DATA SCIENCE

PROJECT 1

PROJECT TITLE: TITANIC SURVIVAL ANALYSIS

OBJECTIVE: Analyze the Titanic dataset to understand survival rate, compute statistics, and visualize the results using Pandas and Matplotlib.

CODE:

```
#using pandas to read the data:

import pandas as di

import matplotlib.pyplot as mp

print("-----IMPORTING THE DATASET-----")

titanic=di.read_csv("E:\DATA SCIENCE\week1\gender_submission.csv")

print("Titanic data set:",titanic)

print("DATASET IMPORTED SUCCESSFULLY!...")


print("\n-----RENAMING THE COLUMN NAME-----")

titanic.rename(columns={"PassengerId":"Id","Survived":"Status for clarity"},inplace=True),

print("Renamed column name:",titanic)

print("COLUMN NAME RENAMED SUCCESSFULLY!...\n")


print("\n-----EXPLORING THE DATASET-----")

print("The Titanic data from the Top:",titanic.head())

print("The Titanic data from the Bottom:",titanic.tail())

print("Titanic dataset information:",titanic.info())

print("Count:",titanic.shape)

print("DATASET EXPLORED SUCCESSFULLY!...\n")


print("\n-----FILTER DATA-----")
```

```
Non_survived=titanic[titanic["Status for clarity"]==0].head(10)
print("Top 10 Non-Survived:",Non_survived)
survived=titanic[titanic["Status for clarity"]==1].head(10)
print("Top 10 Survived:",survived)
print("DATA FILTERED SUCCESSFULLY!...\n")

print("\n-----COUNT THE TOTAL PASSENGERS-----")
Total_count=titanic["Status for clarity"].value_counts()
survived_count=Total_count[1]
Nonsurvived_count=Total_count[0]
print("Total Survived:",survived_count)
print("Total Not Survived:",Nonsurvived_count)
print("TOTAL NO.OF PASSENGERS CALCULATED SUCCESSFULLY!...\n")

print("\n-----PERCENTAGE OF THE TOTAL PASSENGERS-----")
Total_passengers=len(titanic)
percentage_of_survived_passengers=(survived_count/Total_passengers)*100
percentage_of_Nonsurvived_passengers=(Nonsurvived_count/Total_passengers)*100
print("Percentage Survived:",percentage_of_survived_passengers)
print("Percentage Not Survived:",percentage_of_Nonsurvived_passengers)
print("PERCENTAGE CALCULATED SUCCESSFULLY!...\n")

print("\n-----CREATING BARCHARTS-----")
Total_count.plot(kind='bar',color=["yellow","pink"])
mp.xlabel("Status 0=Not survived,1=survived")
mp.ylabel("No of persons")
mp.title("Survived vs non survived")
mp.xticks(rotation=0)
mp.show()
```

print("BARCHART CREATED SUCCESSFULLY!...\n")

OUTPUT:

```
PS E:\DATA SCIENCE> & E:/Python/python.exe "e:/DATA SCIENCE/titanic1.py"
e:\DATA SCIENCE\titanic1.py:4: SyntaxWarning: invalid escape sequence '\D'
  titanic=di.read_csv("E:\DATA SCIENCE\week1\gender_submission.csv")
-----IMPORTING THE DATASET-----
Titanic data set:      PassengerId  Survived
0           892         0
1           893         1
2           894         0
3           895         0
4           896         1
..          ...       ...
413        1305         0
414        1306         1
415        1307         0
416        1308         0
417        1309         0

[418 rows x 2 columns]
DATASET IMPORTED SUCCESSFULLY!...

-----RENAMING THE COLUMN NAME-----
Renamed column name:        Id  Status for clarity
0     892                  0
1     893                  1
2     894                  0
3     895                  0
4     896                  1
..    ...                ...
413  1305                  0
414  1306                  1
415  1307                  0
416  1308                  0
417  1309                  0

[418 rows x 2 columns]
COLUMN NAME RENAMED SUCCESSFULLY!...
```

```
-----EXPLORING THE DATASET-----
The Titanic data from the Top:     Id  Status for clarity
0  892                   0
1  893                   1
2  894                   0
3  895                   0
4  896                   1
The Titanic data from the Bottom:        Id  Status for clarity
413  1305                   0
414  1306                   1
415  1307                   0
416  1308                   0
417  1309                   0
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 418 entries, 0 to 417
Data columns (total 2 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Id                  418 non-null    int64
 1   Status for clarity  418 non-null    int64
dtypes: int64(2)
memory usage: 6.7 KB
Titanic dataset information: None
Count: (418, 2)
DATASET EXPLORED SUCCESSFULLY!...
```

```
-----FILTER DATA-----
Top 10 Non-Survived:      Id  Status for clarity
0   892                   0
2   894                   0
3   895                   0
5   897                   0
7   899                   0
9   901                   0
10  902                   0
11  903                   0
13  905                   0
16  908                   0
Top 10 Survived:      Id  Status for clarity
1   893                   1
4   896                   1
6   898                   1
8   900                   1
12  904                   1
14  906                   1
15  907                   1
18  910                   1
19  911                   1
22  914                   1
DATA FILTERED SUCCESSFULLY!...


-----COUNT THE TOTAL PASSENGERS-----
Total Survived: 152
Total Not Survived: 266
TOTAL NO.OF PASSENGERS CALCULATED SUCCESSFULLY!...


-----PERCENTAGE OF THE TOTAL PASSENGERS-----
Percentage Survived: 36.36363636363637
Percentage Not Survived: 63.63636363636363
PERCENTAGE CALCULATED SUCCESSFULLY!...


-----CREATING BARCHARTS-----
BARCHART CREATED SUCCESSFULLY!...
```