

Machine Learning Lab

Reg No: 23MCA1030

Name : Vinayak Kumar Singh

Exercise 3_ Exploratory Analysis for any dataset of your Choice

- Data preprocessing
- Visualization
- Inferences

Dataset url : <https://www.kaggle.com/datasets/mauryansshivam/paytm-revenue-users-transactions/data?select=Paytm.csv>

Collab url :

<https://colab.research.google.com/drive/1OzNRK2GuFw4K5EyekoJlWZ9ocGTyEhd?usp=sharing>

```
import numpy as np
import pandas as pd
import numpy as np
import seaborn as sns
```

```
df = pd.read_csv("/content/Paytm.csv")
```

```
df.head(5)
```

	Date	Gross Merchandise Value	Merchant Transactions \
0	01-06-2020	0.70	969
1	01-09-2020	0.95	1232
2	01-12-2020	1.12	1587
3	01-03-2021	1.27	2084
4	01-06-2021	1.47	2287

	Total Transactions	Average Monthly Transacting Users (MTU) \
0	1231	39.7
1	1578	43.0
2	2038	47.1
3	2558	50.4
4	2785	50.4

	Registered Merchants (end of period) \
0	17.0
1	18.5
2	20.0
3	21.1
4	21.8

	Payment Devices (cumulative, end of period) distributed \	Volume of loans
--	---	-----------------

0	0.0
23	
1	0.3
349	
2	0.6
881	
3	0.8
1381	
4	0.9
1433	

	Postpaid Loans	Personal Loans	...	Value of loans distributed
\				
0	20	0	...	44
1	332	0	...	206
2	851	5	...	468
3	1355	9	...	687
4	1413	13	...	632

	Postpaid Loans.1	Personal Loans.1	Merchant Loans.1	\
0	19	1	23	
1	74	2	129	
2	234	25	208	
3	416	68	203	
4	447	111	73	

	Payments Services to Consumers	Payments Services to Merchants	\
0	NaN		NaN
1	NaN		NaN
2	NaN		NaN
3	NaN		NaN
4	301.0		334.0

	Other Operating Revenue	Total Payments Revenue	Payment processing charges	\
0	NaN		NaN	
NaN				
1	NaN		NaN	
NaN				
2	NaN		NaN	
NaN				
3	NaN		NaN	
NaN				
4	-	635.0		
-527.0				

	Net Payments Margin
0	NaN
1	NaN
2	NaN
3	NaN
4	108.0

[5 rows x 21 columns]

df.tail(5)

	Date	Gross Merchandise Value	Merchant Transactions \
7	01-03-2022	2.59	4142
8	01-06-2022	2.96	5124
9	01-09-2022	3.18	5752
10	01-12-2022	3.46	6284
11	01-03-2023	3.62	6851

	Total Transactions	Average Monthly Transacting Users (MTU) \
7	5029	70.9
8	6126	74.8
9	6885	79.7
10	7634	84.9
11	8350	89.9

	Registered Merchants (end of period) \
7	26.7
8	28.3
9	29.5
10	31.4
11	33.5

	Payment Devices (cumulative, end of period) distributed \	Volume of loans
7	6544	2.9
8	8478	3.8
9	9192	4.8
10	10473	5.8
11	11910	6.8

	Postpaid Loans	Personal Loans	...	Value of loans distributed \
7	6414	92	...	3553

8	8286	132	...
5554			
9	8924	189	...
7313			
10	10105	244	...
9958			
11	11503	269	...
12554			

	Postpaid Loans.1	Personal Loans.1	Merchant Loans.1	\
7	2183	805	565	
8	3383	1344	827	
9	4050	2055	1208	
10	5202	2931	1825	
11	6794	3447	2313	

	Payments Services to Consumers	Payments Services to Merchants	\
7	469.0	572.0	
8	519.0	557.0	
9	549.0	624.0	
10	513.0	640.0	
11	524.0	918.0	

	Other Operating Revenue	Total Payments	Revenue	\
7	-		1041.0	
8	-		1075.0	
9	15		1188.0	
10	44		1197.0	
11	25		1467.0	

	Payment processing charges	Net Payments	Margin
7	-774.0		266.0
8	-694.0		382.0
9	-746.0		443.0
10	-738.0		459.0
11	-780.0		687.0

[5 rows x 21 columns]

Data Preprocessing

df.dtypes *#Checking Data Types*

Date	object
Gross Merchandise Value	float64
Merchant Transactions	int64
Total Transactions	int64
Average Monthly Transacting Users (MTU)	float64
Registered Merchants (end of period)	float64
Payment Devices (cumulative, end of period)	float64

```

Volume of loans distributed      int64
  Postpaid Loans                int64
  Personal Loans                int64
  Merchant Loans                int64
Value of loans distributed      int64
  Postpaid Loans.1             int64
  Personal Loans.1             int64
  Merchant Loans.1             int64
Payments Services to Consumers  float64
Payments Services to Merchants float64
Other Operating Revenue        object
Total Payments Revenue         float64
Payment processing charges      float64
Net Payments Margin            float64
dtype: object

```

```

df['Date '] = pd.to_datetime(df['Date '],format = '%d-%m-%Y')
df.replace('-',np.NaN,inplace=True) #Replacing dash by numpy NaN
df['Other Operating Revenue'] = df['Other Operating
Revenue'].astype('float') #Convert Default object to Float:
df

```

	Date	Gross Merchandise Value	Merchant Transactions	\
0	2020-06-01	0.70	969	
1	2020-09-01	0.95	1232	
2	2020-12-01	1.12	1587	
3	2021-03-01	1.27	2084	
4	2021-06-01	1.47	2287	
5	2021-09-01	1.96	2692	
6	2021-12-01	2.50	3477	
7	2022-03-01	2.59	4142	
8	2022-06-01	2.96	5124	
9	2022-09-01	3.18	5752	
10	2022-12-01	3.46	6284	
11	2023-03-01	3.62	6851	

	Total Transactions	Average Monthly Transacting Users (MTU)	\
0	1231	39.7	
1	1578	43.0	
2	2038	47.1	
3	2558	50.4	
4	2785	50.4	
5	3316	57.4	
6	4266	64.4	
7	5029	70.9	
8	6126	74.8	
9	6885	79.7	

10	7634	84.9
11	8350	89.9

Registered Merchants (end of period) \	
0	17.0
1	18.5
2	20.0
3	21.1
4	21.8
5	23.0
6	24.9
7	26.7
8	28.3
9	29.5
10	31.4
11	33.5

Payment Devices (cumulative, end of period) distributed \		Volume of loans
0		0.0
23		
1		0.3
349		
2		0.6
881		
3		0.8
1381		
4		0.9
1433		
5		1.3
2841		
6		2.0
4414		
7		2.9
6544		
8		3.8
8478		
9		4.8
9192		
10		5.8
10473		
11		6.8
11910		

Postpaid Loans distributed \		Personal Loans	...	Value of loans
0	20	0	...	
44				
1	332	0	...	
206				

2	851	5	...
468			
3	1355	9	...
687			
4	1413	13	...
632			
5	2790	29	...
1257			
6	4320	60	...
2181			
7	6414	92	...
3553			
8	8286	132	...
5554			
9	8924	189	...
7313			
10	10105	244	...
9958			
11	11503	269	...
12554			

	Postpaid Loans.1	Personal Loans.1	Merchant Loans.1 \
0	19	1	23
1	74	2	129
2	234	25	208
3	416	68	203
4	447	111	73
5	738	246	273
6	1190	516	474
7	2183	805	565
8	3383	1344	827
9	4050	2055	1208
10	5202	2931	1825
11	6794	3447	2313

	Payments Services to Consumers	Payments Services to Merchants \
0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	301.0	334.0
5	354.0	400.0
6	406.0	586.0
7	469.0	572.0
8	519.0	557.0
9	549.0	624.0
10	513.0	640.0
11	524.0	918.0

Other Operating Revenue Total Payments Revenue \

0	NaN	NaN
1	NaN	NaN
2	NaN	NaN
3	NaN	NaN
4	NaN	635.0
5	NaN	754.0
6	NaN	992.0
7	NaN	1041.0
8	NaN	1075.0
9	15.0	1188.0
10	44.0	1197.0
11	25.0	1467.0

	Payment processing charges	Net Payments	Margin
0	NaN		NaN
1	NaN		NaN
2	NaN		NaN
3	NaN		NaN
4	-527.0		108.0
5	-670.0		84.0
6	-783.0		209.0
7	-774.0		266.0
8	-694.0		382.0
9	-746.0		443.0
10	-738.0		459.0
11	-780.0		687.0

[12 rows x 21 columns]

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

Date	0
Gross Merchandise Value	0
Merchant Transactions	0
Total Transactions	0
Average Monthly Transacting Users (MTU)	0
Registered Merchants (end of period)	0
Payment Devices (cumulative, end of period)	0
Volume of loans distributed	0
Postpaid Loans	0
Personal Loans	0
Merchant Loans	0
Value of loans distributed	0
Postpaid Loans.1	0
Personal Loans.1	0
Merchant Loans.1	0
Payments Services to Consumers	4
Payments Services to Merchants	4
Other Operating Revenue	9
Total Payments Revenue	4


```

Payment processing charges          4
Net Payments Margin                 4
dtype: int64

df.columns

Index(['Date ', 'Gross Merchandise Value', 'Merchant Transactions',
      'Total Transactions', 'Average Monthly Transacting Users
      (MTU)', 'Registered Merchants (end of period)',
      'Payment Devices (cumulative, end of period)',
      'Volume of loans distributed', '  Postpaid Loans', '  Personal
      Loans', '  Merchant Loans', 'Value of loans distributed', '  Postpaid
      Loans.1', '  Personal Loans.1', '  Merchant Loans.1',
      'Payments Services to Consumers', 'Payments Services to
      Merchants', 'Other Operating Revenue', 'Total Payments Revenue',
      'Payment processing charges', 'Net Payments Margin'],
      dtype='object')

```

Data Analysis

```

df['person to person transactions'] = df['Total Transactions'] -
df['Merchant Transactions']
df['person to person transactions']

0      262
1      346
2      451
3      474
4      498
5      624
6      789
7      887
8     1002
9     1133
10     1350
11     1499
Name: person to person transactions, dtype: int64

```

Data Visualization

```

import matplotlib.pyplot as plt

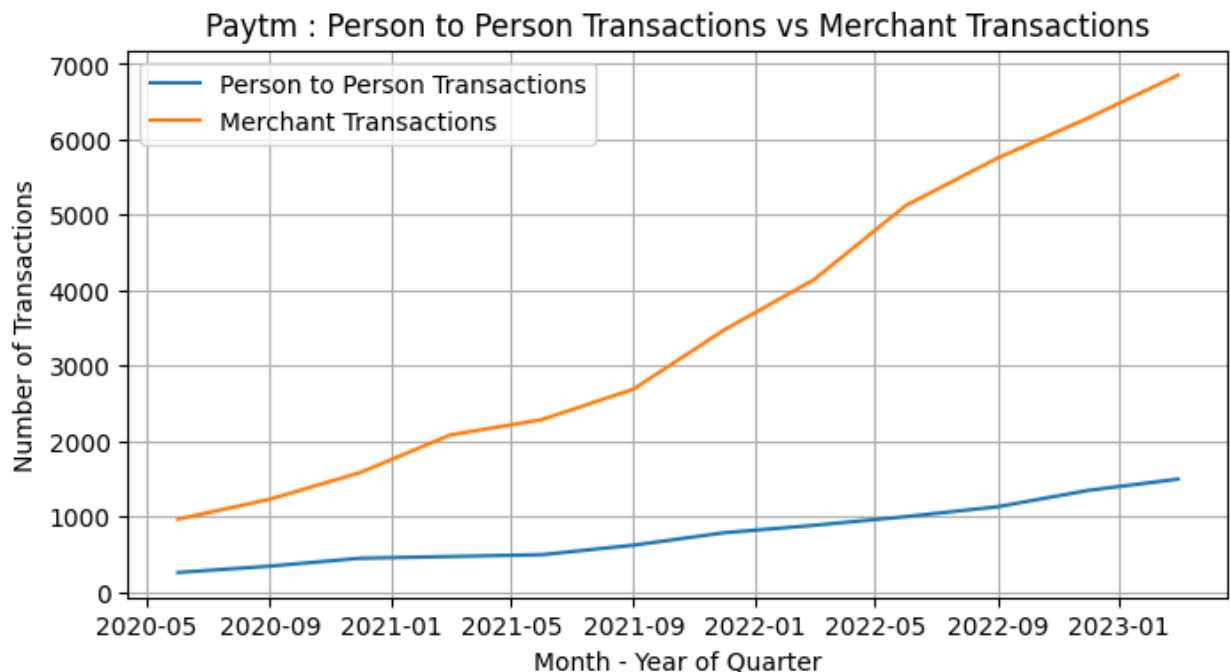
plt.figure(figsize=(8,4))
plt.plot(df['Date '], df['person to person transactions'],
label='Person to Person Transactions')

```

```

plt.plot(df['Date '], df['Merchant Transactions'], label='Merchant Transactions')
plt.title('Paytm : Person to Person Transactions vs Merchant Transactions')
plt.xlabel('Month - Year of Quarter')
plt.ylabel('Number of Transactions')
plt.grid(True)
plt.legend()
plt.show()

```



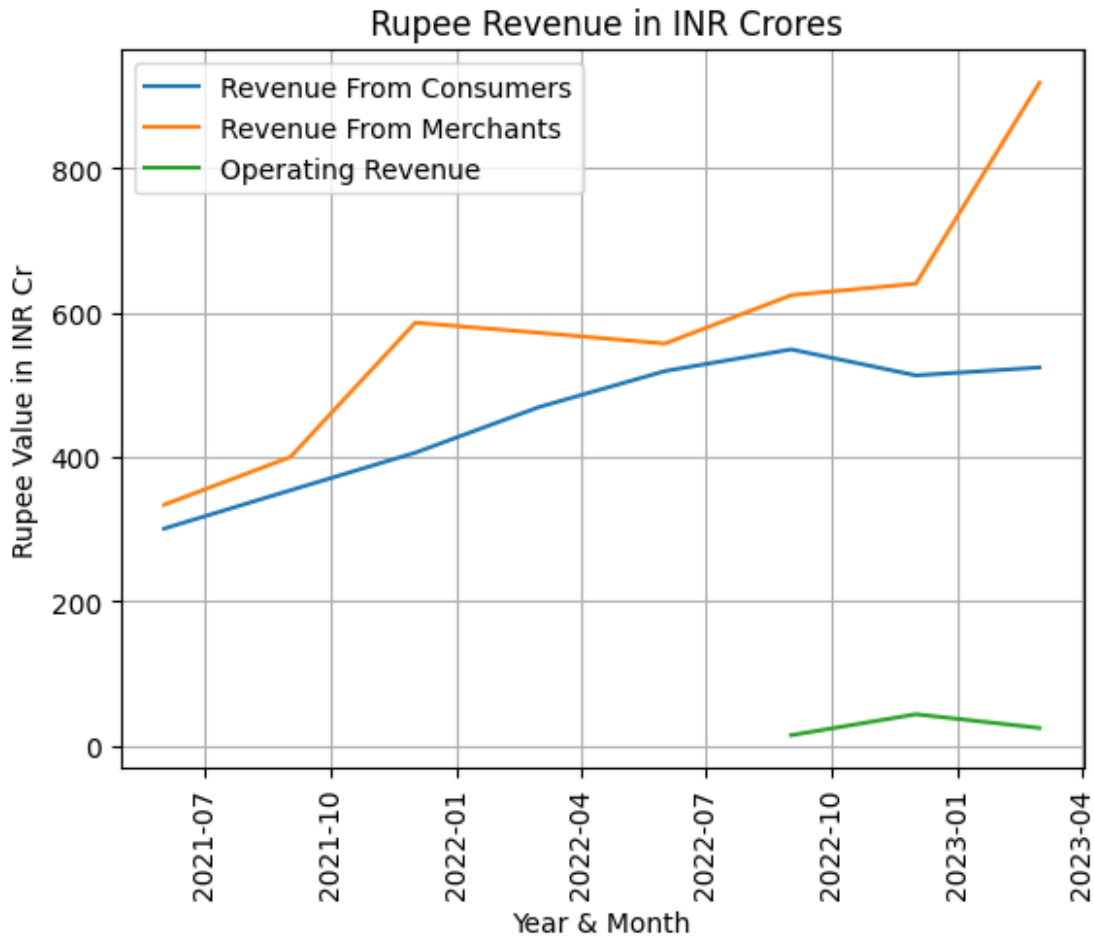
Revenue in INR Crores

```

plt.plot(df['Date '],df['Payments Services to Consumers'],label='Revenue From Consumers')
plt.plot(df['Date '],df['Payments Services to Merchants'],label='Revenue From Merchants')
plt.plot(df['Date '],df['Other Operating Revenue'],label='Operating Revenue')
plt.title(' Rupee Revenue in INR Crores')
plt.xlabel('Year & Month')
plt.ylabel('Rupee Value in INR Cr')
plt.xticks(rotation = 90)
plt.grid(True)
plt.legend()

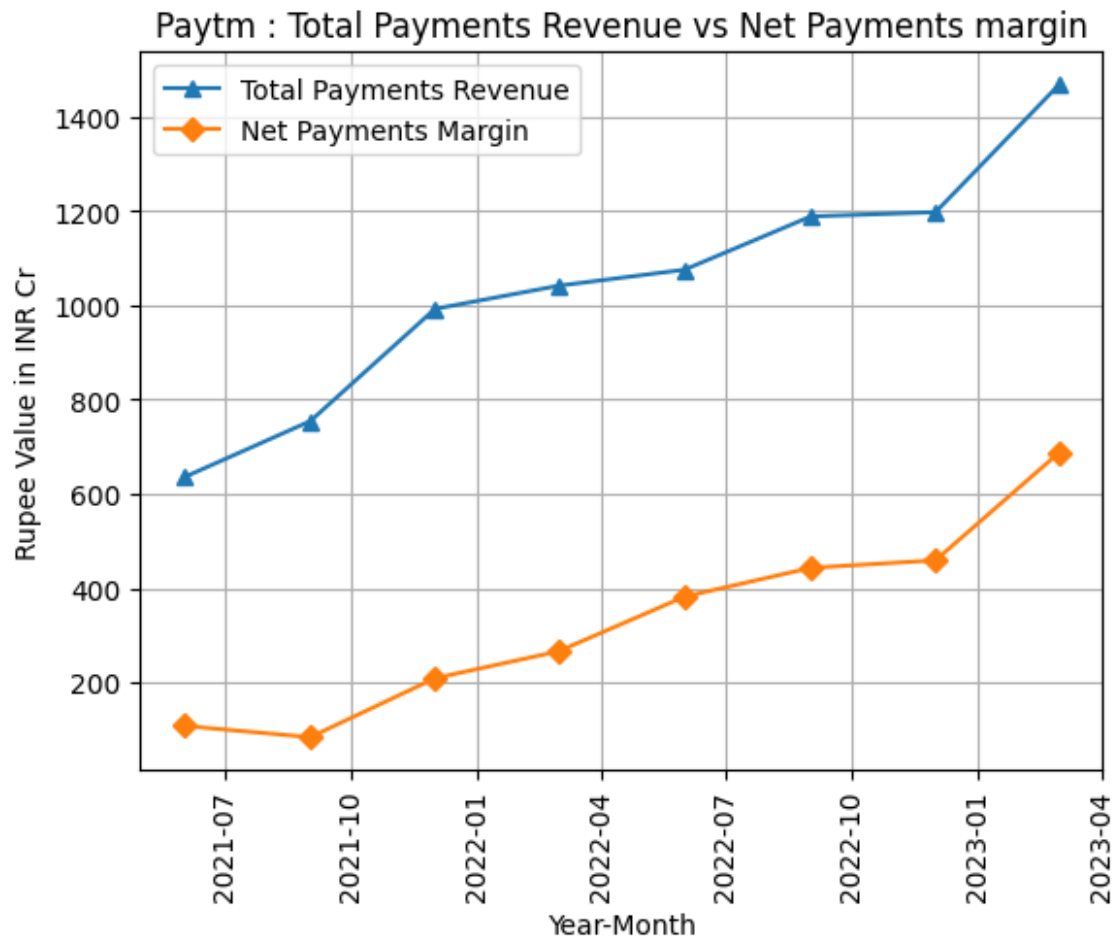
plt.show()

```

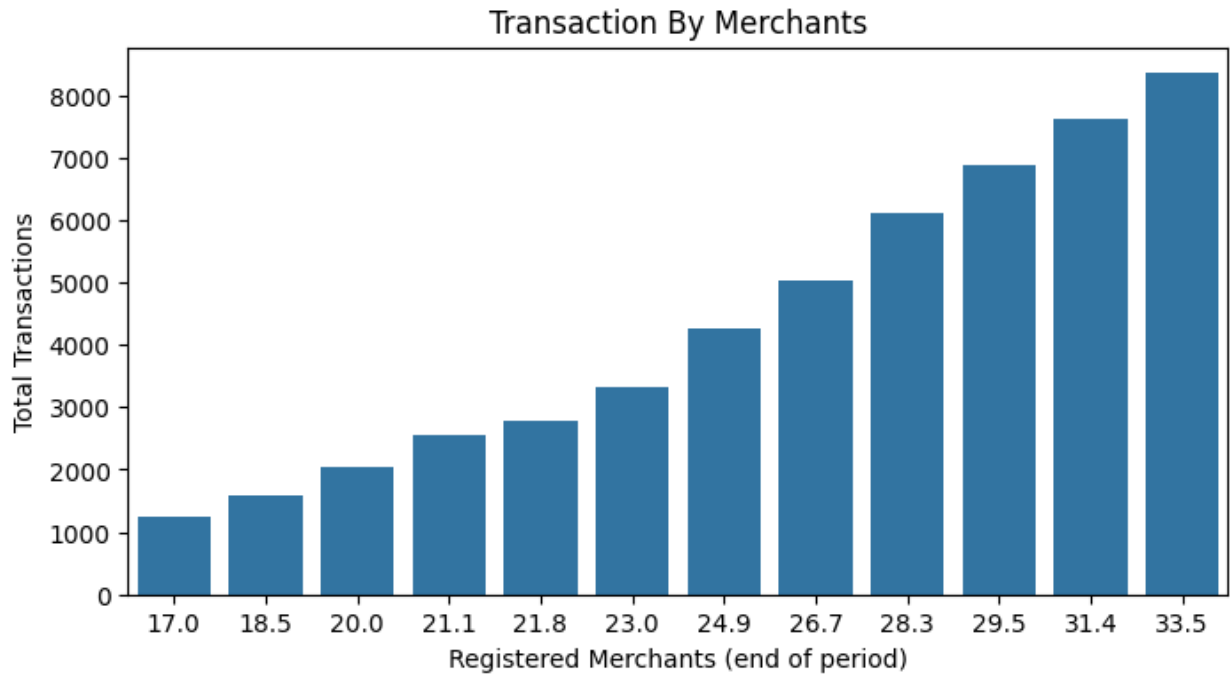


Inference : Overall revenue is increasing, with consumer revenue growing faster than merchant revenue.

```
plt.plot(df['Date'], df['Total Payments Revenue'], label='Total Payments Revenue', marker='^')
plt.plot(df['Date'], df['Net Payments Margin'], label='Net Payments Margin', marker='D')
plt.title('Paytm : Total Payments Revenue vs Net Payments margin')
plt.xlabel('Year-Month')
plt.ylabel('Rupee Value in INR Cr')
plt.grid(True)
plt.xticks(rotation = 90)
plt.legend()
plt.show()
```

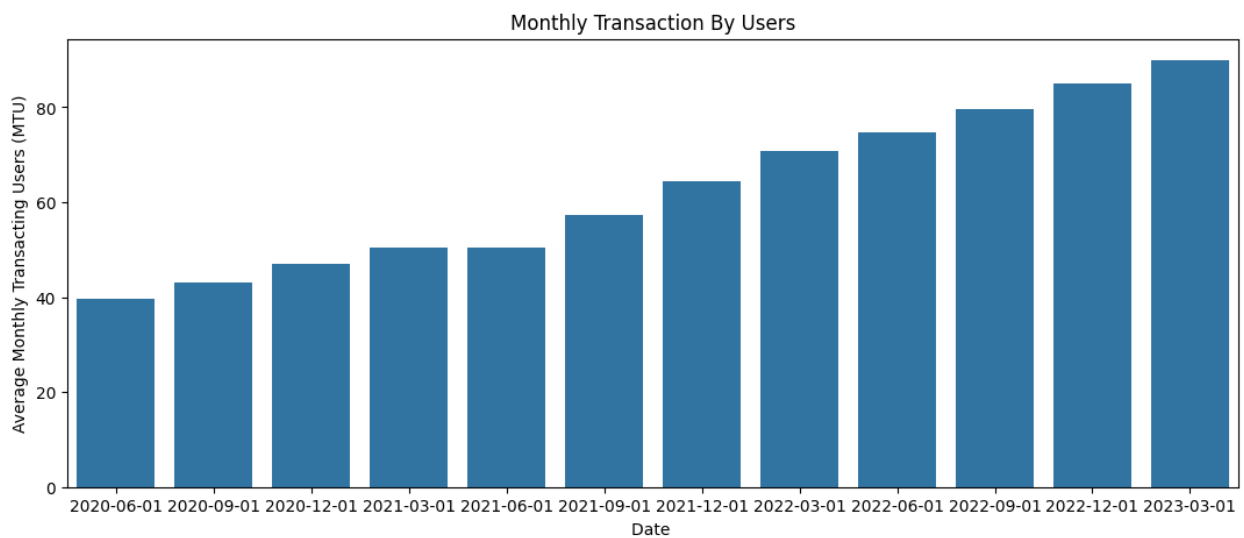


```
plt.figure(figsize=(8,4))
sns.barplot(data=df,x=df['Registered Merchants (end of
period)'],y=df['Total Transactions'])
plt.title("Transaction By Merchants")
plt.show()
```



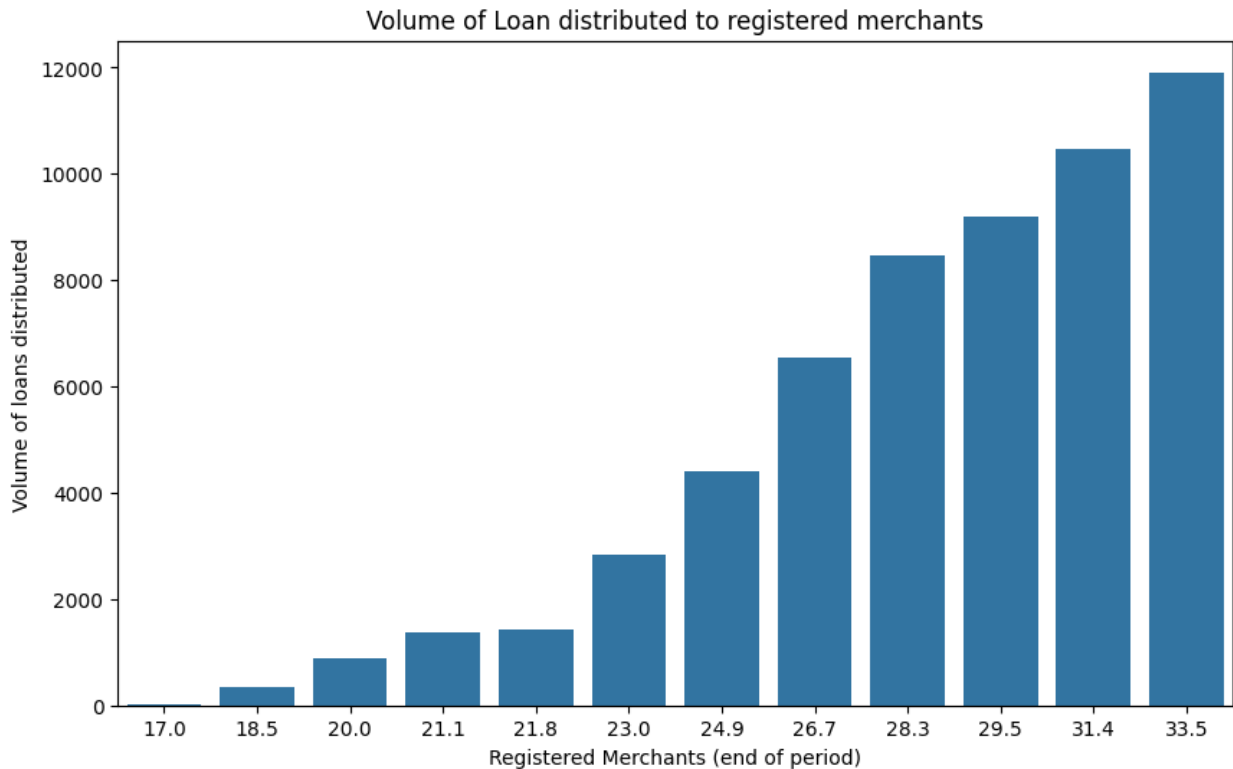
Inference : The number of Transaction by merchants on the platform appears to be increasing steadily over time. This suggests that the platform is growing and attracting new businesses.

```
plt.figure(figsize=(13,5))
sns.barplot(data=df,x=df['Date'],y=df['Average Monthly Transacting Users (MTU)'])
plt.title("Monthly Transaction By Users")
plt.show()
```



Inference : The average monthly transaction by users has increased steadily over time. This suggests that users are making more transactions on average each month.

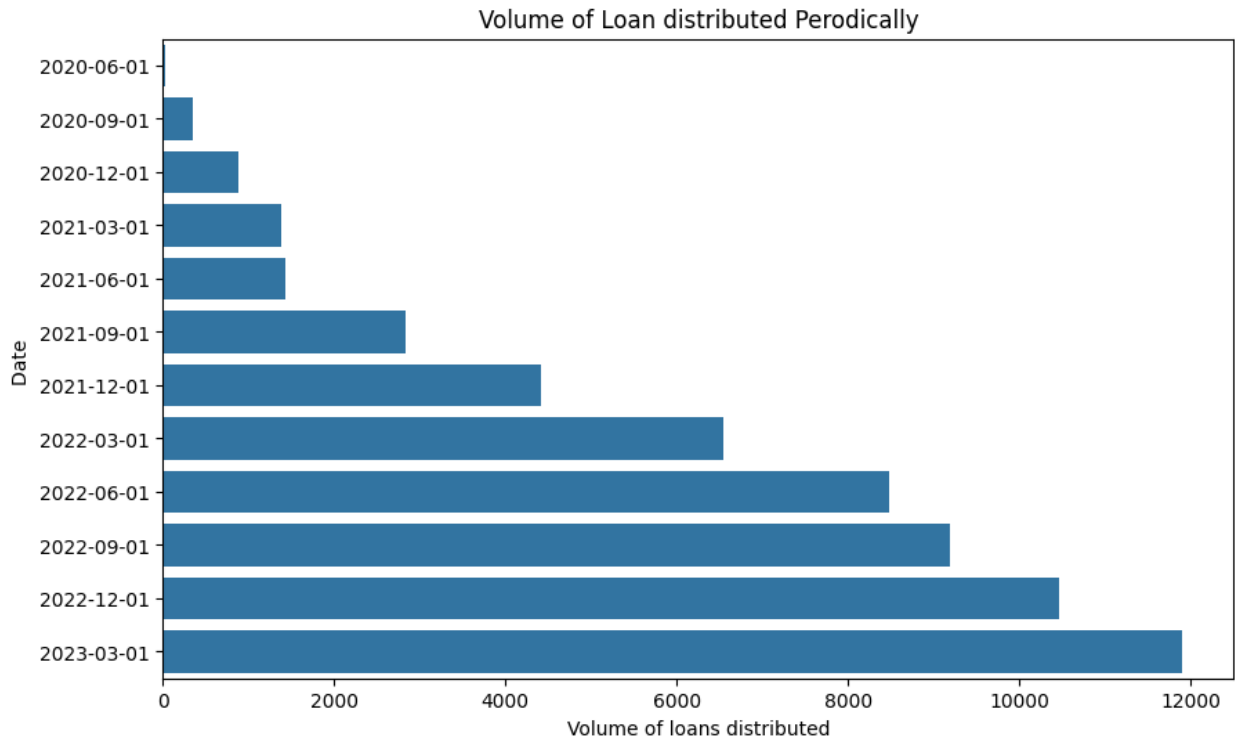
```
plt.figure(figsize=(10,6))
sns.barplot(data=df,x=df['Registered Merchants (end of
period)'],y=df['Volume of loans distributed'])
plt.title("Volume of Loan distributed to registered merchants")
plt.show()
```



Inference : The amount of loan distributed to registered merchants has been increasing steadily over time. This suggests that there is a growing demand for loans from merchants on the platform.

Volume of Loan distributed Periodically

```
plt.figure(figsize=(10,6))
sns.barplot(data=df,x=df['Volume of loans distributed'],y=df['Date '])
plt.title("Volume of Loan distributed Periodically")
plt.show()
```



Inference : The loan volume distributed fluctuates throughout the year, but there appears to be an upward trend overall. The graph shows several peaks and valleys, but the overall trend is positive. This suggests that the total amount of loans being distributed is increasing over time.