

```
In [1]: # importing the packages
import os

import pandas as pd
import numpy as np
import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.impute import SimpleImputer
from sklearn.preprocessing import OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import cross_val_score
```

```
In [3]: # Loading the dataset:
train_df = pd.read_csv("train (3).csv")
test_df = pd.read_csv("test (3).csv")
```

```
In [4]: print(train_df)
```

Passenger ID	Name		Sex	Age	Survived?
	Parch	Ticket	Fare	Cabin	Embarked
887	Graham, Miss. Margaret Edith		female	19.0	0
888	Johnston, Miss. Catherine Helen "Carrie"		female	NaN	1
889		Behr, Mr. Karl Howell	male	26.0	0
890		Dooley, Mr. Patrick	male	32.0	0
..
886	0	211536	13.0000	NaN	S
887	0	112053	30.0000	B42	S
888	2	W./C. 6607	23.4500	NaN	S
889	0	111369	30.0000	C148	C
890	0	370376	7.7500	NaN	Q

[891 rows x 12 columns]

```
In [5]: print(test_df)
```

	PassengerId	Pclass		Name	\				
0	892	3		Kelly, Mr. James					
1	893	3		Wilkes, Mrs. James (Ellen Needs)					
2	894	2		Myles, Mr. Thomas Francis					
3	895	3		Wirz, Mr. Albert					
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)						
..					
413	1305	3		Spector, Mr. Woolf					
414	1306	1		Oliva y Ocana, Dona. Fermina					
415	1307	3		Saether, Mr. Simon Sivertsen					
416	1308	3		Ware, Mr. Frederick					
417	1309	3		Peter, Master. Michael J					
	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	
0	male	34.5	0	0	330911	7.8292	NaN	Q	
1	female	47.0	1	0	363272	7.0000	NaN	S	
2	male	62.0	0	0	240276	9.6875	NaN	Q	
3	male	27.0	0	0	315154	8.6625	NaN	S	
4	female	22.0	1	1	3101298	12.2875	NaN	S	
..
413	male	NaN	0	0	A.5. 3236	8.0500	NaN	S	
414	female	39.0	0	0	PC 17758	108.9000	C105	C	
415	male	38.5	0	0	SOTON/O.Q.	3101262	7.2500	NaN	S
416	male	NaN	0	0		359309	8.0500	NaN	S
417	male	NaN	1	1		2668	22.3583	NaN	C

[418 rows x 11 columns]

```
In [6]: train_df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel))	female	35.0	1	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S

```
In [7]: train_df.tail()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536	13.00	NaN	S
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.00	B42	S
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607	23.45	NaN	S
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.00	C148	C
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376	7.75	NaN	Q

```
In [9]: test_df.head()
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	0	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	1	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	0	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	0	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	1	1	3101298	12.2875	NaN	S

```
In [10]: test_df.tail()
```

	PassengerId	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
413	1305	3	Spector, Mr. Woolf	male	NaN	0	0	A.5. 3236	8.0500	NaN	S
414	1306	1	Oliva y Ocana, Dona. Fermina	female	39.0	0	0	PC 17758	108.9000	C105	C
415	1307	3	Saether, Mr. Simon Sivertsen	male	38.5	0	0	SOTON/O.Q. 3101262	7.2500	NaN	S
416	1308	3	Ware, Mr. Frederick	male	NaN	0	0	359309	8.0500	NaN	S
417	1309	3	Peter, Master. Michael J	male	NaN	1	1	2668	22.3583	NaN	C

```
In [15]: # Exploratory Data Analysis(EDA)  
#checking missing values
```

```
def plot_missing_data(dataset, title):  
    fig, ax = plt.subplots(figsize=(5,5))  
    plt.title(title)  
    sns.heatmap(dataset.isnull(), cbar=False)
```

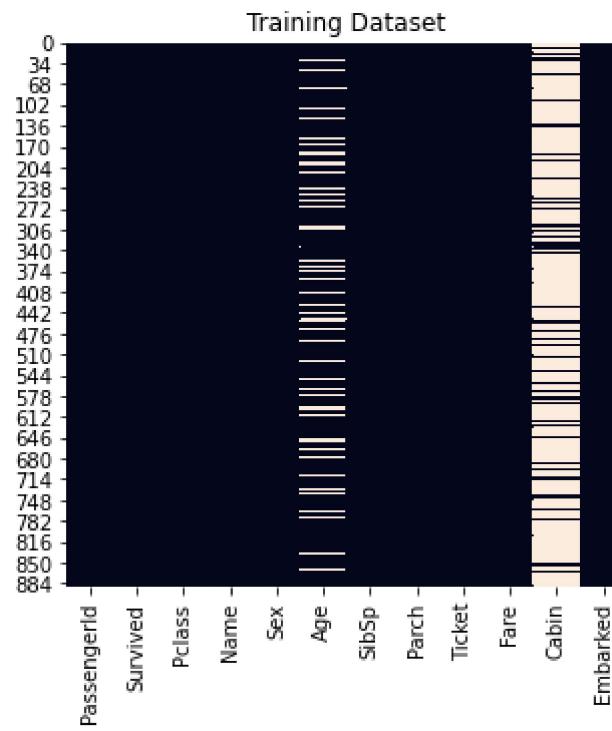
```
In [16]: train_df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 12 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PassengerId  891 non-null    int64  
 1   Survived     891 non-null    int64  
 2   Pclass       891 non-null    int64  
 3   Name         891 non-null    object  
 4   Sex          891 non-null    object  
 5   Age          714 non-null    float64 
 6   SibSp        891 non-null    int64  
 7   Parch        891 non-null    int64  
 8   Ticket       891 non-null    object  
 9   Fare          891 non-null    float64 
 10  Cabin         204 non-null    object  
 11  Embarked     889 non-null    object  
dtypes: float64(2), int64(5), object(5)
memory usage: 83.7+ KB
```

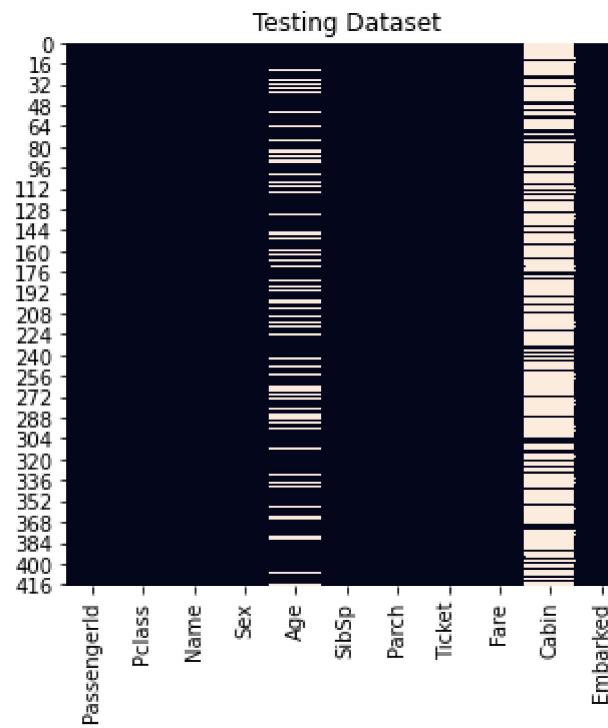
In [17]: test_df.info()

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 11 columns):
 #   Column      Non-Null Count  Dtype  
--- 
 0   PassengerId  418 non-null    int64  
 1   Pclass        418 non-null    int64  
 2   Name          418 non-null    object  
 3   Sex           418 non-null    object  
 4   Age           332 non-null    float64 
 5   SibSp         418 non-null    int64  
 6   Parch         418 non-null    int64  
 7   Ticket        418 non-null    object  
 8   Fare          417 non-null    float64 
 9   Cabin         91 non-null    object  
 10  Embarked      418 non-null    object  
dtypes: float64(2), int64(4), object(5)
memory usage: 36.0+ KB
```

```
In [20]: plot_missing_data(train_df, "Training Dataset")
```



```
In [23]: plot_missing_data(test_df, "Testing Dataset")
```



```
In [24]: # observations
def bar_chart_stacked(dataset, feature, stacked = True):
    survived = dataset[dataset['Survived']==1][feature].value_counts()
    dead = dataset[dataset['Survived']==0][feature].value_counts()
    df_survived_dead = pd.DataFrame([survived,dead])
    df_survived_dead.index = ['Passengers Survived','Passengers Died']
    ax = df_survived_dead.plot(kind='bar',stacked=stacked, figsize=(5,5))
```

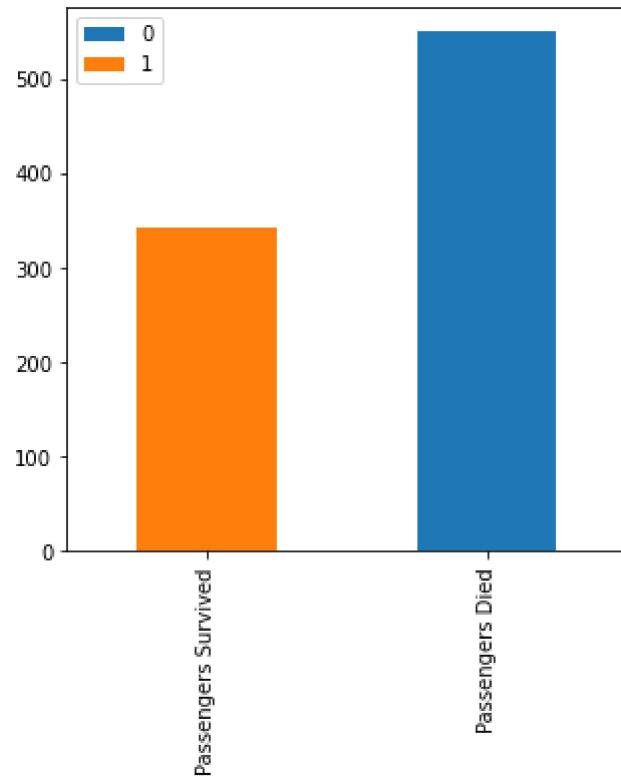
```
In [25]: # Analysed Feature Survival
train_df['Survived'].value_counts()
```

```
Out[25]: 0    549
1    342
Name: Survived, dtype: int64
```

```
In [26]: train_df['Survived'].value_counts(normalize=True)
```

```
Out[26]: 0    0.616162  
1    0.383838  
Name: Survived, dtype: float64
```

```
In [29]: # data visualization  
bar_chart_stacked(train_df, "Survived")
```



Observation:

As expected the majority of passengers in the training data died. Only 38% survived the disaster. So the training data suffers from data imbalance but it is not severe which is why I will not consider techniques like sampling to tackle the imbalance.

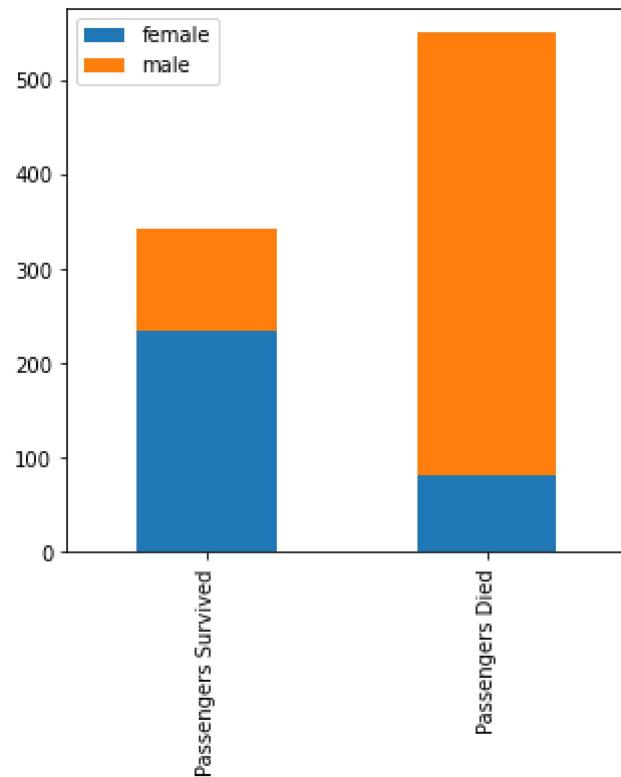
```
In [30]: # Analyse Feature Sex:  
train_df['Sex'].value_counts().to_frame()
```

```
Out[30]:  
Sex  
---  
male 577  
female 314
```

Observation:

As the majority of passenger were we could infer that the majority of survivors were men.

```
In [31]: bar_chart_stacked(train_df, "Sex")
```



```
In [32]: train_df.groupby('Sex').Survived.mean()
```

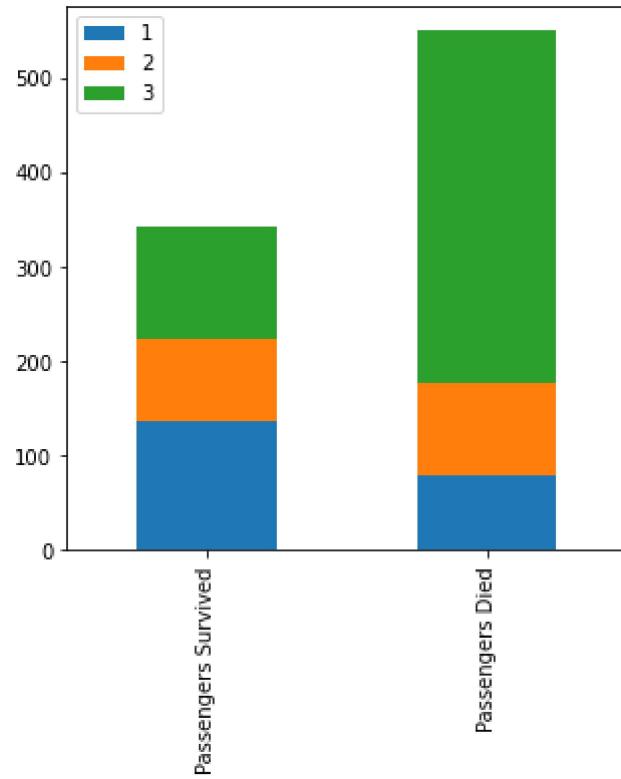
```
Out[32]: Sex  
female    0.742038  
male      0.188908  
Name: Survived, dtype: float64
```

Observation:

Here, we can see even though the majority of the passenger were male, the majority of survivors were female. The key observation here is that the survival rate for female passengers is 4 times higher than the survival rate of male passengers.

Analyze Feature Pclass:

```
In [33]: bar_chart_stacked(train_df, 'Pclass')
```



```
In [34]: pd.pivot_table(train_df, index = 'Survived', columns = 'Pclass', values = 'Ticket' ,aggfunc ='count')
```

Out[34]:

	Pclass	1	2	3
Survived				
0	80	97	372	
1	136	87	119	

```
In [35]: train_df.groupby(['Pclass']).Survived.mean().to_frame()
```

```
Out[35]:
```

Pclass	Survived
1	0.629630
2	0.472826
3	0.242363

Observation:

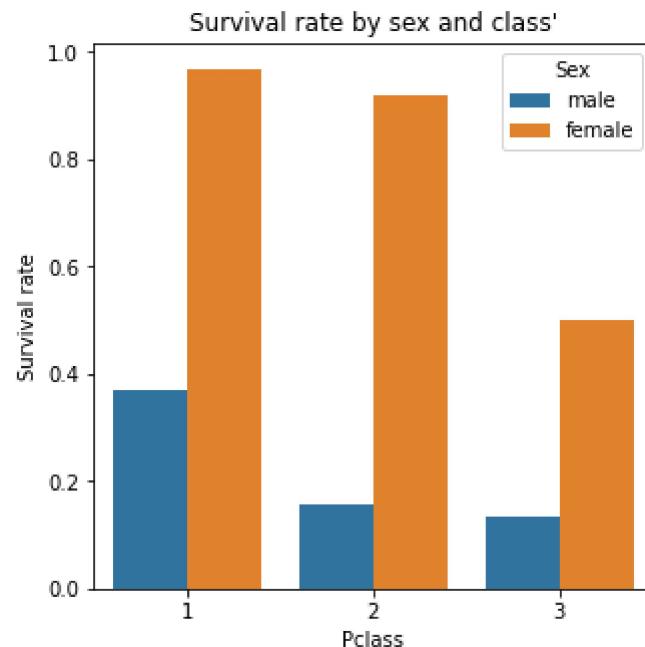
From the plots and tables above it becomes clear that the Pclass is an important factor to consider.

Most passenger had class 3 tickets, yet only 24% of class 3 passengers survived. Almost 63% of the passenger from class 1 survived. Also approx 50% of the class 2 passenger survived. One questions that comes into mind is:

Is the class or sex the deciding factor?

```
In [36]: def bar_chart_compare(dataset, feature1, feature2=None, title = "Survival rate by sex and class"):  
    plt.figure(figsize = [5,5])  
    plt.title(title)  
    g = sns.barplot(x=feature1, y='Survived', hue=feature2, ci=None, data=dataset).set_ylabel('Survival rate')
```

```
In [38]: bar_chart_compare(train_df, "Pclass", "Sex")
```



```
In [39]: pd.pivot_table(train_df, index = 'Survived', columns = ['Pclass', "Sex"], values = 'Ticket' ,aggfunc = 'count')
```

```
Out[39]:
```

Survived	Pclass		1		2		3	
	Sex	female	male	female	male	female	male	
0	3	77	6	91	72	300		
1	91	45	70	17	72	47		

```
In [40]: train_df.groupby(['Pclass', "Sex"]).Survived.mean().to_frame()
```

```
Out[40]:
```

		Survived
Pclass	Sex	
1	female	0.968085
	male	0.368852
2	female	0.921053
	male	0.157407
3	female	0.500000
	male	0.135447

Observation:

Here, we can clearly see that the question was justified. Irrespective of the class the most important factor when it comes to surviving was gender. (At least between Sex and Pclass) However, men in class 1 had a significantly higher chance of survival when they bought class 1 tickets. This just shows to say that we should keep both features as both yield insightful information that should help our model.

Survival Rate females 1. Class: 96,8% Survival Rate females 2. Class: 92,1% Survival Rate females 3. Class: 50%

Survival Rate male 1. Class: 36.8% (still significantly lower than 3. class females)

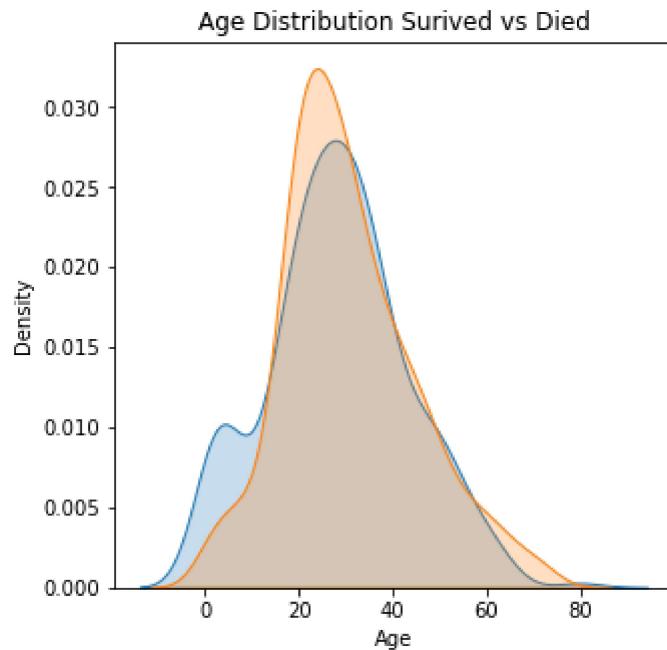
Analyze Feature Age

```
In [42]: def plot_distribution(dataset, feature, title, bins = 30, hist = True, fsize = (5,5)):  
    fig, ax = plt.subplots(figsize=fsize)  
    ax.set_title(title)  
    sns.distplot(train_df[feature], color='g', bins=bins, ax=ax)
```

```
In [43]: def plot_kernel_density_estimate_survivors(dataset, feature1, title, fsize = (5,5)):
    fig, ax = plt.subplots(figsize=fsize)
    ax.set_title(title)
    sns.kdeplot(dataset[feature1].loc[train_df["Survived"] == 1],
                shade= True, ax=ax, label='Survived').set_xlabel(feature1)
    sns.kdeplot(dataset[feature1].loc[train_df["Survived"] == 0],
                shade=True, ax=ax, label="Died")
```

plot_distribution(train_df, "Age", "Age Distribution Passengers")

```
In [45]: plot_kernel_density_estimate_survivors(train_df, "Age", "Age Distribution Survived vs Died")
```



Observation:

The Histogram above shows that age follows a fairly normal distribution.

Also investigating the kernel density estimate does not provide additional information except a raise in survivors at a very young age

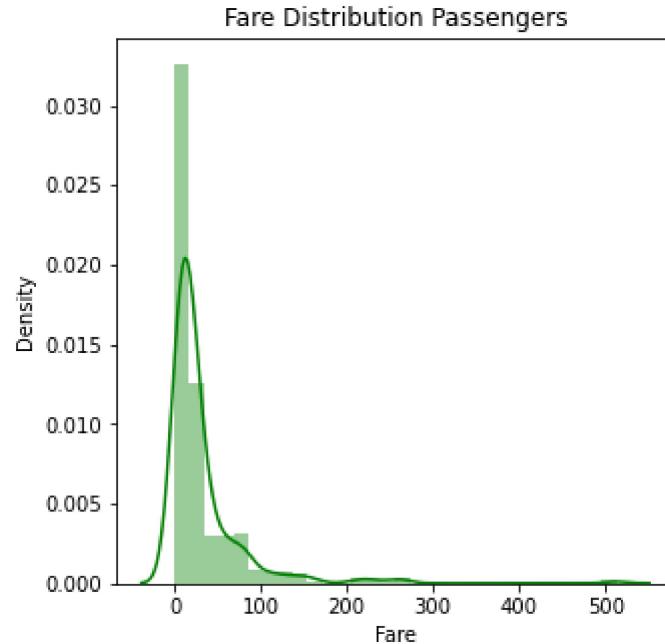
Analyze Features Age and Pclass together

```
In [46]: train_df.Fare.describe()
```

```
Out[46]: count    891.000000
mean      32.204208
std       49.693429
min       0.000000
25%      7.910400
50%     14.454200
75%     31.000000
max     512.329200
Name: Fare, dtype: float64
```

```
In [47]: plot_distribution(train_df, "Fare", "Fare Distribution Passengers")
```

```
C:\Users\karth\anaconda3\lib\site-packages\seaborn\distributions.py:2557: FutureWarning: `distplot` is a deprecated function and will be removed in a future version. Please adapt your code to use either `displot` (a figure-level function with similar flexibility) or `histplot` (an axes-level function for histograms).  
warnings.warn(msg, FutureWarning)
```



Observation:

Fare does not follow a normal distribution and has a huge spike at the price range [0-100\$].

The distribution is skewed to the left with 75% of the fare paid under 31% and a max paid fare of 512\$. Depending on the model that I'm going to use it might make sense to normalize this feature. However, this aspect will be tackled later in the feature engineering section.

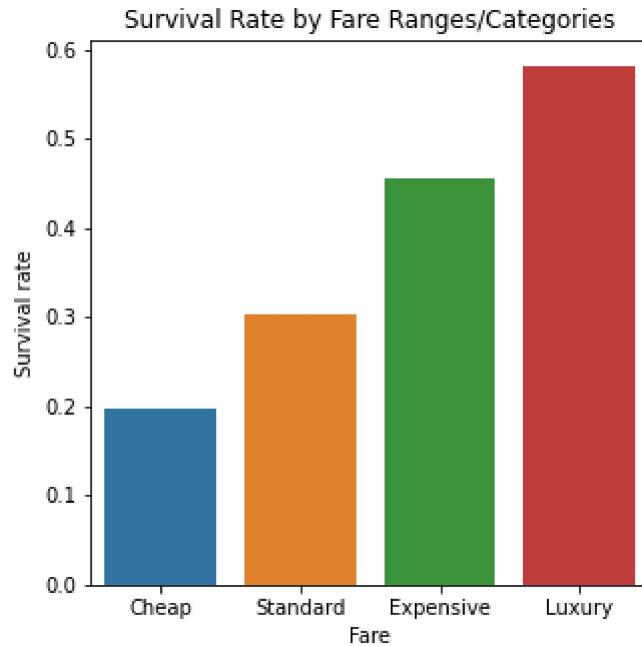
To better understand how this feature influences the survival rate, we could plot bar plots of Fare vs Survived. However, due to the large range of fares such as plot would not be useful for inferring useful information.

A more suited visualization would be to combine fares into categories and then plot the categories vs Survived.

```
In [48]: def plot_quartiles(dataset, feature, title, categories):
    fig, axarr = plt.subplots(figsize=(5,5))
    fare_ranges = pd.qcut(dataset[feature], len(categories), labels = categories) #. [0, .25, .5, .75, 1.]
    axarr.set_title(title)
    sns.barplot(x=fare_ranges, y=dataset.Survived, ci=None, ax=axarr).set_ylabel('Survival rate')
```

```
In [49]: categories = ['Cheap', 'Standard', 'Expensive', 'Luxury']

plot_quartiles(train_df, "Fare", "Survival Rate by Fare Ranges/Categories", categories)
```



Observation:

As we can see the likelihood of survival is definitely influenced by the price paid.

Cheap (0-25% of max Price): Survival Rate = 0.2 (approx)
Cheap (25%-50% of max Price): Survival Rate = 0.3 (approx)
Cheap (50%-75% of max Price): Survival Rate = 0.45 (approx)
Cheap (75%-100% of max Price): Survival Rate = 0.55 (approx)
Additionally, we can investigate the relationship between fare, sex and survived to further understand the importance of the feature.

```
In [52]: plot_swarm_survivors(train_df, "Fare", "Sex", "Survivor Swarmplot for Age and Gender")
```

```
NameError Traceback (most recent call last)
<ipython-input-52-db84dd8c797a> in <module>
----> 1 plot_swarm_survivors(train_df, "Fare", "Sex", "Survivor Swarmplot for Age and Gender")

NameError: name 'plot_swarm_survivors' is not defined
```

Observation:

Here, some important observations can be made:

Irrespective of gender all passengers with a fare above 500\$ survived.

All male passenger that paid between 200-300\$ died

All female passenger that paid between 200-300\$ died

This could be a feature a classifier might pick up.

One thing that caught my attention is that the minimum fare paid was 0.0 \$.

This seems highly unlikely. We can investigate who these people were:

```
In [51]: train_df.loc[train_df.Fare==0]
```

Out[51]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
179	180	0	3	Leonard, Mr. Lionel	male	36.0	0	0	LINE	0.0	NaN	S
263	264	0	1	Harrison, Mr. William	male	40.0	0	0	112059	0.0	B94	S
271	272	1	3	Tornquist, Mr. William Henry	male	25.0	0	0	LINE	0.0	NaN	S
277	278	0	2	Parkes, Mr. Francis "Frank"	male	NaN	0	0	239853	0.0	NaN	S
302	303	0	3	Johnson, Mr. William Cahoon Jr	male	19.0	0	0	LINE	0.0	NaN	S
413	414	0	2	Cunningham, Mr. Alfred Fleming	male	NaN	0	0	239853	0.0	NaN	S
466	467	0	2	Campbell, Mr. William	male	NaN	0	0	239853	0.0	NaN	S
481	482	0	2	Frost, Mr. Anthony Wood "Archie"	male	NaN	0	0	239854	0.0	NaN	S
597	598	0	3	Johnson, Mr. Alfred	male	49.0	0	0	LINE	0.0	NaN	S
633	634	0	1	Parr, Mr. William Henry Marsh	male	NaN	0	0	112052	0.0	NaN	S
674	675	0	2	Watson, Mr. Ennis Hastings	male	NaN	0	0	239856	0.0	NaN	S
732	733	0	2	Knight, Mr. Robert J	male	NaN	0	0	239855	0.0	NaN	S
806	807	0	1	Andrews, Mr. Thomas Jr	male	39.0	0	0	112050	0.0	A36	S
815	816	0	1	Fry, Mr. Richard	male	NaN	0	0	112058	0.0	B102	S
822	823	0	1	Reuchlin, Jonkheer. John George	male	38.0	0	0	19972	0.0	NaN	S

Observation:

15 Passengers paid no fare. As this is nonsensical I decide to replace 0 values with nan and then later to think about how to impute these values.

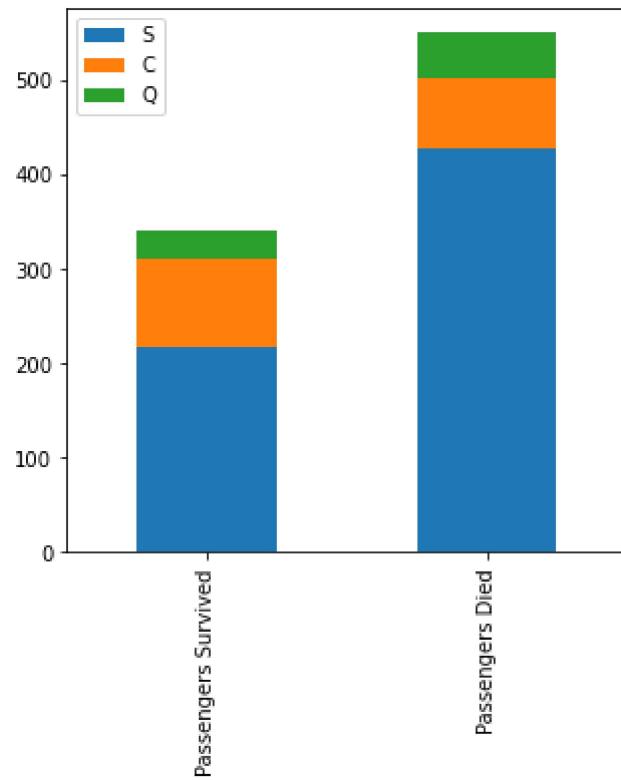
```
In [55]: # Replace Fare == 0 with nan
train_df.loc[train_df['Fare'] == 0, 'Fare'] = np.NaN
test_df.loc[train_df['Fare'] == 0, 'Fare'] = np.NaN
```

Analyze Feature Embarked

```
In [56]: def show_countplot(dataset, feature, title, fsize = (5,5)):
    fig, ax = plt.subplots(figsize=fsize)
    sns.countplot(dataset[feature], ax=ax).set_title(title)

def show_compare_countplot(dataset, feature1, feature2, title):
    fig, ax = plt.subplots(figsize=(5,5))
    p = sns.countplot(x = feature1, hue = feature2, data = dataset, ax=ax).set_title(title)
```

```
In [57]: bar_chart_stacked(train_df, 'Embarked')
```

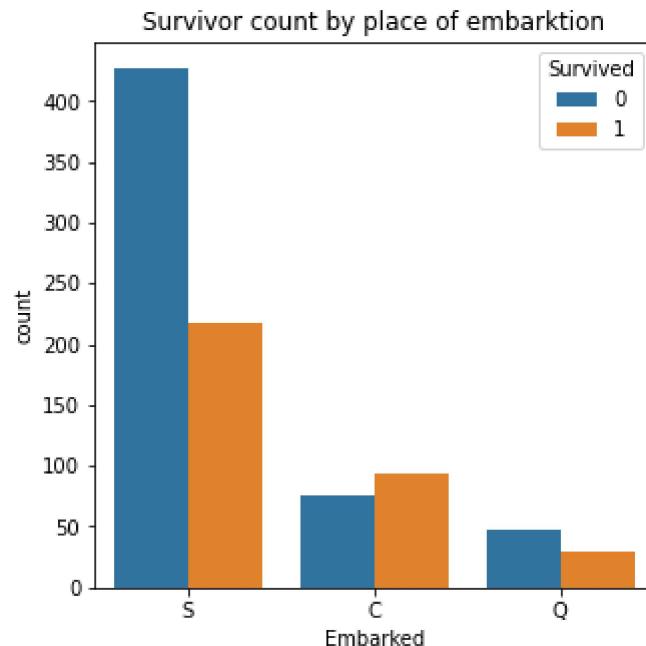


```
In [58]: train_df['Embarked'].value_counts().to_frame()
```

```
Out[58]: Embarked
```

S	644
C	168
Q	77

```
In [59]: show_compare_countplot(train_df, "Embarked", "Survived", "Survivor count by place of embarktion")
```



```
In [60]: pd.pivot_table(train_df, index = 'Survived', columns = 'Embarked', values = 'Ticket' ,aggfunc ='count')
```

```
Out[60]: Embarked  C   Q   S
```

Survived			
	C	Q	S
0	75	47	427
1	93	30	217

```
In [61]: train_df.groupby(['Embarked']).Survived.mean().to_frame()
```

```
Out[61]:
```

Survived

Embarked	Survived
C	0.553571
Q	0.389610
S	0.336957

Observation:

There are three possible values for the Embarked Feature:

Southampton, Cherbourg, and Queenstown. We can see that the majority of passengers embarked from Southampton. However, only 33% survived the sinking of the titanic.

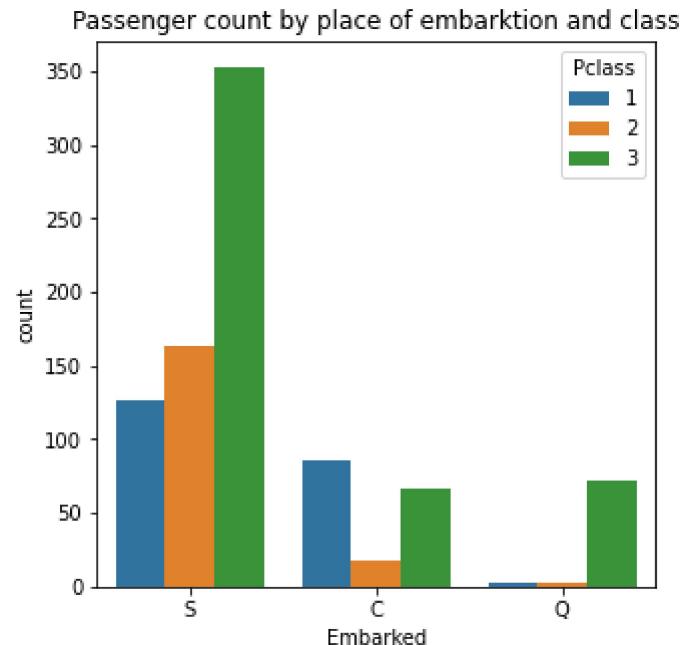
The highest survival rate of 55% is in the group of passengers that embarked from Cherbourg

Using common sense I would not expect that the place of border changes the likelihood of survival.

Why is the likelihood of survival higher at Cherbourg? As we know from above a better class increases the survival rate drastically. One indicator may be the percentage of 1. Class passengers that embarked at Cherbourg:

Analyze Features Embarked & Pclass together

```
In [62]: show_compare_countplot(train_df, "Embarked", "Pclass", "Passenger count by place of embarktion and class")
```



Observation:

As expected twice as many male passengers embarked from Southam were roughly the same number of male and female passengers embarked from Queenstown. This just shows the importance of the Sex feature.

Analyze Feature SibSp

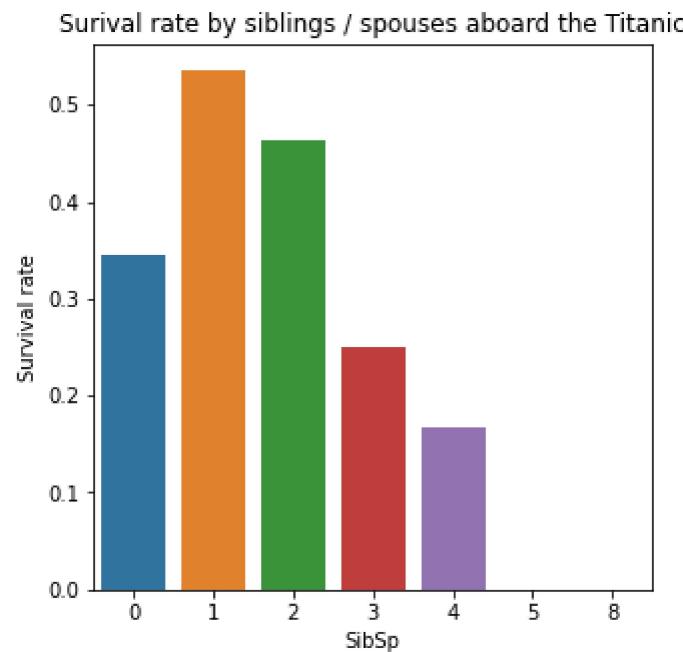
SibSp is the number of siblings or spouses of a person aboard the Titanic

```
In [63]: train_df['SibSp'].value_counts().to_frame()
```

Out[63]:

SibSp	
0	608
1	209
2	28
4	18
3	16
8	7
5	5

```
In [64]: bar_chart_compare(train_df, "SibSp", title = "Survival rate by siblings / spouses aboard the Titanic")
```



```
In [65]: train_df.groupby(['SibSp']).Survived.mean().to_frame()
```

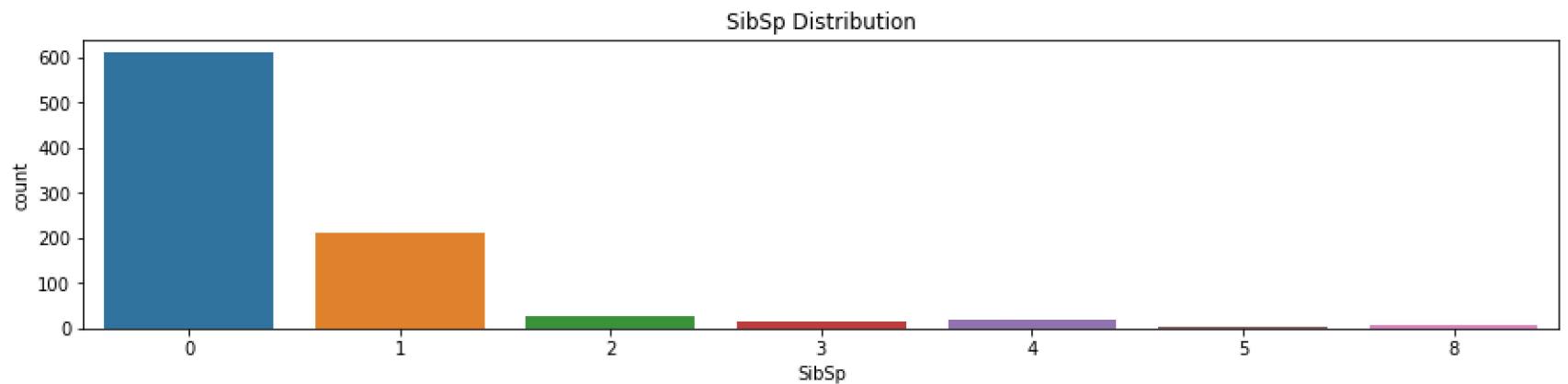
```
Out[65]:      Survived
```

SibSp	Survived
0	0.345395
1	0.535885
2	0.464286
3	0.250000
4	0.166667
5	0.000000
8	0.000000

```
In [66]: show_countplot(train_df, "SibSp", 'SibSp Distribution', (15,3))
```

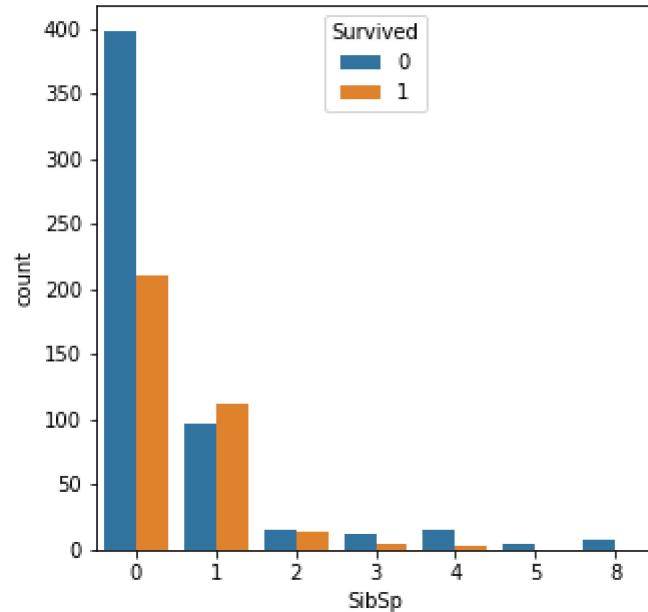
C:\Users\karth\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
warnings.warn(
```



```
In [67]: show_compare_countplot(train_df, "SibSp", "Survived", "Survivor count by number of siblings / spouses aboard the
```

Survivor count by number of siblings / spouses aboard the Titanic



Observation:

The majority traveled alone. It seems that the more Sibling a passenger has the less chance of survival there is.

Analyze Feature Parch

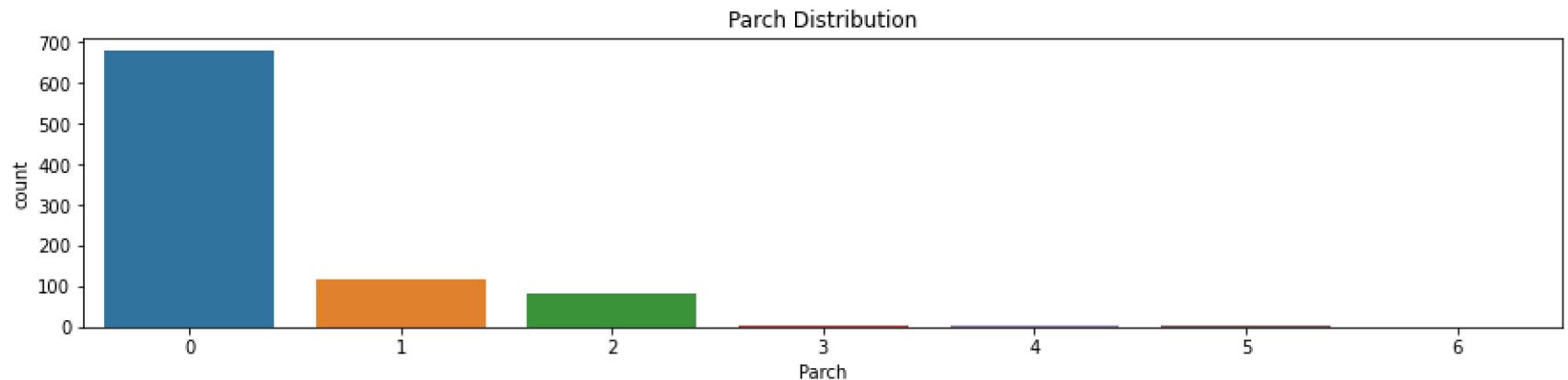
Analyze Feature Parch Similar to the SibSp column, this feature contains the number of parents or children each passenger was traveling with.

Here we draw the same conclusions as SibSp; we see again that small families had more chances to survive than bigger ones and passengers who traveled alone.

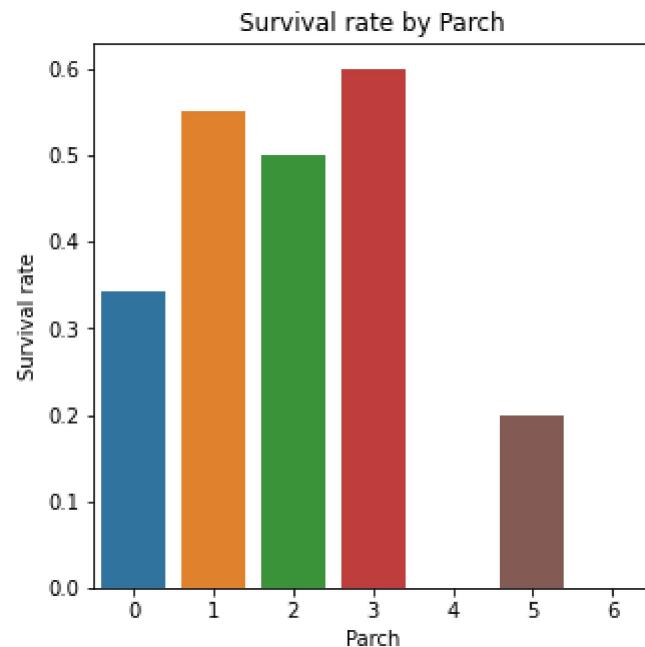
```
In [68]: show_countplot(train_df, "Parch", 'Parch Distribution', (15,3))
```

C:\Users\karth\anaconda3\lib\site-packages\seaborn_decorators.py:36: FutureWarning: Pass the following variable as a keyword arg: x. From version 0.12, the only valid positional argument will be `data`, and passing other arguments without an explicit keyword will result in an error or misinterpretation.

```
    warnings.warn(
```



```
In [69]: bar_chart_compare(train_df, "Parch", title = "Survival rate by Parch")
```

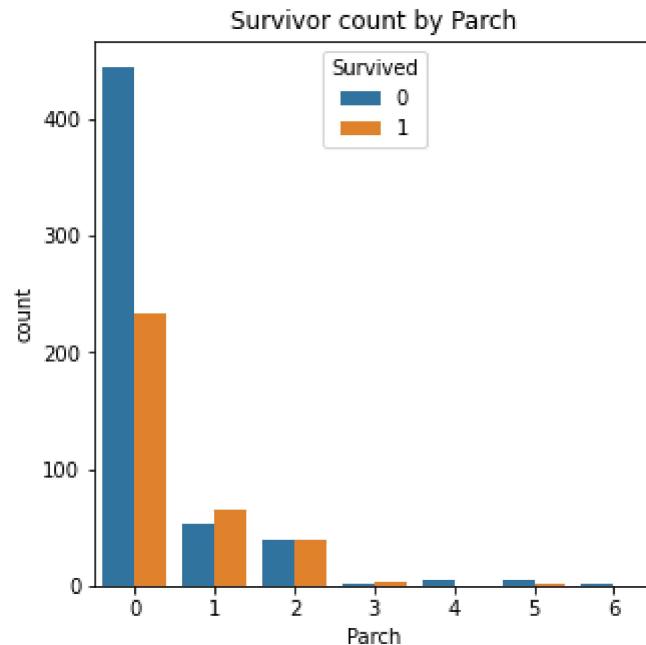


```
In [70]: train_df.groupby(['Parch']).Survived.mean().to_frame()
```

Out[70]: **Survived**

Parch	Survived
0	0.343658
1	0.550847
2	0.500000
3	0.600000
4	0.000000
5	0.200000
6	0.000000

```
In [71]: show_compare_countplot(train_df, "Parch", "Survived", "Survivor count by Parch")
```



Observation:

Similar to the SibSp column, this feature contains the number of parents or children each passenger was traveling with. Here we draw the same conclusions as SibSp; we see again that small families had more chances to survive than bigger ones and passengers who traveled alone.

Both features do not seem to be very informative, however, higher abstraction levels trees might pick up certain groups. For faster training, it might be beneficial to combine both features into 1 and create a family size feature.

Feature engineering

Feature Name:

```
In [72]: pd.unique(train_df['Name'])
```

```
Out[72]: array(['Braund, Mr. Owen Harris',
   'Cumings, Mrs. John Bradley (Florence Briggs Thayer)',
   'Heikkinen, Miss. Laina',
   'Futrelle, Mrs. Jacques Heath (Lily May Peel)',
   'Allen, Mr. William Henry', 'Moran, Mr. James',
   'McCarthy, Mr. Timothy J', 'Palsson, Master. Gosta Leonard',
   'Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)',
   'Nasser, Mrs. Nicholas (Adele Achem)',
   'Sandstrom, Miss. Marguerite Rut', 'Bonnell, Miss. Elizabeth',
   'Saundercock, Mr. William Henry', 'Andersson, Mr. Anders Johan',
   'Vestrom, Miss. Hulda Amanda Adolfina',
   'Hewlett, Mrs. (Mary D Kingcome)', 'Rice, Master. Eugene',
   'Williams, Mr. Charles Eugene',
   'Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)',
   'Masselmani, Mrs. Fatima', 'Fynney, Mr. Joseph J',
   'Beesley, Mr. Lawrence', 'McGowan, Miss. Anna "Annie"',
   'Sloper, Mr. William Thompson', 'Palsson, Miss. Torborg Danira',
   'Asplund, Mrs. Carl Oscar (Selma Augusta Emilia Johansson)',
   'Emir, Mr. Farred Chehab', 'Fortune, Mr. Charles Alexander',
   'O\'Dwyer, Miss. Ellen "Nellie"', 'Todoroff, Mr. Lalio',
   'Uruchurtu, Don. Manuel E',
   'Spencer, Mrs. William Augustus (Marie Eugenie)',
   'Glynn, Miss. Mary Agatha', 'Wheadon, Mr. Edward H',
   'Meyer, Mr. Edgar Joseph', 'Holverson, Mr. Alexander Oskar',
   'Mamee, Mr. Hanna', 'Cann, Mr. Ernest Charles',
   'Vander Planke, Miss. Augusta Maria',
   'Nicola-Yarred, Miss. Jamila',
   'Ahlin, Mrs. Johan (Johanna Persdotter Larsson)',
   'Turpin, Mrs. William John Robert (Dorothy Ann Wonnacott)',
   'Kraeff, Mr. Theodor', 'Laroche, Miss. Simonne Marie Anne Andree',
   'Devaney, Miss. Margaret Delia', 'Rogers, Mr. William John',
   'Lennon, Mr. Denis', "O'Driscoll, Miss. Bridget",
   'Samaan, Mr. Youssef',
   'Arnold-Franchi, Mrs. Josef (Josefine Franchi)',
   'Panula, Master. Juha Niilo', 'Nosworthy, Mr. Richard Cater',
   'Harper, Mrs. Henry Sleeper (Myra Haxton)',
   'Faunthorpe, Mrs. Lizzie (Elizabeth Anne Wilkinson)',
   'Ostby, Mr. Engelhart Cornelius', 'Woolner, Mr. Hugh',
   'Rugg, Miss. Emily', 'Novel, Mr. Mansouer',
   'West, Miss. Constance Mirium',
   'Goodwin, Master. William Frederick', 'Sirayanian, Mr. Orsen',
   'Icard, Miss. Amelie', 'Harris, Mr. Henry Birkhardt',
   'Skoog, Master. Harald', 'Stewart, Mr. Albert A',
```

'Moubarek, Master. Gerios', 'Nye, Mrs. (Elizabeth Ramell)',
'Crease, Mr. Ernest James', 'Andersson, Miss. Erna Alexandra',
'Kink, Mr. Vincenz', 'Jenkin, Mr. Stephen Curnow',
'Goodwin, Miss. Lillian Amy', 'Hood, Mr. Ambrose Jr',
'Chronopoulos, Mr. Apostolos', 'Bing, Mr. Lee',
'Moen, Mr. Sigurd Hansen', 'Stanoff, Mr. Ivan',
'Moutal, Mr. Rahamin Haim', 'Caldwell, Master. Alden Gates',
'Dowdell, Miss. Elizabeth', 'Waelens, Mr. Achille',
'Sheerlinck, Mr. Jan Baptist', 'McDermott, Miss. Brigdet Delia',
'Carrau, Mr. Francisco M', 'Ilett, Miss. Bertha',
'Backstrom, Mrs. Karl Alfred (Maria Mathilda Gustafsson)',
'Ford, Mr. William Neal', 'Slocovski, Mr. Selman Francis',
'Fortune, Miss. Mabel Helen', 'Celotti, Mr. Francesco',
'Christmann, Mr. Emil', 'Andreasson, Mr. Paul Edvin',
'Chaffee, Mr. Herbert Fuller', 'Dean, Mr. Bertram Frank',
'Coxon, Mr. Daniel', 'Shorney, Mr. Charles Joseph',
'Goldschmidt, Mr. George B', 'Greenfield, Mr. William Bertram',
'Doling, Mrs. John T (Ada Julia Bone)', 'Kantor, Mr. Sinai',
'Petranec, Miss. Matilda', 'Petroff, Mr. Pastcho ("Pentcho")',
'White, Mr. Richard Frasar', 'Johansson, Mr. Gustaf Joel',
'Gustafsson, Mr. Anders Vilhelm', 'Mionoff, Mr. Stoytcho',
'Salkjelsvik, Miss. Anna Kristine', 'Moss, Mr. Albert Johan',
'Rekic, Mr. Tido', 'Moran, Miss. Bertha',
'Porter, Mr. Walter Chamberlain', 'Zabour, Miss. Hileni',
'Barton, Mr. David John', 'Jussila, Miss. Katriina',
'Attalah, Miss. Malake', 'Pekoniemi, Mr. Edvard',
'Connors, Mr. Patrick', 'Turpin, Mr. William John Robert',
'Baxter, Mr. Quigg Edmond', 'Andersson, Miss. Ellis Anna Maria',
'Hickman, Mr. Stanley George', 'Moore, Mr. Leonard Charles',
'Nasser, Mr. Nicholas', 'Webber, Miss. Susan',
'White, Mr. Percival Wayland', 'Nicola-Yarred, Master. Elias',
'McMahon, Mr. Martin', 'Madsen, Mr. Fridtjof Arne',
'Peter, Miss. Anna', 'Ekstrom, Mr. Johan', 'Drazenoic, Mr. Jozef',
'Coelho, Mr. Domingos Fernandeo',
'Robins, Mrs. Alexander A (Grace Charity Laury)',
'Weisz, Mrs. Leopold (Mathilde Francoise Pede)',
'Sobey, Mr. Samuel James Hayden', 'Richard, Mr. Emile',
'Newsom, Miss. Helen Monypeny', 'Futrelle, Mr. Jacques Heath',
'Osen, Mr. Olaf Elon', 'Giglio, Mr. Victor',
'Boulos, Mrs. Joseph (Sultana)', 'Nysten, Miss. Anna Sofia',
'Hakkarainen, Mrs. Pekka Pietari (Elin Matilda Dolck)',
'Burke, Mr. Jeremiah', 'Andrew, Mr. Edgardo Samuel',
'Nicholls, Mr. Joseph Charles',

'Andersson, Mr. August Edvard ("Wennerstrom")',
'Ford, Miss. Robina Maggie "Ruby"',
'Navratil, Mr. Michel ("Louis M Hoffman")',
'Byles, Rev. Thomas Roussel Davids', 'Bateman, Rev. Robert James',
'Pears, Mrs. Thomas (Edith Wearne)', 'Meo, Mr. Alfonzo',
'van Billiard, Mr. Austin Blyler', 'Olsen, Mr. Ole Martin',
'Williams, Mr. Charles Duane', 'Gilnagh, Miss. Katherine "Katie"',
'Corn, Mr. Harry', 'Smiljanic, Mr. Mile',
'Sage, Master. Thomas Henry', 'Cribb, Mr. John Hatfield',
'Watt, Mrs. James (Elizabeth "Bessie" Inglis Milne)',
'Bengtsson, Mr. John Viktor', 'Calic, Mr. Jovo',
'Panula, Master. Eino Viljami',
'Goldsmith, Master. Frank John William "Frankie"',
'Chibnall, Mrs. (Edith Martha Bowerman)',
'Skoog, Mrs. William (Anna Bernhardina Karlsson)',
'Baumann, Mr. John D', 'Ling, Mr. Lee',
'Van der hoef, Mr. Wyckoff', 'Rice, Master. Arthur',
'Johnson, Miss. Eleanor Ileen', 'Sivola, Mr. Antti Wilhelm',
'Smith, Mr. James Clinch', 'Klasen, Mr. Klas Albin',
'Lefebre, Master. Henry Forbes', 'Isham, Miss. Ann Elizabeth',
'Hale, Mr. Reginald', 'Leonard, Mr. Lionel',
'Sage, Miss. Constance Gladys', 'Pernot, Mr. Rene',
'Asplund, Master. Clarence Gustaf Hugo',
'Becker, Master. Richard F', 'Kink-Heilmann, Miss. Luise Gretchen',
'Rood, Mr. Hugh Roscoe',
'O'Brien, Mrs. Thomas (Johanna "Hannah" Godfrey)',
'Romaine, Mr. Charles Hallace ("Mr C Rolmane")',
'Bourke, Mr. John', 'Turcin, Mr. Stjepan', 'Pinsky, Mrs. (Rosa)',
'Carbines, Mr. William',
'Andersen-Jensen, Miss. Carla Christine Nielsine',
'Navratil, Master. Michel M',
'Brown, Mrs. James Joseph (Margaret Tobin)',
'Lurette, Miss. Elise', 'Mernagh, Mr. Robert',
'Olsen, Mr. Karl Siegwart Andreas',
'Madigan, Miss. Margaret "Maggie"',
'Yrois, Miss. Henriette ("Mrs Harbeck")',
'Vande Walle, Mr. Nestor Cyriel', 'Sage, Mr. Frederick',
'Johanson, Mr. Jakob Alfred', 'Youseff, Mr. Gerious',
'Cohen, Mr. Gurshon "Gus"', 'Strom, Miss. Telma Matilda',
'Backstrom, Mr. Karl Alfred', 'Albimona, Mr. Nassef Cassem',
'Carr, Miss. Helen "Ellen"', 'Blank, Mr. Henry', 'Ali, Mr. Ahmed',
'Cameron, Miss. Clear Annie', 'Perkin, Mr. John Henry',
'Givard, Mr. Hans Kristensen', 'Kiernan, Mr. Philip',

'Newell, Miss. Madeleine', 'Honkanen, Miss. Eliina',
'Jacobsohn, Mr. Sidney Samuel', 'Bazzani, Miss. Albina',
'Harris, Mr. Walter', 'Sunderland, Mr. Victor Francis',
'Bracken, Mr. James H', 'Green, Mr. George Henry',
'Nenkoff, Mr. Christo', 'Hoyt, Mr. Frederick Maxfield',
'Berglund, Mr. Karl Ivar Sven', 'Mellors, Mr. William John',
'Lovell, Mr. John Hall ("Henry")', 'Fahlstrom, Mr. Arne Jonas',
'Lefebre, Miss. Mathilde',
'Harris, Mrs. Henry Birkhardt (Irene Wallach)',
'Larsson, Mr. Bengt Edvin', 'Sjostedt, Mr. Ernst Adolf',
'Asplund, Miss. Lillian Gertrud',
'Leyson, Mr. Robert William Norman',
'Harknett, Miss. Alice Phoebe', 'Hold, Mr. Stephen',
'Collyer, Miss. Marjorie "Lottie"',
'Pengelly, Mr. Frederick William', 'Hunt, Mr. George Henry',
'Zabour, Miss. Thamine', 'Murphy, Miss. Katherine "Kate"',
'Coleridge, Mr. Reginald Charles', 'Maenpaa, Mr. Matti Alexanderi',
'Attalah, Mr. Sleiman', 'Minahan, Dr. William Edward',
'Lindahl, Miss. Agda Thorilda Viktoria',
'Hamalainen, Mrs. William (Anna)', 'Beckwith, Mr. Richard Leonard',
'Carter, Rev. Ernest Courtenay', 'Reed, Mr. James George',
'Strom, Mrs. Wilhelm (Elna Matilda Persson)',
'Stead, Mr. William Thomas', 'Lobb, Mr. William Arthur',
'Rosblom, Mrs. Viktor (Helena Wilhelmina)',
'Touma, Mrs. Darwis (Hanne Youssef Razi)',
'Thorne, Mrs. Gertrude Maybelle', 'Cherry, Miss. Gladys',
'Ward, Miss. Anna', 'Parrish, Mrs. (Lutie Davis)',
'Smith, Mr. Thomas', 'Asplund, Master. Edvin Rojj Felix',
'Taussig, Mr. Emil', 'Harrison, Mr. William', 'Henry, Miss. Delia',
'Reeves, Mr. David', 'Panula, Mr. Ernesti Arvid',
'Persson, Mr. Ernst Ulrik',
'Graham, Mrs. William Thompson (Edith Junkins)',
'Bissette, Miss. Amelia', 'Cairns, Mr. Alexander',
'Tornquist, Mr. William Henry',
'Mellinger, Mrs. (Elizabeth Anne Maidment)',
'Natsch, Mr. Charles H', 'Healy, Miss. Hanora "Nora"',
'Andrews, Miss. Kornelia Theodosia',
'Lindblom, Miss. Augusta Charlotta', 'Parkes, Mr. Francis "Frank"',
'Rice, Master. Eric', 'Abbott, Mrs. Stanton (Rosa Hunt)',
'Duane, Mr. Frank', 'Olsson, Mr. Nils Johan Goransson',
'de Pelsmaeker, Mr. Alfons', 'Dorking, Mr. Edward Arthur',
'Smith, Mr. Richard William', 'Stankovic, Mr. Ivan',
'de Mulder, Mr. Theodore', 'Naidenoff, Mr. Penko',

'Hosono, Mr. Masabumi', 'Connolly, Miss. Kate',
'Barber, Miss. Ellen "Nellie"',
'Bishop, Mrs. Dickinson H (Helen Walton)',
'Levy, Mr. Rene Jacques', 'Haas, Miss. Aloisia',
'Mineff, Mr. Ivan', 'Lewy, Mr. Ervin G', 'Hanna, Mr. Mansour',
'Allison, Miss. Helen Loraine', 'Saalfeld, Mr. Adolphe',
'Baxter, Mrs. James (Helene DeLaudeniere Chaput)',
'Kelly, Miss. Anna Katherine "Annie Kate"', 'McCoy, Mr. Bernard',
'Johnson, Mr. William Cahone Jr', 'Keane, Miss. Nora A',
'Williams, Mr. Howard Hugh "Harry"',
'Allison, Master. Hudson Trevor', 'Fleming, Miss. Margaret',
'Penasco y Castellana, Mrs. Victor de Satode (Maria Josefa Perez de Soto y Vallejo)',
'Abelson, Mr. Samuel', 'Francatelli, Miss. Laura Mabel',
'Hays, Miss. Margaret Bechstein', 'Ryerson, Miss. Emily Borie',
'Lahtinen, Mrs. William (Anna Sylfven)', 'Hendekovic, Mr. Ignjac',
'Hart, Mr. Benjamin', 'Nilsson, Miss. Helmina Josefina',
'Kantor, Mrs. Sinai (Miriam Sternin)', 'Moraweck, Dr. Ernest',
'Wick, Miss. Mary Natalie',
'Spedden, Mrs. Frederic Oakley (Margaretta Corning Stone)',
'Dennis, Mr. Samuel', 'Danoff, Mr. Yoto',
'Slayter, Miss. Hilda Mary',
'Caldwell, Mrs. Albert Francis (Sylvia Mae Harbaugh)',
'Sage, Mr. George John Jr', 'Young, Miss. Marie Grice',
'Nysveen, Mr. Johan Hansen', 'Ball, Mrs. (Ada E Hall)',
'Goldsmith, Mrs. Frank John (Emily Alice Brown)',
'Hippach, Miss. Jean Gertrude', 'McCoy, Miss. Agnes',
'Partner, Mr. Austen', 'Graham, Mr. George Edward',
'Vander Planke, Mr. Leo Edmondus',
'Frauenthal, Mrs. Henry William (Clara Heinsheimer)',
'Denkoff, Mr. Mitto', 'Pears, Mr. Thomas Clinton',
'Burns, Miss. Elizabeth Margaret', 'Dahl, Mr. Karl Edward',
'Blackwell, Mr. Stephen Weart', 'Navratil, Master. Edmond Roger',
'Fortune, Miss. Alice Elizabeth', 'Collander, Mr. Erik Gustaf',
'Sedgwick, Mr. Charles Frederick Waddington',
'Fox, Mr. Stanley Hubert', 'Brown, Miss. Amelia "Mildred"',
'Smith, Miss. Marion Elsie',
'Davison, Mrs. Thomas Henry (Mary E Finck)',
'Coutts, Master. William Loch "William"', 'Dimic, Mr. Jovan',
'Odahl, Mr. Nils Martin', 'Williams-Lambert, Mr. Fletcher Fellows',
'Elias, Mr. Tannous', 'Arnold-Franchi, Mr. Josef',
'Yousif, Mr. Wazli', 'Vanden Steen, Mr. Leo Peter',
'Bowerman, Miss. Elsie Edith', 'Funk, Miss. Annie Clemmer',
'McGovern, Miss. Mary', 'Mockler, Miss. Helen Mary "Ellie"',

'Skoog, Mr. Wilhelm', 'del Carlo, Mr. Sebastiano',
'Barbara, Mrs. (Catherine David)', 'Asim, Mr. Adola',
"O'Brien, Mr. Thomas", 'Adahl, Mr. Mauritz Nils Martin',
'Warren, Mrs. Frank Manley (Anna Sophia Atkinson)',
'Moussa, Mrs. (Mantoura Boulos)', 'Jermyn, Miss. Annie',
'Aubart, Mme. Leontine Pauline', 'Harder, Mr. George Achilles',
'Wiklund, Mr. Jakob Alfred', 'Beavan, Mr. William Thomas',
'Ringhini, Mr. Sante', 'Palsson, Miss. Stina Viola',
'Meyer, Mrs. Edgar Joseph (Leila Saks)',
'Landergren, Miss. Aurora Adelia', 'Widener, Mr. Harry Elkins',
'Betros, Mr. Tannous', 'Gustafsson, Mr. Karl Gideon',
'Bidois, Miss. Rosalie', 'Nakid, Miss. Maria ("Mary")',
'Tikkanen, Mr. Juho',
'Holverson, Mrs. Alexander Oskar (Mary Aline Towner)',
'Plotcharsky, Mr. Vasil', 'Davies, Mr. Charles Henry',
'Goodwin, Master. Sidney Leonard', 'Buss, Miss. Kate',
'Sadlier, Mr. Matthew', 'Lehmann, Miss. Bertha',
'Carter, Mr. William Ernest', 'Jansson, Mr. Carl Olof',
'Gustafsson, Mr. Johan Birger', 'Newell, Miss. Marjorie',
'Sandstrom, Mrs. Hjalmar (Agnes Charlotta Bengtsson)',
'Johansson, Mr. Erik', 'Olsson, Miss. Elina',
'McKane, Mr. Peter David', 'Pain, Dr. Alfred',
'Trout, Mrs. William H (Jessie L)', 'Niskanen, Mr. Juha',
'Adams, Mr. John', 'Jussila, Miss. Mari Aina',
'Hakkarainen, Mr. Pekka Pietari', 'Oreskovic, Miss. Marija',
'Gale, Mr. Shadrach', 'Widegren, Mr. Carl/Charles Peter',
'Richards, Master. William Rowe',
'Birkeland, Mr. Hans Martin Monsen', 'Lefebre, Miss. Ida',
'Sdycoff, Mr. Todor', 'Hart, Mr. Henry', 'Minahan, Miss. Daisy E',
'Cunningham, Mr. Alfred Fleming', 'Sundman, Mr. Johan Julian',
'Meek, Mrs. Thomas (Annie Louise Rowley)',
'Drew, Mrs. James Vivian (Lulu Thorne Christian)',
'Silven, Miss. Lyyli Karoliina', 'Matthews, Mr. William John',
'Van Impe, Miss. Catharina', 'Gheorgheff, Mr. Stanio',
'Charters, Mr. David', 'Zimmerman, Mr. Leo',
'Danbom, Mrs. Ernst Gilbert (Anna Sigrid Maria Brogren)',
'Rosblom, Mr. Viktor Richard', 'Wiseman, Mr. Phillippe',
'Clarke, Mrs. Charles V (Ada Maria Winfield)',
'Phillips, Miss. Kate Florence ("Mrs Kate Louise Phillips Marshall")',
'Flynn, Mr. James', 'Pickard, Mr. Berk (Berk Trembisky)',
'Bjornstrom-Steffansson, Mr. Mauritz Hakan',
'Thorncroft, Mrs. Percival (Florence Kate White)',
'Louch, Mrs. Charles Alexander (Alice Adelaide Slow)',

'Kallio, Mr. Nikolai Erland', 'Silvey, Mr. William Baird',
'Carter, Miss. Lucile Polk',
'Ford, Miss. Doolina Margaret "Daisy"',
'Richards, Mrs. Sidney (Emily Hocking)', 'Fortune, Mr. Mark',
'Kvillner, Mr. Johan Henrik Johannesson',
'Hart, Mrs. Benjamin (Esther Ada Bloomfield)', 'Hampe, Mr. Leon',
'Petterson, Mr. Johan Emil', 'Reynaldo, Ms. Encarnacion',
'Johannesen-Bratthammer, Mr. Bernt', 'Dodge, Master. Washington',
'Mellinger, Miss. Madeleine Violet', 'Seward, Mr. Frederic Kimber',
'Baclini, Miss. Marie Catherine', 'Peuchen, Major. Arthur Godfrey',
'West, Mr. Edwy Arthur', 'Hagland, Mr. Ingvald Olai Olsen',
'Foreman, Mr. Benjamin Laventall', 'Goldenberg, Mr. Samuel L',
'Peduzzi, Mr. Joseph', 'Jalsevac, Mr. Ivan',
'Millet, Mr. Francis Davis', 'Kenyon, Mrs. Frederick R (Marion)',
'Toomey, Miss. Ellen', "O'Connor, Mr. Maurice",
'Anderson, Mr. Harry', 'Morley, Mr. William', 'Gee, Mr. Arthur H',
'Milling, Mr. Jacob Christian', 'Maisner, Mr. Simon',
'Goncalves, Mr. Manuel Estanslas', 'Campbell, Mr. William',
'Smart, Mr. John Montgomery', 'Scanlan, Mr. James',
'Baclini, Miss. Helene Barbara', 'Keefe, Mr. Arthur',
'Cacic, Mr. Luka', 'West, Mrs. Edwy Arthur (Ada Mary Worth)',
'Jerwan, Mrs. Amin S (Marie Marthe Thuillard)',
'Strandberg, Miss. Ida Sofia', 'Clifford, Mr. George Quincy',
'Renouf, Mr. Peter Henry', 'Braund, Mr. Lewis Richard',
'Karlsson, Mr. Nils August', 'Hirvonen, Miss. Hildur E',
'Goodwin, Master. Harold Victor',
'Frost, Mr. Anthony Wood "Archie"', 'Rouse, Mr. Richard Henry',
'Turkula, Mrs. (Hedwig)', 'Bishop, Mr. Dickinson H',
'Lefebre, Miss. Jeannie',
'Hoyt, Mrs. Frederick Maxfield (Jane Anne Forby)',
'Kent, Mr. Edward Austin', 'Somerton, Mr. Francis William',
'Coutts, Master. Eden Leslie "Neville"',
'Hagland, Mr. Konrad Mathias Reiersen', 'Windelov, Mr. Einar',
'Molson, Mr. Harry Markland', 'Artagaveytia, Mr. Ramon',
'Stanley, Mr. Edward Roland', 'Yousseff, Mr. Gerious',
'Eustis, Miss. Elizabeth Mussey',
'Shellard, Mr. Frederick William',
'Allison, Mrs. Hudson J C (Bessie Waldo Daniels)',
'Svensson, Mr. Olof', 'Calic, Mr. Petar', 'Canavan, Miss. Mary',
"O'Sullivan, Miss. Bridget Mary", 'Laitinen, Miss. Kristina Sofia',
'Maioni, Miss. Roberta',
'Penasco y Castellana, Mr. Victor de Satode',
'Quick, Mrs. Frederick Charles (Jane Richards)',

'Bradley, Mr. George ("George Arthur Brayton")',
'Olsen, Mr. Henry Margido', 'Lang, Mr. Fang',
'Daly, Mr. Eugene Patrick', 'Webber, Mr. James',
'McGough, Mr. James Robert',
'Rothschild, Mrs. Martin (Elizabeth L. Barrett)',
'Coleff, Mr. Satio', 'Walker, Mr. William Anderson',
'Lemore, Mrs. (Amelia Milley)', 'Ryan, Mr. Patrick',
'Angle, Mrs. William A (Florence "Mary" Agnes Hughes)',
'Pavlovic, Mr. Stefo', 'Perreault, Miss. Anne', 'Vovk, Mr. Janko',
'Lahoud, Mr. Sarkis',
'Hippach, Mrs. Louis Albert (Ida Sophia Fischer)',
'Kassem, Mr. Fared', 'Farrell, Mr. James', 'Ridsdale, Miss. Lucy',
'Farthing, Mr. John', 'Salonen, Mr. Johan Werner',
'Hocking, Mr. Richard George', 'Quick, Miss. Phyllis May',
'Toufik, Mr. Nakli', 'Elias, Mr. Joseph Jr',
'Peter, Mrs. Catherine (Catherine Rizk)', 'Cacic, Miss. Marija',
'Hart, Miss. Eva Miriam', 'Butt, Major. Archibald Willingham',
'LeRoy, Miss. Bertha', 'Risien, Mr. Samuel Beard',
'Frolicher, Miss. Hedwig Margaritha', 'Crosby, Miss. Harriet R',
'Andersson, Miss. Ingeborg Constanzia',
'Andersson, Miss. Sigrid Elisabeth', 'Beane, Mr. Edward',
'Douglas, Mr. Walter Donald', 'Nicholson, Mr. Arthur Ernest',
'Beane, Mrs. Edward (Ethel Clarke)', 'Padro y Manent, Mr. Julian',
'Goldsmith, Mr. Frank John', 'Davies, Master. John Morgan Jr',
'Thayer, Mr. John Borland Jr', 'Sharp, Mr. Percival James R',
"O'Brien, Mr. Timothy", 'Leeni, Mr. Fahim ("Philip Zenni")',
'Ohman, Miss. Velin', 'Wright, Mr. George',
'Duff Gordon, Lady. (Lucille Christiana Sutherland) ("Mrs Morgan")',
'Robbins, Mr. Victor', 'Taussig, Mrs. Emil (Tillie Mandelbaum)',
'de Messemaeker, Mrs. Guillaume Joseph (Emma)',
'Morrow, Mr. Thomas Rowan', 'Sivic, Mr. Husein',
'Norman, Mr. Robert Douglas', 'Simmons, Mr. John',
'Meanwell, Miss. (Marion Ogden)', 'Davies, Mr. Alfred J',
'Stoytcheff, Mr. Ilia',
'Palsson, Mrs. Nils (Alma Cornelia Berglund)',
'Doharr, Mr. Tannous', 'Jonsson, Mr. Carl', 'Harris, Mr. George',
'Appleton, Mrs. Edward Dale (Charlotte Lamson)',
'Flynn, Mr. John Irwin ("Irving")', 'Kelly, Miss. Mary',
'Rush, Mr. Alfred George John', 'Patchett, Mr. George',
'Garside, Miss. Ethel',
'Silvey, Mrs. William Baird (Alice Munger)',
'Caram, Mrs. Joseph (Maria Elias)', 'Jussila, Mr. Eiriik',
'Christy, Miss. Julie Rachel',

'Thayer, Mrs. John Borland (Marian Longstreth Morris)',
'Downton, Mr. William James', 'Ross, Mr. John Hugo',
'Paulner, Mr. Uscher', 'Taussig, Miss. Ruth',
'Jarvis, Mr. John Denzil', 'Frolicher-Stehli, Mr. Maxmillian',
'Gilinski, Mr. Eliezer', 'Murdlin, Mr. Joseph',
'Rintamaki, Mr. Matti',
'Stephenson, Mrs. Walter Bertram (Martha Eustis)',
'Elsbury, Mr. William James', 'Bourke, Miss. Mary',
'Chapman, Mr. John Henry', 'Van Impe, Mr. Jean Baptiste',
'Leitch, Miss. Jessie Wills', 'Johnson, Mr. Alfred',
'Boulos, Mr. Hanna',
'Duff Gordon, Sir. Cosmo Edmund ("Mr Morgan")',
'Jacobsohn, Mrs. Sidney Samuel (Amy Frances Christy)',
'Slabenoff, Mr. Petco', 'Harrington, Mr. Charles H',
'Torber, Mr. Ernst William', 'Homer, Mr. Harry ("Mr E Haven")',
'Lindell, Mr. Edvard Bengtsson', 'Karaic, Mr. Milan',
'Daniel, Mr. Robert Williams',
'Laroche, Mrs. Joseph (Juliette Marie Louise Lafargue)',
'Shutes, Miss. Elizabeth W',
'Andersson, Mrs. Anders Johan (Alfrida Konstantia Brogren)',
'Jardin, Mr. Jose Neto', 'Murphy, Miss. Margaret Jane',
'Horgan, Mr. John', 'Brocklebank, Mr. William Alfred',
'Herman, Miss. Alice', 'Danbom, Mr. Ernst Gilbert',
'Lobb, Mrs. William Arthur (Cordelia K Stanlick)',
'Becker, Miss. Marion Louise', 'Gavey, Mr. Lawrence',
'Yasbeck, Mr. Antoni', 'Kimball, Mr. Edwin Nelson Jr',
'Nakid, Mr. Sahid', 'Hansen, Mr. Henry Damsgaard',
'Bowen, Mr. David John "Dai"', 'Sutton, Mr. Frederick',
'Kirkland, Rev. Charles Leonard', 'Longley, Miss. Gretchen Fiske',
'Bostandyeff, Mr. Guentcho', "O'Connell, Mr. Patrick D",
'Barkworth, Mr. Algernon Henry Wilson',
'Lundahl, Mr. Johan Svensson', 'Stahelin-Maeglin, Dr. Max',
'Parr, Mr. William Henry Marsh', 'Skoog, Miss. Mabel',
'Davis, Miss. Mary', 'Leinonen, Mr. Antti Gustaf',
'Collyer, Mr. Harvey', 'Panula, Mrs. Juha (Maria Emilia Ojala)',
'Thorneycroft, Mr. Percival', 'Jensen, Mr. Hans Peder',
'Sagesser, Mlle. Emma', 'Skoog, Miss. Margit Elizabeth',
'Foo, Mr. Choong', 'Baolini, Miss. Eugenie',
'Harper, Mr. Henry Sleeper', 'Cor, Mr. Liudevit',
'Simonius-Blumer, Col. Oberst Alfons', 'Willey, Mr. Edward',
'Stanley, Miss. Amy Zillah Elsie', 'Mitkoff, Mr. Mito',
'Doling, Miss. Elsie', 'Kalvik, Mr. Johannes Halvorsen',
'O'Leary, Miss. Hanora "Norah"', 'Hegarty, Miss. Hanora "Nora"',

'Hickman, Mr. Leonard Mark', 'Radeff, Mr. Alexander',
'Bourke, Mrs. John (Catherine)', 'Eitemiller, Mr. George Floyd',
'Newell, Mr. Arthur Webster', 'Frauenthal, Dr. Henry William',
'Badt, Mr. Mohamed', 'Colley, Mr. Edward Pomeroy',
'Coleff, Mr. Peju', 'Lindqvist, Mr. Eino William',
'Hickman, Mr. Lewis', 'Butler, Mr. Reginald Fenton',
'Rommelvedt, Mr. Knud Paust', 'Cook, Mr. Jacob',
'Taylor, Mrs. Elmer Zebley (Juliet Cummins Wright)',
'Brown, Mrs. Thomas William Solomon (Elizabeth Catherine Ford)',
'Davidson, Mr. Thornton', 'Mitchell, Mr. Henry Michael',
'Wilhelms, Mr. Charles', 'Watson, Mr. Ennis Hastings',
'Edvardsson, Mr. Gustaf Hjalmar', 'Sawyer, Mr. Frederick Charles',
'Turja, Miss. Anna Sofia',
'Goodwin, Mrs. Frederick (Augusta Tyler)',
'Cardeza, Mr. Thomas Drake Martinez', 'Peters, Miss. Katie',
'Hassab, Mr. Hammad', 'Olsvigen, Mr. Thor Anderson',
'Goodwin, Mr. Charles Edward', 'Brown, Mr. Thomas William Solomon',
'Laroche, Mr. Joseph Philippe Lemercier',
'Panula, Mr. Jaako Arnold', 'Dakic, Mr. Branko',
'Fischer, Mr. Eberhard Thelander',
'Madill, Miss. Georgette Alexandra', 'Dick, Mr. Albert Adrian',
'Karun, Miss. Manca', 'Lam, Mr. Ali', 'Saad, Mr. Khalil',
'Weir, Col. John', 'Chapman, Mr. Charles Henry',
'Kelly, Mr. James', 'Mullens, Miss. Katherine "Katie"',
'Thayer, Mr. John Borland',
'Humblen, Mr. Adolf Mathias Nicolai Olsen',
'Astor, Mrs. John Jacob (Madeleine Talmadge Force)',
'Silverthorne, Mr. Spencer Victor', 'Barbara, Miss. Saiide',
'Gallagher, Mr. Martin', 'Hansen, Mr. Henrik Juul',
'Morley, Mr. Henry Samuel ("Mr Henry Marshall")',
'Kelly, Mrs. Florence "Fannie"',
'Calderhead, Mr. Edward Pennington', 'Cleaver, Miss. Alice',
'Moubarek, Master. Halim Gonios ("William George")',
'Mayne, Mlle. Berthe Antonine ("Mrs de Villiers")',
'Klaber, Mr. Herman', 'Taylor, Mr. Elmer Zebley',
'Larsson, Mr. August Viktor', 'Greenberg, Mr. Samuel',
'Soholt, Mr. Peter Andreas Lauritz Andersen',
'Endres, Miss. Caroline Louise',
'Troutt, Miss. Edwina Celia "Winnie"', 'McEvoy, Mr. Michael',
'Johnson, Mr. Malkolm Joackim',
'Harper, Miss. Annie Jessie "Nina"', 'Jensen, Mr. Svend Lauritz',
'Gillespie, Mr. William Henry', 'Hodges, Mr. Henry Price',
'Chambers, Mr. Norman Campbell', 'Oreskovic, Mr. Luka',

'Renouf, Mrs. Peter Henry (Lillian Jefferys)',
'Mannion, Miss. Margareth', 'Bryhl, Mr. Kurt Arnold Gottfrid',
'Ilmakangas, Miss. Pieta Sofia', 'Allen, Miss. Elisabeth Walton',
'Hassan, Mr. Houssein G N', 'Knight, Mr. Robert J',
'Berriman, Mr. William John', 'Troupiansky, Mr. Moses Aaron',
'Williams, Mr. Leslie', 'Ford, Mrs. Edward (Margaret Ann Watson)',
'Lesurer, Mr. Gustave J', 'Ivanoff, Mr. Kanio',
'Nankoff, Mr. Minko', 'Hawksford, Mr. Walter James',
'Cavendish, Mr. Tyrell William',
'Ryerson, Miss. Susan Parker "Suzette"', 'McNamee, Mr. Neal',
'Stranden, Mr. Juho', 'Crosby, Capt. Edward Gifford',
'Abbott, Mr. Rossmore Edward', 'Sinkkonen, Miss. Anna',
'Marvin, Mr. Daniel Warner', 'Connaghton, Mr. Michael',
'Wells, Miss. Joan', 'Moor, Master. Meier',
'Vande Velde, Mr. Johannes Joseph', 'Jonkoff, Mr. Lilio',
'Herman, Mrs. Samuel (Jane Laver)', 'Hamalainen, Master. Viljo',
'Carlsson, Mr. August Sigfrid', 'Bailey, Mr. Percy Andrew',
'Theobald, Mr. Thomas Leonard',
'Rothes, the Countess. of (Lucy Noel Martha Dyer-Edwards)',
'Garfirth, Mr. John', 'Nirva, Mr. Iisakki Antino Aijo',
'Barah, Mr. Hanna Assi',
'Carter, Mrs. William Ernest (Lucile Polk)',
'Eklund, Mr. Hans Linus', 'Hogeboom, Mrs. John C (Anna Andrews)',
'Brewe, Dr. Arthur Jackson', 'Mangan, Miss. Mary',
'Moran, Mr. Daniel J', 'Gronnestad, Mr. Daniel Danielsen',
'Lievens, Mr. Rene Aime', 'Jensen, Mr. Niels Peder',
'Mack, Mrs. (Mary)', 'Elias, Mr. Dibo',
'Hocking, Mrs. Elizabeth (Eliza Needs)',
'Myhrman, Mr. Pehr Fabian Oliver Malkom', 'Tobin, Mr. Roger',
'Emanuel, Miss. Virginia Ethel', 'Kilgannon, Mr. Thomas J',
'Robert, Mrs. Edward Scott (Elisabeth Walton McMillan)',
'Ayoub, Miss. Banoura',
'Dick, Mrs. Albert Adrian (Vera Gillespie)',
'Long, Mr. Milton Clyde', 'Johnston, Mr. Andrew G',
'Ali, Mr. William', 'Harmer, Mr. Abraham (David Lishin)',
'Sjoblom, Miss. Anna Sofia', 'Rice, Master. George Hugh',
'Dean, Master. Bertram Vere', 'Guggenheim, Mr. Benjamin',
'Keane, Mr. Andrew "Andy"', 'Gaskell, Mr. Alfred',
'Sage, Miss. Stella Anna', 'Hoyt, Mr. William Fisher',
'Dantcheff, Mr. Ristiu', 'Otter, Mr. Richard',
'Leader, Dr. Alice (Farnham)', 'Osman, Mrs. Mara',
'Ibrahim Shawah, Mr. Yousseff',
'Van Impe, Mrs. Jean Baptiste (Rosalie Paula Govaert)',

'Ponesell, Mr. Martin',
'Collyer, Mrs. Harvey (Charlotte Annie Tate)',
'Carter, Master. William Thornton II',
'Thomas, Master. Assad Alexander', 'Hedman, Mr. Oskar Arvid',
'Johansson, Mr. Karl Johan', 'Andrews, Mr. Thomas Jr',
'Pettersson, Miss. Ellen Natalia', 'Meyer, Mr. August',
'Chambers, Mrs. Norman Campbell (Bertha Griggs)',
'Alexander, Mr. William', 'Lester, Mr. James',
'Slemen, Mr. Richard James', 'Andersson, Miss. Ebba Iris Alfrida',
'Tomlin, Mr. Ernest Portage', 'Fry, Mr. Richard',
'Heininen, Miss. Wendla Maria', 'Mallet, Mr. Albert',
'Holm, Mr. John Fredrik Alexander', 'Skoog, Master. Karl Thorsten',
'Hays, Mrs. Charles Melville (Clara Jennings Gregg)',
'Lulic, Mr. Nikola', 'Reuchlin, Jonkheer. John George',
'Moor, Mrs. (Beila)', 'Panula, Master. Urho Abraham',
'Flynn, Mr. John', 'Lam, Mr. Len', 'Mallet, Master. Andre',
'McCormack, Mr. Thomas Joseph',
'Stone, Mrs. George Nelson (Martha Evelyn)',
'Yasbeck, Mrs. Antoni (Selini Alexander)',
'Richards, Master. George Sibley', 'Saad, Mr. Amin',
'Augustsson, Mr. Albert', 'Allum, Mr. Owen George',
'Compton, Miss. Sara Rebecca', 'Pasic, Mr. Jakob',
'Sirota, Mr. Maurice', 'Chip, Mr. Chang', 'Marechal, Mr. Pierre',
'Alhomaki, Mr. Ilmari Rudolf', 'Mudd, Mr. Thomas Charles',
'Serepeca, Miss. Augusta', 'Lemberopolous, Mr. Peter L',
'Culumovic, Mr. Jeso', 'Abbing, Mr. Anthony',
'Sage, Mr. Douglas Bullen', 'Markoff, Mr. Marin',
'Harper, Rev. John',
'Goldenberg, Mrs. Samuel L (Edwiga Grabowska)',
'Andersson, Master. Sigvard Harald Elias', 'Svensson, Mr. Johan',
'Boulos, Miss. Nourelain', 'Lines, Miss. Mary Conover',
'Carter, Mrs. Ernest Courtenay (Lilian Hughes)',
'Aks, Mrs. Sam (Leah Rosen)',
'Wick, Mrs. George Dennick (Mary Hitchcock)',
'Daly, Mr. Peter Denis ', 'Baolini, Mrs. Solomon (Latifa Quran)',
'Razi, Mr. Raihed', 'Hansen, Mr. Claus Peter',
'Giles, Mr. Frederick Edward',
'Swift, Mrs. Frederick Joel (Margaret Welles Barron)',
'Sage, Miss. Dorothy Edith "Dolly"', 'Gill, Mr. John William',
'Bystrom, Mrs. (Karolina)', 'Duran y More, Miss. Asuncion',
'Roebling, Mr. Washington Augustus II',
'van Melkebeke, Mr. Philemon', 'Johnson, Master. Harold Theodor',
'Balkic, Mr. Cerin',

```
'Beckwith, Mrs. Richard Leonard (Sallie Monypeny)',  

'Carlsson, Mr. Frans Olof', 'Vander Cruyssen, Mr. Victor',  

'Abelson, Mrs. Samuel (Hannah Wizosky)',  

'Najib, Miss. Adele Kiamie "Jane"',  

'Gustafsson, Mr. Alfred Ossian', 'Petroff, Mr. Nedelio',  

'Laleff, Mr. Kristo',  

'Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)',  

'Shelley, Mrs. William (Imanita Parrish Hall)',  

'Markun, Mr. Johann', 'Dahlberg, Miss. Gerda Ulrika',  

'Banfield, Mr. Frederick James', 'Suttehall, Mr. Henry Jr',  

'Rice, Mrs. William (Margaret Norton)', 'Montvila, Rev. Juozas',  

'Graham, Miss. Margaret Edith',  

'Johnston, Miss. Catherine Helen "Carrie"',  

'Behr, Mr. Karl Howell', 'Dooley, Mr. Patrick'], dtype=object)
```

Observation

To create a new feature we can extract the Titles from the name.

```
In [73]: train_df['Title'] = train_df['Name'].apply(lambda x: x.split(',')[1].split('.')[0].strip())
test_df['Title'] = test_df['Name'].apply(lambda x: x.split(',')[1].split('.')[0].strip())
```

```
In [74]: train_df.head()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Title
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.2500	NaN	S	Mr
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...)	female	38.0	1	0	PC 17599	71.2833	C85	C	Mrs
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282	7.9250	NaN	S	Miss
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1000	C123	S	Mrs
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450	8.0500	NaN	S	Mr

```
In [75]: train_df['Title'].value_counts().to_frame()
```

Out[75]:

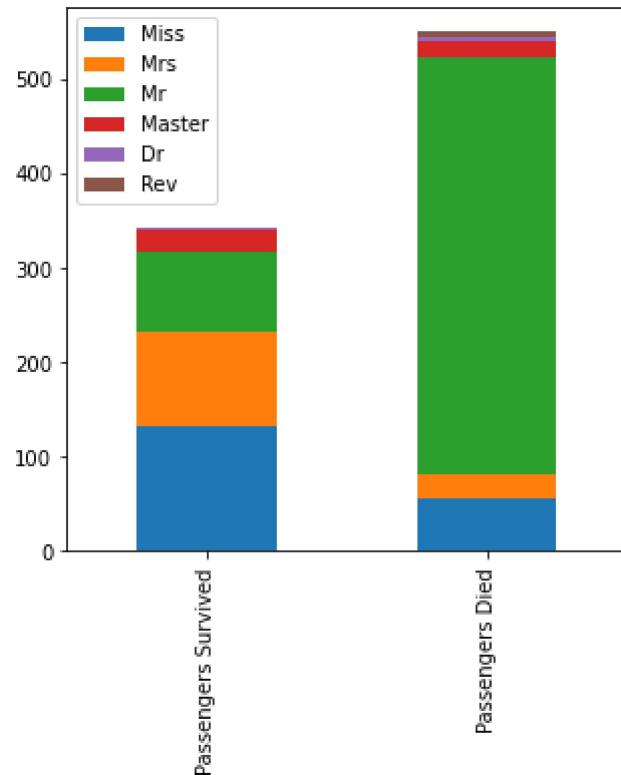
	Title
Mr	517
Miss	182
Mrs	125
Master	40
Dr	7
Rev	6
Mlle	2
Col	2
Major	2
Don	1
Mme	1
Capt	1
Ms	1
the Countess	1
Lady	1
Jonkheer	1
Sir	1

Observation:

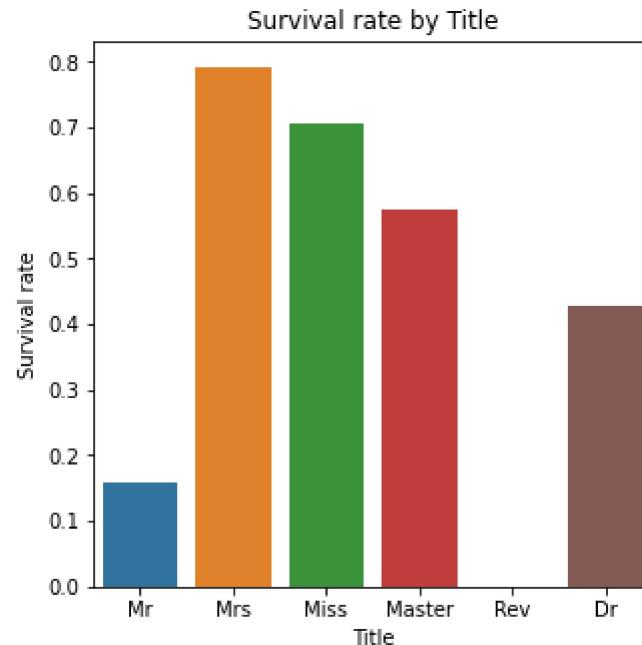
As can see the Titles are severely imbalanced. It makes sense to group less frequent Titles together. I will substitute male titles into Mr and female Titles into Mrs:

```
In [76]: train_df['Title'].replace(['Mme', 'Ms', 'Lady', 'Mlle', 'the Countess', 'Dona'], 'Miss', inplace=True)
test_df['Title'].replace(['Mme', 'Ms', 'Lady', 'Mlle', 'the Countess', 'Dona'], 'Miss', inplace=True)
train_df['Title'].replace(['Major', 'Col', 'Capt', 'Don', 'Sir', 'Jonkheer'], 'Mr', inplace=True)
test_df['Title'].replace(['Major', 'Col', 'Capt', 'Don', 'Sir', 'Jonkheer'], 'Mr', inplace=True)
```

```
In [77]: bar_chart_stacked(train_df, 'Title')
```



```
In [78]: bar_chart_compare(train_df, "Title", title = "Survival rate by Title")
```



Observation:

As expected female Titles result in a higher survival rate. Master and Dr are slightly have a surprisingly high survival rate even though both are male titles. Being "just" a Mr comes with a compromised survival rate of approx 15%. It is interesting that all 6 Reverends died. Maybe they decided to accept their destiny and wanted to die with dignity.

Cabin and Ticket

```
In [79]: train_df[['Cabin', 'Ticket']]
```

```
Out[79]:
```

	Cabin	Ticket
0	NaN	A/5 21171
1	C85	PC 17599
2	NaN	STON/O2. 3101282
3	C123	113803
4	NaN	373450
...
886	NaN	211536
887	B42	112053
888	NaN	W./C. 6607
889	C148	111369
890	NaN	370376

891 rows × 2 columns

Observation:

As we can see both features are not easy to deal with. Cabin contains a lot of Nans and the ticket seems not to provide any useful information.

We can try different ideas:

Extract two leading letters to create a new feature
Extract number of letter in ticket to create a new features
Extract number of cabins used
Extract Cabin letter

```
In [80]: # Extract Leading Letter:  
train_df['Ticket_2letter'] = train_df.Ticket.apply(lambda x: x[:2])  
test_df['Ticket_2letter'] = test_df.Ticket.apply(lambda x: x[:2])
```

```
In [81]: # Extract Ticket Length:  
train_df['Ticket_len'] = train_df.Ticket.apply(lambda x: len(x))  
test_df['Ticket_len'] = test_df.Ticket.apply(lambda x: len(x))
```

```
In [82]: # Extract Number of Cabins:  
train_df['Cabin_num'] = train_df.Ticket.apply(lambda x: len(x.split()))  
test_df['Cabin_num'] = test_df.Ticket.apply(lambda x: len(x.split()))
```

```
In [83]: # Extract Leading Letter:  
train_df['Cabin_1letter'] = train_df.Ticket.apply(lambda x: x[:1])  
test_df['Cabin_1letter'] = test_df.Ticket.apply(lambda x: x[:1])
```

Investigate Results:

Ticket: