PROJECT TITLE : ACCIDENTS

GROUP NO. 10

SECTION : 7CSE05

GROUP MEMBERS-

KAPA SAI TEJA - 20181CSE0312

K.LIKHITHA - 20181CSE0327

K.SAI KRISHNA - 20181CSE0311

K.VENU SUNIL - 20181CSE0321

```
from google.colab import drive
drive.mount('/content/drive')
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.m
```

IMPORTING THE LIBRARIES

```
!pip install pywaffle
```

```
Requirement already satisfied: pywaffle in /usr/local/lib/python3.7/dist-packages (0
Requirement already satisfied: matplotlib in /usr/local/lib/python3.7/dist-packages (
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.7/dist-pac
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in /usr/local
Requirement already satisfied: python-dateutil>=2.1 in /usr/local/lib/python3.7/dist-
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: numpy>=1.11 in /usr/local/lib/python3.7/dist-packages
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.7/dist-packages (fr
```

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

LOADING THE DATA

```
df=pd.read_csv("/content/drive/MyDrive/train (3).csv")
```

TOP 5 ROWS

```
df.head(5)
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171	7.
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th	female	38.0	1	0	PC 17599	71.

SHAPE OF THE DATA

```
df.shape
```

```
(891, 12)
```

DATA INFO

```
df.info
```

```
<bound method DataFrame.info of      PassengerId  Survived  Pclass  ...     Fare Cabi
0              1         0       3  ...   7.2500   NaN         S
1              2         1       1  ...  71.2833   C85         C
2              3         1       3  ...   7.9250   NaN         S
3              4         1       1  ...  53.1000  C123         S
4              5         0       3  ...   8.0500   NaN         S
..           ...       ...     ...  ...      ...   ...       ...
886          887         0       2  ...  13.0000   NaN         S
887          888         1       1  ...  30.0000   B42         S
888          889         0       3  ...  23.4500   NaN         S
889          890         1       1  ...  30.0000  C148         C
890          891         0       3  ...   7.7500   NaN         Q

[891 rows x 12 columns]>
```

DATA DESCRIPTION

```
df.describe()
```

	PassengerId	Survived	Pclass	Age	SibSp	Parch	F
count	891.000000	891.000000	891.000000	714.000000	891.000000	891.000000	891.000
mean	446.000000	0.383838	2.308642	29.699118	0.523008	0.381594	32.204
std	257.353842	0.486592	0.836071	14.526497	1.102743	0.806057	49.693
min	1.000000	0.000000	1.000000	0.420000	0.000000	0.000000	0.000
25%	223.500000	0.000000	2.000000	20.125000	0.000000	0.000000	7.910
50%	446.000000	0.000000	3.000000	28.000000	0.000000	0.000000	14.454
75%	668.500000	1.000000	3.000000	38.000000	1.000000	0.000000	31.000
max	891.000000	1.000000	3.000000	80.000000	8.000000	6.000000	512.329

IF THERE IS ANY NULL VALUE IN THE COLUMN

```
df.isnull()
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
0	False	False	False	False	False	False	False	False	False	False	
1	False	False	False	False	False	False	False	False	False	False	
2	False	False	False	False	False	False	False	False	False	False	
3	False	False	False	False	False	False	False	False	False	False	
4	False	False	False	False	False	False	False	False	False	False	
...	...	...	...	...	...	...	...	...	...	...	
886	False	False	False	False	False	False	False	False	False	False	
887	False	False	False	False	False	False	False	False	False	False	
888	False	False	False	False	False	True	False	False	False	False	
889	False	False	False	False	False	False	False	False	False	False	
890	False	False	False	False	False	False	False	False	False	False	

891 rows × 12 columns

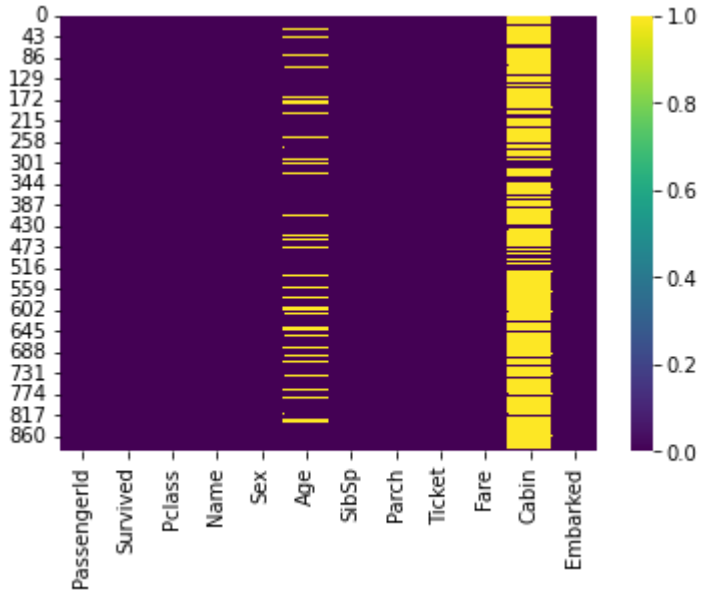
```
df.isnull().sum()
```

```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

We can use seaborn to create a simple heatmap to see where we are missing data!

```
sns.heatmap(df.isnull(),cmap='viridis')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a2a685350>
```



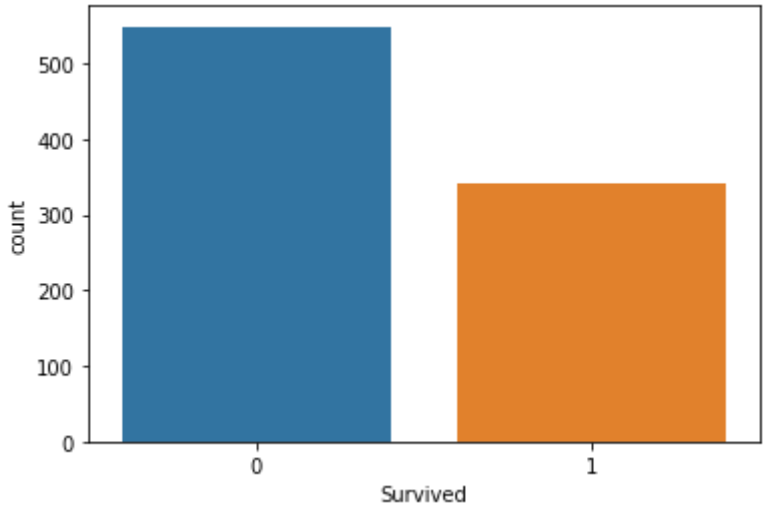

```
df['Survived'].value_counts()
```

```
0    549
1    342
Name: Survived, dtype: int64
```

200 less count in survived compared to not survived

```
sns.countplot(x='Survived',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a2a57fdd0>
```




```
df.Sex.value_counts()
```

```
male      577
female    314
Name: Sex, dtype: int64
```

How many Males and Females were Survived?

```
df.groupby('Sex')['Survived'].value_counts()
```

```
Sex     Survived
female  1           233
        0            81
male    0           468
        1           109
Name: Survived, dtype: int64
```

```
df.groupby('Survived')['Sex'].value_counts()
```

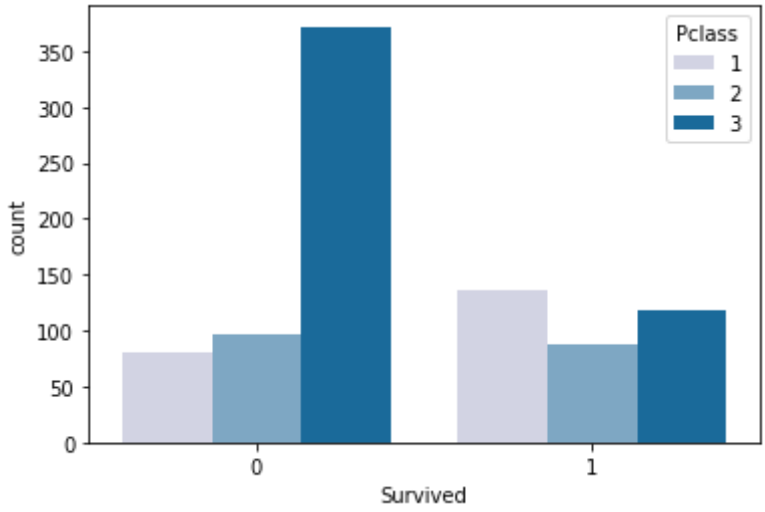
```
Survived  Sex
0         male      468
          female     81
1         female    233
          male      109
Name: Sex, dtype: int64
```

```
sns.countplot(x='Sex',hue='Survived',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a21fd49d0>
```


```
#no of people survived and not survived in various classes
sns.countplot(x='Survived',hue='Pclass',data=df,palette='PuBu')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a215108d0>
```



Pclass vs Fare

```
df.groupby('Pclass')['Fare'].mean()
```

```
Pclass
1    84.154687
2    20.662183
3    13.675550
Name: Fare, dtype: float64
```

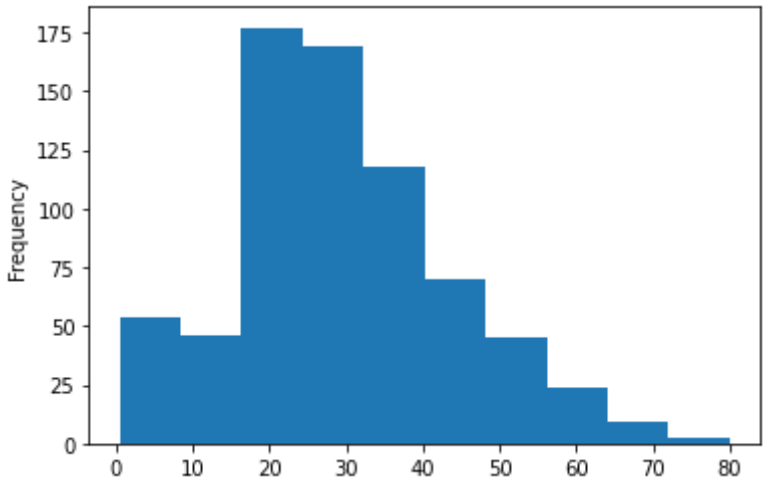
```
df.groupby('Pclass')['Fare'].mean().plot(kind='bar')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a2145f9d0>
```

histogram of age


```
df['Age'].plot.hist()
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a213ca990>
```



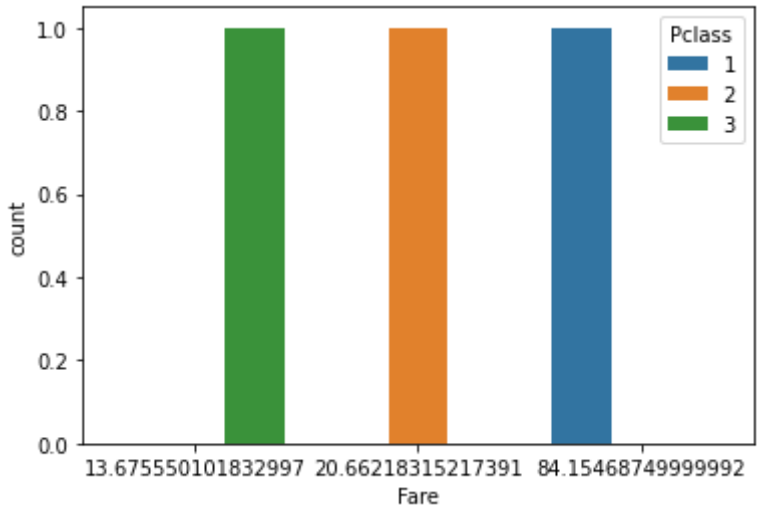

```
df.groupby('Pclass')['Fare'].mean().reset_index()
```

	Pclass	Fare
0	1	84.154687
1	2	20.662183
2	3	13.675550

```
d=df.groupby('Pclass')['Fare'].mean().reset_index()
```

```
sns.countplot(x='Fare',hue='Pclass',data=d)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a212fdb50>
```



How many People Were travelling Alone? And what's thier Survival Rate?

```
df[(df['SibSp']==0) & (df['Parch']==0)].shape[0]
```

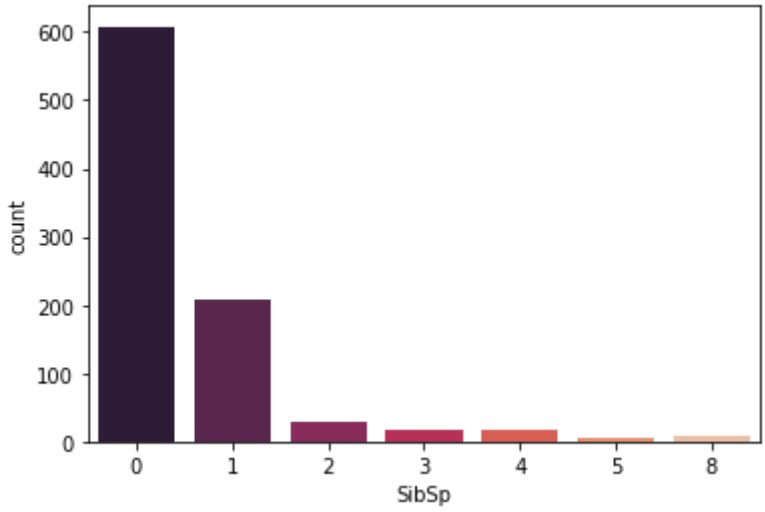
```
537
```

```
df[(df['SibSp']==0) & (df['Parch']==0)]['Survived'].value_counts(normalize=True)
```

```
0    0.696462
1    0.303538
Name: Survived, dtype: float64
```

```
#depicting siblings
sns.countplot(x='SibSp',data=df,palette='rocket')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a2128dc90>
```



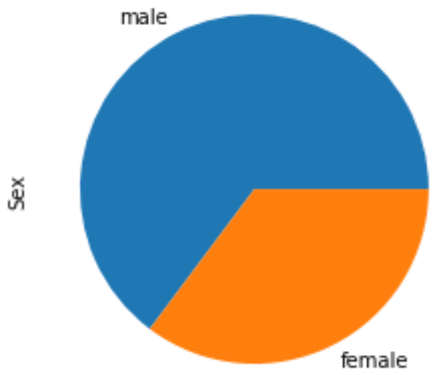

```
#of parents or children
sns.countplot(x='Parch',data=df,palette='summer')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a2120cb90>
```

Sex count representation in pie

```
df.Sex.value_counts().plot(kind='pie')
```

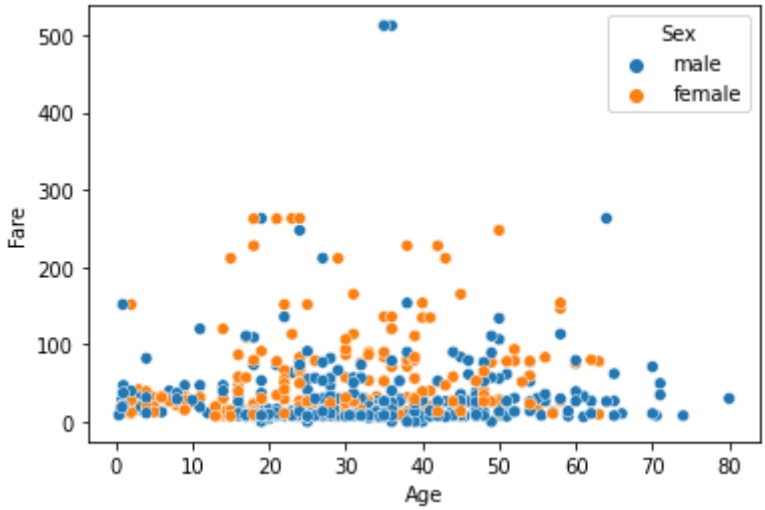
```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a21194890>
```



Scatter plot with the fare payed and the age

```
sns.scatterplot(x='Age',y='Fare',hue='Sex',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a21154490>
```



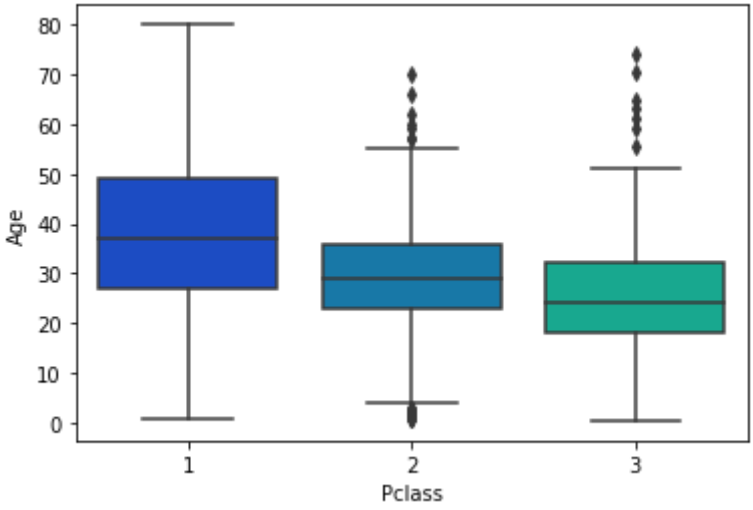

```
sns.distplot(df.Fare)
```

```
/usr/local/lib/python3.7/dist-packages/seaborn/distributions.py:2619: FutureWarning:
  warnings.warn(msg, FutureWarning)
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a210ab590>
```

we observe that older age group are travelling more in class1 and 2 compared to 3

```
sns.boxplot(x='Pclass',y='Age',data=df,palette='winter')
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a20f9b350>
```



fare comparision by passangers

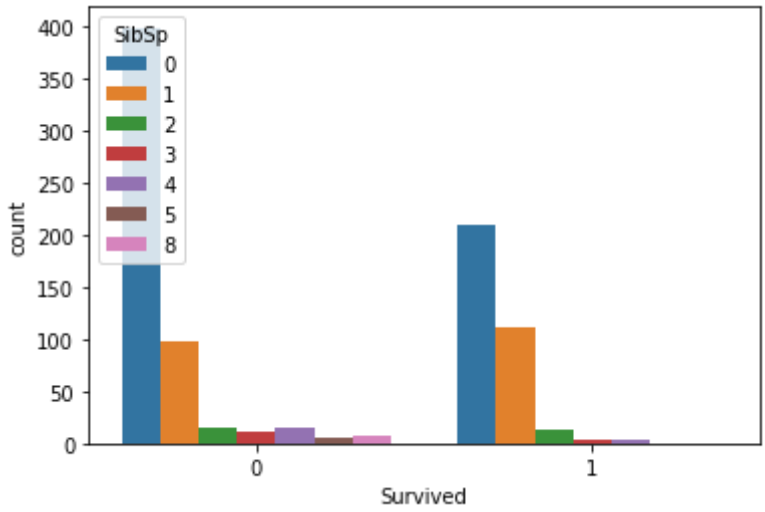
```
df['Fare'].hist(color='green',bins=40)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a21275450>
```

people having siblings r spouse on board


```
sns.countplot(x='Survived',hue='SibSp',data=df)
```

```
<matplotlib.axes._subplots.AxesSubplot at 0x7f6a20fa10d0>
```



✓ 0s completed at 10:42 PM