STATISTICS PROJECT
ON
UPI PAYMENT

SUBMITTED TO
Praxis Business School

Submitted By –
Akash Dey – A22005
Kishore Dipra Dutta – A22019
Subhrajyoti Basak – A22036

Exploratory Data Analysis and Descriptive Statistical Analysis on UPI Payment

Introduction:

The Unified Payments Interface is a digital payment system that allows users to transfer funds between bank accounts across India.

Unlike other methods of digital payment and money transfer, to use the UPI payment system, you don't need to enter your bank account number or the recipient's bank details in the app for every transaction. Money transfers through the UPI are processed with a unique virtual ID similar to an e-mail ID. All you need to do is share this ID with anyone who wants to transfer funds to you.

After the implementation of UPI. There is a dramatic change in the payment method. Hence it is important to know the benefits and drawbacks of UPI services and customer perception. This study aims to identify the preference and the impact of UPI on customer satisfaction.

This study will approach to understand, discuss and bring out the issue relevant to the title.

Data Collection and Methodology:

The data is based on primary data and the tool used to get the data was a structured questionnaire with a set of questions. This was circulated online and the respondents consist of people of various age groups, gender, occupations, etc.

Questionnaire:

The set of questions prepared is as follows-

Market Research On Digital Payment

* Required

1. Name *

2. Age *

3. Gender *

Mark only one oval.

- Male
- Female
- Other

4. Occupation *

Mark only one oval.

- Service
- Homemaker
- Student
- Business
- Other

5. Location *

Please mention the state you are currently residing in.

6. Preferred Mode of Payment *

To mention the Mode being used presently

Mark only one oval.

- Card
- BHIM/UPI
- Cash
- Net Banking

7. Mode of Payment pre covid *

To mention the Mode being used before COVID

Mark only one oval.

- Card
- BHIM/UPI
- Cash
- Net Banking

8. Preferred platform for UPI payment *

Mark only one oval.

- Google Pay
- PhonePe
- Paytm Payments Bank App
- Amazon Pay
- Others
- None

9. Average Amount Paid Monthly through the preferred mode *

_____

10. Mode of payment for small amount of transactions (< Rs. 500) *

Mark only one oval.

- UPI
- Cash

11. Number of unsuccessful transactions per 5 transactions *

_____

12. Feedback regarding online payment *

Mark only one oval.

- Unsatisfied
- Average
- Satisfied

Exploratory Data Analysis and Descriptive Statistics of UPI Payment

Importing all the libraries needed for the analysis

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sn
import plotly.express as px
%matplotlib inline
```

In [2]:
```
import warnings
warnings.filterwarnings("ignore")
```

Importing the UPI dataset

In [3]:
```
df = pd.read_csv('UPI_Primary_Data.csv')
```

In [4]:
```
df
```

Out[4]:

	Timestamp	Name	Age	Gender	Occupation	Location	Preferred Mode of Payment	Mode of Payment pre covid	Preferred platform for UPI payment	Average Amount Paid Monthly through the preferred mode	Mode of payment for small amount of transactions (< Rs. 500)	Number of unsuccessful transactions per 5 transactions	Feedback regarding online payment
0	9/14/2022 16:00:46	Debaleena Deb	21	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	7000	UPI	0.0	Satisfied
1	9/13/2022 16:49:58	Niloy Saha	21	Male	Student	Assam	BHIM/UPI	Cash	PhonePe	3000	UPI	0.0	Satisfied
2	9/14/2022 6:28:50	Supriya Deb	21	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	500	UPI	1.0	Satisfied
3	9/13/2022 16:36:25	Papree Ghose	22	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	2000	UPI	2.0	Satisfied
4	9/13/2022 16:45:35	Sourav Das	22	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	2000	UPI	0.0	Satisfied
...	...	...	...	...	...	...	...	...	...	...	...	...	...
114	9/13/2022 19:12:09	Siddhartha Sankar Ghosh	40	Male	Service	West Bengal	BHIM/UPI	Net Banking	Google Pay	10000	UPI	0.0	Satisfied
115	9/13/2022	DRai	42	Male	Service	West	BHIM/UPI	Card	PhonePe	20000	Cash	1.0	Satisfied

115	9/13/2022 18:50:04	DRaj	42	Male	Service	West Bengal	BHIM/UPI	Card	PhonePe	20000	Cash	1.0	Satisfied
116	9/13/2022 16:49:10	Mithun Ranjan Bakchi	43	Male	Service	West Bengal	Card	Card	PhonePe	30000	UPI	1.0	Satisfied
117	9/13/2022 18:22:30	Sulata mondal	43	Female	Other	West Bengal	Cash	Net Banking	Google Pay	500	Cash	2.0	Satisfied
118	9/14/2022 4:15:01	Pradeep Kumar Sarkar	54	Male	Service	West Bengal	Card	Cash	PhonePe	20000	Cash	1.0	Average

119 rows × 13 columns

Data Cleaning

In [5]:
```
df.columns
```

Out[5]:
```
Index(['Timestamp', 'Name', 'Age', 'Gender', 'Occupation', 'Location',
       'Preferred Mode of Payment', 'Mode of Payment pre covid',
       'Preferred platform for UPI payment',
       'Average Amount Paid Monthly through the preferred mode',
       'Mode of payment for small amount of transactions (< Rs. 500 )',
       'Number of unsuccessful transactions per 5 transactions',
       'Feedback regarding online payment'],
      dtype='object')
```

In [6]:
```
df.head()
```

Out[6]:

	Timestamp	Name	Age	Gender	Occupation	Location	Preferred Mode of Payment	Mode of Payment pre covid	Preferred platform for UPI payment	Average Amount Paid Monthly through the preferred mode	Mode of payment for small amount of transactions (< Rs. 500)	Number of unsuccessful transactions per 5 transactions	Feedback regarding online payment
0	9/14/2022 16:00:46	Debaleena Deb	21	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	7000	UPI	0.0	Satisfied
1	9/13/2022 16:49:58	Niloy Saha	21	Male	Student	Assam	BHIM/UPI	Cash	PhonePe	3000	UPI	0.0	Satisfied
2	9/14/2022 6:28:50	Supriya Deb	21	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	500	UPI	1.0	Satisfied
3	9/13/2022 16:36:25	Papree Ghose	22	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	2000	UPI	2.0	Satisfied
4	9/13/2022 16:45:35	Sourav Das	22	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	2000	UPI	0.0	Satisfied

In [7]:
```
UPI_DS = df.drop(['Timestamp', 'Name'], axis = 1)
```

In [8]:
```
UPI_DS.head()
```

Out[8]:

	Age	Gender	Occupation	Location	Preferred Mode of Payment	Mode of Payment pre covid	Preferred platform for UPI payment	Average Amount Paid Monthly through the preferred mode	Mode of payment for small amount of transactions (< Rs. 500)	Number of unsuccessful transactions per 5 transactions	Feedback regarding online payment
0	21	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	7000	UPI	0.0	Satisfied
1	21	Male	Student	Assam	BHIM/UPI	Cash	PhonePe	3000	UPI	0.0	Satisfied
2	21	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	500	UPI	1.0	Satisfied
3	22	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	2000	UPI	2.0	Satisfied
4	22	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	2000	UPI	0.0	Satisfied

Checking the values in all the columns

In [9]:
```
UPI_DS['Age'].value_counts().sort_values()
```

Out[9]:
```
32     1
31     1
20     1
74     1
64     1
60     1
40     1
45     2
36     2
35     2
43     2
62     2
34     2
39     2
54     3
42     3
21     4
33     5
30     5
23     6
29     7
28     9
25     9
22    10
```

```
25     9
22    10
24    10
27    12
26    15
Name: Age, dtype: int64
```

In [10]:
```
UPI_DS['Gender'].value_counts()
```

Out[10]:
```
Male      79
Female    40
Name: Gender, dtype: int64
```

In [11]:
```
UPI_DS['Location'].value_counts()
```

Out[11]:
```
West Bengal        64
Assam              19
Karnataka           5
west Bengal         3
Tamil Nadu          3
Guwahati            2
WB                  2
Uttar Pradesh       2
Telangana           2
Odisha              2
WEST BENGAL         2
Maharastra          2
London, UK          2
Delhi               2
Ontario, Canada     1
Kerala              1
Jharkhand           1
Gujarat             1
Bihar               1
BARASAT             1
New York, USA       1
Name: Location, dtype: int64
```

Correcting the different values which are of same type: like that of West Bengal & Assam

In [12]:
```
UPI_DS['Location'] = UPI_DS['Location'].replace(['WB','west Bengal', 'WEST BENGAL', 'BARASAT', 'West Bengal'],'West Bengal')
```

In [13]:
```
UPI_DS['Location'] = UPI_DS['Location'].replace(['Guwahati','Assam'],'Assam')
```

In [14]:
```
UPI_DS['Location'].value_counts().sort_values()
```

Out[14]:
```
Bihar              1
Gujarat            1
Jharkhand          1
Kerala             1
New York, USA      1
Ontario, Canada    1
Delhi              2
London, UK         2
Maharastra         2
Odisha             2
Telangana          2
Uttar Pradesh      2
Tamil Nadu         3
Karnataka          5
Assam             21
West Bengal       72
Name: Location, dtype: int64
```

Data collected from West Bengal are much higher compared to the rest

In [15]:
```
UPI_DS['Occupation'].value_counts().sort_values()
```

Out[15]:
```
Business      2
Homemaker     4
Other        11
Student      33
Service      69
Name: Occupation, dtype: int64
```

In [16]:
```
UPI_DS['Preferred Mode of Payment'].value_counts().sort_values()
```

Out[16]:
```
Net Banking     4
Cash           10
Card           11
BHIM/UPI       94
Name: Preferred Mode of Payment, dtype: int64
```

In [17]:
```
UPI_DS['Mode of Payment pre covid'].value_counts().sort_values()
```

Out[17]:
```
Net Banking     4
Card           27
BHIM/UPI       43
Cash           45
Name: Mode of Payment pre covid, dtype: int64
```

In [18]:
```
UPI_DS['Preferred platform for UPI payment'].value_counts().sort_values()
```

Out[18]:
```
Amazon Pay                 2
Others                     3
Paytm Payments Bank App    8
None                      10
PhonePe                   38
Google Pay                58
Name: Preferred platform for UPI payment, dtype: int64
```

In [19]:
```
UPI_DS['Average Amount Paid Monthly through the preferred mode'].value_counts().sort_values()
```

Out[19]:
```
1000000     1
1009        1
300         1
80000       1
1500        1
4000        1
75000       1
8000        2
12000       2
7000        2
50000       2
40000       2
25000       3
100000      3
30000       4
100         4
3000        5
500         5
0           6
1000        6
2000        7
20000       8
15000      11
5000       18
10000      22
Name: Average Amount Paid Monthly through the preferred mode, dtype: int64
```

In [20]:
```
UPI_DS['Mode of payment for small amount of transactions (< Rs. 500 )'].value_counts().sort_values()
```

Out[20]:
```
Cash    32
UPI     87
Name: Mode of payment for small amount of transactions (< Rs. 500 ), dtype: int64
```

In [21]:
```
UPI_DS['Number of unsuccessful transactions per 5 transactions'].value_counts().sort_values()
```

Out[21]:
```
0.2     1
3.0     1
5.0     2
4.0     4
2.0     9
0.0    35
1.0    67
Name: Number of unsuccessful transactions per 5 transactions, dtype: int64
```

In [22]:
```
UPI_DS['Feedback regarding online payment'].value_counts().sort_values()
```

Out[22]:
```
Unsatisfied    10
Average        23
Satisfied      86
Name: Feedback regarding online payment, dtype: int64
```

Checking for null values & imputting the null values if any with the mode

In [23]:
```
UPI_DS.isnull().sum()
```

Out[23]:
```
Age                                                    0
Gender                                                 0
Occupation                                             0
Location                                               0
Preferred Mode of Payment                              0
Mode of Payment pre covid                              0
Preferred platform for UPI payment                     0
Average Amount Paid Monthly through the preferred mode 0
Mode of payment for small amount of transactions (< Rs. 500 )    0
Number of unsuccessful transactions per 5 transactions 0
Feedback regarding online payment                      0
dtype: int64
```

In [24]:
```
def replacing_null_values(field_name):
    mode = UPI_DS[field_name].mode()
    UPI_DS[field_name].fillna(mode[0], inplace=True)
```

In [25]:
```
for i in UPI_DS.columns:
    replacing_null_values(i)
```

In [26]:
```
Clean_UPI_DS = UPI_DS
```

In [27]:
```
Clean_UPI_DS.isnull().sum()
```

Out[27]:
```
Age                                                    0
Gender                                                 0
Occupation                                             0
Location                                               0
Preferred Mode of Payment                              0
Mode of Payment pre covid                              0
Preferred platform for UPI payment                     0
Average Amount Paid Monthly through the preferred mode 0
Mode of payment for small amount of transactions (< Rs. 500 )  0
Number of unsuccessful transactions per 5 transactions 0
Feedback regarding online payment                      0
dtype: int64
```

Here we can see there are no null values

In [28]:
```
Clean_UPI_DS
```

Out[28]:

	Age	Gender	Occupation	Location	Preferred Mode of Payment	Mode of Payment pre covid	Preferred platform for UPI payment	Average Amount Paid Monthly through the preferred mode	Mode of payment for small amount of transactions (< Rs. 500)	Number of unsuccessful transactions per 5 transactions	Feedback regarding online payment
0	21	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	7000	UPI	0.0	Satisfied
1	21	Male	Student	Assam	BHIM/UPI	Cash	PhonePe	3000	UPI	0.0	Satisfied
2	21	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	500	UPI	1.0	Satisfied
3	22	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	2000	UPI	2.0	Satisfied
4	22	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	2000	UPI	0.0	Satisfied
...	...	...	...	...	...	...	...	...	...	...	...
114	40	Male	Service	West Bengal	BHIM/UPI	Net Banking	Google Pay	10000	UPI	0.0	Satisfied
115	42	Male	Service	West Bengal	BHIM/UPI	Card	PhonePe	20000	Cash	1.0	Satisfied
116	43	Male	Service	West Bengal	Card	Card	PhonePe	30000	UPI	1.0	Satisfied
117	43	Female	Other	West Bengal	Cash	Net Banking	Google Pay	500	Cash	2.0	Satisfied
118	54	Male	Service	West Bengal	Card	Cash	PhonePe	20000	Cash	1.0	Average

119 rows × 11 columns

Checking datatypes for each column

In [29]:
```
Clean_UPI_DS.dtypes
```

Out[29]:
```
Age                                                 int64
Gender                                             object
Occupation                                         object
Location                                           object
Preferred Mode of Payment                          object
Mode of Payment pre covid                          object
Preferred platform for UPI payment                 object
Average Amount Paid Monthly through the preferred mode      int64
Mode of payment for small amount of transactions (< Rs. 500 )     object
Number of unsuccessful transactions per 5 transactions          float64
Feedback regarding online payment                  object
dtype: object
```

Breaking the columns into categorical and continuous

In [30]:
```
DS_catagorical = Clean_UPI_DS[Clean_UPI_DS.select_dtypes(exclude=['int64','float64']).columns]
```

In [31]:
```
DS_catagorical.dtypes
```

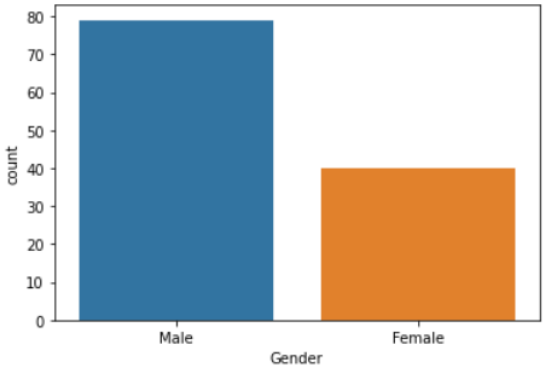
Out[31]:
```
Gender                                             object
Occupation                                         object
Location                                           object
Preferred Mode of Payment                          object
Mode of Payment pre covid                          object
Preferred platform for UPI payment                 object
Mode of payment for small amount of transactions (< Rs. 500 )    object
Feedback regarding online payment                  object
dtype: object
```

In [32]:
```
DS_continuous = Clean_UPI_DS[Clean_UPI_DS.select_dtypes(include=['int64','float64']).columns]
```

In [33]:
```
DS_continuous.dtypes
```

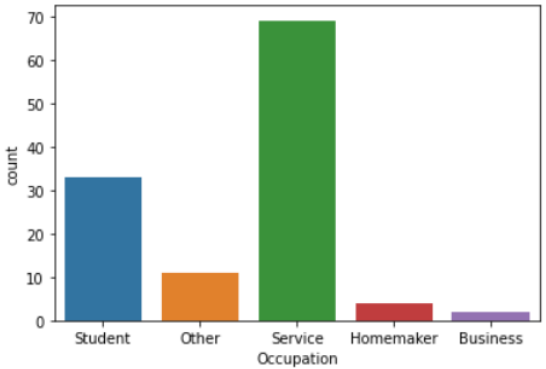
Out[33]:
```
Age                                                 int64
Average Amount Paid Monthly through the preferred mode      int64
Number of unsuccessful transactions per 5 transactions    float64
dtype: object
```

In [34]:
```
sn.countplot(x='Gender',data = Clean_UPI_DS);
```



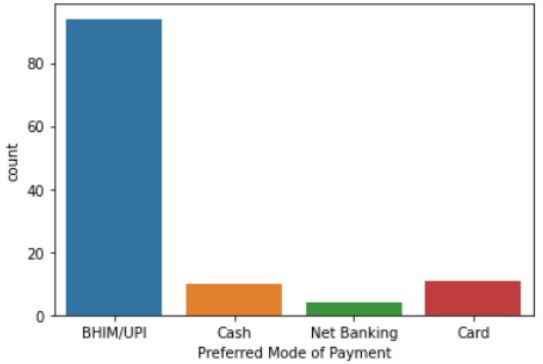
From the data collected, count of Men is higher than that of Female

In [35]:
```
sn.countplot(x='Occupation',data = Clean_UPI_DS);
```



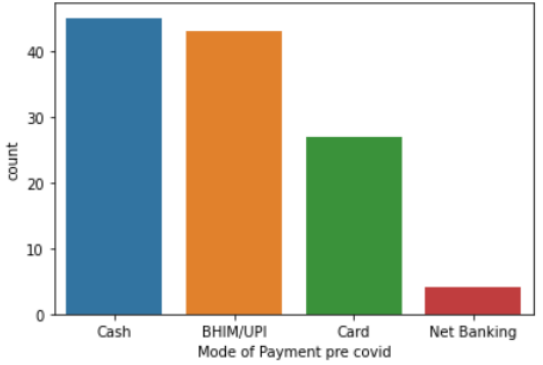
Service/ Occupation has much higher count

In [36]:
```
sn.countplot(x='Preferred Mode of Payment',data = Clean_UPI_DS);
```



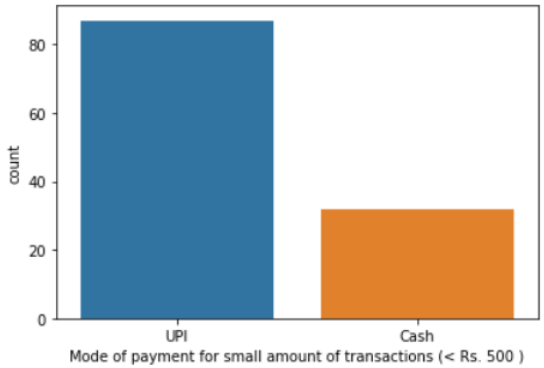
Currently we can see, people are using more BHIm/UPI payments in recent times, i.e. post covid

In [37]:
```
sn.countplot(x='Mode of Payment pre covid',data = Clean_UPI_DS);
```



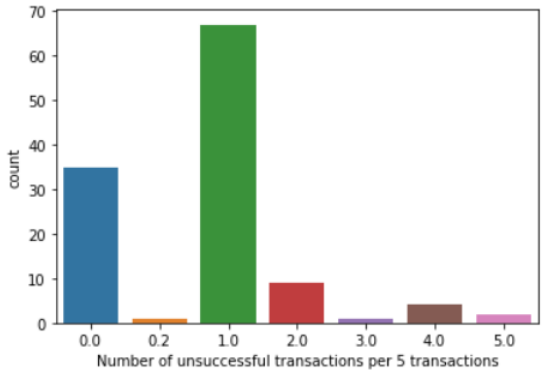
During pre covid times, people were using cash more than the rest payments. Also people were using BHIM/UPI as second payment optin

In [38]:
```
sn.countplot(x='Mode of payment for small amount of transactions (< Rs. 500 )',data = Clean_UPI_DS);
```



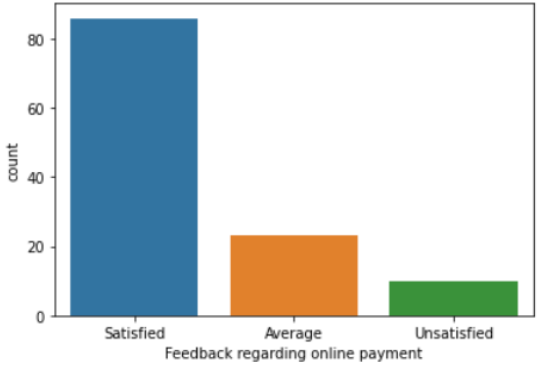
Our dataset shows people are preffering to use UPI payments even for small amount of transactions.

In [39]:
```
sn.countplot(x='Number of unsuccessful transactions per 5 transactions',data = Clean_UPI_DS);
```



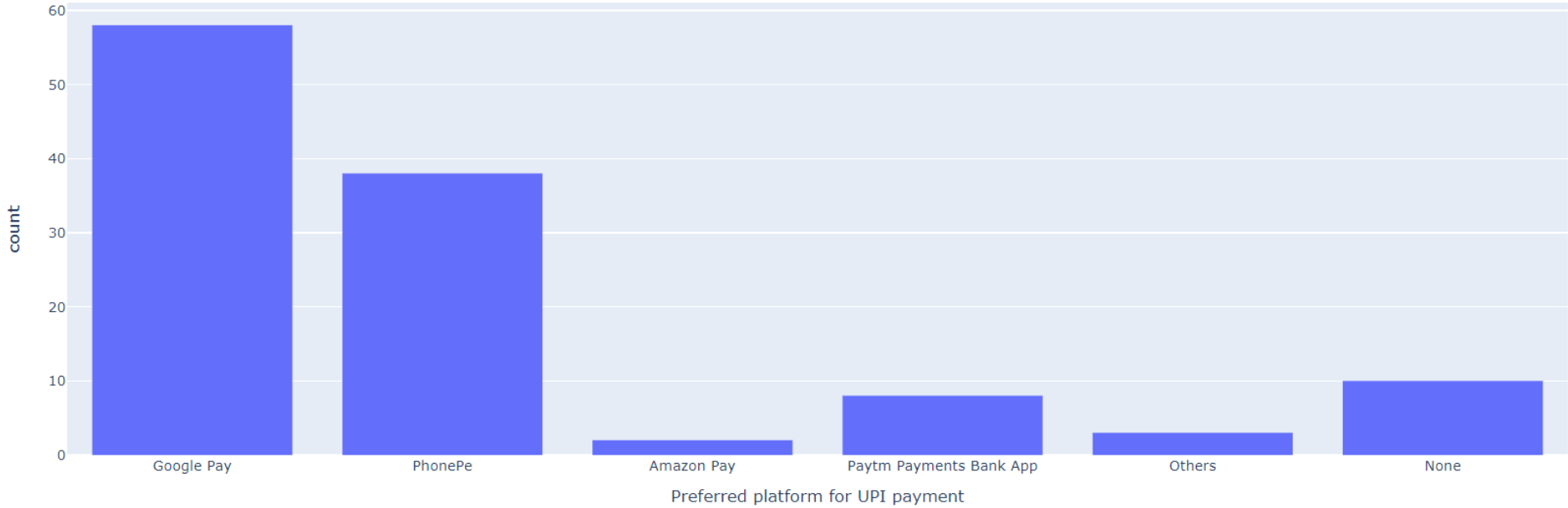
This graph shows the number of times there was unsuccessful transactions per 5 transactios.

In [40]:
```
sn.countplot(x='Feedback regarding online payment',data = Clean_UPI_DS);
```



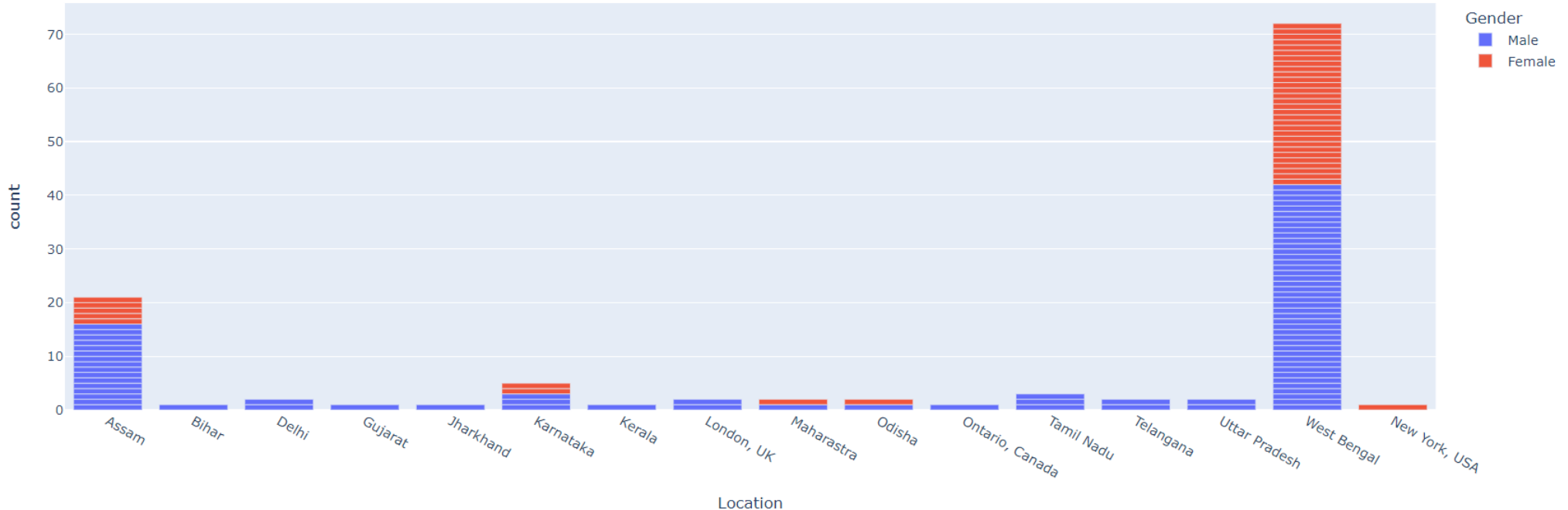
The data suggests that people are highly satisfied with UPI payment platform

In [41]:
```
px.histogram(Clean_UPI_DS, x= 'Preferred platform for UPI payment', nbins = 10)
```

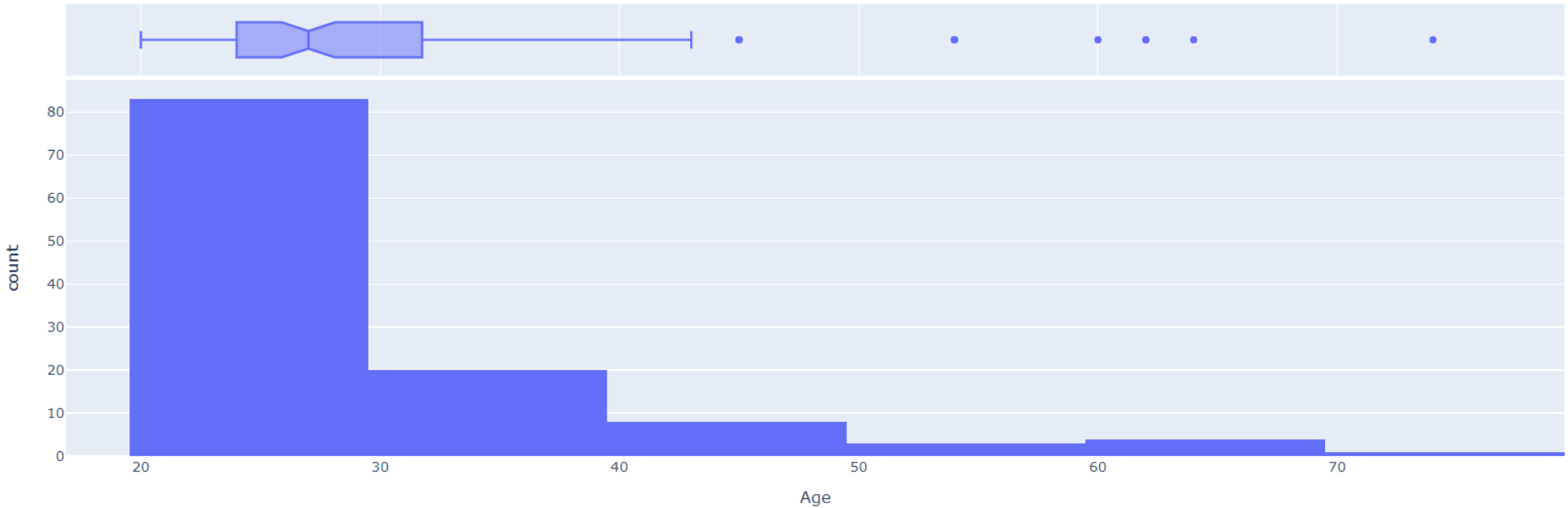


Data shows Google UPI payment as the most preffered platform for UPI payment

In [42]:
```
px.bar(Clean_UPI_DS, x= 'Location', color = 'Gender')
```



In [43]:
```
px.histogram(Clean_UPI_DS, x= 'Age', marginal='box', nbins = 10)
```



The Median Value for Age in the dataset is 27, max is 74 and min is 20. Q1 is 24, Q3 is 31.75, so IQR is (Q3 - Q1) 7.75

Finding the mean, median, min, max of Age by the category 'Gender'

In [44]:
```
score_df1 = Clean_UPI_DS.groupby('Gender').agg({'Age': ['mean', 'median', 'min', 'max']})
score_df1
```

Out[44]:

	Age			
	mean	median	min	max
Gender				
Female	30.550000	26.0	20	74
Male	29.949367	27.0	21	64

Finding the mean, median, min, max of Age by the category 'Preferred Mode of Payment'

In [45]:
```
score_df2 = Clean_UPI_DS.groupby('Preferred Mode of Payment').agg({'Age': ['mean', 'median', 'min', 'max']})
score_df2
```

Out[45]:

	Age			
	mean	median	min	max
Preferred Mode of Payment				
BHIM/UPI	27.776596	26.0	20	62
Card	33.181818	29.0	25	54
Cash	47.900000	48.5	24	74
Net Banking	33.250000	30.5	27	45

Finding the mean, median, min, max of Age by the category 'Mode of Payment pre covid'

Finding the mean, median, min, max of Age by the category 'Mode of Payment pre covid'

In [46]:
```
score_df3 = Clean_UPI_DS.groupby('Mode of Payment pre covid').agg({'Age': ['mean', 'median', 'min', 'max']})
score_df3
```

Out[46]:

	Age			
	mean	median	min	max
Mode of Payment pre covid				
BHIM/UPI	26.488372	26.0	21	39
Card	32.740741	29.0	20	62
Cash	31.444444	26.0	21	74
Net Banking	37.500000	37.0	33	43

Finding the mean, median, min, max of Age by the category 'Preferred platform for UPI payment'

In [47]:
```
score_df4 = Clean_UPI_DS.groupby('Preferred platform for UPI payment').agg({'Age': ['mean', 'median', 'min', 'max']})
score_df4
```

Out[47]:

	Age			
	mean	median	min	max
Preferred platform for UPI payment				
Amazon Pay	24.500000	24.5	22	27
Google Pay	29.103448	27.0	20	64
None	43.100000	40.5	24	74
Others	33.333333	30.0	28	42
Paytm Payments Bank App	30.250000	28.5	22	45
PhonePe	28.368421	26.0	21	54

Finding the mean, median, min, max of Age by the category 'Feedback regarding online payment'

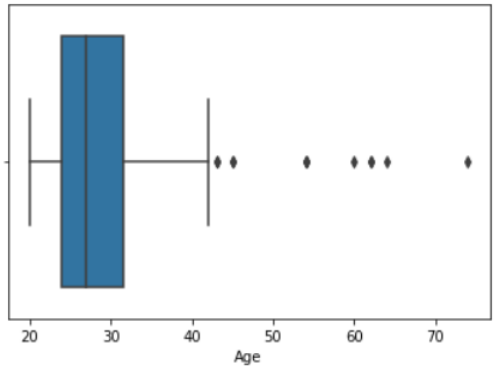
In [48]:
```
score_df6 = Clean_UPI_DS.groupby('Feedback regarding online payment').agg({'Age': ['mean', 'median', 'min', 'max']})
score_df6
```

Out[48]:

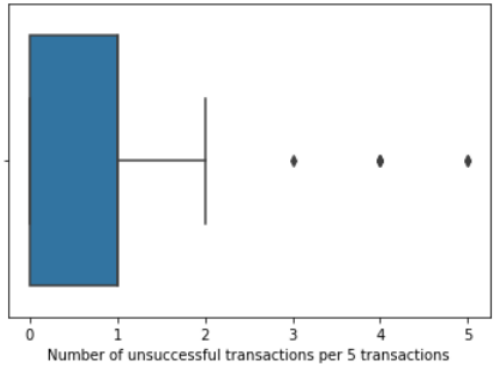
	Age			
	mean	median	min	max
Feedback regarding online payment				
Average	31.826087	27.0	22	64
Satisfied	28.360465	27.0	20	54
Unsatisfied	41.700000	35.5	22	74

Plotting box plot for continuous output

In [49]:
```
sn.boxplot(Clean_UPI_DS['Age']);
```



In [50]:
```
sn.boxplot(Clean_UPI_DS['Number of unsuccessful transactions per 5 transactions']);
```



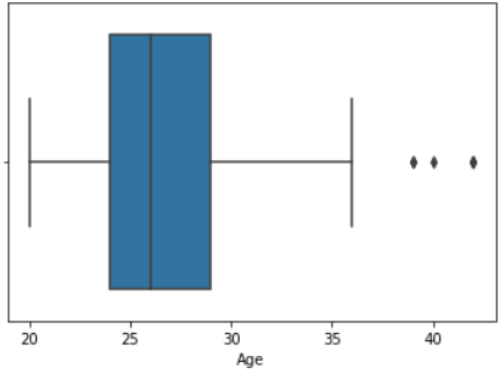
Detecting and Removing Outliers

In [51]:
```
def removing_outliers(field_name, DS_continuous = DS_continuous):
    iqr = DS_continuous[field_name].quantile(0.75) - DS_continuous[field_name].quantile(0.25)
    q1 = DS_continuous[field_name].quantile(0.25)
    q3 = DS_continuous[field_name].quantile(0.75)
    upp_lim = q3 + 1.5*iqr
#     print(field_name upper limit =' + upp_lim)
    low_lim = q1 - 1.5*iqr
#     print(field_name + ' lower limit ' + '=' + low_lim)
    print(field_name + ' - before removing outliers ')
    print(DS_continuous[field_name].count())
    DS_continuous = DS_continuous[low_lim <= DS_continuous[field_name]]
    DS_continuous = DS_continuous[DS_continuous[field_name] <= upp_lim]
    print(field_name + ' - after removing outliers ')
    print(DS_continuous[field_name].count())
    return DS_continuous
```

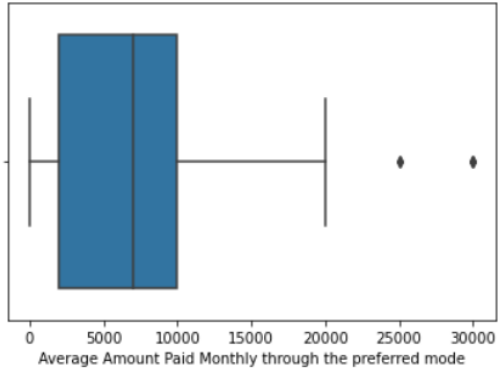
In [52]:
```
for i in DS_continuous.columns:
    DS_continuous = removing_outliers(i, DS_continuous)
```

```
Age - before removing outliers
119
Age - after removing outliers
107
Average Amount Paid Monthly through the preferred mode - before removing outliers
107
Average Amount Paid Monthly through the preferred mode - after removing outliers
97
Number of unsuccessful transactions per 5 transactions - before removing outliers
97
Number of unsuccessful transactions per 5 transactions - after removing outliers
93
```

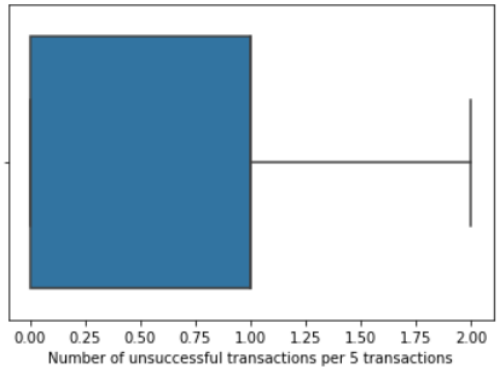
In [53]:
```
sn.boxplot(DS_continuous['Age']);
```



In [54]:
```
sn.boxplot(DS_continuous['Average Amount Paid Monthly through the preferred mode']);
```



In [55]:
```
sn.boxplot(DS_continuous['Number of unsuccessful transactions per 5 transactions']);
```



Dividing the dataset by Age.

- Age <= 30 DS_subset1
- Age = 30 DS_subset2

In [56]:
```
DS_subset1 = Clean_UPI_DS[Clean_UPI_DS['Age'] <= 30]
```

In [57]:
```
DS_subset2 = Clean_UPI_DS[Clean_UPI_DS['Age'] > 30]
```

In [58]:
```
DS_subset1
```

Out[58]:

	Age	Gender	Occupation	Location	Preferred Mode of Payment	Mode of Payment pre covid	Preferred platform for UPI payment	Average Amount Paid Monthly through the preferred mode	Mode of payment for small amount of transactions (< Rs. 500)	Number of unsuccessful transactions per 5 transactions	Feedback regarding online payment
0	21	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	7000	UPI	0.0	Satisfied
1	21	Male	Student	Assam	BHIM/UPI	Cash	PhonePe	3000	UPI	0.0	Satisfied
2	21	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	500	UPI	1.0	Satisfied
3	22	Female	Student	Assam	BHIM/UPI	Cash	PhonePe	2000	UPI	2.0	Satisfied
4	22	Male	Student	Assam	BHIM/UPI	Cash	Google Pay	2000	UPI	0.0	Satisfied
...	...	...	...	...	...	...	...	...	...	...	...
102	28	Male	Service	West Bengal	BHIM/UPI	Card	Paytm Payments Bank App	1000000	UPI	5.0	Satisfied

103	28	Male	Service	West Bengal	BHIM/UPI	BHIM/UPI	Google Pay	10000	UPI	1.0	Satisfied
104	29	Male	Service	West Bengal	BHIM/UPI	Card	Google Pay	40000	Cash	0.0	Satisfied
105	29	Male	Service	West Bengal	Card	Card	PhonePe	15000	UPI	1.0	Satisfied
106	30	Male	Student	West Bengal	BHIM/UPI	BHIM/UPI	PhonePe	25000	UPI	1.0	Satisfied

88 rows × 11 columns

In [59]:
```
DS_subset2
```

Out[59]:

	Age	Gender	Occupation	Location	Preferred Mode of Payment	Mode of Payment pre covid	Preferred platform for UPI payment	Average Amount Paid Monthly through the preferred mode	Mode of payment for small amount of transactions (< Rs. 500)	Number of unsuccessful transactions per 5 transactions	Feedback regarding online payment
17	45	Male	Service	Assam	Net Banking	Cash	Paytm Payments Bank App	25000	Cash	0.0	Satisfied
18	54	Male	Service	Assam	BHIM/UPI	Card	Google Pay	7000	Cash	0.0	Satisfied
20	45	Male	Service	Assam	BHIM/UPI	Card	Google Pay	10000	UPI	1.0	Average
23	36	Male	Service	Delhi	BHIM/UPI	Card	Paytm Payments Bank App	12000	UPI	0.0	Satisfied
30	35	Male	Service	Karnataka	BHIM/UPI	Card	PhonePe	100000	UPI	1.0	Average
33	32	Male	Service	London, UK	BHIM/UPI	BHIM/UPI	PhonePe	5000	UPI	0.0	Satisfied
42	33	Male	Service	Tamil Nadu	BHIM/UPI	Net Banking	Google Pay	1009	UPI	2.0	Satisfied
46	33	Male	Student	Uttar Pradesh	BHIM/UPI	Cash	Paytm Payments Bank App	10000	UPI	0.2	Average
63	33	Male	Service	West Bengal	BHIM/UPI	BHIM/UPI	Google Pay	5000	UPI	0.0	Satisfied
64	33	Male	Service	West Bengal	BHIM/UPI	Cash	Google Pay	10000	UPI	1.0	Satisfied
65	39	Male	Service	West Bengal	BHIM/UPI	Cash	None	0	UPI	0.0	Satisfied

66	42	Female	Service	West Bengal	Card	Cash	Others	1000	Cash	1.0	Satisfied
67	42	Female	Homemaker	West Bengal	Cash	Cash	None	100	Cash	2.0	Unsatisfied
68	54	Female	Homemaker	West Bengal	Cash	Cash	None	0	Cash	0.0	Unsatisfied
69	60	Female	Homemaker	West Bengal	Cash	Cash	None	0	Cash	1.0	Unsatisfied
70	62	Male	Service	West Bengal	BHIM/UPI	Card	Google Pay	15000	Cash	1.0	Average
71	62	Female	Homemaker	West Bengal	Cash	Cash	None	0	Cash	1.0	Unsatisfied
72	64	Male	Other	West Bengal	Cash	Cash	Google Pay	5000	Cash	2.0	Average
73	74	Female	Other	West Bengal	Cash	Cash	None	0	Cash	0.0	Unsatisfied
107	31	Female	Service	West Bengal	BHIM/UPI	BHIM/UPI	Google Pay	15000	UPI	1.0	Satisfied
108	33	Male	Service	West Bengal	BHIM/UPI	BHIM/UPI	Google Pay	5000	UPI	1.0	Satisfied
109	34	Male	Service	West Bengal	Net Banking	Net Banking	PhonePe	15000	Cash	1.0	Satisfied
110	34	Male	Service	West Bengal	Card	Card	Google Pay	5000	UPI	1.0	Satisfied
111	35	Male	Service	West Bengal	BHIM/UPI	Card	PhonePe	20000	UPI	1.0	Satisfied
112	36	Male	Service	West Bengal	BHIM/UPI	Card	Google Pay	50000	Cash	0.0	Satisfied
113	39	Male	Service	West Bengal	BHIM/UPI	BHIM/UPI	PhonePe	5000	UPI	0.0	Satisfied
114	40	Male	Service	West Bengal	BHIM/UPI	Net Banking	Google Pay	10000	UPI	0.0	Satisfied
115	42	Male	Service	West Bengal	BHIM/UPI	Card	PhonePe	20000	Cash	1.0	Satisfied
116	43	Male	Service	West Bengal	Card	Card	PhonePe	30000	UPI	1.0	Satisfied

117	43	Female	Other	West Bengal	Cash	Net Banking	Google Pay	500	Cash	2.0	Satisfied
118	54	Male	Service	West Bengal	Card	Cash	PhonePe	20000	Cash	1.0	Average

In [60]:
```
DS_subset1_continuous = DS_subset1[DS_subset1.select_dtypes(include=['int64','float64']).columns]
DS_subset1_catagorical = DS_subset1[DS_subset1.select_dtypes(exclude=['int64','float64']).columns]
```

In [61]:
```
DS_subset2_continuous = DS_subset2[DS_subset2.select_dtypes(include=['int64','float64']).columns]
DS_subset2_catagorical = DS_subset2[DS_subset2.select_dtypes(exclude=['int64','float64']).columns]
```

Analysis on Continuous dataset

In [62]:
```
DS_continuous.describe()
```

Out[62]:

	Age	Average Amount Paid Monthly through the preferred mode	Number of unsuccessful transactions per 5 transactions
count	93.000000	93.000000	93.000000
mean	27.279570	8517.301075	0.776344
std	4.985557	7428.468480	0.570926
min	20.000000	0.000000	0.000000
25%	24.000000	2000.000000	0.000000
50%	26.000000	7000.000000	1.000000
75%	29.000000	10000.000000	1.000000
max	42.000000	30000.000000	2.000000

In [63]:
```
DS_continuous.dtypes
```

Out[63]:
```
Age                                                  int64
Average Amount Paid Monthly through the preferred mode     int64
Number of unsuccessful transactions per 5 transactions    float64
dtype: object
```

In [64]:
```
DS_continuous['Number of unsuccessful transactions per 5 transactions'] = DS_continuous['Number of unsuccessful transactions per 5 transactions'].values.astype(np.int64)
```

In [65]:
```
DS_continuous.dtypes
```

Out[65]:
```
Age                                                 int64
Average Amount Paid Monthly through the preferred mode    int64
Number of unsuccessful transactions per 5 transactions    int64
dtype: object
```

In [66]:
```
DS_continuous.describe()
```

Out[66]:

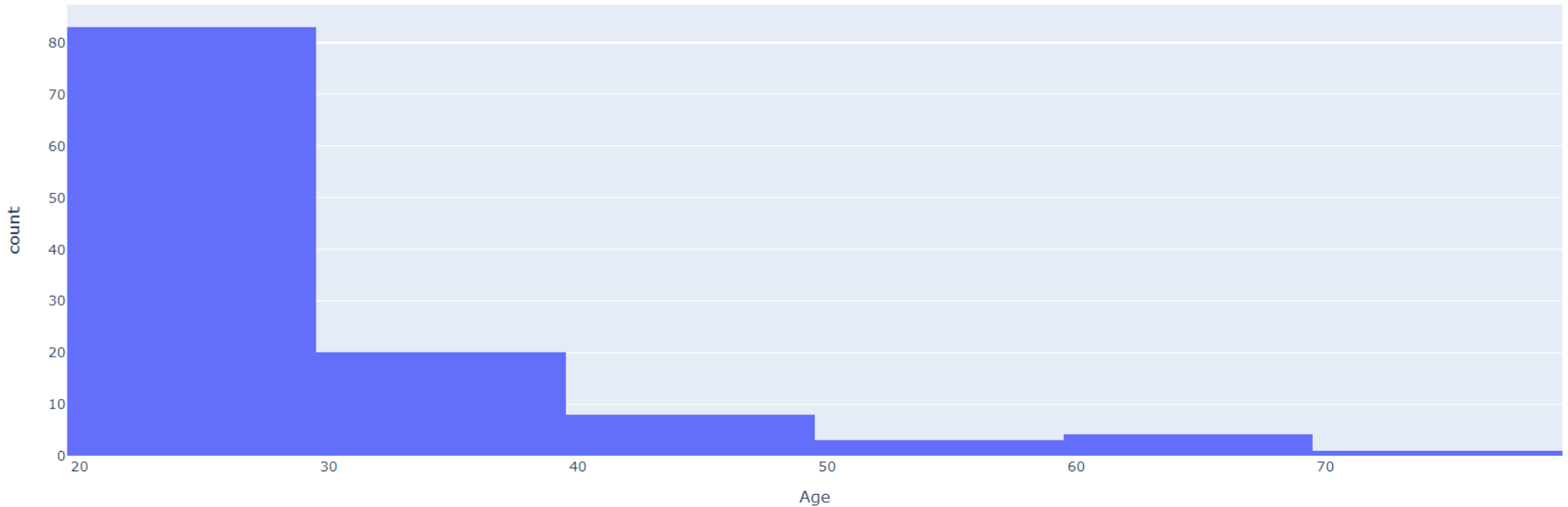
	Age	Average Amount Paid Monthly through the preferred mode	Number of unsuccessful transactions per 5 transactions
count	93.000000	93.000000	93.000000
mean	27.279570	8517.301075	0.774194
std	4.985557	7428.468480	0.573491
min	20.000000	0.000000	0.000000
25%	24.000000	2000.000000	0.000000
50%	26.000000	7000.000000	1.000000
75%	29.000000	10000.000000	1.000000
max	42.000000	30000.000000	2.000000

In [67]:
```
DS_continuous.count()
```

Out[67]:
```
Age                                                 93
Average Amount Paid Monthly through the preferred mode    93
Number of unsuccessful transactions per 5 transactions    93
dtype: int64
```

Univariate Analysis

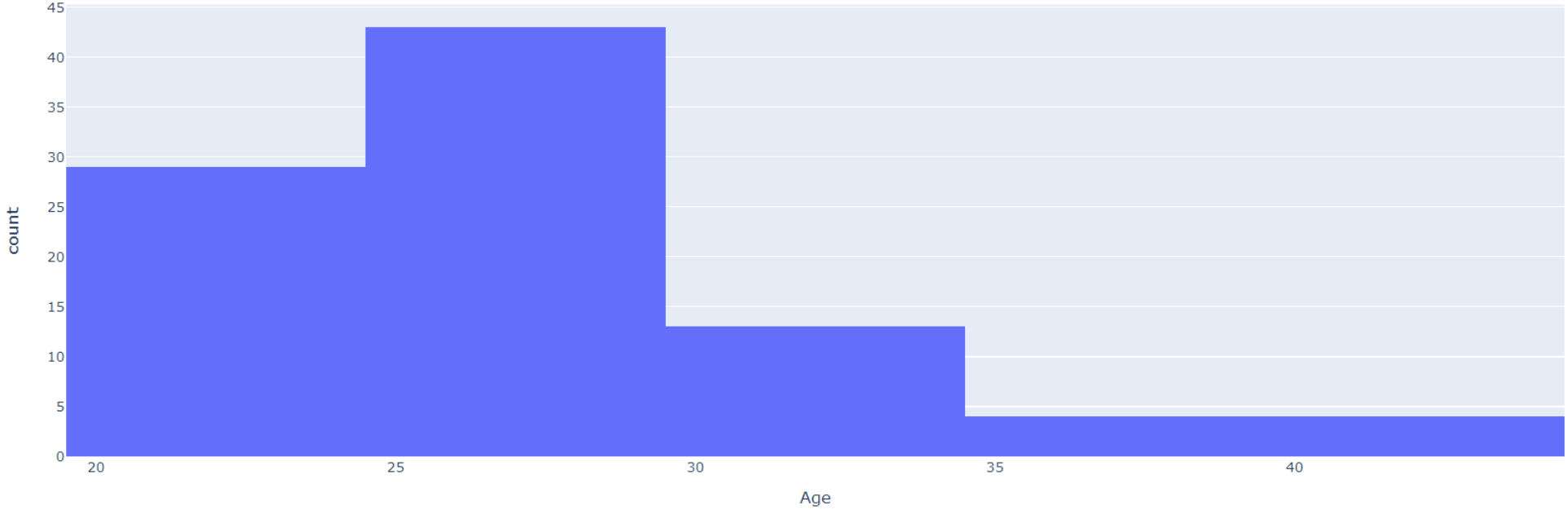
In [68]:
```
px.histogram(Clean_UPI_DS, x= 'Age',labels= {'Age':'Age'}, nbins = 10)
```



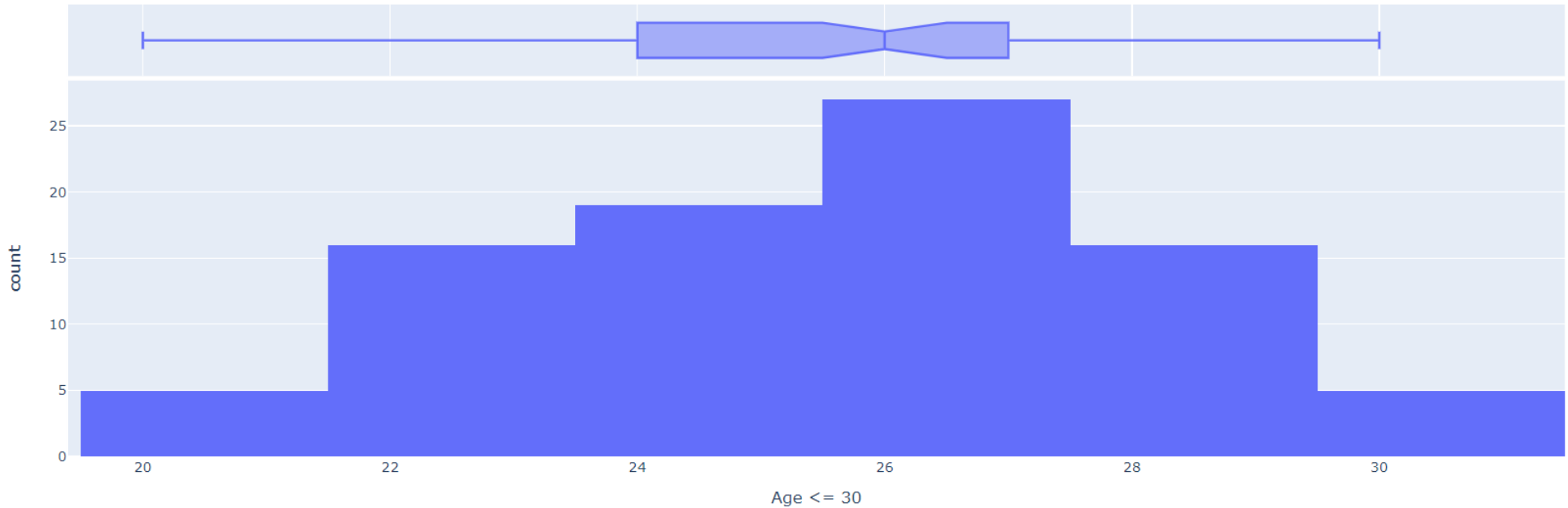
The plot is skewed right (positively skewed) that means younger people have responded more to our question

Plotting the age graph after removing outliers

In [69]:
```
px.histogram(DS_continuous, x= 'Age',labels= {'Age':'Age'}, nbins = 10)
```



In [70]:
```
px.histogram(DS_subset1, x= 'Age', labels= {'Age':'Age <= 30'}, marginal='box', nbins = 10)
```



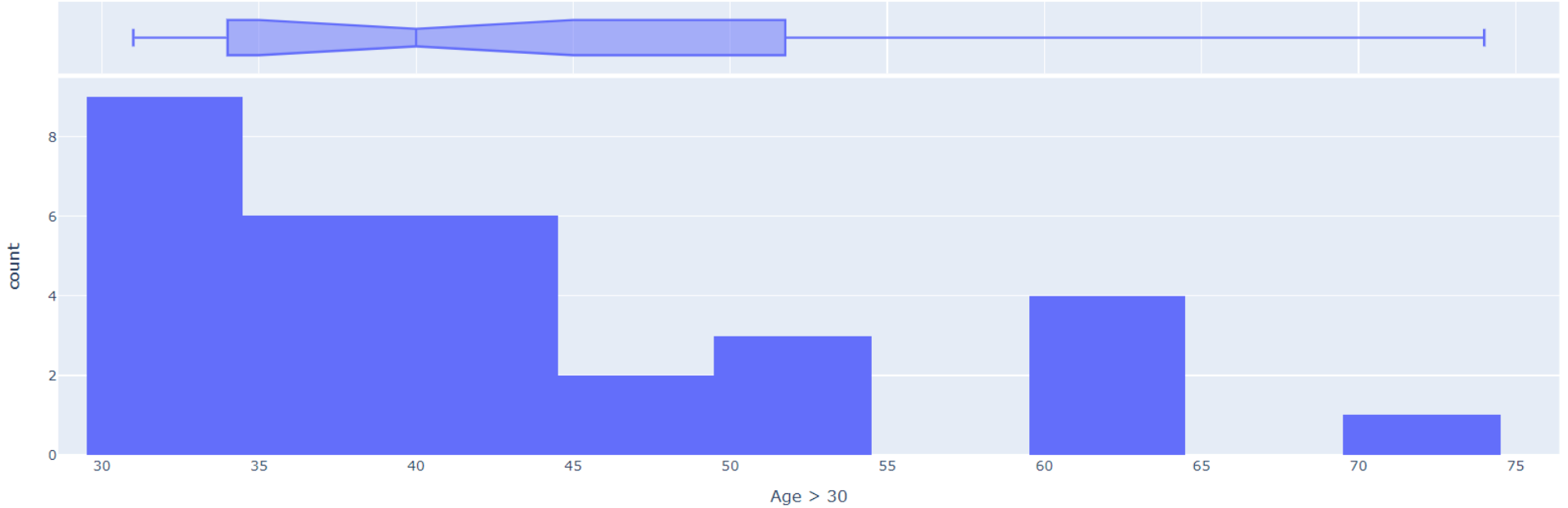
Q1 is 24, Q3 is 27, IQR is (Q3 -Q1) 3. The median is 26. Minimum is 20 and Max is 30. So it is clearly visible graph is slightly Negatively skewed

In [71]:
```
DS_subset1.describe()
```

Out[71]:

	Age	Average Amount Paid Monthly through the preferred mode	Number of unsuccessful transactions per 5 transactions
count	88.000000	88.000000	88.000000
mean	25.522727	24569.318182	1.034091
std	2.559532	106936.013049	1.076914
min	20.000000	0.000000	0.000000
25%	24.000000	2750.000000	0.000000
50%	26.000000	10000.000000	1.000000
75%	27.000000	15000.000000	1.000000
max	30.000000	1000000.000000	5.000000

In [72]:
```
px.histogram(DS_subset2, x= 'Age', labels= {'Age':'Age > 30'}, marginal='box', nbins = 10)
```



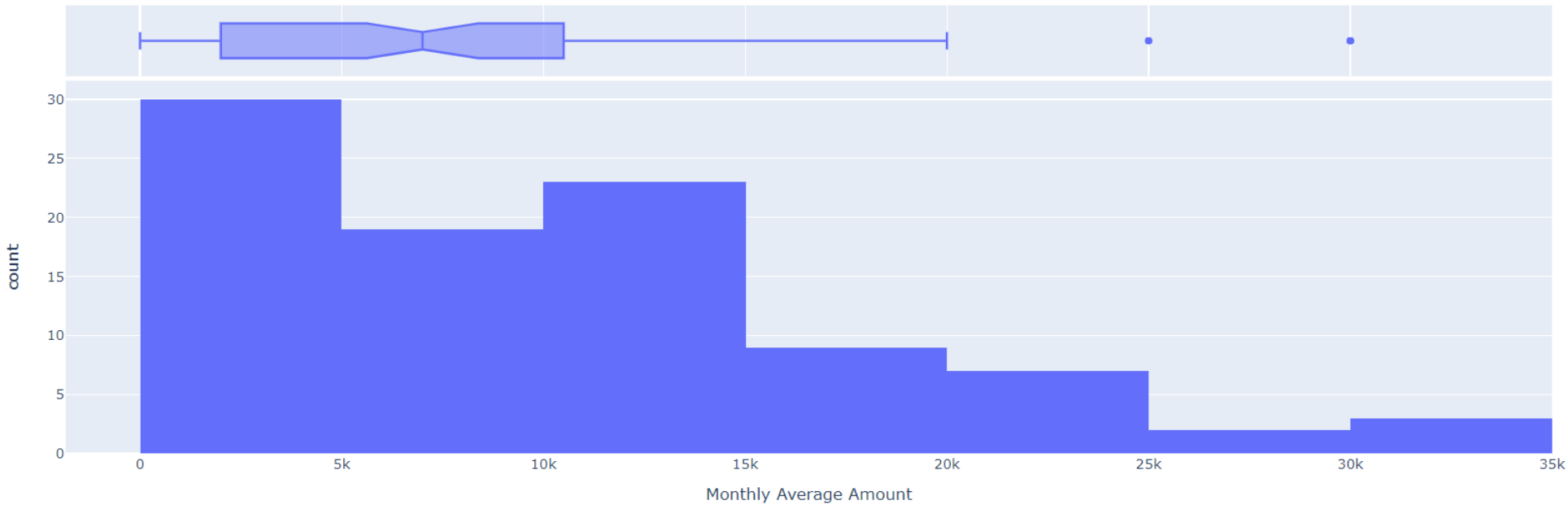
Q1 is 34, Q3 is 51.75, IQR is (Q3 -Q1) 17.75. The median is 40. Minimum is 31 and Max is 74. So it is clearly visible, graph is Positively skewed

In [73]:
```
DS_subset2.describe()
```

Out[73]:

	Age	Average Amount Paid Monthly through the preferred mode	Number of unsuccessful transactions per 5 transactions
count	31.000000	31.000000	31.000000
mean	43.290323	12955.129032	0.748387
std	11.515768	19529.790633	0.675214
min	31.000000	0.000000	0.000000
25%	34.000000	1004.500000	0.000000
50%	40.000000	7000.000000	1.000000
75%	49.500000	15000.000000	1.000000
max	74.000000	100000.000000	2.000000

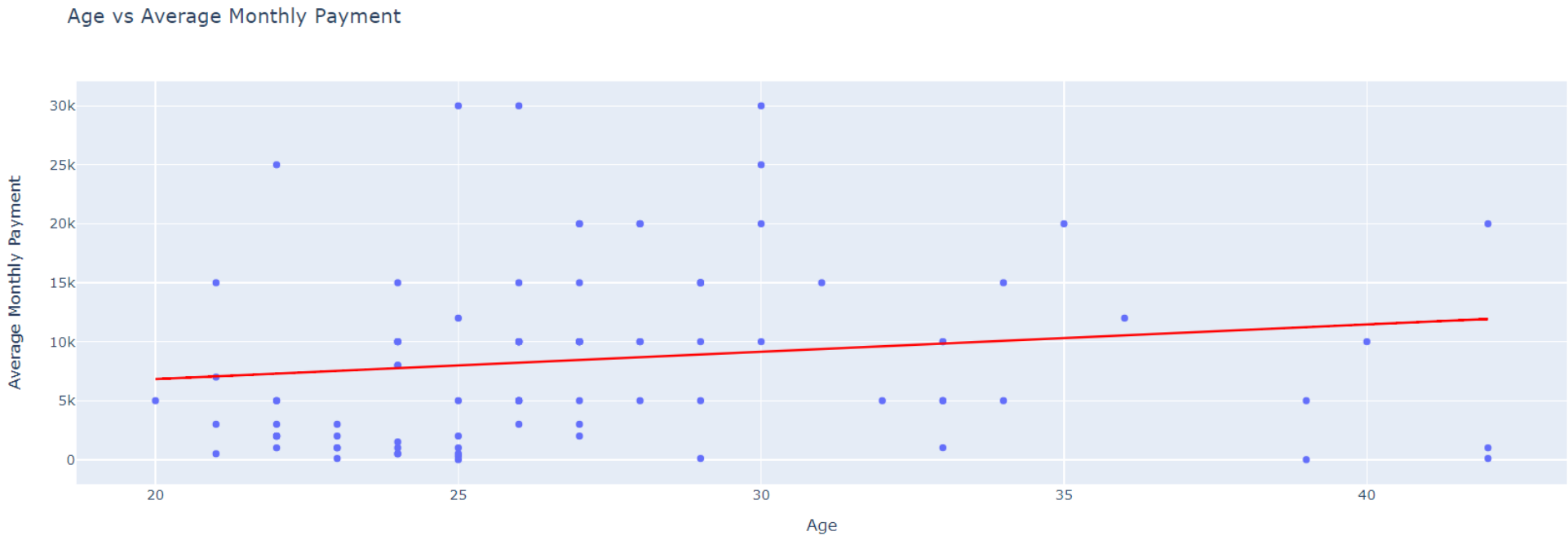
In [74]:
```
px.histogram(DS_continuous, x= 'Average Amount Paid Monthly through the preferred mode',labels= {'Average Amount Paid Monthly through the preferred mode':'Monthly Average Amount'}, marginal='box', nbins
```



Q1 is 2000, Q3 is 10.5K, IQR is (Q3 -Q1) 8.5K . The median is 7000. Minimum is 0 and Max is 30K. So it is clearly visible, graph is Positively skewed

Bivariate Analysis

In [75]:
```
px.scatter(DS_continuous,
           title='Age vs Average Monthly Payment',
           x='Age',
           y='Average Amount Paid Monthly through the preferred mode',
           labels = {'Age': 'Age', 'Average Amount Paid Monthly through the preferred mode': 'Average Monthly Payment'}, trendline = 'ols', trendline_color_override = 'red' )
```



In [76]:
```
DS_continuous.cov()
```

Out[76]:

	Age	Average Amount Paid Monthly through the preferred mode	Number of unsuccessful transactions per 5 transactions
Age	24.855774	5.758741e+03	-0.044881
Average Amount Paid Monthly through the preferred mode	5758.741000	5.518214e+07	99.699158
Number of unsuccessful transactions per 5 transactions	-0.044881	9.969916e+01	0.328892

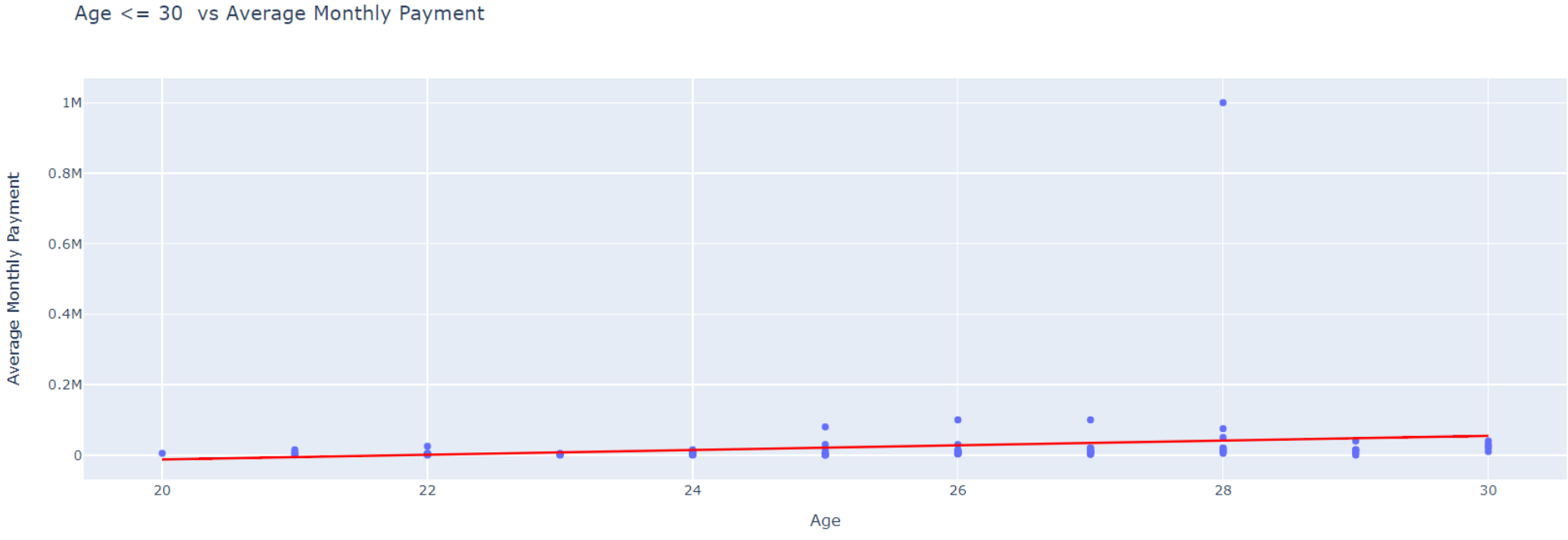
In [77]:
```
DS_continuous.corr()
```

Out[77]:

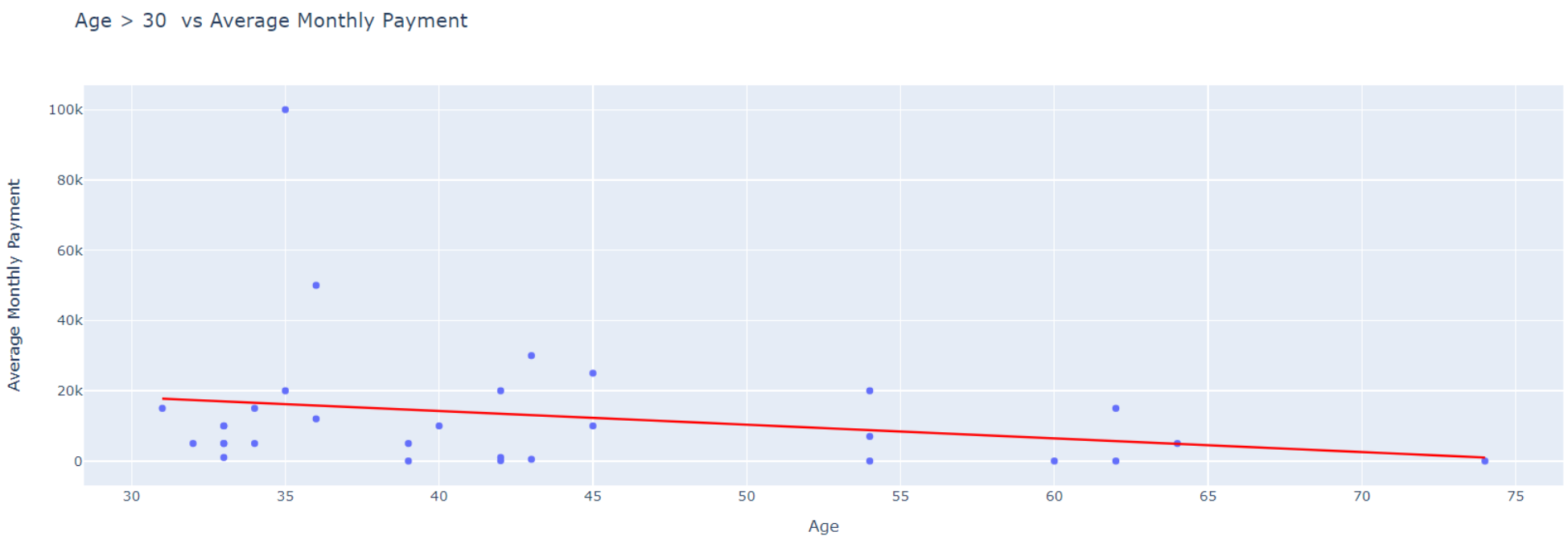
	Age	Average Amount Paid Monthly through the preferred mode	Number of unsuccessful transactions per 5 transactions
Age	1.000000	0.155494	-0.015697
Average Amount Paid Monthly through the preferred mode	0.155494	1.000000	0.023403
Number of unsuccessful transactions per 5 transactions	-0.015697	0.023403	1.000000

With age and Average monthly Payment we got the correlation value 0.155494. Therefore these two are weakly corelated. So we can say with increasing Age, Avg monthly payment is also getting affected.

In [78]:
```
px.scatter(DS_subset1_continuous,
           title='Age <= 30  vs Average Monthly Payment',
           x='Age',
           y='Average Amount Paid Monthly through the preferred mode',
           labels = {'Age': 'Age', 'Average Amount Paid Monthly through the preferred mode': 'Average Monthly Payment'}, trendline = 'ols', trendline_color_override = 'red' )
```



In [79]:
```
px.scatter(DS_subset2_continuous,
           title='Age > 30  vs Average Monthly Payment',
           x='Age',
           y='Average Amount Paid Monthly through the preferred mode',
           labels = {'Age': 'Age', 'Average Amount Paid Monthly through the preferred mode': 'Average Monthly Payment'}, trendline = 'ols', trendline_color_override = 'red' )
```



Observation:

After our survey and analysis, we found out that UPI has become an integral part of our life that too among middle age people, the use of cash has fallen significantly and people are using UPI even for a small number (amount) of transactions.

Post covid, the rise of UPI has been significantly huge and it is evident mainly because it allows contactless payment.
Among all other UPI apps, **G-Pay** is the most preferred one maybe because it gives better cashback and offers than others.

People of the age group 20-40 are active users of UPI payments while the slightly older age group (>=40) uses more cash.
From our dataset, we can see that out of **119** people 86 people are satisfied with UPI payment which means **72.76%** of people are satisfied.

Conclusion:

Thus, having done the statistical analysis on **UPI Payments**, we can say that UPI payments are the future of Digital India.