


```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
```


```
1 df=pd.read_csv("/content/creditcard.csv")
2 df
```

 <ipython-input-51-b9f23f3575cf>:1: DtypeWarning: Columns (16) have mixed types. Specify dtype option on import or setting low\_resolution=False option on plot

	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIncome
0	1.0	0.766127	45.0	Male	South	9120.0
1	0.0	0.957151	40.0	Female	South	2600.0
2	0.0	0.658180	38.0	Female	South	3042.0
3	0.0	0.233810	30.0	Female	South	3300.0
4	0.0	0.907239	49.0	Male	South	63588.0
...	...	...	...	...	...	...
149997	0.0	0.246044	58.0	Male	North	...
149998	0.0	0.000000	30.0	Male	North	5700.0
149999	0.0	0.850283	64.0	Male	North	8100.0
150000	NaN	NaN	NaN	NaN	NaN	...
150001	NaN	NaN	NaN	NaN	NaN	...

150002 rows × 18 columns

```
1 df.head()
```

 DtypeWarning: Columns (16) have mixed types. Specify dtype option on import or setting low\_resolution=False option on plot

	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIncome
0	1.0	0.766127	45.0	Male	South	9120.0
1	0.0	0.957151	40.0	Female	South	2600.0
2	0.0	0.658180	38.0	Female	South	3042.0
3	0.0	0.233810	30.0	Female	South	3300.0
4	0.0	0.907239	49.0	Male	South	63588.0

```
1 df.tail()
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
<b>149997</b>	0.0	0.246044	58.0	Male	North	
<b>149998</b>	0.0	0.000000	30.0	Male	North	57
<b>149999</b>	0.0	0.850283	64.0	Male	North	81
<b>150000</b>	NaN	NaN	NaN	NaN	NaN	
<b>150001</b>	NaN	NaN	NaN	NaN	NaN	

```
1 df.sample(3)
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyInc
<b>25361</b>	0.0	0.392723	32.0	Female	North	440
<b>82689</b>	0.0	0.886759	44.0	Male	Central	1
<b>31071</b>	0.0	0.580165	66.0	Female	North	628

```
1 df.isnull().sum()
```



	0
<b>NPA Status</b>	2
<b>RevolvingUtilizationOfUnsecuredLines</b>	2
<b>age</b>	2
<b>Gender</b>	2
<b>Region</b>	2
<b>MonthlyIncome</b>	29733
<b>Rented_OwnHouse</b>	2
<b>Occupation</b>	2
<b>Education</b>	2
<b>NumberOfTime30-59DaysPastDueNotWorse</b>	2
<b>DebtRatio</b>	2
<b>MonthlyIncome.1</b>	29733
<b>NumberOfOpenCreditLinesAndLoans</b>	2
<b>NumberOfTimes90DaysLate</b>	2
<b>NumberRealEstateLoansOrLines</b>	2
<b>NumberOfTime60-89DaysPastDueNotWorse</b>	2
<b>NumberOfDependents</b>	3924
<b>Good_Bad</b>	2

**dtype:** int64

1 df.columns



```
Index(['NPA Status', 'RevolvingUtilizationOfUnsecuredLines', 'age', 'Gender',
      'Region', 'MonthlyIncome', 'Rented_OwnHouse', 'Occupation', 'Education',
      'NumberOfTime30-59DaysPastDueNotWorse', 'DebtRatio', 'MonthlyIncome.1',
      'NumberOfOpenCreditLinesAndLoans', 'NumberOfTimes90DaysLate',
      'NumberRealEstateLoansOrLines', 'NumberOfTime60-89DaysPastDueNotWorse',
      'NumberOfDependents', 'Good_Bad'],
      dtype='object')
```

```
1 df['MonthlyIncome'] = df['MonthlyIncome'].fillna(df['MonthlyIncome'].mean())
```

```
1 df.sample(10)
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
74479	0.0	1.061292	40.0	Female	Central	6670.22
28591	0.0	0.011196	61.0	Male	North	8580.00
137673	0.0	1.000000	47.0	Male	West	6670.22
35536	0.0	0.066812	82.0	Female	North	2083.00
110165	0.0	0.084765	51.0	Male	South	4000.00
87707	0.0	0.042401	74.0	Male	Central	10171.00
147126	0.0	0.000345	44.0	Male	South	11000.00
34993	0.0	1.000000	39.0	Female	North	5820.00
7771	0.0	0.738082	47.0	Male	South	6890.00
84049	0.0	0.000000	44.0	Male	Central	13000.00

```
1 df['MonthlyIncome.1']=df['MonthlyIncome.1'].fillna(df['MonthlyIncome.1'].mean())
```

```
1 df
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
0	1.0	0.766127	45.0	Male	South	9120.00
1	0.0	0.957151	40.0	Female	South	2600.00
2	0.0	0.658180	38.0	Female	South	3042.00
3	0.0	0.233810	30.0	Female	South	3300.00
4	0.0	0.907239	49.0	Male	South	63588.00
...	...	...	...	...	...	...
149997	0.0	0.246044	58.0	Male	North	6670.22
149998	0.0	0.000000	30.0	Male	North	5716.00
149999	0.0	0.850283	64.0	Male	North	8158.00
150000	NaN	NaN	NaN	NaN	NaN	6670.22
150001	NaN	NaN	NaN	NaN	NaN	6670.22

150002 rows × 18 columns

```
1 df.tail()
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
149997	0.0	0.246044	58.0	Male	North	6670.22
149998	0.0	0.000000	30.0	Male	North	5716.00
149999	0.0	0.850283	64.0	Male	North	8158.00
150000	NaN	NaN	NaN	NaN	NaN	6670.22
150001	NaN	NaN	NaN	NaN	NaN	6670.22

```
1 df.isnull().sum()
```



	0
NPA Status	2
RevolvingUtilizationOfUnsecuredLines	2
age	2
Gender	2
Region	2
MonthlyIncome	0
Rented_OwnHouse	2
Occupation	2
Education	2
NumberOfTime30-59DaysPastDueNotWorse	2
DebtRatio	2
MonthlyIncome.1	0
NumberOfOpenCreditLinesAndLoans	2
NumberOfTimes90DaysLate	2
NumberRealEstateLoansOrLines	2
NumberOfTime60-89DaysPastDueNotWorse	2
NumberOfDependents	3924
Good_Bad	2

dtype: int64

```
1 df.tail()
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
<b>149997</b>	0.0	0.246044	58.0	Male	North	6670.22
<b>149998</b>	0.0	0.000000	30.0	Male	North	5716.00
<b>149999</b>	0.0	0.850283	64.0	Male	North	8158.00
<b>150000</b>	NaN	NaN	NaN	NaN	NaN	6670.22
<b>150001</b>	NaN	NaN	NaN	NaN	NaN	6670.22

```
1 df=df.dropna(axis=0)
```

```
1 df.tail()
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
<b>149995</b>	0.0	0.040674	74.0	Male	North	2100.00
<b>149996</b>	0.0	0.299745	44.0	Male	North	5584.00
<b>149997</b>	0.0	0.246044	58.0	Male	North	6670.22
<b>149998</b>	0.0	0.000000	30.0	Male	North	5716.00
<b>149999</b>	0.0	0.850283	64.0	Male	North	8158.00

```
1 df.isnull().sum()
```



	0
<hr/>	
NPA Status	0
RevolvingUtilizationOfUnsecuredLines	0
age	0
Gender	0
Region	0
MonthlyIncome	0
Rented_OwnHouse	0
Occupation	0
Education	0
NumberOfTime30-59DaysPastDueNotWorse	0
DebtRatio	0
MonthlyIncome.1	0
NumberOfOpenCreditLinesAndLoans	0
NumberOfTimes90DaysLate	0
NumberRealEstateLoansOrLines	0
NumberOfTime60-89DaysPastDueNotWorse	0
NumberOfDependents	0
Good_Bad	0

**dtype:** int64

```
1 df_=df.drop(['MonthlyIncome.1'],axis=1)
```

```
1 df_
```



	NPA Status	RevolvingUtilizationOfUnsecuredLines	age	Gender	Region	MonthlyIn
0	1.0	0.766127	45.0	Male	South	9120.00
1	0.0	0.957151	40.0	Female	South	2600.00
2	0.0	0.658180	38.0	Female	South	3042.00
3	0.0	0.233810	30.0	Female	South	3300.00
4	0.0	0.907239	49.0	Male	South	63588.00
...	...	...	...	...	...	...
149995	0.0	0.040674	74.0	Male	North	2100.00
149996	0.0	0.299745	44.0	Male	North	5584.00
149997	0.0	0.246044	58.0	Male	North	6670.22
149998	0.0	0.000000	30.0	Male	North	5716.00
149999	0.0	0.850283	64.0	Male	North	8158.00

146076 rows × 7 columns

1

```
2 df_new=df_.replace(to_replace="0.0,(df_['MonthlyIncome'].mean()))
```



<ipython-input-69-01a200862ba4>:1: FutureWarning: DataFrame.replace without 'value' a  
df\_new=df\_.replace(to\_replace="0.0,(df\_['MonthlyIncome'].mean()))



```
1 df['MonthlyIncome'].mode()
```



	MonthlyIncome
0	6670.221237

**dtype:** float64

```
1 df_new['MonthlyIncome'].sample(10)
```





	MonthlyIncome
94451	4300.000000
36711	2000.000000
92970	9763.000000
123629	10666.000000
101712	6670.221237
132043	4500.000000
94092	8500.000000
104403	5775.000000
68497	29100.000000
72584	5400.000000

**dtype:** float64

```
1 df_new.to_csv("Credit.csv",index=False)
```

```
1 df_.columns
```

```
2
```



```
Index(['NPA Status', 'RevolvingUtilizationOfUnsecuredLines', 'age', 'Gender',
      'Region', 'MonthlyIncome', 'Rented_OwnHouse', 'Occupation', 'Education',
      'NumberOfTime30-59DaysPastDueNotWorse', 'DebtRatio',
      'NumberOfOpenCreditLinesAndLoans', 'NumberOfTimes90DaysLate',
      'NumberRealEstateLoansOrLines', 'NumberOfTime60-89DaysPastDueNotWorse',
      'NumberOfDependents', 'Good_Bad'],
      dtype='object')
```

```
1 df_new['age']
```



	age
0	45.0
1	40.0
2	38.0

```
1 plt.figure(figsize=(5,5))
2 plt.hist(df_new['age'],bins=5,rwidth=0.7,color='black')
3 plt.title("Age Wise Count")
4 plt.xlabel("Age")
5
6 plt.show()
7 plt.savefig("HistoGram.png")
8 pad_inches=2
9
```

