

MINI PROJECT - 1

```
#MINI PROJECT 1

#NAME - AYUSH KUMAR SHARMA
#YEAR - 2ND DEPT - ECE
#email - sharmaayushKv@gmail.com
#KALYANI GOVERNMENT ENGINEERING COLLEGE
#GITHUB - https://github.com/PHENOL47/TITANIC-EDA

#PERFORMING EDA ON TITANIC DATASET

#PROGRAM PERFORMING EDA ON TITANIC DATASET
import pandas as pd
import seaborn as sns
import numpy as np
import matplotlib.pyplot as plt

#TAKING THE TITANIC DATA
df = pd.read_csv(r"https://raw.githubusercontent.com/ameenmanna8824/DATASETS/main/train.csv")

#PRINTING THE DATASET

print('TITANIC DATASET'.center(120, '-'))
print(df)
print('.'.center(120, '-'))

#PERFORMING EDA

#EDA 1 :-
#GRAPH SHOWING THE SURVIVOR AND NON SURVIVOR UNDER AGE 20

new = df.where(df['Age']<20) #groupng data under the age of 20

sns.countplot(x='Survived',data=new)
plt.title('BAR GRAPH FOR SURVIVOR UNDER AGE OF 20')

plt.show()
```

```
#EDA 2 :-
```

```
print('DATA FOR SURVIVOR FOR PEOPLE UNDER AGE 20:-\n',new.groupby('Survived').size())  
#hence , we conclude that 85 young kids died and 79 young kids survived  
print('-'.center(120,'-'))
```

```
#EDA 3 :-
```

```
print('DATA FOR SURVIVOR FOR MALE AND FEMALE UNDER AGE 20:-  
\n',new.groupby(['Survived','Sex']).size())  
#hence , we conclude 22 young female died and 63 young male died  
#whereas 53 young female survived and 26 young male survived  
print('-'.center(120,'-'))
```

```
#EDA 4 :-
```

```
#minimum price of fare
```

```
print('THE MINIMUM PRICE OF FARE :-',np.min(df['Fare']))  
#so, according to our dataset some people payed no price for fare  
print('-'.center(120,'-'))
```

```
#EDA 5 :-
```

```
#maximum price of fare
```

```
print('THE MAXIMUM PRICE OF FARE :-',np.max(df['Fare']))  
#so, according to the dataset maximum price of 512.3292 was paid for fare  
print('-'.center(120,'-'))
```

```
#EDA 6 :-
```

```
#BARGRAPH FOR SURVIVOR FROM DIFFRENT BOARDING LOCATION
```

```
sns.countplot(data=new,x= 'Embarked', hue= 'Survived')  
plt.title('BAR GRAPH FOR SURVIVOR FROM BOARDING LOCATION')#S-Southampton, C- Cherbourg, Q Queenstown  
plt.show()  
#HENCE , WE CONCLUDE THAT MOST PEOPLE DIED WHO WERE BOARDED FROM SOUTHAMPTON
```

```
#EDA 7 :-
```

```
#BARGRAPH FOR SURVIVOR FROM DIFFERENT CLASS
```

```
sns.countplot(data=new,x= 'Pclass', hue= 'Survived')  
plt.title('BAR GRAPH FOR SURVIVOR FROM DIFFERENT CLASS')  
plt.show()  
#HENCE , FROM THIS GRAPH WE CONCLUDE THAT MOST PEOPLE THAT DIED IN TINATIC WERE FROM 3RD CLASS
```

OUTPUT –

-----TITANIC DATASET-----

PassengerId	Survived	Pclass	Name	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	A/5 21171	7.2500	NaN
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	PC 17599	71.2833	C85
2	3	1	3	Heikkinen, Miss. Laina	STON/O2. 3101282	7.9250	NaN
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	113803	53.1000	C123
4	5	0	3	Allen, Mr. William Henry	373450	8.0500	NaN
..
886	887	0	2	Montvila, Rev. Juozas	211536	13.0000	NaN
887	888	1	1	Graham, Miss. Margaret Edith	112053	30.0000	B42
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	W./C. 6607	23.4500	NaN
889	890	1	1	Behr, Mr. Karl Howell	111369	30.0000	C148
890	891	0	3	Dooley, Mr. Patrick	370376	7.7500	NaN
Q							

[891 rows x 12 columns]

-----DATA FOR SURVIVOR FOR PEOPLE UNDER AGE 20:-----

Survived
0.0 85
1.0 79
dtype: int64

-----DATA FOR SURVIVOR FOR MALE AND FEMALE UNDER AGE 20:-----

Survived Sex
0.0 female 22
male 63
1.0 female 53
male 26
dtype: int64

-----THE MINIMUM PRICE OF FARE :- 0.0-----

-----THE MAXIMUM PRICE OF FARE :- 512.3292-----



