In [1]:
```
#!pip install plotly
```

In [2]:
```
#import libraries
import pandas as pd
import plotly.express as px
import matplotlib.pyplot as plt
```

In [3]:
```
#load dataset
```

In [4]:
```
dataset=pd.read_csv("preprocessed_data.csv")
dataset
```

Out[4]:

	transaction_id	transaction_amount	location	merchant	age	gender	fraud_label
0	1	1000.0	New York	ABC Corp	35	M	0
1	2	500.0	Chicago	XYZ Inc	45	F	0
2	3	2000.0	Los Angeles	ABC Corp	28	M	1
3	4	1500.0	San Francisco	XYZ Inc	30	F	0
4	5	800.0	Chicago	ABC Corp	50	F	0
...	...	...	...	...	...	...	...
81	82	1500.0	Los Angeles	XYZ Inc	31	M	0
82	83	2800.0	San Francisco	ABC Corp	50	F	1
83	84	1350.0	Chicago	XYZ Inc	28	M	0
84	85	920.0	New York	ABC Corp	47	F	0
85	86	2000.0	Los Angeles	XYZ Inc	36	M	0

86 rows × 7 columns

In [5]:
```
#First we check null values
dataset.isnull().sum()
```

Out[5]:
```
transaction_id        0
transaction_amount    0
location              0
merchant              0
age                   0
gender                0
fraud_label           0
dtype: int64
```

In [6]:
```
dataset.shape
```

Out[6]:
```
(86, 7)
```

In [7]:
```
dataset['gender'].value_counts()['M']
```

Out[7]:
```
43
```

In [8]:
```
dataset['gender'].value_counts()['F']
```

Out[8]:
```
43
```

In [9]:
```
dataset['location'].value_counts()
```

Out[9]:
```
Chicago          22
New York         22
San Francisco    21
Los Angeles      21
Name: location, dtype: int64
```

In [10]:
```
dataset['merchant'].value_counts()
```

Out[10]:
```
XYZ Inc     43
ABC Corp    43
Name: merchant, dtype: int64
```

In [11]:
```
df= dataset[dataset['gender'] == 'M']
df
```

Out[11]:

	transaction_id	transaction_amount	location	merchant	age	gender	fraud_label
0	1	1000.0	New York	ABC Corp	35	M	0
2	3	2000.0	Los Angeles	ABC Corp	28	M	1
5	6	3000.0	New York	XYZ Inc	42	M	1
7	8	900.0	Los Angeles	XYZ Inc	37	M	0
9	10	1800.0	New York	XYZ Inc	48	M	0
11	12	2200.0	Chicago	XYZ Inc	51	M	0
13	14	1600.0	Los Angeles	XYZ Inc	26	M	0
15	16	1200.0	Chicago	XYZ Inc	34	M	0
17	18	1900.0	Los Angeles	XYZ Inc	32	M	0
19	20	4000.0	Chicago	XYZ Inc	38	M	1
21	22	1700.0	Los Angeles	XYZ Inc	49	M	0
23	24	2300.0	Chicago	XYZ Inc	27	M	1
25	26	1400.0	Los Angeles	XYZ Inc	54	M	0
27	28	1100.0	Chicago	XYZ Inc	44	M	0
29	30	2000.0	Los Angeles	XYZ Inc	46	M	0
31	32	2100.0	Chicago	XYZ Inc	43	M	0
33	34	1800.0	Los Angeles	XYZ Inc	29	M	0
35	36	1300.0	Chicago	XYZ Inc	37	M	0
37	38	2000.0	Los Angeles	XYZ Inc	33	M	0
39	40	2400.0	Chicago	XYZ Inc	26	M	0
41	42	1500.0	Los Angeles	XYZ Inc	31	M	0
43	44	1350.0	Chicago	XYZ Inc	28	M	0
45	46	2000.0	Los Angeles	XYZ Inc	36	M	0
47	48	1900.0	Chicago	XYZ Inc	38	M	1
49	50	1750.0	Los Angeles	XYZ Inc	49	M	0
51	52	2300.0	Chicago	XYZ Inc	41	M	0

	transaction_id	transaction_amount	location	merchant	age	gender	fraud_label
53	54	1600.0	Los Angeles	XYZ Inc	39	M	0
55	56	1250.0	Chicago	XYZ Inc	35	M	0
57	58	2200.0	Los Angeles	XYZ Inc	29	M	0
59	60	4000.0	Chicago	XYZ Inc	37	M	1
61	62	1700.0	Los Angeles	XYZ Inc	49	M	0
63	64	2800.0	Chicago	XYZ Inc	27	M	1
65	66	1400.0	Los Angeles	XYZ Inc	54	M	0
67	68	1100.0	Chicago	XYZ Inc	44	M	0
69	70	2000.0	Los Angeles	XYZ Inc	46	M	0
71	72	2100.0	Chicago	XYZ Inc	43	M	0
73	74	1800.0	Los Angeles	XYZ Inc	29	M	0
75	76	1300.0	Chicago	XYZ Inc	37	M	0
77	78	2000.0	Los Angeles	XYZ Inc	33	M	0
79	80	2400.0	Chicago	XYZ Inc	26	M	0
81	82	1500.0	Los Angeles	XYZ Inc	31	M	0
83	84	1350.0	Chicago	XYZ Inc	28	M	0
85	86	2000.0	Los Angeles	XYZ Inc	36	M	0

In [12]:
```
dataset.columns
```

Out[12]:
```
Index(['transaction_id', 'transaction_amount', 'location', 'merchant', 'age',
       'gender', 'fraud_label'],
      dtype='object')
```

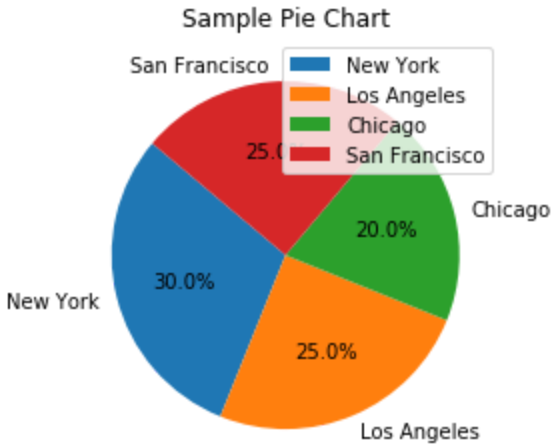
Fraud and Non Fraud Customer details

In [13]:
```
fig = px.histogram(dataset, x='fraud_label', color='fraud_label',
 title= 'Fraud and Non Fraud Customer',
 labels={'TARGET':'fraud_label'})
fig.show()
```

Percentage of Customer loacationWise

In [14]:
```
labels = ['New York', 'Los Angeles', 'Chicago', 'San Francisco']
sizes = [30, 25, 20 ,25]

plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140)
plt.title('Sample Pie Chart')
plt.legend(labels, loc='upper right')
plt.show()
```



What kind of relation between Transacation amount and Age

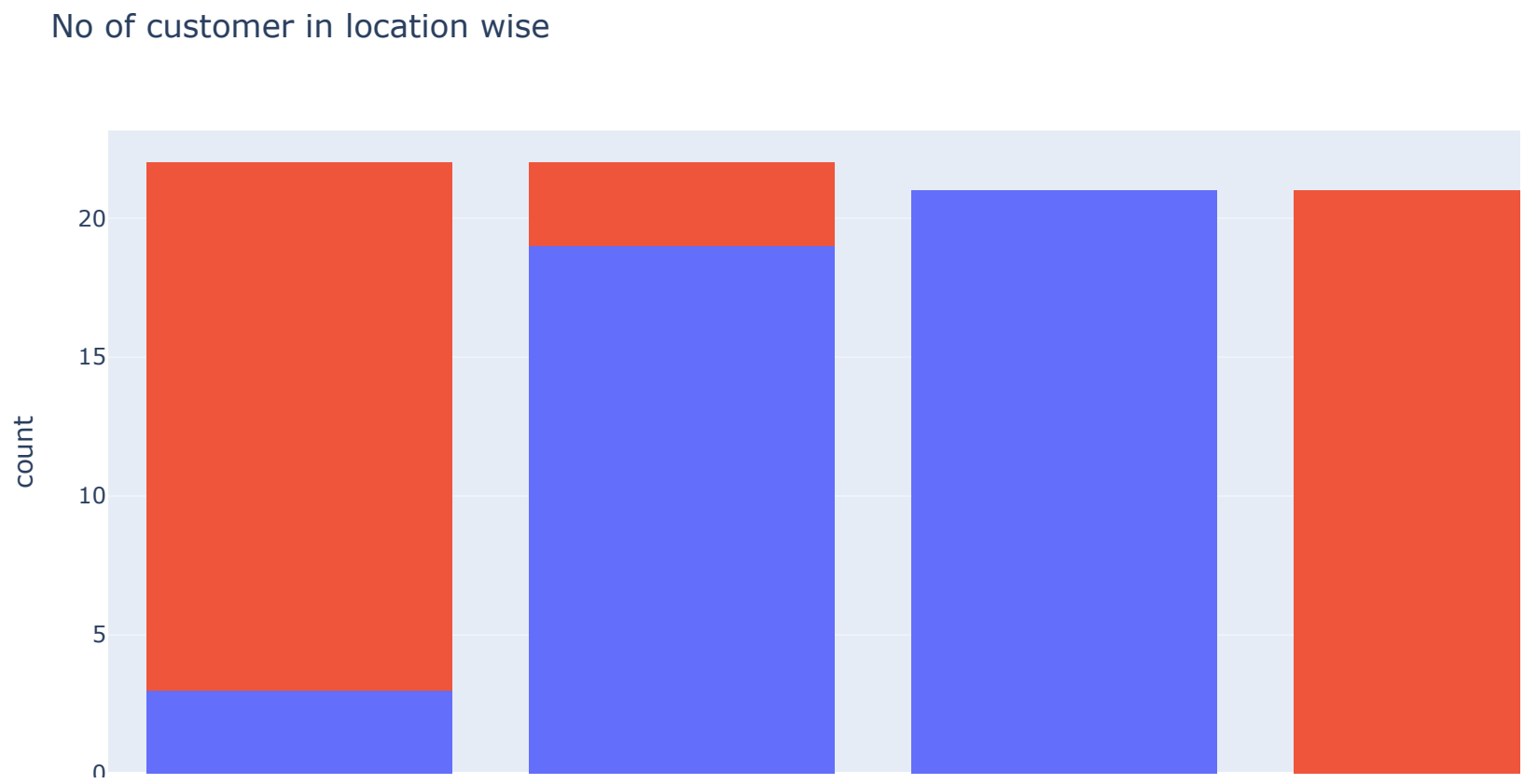
In [15]:
```
dataset.corr()
```

Out[15]:

	transaction_id	transaction_amount	age	fraud_label
transaction_id	1.000000	0.048031	0.009045	-0.074628
transaction_amount	0.048031	1.000000	-0.096590	0.771904
age	0.009045	-0.096590	1.000000	-0.021454
fraud_label	-0.074628	0.771904	-0.021454	1.000000

#Relationship between transaction_amount and age is Negative correlation but its a poor Neagtive value -0.096590

In [16]:
```
fig = px.histogram(dataset, x='location', color='gender',
 title= 'No of customer in location wise').update_xaxes(categoryorder="total descending")
fig.show()
```

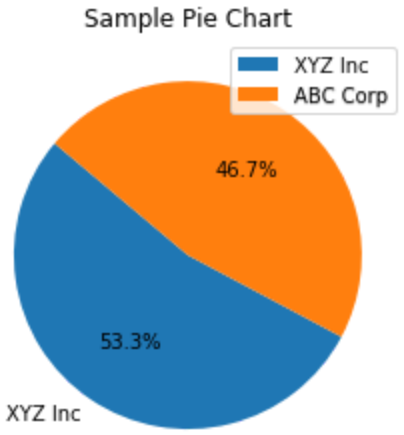


For Los Angeles location we only have a male candidate(21) & san fransisco location we only have a female candidates (21)

Bank Details

In [17]:
```
labels = ['XYZ Inc', 'ABC Corp',]
sizes = [40, 35,]

plt.pie(sizes, labels=labels, autopct='%1.1f%%', startangle=140)
plt.title('Sample Pie Chart')
plt.legend(labels, loc='upper right')
plt.show()
```



In [18]:
```
fig = px.bar(dataset, x='location', y='transaction_amount', color='gender', title= 'locationwise transaction_a
fig.show()
```

In [ ]: