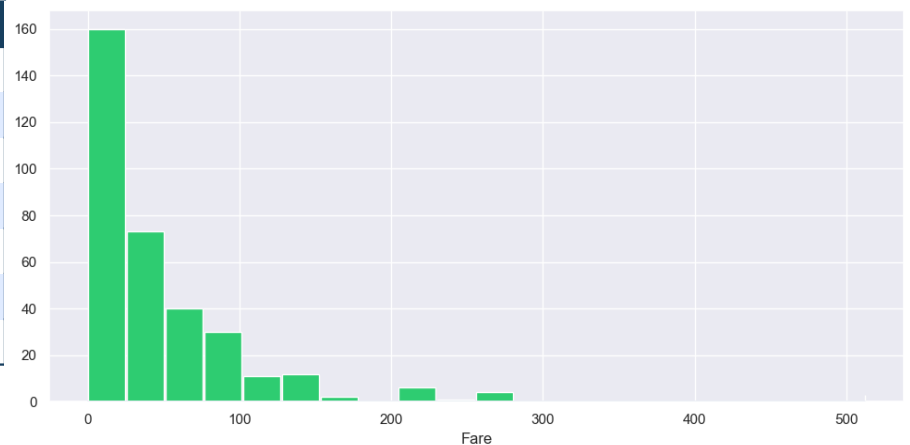
- Distribution for 0

Statistical Field	Value
Total valid count	549
Min	0
Max	263
Mean	22
Median	10
Standard deviation	31
NAN count	0



- Distribution for 1

Statistical Field	Value
Total valid count	342
Min	0
Max	512
Mean	48
Median	26
Standard deviation	66
NAN count	0



4.1.6 Text Field Distribution

Distribution for Name

Report for Titanic Dataset

created on 2020-06-11 22:12

