random_data_generation_code

September 13, 2021

[355]:
```
import numpy
import pandas
```

[356]:
```
myData = pandas.DataFrame()

myData
```

[356]:
```
Empty DataFrame
Columns: []
Index: []
```

[357]:
```
# count of the current col
count = 0
```

[358]:
```
# inserting the roll no col
# index to insert , col name , what to insert
myData.insert(count , "rollNo" , list(range(100000 , 100000 + 50000)))

# icreamenting the count so that we will insert next col in next index
count = count + 1

myData
```

[358]:

	rollNo
0	100000
1	100001
2	100002
3	100003
4	100004
…	…
49995	149995
49996	149996
49997	149997
49998	149998
49999	149999

[50000 rows x 1 columns]

[359]:
```
# importing names , external csv from kaggle
namesDf = pandas.read_csv("names.csv")
namesDf
```

[359]:

	last name	first name	gender	race
0	aaron	aaric a	m	black
1	aaron	dominique j	m	black
2	aaron	fredrick r	m	black
3	aaron	jarvis	m	black
4	aaron	lorenzo	m	black
…	…	…	…	…
35076	rogers	freddie l	m	black
35077	rogers	fredrick c	m	black
35078	rogers	gregory	m	black
35079	rogers	irvin c	m	black
35080	rogers	isaiah f	m	b

[35081 rows x 4 columns]

[360]:
```
# inserting the names
myData.insert(count , "first name" , namesDf["first name"])

count = count + 1
```

[361]:
```
myData
```

[361]:

	rollNo	first name
0	100000	aaric a
1	100001	dominique j
2	100002	fredrick r
3	100003	jarvis
4	100004	lorenzo
…	…	…
49995	149995	NaN
49996	149996	NaN
49997	149997	NaN
49998	149998	NaN
49999	149999	NaN

[50000 rows x 2 columns]

[362]:
```
# making a list of first names
import random

firstNameList = list(namesDf["first name"])
```

[363]:
```
# names were only 37000 we needed 50000
# so were are insert random names from he names list above in the cols having␣
 ↪nan value
for i in myData.index:

    # to check if the dataframe col is Nan
    if(pandas.isnull(myData.at[i, "first name"])):
        myData.at[i, "first name"] = random.choice(firstNameList)
```

[364]:
```
myData
```

[364]:

	rollNo	first name
0	100000	aaric a
1	100001	dominique j
2	100002	fredrick r
3	100003	jarvis
4	100004	lorenzo
…	…	…
49995	149995	leonard jr
49996	149996	john
49997	149997	demoy p
49998	149998	shedrick s
49999	149999	dennis l

[50000 rows x 2 columns]

[365]:
```
# inserting the last names
myData.insert(count , "last name" , namesDf["last name"])

count = count + 1
```

[366]:
```
myData
```

[366]:

	rollNo	first name	last name
0	100000	aaric a	aaron
1	100001	dominique j	aaron
2	100002	fredrick r	aaron
3	100003	jarvis	aaron
4	100004	lorenzo	aaron
…	…	…	…
49995	149995	leonard jr	NaN
49996	149996	john	NaN
49997	149997	demoy p	NaN
49998	149998	shedrick s	NaN
49999	149999	dennis l	NaN

[50000 rows x 3 columns]

[367]:
```
# making a list of first names

lastNameList = list(namesDf["last name"])
```

[368]:
```
# names were only 37000 we needed 50000
# so were are insert random names from he names list above in the cols having␣
 ↪nan value
for i in myData.index:

    # to check if the dataframe col is Nan
    if(pandas.isnull(myData.at[i, "last name"])):
        myData.at[i, "last name"] = random.choice(lastNameList)
```

[369]:
```
myData
```

[369]:

	rollNo	first name	last name
0	100000	aaric a	aaron
1	100001	dominique j	aaron
2	100002	fredrick r	aaron
3	100003	jarvis	aaron
4	100004	lorenzo	aaron
…	…	…	…
49995	149995	leonard jr	hanton
49996	149996	john	clark
49997	149997	demoy p	johnson
49998	149998	shedrick s	jean
49999	149999	dennis l	mustafa

[50000 rows x 3 columns]

[370]:
```
# inserting full name col
myData.insert(count , "fullName" , [i + j for i,j in zip(myData["first name"] ,␣
 ↪myData["last name"])])

# icreamenting the count so that we will insert next col in next index
count = count + 1

myData
```

[370]:

	rollNo	first name	last name	fullName
0	100000	aaric a	aaron	aaric a aaron
1	100001	dominique j	aaron	dominique j aaron
2	100002	fredrick r	aaron	fredrick r aaron
3	100003	jarvis	aaron	jarvis aaron
4	100004	lorenzo	aaron	lorenzo aaron
…	…	…	…	…
49995	149995	leonard jr	hanton	leonard jrhanton

```
49996  149996         john       clark         john  clark
49997  149997       demoy p    johnson     demoy p johnson
49998  149998    shedrick s       jean     shedrick s jean
49999  149999      dennis l    mustafa    dennis l mustafa

[50000 rows x 4 columns]
```

[371]:
```
# adding branches col name
branches = ["CSE" , "COE" , "ECE" , "MEC" , "BOT", "ENC" , "CSE-MBA" ,␣
 ↪"ECE-MBA" , "MEC-MBA"]

randomList50000 = []
for i in range(50000):
    randomList50000.append(random.choice(branches))

myData.insert(count , "branch" , randomList50000)

count = count + 1

myData
```

[371]:
```
       rollNo     first name last name            fullName   branch
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …             …       …                 …       …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

[50000 rows x 5 columns]
```

[372]:
```
# inserting major list
CSE_COE_Majors = ["AI" , "Data Science" , "cloud" , "cyber"]
ECE_ENC_Majors = ["Micro electronic" , "IOT" , "Robotics"]

CSE_COE_Majors_applicables = ["CSE" , "COE" , "CSE-MBA"]
ECE_ENC_Majors_applicables = ["ECE" , "ENC" , "ECE-MBA"]

for i in myData.index:
    if(myData.at[i, "branch"] in CSE_COE_Majors_applicables):
        myData.at[i, "major"] = random.choice(CSE_COE_Majors)
    elif(myData.at[i, "branch"] in ECE_ENC_Majors_applicables):
```

```
        myData.at[i, "major"] = random.choice(ECE_ENC_Majors)
    else:
        myData.at[i, "major"] = numpy.nan

count = count + 1

myData
```

[372]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                 …      …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                  major
0              Robotics
1                 cyber
2                 cyber
3                   NaN
4      Micro electronic
…                 …
49995               NaN
49996                AI
49997          Robotics
49998               NaN
49999          Robotics

[50000 rows x 6 columns]
```

[373]:
```
# adding year col name
for i in myData.index:
        myData.at[i, "year"] = int(random.randint(1 , 4))

count = count + 1

myData
```

[373]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
```

```
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                …       …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                  major  year
0              Robotics   3.0
1                 cyber   2.0
2                 cyber   1.0
3                   NaN   4.0
4      Micro electronic   4.0
…                   …    …
49995               NaN   2.0
49996                AI   4.0
49997          Robotics   2.0
49998               NaN   4.0
49999          Robotics   1.0

[50000 rows x 7 columns]
```

[374]:
```
# adding year col name
for i in myData.index:
        # semester can be either current year * 2 or current year * 2 - 1
        myData.at[i, "sem"] = random.choice([(myData.at[i, "year"] * 2) - 1 ,␣
 ↪myData.at[i, "year"] * 2])

count = count + 1

myData
```

[374]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                …       …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC
```

```
                  major  year  sem
0              Robotics   3.0  6.0
1                 cyber   2.0  3.0
2                 cyber   1.0  2.0
3                   NaN   4.0  8.0
4      Micro electronic   4.0  7.0
…                   …   …  …
49995               NaN   2.0  3.0
49996                AI   4.0  8.0
49997          Robotics   2.0  4.0
49998               NaN   4.0  7.0
49999          Robotics   1.0  1.0

[50000 rows x 8 columns]
```

[375]:
```
# inserting fees col
fees = {
    "CSE" : 200000 , "COE" : 190000 , "ECE" : 170000 , "MEC" : 130000 , "BOT" :␣
 ↪140000, "ENC" : 185000 , "CSE-MBA" : 215000 , "ECE-MBA" : 200000 , "MEC-MBA"␣
 ↪: 180000
}

for i in myData.index:
    myData.at[i, "fees"] = fees.get(myData.at[i, "branch"] , numpy.nan)

count = count + 1

myData
```

[375]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                …       …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                  major  year  sem      fees
0              Robotics   3.0  6.0  170000.0
1                 cyber   2.0  3.0  190000.0
2                 cyber   1.0  2.0  200000.0
```

```
3                   NaN   4.0  8.0  180000.0
4      Micro electronic   4.0  7.0  185000.0
…                    …    …  …       …
49995               NaN   2.0  3.0  180000.0
49996                AI   4.0  8.0  200000.0
49997          Robotics   2.0  4.0  170000.0
49998               NaN   4.0  7.0  130000.0
49999          Robotics   1.0  1.0  185000.0

[50000 rows x 9 columns]
```

[376]:
```
# inserting total program fee col and program duration col

# MBA is of 5 years
for i in myData.index:
    if(str(myData.at[i, "branch"]).find("MBA") != -1):
        myData.at[i, "program duration"] = 5
        myData.at[i, "total program fee"] = myData.at[i, "fees"] * 5 * 2
    else:
        myData.at[i, "program duration"] = 4
        myData.at[i, "total program fee"] = myData.at[i, "fees"] * 4 * 2

count = count + 1

myData
```

[376]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                …      …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                   major  year  sem      fees  program duration  \
0              Robotics   3.0  6.0  170000.0               4.0
1                 cyber   2.0  3.0  190000.0               4.0
2                 cyber   1.0  2.0  200000.0               4.0
3                   NaN   4.0  8.0  180000.0               5.0
4      Micro electronic   4.0  7.0  185000.0               4.0
…                    …    …  …       …                 …
49995               NaN   2.0  3.0  180000.0               5.0
```

```
49996                AI   4.0  8.0  200000.0              4.0
49997          Robotics   2.0  4.0  170000.0              4.0
49998               NaN   4.0  7.0  130000.0              4.0
49999          Robotics   1.0  1.0  185000.0              4.0

       total program fee
0              1360000.0
1              1520000.0
2              1600000.0
3              1800000.0
4              1480000.0
…                    …
49995          1800000.0
49996          1600000.0
49997          1360000.0
49998          1040000.0
49999          1480000.0

[50000 rows x 11 columns]
```

[377]:
```
# inserting paid fee col

for i in myData.index:
       myData.at[i, "paid fee"] = bool(random.randint(0,1))

count = count + 1

myData
```

[377]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                 …      …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                  major  year  sem      fees  program duration  \
0             Robotics   3.0  6.0  170000.0              4.0
1                cyber   2.0  3.0  190000.0              4.0
2                cyber   1.0  2.0  200000.0              4.0
3                  NaN   4.0  8.0  180000.0              5.0
```

```
4      Micro electronic   4.0  7.0  185000.0              4.0
…                   …   …  …       …                 …
49995              NaN   2.0  3.0  180000.0              5.0
49996               AI   4.0  8.0  200000.0              4.0
49997         Robotics   2.0  4.0  170000.0              4.0
49998              NaN   4.0  7.0  130000.0              4.0
49999         Robotics   1.0  1.0  185000.0              4.0

       total program fee paid fee
0              1360000.0     True
1              1520000.0     True
2              1600000.0    False
3              1800000.0     True
4              1480000.0     True
…                    …      …
49995          1800000.0     True
49996          1600000.0     True
49997          1360000.0     True
49998          1040000.0     True
49999          1480000.0     True

[50000 rows x 12 columns]
```

[378]:
```
# inserting CGPA col

for i in myData.index:
       myData.at[i, "CGPA"] = random.random() + random.randint(5 , 9)

count = count + 1

myData
```

[378]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                 …      …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                  major  year  sem      fees  program duration  \
0             Robotics   3.0  6.0  170000.0               4.0
```

```
1                 cyber   2.0  3.0  190000.0              4.0
2                 cyber   1.0  2.0  200000.0              4.0
3                   NaN   4.0  8.0  180000.0              5.0
4      Micro electronic   4.0  7.0  185000.0              4.0
…                   …   …  …       …                 …
49995               NaN   2.0  3.0  180000.0              5.0
49996                AI   4.0  8.0  200000.0              4.0
49997          Robotics   2.0  4.0  170000.0              4.0
49998               NaN   4.0  7.0  130000.0              4.0
49999          Robotics   1.0  1.0  185000.0              4.0

       total program fee paid fee      CGPA
0              1360000.0     True  8.002752
1              1520000.0     True  5.379498
2              1600000.0    False  9.273052
3              1800000.0     True  7.998828
4              1480000.0     True  8.952546
…                    …      …       …
49995          1800000.0     True  8.395532
49996          1600000.0     True  8.578161
49997          1360000.0     True  7.504524
49998          1040000.0     True  8.774756
49999          1480000.0     True  9.961632

[50000 rows x 13 columns]
```

[379]:
```
# inserting grade col

for i in myData.index:
    if(myData.at[i, "CGPA"] > 9):
        myData.at[i, "grade"] = "A"
    elif(myData.at[i, "CGPA"] > 8):
        myData.at[i, "grade"] = "A-"
    elif(myData.at[i, "CGPA"] > 7):
        myData.at[i, "grade"] = "B"
    elif(myData.at[i, "CGPA"] > 6):
        myData.at[i, "grade"] = "B-"
    else:
        myData.at[i, "grade"] = "C"

count = count + 1

myData
```

[379]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
```

```
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                 …      …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                  major  year  sem      fees  program duration  \
0              Robotics   3.0  6.0  170000.0               4.0
1                 cyber   2.0  3.0  190000.0               4.0
2                 cyber   1.0  2.0  200000.0               4.0
3                   NaN   4.0  8.0  180000.0               5.0
4      Micro electronic   4.0  7.0  185000.0               4.0
…                   …   …  …       …                 …
49995               NaN   2.0  3.0  180000.0               5.0
49996                AI   4.0  8.0  200000.0               4.0
49997          Robotics   2.0  4.0  170000.0               4.0
49998               NaN   4.0  7.0  130000.0               4.0
49999          Robotics   1.0  1.0  185000.0               4.0

       total program fee paid fee      CGPA grade
0              1360000.0     True  8.002752    A-
1              1520000.0     True  5.379498     C
2              1600000.0    False  9.273052     A
3              1800000.0     True  7.998828     B
4              1480000.0     True  8.952546    A-
…                    …       …       …   …
49995          1800000.0     True  8.395532    A-
49996          1600000.0     True  8.578161    A-
49997          1360000.0     True  7.504524     B
49998          1040000.0     True  8.774756    A-
49999          1480000.0     True  9.961632     A

[50000 rows x 14 columns]
```

[381]:
```
# inserting hostel
import string

for i in myData.index:
        myData.at[i, "grade"] = random.choice(list(string.ascii_lowercase)[:15])

count = count + 1

myData
```

[381]:
```
       rollNo     first name last name            fullName   branch  \
0      100000       aaric a      aaron       aaric a aaron      ECE
1      100001   dominique j      aaron   dominique j aaron      COE
2      100002    fredrick r      aaron    fredrick r aaron      CSE
3      100003       jarvis       aaron       jarvis  aaron  MEC-MBA
4      100004      lorenzo       aaron      lorenzo  aaron      ENC
…       …            …       …                …      …
49995  149995    leonard  jr    hanton   leonard  jrhanton  MEC-MBA
49996  149996         john       clark         john  clark      CSE
49997  149997       demoy p    johnson     demoy p johnson      ECE
49998  149998    shedrick s       jean     shedrick s jean      MEC
49999  149999      dennis l    mustafa    dennis l mustafa      ENC

                   major  year  sem      fees  program duration  \
0              Robotics   3.0  6.0  170000.0               4.0
1                 cyber   2.0  3.0  190000.0               4.0
2                 cyber   1.0  2.0  200000.0               4.0
3                   NaN   4.0  8.0  180000.0               5.0
4      Micro electronic   4.0  7.0  185000.0               4.0
…                  …   …  …       …                …
49995               NaN   2.0  3.0  180000.0               5.0
49996                AI   4.0  8.0  200000.0               4.0
49997          Robotics   2.0  4.0  170000.0               4.0
49998               NaN   4.0  7.0  130000.0               4.0
49999          Robotics   1.0  1.0  185000.0               4.0

       total program fee paid fee      CGPA grade
0              1360000.0     True  8.002752     m
1              1520000.0     True  5.379498     k
2              1600000.0    False  9.273052     g
3              1800000.0     True  7.998828     l
4              1480000.0     True  8.952546     k
…                    …      …       …   …
49995          1800000.0     True  8.395532     h
49996          1600000.0     True  8.578161     m
49997          1360000.0     True  7.504524     b
49998          1040000.0     True  8.774756     l
49999          1480000.0     True  9.961632     c

[50000 rows x 14 columns]
```

[382]:
```
# exporting to csv

myData.to_csv("studentsData.csv")
```

[ ]:
```

```