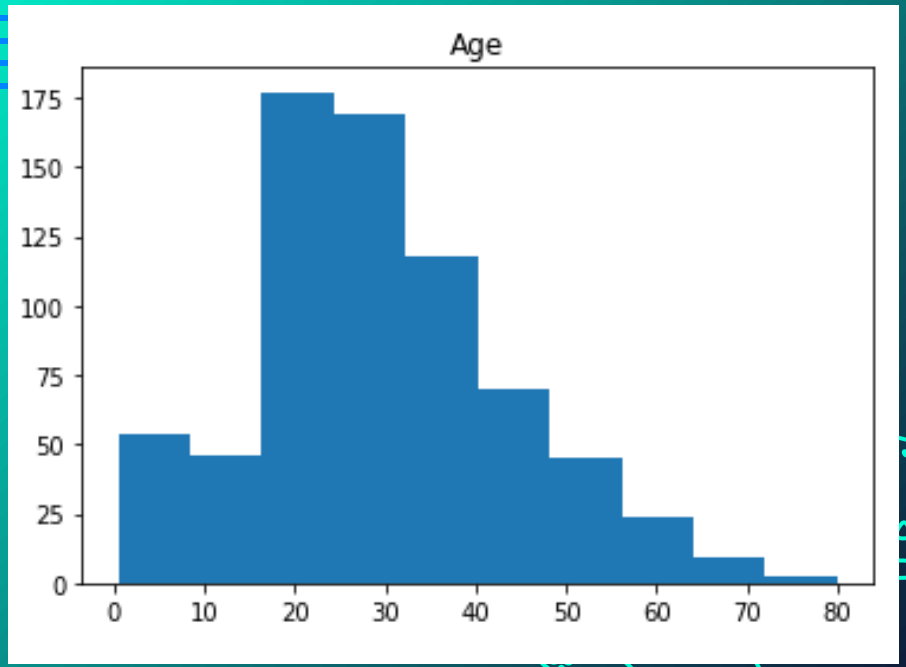
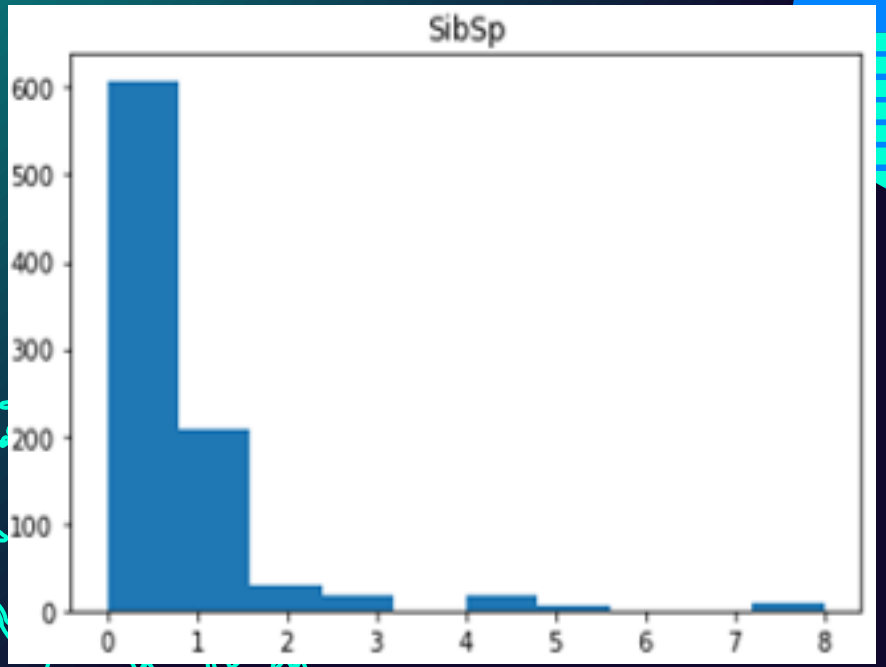
EDA on Titanic Dataset using Voting Classifier

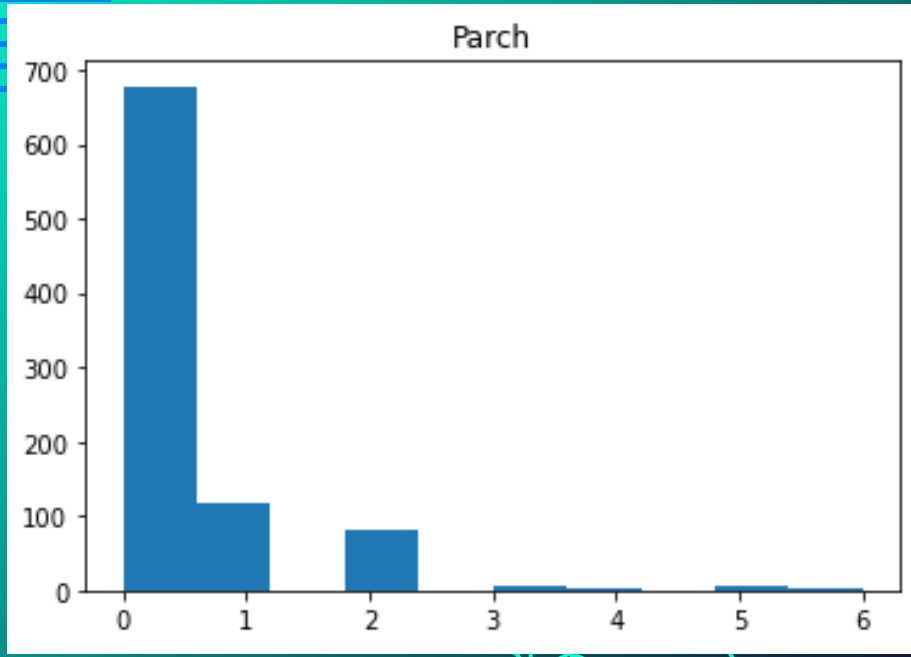
01.

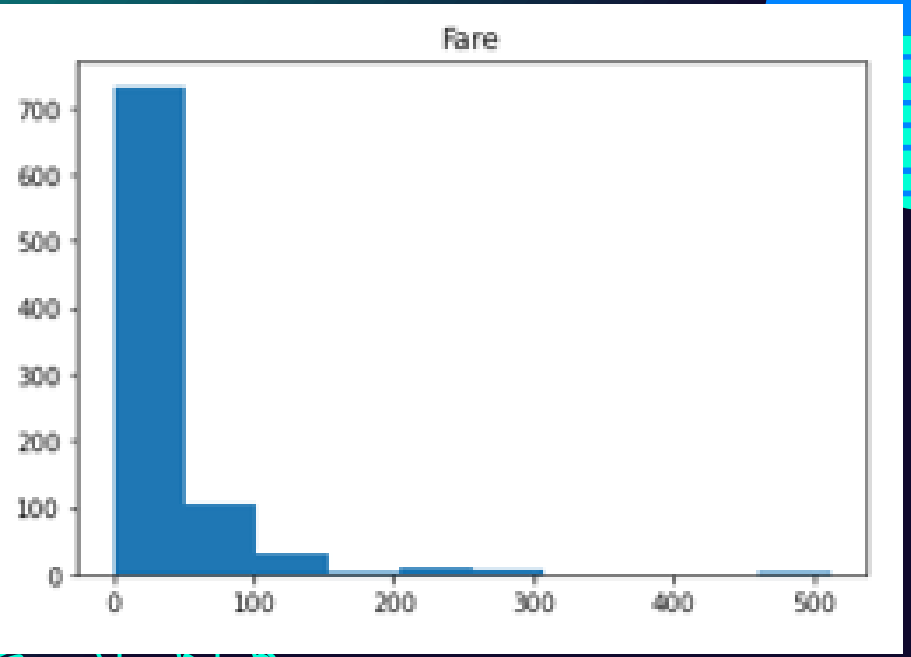
E D A

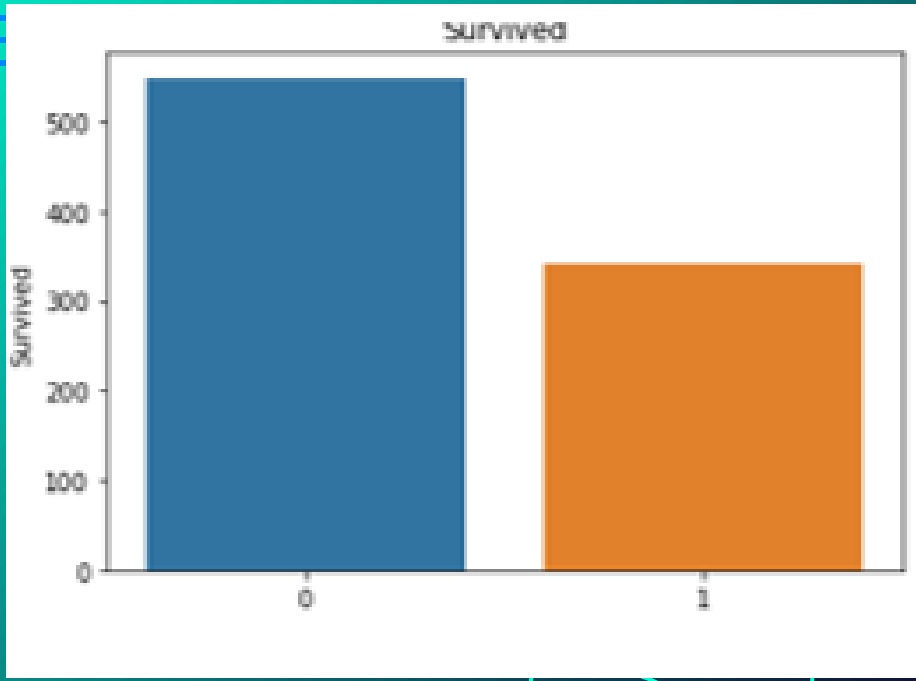
Exploratory Data Analysis

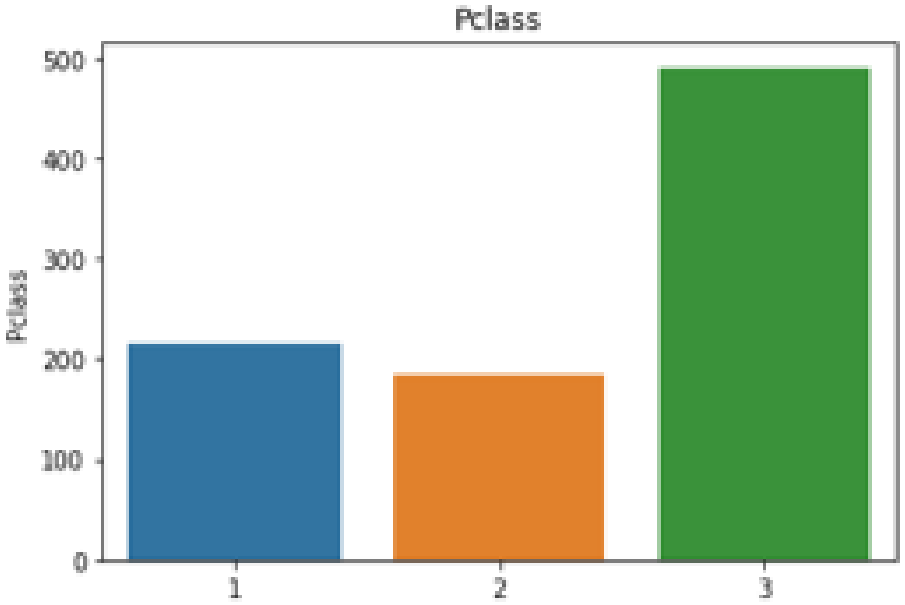

Age
175
150
125
100
75
50
25
0
0
10
20
30
40
50
60
70
80

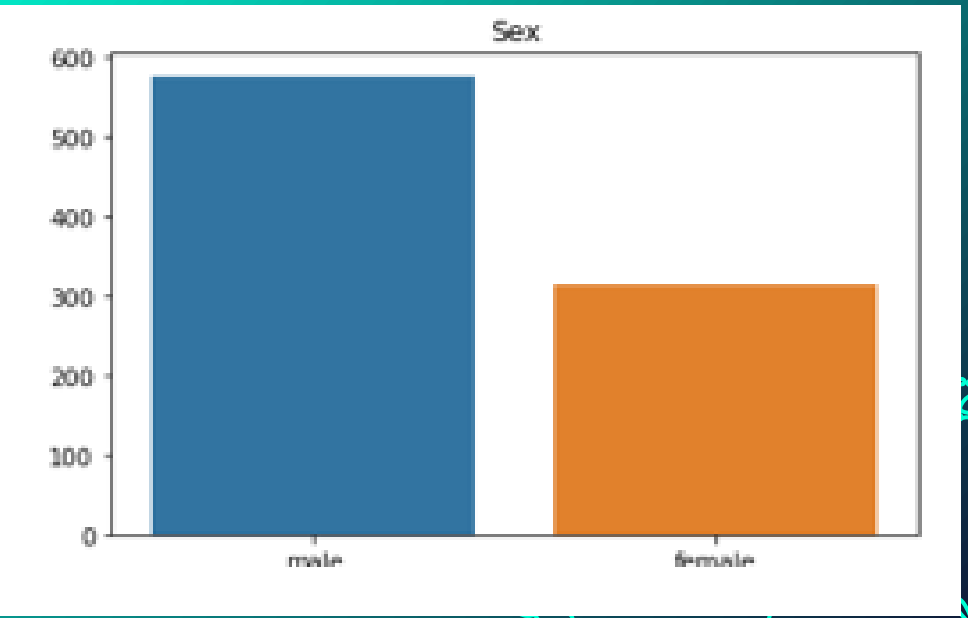

SibSp
600
500
400
300
200
100
0
0
1
2
3
4
5
6
7
8

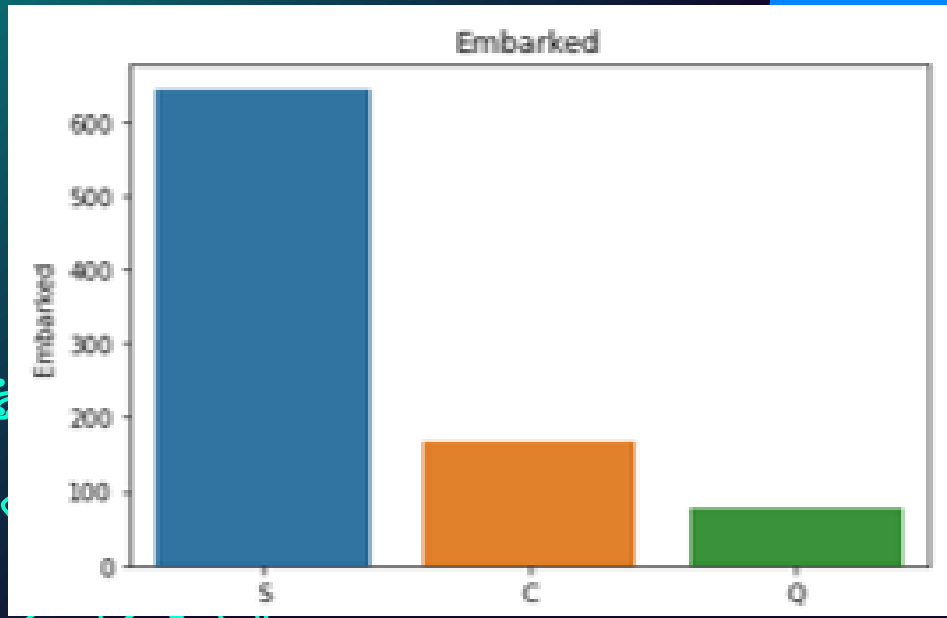

Parch
700
600
500
400
300
200
100
0
0
1
2
3
4
5
6

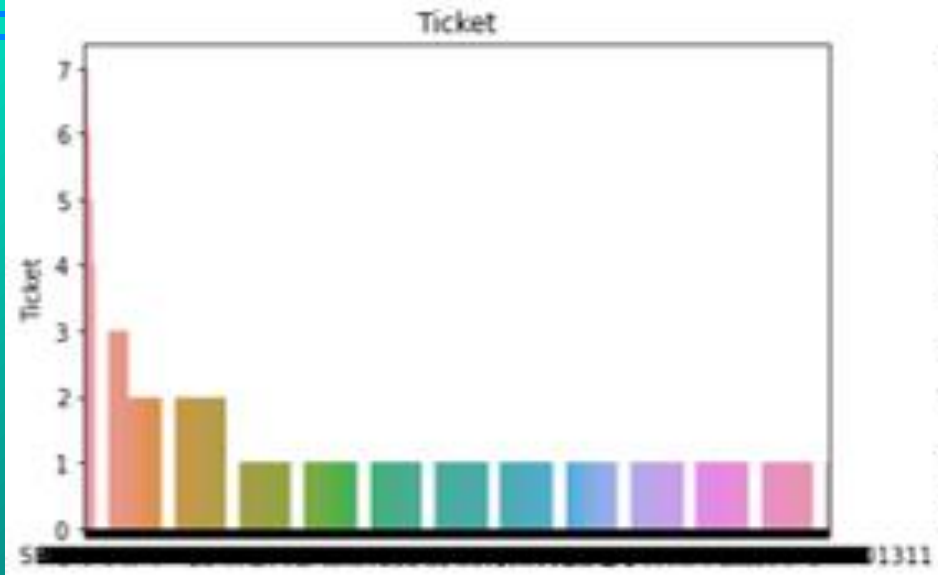

Fare
700
600
500
400
300
200
100
0
0
100
200
300
400
500

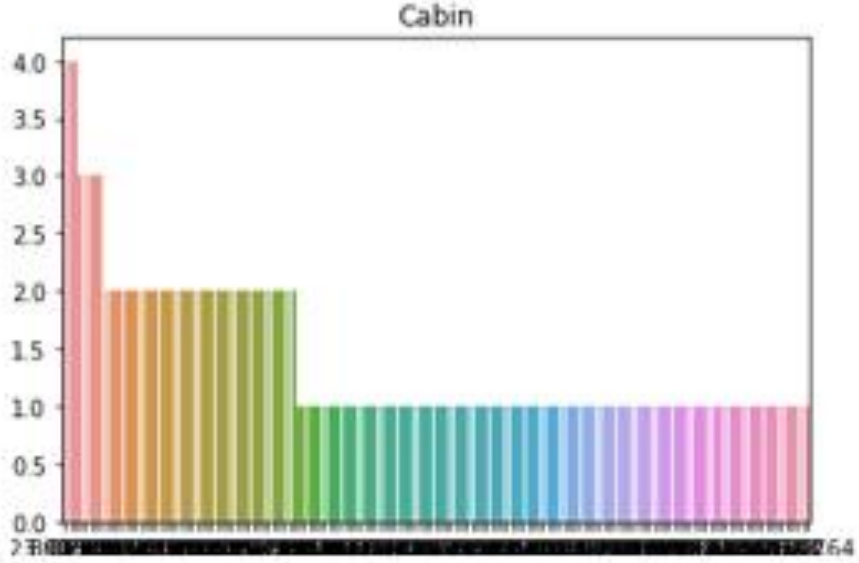

Survived
Survived
500
400
300
200
100
0
0
1


Pclass
Pclass
500
400
300
200
100
0
1
2
3


Sex
600
500
400
300
200
100
0
male
female


Embarked
Embarked
600
500
400
300
200
100
0
S
C
Q


Ticket
Ticket
7
6
5
4
3
2
1
0
1311

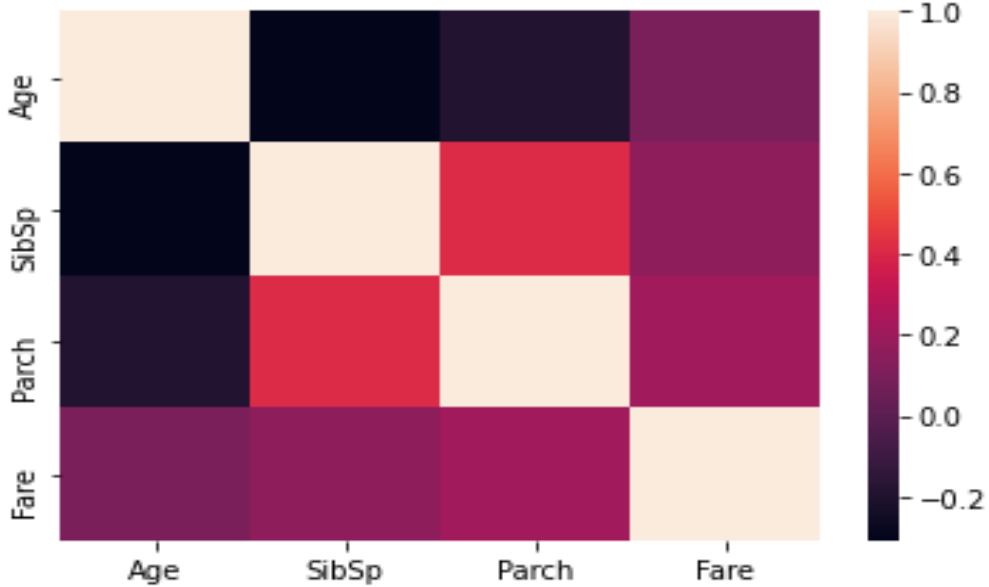

Cabin
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0

02.

Heatmap

Correlation Analysis

	Age	SibSp	Parch	Fare
Age	1.000000	-0.308247	-0.189119	0.096067
SibSp	-0.308247	1.000000	0.414838	0.159651
Parch	-0.189119	0.414838	1.000000	0.216225
Fare	0.096067	0.159651	0.216225	1.000000



03.

Machine Learning

Feature Engineering & Tuning

Model	Baseline	Tuned Performance
Naive Bayes	72.6%	NA
Logistic Regression	82.1%	82.6%
Decision Tree	77.6%	NA
K Nearest Neighbor	80.5%	83.0%
Random Forest	80.6%	83.6
Support Vector Classifier	83.2%	83.2%
Xtreme Gradient Boosting	81.8%	85.3%

```
voting_clf_hard : [0.79213483 0.81460674 0.82022472 0.79775281 0.836
15819]
voting_clf_hard mean : 0.8121754586427983
voting_clf_soft : [0.78651685 0.82022472 0.81460674 0.79775281 0.853
10734]
voting_clf_soft mean : 0.8155652891512728
voting_clf_all : [0.80898876 0.83146067 0.8258427  0.80898876 0.8587
5706]
voting_clf_all mean : 0.8268075922046595
voting_clf_xgb : [0.82022472 0.8258427  0.83146067 0.80898876 0.8700
565 ]
voting_clf_xgb mean : 0.8313146702215451
```

```
Fitting 5 folds for each of 7 candidates, totalling 35 fits

[Parallel(n_jobs=-1)]: Using backend LokyBackend with 4 concurrent w
orkers.
[Parallel(n_jobs=-1)]: Done  35 out of  35 | elapsed:   22.9s finish
ed

VC Weights
Best Score: 0.831321018218752
Best Parameters: {'weights': [1, 1, 1]}
```

```
0    410
1      8
Name: difference_hard_all, dtype: int64
```

THANKS

@achdiyat.kusuma

https://www.linkedin.com/in/achdiyat-kusuma-9b3141121/