

```
importing libraries

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

[3]: df = pd.read_csv("expense_data_1.csv")

start

[7]: df.head()

[7]:
```

	Date	Account	Category	Subcategory	Note	INR	Income/Expense	Note.1	Amount	Currency	Account.1
0	3/2/2022 10:11	CUB - online payment	Food	NaN	Brownie	50.0	Expense	NaN	50.0	INR	50.0
1	3/2/2022 10:11	CUB - online payment	Other	NaN	To lended people	300.0	Expense	NaN	300.0	INR	300.0
2	3/1/2022 19:50	CUB - online payment	Food	NaN	Dinner	78.0	Expense	NaN	78.0	INR	78.0
3	3/1/2022 18:56	CUB - online payment	Transportation	NaN	Metro	30.0	Expense	NaN	30.0	INR	30.0
4	3/1/2022 18:22	CUB - online payment	Food	NaN	Snacks	67.0	Expense	NaN	67.0	INR	67.0

```
[11]: df.drop_duplicates()
```

```
Untitled.ipynb
277 rows x 11 columns

[87]: df.isna().sum()

[87]: Date      0
Account      0
Category      0
Subcategory  277
Note         4
INR          0
Income/Expense  0
Note.1       277
Amount       0
Currency      0
Account.1     0
dtype: int64

[89]: df.drop(["Subcategory", "Note.1"], axis=1, inplace=True)

[91]: df.head()

[91]:
```

	Date	Account	Category	Note	INR	Income/Expense	Amount	Currency	Account.1
0	3/2/2022 10:11	CUB - online payment	Food	Brownie	50.0	Expense	50.0	INR	50.0
1	3/2/2022 10:11	CUB - online payment	Other	To lended people	300.0	Expense	300.0	INR	300.0
2	3/1/2022 19:50	CUB - online payment	Food	Dinner	78.0	Expense	78.0	INR	78.0
3	3/1/2022 18:56	CUB - online payment	Transportation	Metro	30.0	Expense	30.0	INR	30.0
4	3/1/2022 18:22	CUB - online payment	Food	Snacks	67.0	Expense	67.0	INR	67.0

```
Untitled.ipynb

[93]: df.describe()

[93]:
```

	INR	Amount	Account.1
count	277.000000	277.000000	277.000000
mean	410.750903	406.759134	406.759134
std	1065.756569	1065.158318	1065.158318
min	3.000000	3.000000	3.000000
25%	50.000000	50.000000	50.000000
50%	128.000000	125.000000	125.000000
75%	301.150000	300.000000	300.000000
max	10000.000000	10000.000000	10000.000000

```
[95]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 277 entries, 0 to 276
Data columns (total 9 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Date        277 non-null    object
1   Account     277 non-null    object
2   Category    277 non-null    object
3   Note        273 non-null    object
4   INR         277 non-null    float64
5   Income/Expense 277 non-null    object
```

```
df['time'] = df['Date'].str.split(" ").str[1]
df['Date'] = df['Date'].str.split(" ").str[0]

df.head()

[13]:
```

	Date	Account	Category	Subcategory	Note	INR	Income/Expense	Note.1	Amount	Currency	Account.1
0	3/2/2022 10:11	CUB - online payment	Food	NaN	Brownie	50.0	Expense	NaN	50.0	INR	50.0
1	3/2/2022 10:11	CUB - online payment	Other	NaN	To lend people	300.0	Expense	NaN	300.0	INR	300.0
2	3/1/2022 19:50	CUB - online payment	Food	NaN	Dinner	78.0	Expense	NaN	78.0	INR	78.0
3	3/1/2022 18:56	CUB - online payment	Transportation	NaN	Metro	30.0	Expense	NaN	30.0	INR	30.0
4	3/1/2022 18:22	CUB - online payment	Food	NaN	Snacks	67.0	Expense	NaN	67.0	INR	67.0

```
df['Currency'].unique()

[15]: array(['INR', 'USD'], dtype=object)

df['Amount'] = df.apply(lambda row: row['Amount'] * 93 if row['Currency'] == 'USD' else row['Amount'], axis=1)

df.drop(['Currency', 'Account.1'], axis=1, inplace=True)

df.head()

[21]:
```

	Date	Account	Category	Subcategory	Note	INR	Income/Expense	Note.1	Amount
0	3/2/2022 10:11	CUB - online payment	Food	NaN	Brownie	50.0	Expense	NaN	50.0
1	3/2/2022 10:11	CUB - online payment	Other	NaN	To lend people	300.0	Expense	NaN	300.0

```
df['Category'].unique()

[23]: array(['Food', 'Other', 'Transportation', 'Social Life', 'Household',
        'Apparel', 'Education', 'Salary', 'Allowance', 'Self-development',
        'Beauty', 'Gift', 'Petty cash'], dtype=object)

df['Account'].unique()

[25]: array(['CUB - online payment', 'Cash'], dtype=object)

df['Account'] = df.apply(lambda row: 'online' if row['Account'] == 'CUB - online payment' else row['Account'], axis=1)

df['Account'].unique()

[29]: array(['online', 'Cash'], dtype=object)
```

