Notebook

November 9, 2024

Work 1-Analysis using python

[1]:
```
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

[19]:
```
df=pd.read_csv("C:\\Users\\i\\Downloads\\Data set 1.csv")
df
```

[19]:

	Date	App	Usage (minutes)	Notifications	Times Opened
0	07-08-2024	Instagram	81	24	57
1	08-08-2024	Instagram	90	30	53
2	26-08-2024	Instagram	112	33	17
3	22-08-2024	Instagram	82	11	38
4	12-08-2024	Instagram	59	47	16
..	…	…	…	…	…
195	10-08-2024	LinkedIn	22	12	5
196	23-08-2024	LinkedIn	5	7	1
197	18-08-2024	LinkedIn	19	2	5
198	26-08-2024	LinkedIn	21	14	1
199	02-08-2024	LinkedIn	13	4	1

[200 rows x 5 columns]

Understanding Attributes in dataset

[20]:
```
column_names=df.columns
column_names
```

[20]:
```
Index(['Date', 'App', 'Usage (minutes)', 'Notifications', 'Times Opened'],
dtype='object')
```

Exploring the attribute APP

[21]:
```
app_col=df["App"].value_counts()
app_col
```

[21]:
```
App
Instagram      25
X              25
WhatsApp       25
8 Ball Pool    25
Safari         25
Netflix        25
Facebook       25
LinkedIn       25
Name: count, dtype: int64
```

Understanding: The dataset contains total of 200 entries of 7 different apps, their usage time in minutes,number of times the app opened and the number of notifications of each app.

Notification count of each App

[22]:
```
sumN_Instagram=df[df["App"]=='Instagram']['Notifications'].sum()
sumN_X=df[df["App"]=="X"]["Notifications"].sum()
sumN_WhatsApp=df[df["App"]=="WhatsApp"]["Notifications"].sum()
sumN_8BallPool=df[df["App"]=="8 Ball Pool"]["Notifications"].sum()
sumN_Safari=df[df["App"]=="Safari"]["Notifications"].sum()
sumN_Netflix=df[df["App"]=="Netflix"]["Notifications"].sum()
sumN_Facebook=df[df["App"]=="Facebook"]["Notifications"].sum()
sumN_LinkedIn=df[df["App"]=="LinkedIn"]["Notifications"].sum()

print("Instagram=",sumN_Instagram,"\nX=",sumN_X,"\nWhatsapp=",sumN_WhatsApp,"\n8␣
 ↪Ball Pool",sumN_8BallPool,
    ␣
 ↪"\nSafari=",sumN_Safari,"\nNetflix=",sumN_Netflix,"\nFacebook=",sumN_Facebook,"\nLinkedIn="
```

```
Instagram= 1245
X= 646
Whatsapp= 2498
8 Ball Pool 113
Safari= 18
Netflix= 11
Facebook= 993
LinkedIn= 223
```

[36]:
```
#sns.barplot(data=df,x='App',y='Notifications')
```

Number of Times each app opened

[24]:
```
sumT_Instagram=df[df['App']=='Instagram']['Times Opened'].sum()
sumT_X=df[df['App']=='X']['Times Opened'].sum()
sumT_WhatsApp=df[df['App']=='WhatsApp']['Times Opened'].sum()
sumT_8BallPool=df[df['App']=='8 Ball Pool']['Times Opened'].sum()
sumT_Safari=df[df['App']=='Safari']['Times Opened'].sum()
sumT_Netflix=df[df['App']=='Netflix']['Times Opened'].sum()
```

```
sumT_Facebook=df[df['App']=='Facebook']['Times Opened'].sum()
sumT_LinkedIn=df[df['App']=='LinkedIn']['Times Opened'].sum()

print("Instagram=",sumT_Instagram,"\nX=",sumT_X,"\nWhatsapp=",sumT_WhatsApp,"\n8␣
 ↪Ball Pool",sumT_8BallPool,
    ␣
 ↪"\nSafari=",sumT_Safari,"\nNetflix=",sumT_Netflix,"\nFacebook=",sumT_Facebook,"\nLinkedIn="

#total_instagram_usage = df[df['App'] == 'Instagram']['Usage (minutes)'].sum()
#total_instagram_usage
```

```
Instagram= 1039
X= 329
Whatsapp= 1706
8 Ball Pool 182
Safari= 132
Netflix= 64
Facebook= 755
LinkedIn= 119
```

[34]:
```
#sns.barplot(data=df,x='App',y='Times Opened')
```

Usage of Apps

[26]:
```
sumU_Instagram=df[df['App']=='Instagram']['Usage (minutes)'].sum()
sumU_X=df[df['App']=='X']['Usage (minutes)'].sum()
sumU_WhatsApp=df[df['App']=='WhatsApp']['Usage (minutes)'].sum()
sumU_8BallPool=df[df['App']=='8 Ball Pool']['Usage (minutes)'].sum()
sumU_Safari=df[df['App']=='Safari']['Usage (minutes)'].sum()
sumU_Netflix=df[df['App']=='Netflix']['Usage (minutes)'].sum()
sumU_Facebook=df[df['App']=='Facebook']['Usage (minutes)'].sum()
sumU_LinkedIn=df[df['App']=='LinkedIn']['Usage (minutes)'].sum()

print("Instagram=",sumU_Instagram,"\nX=",sumU_X,"\nWhatsapp=",sumU_WhatsApp,"\n8␣
 ↪Ball Pool",sumU_8BallPool,
    ␣
 ↪"\nSafari=",sumU_Safari,"\nNetflix=",sumU_Netflix,"\nFacebook=",sumU_Facebook,"\nLinkedIn="
```

```
Instagram= 1898
X= 675
Whatsapp= 1204
8 Ball Pool 452
Safari= 270
Netflix= 1819
Facebook= 842
LinkedIn= 390
```

[35]:
```
#sns.barplot(data=df,x='App',y='Usage (minutes)')
```

Creating a new dataframe to combine and store every values

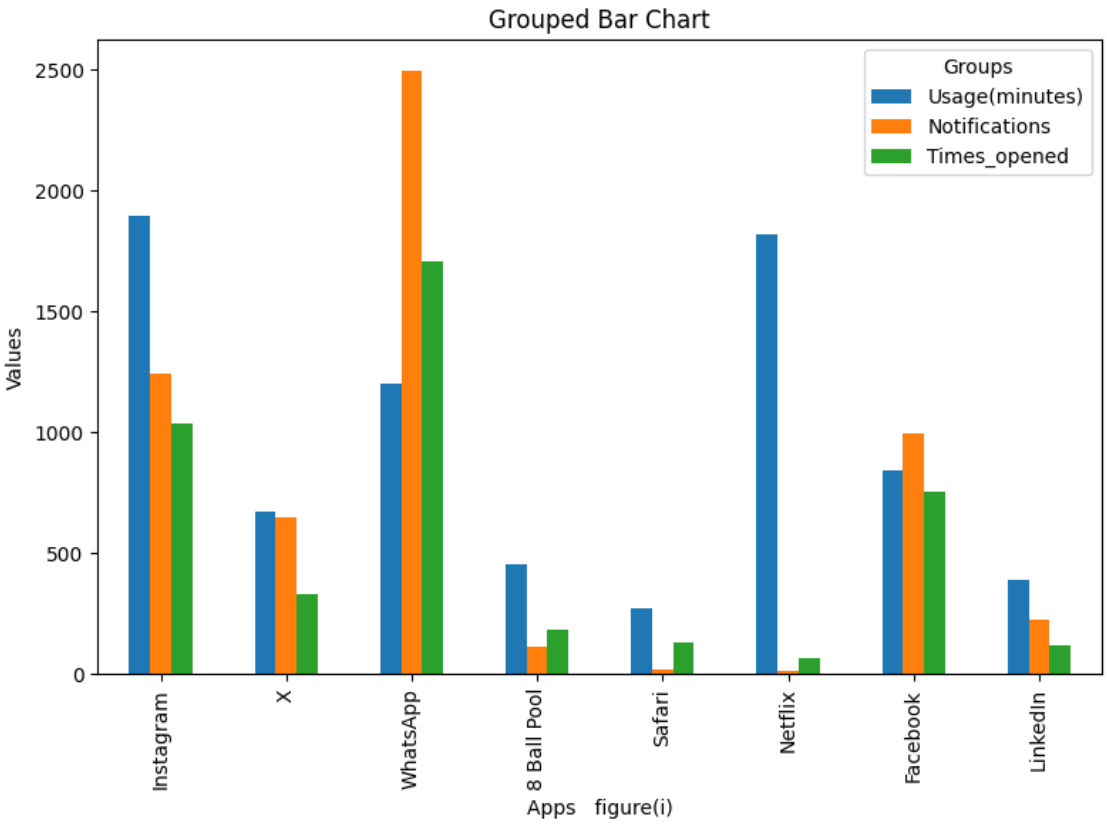
[32]:
```
new_df=pd.DataFrame({
    'App':['Instagram', 'X', 'WhatsApp', '8 Ball Pool', 'Safari', 'Netflix',
       'Facebook', 'LinkedIn'],
    'Usage(minutes)':
 ↪[sumU_Instagram,sumU_X,sumU_WhatsApp,sumU_8BallPool,sumU_Safari,sumU_Netflix,sumU_Facebook,
    'Notifications':
 ↪[sumN_Instagram,sumN_X,sumN_WhatsApp,sumN_8BallPool,sumN_Safari,sumN_Netflix,sumN_Facebook,
    'Times_opened':
 ↪[sumT_Instagram,sumT_X,sumT_WhatsApp,sumT_8BallPool,sumT_Safari,sumT_Netflix,sumT_Facebook,
}
)
new_df
```

[32]:

	App	Usage(minutes)	Notifications	Times_opened
0	Instagram	1898	1245	1039
1	X	675	646	329
2	WhatsApp	1204	2498	1706
3	8 Ball Pool	452	113	182
4	Safari	270	18	132
5	Netflix	1819	11	64
6	Facebook	842	993	755
7	LinkedIn	390	223	119

Grouped Bar Chart to compare between apps

[33]:
```
new_df.set_index('App', inplace=True)
new_df.plot(kind='bar', figsize=(8, 6))
plt.title('Grouped Bar Chart')
plt.xlabel('Apps   figure(i)')
plt.ylabel('Values')
plt.legend(title='Groups')
plt.tight_layout()
plt.show()
```



Correlation matrix to identify whether there are any relation between the attributes

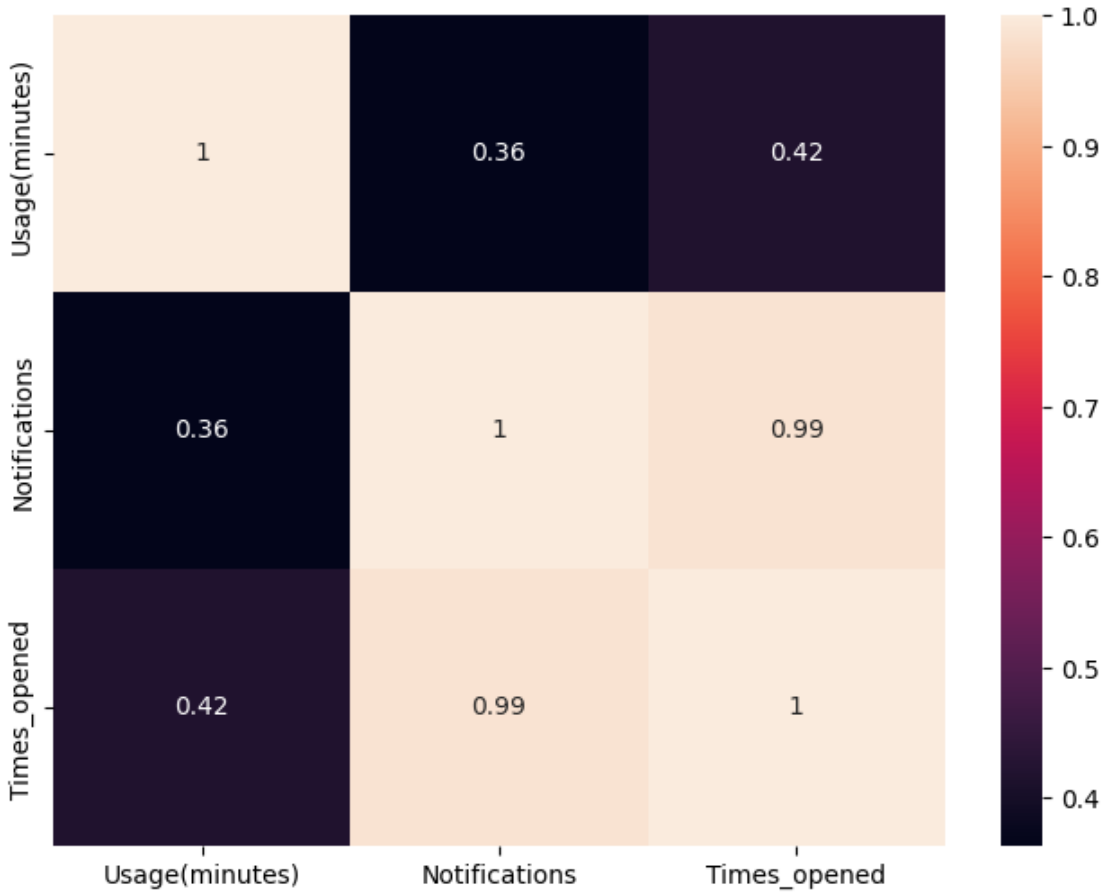
[30]:
```
CorrelationMatrix=new_df[['Usage(minutes)','Notifications','Times_opened']].
↪corr()
CorrelationMatrix
```

[30]:

	Usage(minutes)	Notifications	Times_opened
Usage(minutes)	1.000000	0.362809	0.415376
Notifications	0.362809	1.000000	0.986718
Times_opened	0.415376	0.986718	1.000000

Displaying the correlation matrix in the form of heatmap

[31]:
```
plt.figure(figsize=(8,6))
sns.heatmap(CorrelationMatrix,annot=True)
plt.show()
```



Inferences Based on the dataset provided and by Analysing figure(i), we can infer the bellow findings:

1.The most used app is Instagram with a total usage of 1898 minutes

2.WhatsApp generates most Notifications

3.8 Ball Pool,Safari and Netflix are being opened more without a strong influence of Notifications

4.Netflix has a total usage time of 1819 minutes without a strong influence of Notifications.

5.Instagram usage is very much more than compared with facebook, the reason can be because the user's might be senior citizens.

6.The number of Notifications and Times Opened have a strong correlation, meaning the notifications are the main reason for opening the apps.

This notebook was converted with convert.ploomber.io