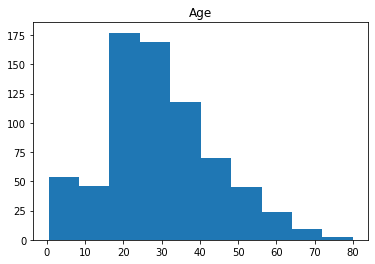
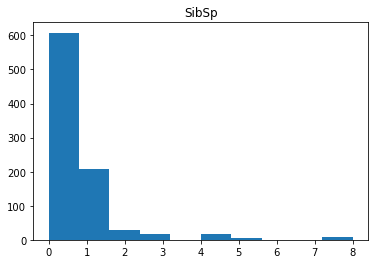
Project #3 : Titanic Survival Prediction

* Observation from numerical column:

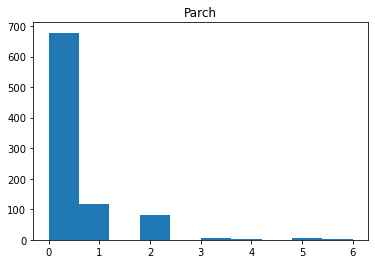
1. Age - fairly normal distribution



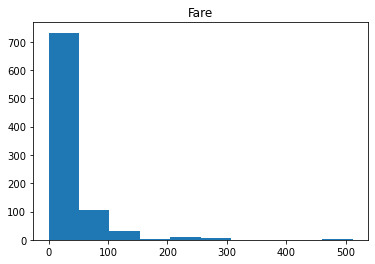
1. Sibling/spouse

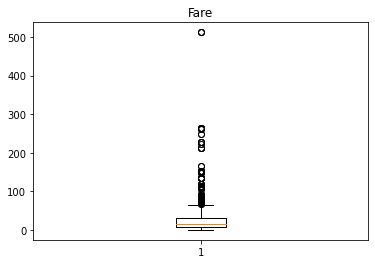


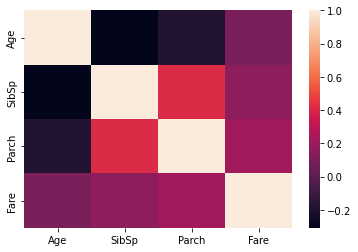
1. Parent+child



1. Passenger Fare: should be normalized due to spike at one interval, too many outlier





1. Correlation between the numerical variables:

Age SibSp Parch Fare

Age 1.000000 -0.308247 -0.189119 0.096067

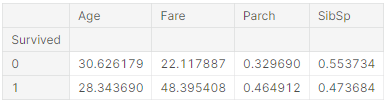
SibSp -0.308247 1.000000 0.414838 0.159651

Parch -0.189119 0.414838 1.000000 0.216225

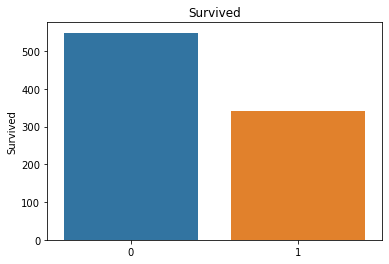
Fare 0.096067 0.159651 0.216225 1.000000

Correlation sibling with age ; Parent + child with sibling

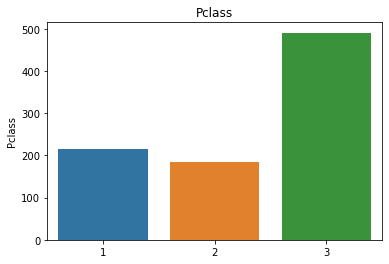
* Observation from Pivot table (numerical variable)



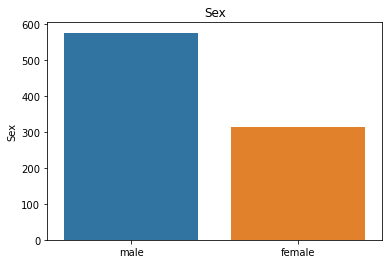
* Average age that survive is young, around 28 yrs
* People that pay more, more likely to survive
* Parent with child more likely to survive
* Observation from categorical variable:

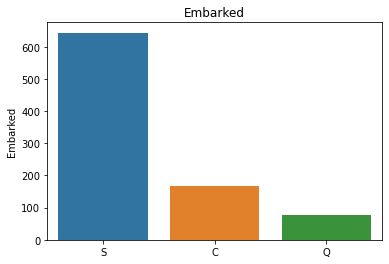


More people did not survive

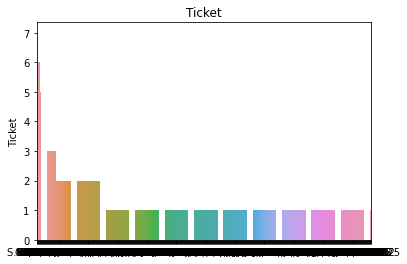
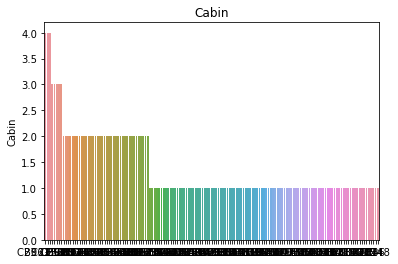


Majority 3rd class passenger, then 1st class , then 2nd class

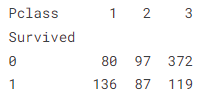




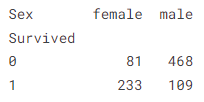
More people embark from Southampton Port



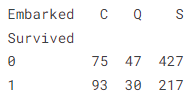
Ticket dgn Cabin perlukan Feature Engineering sbb dari bar graph ni xde meaning sgt



More 1st class passenger survived. Most 3rd class passenger did not survive.



More female survived compared to male.

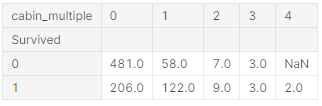


Since ramai yg embark dari Southampton, so ramai yg survive dari port tu. So, x relevant utk prediction.

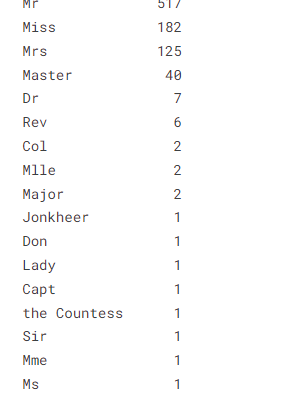
Feature Engineering – utk column yg data nmpk mcm x relevant dekat distribution

Doubts:

1. Cabin data byk yg xde:   
   - Nak tau kalau org yg ada rekod cabin number ni survive ke  
   - Maybe cabin floor bawah semua x survive
2. Nak tgk brapa org yg ada sewa lebih dari satu cabin ni survived ke tak:



Yg 4 cabin survived, yg 3 cabin half-half. Yg 0 tu xde rekod sbb majority x survive.

1. Maybe yg ada status royalty ada higher chance of survival.

