Report for Titanic Dataset

created on 2020-06-13 11:45

Content Table

Report for Titanic Dataset

created on 2020-06-13 11:45

1 Training Result

This section summarized the training performance

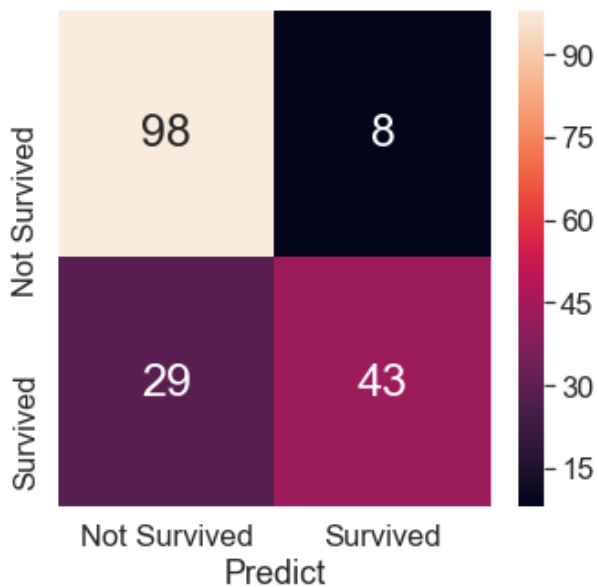### 1.1 Training Result

1.1.1 Metric Scores

This section shows binary-class evaluation metric.

Overall Result

Split/Round	Testing
Precision	0.8431
Recall	0.5972
F1	0.6992
Accuracy	0.7921
Auc	0.4068

1.1.2 Confusion Matrix
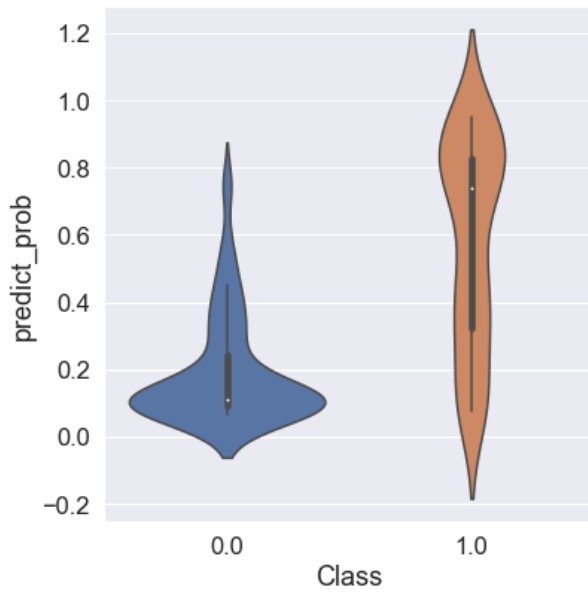
This section shows confusion matrix result.

1.1.3 Confidence Distribution

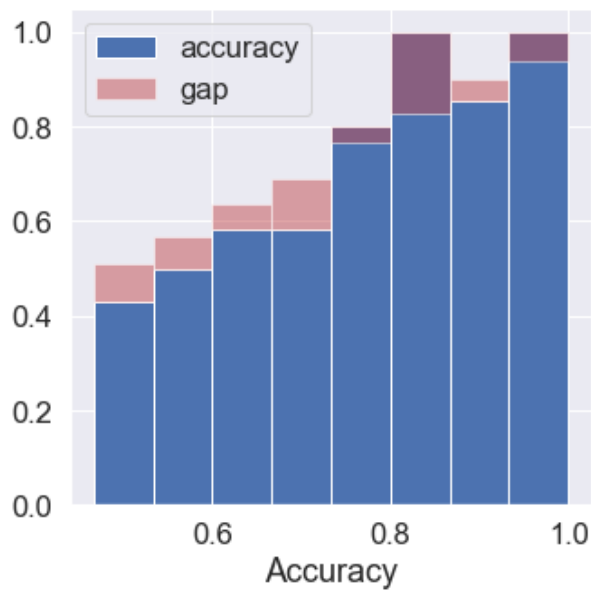This section shows binary-class confidence distribution result.

Report for Titanic Dataset

created on 2020-06-13 11:45

1.1.4 Reliability Diagram

This section shows binary-class reliability diagram result.
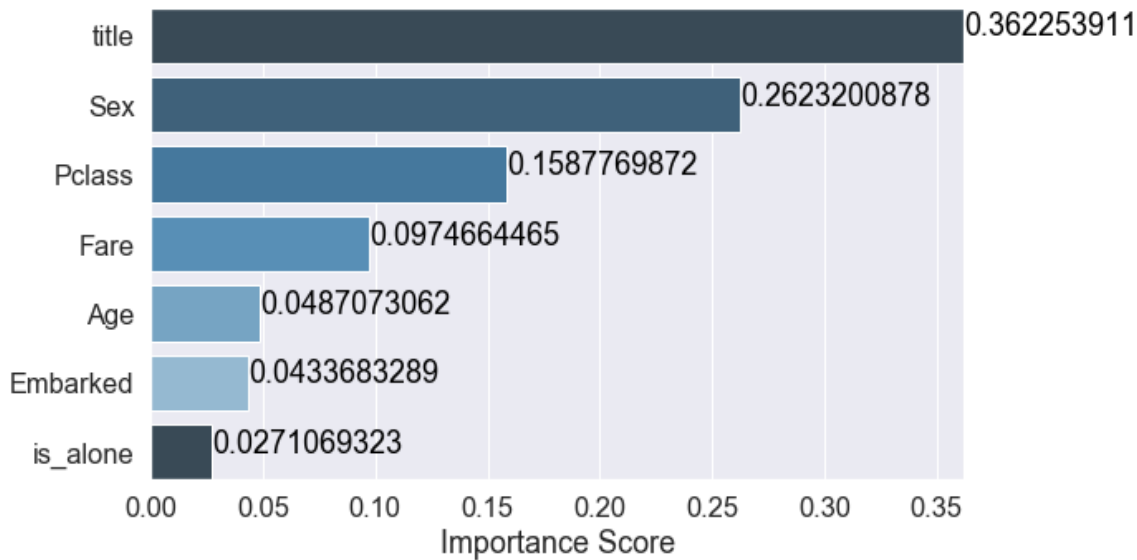
Report for Titanic Dataset

created on 2020-06-13 11:45

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
title	0.362253911
Sex	0.2623200878
Pclass	0.1587769872
Fare	0.0974664465
Age	0.0487073062
Embarked	0.0433683289
is_alone	0.0271069323

Report for Titanic Dataset

created on 2020-06-13 11:45

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: 0.066134612311668
- Fare=lower than 7.91: **-0.054**
- Embarked=S: **-0.067**
- Pclass=3: **-0.129**
- Sex=male: **-0.221**
- title=Mr: **-0.244**

Class Not Survived - Confidence: 0.933865387688332
- title=Mr: **0.244**
- Sex=male: **0.221**
- Pclass=3: **0.129**
- Embarked=S: **0.067**
- Fare=lower than 7.91: **0.054**

Example 1:
Class Not Survived - Confidence: 0.07916835228833646
- Embarked=S: **0.074**
- Fare=higher than 31: **-0.057**
- Pclass=1: **-0.150**
- title=Mrs: **-0.195**
- Sex=female: **-0.222**

Class Survived - Confidence: 0.9208316477116637
- Sex=female: **0.222**
- title=Mrs: **0.195**
- Pclass=1: **0.150**
- Fare=higher than 31: **0.057**
- Embarked=S: **-0.074**

4   Data Statistics Analysis
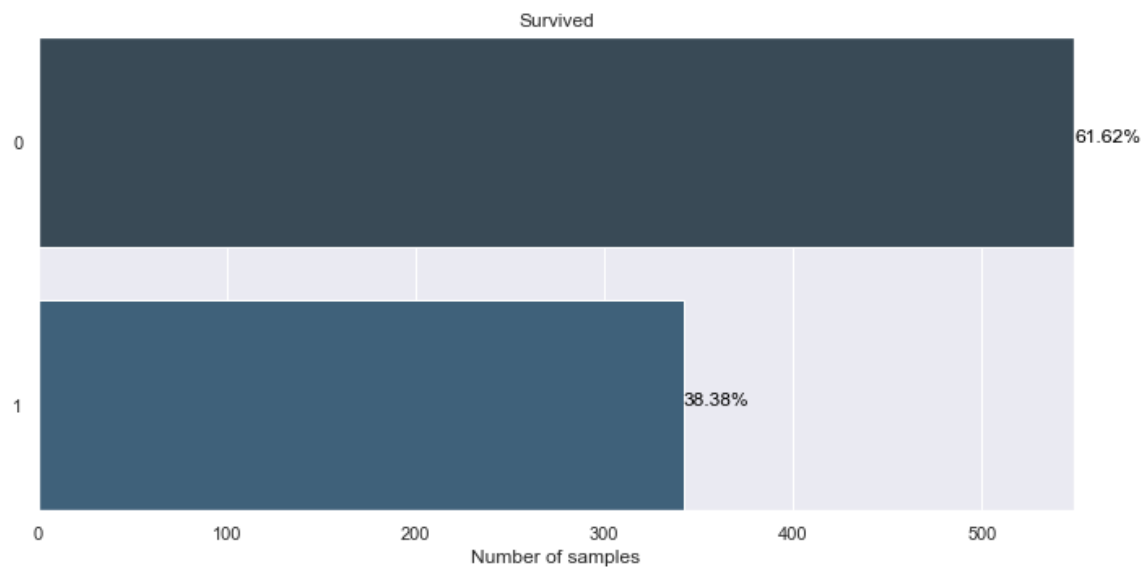
This section provides the analysis on data

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.
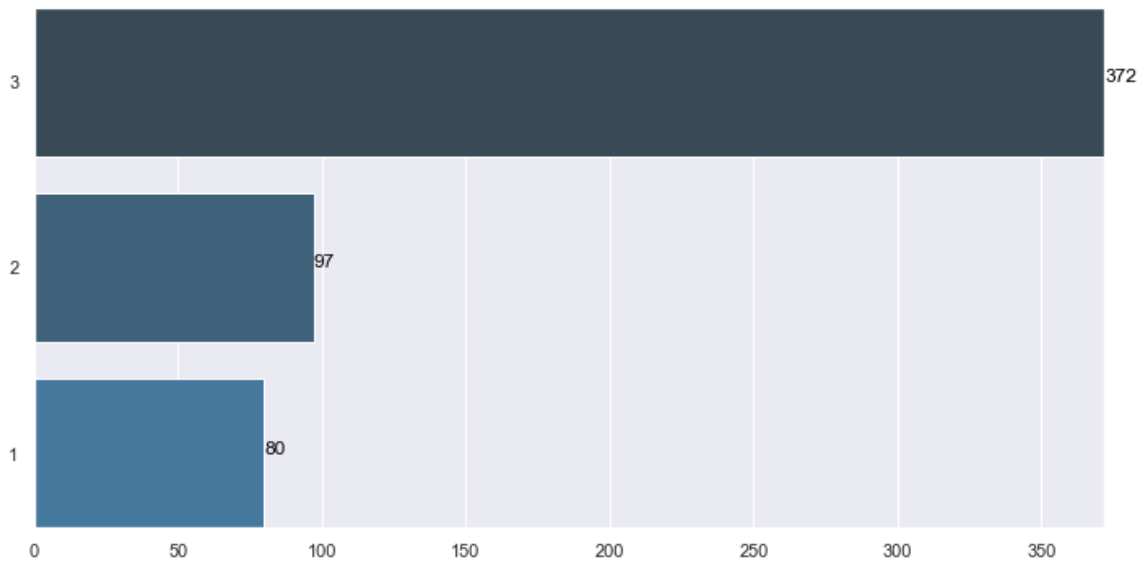
Report for Titanic Dataset

created on 2020-06-13 11:45

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

4.1.4 Categorical Field Distribution

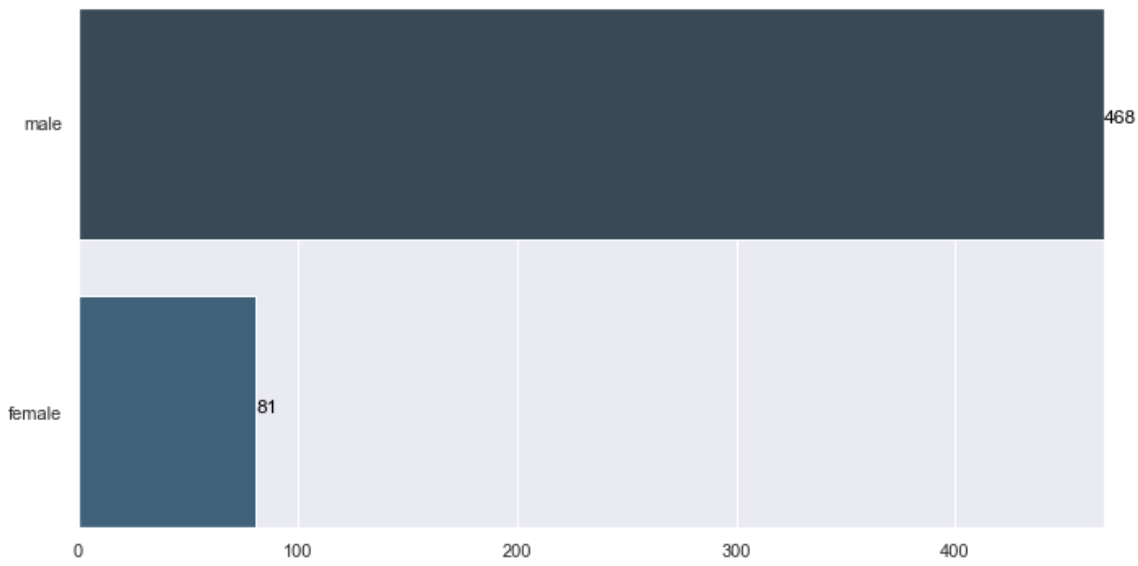Distribution for Pclass

- *For 0 samples*

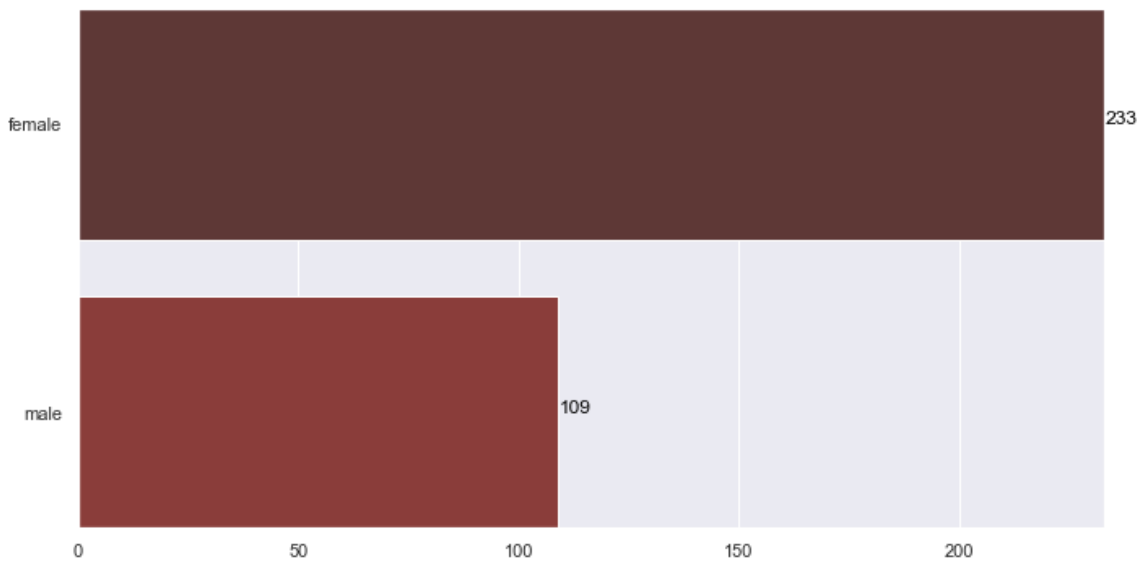- *For 1 samples*
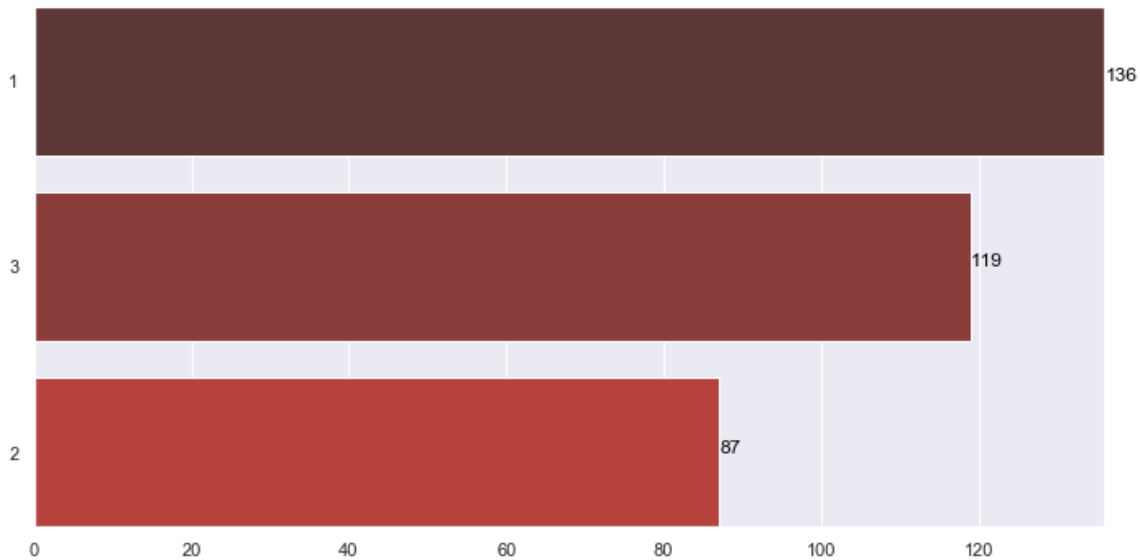
Distribution for Sex

Report for Titanic Dataset

created on 2020-06-13 11:45

- *For 0 samples*

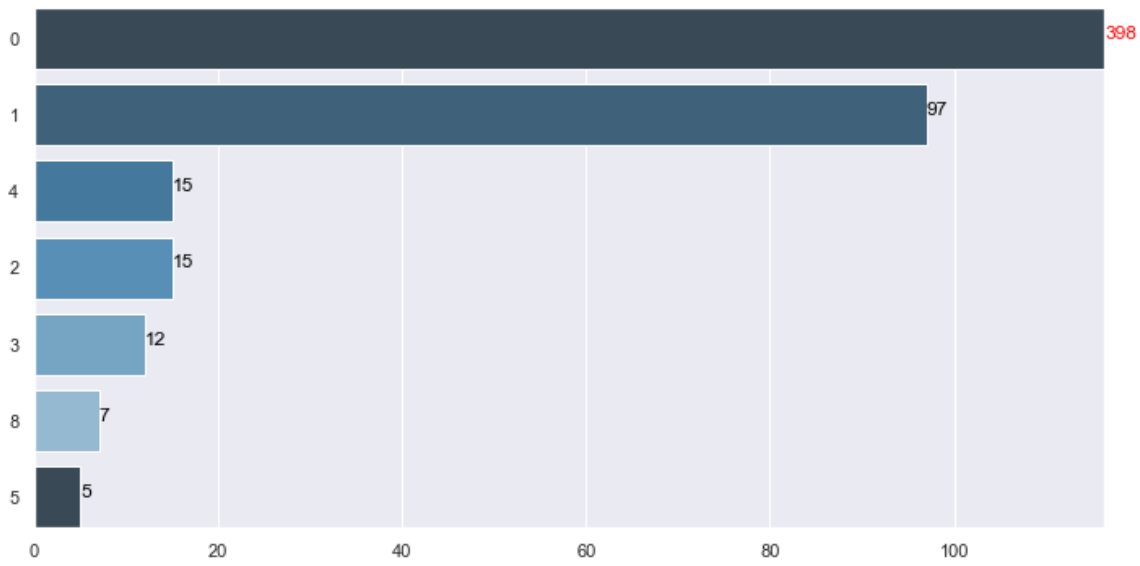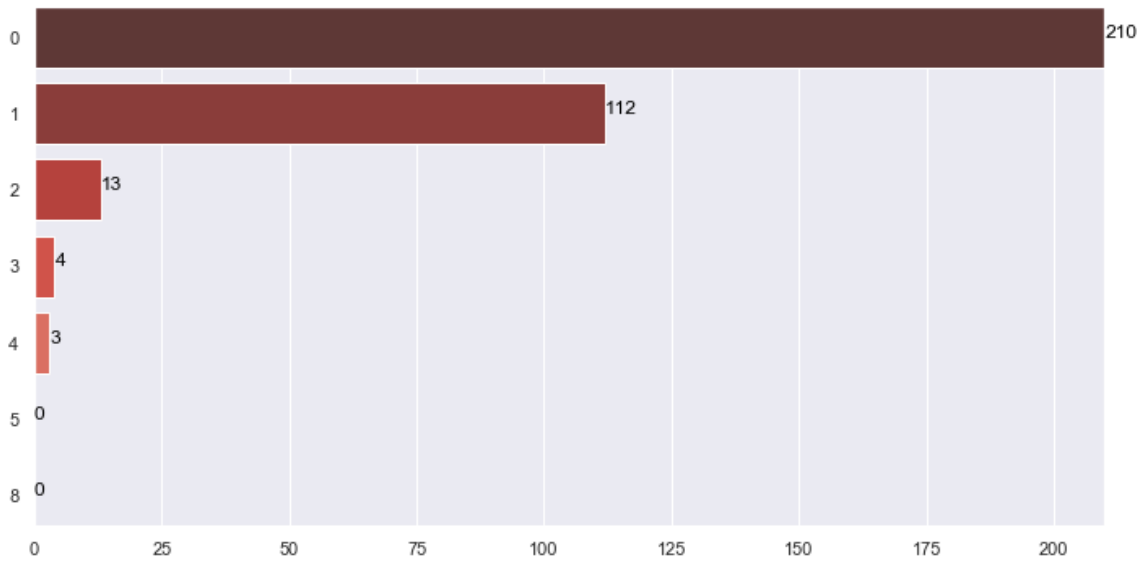- *For 1 samples*

Distribution for SibSp

Report for Titanic Dataset

created on 2020-06-13 11:45

- For 0 samples
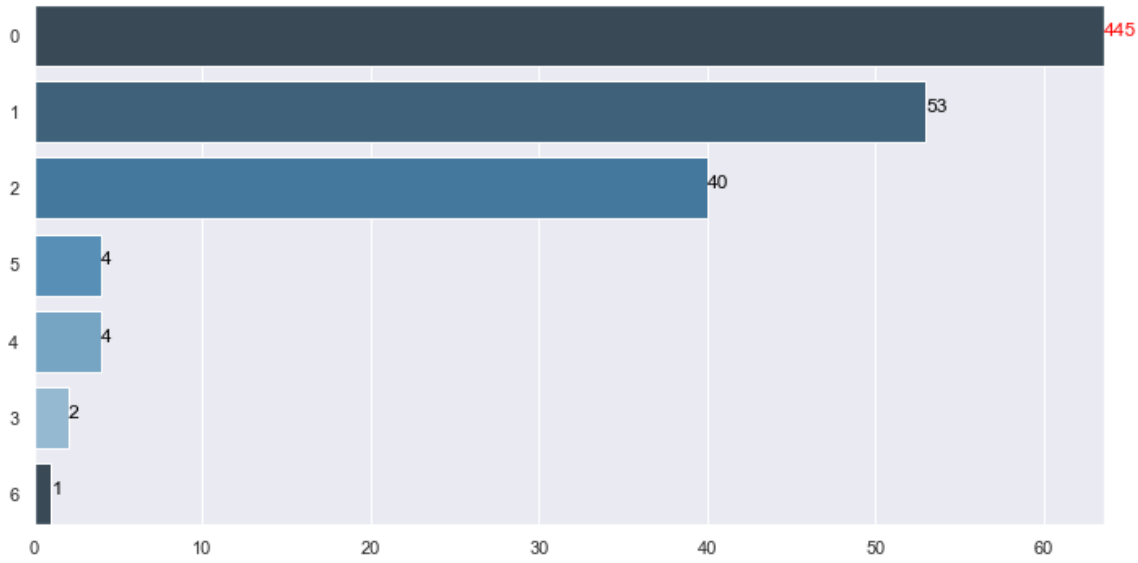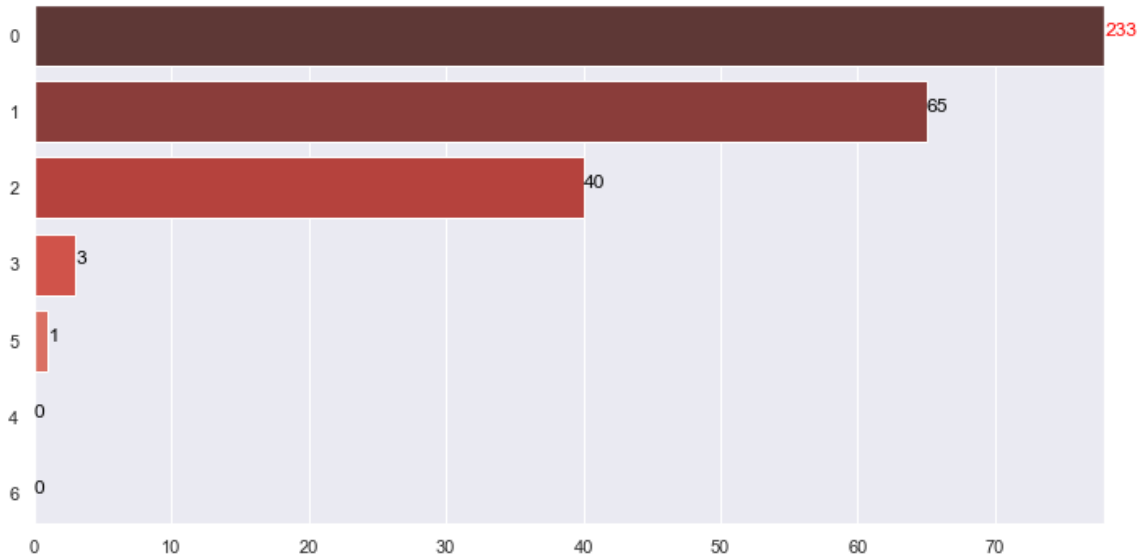

- For 1 samples

Distribution for Parch

Report for Titanic Dataset

created on 2020-06-13 11:45

- *For 0 samples*

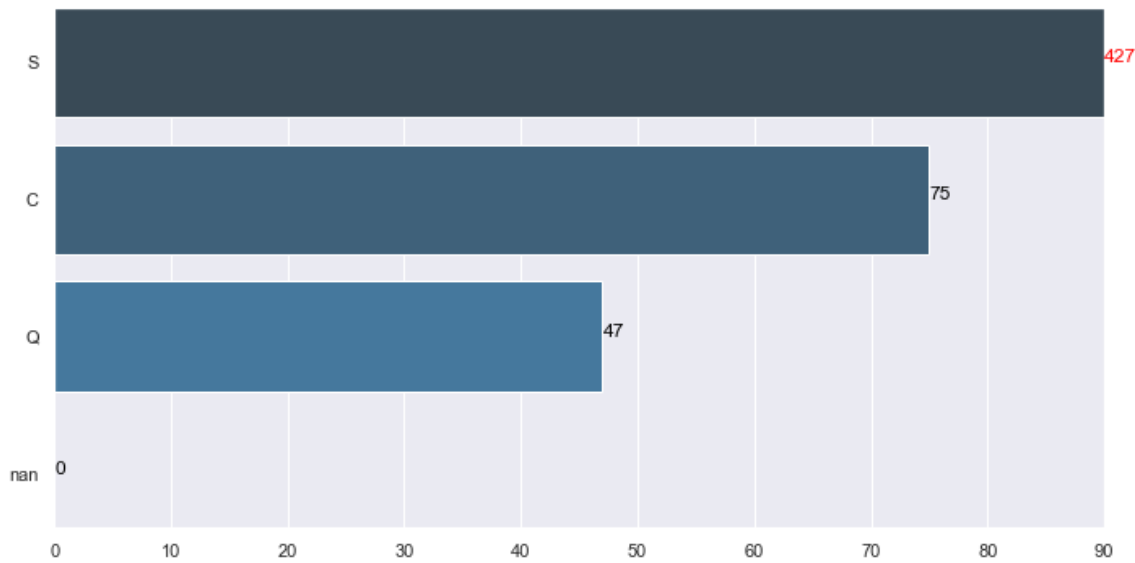- *For 1 samples*
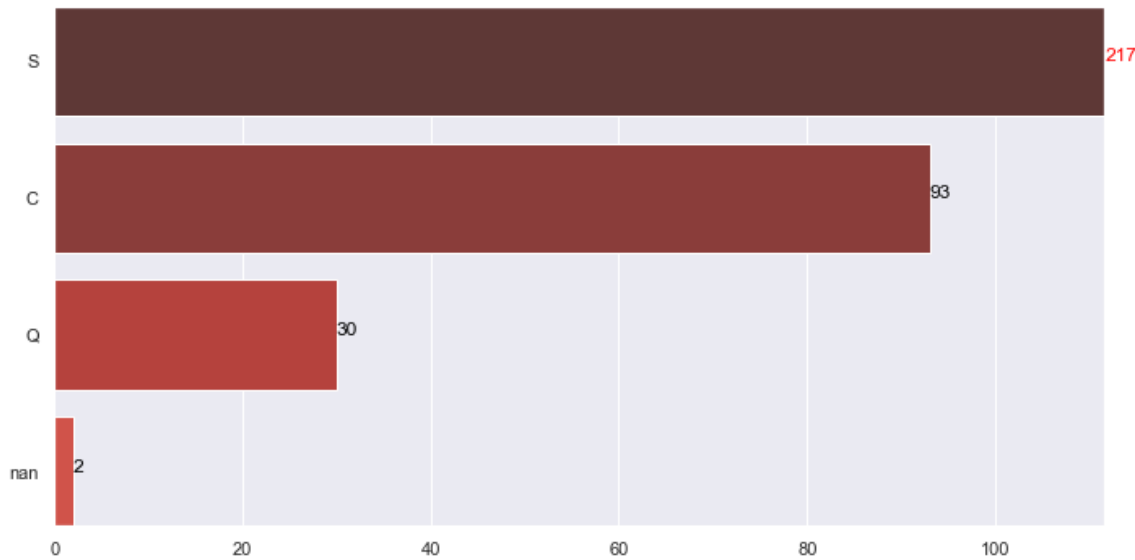

Distribution for Embarked

Report for Titanic Dataset

created on 2020-06-13 11:45
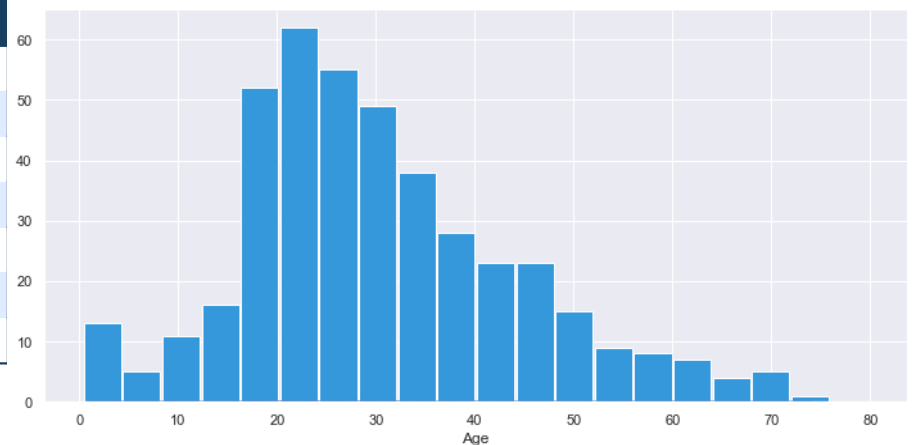
- For 0 samples

- 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
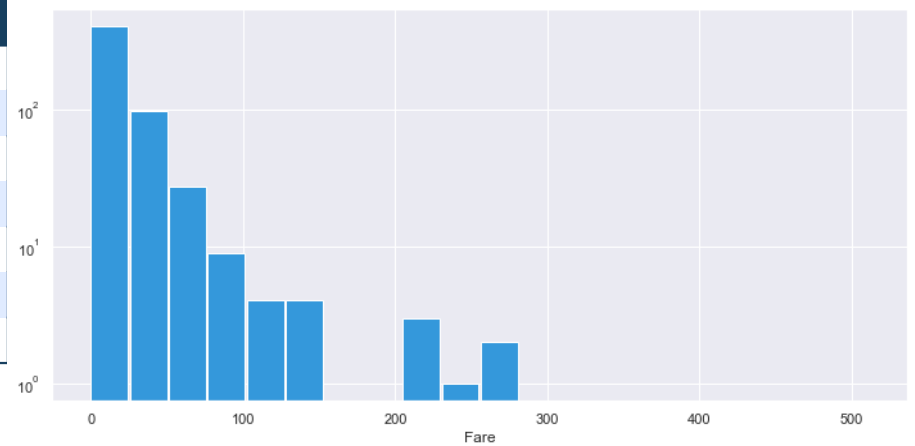
created on 2020-06-13 11:45

- Distribution for 1

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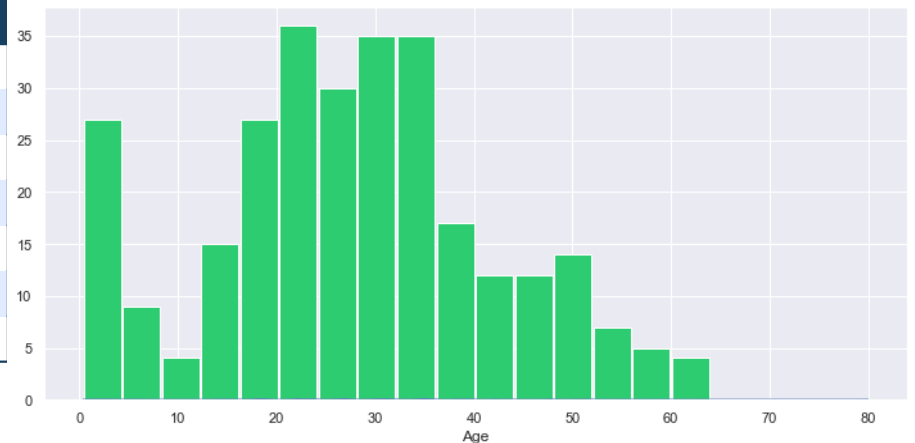
Distribution for Fare

- 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
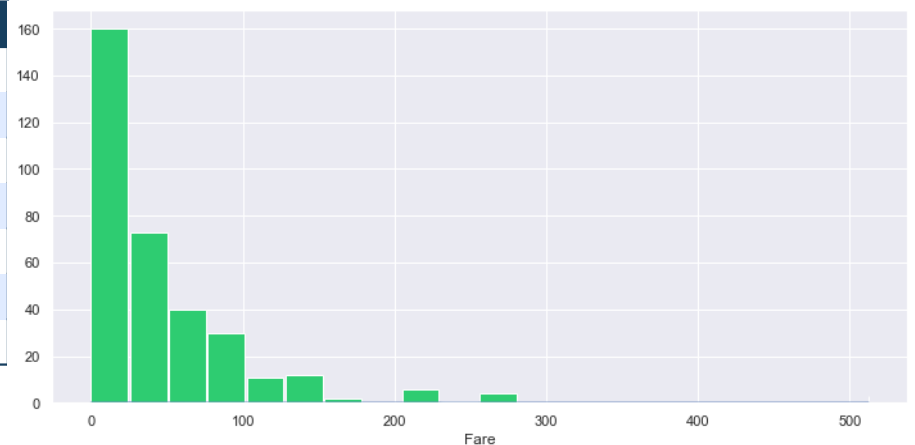
4.1.6 Text Field Distribution

Distribution for Name

Report for Titanic Dataset

created on 2020-06-13 11:45