23mcs1008-healthcare-system-eda

March 1, 2024

[ ]:
```
import matplotlib.pyplot as plt
plt.style.use("seaborn")
import pandas as pd
import numpy as np
```

[ ]:
```
import seaborn as sns
```

[ ]:
```
df=pd.read_csv("patient.csv",header=None)
df.rename(columns={0: 'Patient_id', 1: 'Patient_name', 2: 'patient_gender',3:␣
 ↪'patient_birth_date',4: 'patient_phone',5: 'disease_name',6:'city',7:
 ↪'hospital_id'}, inplace=True)
```

[ ]:
```
df
```

[ ]:
```
    Patient_id Patient_name patient_gender patient_birth_date   patient_phone  \
0       187158       Harbir         Female         1960-02-24  +91 0112009318
1       112766     Brahmdev         Female         1955-05-30  +91 1727749552
2       199252      Ujjawal           Male         1965-12-31  +91 8547451606
3       133424      Ballari         Female         1979-06-11  +91 0106026841
4       172579      Devnath         Female         1982-02-22  +91 1868774631
..         …           …              …                 …             …
65      191132       Dipesh         Female         2014-05-21  +91 5851958964
66      105686          NaN           Male         2012-01-25  +91 7061843400
67      160140       Kishan           Male         1955-06-30  +91 9067652693
68      114252          NaN         Female         1965-05-30  +91 4984346995
69      188365  Bhageeratha           Male         1957-11-14  +91 0590662722

      disease_name          city hospital_id
0     Galactosemia      Rourkela       H1001
1   Bladder cancer  Tiruvottiyur       H1016
2    Kidney cancer     Berhampur       H1009
3          Suicide  Bihar Sharif       H1017
4     Food allergy   Bidhannagar       H1019
..             …           …         …
65       Glaucoma         Kochi       H1016
66      Hepatitis      Kolhapur       H1008
67   Rett Syndrome    Srikakulam       H1002
```

```
68        Diabetes     Ambarnath       H1014
69     Pet allergy       Sonipat       H1017

[70 rows x 8 columns]
```

[ ]:
```
def age(born):
    born = datetime.strptime(born, "%Y-%m-%d").date()
    today = date.today()
    return today.year - born.year - ((today.month, today.day) < (born.month,␣
 ↪born.day))
```

[ ]:
```
from datetime import datetime, date
```

[ ]:
```
df['Age'] = df['patient_birth_date'].apply(age)
```

[ ]:
```
df
```

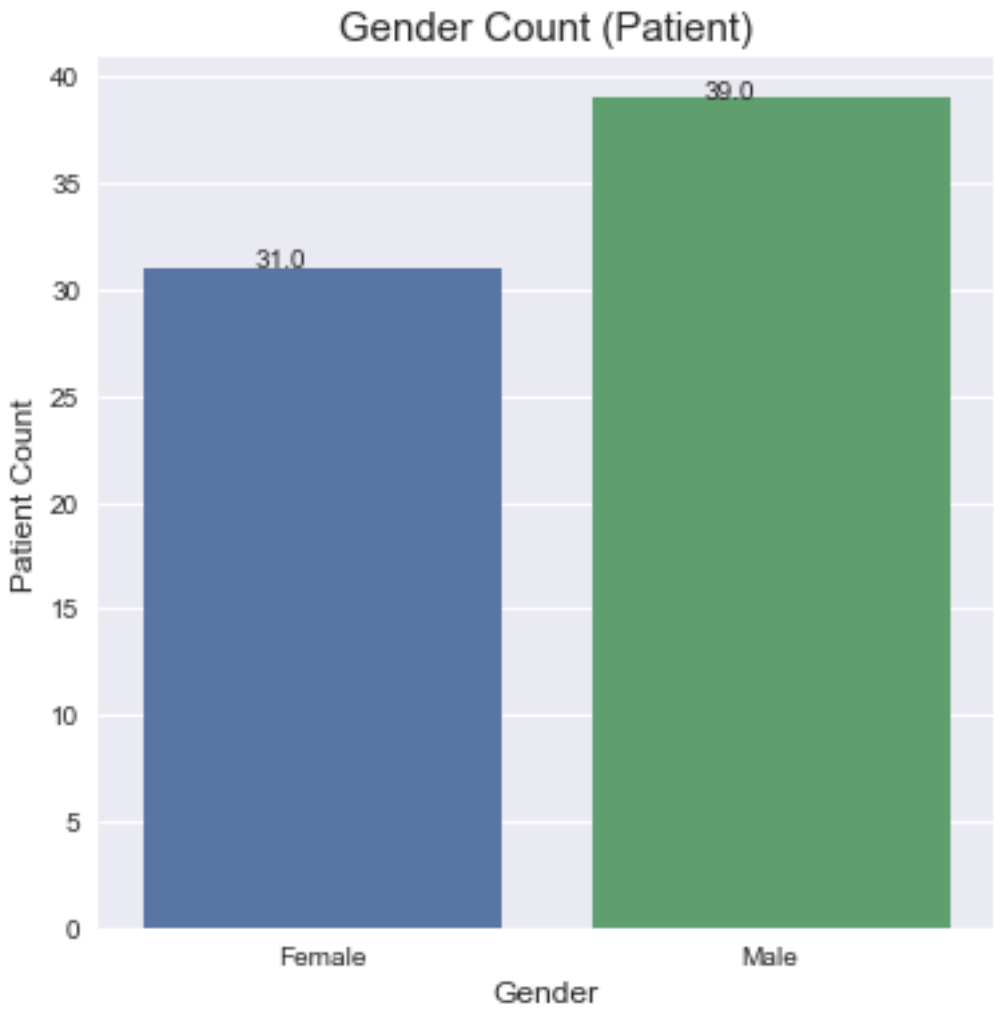
[ ]:
```
    Patient_id Patient_name patient_gender patient_birth_date   patient_phone  \
0       187158       Harbir         Female         1960-02-24  +91 0112009318
1       112766     Brahmdev         Female         1955-05-30  +91 1727749552
2       199252      Ujjawal           Male         1965-12-31  +91 8547451606
3       133424      Ballari         Female         1979-06-11  +91 0106026841
4       172579      Devnath         Female         1982-02-22  +91 1868774631
..         …           …              …                 …              …
65      191132       Dipesh         Female         2014-05-21  +91 5851958964
66      105686          NaN           Male         2012-01-25  +91 7061843400
67      160140       Kishan           Male         1955-06-30  +91 9067652693
68      114252          NaN         Female         1965-05-30  +91 4984346995
69      188365  Bhageeratha           Male         1957-11-14  +91 0590662722

      disease_name          city hospital_id  Age
0     Galactosemia      Rourkela       H1001   62
1   Bladder cancer  Tiruvottiyur       H1016   66
2    Kidney cancer     Berhampur       H1009   56
3          Suicide  Bihar Sharif       H1017   42
4     Food allergy   Bidhannagar       H1019   40
..             …           …         …  …
65        Glaucoma         Kochi       H1016    8
66       Hepatitis      Kolhapur       H1008   10
67   Rett Syndrome    Srikakulam       H1002   66
68        Diabetes     Ambarnath       H1014   56
69     Pet allergy       Sonipat       H1017   64

[70 rows x 9 columns]
```

[ ]:
```
df['Age']
```

[ ]:
```
0     62
1     66
2     56
3     42
4     40
      ..
65     8
66    10
67    66
68    56
69    64
Name: Age, Length: 70, dtype: int64
```

[ ]:
```
plt.figure(figsize=(6,6))
bins=[0,10,20,30,40,50,60,70,80,90,100,110]
a=df['Age']
plt.hist(a, bins, facecolor='crimson', alpha=0.8, edgecolor='k', linewidth=1)
plt.xlabel("Age",fontsize=12)
plt.ylabel("Patient Count",fontsize=12)
plt.locator_params('y', nbins=12)
plt.ylim(ymin=0,ymax=20)
plt.locator_params('x', nbins=10)
plt.xlim(xmin=0,xmax=100)
plt.title("Patient Count (According to Age)",fontsize=15)
plt.savefig('processed_data_graph/PatientCount(Age).png')
```

[ ]:
```
plt.figure(figsize=(6,6))
ax=sns.countplot(x='patient_gender',data=df)
ax.set_ylabel("Patient Count", fontsize = 12)
ax.set_xlabel("Gender", fontsize = 12)
ax.set_title("Gender Count (Patient)", fontsize=15)
for p in ax.patches:
    ax.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+0.25, p.
 ↪get_height()+0.01))
plt.savefig('processed_data_graph/GenderCount(Patient).png')
```



[ ]:
```
js=pd.read_json('claims.json')
```

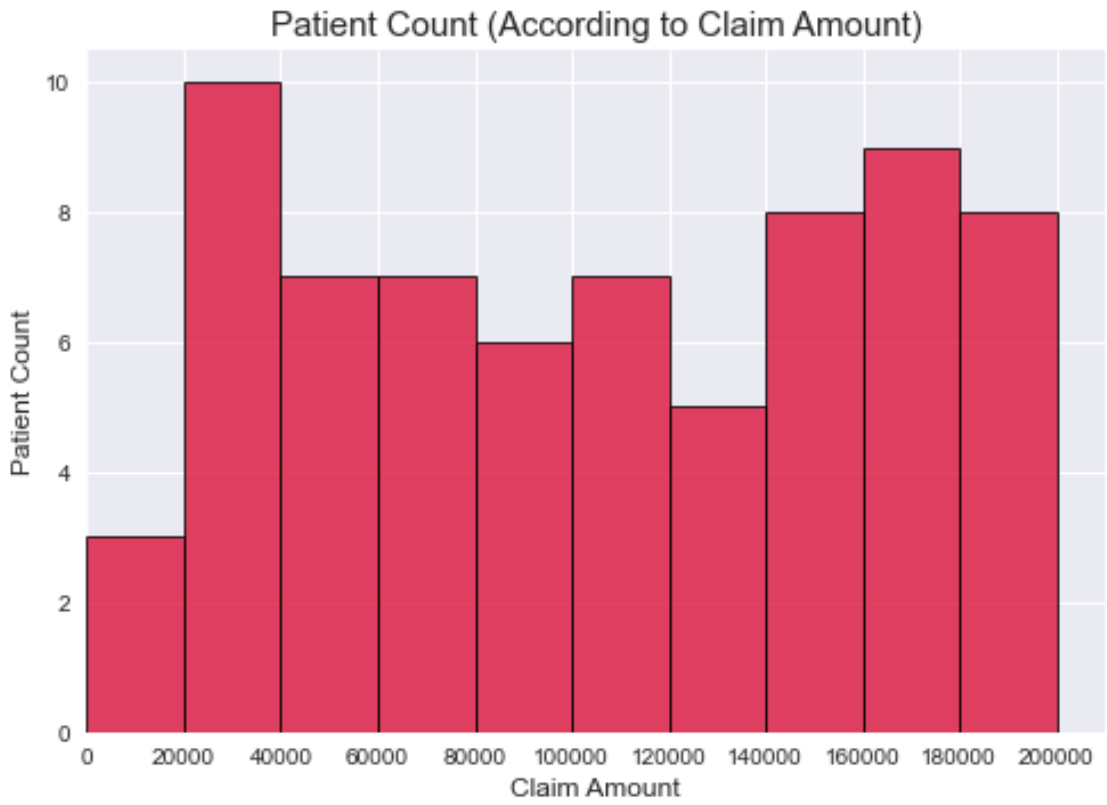
[ ]:
```
js
```

[ ]:
```
    claim_id  patient_id    disease_name      SUB_ID Claim_Or_Rejected  \
0          0      187158    Galactosemia  SUBID10000                 N
1          1      112766  Bladder cancer  SUBID10001                 N
2          2      199252   Kidney cancer  SUBID10002                 N
3          3      133424         Suicide  SUBID10003                 N
4          4      172579    Food allergy  SUBID10004                 Y
..       ...         ...             ...         ...               ...
65        65      191132        Glaucoma   SUBID1065                 Y
66        66      105686       Hepatitis  SUBID10066                 N
67        67      160140   Rett Syndrome   SUBID1067                 N
68        68      114252        Diabetes  SUBID10068                 N
```

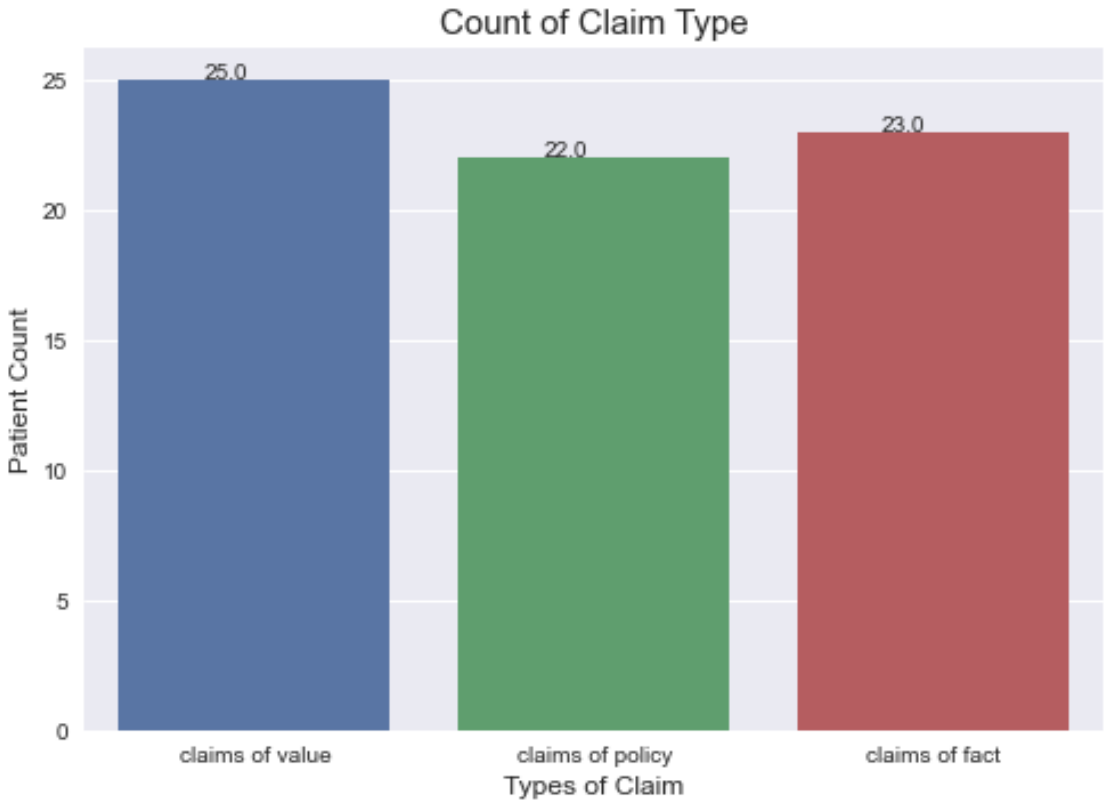
```
69        69      188365     Pet allergy  SUBID10069                N

          claim_type  claim_amount  claim_date
0    claims of value         79874  1949-03-14
1   claims of policy        151142  1970-03-16
2    claims of value         59924  2008-02-03
3     claims of fact        143120  1995-02-08
4    claims of value        168634  1967-05-23
..               …           …         …
65  claims of policy         81980  1969-05-31
66    claims of fact         13667  1957-09-12
67   claims of value        109433  1944-12-25
68  claims of policy        152901  1948-02-13
69    claims of fact         99313  1994-08-25

[70 rows x 8 columns]
```

[ ]:
```
plt.figure(figsize=(5,5))
ax=sns.countplot(x='Claim_Or_Rejected',data=js,hue='Claim_Or_Rejected')
ax.set_ylabel("Count",fontsize=12)
ax.set_xlabel("Status", fontsize=12)
ax.set_title("Claim Status",fontsize=15)
for p in ax.patches:
    ax.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+0.25, p.
 ↪get_height()+0.01))
plt.savefig('processed_data_graph/ClaimStatus.png')
```

[ ]:
```
bins=[0,20000,40000,60000,80000,100000,120000,140000,160000,180000,200000,220000,240000]
a=js['claim_amount']
plt.hist(a, bins, facecolor='crimson', alpha=0.8, edgecolor='k', linewidth=1)
plt.xlabel("Claim Amount",fontsize=12)
plt.ylabel("Patient Count",fontsize=12)
plt.locator_params('x', nbins=12)
plt.xlim(xmin=0,xmax=210000)
plt.title("Patient Count (According to Claim Amount)",fontsize=15)
plt.savefig('processed_data_graph/Patient_count(claimAmt).png')
```



[ ]:
```
ax=sns.countplot(x='claim_type',data=js)
ax.set_ylabel("Patient Count",fontsize=12)
ax.set_xlabel("Types of Claim", fontsize=12)
ax.set_title("Count of Claim Type",fontsize=15)
for p in ax.patches:
    ax.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+0.25, p.
 ↪get_height()+0.01))
plt.savefig('processed_data_graph/CountofClaimType.png')
```



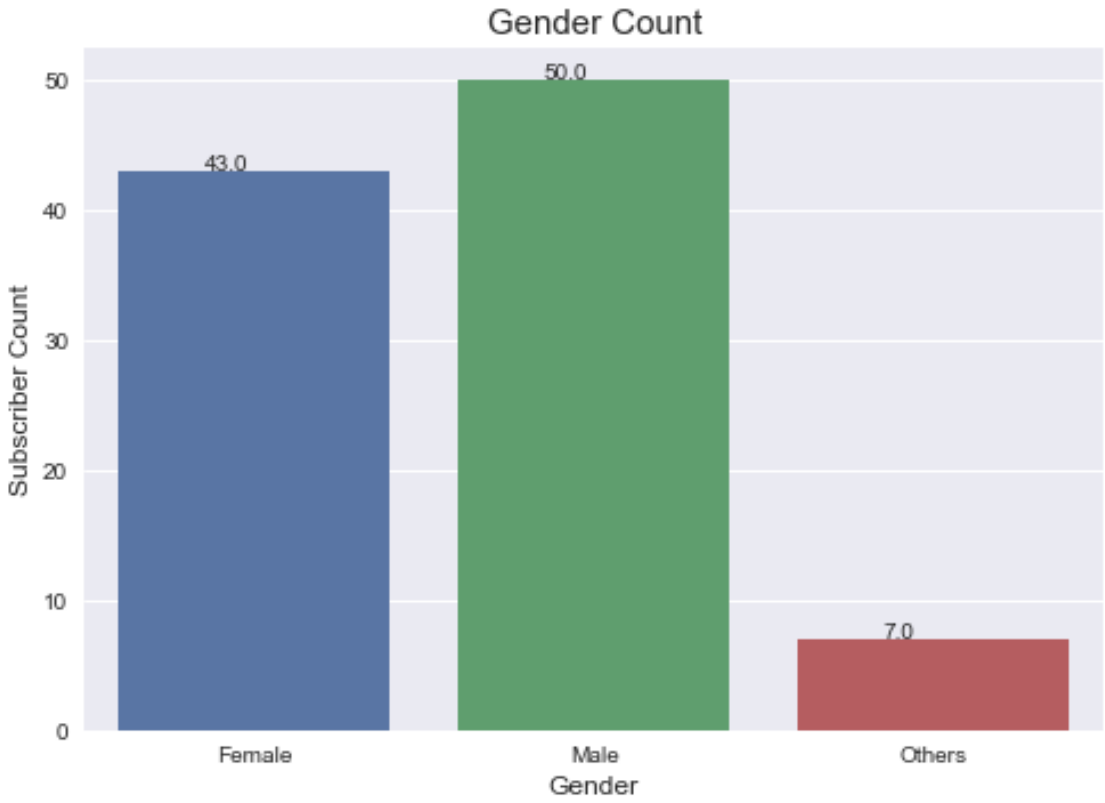
[ ]:
```
df1=pd.read_csv('subscriber.csv')
```

[ ]:
```
df1
```

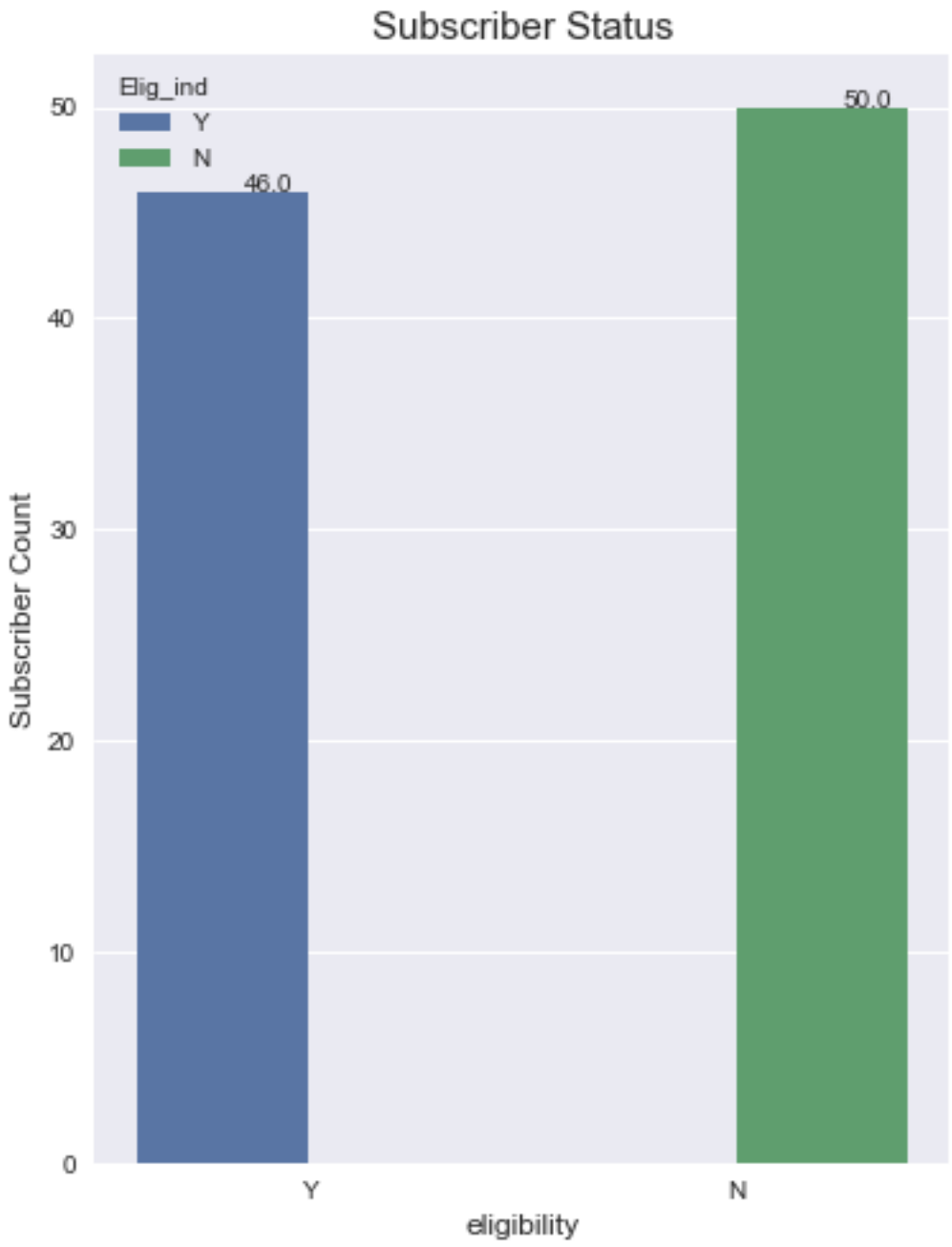
[ ]:
```
     Unnamed: 0     sub _id first_name    last_name       Street  Birth_date  \
0             0  SUBID10000     Harbir  Vishwakarma   Baria Marg  30-06-1924
1             1  SUBID10001   Brahmdev       Sonkar    Lala Marg  20-12-1948
2             2  SUBID10002    Ujjawal         Devi  Mammen Zila  16-04-1980
3             3  SUBID10003    Ballari       Mishra   Sahni Zila  25-09-1969
4             4  SUBID10004    Devnath    Srivastav   Magar Zila  01-05-1946
..          ...         ...        ...          ...          ...         ...
95           95  SUBID10095    Ekaaksh          Rai  Bansal Ganj  02-12-1933
96           96  SUBID10096     Chanak       Sonkar         Kaur  07-04-1959
97           97  SUBID10097        NaN       Sonkar    Rana Ganj  04-02-1940
98           98   SUBID1098    Pushkar        Kumar   Sodhi Zila  05-10-1934
99           99  SUBID10099     Shikha    Srivastav   Ahuja Road  06-09-1970

    Gender           Phone Country                      City  Zip Code  \
0   Female  +91 0112009318   India                   Rourkela    767058
1   Female  +91 1727749552   India                 Tiruvottiyur    34639
2     Male  +91 8547451606   India                    Berhampur    914455
```

```
3   Female  +91 0106026841   India                  Bihar Sharif     91481
4   Female  +91 1868774631   India                   Bidhannagar    531742
..     …              …     …                             …       …
95  Others             NaN   India              Pimpri-Chinchwad    158186
96  Others  +91 7284540687   India  Raurkela Industrial Township    899590
97  Others  +91 8908240160   India                Mira-Bhayandar    896586
98  Others  +91 8956368286   India                         Korba    910732
99  Others  +91 3042509956   India                        Nanded    101500

   Subgrp_id Elig_ind    eff_date   term_date
0       S107        Y  30-06-1944  14-01-1954
1       S105        Y  20-12-1968  16-05-1970
2       S106        N  16-04-2000  04-05-2008
3       S104        N  25-09-1989  05-06-1995
4       S110        N  01-05-1966  09-12-1970
..       …       …         …          …
95      S107        N  02-12-1953  29-07-1960
96      S101        Y  07-04-1979  07-03-1986
97      S107        Y  04-02-1960  12-01-1965
98      S107        Y  05-10-1954  05-04-1961
99      S109        Y  06-09-1990  27-11-1997

[100 rows x 15 columns]
```

[ ]:
```
ax=sns.countplot(x='Gender',data=df1)
ax.set_ylabel("Subscriber Count", fontsize = 12)
ax.set_xlabel("Gender", fontsize = 12)
ax.set_title("Gender Count", fontsize=15)
for p in ax.patches:
    ax.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+0.25, p.
 ↪get_height()+0.01))

plt.savefig('processed_data_graph/Subs_GenderCount.png')
```



[ ]:
```
plt.figure(figsize = (6,8))
ax=sns.countplot(x='Elig_ind',data=df1,hue='Elig_ind')
ax.set_ylabel("Subscriber Count", fontsize = 12)
ax.set_xlabel("eligibility", fontsize = 12)
ax.set_title("Subscriber Status", fontsize=15)
for p in ax.patches:
    ax.annotate('{:.1f}'.format(p.get_height()), (p.get_x()+0.25, p.
 ↪get_height()+0.01))
plt.savefig('processed_data_graph/Subs_Status.png')
```



[ ]:
```
def age1(born):
    born = datetime.strptime(born, "%d-%m-%Y").date()
    today = date.today()
    return today.year - born.year - ((today.month,
                                      today.day) < (born.month,
                                                    born.day))
```

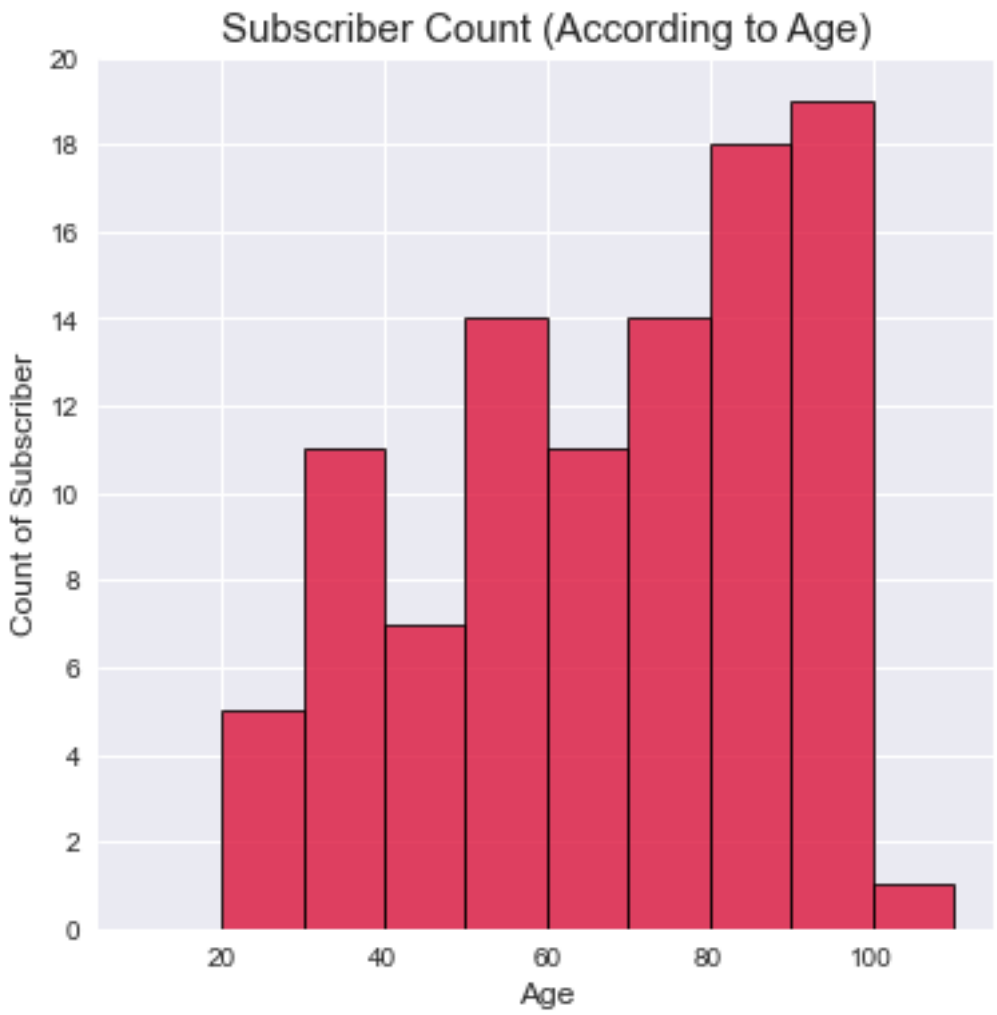
[ ]:
```
df1['Age'] = df1['Birth_date'].apply(age1)
```

[ ]:
```
df1
```

[ ]:
```
    Unnamed: 0     sub _id first_name    last_name       Street  Birth_date  \
0            0  SUBID10000     Harbir  Vishwakarma   Baria Marg  30-06-1924
1            1  SUBID10001   Brahmdev       Sonkar    Lala Marg  20-12-1948
2            2  SUBID10002    Ujjawal         Devi  Mammen Zila  16-04-1980
3            3  SUBID10003    Ballari       Mishra   Sahni Zila  25-09-1969
4            4  SUBID10004    Devnath    Srivastav   Magar Zila  01-05-1946
..         …         …        …           …           …          …
95          95  SUBID10095    Ekaaksh          Rai  Bansal Ganj  02-12-1933
96          96  SUBID10096     Chanak       Sonkar         Kaur  07-04-1959
97          97  SUBID10097        NaN       Sonkar    Rana Ganj  04-02-1940
98          98   SUBID1098    Pushkar        Kumar   Sodhi Zila  05-10-1934
99          99  SUBID10099     Shikha    Srivastav   Ahuja Road  06-09-1970

    Gender           Phone Country                         City  Zip Code  \
0   Female  +91 0112009318   India                      Rourkela    767058
1   Female  +91 1727749552   India                   Tiruvottiyur     34639
2     Male  +91 8547451606   India                      Berhampur    914455
3   Female  +91 0106026841   India                   Bihar Sharif     91481
4   Female  +91 1868774631   India                    Bidhannagar    531742
..     …              …     …                                 …       …
95  Others             NaN   India              Pimpri-Chinchwad    158186
96  Others  +91 7284540687   India  Raurkela Industrial Township    899590
97  Others  +91 8908240160   India               Mira-Bhayandar    896586
98  Others  +91 8956368286   India                         Korba    910732
99  Others  +91 3042509956   India                       Nanded    101500

   Subgrp_id Elig_ind    eff_date   term_date  Age
0       S107        Y  30-06-1944  14-01-1954   97
1       S105        Y  20-12-1968  16-05-1970   73
2       S106        N  16-04-2000  04-05-2008   42
3       S104        N  25-09-1989  05-06-1995   52
4       S110        N  01-05-1966  09-12-1970   76
..       …       …          …          …  …
95      S107        N  02-12-1953  29-07-1960   88
96      S101        Y  07-04-1979  07-03-1986   63
97      S107        Y  04-02-1960  12-01-1965   82
98      S107        Y  05-10-1954  05-04-1961   87
99      S109        Y  06-09-1990  27-11-1997   51

[100 rows x 16 columns]
```

[ ]:
```
plt.figure(figsize=(6,6))
bins=[10,20,30,40,50,60,70,80,90,100,110]
```

```
a=df1['Age']
plt.hist(a, bins, facecolor='crimson', alpha=0.8, edgecolor='k', linewidth=1)
plt.xlabel("Age",fontsize=12)
plt.ylabel("Count of Subscriber",fontsize=12)
plt.locator_params('y', nbins=12)
plt.ylim(ymin=0,ymax=20)
plt.title("Subscriber Count (According to Age)",fontsize=15)
plt.savefig('processed_data_graph/Subs_Count(age).png')
```



[ ]: