

Name: Priyanka WarudeIn []: `import pandas as pd`In [2]: `data=pd.read_csv("E:\Data files\Brain Stroke.csv")`In [3]: `data`

Out[3]:

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level
0	Male	67.0	0	1	Yes	Private	Urban	
1	Male	80.0	0	1	Yes	Private	Rural	
2	Female	49.0	0	0	Yes	Private	Urban	
3	Female	79.0	1	0	Yes	Self-employed	Rural	
4	Male	81.0	0	0	Yes	Private	Urban	
...
4976	Male	41.0	0	0	No	Private	Rural	
4977	Male	40.0	0	0	Yes	Private	Urban	
4978	Female	45.0	1	0	Yes	Govt_job	Rural	
4979	Male	40.0	0	0	Yes	Private	Rural	
4980	Female	80.0	1	0	Yes	Private	Urban	

4981 rows × 11 columns



1. Find the male and female those age are below 50 and they working as self & government employees

In [4]: `data["work_type"].unique()`Out[4]: `array(['Private', 'Self-employed', 'Govt_job', 'children'], dtype=object)`

```
In [5]: Age=[]
         for x in data["age"]:
             if x<50:
                 Age.append(0)
             else:
                 Age.append(1)
         print(Age)
```

```
In [8]: ad=data.iloc[:,[0,1,6]]
ad
```

```
Out[8]:
```

	gender	age1	work_type
0	Male	1	Private
1	Male	1	Private
2	Female	0	Private
3	Female	1	Self-employed
4	Male	1	Private
...
4976	Male	0	Private
4977	Male	0	Private
4978	Female	0	Govt_job
4979	Male	0	Private
4980	Female	1	Private

4981 rows × 3 columns

```
In [9]: Result=ad.loc[(ad["age1"]==0)&((ad["work_type"]=="Govt_job")|(ad["work_type"]=="Self-employed"))]
```

```
In [10]: Result
```

```
Out[10]:
```

	gender	age1	work_type
28	Male	0	Govt_job
64	Male	0	Self-employed
98	Female	0	Self-employed
188	Female	0	Self-employed
223	Female	0	Govt_job
...
4940	Female	0	Self-employed
4949	Female	0	Govt_job
4950	Male	0	Govt_job
4963	Female	0	Govt_job
4978	Female	0	Govt_job

497 rows × 3 columns

```
In [11]: Result["work_type"].unique()
```

```
Out[11]: array(['Govt_job', 'Self-employed'], dtype=object)
```



```
In [15]: ak=data.loc[(data["New Gender"]==0)&(data["heart_disease"]==1)]
ak
```

Out[15]:

	New Gender	gender	age1	age	hypertension	heart_disease	ever_married	work_type	Residence
9	0	Female	1	61.0	0	1	Yes	Govt_job	
11	0	Female	1	79.0	0	1	Yes	Private	
29	0	Female	1	82.0	1	1	No	Private	
42	0	Female	1	79.0	0	1	Yes	Private	
61	0	Female	1	80.0	0	1	Yes	Self-employed	
...
4955	0	Female	1	69.0	0	1	Yes	Private	
4966	0	Female	1	65.0	0	1	Yes	Private	
4968	0	Female	1	78.0	1	1	Yes	Private	
4969	0	Female	1	70.0	0	1	Yes	Self-employed	
4970	0	Female	1	70.0	0	1	Yes	Self-employed	

112 rows × 13 columns



```
In [16]: D=ak.iloc[:,[0,5,10]]
D
```

Out[16]:

	New Gender	heart_disease	bmi
9	0	1	36.8
11	0	1	28.2
29	0	1	26.5
42	0	1	27.7
61	0	1	21.7
...
4955	0	1	31.2
4966	0	1	29.4
4968	0	1	31.2
4969	0	1	28.6
4970	0	1	30.9

112 rows × 3 columns

```
In [17]: Bmi=[]
for x in D["bmi"]:
    if x>=20 and x<35:
        Bmi.append(x)
    else:
        Bmi.append(0)
print(Bmi)
```

```
[0, 28.2, 26.5, 27.7, 21.7, 32.3, 29.8, 0, 0, 26.1, 0, 27.9, 31.0, 0, 31.0, 0,
27.0, 25.2, 24.9, 30.0, 24.8, 26.0, 34.5, 29.7, 22.8, 32.3, 21.7, 0, 29.4, 34.
4, 22.2, 22.1, 26.0, 26.7, 0, 30.7, 34.2, 0, 31.8, 0, 0, 31.0, 28.7, 27.3, 34.
3, 0, 24.9, 25.1, 29.2, 25.3, 28.8, 28.2, 0, 22.2, 23.7, 28.1, 22.9, 25.4, 25.
6, 29.6, 28.3, 33.4, 29.7, 31.1, 32.2, 0, 32.8, 28.1, 34.6, 0, 29.7, 23.4, 21.
4, 33.5, 24.3, 0, 31.3, 34.5, 0, 0, 0, 26.1, 30.6, 0, 32.5, 25.5, 20.7, 32.9, 2
7.7, 22.7, 0, 28.6, 24.2, 23.3, 31.5, 27.4, 0, 26.4, 24.5, 28.3, 28.8, 31.7, 2
9.0, 33.7, 30.0, 33.5, 28.4, 31.2, 29.4, 31.2, 28.6, 30.9]
```

```
In [18]: D.insert(2,"BMI",Bmi)
D
```

```
Out[18]:
```

	New Gender	heart_disease	BMI	bmi
9	0	1	0.0	36.8
11	0	1	28.2	28.2
29	0	1	26.5	26.5
42	0	1	27.7	27.7
61	0	1	21.7	21.7
...
4955	0	1	31.2	31.2
4966	0	1	29.4	29.4
4968	0	1	31.2	31.2
4969	0	1	28.6	28.6
4970	0	1	30.9	30.9

112 rows × 4 columns

3.Add the new column as per age group.

```
In [19]: data
```

Out[19]:

	New Gender	gender	age1	age	hypertension	heart_disease	ever_married	work_type	Residence
0	1	Male	1	67.0	0	1	Yes	Private	
1	1	Male	1	80.0	0	1	Yes	Private	
2	0	Female	0	49.0	0	0	Yes	Private	
3	0	Female	1	79.0	1	0	Yes	Self-employed	
4	1	Male	1	81.0	0	0	Yes	Private	
...
4976	1	Male	0	41.0	0	0	No	Private	
4977	1	Male	0	40.0	0	0	Yes	Private	
4978	0	Female	0	45.0	1	0	Yes	Govt_job	
4979	1	Male	0	40.0	0	0	Yes	Private	
4980	0	Female	1	80.0	1	0	Yes	Private	

4981 rows × 13 columns




[illegible]


```
In [21]: data.insert(2, "Age1", Age)
data
```

Out[21]:

	New Gender	gender	Age1	age1	age	hypertension	heart_disease	ever_married	work_type
0	1	Male	Oldage	1	67.0	0	1	Yes	Private
1	1	Male	Oldage	1	80.0	0	1	Yes	Private
2	0	Female	Adultage	0	49.0	0	0	Yes	Private
3	0	Female	Oldage	1	79.0	1	0	Yes	Self-employed
4	1	Male	Oldage	1	81.0	0	0	Yes	Private
...
4976	1	Male	Adultage	0	41.0	0	0	No	Private
4977	1	Male	Adultage	0	40.0	0	0	Yes	Private
4978	0	Female	Adultage	0	45.0	1	0	Yes	Govt_job
4979	1	Male	Adultage	0	40.0	0	0	Yes	Private
4980	0	Female	Oldage	1	80.0	1	0	Yes	Private

4981 rows × 14 columns



```
In [22]: del data["New Gender"]
```

```
In [23]: del data["age1"]
```

In [24]: data

Out[24]:

	gender	Age1	age	hypertension	heart_disease	ever_married	work_type	Residence_type
0	Male	Oldage	67.0	0	1	Yes	Private	Urban
1	Male	Oldage	80.0	0	1	Yes	Private	Rural
2	Female	Adulage	49.0	0	0	Yes	Private	Urban
3	Female	Oldage	79.0	1	0	Yes	Self-employed	Rural
4	Male	Oldage	81.0	0	0	Yes	Private	Urban
...
4976	Male	Adulage	41.0	0	0	No	Private	Rural
4977	Male	Adulage	40.0	0	0	Yes	Private	Urban
4978	Female	Adulage	45.0	1	0	Yes	Govt_job	Rural
4979	Male	Adulage	40.0	0	0	Yes	Private	Rural
4980	Female	Oldage	80.0	1	0	Yes	Private	Urban

4981 rows × 12 columns



4. Oldage peoples those are smoking continously and average glucose level above 150

In [25]: `part=data.iloc[:,[1,8,10]]`
`part`

Out[25]:

	Age1	avg_glucose_level	smoking_status
0	Oldage	228.69	formerly smoked
1	Oldage	105.92	never smoked
2	Adulage	171.23	smokes
3	Oldage	174.12	never smoked
4	Oldage	186.21	formerly smoked
...
4976	Adulage	70.15	formerly smoked
4977	Adulage	191.15	smokes
4978	Adulage	95.02	smokes
4979	Adulage	83.94	smokes
4980	Oldage	83.75	never smoked

4981 rows × 3 columns

Out[46]:

75 rows \times 3 columns

b) Adult & oldage females count.

In [27]: data

Out[27]:

	gender	Age1	age	hypertension	heart_disease	ever_married	work_type	Residence_
0	Male	Oldage	67.0	0	1	Yes	Private	U
1	Male	Oldage	80.0	0	1	Yes	Private	U
2	Female	Adultage	49.0	0	0	Yes	Private	U
3	Female	Oldage	79.0	1	0	Yes	Self-employed	U
4	Male	Oldage	81.0	0	0	Yes	Private	U
...
4976	Male	Adultage	41.0	0	0	No	Private	U
4977	Male	Adultage	40.0	0	0	Yes	Private	U
4978	Female	Adultage	45.0	1	0	Yes	Govt_job	U
4979	Male	Adultage	40.0	0	0	Yes	Private	U
4980	Female	Oldage	80.0	1	0	Yes	Private	U

4981 rows × 12 columns

In [28]: N=data.iloc[:,[0,1,6,7]]
N

Out[28]:

	gender	Age1	work_type	Residence_type
0	Male	Oldage	Private	Urban
1	Male	Oldage	Private	Rural
2	Female	Adultage	Private	Urban
3	Female	Oldage	Self-employed	Rural
4	Male	Oldage	Private	Urban
...
4976	Male	Adultage	Private	Rural
4977	Male	Adultage	Private	Urban
4978	Female	Adultage	Govt_job	Rural
4979	Male	Adultage	Private	Rural
4980	Female	Oldage	Private	Urban

4981 rows × 4 columns

```
In [29]: M=N.loc[(N["gender"]=="Female")]
M
```

```
Out[29]:
```

	gender	Age1	work_type	Residence_type
2	Female	Adultage	Private	Urban
3	Female	Oldage	Self-employed	Rural
6	Female	Oldage	Private	Urban
7	Female	Oldage	Private	Urban
8	Female	Oldage	Private	Rural
...
4968	Female	Oldage	Private	Rural
4969	Female	Oldage	Self-employed	Rural
4970	Female	Oldage	Self-employed	Urban
4978	Female	Adultage	Govt_job	Rural
4980	Female	Oldage	Private	Urban

2907 rows × 4 columns

```
In [30]: P=M.loc[(N["Age1"]=="Oldage")|(N["Age1"]=="Adultage")]
P
```

```
Out[30]:
```

	gender	Age1	work_type	Residence_type
2	Female	Adultage	Private	Urban
3	Female	Oldage	Self-employed	Rural
6	Female	Oldage	Private	Urban
7	Female	Oldage	Private	Urban
8	Female	Oldage	Private	Rural
...
4968	Female	Oldage	Private	Rural
4969	Female	Oldage	Self-employed	Rural
4970	Female	Oldage	Self-employed	Urban
4978	Female	Adultage	Govt_job	Rural
4980	Female	Oldage	Private	Urban

2048 rows × 4 columns

```
In [31]: P["Age1"].value_counts()
```

```
Out[31]: Oldage      1191  
         Adultage    857  
         Name: Age1, dtype: int64
```

```
In [32]: A=N.loc[(N["gender"]=="Female")&(N["work_type"]=="Private")&(N["Residence_type"]=="Urban")]  
         A
```

```
Out[32]:
```

	gender	Age1	work_type	Residence_type
2	Female	Adultage	Private	Urban
6	Female	Oldage	Private	Urban
7	Female	Oldage	Private	Urban
10	Female	Oldage	Private	Urban
11	Female	Oldage	Private	Urban
...
4944	Female	Adultage	Private	Urban
4955	Female	Oldage	Private	Urban
4957	Female	Oldage	Private	Urban
4959	Female	Teenage	Private	Urban
4980	Female	Oldage	Private	Urban

858 rows × 4 columns

```
In [33]: del A["Age1"]
```

In [34]: A

Out[34]:

	gender	work_type	Residence_type
2	Female	Private	Urban
6	Female	Private	Urban
7	Female	Private	Urban
10	Female	Private	Urban
11	Female	Private	Urban
...
4944	Female	Private	Urban
4955	Female	Private	Urban
4957	Female	Private	Urban
4959	Female	Private	Urban
4980	Female	Private	Urban

858 rows × 3 columns

6. Self employee peoples those are never smoked have brain stroke.

In [35]: data

Out[35]:

	gender	Age1	age	hypertension	heart_disease	ever_married	work_type	Residence_type
0	Male	Oldage	67.0	0	1	Yes	Private	Urban
1	Male	Oldage	80.0	0	1	Yes	Private	Rural
2	Female	Adultage	49.0	0	0	Yes	Private	Urban
3	Female	Oldage	79.0	1	0	Yes	Self-employed	Rural
4	Male	Oldage	81.0	0	0	Yes	Private	Urban
...
4976	Male	Adultage	41.0	0	0	No	Private	Rural
4977	Male	Adultage	40.0	0	0	Yes	Private	Urban
4978	Female	Adultage	45.0	1	0	Yes	Govt_job	Rural
4979	Male	Adultage	40.0	0	0	Yes	Private	Rural
4980	Female	Oldage	80.0	1	0	Yes	Private	Urban

4981 rows × 12 columns



```
In [36]: Glu=[]
for x in data["avg_glucose_level"]:
    if x>0 and x<=75:
        Glu.append(0)
    elif x>75 and x<=150:
        Glu.append(1)
    else:
        Glu.append(2)
print(Glu)
```

```
[2, 1, 2, 2, 2, 0, 1, 0, 1, 1, 1, 2, 2, 2, 2, 1, 2, 2, 2, 2, 1, 1, 1, 2, 2,
1, 2, 2, 1, 1, 2, 0, 1, 0, 1, 0, 1, 2, 2, 1, 1, 0, 1, 1, 2, 0, 2, 2, 2, 0, 0,
1, 1, 2, 1, 1, 0, 2, 1, 2, 1, 0, 2, 1, 1, 1, 1, 1, 1, 0, 1, 1, 2, 0, 0, 0,
0, 2, 1, 2, 1, 1, 0, 2, 2, 1, 1, 0, 2, 1, 0, 1, 0, 1, 2, 1, 0, 1, 1, 0, 0, 2,
2, 1, 0, 2, 2, 1, 2, 2, 2, 1, 2, 2, 2, 1, 1, 0, 0, 2, 2, 0, 0, 2, 2, 1, 1, 1,
2, 1, 2, 2, 0, 2, 1, 1, 2, 1, 1, 1, 1, 2, 0, 0, 2, 1, 1, 1, 1, 1, 2, 1, 2,
1, 1, 2, 1, 2, 2, 1, 0, 2, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 0, 1, 2, 1, 2, 2,
2, 0, 2, 0, 1, 2, 2, 1, 0, 1, 1, 0, 2, 2, 2, 1, 2, 1, 1, 1, 2, 2, 2, 1, 0, 1,
1, 1, 1, 1, 0, 1, 2, 1, 1, 0, 2, 0, 1, 1, 1, 1, 0, 0, 1, 0, 2, 1, 0, 1, 1, 2,
2, 1, 1, 0, 0, 1, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 2, 1, 1, 1, 0, 2, 1, 1, 1, 2,
1, 1, 1, 2, 0, 1, 1, 1, 1, 1, 2, 0, 0, 1, 2, 0, 1, 1, 0, 1, 1, 2, 2, 1, 1, 1,
2, 2, 2, 1, 1, 1, 2, 1, 1, 1, 1, 2, 0, 1, 1, 1, 0, 1, 2, 1, 1, 0, 1, 2, 1, 1,
0, 1, 1, 1, 1, 2, 1, 0, 1, 1, 0, 1, 1, 1, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1,
2, 1, 1, 0, 1, 1, 1, 0, 1, 1, 2, 1, 1, 2, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0, 0, 2,
1, 1, 1, 0, 0, 2, 1, 1, 1, 1, 1, 0, 1, 2, 0, 1, 2, 1, 2, 1, 1, 2, 1, 1, 1, 1,
1, 1, 1, 2, 0, 0, 1, 1, 0, 1, 1, 1, 1, 1, 1, 2, 0, 1, 1, 0, 1, 1, 1, 1, 2, 2,
1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 0, 1, 2, 1, 1, 1, 1,
1, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 2, 1, 1, 1, 1, 0, 0, 1, 0,
1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 2, 1, 0, 1, 1, 0, 1, 1, 1, 2, 1, 2, 2,
```



```
In [37]: data.insert(8, "Glucose_level", Glu)
data
```

```
Out[37]:
```

	gender	Age1	age	hypertension	heart_disease	ever_married	work_type	Residence_type
0	Male	Oldage	67.0	0	1	Yes	Private	Urban
1	Male	Oldage	80.0	0	1	Yes	Private	Rural
2	Female	Adultage	49.0	0	0	Yes	Private	Urban
3	Female	Oldage	79.0	1	0	Yes	Self-employed	Rural
4	Male	Oldage	81.0	0	0	Yes	Private	Urban
...
4976	Male	Adultage	41.0	0	0	No	Private	Rural
4977	Male	Adultage	40.0	0	0	Yes	Private	Urban
4978	Female	Adultage	45.0	1	0	Yes	Govt_job	Rural
4979	Male	Adultage	40.0	0	0	Yes	Private	Rural
4980	Female	Oldage	80.0	1	0	Yes	Private	Urban

4981 rows × 13 columns



```
In [38]: E=data.iloc[:, [6,8,11,12]]
E
```

```
Out[38]:
```

	work_type	Glucose_level	smoking_status	stroke
0	Private	2	formerly smoked	1
1	Private	1	never smoked	1
2	Private	2	smokes	1
3	Self-employed	2	never smoked	1
4	Private	2	formerly smoked	1
...
4976	Private	0	formerly smoked	0
4977	Private	2	smokes	0
4978	Govt_job	1	smokes	0
4979	Private	1	smokes	0
4980	Private	1	never smoked	0

4981 rows × 4 columns

```
In [39]: F=E.loc[(E["work_type"]=="Self-employed")&(E["smoking_status"]=="never smoked")&(
F
```

```
Out[39]:
```

	work_type	Glucose_level	smoking_status	stroke
3	Self-employed	2	never smoked	1
12	Self-employed	2	never smoked	1
17	Self-employed	2	never smoked	1
18	Self-employed	2	never smoked	1
21	Self-employed	1	never smoked	1
26	Self-employed	2	never smoked	1
40	Self-employed	1	never smoked	1
43	Self-employed	1	never smoked	1
46	Self-employed	2	never smoked	1
56	Self-employed	0	never smoked	1
67	Self-employed	1	never smoked	1
75	Self-employed	0	never smoked	1
110	Self-employed	2	never smoked	1
119	Self-employed	0	never smoked	1
126	Self-employed	1	never smoked	1
128	Self-employed	1	never smoked	1
133	Self-employed	0	never smoked	1
139	Self-employed	1	never smoked	1
142	Self-employed	2	never smoked	1
162	Self-employed	0	never smoked	1
164	Self-employed	1	never smoked	1
165	Self-employed	1	never smoked	1
169	Self-employed	1	never smoked	1
176	Self-employed	1	never smoked	1
178	Self-employed	1	never smoked	1
190	Self-employed	1	never smoked	1
192	Self-employed	0	never smoked	1
4780	Self-employed	2	never smoked	1
4810	Self-employed	0	never smoked	1
4818	Self-employed	0	never smoked	1

7. Age categories having brain strokes

In [40]: data

Out[40]:

	gender	Age1	age	hypertension	heart_disease	ever_married	work_type	Residence_type
0	Male	Oldage	67.0	0	1	Yes	Private	Urban
1	Male	Oldage	80.0	0	1	Yes	Private	Rural
2	Female	Adultage	49.0	0	0	Yes	Private	Urban
3	Female	Oldage	79.0	1	0	Yes	Self-employed	Rural
4	Male	Oldage	81.0	0	0	Yes	Private	Urban
...
4976	Male	Adultage	41.0	0	0	No	Private	Rural
4977	Male	Adultage	40.0	0	0	Yes	Private	Urban
4978	Female	Adultage	45.0	1	0	Yes	Govt_job	Rural
4979	Male	Adultage	40.0	0	0	Yes	Private	Rural
4980	Female	Oldage	80.0	1	0	Yes	Private	Urban

4981 rows × 13 columns



In [41]: z=data.iloc[:,[1,12]]
z

Out[41]:

	Age1	stroke
0	Oldage	1
1	Oldage	1
2	Adultage	1
3	Oldage	1
4	Oldage	1
...
4976	Adultage	0
4977	Adultage	0
4978	Adultage	0
4979	Adultage	0
4980	Oldage	0

4981 rows × 2 columns

```
In [50]: z1=z.loc[(z["stroke"]==1)]  
z1
```

```
Out[50]:
```

	Age1	stroke
0	Oldage	1
1	Oldage	1
2	Adultage	1
3	Oldage	1
4	Oldage	1
...
4815	Oldage	1
4816	Oldage	1
4817	Oldage	1
4818	Oldage	1
4819	Oldage	1

248 rows × 2 columns

```
In [43]: z1["Age1"].value_counts()
```

```
Out[43]: Oldage      226  
Adultage    20  
Child         2  
Name: Age1, dtype: int64
```

```
In [ ]:
```