

```
In [2]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sb
```

```
In [3]: df = pd.read_csv(r'C:\Users\VARBRU\OneDrive\Desktop\FFINAL_FILES\FFINAL_FILES_CSV\FFINAL_FILES_CSV\3_cons_Dummy_datas.csv')
df=pd.DataFrame(df)
(df)
```

```
Out[3]:
```

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submiti Meth
0	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Interi
1	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Interi
2	45150.0	1002.0	4051.0	76891.0	NaN	31/12/2021	High	Monthly	NaN	Complex	In Process	ACTIVE	N
3	45150.0	1002.0	4053.0	76891.0	Cashless	22/5/2021	Low	Monthly	35,000.00	Standard	Settled	RECOVERED	Interi
4	45150.0	1002.0	4101.0	76891.0	Cashless	8/9/2021	Low	Monthly	36,000.00	Standard	Settled	RECOVERED	Interi
...
79	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Interi
80	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Interi
81	56035.0	1001.0	5035.0	76894.0	Cashless	5/5/2022	Low	Quarterly	19,999.00	Express	In Process	ACTIVE	M
82	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	M
83	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	M

84 rows × 14 columns

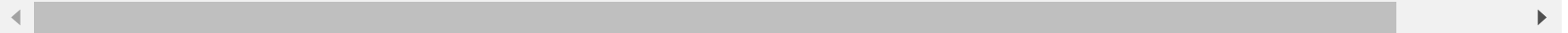


```
In [4]: df = pd.read_csv(r'C:\Users\VARBRU\OneDrive\Desktop\FFINAL_FILES\FFINAL_FILES_CSV\FFINAL_FILES_CSV\3_cons_Dummy_datas.csv')
df=pd.DataFrame(df)
(df)
```

Out[4]:

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submitl Meth
0	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Interi
1	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Interi
2	45150.0	1002.0	4051.0	76891.0	NaN	31/12/2021	High	Monthly	NaN	Complex	In Process	ACTIVE	N
3	45150.0	1002.0	4053.0	76891.0	Cashless	22/5/2021	Low	Monthly	35,000.00	Standard	Settled	RECOVERED	Interi
4	45150.0	1002.0	4101.0	76891.0	Cashless	8/9/2021	Low	Monthly	36,000.00	Standard	Settled	RECOVERED	Interi
...
79	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Interi
80	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Interi
81	56035.0	1001.0	5035.0	76894.0	Cashless	5/5/2022	Low	Quarterly	19,999.00	Express	In Process	ACTIVE	N
82	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	N
83	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	N

84 rows × 14 columns



In [5]:

```
df.isnull() #checking the null values
```

Out[5]:

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submission Method	Country
0	False	False	False	False	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	True	False	False	False	True	False	False	False	True	False
3	False	False	False	False	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	False	False	False	False
...
79	False	False	False	False	False	False	False	False	False	False	False	False	False	False

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submission Method	Country
80	False	False	False	False	False	False	False	False	False	False	False	False	False	False
81	False	False	False	False	False	False	False	False	False	False	False	False	False	False
82	False	False	False	False	False	False	False	False	False	False	False	False	False	False
83	False	False	False	False	False	False	False	False	False	False	False	False	False	False

84 rows × 14 columns

In [6]:

```
duplicate=df[df.duplicated()]
duplicate
```

Out[6]:

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submission Method
1	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Interim
21	45183.0	1002.0	4086.0	76895.0	Cashless	23/1/2022	High	Semi-annually	45,000.00	Accelerated	Settled	RECOVERED	
26	45197.0	1002.0	4100.0	76892.0	Cashless	5/12/2022	High	Semi-annually	23,000.00	Standard	Process	ACTIVE	
36	45123.0	1003.0	4026.0	76891.0	Cashless	18/5/2022	High	Semi-annually	49,500.00	Accelerated	In Process	ACTIVE	
42	45135.0	1003.0	4081.0	76894.0	Reimbursement	11/2/2022	High	Semi-annually	25,500.00	Express	Invalid	ACTIVE	
53	45147.0	1003.0	4050.0	76893.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	
54	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
58	56005.0	1001.0	5055.0	76894.0	Cashless	14/5/2022	Low	Quarterly	45,000.00	Accelerated	Invalid	RECOVERED	
62	56022.0	1001.0	5022.0	76892.0	Reimbursement	12/4/2021	Low	Quarterly	24,100.00	Standard	Settled	RECOVERED	
71	56038.0	1001.0	5067.0	76894.0	Cashless	14/4/2021	High	Quarterly	78,000.00	Standard	In Process	ACTIVE	
75	56047.0	1001.0	5047.0	76891.0	Cashless	10/5/2022	Low	Monthly	37,000.00	Standard	Settled	RECOVERED	Interim

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submit Method
80	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Inter
83	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	

In [7]:

```
remove_duplicates=df.drop_duplicates()
remove_duplicates #drop duplicates
```

Out[7]:

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submit Method
0	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Inter
2	45150.0	1002.0	4051.0	76891.0	NaN	31/12/2021	High	Monthly	NaN	Complex	In Process	ACTIVE	N
3	45150.0	1002.0	4053.0	76891.0	Cashless	22/5/2021	Low	Monthly	35,000.00	Standard	Settled	RECOVERED	Inter
4	45150.0	1002.0	4101.0	76891.0	Cashless	8/9/2021	Low	Monthly	36,000.00	Standard	Settled	RECOVERED	Inter
5	45152.0	1002.0	4055.0	76891.0	Reimbursement	2/2/2022	Low	Quarterly	24,100.00	Standard	Settled	RECOVERED	M
...
77	56013.0	1001.0	5013.0	76892.0	Cashless	5/6/2022	High	Semi-annually	41,000.00	Standard	Invalid	INVALID	M
78	56018.0	1001.0	5018.0	76893.0	Reimbursement	10/10/2021	Low	Semi-annually	59,000.00	Standard	Settled	RECOVERED	M
79	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Inter
81	56035.0	1001.0	5035.0	76894.0	Cashless	5/5/2022	Low	Quarterly	19,999.00	Express	In Process	ACTIVE	M
82	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	M

71 rows × 14 columns

In [8]:

```
drop_null=remove_duplicates.dropna()
```

```
drop_null #Drop Null Values
```

Out[8]:

	Patient ID	Insurance ID	Claim ID	Group ID	Claim Type	Issue Date	Risk Category	Premium Plan Type	Amount	Handling Type	Settlement	Claim Status	Submit Meth
0	45148.0	1002.0	4110.0	76892.0	Cashless	2/4/2022	High	Monthly	40000	Complex	In Process	ACTIVE	Interi
3	45150.0	1002.0	4053.0	76891.0	Cashless	22/5/2021	Low	Monthly	35,000.00	Standard	Settled	RECOVERED	Interi
4	45150.0	1002.0	4101.0	76891.0	Cashless	8/9/2021	Low	Monthly	36,000.00	Standard	Settled	RECOVERED	Interi
5	45152.0	1002.0	4055.0	76891.0	Reimbursement	2/2/2022	Low	Quarterly	24,100.00	Standard	Settled	RECOVERED	N
7	45152.0	1002.0	4103.0	76891.0	Reimbursement	4//5/2022	Low	Quarterly	29,100.00	Standard	Settled	RECOVERED	N
...	
77	56013.0	1001.0	5013.0	76892.0	Cashless	5/6/2022	High	Semi-annually	41,000.00	Standard	Invalid	INVALID	N
78	56018.0	1001.0	5018.0	76893.0	Reimbursement	10/10/2021	Low	Semi-annually	59,000.00	Standard	Settled	RECOVERED	N
79	56021.0	1001.0	5021.0	76891.0	Reimbursement	12/11/2021	High	Monthly	20,500.00	Standard	Settled	ACTIVE	Interi
81	56035.0	1001.0	5035.0	76894.0	Cashless	5/5/2022	Low	Quarterly	19,999.00	Express	In Process	ACTIVE	N
82	56003.0	1001.0	5003.0	76894.0	Reimbursement	23/4/2022	Low	Quarterly	22,500.00	Standard	In Process	RECOVERED	N

61 rows × 14 columns

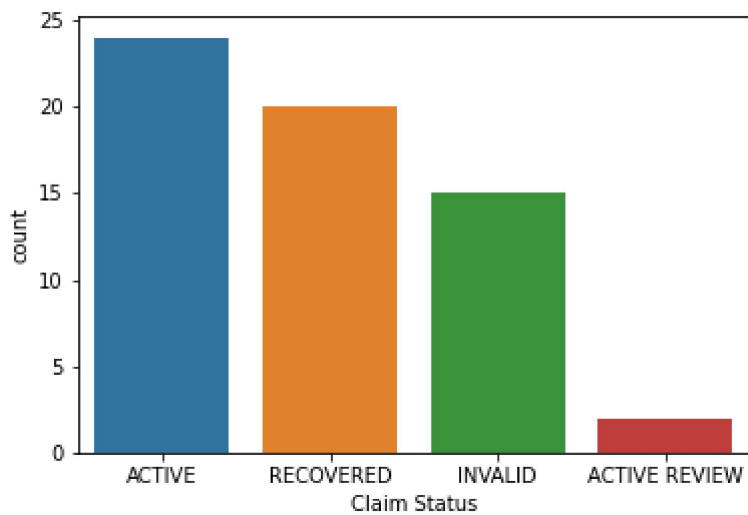


In [9]:

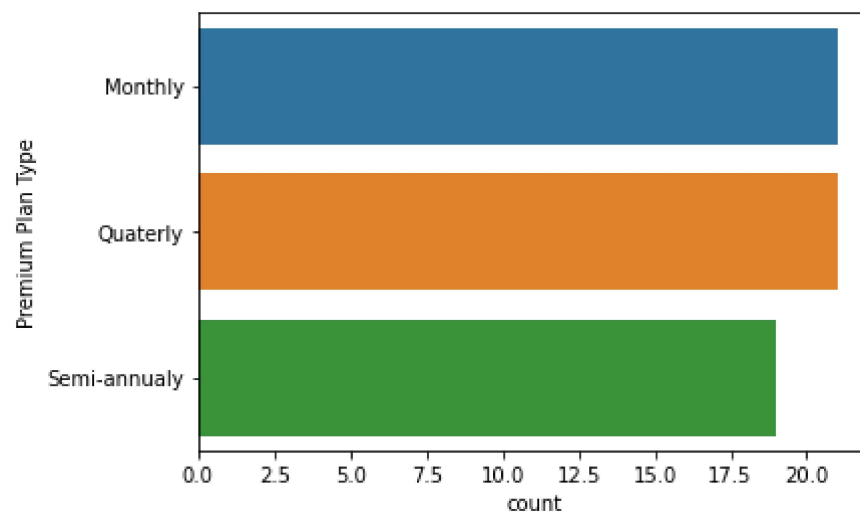
```
drop_null.to_csv('Claims_Final')
```

In [10]:

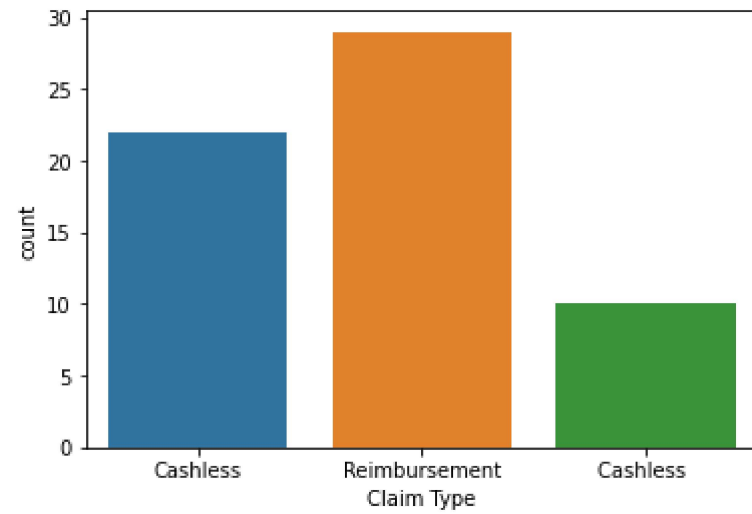
```
sb.countplot(x='Claim Status', data=drop_null)
plt.show()
```



```
In [13]: sb.countplot(y='Premium Plan Type',data=drop_null)
         plp.show()
```



```
In [15]: sb.countplot(x='Claim Type',data=drop_null)
         plp.show()
```



In []: