#Name : Nikhil Irkar

#Roll No: 518

#PRN No:202201090076

#*THEORY ASSIGNMENT DATASET4*

'''4) Find the statistical Analysis on Employee Info (Refer Data Set 4)

a. Find the total no of male and female employee

b. Find the total no of single married and divorced employee

c. Find the employee ID who is manager

d. Find the all employee ID who is Supervisor

e. Clean the dataset if record is empty with 0 values or delete incomplete data row

f. Find the name of employee who is working as manager and from Pune

g. Find the employee who's salary is greater than 1,00,000/-'''

```
import pandas as pd import
numpy as np
df=pd.read_csv("dataset4.cs
v")
print(df)
```

```
#Q1 Total no of male & female in dataset
female_count=df[df['Gender']=='female'].value_counts().s
um() print("Total no of female in company
are:",female_count) male_count=10-female_count
print("Total no male in company are: ",male_count)
```

OUTPUT: Total no of female in company are: 3
Total no male in company are: 7

```
#Q2 Total no of single married and divorced employee
single_emp_count=df[df['Status']=='single'].value_counts().sum()
married_emp_count=df[df['Status']=='married'].value_counts().sum()
divorced_emp_count=10-single_emp_count-married_emp_count

print("Total no of single, married and divorced employee are
```

:",single_emp_count,married_emp_count,divorced_emp_count)

OUTPUT: Total no of single, married and divorced employee are : 5 3 2

```
#Q3 employee ID who is manager
emp=df[df['Designation']=='Manager']['Emp_id']
emp_details=df[df['Designation']=='Manager']

print("The employee Ids who is manager are:\n ",emp)
print("\n& their details is:\n",emp_details)
```

OUTPUT:

```
The employee Ids who is manager are:
 0   1
2   3
5   6
7   8
Name: Emp_id, dtype: int64

& their details is:
   Emp_id Employee_name    City Designation  Salary  Gender    Status
0       1        Sanvi    Pune     Manager  100000  female    single
2       3       Jayesh  Nashik     Manager   90500    male    single
5       6       Pranav    Pune     Manager  100000    male  divorced 7       8         Raja  Nashik
Manager   90500    male   married
```

```
#Q4 All employee ID who is Supervisor

#1st way using boolean index
supervisor_ids = df[df['Designation'] == 'Supervisor']['Emp_id'] print("All the employee Ids who is
supervisor are:
\n",supervisor_ids.to_string(index=False))

#2nd way using loc function
supervisor_ids = df.loc[df['Designation'] == 'Supervisor', 'Emp_id'] print("\n All the employee Ids who is
supervisor are:
\n",supervisor_ids.to_string(index=False))

#3rd way using query function
supervisor_ids = df.query("Designation == 'Supervisor'")['Emp_id'] print("\n All the employee Ids who is
supervisor are:
\n",supervisor_ids.to_string(index=False))
```

OUTPUT:

```
All the employee Ids who is supervisor are:
  5
10

 All the employee Ids who is supervisor are:    5
10

 All the employee Ids who is supervisor are:    5
10
```

#Q5 Clean the dataset if record is empty with 0 values or delete incomplete data row

Replace empty records with 0 values df.fillna(0, inplace=True)

Delete incomplete data rows df.dropna(inplace=True)

print("Dataset after cleaning & deleting incompleting data row is:\n ",df)
OUTPUT:
Dataset after cleaning & deleting incompleting data row is:

	Emp_id	Employee_name	City	Designation	Salary	Gender	Status
1	1	Sanvi	Pune	Manager	100000	female	single
2	2	Mrunmayee	Pune	Sr. Manager	150000	male	married
3	3	Jayesh	Nashik	Manager	90500	male	single
4	4	Gouri	Nashik	Sr. Manager	100500	female	married
5	5	Mahesh	Pune	Supervisor	85000	male	single
6	6	Pranav	Pune	Manager	100000	male	divorced
7	7	Saksham	Pune	Sr. Manager	150000	male	single
8	8	Raja	Nashik	Manager	90500	male	married
9	9	Sunil	Nashik	Sr. Manager	100500	male	single 9
10	Radha	Pune	Supervisor	85000	female	divorced	

#Q6 the name of employee who is working as manager and from Pune
emp_name=df[(df['Designation']=='Manager') &
(df['City']=='Pune')]['Employee_name']
print("The empoyee who is workinng as manager & from pune
are:\n",emp_name.to_string(header=False,index=False)) OUTPUT:
the empoyee who is workinng as manager & from pune are:

Sanvi
Pranav

#Q7 the employee who's salary is greater than 1,00,000/-

emp_grt_sal=df[df['Salary']>100000]['Employee_name']

print("The employee who's salary is greater than 1,00,000
are:\n",emp_grt_sal)
OUTPUT:
The employee who's salary is greater than 1,00,000 are:
1 Mrunmayee
3 Gouri
6 Saksham
8 Sunil
Name: Employee_name, dtype: object