p3

February 20, 2024

1 Name: Rutuja Sawant

2 Roll No.: TCOD79

3 Batch: T12

[6]:
```
import pandas as pd
data = pd.read_csv("C:\\Users\\EL18\\Downloads\\archive (1)\\test.csv")
data
```

[6]:

	Age	IsAdult	Job	Qualification	MaritalStatus	Gender	BankBalance	Salary
0	39	1	GovtJob	Bachelor	UnMarried	Male	77516	<=50K
1	50	1	SelfEmployed	Bachelor	Married	Male	83311	<=50K
2	38	1	PrivateJob	Bachelor	Divorced	Male	215646	<=50K
3	53	1	PrivateJob	School	Married	Male	234721	<=50K
4	28	1	PrivateJob	Bachelor	Married	Female	338409	<=50K
…	…	…	…	…	…	…	…	…
32556	27	1	PrivateJob	Bachelor	Married	Female	257302	<=50K
32557	40	1	PrivateJob	Bachelor	Married	Male	154374	>50K
32558	58	1	PrivateJob	Bachelor	Widowed	Female	151910	<=50K
32559	22	1	PrivateJob	Bachelor	UnMarried	Male	201490	<=50K
32560	52	1	SelfEmployed	Bachelor	Married	Female	287927	>50K

[32561 rows x 8 columns]

[7]:
```
data.dropna(inplace=True)
```

[8]:
```
data
```

[8]:
```
        Age  IsAdult           Job Qualification MaritalStatus  Gender  \
0        39        1       GovtJob      Bachelor     UnMarried    Male
1        50        1  SelfEmployed      Bachelor       Married    Male
2        38        1    PrivateJob      Bachelor      Divorced    Male
3        53        1    PrivateJob        School       Married    Male
4        28        1    PrivateJob      Bachelor       Married  Female
…     …       …            …             …           …      …
32556    27        1    PrivateJob      Bachelor       Married  Female
32557    40        1    PrivateJob      Bachelor       Married    Male
32558    58        1    PrivateJob      Bachelor       Widowed  Female
32559    22        1    PrivateJob      Bachelor     UnMarried    Male
32560    52        1  SelfEmployed      Bachelor       Married  Female

        BankBalance Salary
0             77516  <=50K
1             83311  <=50K
2            215646  <=50K
3            234721  <=50K
4            338409  <=50K
…              …    …
32556        257302  <=50K
32557        154374   >50K
32558        151910  <=50K
32559        201490  <=50K
32560        287927   >50K

[32561 rows x 8 columns]
```

[10]:
```
to_drop_col = ["MaritalStatus","Gender"]
data.drop(columns = to_drop_col, inplace=True)
data
```

[10]:
```
        Age  IsAdult           Job Qualification  BankBalance Salary
0        39        1       GovtJob      Bachelor        77516  <=50K
1        50        1  SelfEmployed      Bachelor        83311  <=50K
2        38        1    PrivateJob      Bachelor       215646  <=50K
3        53        1    PrivateJob        School       234721  <=50K
4        28        1    PrivateJob      Bachelor       338409  <=50K
…     …       …            …             …           …     …
32556    27        1    PrivateJob      Bachelor       257302  <=50K
32557    40        1    PrivateJob      Bachelor       154374   >50K
32558    58        1    PrivateJob      Bachelor       151910  <=50K
32559    22        1    PrivateJob      Bachelor       201490  <=50K
```

```
32560   52        1  SelfEmployed      Bachelor       287927   >50K

[32561 rows x 6 columns]
```

[11]:
```
data["BankBalance"].mean()
```

[11]:
```
189778.36651208502
```

[14]:
```
data["BankBalance"].mode()[0]
```

[14]:
```
123011
```

[31]:
```
data["Age"].mode()
```

[31]:
```
0    36
Name: Age, dtype: int64
```

[16]:
```
data["BankBalance"].median()
```

[16]:
```
178356.0
```

[17]:
```
def calculate_mean(col_data):
    total = 0
    for i in col_data:
        total += i
    return (total)/len(col_data)

print(calculate_mean(data["BankBalance"]))
```

```
189778.36651208502
```

[18]:
```
def calculate_median(col_data):
    ns = sorted(col_data)
    mid1, mid2 = 0, 0
    for i in range(len(ns)):
        if(len(ns)%2==0):
            mid1 = ns[(len(ns))//2]
            mid2 = ns[((len(ns))//2) - 1]
            return(mid1+mid2)//2
        else:
            return ns[len(ns)//2]
print(calculate_median(data["BankBalance"]))
```

```
178356
```

[20]:
```
from collections import Counter
def calculate_mode(col_data):
```

```
    cntr = Counter(col_data)
    return max(cntr, key = cntr.get)
print(calculate_mode(data["BankBalance"]))
```

```
123011
```

[24]:
```
data["BankBalance"].min()
```

[24]:
```
12285
```

[25]:
```
data['BankBalance'].max()
```

[25]:
```
1484705
```

[27]:
```
def calculate_min(col_data):
    minval = col_data[0]
    for i in col_data:
        if i < minval:
            minval = i
    print(minval)
calculate_min(data["BankBalance"])
```

```
12285
```

[29]:
```
def calculate_max(col_data):
    maxval = col_data[0]
    for i in col_data:
        if i > maxval:
            maxval = i
    print(maxval)
calculate_max(data["BankBalance"])
```

```
1484705
```

[32]:
```
data["Job"].str.upper()
```

[32]:
```
0             GOVTJOB
1        SELFEMPLOYED
2          PRIVATEJOB
3          PRIVATEJOB
4          PRIVATEJOB
              ...
32556      PRIVATEJOB
32557      PRIVATEJOB
32558      PRIVATEJOB
32559      PRIVATEJOB
32560    SELFEMPLOYED
Name: Job, Length: 32561, dtype: object
```

[33]:
```
data["Job"].str.lower()
```

[33]:
```
0             govtjob
1        selfemployed
2          privatejob
3          privatejob
4          privatejob
             ...
32556      privatejob
32557      privatejob
32558      privatejob
32559      privatejob
32560    selfemployed
Name: Job, Length: 32561, dtype: object
```

[40]:
```
col_data = data["Job"]
cnt=0
for i in col_data:
    if i=='PrivateJob':
        cnt+=1
print("PrivateJob count=", cnt)
```

```
PrivateJob count= 22696
```

[41]:
```
col_data = data["Job"]
cnt=0
for i in col_data:
    if i=='SelfEmployed':
        cnt+=1
print("SelfEmployed count=", cnt)
```

```
SelfEmployed count= 5514
```

[42]:
```
col_data = data["Job"]
cnt=0
for i in col_data:
    if i=='GovtJob':
        cnt+=1
print("GovtJob count=", cnt)
```

```
GovtJob count= 4351
```

[45]:
```
data['Job'].value_counts()
```

[45]:
```
Job
PrivateJob      22696
SelfEmployed     5514
GovtJob          4351
```

```
Name: count, dtype: int64
```

[50]:
```
data[data['Job'].str.contains('GovtJob')]
```

[50]:

	Age	IsAdult	Job	Qualification	BankBalance	Salary
0	39	1	GovtJob	Bachelor	77516	<=50K
11	30	1	GovtJob	Bachelor	141297	>50K
22	35	1	GovtJob	School	76845	<=50K
25	56	1	GovtJob	Bachelor	216851	>50K
30	23	1	GovtJob	Bachelor	190709	<=50K
…	…	…	…	…	…	…
32518	57	1	GovtJob	Bachelor	110417	>50K
32540	45	1	GovtJob	Bachelor	252208	<=50K
32543	45	1	GovtJob	Bachelor	119199	<=50K
32545	39	1	GovtJob	Bachelor	111499	>50K
32549	43	1	GovtJob	Bachelor	255835	<=50K

[4351 rows x 6 columns]

[ ]: