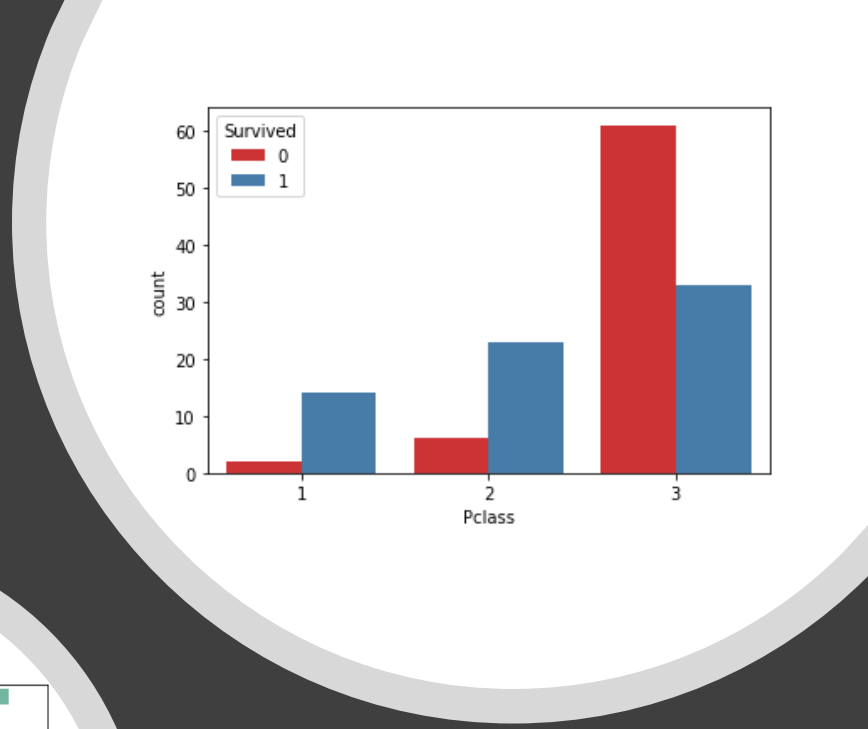
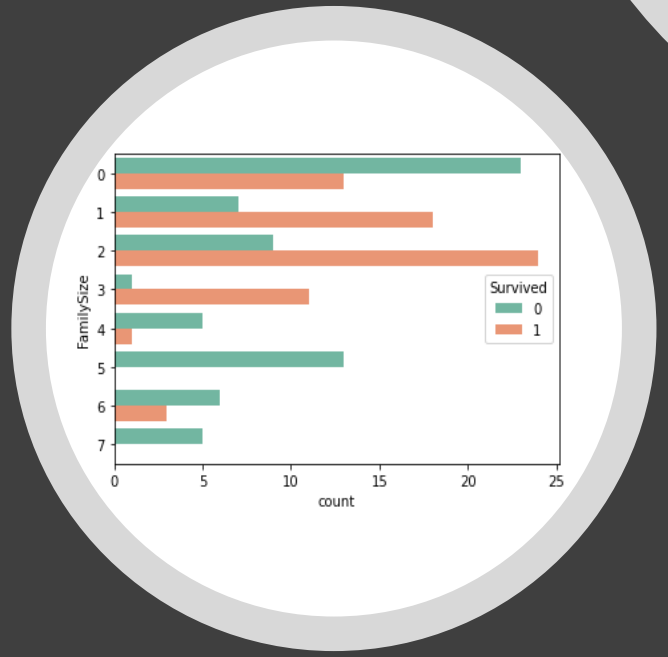
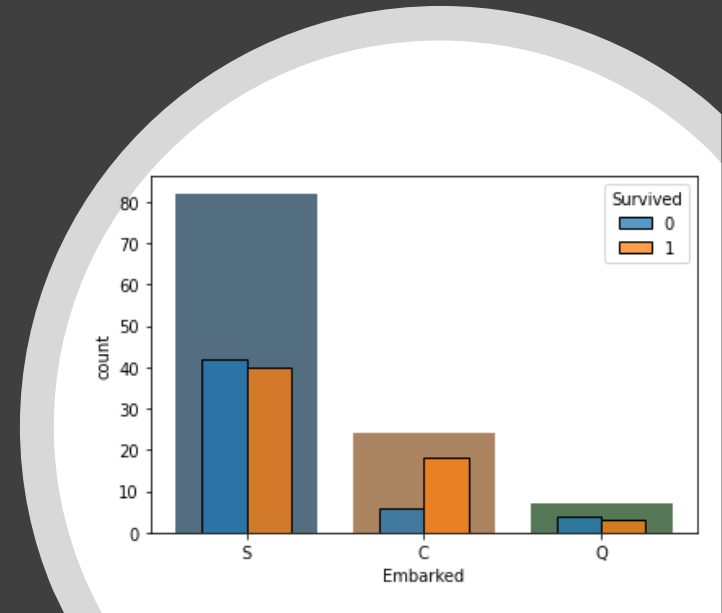
# Titanic Dataset:




- 12 features and 139 observations
- The goal is to be survived after Crash
- Max in age:18 years , Min in age : 4-month, age average : 9 years
- 50% Survived and 50% Not Survived
- Females survived more than males
- The highest survival percentage is among passengers embarked from the port Cherbourg  in France
- Family Size: 1,2,3 => survived the most,  5&7=> not survived, 0(alone),4,6:  a few survival
- Passengers could afford higher class were saved more like passengers in Pclass 1,2



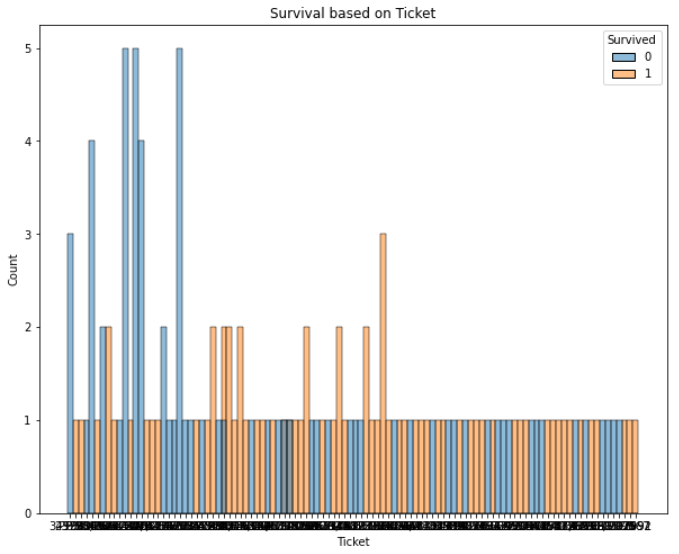
# The question is which factors affect the survival rate of children?
# First need to address with 2 issues

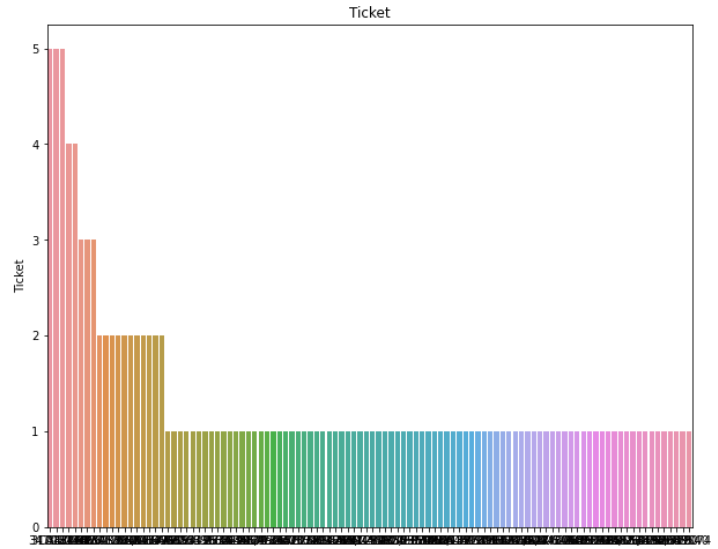
Ticket:

- Thicket determine location of passengers' room. The Survival depends on passenger location
- Ticket value does not make sense
- It includes a lot of unique values
- Examples of ticket value: "SOTon/OQ 392089", "PC 17608", and "34765"

Title:

- Title is an important feature to determine passenger status
- Name includes Name, Title and Last Name
- There is no direct access to title
- Examples of Name value: "Rice, Master. Eugene" and "Najib, Miss.Adele Kiamie "



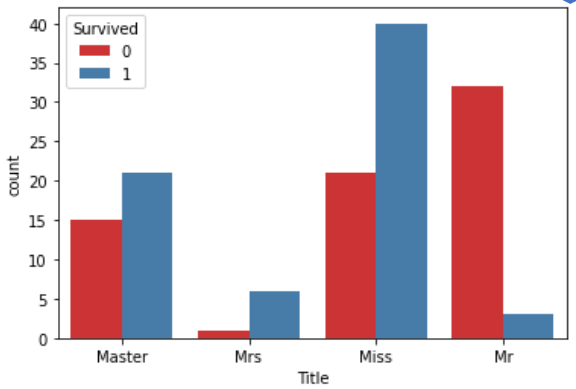
# Feature Engineering

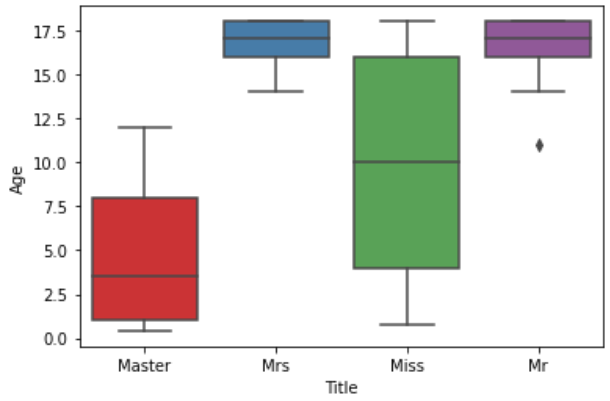
Passenger with title Mrs., Miss.& Master were saved in better percentage respectively

Extract numeric and non-numeric parts from Ticket and create two new features based on them

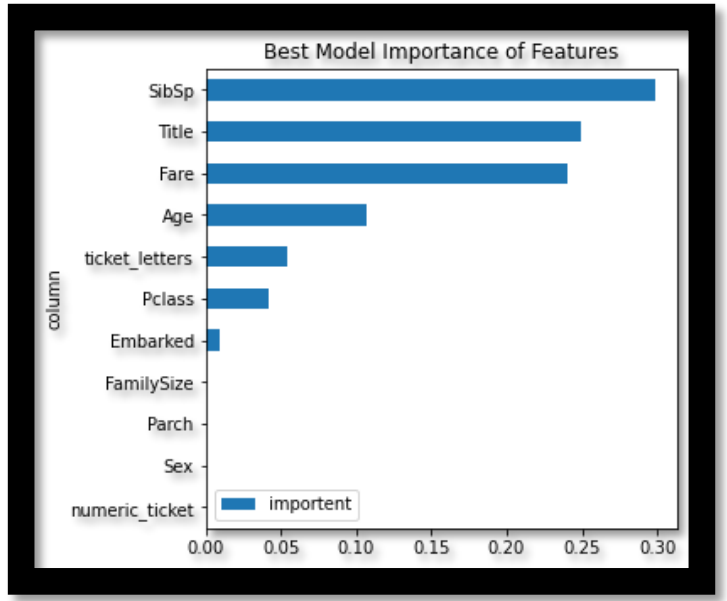
Extract title from Name column to create Title column

Most passenger with Title Mr. did not survive




- Using Machine learning Models: Random Forest and Decision Tree
- Applying different set of features to see the effectiveness of the new feature
- Feature importance plot shows title as second important feature
- ticket_ letters impact the survival rate better than Sex, Pclass and Embarked
- The title indicates a combination of 2 age ranges and gender here, so it takes place above Age and Sex.



- For the future analysis: we need more samples to estimate better set of feature
- Finding other possible features
- Find the best set of features because having a more extensive set doesn’t guarantee better efficiencyRescue policies in crisis show discrimination against passengers based on their ability to afford rooms or what title they had. It is a human right that everyone has equal rights with others regardless of social status