# Assignment 4

## Group Members :-

**Vidya bingi (209)**

**Samruddhi bhujbal(208)**

**Vaibhav Dingalwar(216)**

## **Dataset :** Mumbai university result

## Code:

```
import pandas as pd
df=pd.read_csv("/content/ass4_dataset.csv")
#Printing the dataset
print(df)
#sum of sgpi of all student
print(df['sgpi'].sum())
33290.36
#display the count of successful status
b=df.groupby('status').count()
print(b)
```

```
               seat_no   prn  centre  total_gradepoints  sgpi  \
status
ABS                  1     1       1                  1     1
RR                 497   497     476                497   497
Successful        4227  4227    4227               4227  4227
UM                   1     1       1                  1     1
Unsuccessful       485   485     485                485   485

               year_of_admission  clg_id  gender
status
ABS                            1       1       1
RR                           497     497     497
Successful                  4227    4227    4227
UM                             1       1       1
Unsuccessful                 485     485     485
#the dimension of data frame
```

```
#average of sgpi
b=df.shape
print(b)
a=df['sgpi'].sum()
avg=a/5211
print(avg)
```
```
(5211, 9)
```
```
6.388478219151795
```
```
#COUNTING MAX GRADEPOINTS
a=df['total_gradepoints'].max()
print(a)
```
```
220
```
```
#converting the satus to lower case
print(df['status'].str.lower())
```
```
0       successful
1       successful
2       successful
3       successful
4       successful
          ...
5206    successful
5207    successful
5208    successful
5209    successful
5210    successful
```
```
#calculating Avg year of admission
c=df['year_of_admission'].mean()
i=c.astype(int)
print(i)
```
```
2019
```
```
#sorting the student by seat no
sorted_df = df.sort_values(by=["seat_no"], ascending=False)
print(sorted_df)
```
```
1077  5235851  2.019020e+15  Thane                  0  0.00  Unsuccessful
2071  5235842  2.019020e+15    NaN                  0  0.00            RR
2070  5235841  2.018020e+15    NaN                  0  0.00            RR
2069  5235840  2.020020e+15    NaN                  0  0.00            RR
2068  5235839  2.017020e+15    NaN                  0  0.00            RR
...       ...           ...    ...                ...   ...           ...
2400  5015049  2.019020e+15  Vasai                166  7.55    Successful
2399  5015048  2.019020e+15  Vasai                158  7.18    Successful
2398  5015047  2.020020e+15  Vasai                158  7.18    Successful
2397  5015046  2.020020e+15  Vasai                157  7.14    Successful
1769  5015045  2.022020e+15  Thane                  0  0.00  Unsuccessful

      year_of_admission  clg_id gender
1077               2019     237      M
2071               2019     428      M
2070               2018     428      M
2069               2020     428      M
```

```
2068                2017     428      M
...                  ...     ...    ...
2400                2019     466      M
2399                2019     466      M
2398                2020     466      M
2397                2020     466      M
1769                2022     403      M

[5211 rows x 9 columns]
```

```
#display the record of First ten student
print(df.iloc[1:10])
```

```
  seat_no           prn  centre  total_gradepoints  sgpi         status  \
1  5201542  2.020020e+15  Mumbai                178  8.09    Successful
2  5201543  2.018020e+15  Mumbai                175  7.95    Successful
3  5201544  2.019020e+15  Mumbai                166  7.55    Successful
4  5201545  2.019020e+15  Mumbai                177  8.05    Successful
5  5201546  2.019020e+15  Mumbai                161  7.32    Successful
6  5201547  2.019020e+15  Mumbai                156  7.09    Successful
7  5201548  2.020020e+15  Mumbai                186  8.45    Successful
8  5201549  2.020020e+15  Mumbai                165  7.50    Successful
9  5201550  2.019020e+15  Mumbai                  0  0.00  Unsuccessful

   year_of_admission  clg_id gender
1               2020      10      M
2               2018      10      M
3               2019      10      M
4               2019      10      M
5               2019      10      F
6               2019      10      M
7               2020      10      M
8               2020      10      M
9               2019      10      F
```

```
#Display number of fail student
c=df.groupby('total_gradepoints').get_group(0)
print(c)
d=c.count()
print(d)
```

```
      seat_no           prn  centre  total_gradepoints  sgpi
status  \
9     5201550  2.019020e+15  Mumbai                  0   0.0  Unsuccessful
11    5201552  2.019020e+15  Mumbai                  0   0.0  Unsuccessful
12    5201553  2.019020e+15  Mumbai                  0   0.0  Unsuccessful
41    5201582  2.019020e+15  Mumbai                  0   0.0  Unsuccessful
43    5201584  2.019020e+15  Mumbai                  0   0.0  Unsuccessful
...        ...           ...     ...                ...   ...
...
5155  5018183  2.019020e+15   Thane                  0   0.0  Unsuccessful
5162  5018190  2.019020e+15   Thane                  0   0.0  Unsuccessful
5179  5018207  2.019020e+15   Thane                  0   0.0  Unsuccessful
5198  5018226  2.019020e+15   Thane                  0   0.0            RR
5202  5018230  2.019020e+15   Thane                  0   0.0  Unsuccessful
```

```
      year_of_admission  clg_id gender
9                  2019      10      F
11                 2019      10      M
12                 2019      10      M
41                 2019      10      M
43                 2019      10      M
...                 ...     ...    ...
5155               2019     996      F
5162               2019     996      M
5179               2019     996      F
5198               2019     996      M
5202               2019     996      M

[984 rows x 9 columns]
seat_no              984
prn                  984
centre               963
total_gradepoints    984
sgpi                 984
status               984
year_of_admission    984
clg_id               984
gender               984
dtype: int64
#Missing value in sgpi
c=df['sgpi']
print(c.isnull())
1       False
2       False
3       False
4       False
        ...
5206    False
5207    False
5208    False
5209    False
5210    False
Name: sgpi, Length: 5211, dtype: bool
#display the max of college id
a=df['clg_id'].max()
print(a)
1118
#counting number of female student
a=df.groupby('gender').get_group('F') 0       False


print(a)
      seat_no           prn  centre  total_gradepoints  sgpi
status  \
5     5201546  2.019020e+15  Mumbai                161  7.32    Successful
9     5201550  2.019020e+15  Mumbai                  0  0.00  Unsuccessful
15    5201556  2.019020e+15  Mumbai                172  7.82    Successful
```

```
16    5201557  2.019020e+15  Mumbai                162  7.36    Successful
17    5201558  2.019020e+15  Mumbai                199  9.05    Successful
...       ...           ...     ...                 ...   ...
...
5190  5018218  2.019020e+15   Thane                174  7.91    Successful
5194  5018222  2.019020e+15   Thane                202  9.18    Successful
5200  5018228  2.018020e+15   Thane                205  9.32    Successful
5205  5018233  2.019020e+15   Thane                184  8.36    Successful
5208  5018236  2.019020e+15   Thane                170  7.73    Successful

      year_of_admission  clg_id gender
5                  2019      10      F
9                  2019      10      F
15                 2019      10      F
16                 2019      10      F
17                 2019      10      F
...                 ...     ...    ...
5190               2019     996      F
5194               2019     996      F
5200               2018     996      F
5205               2019     996      F
5208               2019     996      F

[1525 rows x 9 columns]
```

```
#counting number of female student
a=df.groupby('gender').get_group('M')
c=a.count()
print(c)
```

```
seat_no              3686
prn                  3686
centre               3668
total_gradepoints    3686
sgpi                 3686
status               3686
year_of_admission    3686
clg_id               3686
gender               3686
dtype: int64
```

```
#display the records of 2020 student
a=df['year_of_admission']
b=df.groupby('year_of_admission').get_group(2020)
print(b)
```

```
    seat_no           prn  centre  total_gradepoints  sgpi       status  \
0     5201541  2.020020e+15  Mumbai                201  9.14  Successful
1     5201542  2.020020e+15  Mumbai                178  8.09  Successful
7     5201548  2.020020e+15  Mumbai                186  8.45  Successful
8     5201549  2.020020e+15  Mumbai                165  7.50  Successful
21    5201562  2.020020e+15  Mumbai                187  8.50  Successful
...       ...           ...     ...                ...   ...         ...
5178  5018206  2.020020e+15   Thane                159  7.23  Successful
5187  5018215  2.020020e+15   Thane                190  8.64  Successful
5189  5018217  2.020020e+15   Thane                151  6.86  Successful
```

```
5192  5018220  2.020020e+15   Thane                161  7.32  Successful
5197  5018225  2.020020e+15   Thane                193  8.77  Successful

      year_of_admission  clg_id gender
0                  2020      10      M
1                  2020      10      M
7                  2020      10      M
8                  2020      10      M
21                 2020      10      F
...                 ...     ...    ...
5178               2020     996      M
5187               2020     996      M
5189               2020     996      M
5192               2020     996      M
5197               2020     996      M
```

```
[1230 rows x 9 columns]
#maximum Sgpi
c=df['sgpi'].max()
print(c)
#minimum sgpi
d=df['sgpi'].min()
print(d)
```

```
10.0
```

```
0.0
#Mean of grade points
e=df['total_gradepoints'].mean()
print(e)
```

```
140.54557666474764
#Median of grade points
f=df['total_gradepoints'].median()
print(f)
```

```
168.0
#varience of year  of admission
g=df['year_of_admission']
h=g.var()
print(h)
```

```
0.310309617445157
#standard daviation of sgpi
i=df['sgpi']
print(i.std())
```

```
3.1634311479904
#drouping the missing df
print(df.dropna())
```

```
     seat_no           prn  centre  total_gradepoints  sgpi       status
\
0     5201541  2.020020e+15  Mumbai                201  9.14  Successful
1     5201542  2.020020e+15  Mumbai                178  8.09  Successful
2     5201543  2.018020e+15  Mumbai                175  7.95  Successful
```

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 3 | 5201544 | 2.019020e+15 | Mumbai | | 166 | 7.55 | Successful |
| 4 | 5201545 | 2.019020e+15 | Mumbai | | 177 | 8.05 | Successful |
| ... | ... | ... | ... | | ... | ... | ... |
| 5206 | 5018234 | 2.019020e+15 | Thane | | 187 | 8.50 | Successful |
| 5207 | 5018235 | 2.019020e+15 | Thane | | 193 | 8.77 | Successful |
| 5208 | 5018236 | 2.019020e+15 | Thane | | 170 | 7.73 | Successful |
| 5209 | 5018237 | 2.019020e+15 | Thane | | 175 | 7.95 | Successful |
| 5210 | 5018238 | 2.018020e+15 | Thane | | 151 | 6.86 | Successful |

| | year_of_admission | clg_id | gender |
|---|---|---|---|
| 0 | 2020 | 10 | M |
| 1 | 2020 | 10 | M |
| 2 | 2018 | 10 | M |
| 3 | 2019 | 10 | M |
| 4 | 2019 | 10 | M |
| ... | ... | ... | ... |
| 5206 | 2019 | 996 | M |
| 5207 | 2019 | 996 | M |
| 5208 | 2019 | 996 | F |
| 5209 | 2019 | 996 | M |
| 5210 | 2018 | 996 | M |

[5190 rows x 9 columns]