```
DESCRIPTION
    A university wants to categorize the marks scored by the students of each batch in
the quizzes conducted.
Datasets:
    studentlist.csv is a csv file (attached) which includes the student details like
enrolment number, admission no. Batch wise list is a folder (attached) which has multiple
csv files. Each file contains the student details belonging to a batch.  For example,
1.csv contains the student details belonging to batch1 and 2.csv contains the student
details belonging to batch2 and so on.
    quiz is a folder(attached) which has multiple csv files.  Each file represents a quiz
conducted by the university and contains all the important details. The column
'Grade/10.00' contains the marks obtained in that quiz out of maximum marks of 10.
    For example quiz1.csv contains the student details and the marks obtained by them in
the first quiz.
```

In [1]:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import glob
import os
```

In [2]:

```
x = os.listdir()
x
```

Out[2]:

```
['.ipynb_checkpoints',
 'batchwiselist',
 'Output0.txt',
 'Output8.txt',
 'Problem Statement.docx',
 'Py  Ass-3.ipynb',
 'quiz',
 'studentlist.csv',
 'testcaseStudent.txt',
 '~$oblem Statement.docx']
```

In [3]:

```
stu_det = pd.read_csv('studentlist.csv')
print(stu_det.shape)
stu_det.head(3)
```

```
(1017, 4)
```

Out[3]:

	enrollmentnumber	admn	studentname	studentEmail
0	1613101016	16SCSE101757	AAYUSH TYAGI	NaN
1	1613101026	16SCSE101415	ABHISHEK	NaN
2	1613101051	16SCSE101222	ADITI SINGH	NaN

In [4]:

```
# Reading batch wise list
path = r'C:\Users\Anand\Desktop\Data Science - Materials\Home  Works\Assignment\Py Ass-3\ba
all_files = glob.glob(path +"/*.csv")

lst = []

for i in all_files:
    df = pd.read_csv(i)
    lst.append(df)

tot_lst = pd.concat(lst)
print(tot_lst.shape)
tot_lst.head(3)
```

```
(1017, 5)
```

Out[4]:

	Unnamed: 0	srNo	enrollmentNumber	admissionNumber	studentName
0	0	1	1613101016	16SCSE101757	AAYUSH TYAGI
1	1	2	1613101026	16SCSE101415	ABHISHEK
2	2	3	1613101051	16SCSE101222	ADITI SINGH

```
Now you need to perform the following steps on this data:
1.  Read the file studentlist.csv
2.  Read the files present in the folders batch wise  list,  quiz
3.  Manipulate the data and create a Statistics dataframe in such a way that it allocates
marks ` for the quizzes. Statistics dataframe for each batch should include the following
column names with the respective statistics value. The column names infer the meaning of
each column.
["no  of  present", "lessthan50", "between50and60", "between60and70", "between70and80",
"greaterthan80"]
        i)  Rows of Statistics dataframe represents the quiz. Such as, the first row
under the column “between50and60” represents the number of students whose percentage is
in between 50 and 60 in the first quiz.
        ii)  The second row under the column “between50and60” represents the number of
students whose percentage is in between 50 and 60 in the second quiz.
```

```
        iii) The first row under the column "no of present" represents the number of
students who attempted the first quiz from that batch.
        iv) The second row under the column "no of present" represents the number of
students who attempted the second quiz from that batch.
```

In [5]:

```
# Reading batch wise list
path = r'C:\Users\Anand\Desktop\Data Science - Materials\Home  Works\Assignment\Py Ass-3\qu
all_files = glob.glob(path +"/*.csv")

quiz = []

for i in all_files:
    df = pd.read_csv(i)
    quiz.append(df)
#define each csv
quiz1 = quiz[0]
quiz2 = quiz[1]
#2 file merged
bth_quiz = pd.concat(quiz)
bth_quiz.head(3)
quiz1 = quiz1.drop(columns=['ID number','Institution',"Department",'Email address'])
quiz2 = quiz2.drop(columns=['ID number','Institution',"Department",'Email address'])
```

In [6]:

```
print(quiz1.shape)
quiz1.head(2)
```

```
(660, 7)
```

Out[6]:

	Surname	Firstname	State	Started on	Completed	Time taken	Grade/10.00
0	16BTCSES11	SIDDHANT GUPTA	Finished	08-02-2019 09:57	08-02-2019 10:11	14 mins 1 sec	4
1	16BTCSES11	ARPIT SHARMA	Finished	08-02-2019 09:57	08-02-2019 10:11	14 mins 5 secs	5

In [7]:

```
stat_df= pd.DataFrame (columns = ["no of present", "lessthan50", "between50and60",
                                  "between60and70", "between70and80", "greaterthan80"],inde

stat_df
```

Out[7]:

	no of present	lessthan50	between50and60	between60and70	between70and80	greaterthan80
Quiz_1	NaN	NaN	NaN	NaN	NaN	NaN
Quiz_2	NaN	NaN	NaN	NaN	NaN	NaN

In [8]:

```
quiz1["Firstname"] = quiz1["Firstname"].dropna()
quiz2["Firstname"] = quiz2["Firstname"].dropna()
```

In [9]:

```
#quiz1["Grade/10.00"].unique()
#quiz2["Grade/10.00"].unique()
quiz1["Grade/10.00"] = quiz1["Grade/10.00"].str.replace("-","0")
quiz2["Grade/10.00"] = quiz2["Grade/10.00"].str.replace("-","0")
quiz1["Grade/10.00"] = quiz1["Grade/10.00"].apply(pd.to_numeric)
quiz2["Grade/10.00"] = quiz2["Grade/10.00"].apply(pd.to_numeric)
```

In [10]:

```
#quiz11 = quiz1[quiz1["Grade/10.00"]<5].reset_index()
quiz1[quiz1["Grade/10.00"]<5]["Firstname"].count()
quiz2[quiz2["Grade/10.00"]<5]["Firstname"].count()
```

Out[10]:

```
93
```

In [11]:

```
quiz1.count()
```

Out[11]:

```
Surname        660
Firstname      659
State          659
Started on     659
Completed      659
Time taken     659
Grade/10.00    660
dtype: int64
```

In [12]:

```
stat_df["no of present"]= [quiz1["Firstname"].count(),quiz2["Firstname"].count()]
stat_df["lessthan50"] = [quiz1[quiz1["Grade/10.00"].between(0,4.9,inclusive=True)]["Firstna
stat_df["between50and60"] = [quiz1[quiz1["Grade/10.00"].between(5,5.9,inclusive=True)]["Fir
stat_df["between60and70"] = [quiz1[quiz1["Grade/10.00"].between(6,6.9,inclusive=True)]["Fir
stat_df["between70and80"] = [quiz1[quiz1["Grade/10.00"].between(7,7.9,inclusive=True)]["Fir
stat_df["greaterthan80"] = [quiz1[quiz1["Grade/10.00"].between(8,10,inclusive=True)]["First


stat_df
```

Out[12]:

	no of present	lessthan50	between50and60	between60and70	between70and80	greaterthan80
Quiz_1	659	127	69	57	58	348
Quiz_2	680	93	40	39	41	467

```
Input Format
You must read a file named testcaseStudent.txt (attached) which has the following lines:
The first line contains the number of testcase pairs.                    9
 Followed by a set of Batch file and Percentage range pairs.
Output Format
For each testcase pair n, create an output file outputn.txt  where n represents the test
case pair number.
This file should contain the number of students who got marks in the requested percentage
range from that batch.
outputn.txt should consist of the values in a separate line for each quiz.
```

In [27]:

```
path = os.getcwd()

all_files = [f for f in glob.glob(path +"\\batchwiselist" +"/*.csv")]
all_files[0]
#for i,v in enumerate(all_files):
#   print(i,v)
file = []

for i in all_files:
    df = pd.read_csv(i)
    file.append(df)

#file
```

In [14]:

```
#bt0 = file[0]
```

In [15]:

```
#student
quiz_1 = pd.merge(quiz1,stu_det,left_on='Firstname',right_on='studentname',how='right')
quiz_2 = pd.merge(quiz2,stu_det,left_on='Firstname',right_on='studentname',how='right')
print(quiz_1.shape)
quiz_1["Grade/10.00"] = quiz_1["Grade/10.00"].fillna(0)
quiz_2["Grade/10.00"] = quiz_2["Grade/10.00"].fillna(0)
```

```
(1017, 11)
```

In [16]:

```
#quiz_1
#stu1 = pd.merge(quiz2,stu_det,left_on='Firstname',right_on='studentname',how='left')
#stu1
#x = pd.merge(quiz_1,)

#quiz_1['studentname']==file[0]['studentName']
#quiz_1['studentname'].count()
#file[0]['studentName'].count()
#quiz_1['studentname'].str.contains(file[0]['studentName'])
```

In [17]:

```
#qz = pd.merge(file[0],quiz_1,left_on='studentName',right_on='studentname',how='left')
#qz.shape

#quiz-1
n = 18
b_q1 = []
for i in range(n+1):
    qz = pd.merge(file[i],quiz_1,left_on='studentName',right_on='studentname',how='left')
    b_q1.append(qz)

print(b_q1[0].shape)
b_q1[0].head(3)

#quiz-2
b_q2 = []
for i in range(n+1):
    qz = pd.merge(file[i],quiz_2,left_on='studentName',right_on='studentname',how='left')
    b_q2.append(qz)

print(b_q2[0].shape)
b_q1[0].head(3)
```

```
(63, 16)
(63, 16)
```

Out[17]:

	Unnamed: 0	srNo	enrollmentNumber	admissionNumber	studentName	Surname	Firstnam
0	0	1	1613101016	16SCSE101757	AAYUSH TYAGI	16BTCSES1	AAYUSH TYAG
1	1	2	1613101026	16SCSE101415	ABHISHEK	16BTCSES1	ABHISHEI
2	2	3	1613101051	16SCSE101222	ADITI SINGH	16BTCSES1	ADIT SINGI

In [ ]:

In [18]:

```
result1= pd.DataFrame (columns = ["no of present", "lessthan50", "between50and60",
                                  "between60and70", "between70and80", "greaterthan80"],inde
result1
```

Out[18]:

	no of present	lessthan50	between50and60	between60and70	between70and80	greaterthan80
Quiz_1	NaN	NaN	NaN	NaN	NaN	NaN
Quiz_2	NaN	NaN	NaN	NaN	NaN	NaN

In [19]:

```
#batch[0].info()

result1["no of present"]= [b_q1[0]["studentName"].count(),b_q2[0]["studentName"].count()]
result1["lessthan50"] = [b_q1[0][b_q1[0]["Grade/10.00"].between(0,4.9,inclusive=True)]["stu
result1["between50and60"] = [b_q1[0][b_q1[0]["Grade/10.00"].between(5,5.9,inclusive=True)][
result1["between60and70"] = [b_q1[0][b_q1[0]["Grade/10.00"].between(6,6.9,inclusive=True)][
result1["between70and80"] = [b_q1[0][b_q1[0]["Grade/10.00"].between(7,7.9,inclusive=True)][
result1["greaterthan80"] = [b_q1[0][b_q1[0]["Grade/10.00"].between(8,10,inclusive=True)]["s

result1
#b_q1[0][b_q1[0]["Grade/10.00"].between(0,5,inclusive=True)]["studentname"].count()
#b_q1[0]["Grade/10.00"].unique()
```

Out[19]:

	no of present	lessthan50	between50and60	between60and70	between70and80	greaterthan80
Quiz_1	63	24	10	4	5	20
Quiz_2	63	25	3	1	0	34

In [20]:

```
rst = []

for i in range(n+1):
    result= pd.DataFrame (columns = ["no of present", "lessthan50", "between50and60",
                                     "between60and70", "between70and80", "greaterthan80"],inde
    result["no of present"]= [b_q1[i]["studentname"].count(),b_q2[i]["studentName"].count()
    result["lessthan50"] = [b_q1[i][b_q1[i]["Grade/10.00"].between(0,4.9,inclusive=True)]["
    result["between50and60"] = [b_q1[i][b_q1[i]["Grade/10.00"].between(5,5.9,inclusive=True
    result["between60and70"] = [b_q1[i][b_q1[i]["Grade/10.00"].between(6,6.9,inclusive=True
    result["between70and80"] = [b_q1[i][b_q1[i]["Grade/10.00"].between(7,7.9,inclusive=True
    result["greaterthan80"] = [b_q1[i][b_q1[i]["Grade/10.00"].between(8,10,inclusive=True)]
    rst.append(result)

rst[0]
#b_q1[0]["studentName"].count()
```

Out[20]:

	no of present	lessthan50	between50and60	between60and70	between70and80	greaterthan80
Quiz_1	63	24	10	4	5	20
Quiz_2	63	25	3	1	0	34

In [21]:

```
txt = pd.read_csv("testcaseStudent.txt",header=-1,names=["inputs"])
txt
#use range
#rst[0]["greaterthan80"]
```

Out[21]:

	inputs
0	2
1	1.csv
2	greaterthan80
3	9.csv
4	lessthan50

In [22]:

```
tst_cs = int(txt.iloc[0,0])
ind = []
for i in range(tst_cs*2+1):
    ind.append(i)

ind
```

Out[22]:

```
[0, 1, 2, 3, 4]
```

In [23]:

```
file_name = ind[1::2]
column = ind[2::2]
print(file_name)
print(column)
```

```
[1, 3]
[2, 4]
```

In [24]:

```
v = []
for i in file_name:
    x = txt.iloc[i,0]
    y = int(x[0])
    v.append(y)
#sub -1 to match index
a = [i - 1 for i in v]

c = []
for i in column:
    x = txt.iloc[i,0]
    c.append(x)

print(v)
print(c)
```

```
[1, 9]
['greaterthan80', 'lessthan50']
```

In [25]:

```
for (x,y) in zip(a,c):
    print(rst[x][y])
```

```
Quiz_1    20
Quiz_2    34
Name: greaterthan80, dtype: int64
Quiz_1    33
Quiz_2    16
Name: lessthan50, dtype: int64
```

In [26]:

```
for (x,y) in zip(a,c):
    out = "Output{}.txt".format(x)
    f = open(out, "a")
    print(rst[x][y], file=f)
```