Assessment

I am going to provide two .csv files , you are supposed to work on them and have to provide solutions to the following problems

import necessary libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
%matplotlib inline

sns.set_style('whitegrid')
plt.style.use("fivethirtyeight")
```

merge those two csv files (after getting as dataframes, get them as a single dataframe)

+ Code + Text

```
# merging two csv files
df = pd.concat(
    map(pd.read_csv, ['college_1.csv', 'college_2.csv']), ignore_index=True)
print(df)
```

```
                Name  python  mysql  Previous Geekions  CodeKata Score  \
     0       A.Dharani    82.0   20.0              24500           24500
     1     V.JEEVITHA     82.0   20.0              21740           21740
     2     HEMAVATHI.R   100.0  100.0              19680           19680
     3     Mugunthan S   100.0   47.0              10610           10610
     4     Sathammai.S   100.0    8.0               8980            8980
     ..            ...     ...    ...                ...             ...
     114  praveen raj j    24.0    0.0               2380            2380
     115     AMARNATH D    -1.0   12.0               1890            1890
```

```
116           bala    32.0    0.0              1720            1720
117           XY Z    -1.0   -1.0                 0               0
118      Hariharan    -1.0   -1.0                 0               0

                                  Department  Rising  python_en  \
0             Computer Science and Engineering       0        NaN
1             Computer Science and Engineering       0        NaN
2             Computer Science and Engineering       0        NaN
3             Computer Science and Engineering       0        NaN
4             Computer Science and Engineering       0        NaN
..                                         ...     ...        ...
114           Computer Science and Engineering       0       -1.0
115  Electronics and Communication Engineering       0       52.0
116  Electronics and Communication Engineering       0       49.0
117           Computer Science and Engineering       0       20.0
118           Computer Science and Engineering       0       -1.0

     computational_thinking
0                       NaN
1                       NaN
2                       NaN
3                       NaN
4                       NaN
..                      ...
114                     0.0
115                    -1.0
116                    -1.0
117                    -1.0
118                     0.0

[119 rows x 9 columns]
```

```
data = pd.DataFrame(df)
data.head()
```

	Name	python	mysql	Previous Geekions	CodeKata Score	Department	Rising	python_en	computational_thinking
0	A.Dharani	82.0	20.0	24500	24500	Computer Science and Engineering	0	NaN	NaN

```
data.isnull().sum()
```

```
Name                     0
python                   0
mysql                    0
Previous Geekions        0
CodeKata Score           0
Department               0
Rising                   0
python_en               84
computational_thinking  84
dtype: int64
```

```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 119 entries, 0 to 118
Data columns (total 9 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Name                    119 non-null    object
 1   python                  119 non-null    float64
 2   mysql                   119 non-null    float64
 3   Previous Geekions       119 non-null    int64
 4   CodeKata Score          119 non-null    int64
 5   Department              119 non-null    object
 6   Rising                  119 non-null    int64
 7   python_en               35 non-null     float64
 8   computational_thinking  35 non-null     float64
dtypes: float64(4), int64(3), object(2)
memory usage: 8.5+ KB
```

```
cl = ['python_en','computational_thinking']

for col in cl :
    data[col].replace({np.nan : 0} , inplace = True)
```

```
data.head()
```

	Name	python	mysql	Previous Geekions	CodeKata Score	Department	Rising	python_en	com
0	A.Dharani	82.0	20.0	24500	24500	Computer Science and Engineering	0	0.0	
1	V.JEEVITHA	82.0	20.0	21740	21740	Computer Science	0	0.0	

```
data.isnull().sum()
```

```
Name                    0
python                  0
mysql                   0
Previous Geekions       0
CodeKata Score          0
Department              0
Rising                  0
python_en               0
computational_thinking  0
dtype: int64
```

Take each csv file , split that csv file into multiple categories (example csv files are added in the repo)

consider if the codekata score exceeds 15000 points(present week) then make a csv on those observations as Exceeded expectations.csv

if 10000<codekata score<15000 (Reached_expectations.csv)

if 7000<codekata score<10000 (Needs_Improvement.csv)

if codekate score < 7000 (Unsatisfactory.csv)

```
EE = data[(data["CodeKata Score"]>15000)]
print("\nNo. of rows and columns",EE.shape[0],end="\n\n")
#to download the file into a csv format
EE.to_csv('Exceeded_expectations.csv')
print("\nExceeded expectations.csv is created\n")
EE
```

```
No. of rows and columns 4


Exceeded expectations.csv is created
```

	Name	python	mysql	Previous Geekions	CodeKata Score	Department	Rising	python_en	co
0	A.Dharani	82.0	20.0	24500	24500	Computer Science and Engineering	0	0.0	
						Computer			

```
RE = data[(data["CodeKata Score"]<15000) & (data["CodeKata Score"]>10000)]
print("No. of rows and columns",RE.shape[0],end="\n\n")
RE.to_csv('Reached_expectations.csv')
print("\nReached_expectations.csv is created\n")
RE
```

```
No. of rows and columns 4


Reached_expectations.csv is created
```

	Name	python	mysql	Previous Geekions	CodeKata Score	Department	Rising	python_en	co
3	Mugunthan S	100.0	47.0	10610	10610	Computer Science and Engineering	0	0.0	
85	Chandru	0.0	0.0	14150	14490	Computer Science and	340	20.0	

```
NI = data[(data["CodeKata Score"]>7000)&(data["CodeKata Score"]<10000)]
print("No. of rows and columns",NI.shape[0],end="\n\n")
NI.to_csv('Needs_Improvement.csv')
print("\nNeeds_Improvement.csv is created\n")
NI
```

```
No. of rows and columns 18


Needs_Improvement.csv is created
```

	Name	python	mysql	Previous Geekions	CodeKata Score	Department	Rising	python_en	computational_thinking
4	Sathammai.S	100.0	8.0	8980	8980	Computer Science and Engineering	0	0.0	0.0
5	NIVEESHWAR S	100.0	0.0	8200	8200	Computer Science and Engineering	0	0.0	0.0
6	KALAIARASAN K	85.0	47.0	7750	7880	Electronics and Communication Engineering	130	0.0	0.0
7	MOHAMED ZUBAIR AHMED	82.0	4.0	7020	7020	Computer Science and Engineering	0	0.0	0.0
88	VIINU V P	-1.0	24.0	9150	9150	Electronics and Electrical Engineering	0	100.0	-1.0
89	Poojitha Y	0.0	35.0	7210	8970	Computer Science and Engineering	1760	78.0	0.0

```
UF = data[(data["CodeKata Score"]<7000)]
print("No. of rows and columns",UF.shape[0],end="\n\n")
UF.to_csv('Unsatisfactory.csv')
print("\nUnsatisfactory.csv is created\n")
UF
```

```
No. of rows and columns 93


Unsatisfactory.csv is created
```

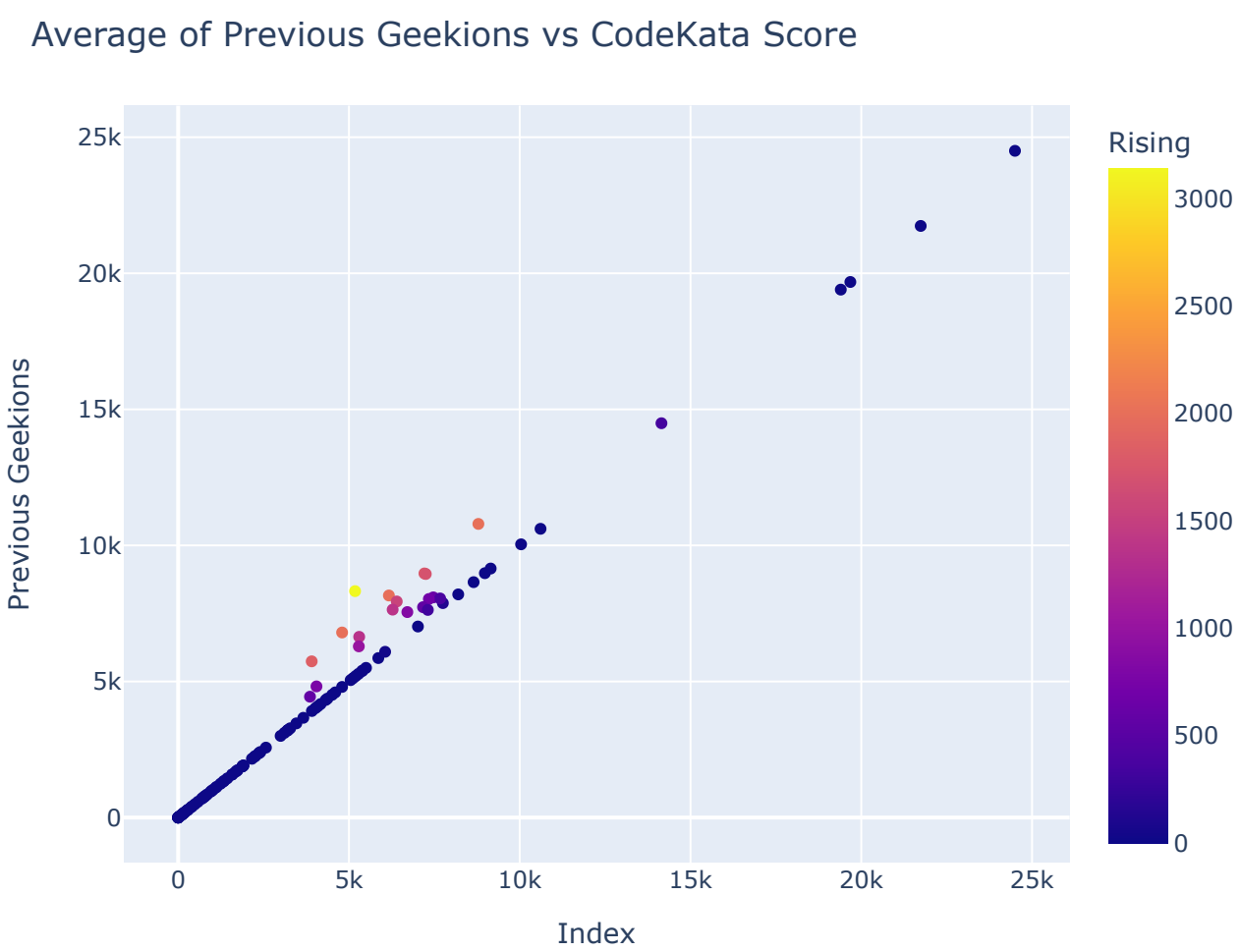
	Name	python	mysql	Previous Geekions	CodeKata Score	Department	Rising	python_en
8	J.SUGANTHI	27.0	50.00	5860	5860	Electronics and Communication Engineering	0	0.0
9	thamizhpaana	29.0	0.00	5500	5500	Computer Science and Engineering	0	0.0
10	Iyappan Samiraj	50.0	24.00	5400	5400	Electronics and Communication Engineering	0	0.0
						Computer		

Average of previous week geekions vs this week geekions (i.e Previous Geekions vs CodeKata Score)

```
print("Average of Previous Geekions is {}".format(data["Previous Geekions"].mean()))
print("Average of Average of CodeKata Score is {}".format(data["CodeKata Score"].mean()))
```

```
Average of Previous Geekions is 3842.3529411764707
Average of Average of CodeKata Score is 4051.764705882353
```

```
m=px.scatter(data,x="Previous Geekions",y='CodeKata Score',color="Rising")
m.update_layout(
    title="Average of Previous Geekions vs CodeKata Score",
    xaxis_title="Index",
    yaxis_title="Previous Geekions",
)
m.show()
```



No of students participated

```
print("\nTotal no. of student participated is {}".format(data["Name"].nunique()))
```

```
Total no. of student participated is 119
```

#Average completion of python course or my_sql or python english or computational thinking

```
x=['python','mysql','python_en','computational_thinking']

for i in x:
  print("\nThe average of the {} is {}".format(i,data[i].mean()))
```

```
The average of the python is 54.35294117647059

The average of the mysql is 23.5

The average of the python_en is 8.722689075630251

The average of the computational_thinking is 0.6470588235294118
```

rising star of the week (top 3 candidate who performed well in that particular week)

```
a=list(data.sort_values(by=["Rising"],ascending=[False])["Name"].head(3))
b=list(data.sort_values(by=["Rising"],ascending=[False])["Rising"].head(3))
print("\nTop 3 rising star of the week \n")
print(*a, sep = "\n")
```

```
Top 3 rising star of the week

shifak N
Narasimhan Y L
Ganesh Ramkumar R
```

Shining stars of the week (top 3 candidates who has highest geekions)

```
c = list(data.sort_values(by=["Previous Geekions"],ascending=[False])["Name"].head(3))
```

```
c = list(data.sort_values(by=["Previous Geekions"],ascending=[False])["Name"].head(3))
print("\nTop 3 rising star of the week \n")
print(*c, sep = "\n")
```

```
Top 3 rising star of the week

A.Dharani
V.JEEVITHA
HEMAVATHI.R
```

Department wise codekata performence (pie chart)

```
d = list(data["Department"].unique())
e = data.groupby("Department")["CodeKata Score"].sum()
e
```

```
Department
Computer Science and Engineering             320025
Electronics and Communication Engineering    108335
Electronics and Electrical Engineering        53800
Name: CodeKata Score, dtype: int64
```

```
plt.pie(e, labels = d, autopct='%1.0f%%', data=True)

plt.show()
```

- Department wise toppers (horizantal bar graph or any visual representations of your choice)

```
t1=px.scatter(data,x="Department",y='CodeKata Score',size='CodeKata Score',color='Name',title="Department wise Topper")
t1.update_layout(
    title="Department wise codekata score and name"
)
t1.show()
```

Department wise codekata score and name

```
toppername = []
```

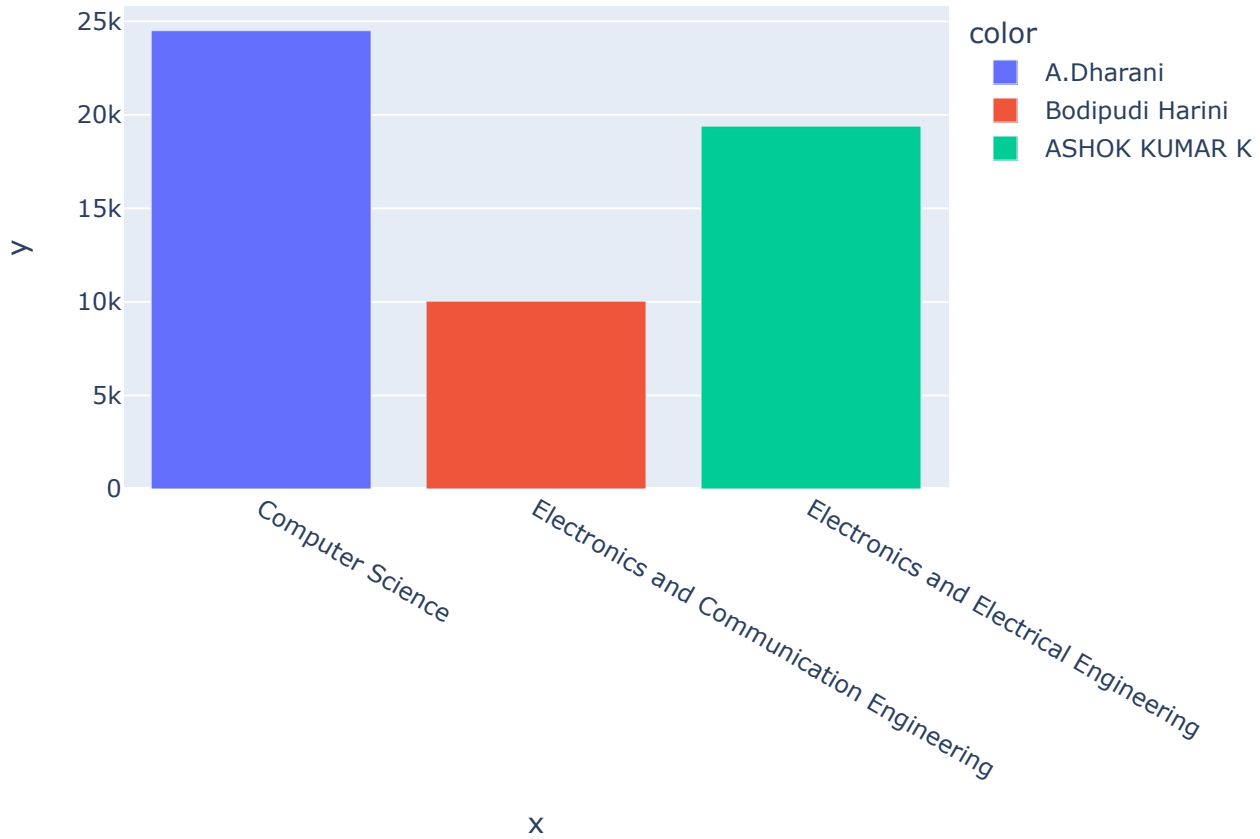
```
x1=data[data["Department"]=="Computer Science and Engineering"]
mcs=x1["CodeKata Score"].max()
toppername.append(x1.Name[x1["CodeKata Score"]==x1["CodeKata Score"].max()].values[0])
x2=data[data["Department"]=="Electronics and Communication Engineering"]
me=x2["CodeKata Score"].max()
toppername.append(x2.Name[x2["CodeKata Score"]==x2["CodeKata Score"].max()].values[0])
x3=data[data["Department"]=="Electronics and Electrical Engineering"]
mee=x3["CodeKata Score"].max()
toppername.append(x3.Name[x3["CodeKata Score"]==x3["CodeKata Score"].max()].values[0])
```

```
print(toppername)
```

```
['A.Dharani', 'Bodipudi Harini', 'ASHOK KUMAR K']
```

```
z1=["Computer Science","Electronics and Communication Engineering","Electronics and Electrical Engineering"]
z2=[mcs,me,mee]

fig=px.bar(x=z1,y=z2, color=toppername,title="Department wise Topper")
fig.update_layout(
    title="Department wise toppers names"
)
fig.show()
```


Department wise toppers names
25k
20k
15k
10k
5k
0
y
x
Computer Science
Electronics and Communication Engineering
Electronics and Electrical Engineering
color
A.Dharani
Bodipudi Harini
ASHOK KUMAR K

Colab paid products - Cancel contracts here

✓ 0s completed at 1:13 PM