

0s

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
from datetime import datetime, timedelta
```

keyboard\_arrow\_down

## Data Loading

[4]

26s

```
df=pd.read_excel("/content/drive/MyDrive/Bright Coffee Shop Sales.xlsx")
```

[5]

0s

```
display(df)
```

keyboard\_arrow\_down

## Exploratory Data Analysis

[6]

0s

```
df.columns
Index(['transaction_id', 'transaction_date', 'transaction_time',
      'transaction_qty', 'store_id', 'store_location', 'product_id',
      'unit_price', 'product_category', 'product_type', 'product_detail'],
      dtype='object')
```

Column names in the data set

[7]

0s

```
df.dtypes
```

Shows the type of data in each column

Double-click (or enter) to edit

[8]

0s

```
df.duplicated()
```

There are no duplicates in the dataset

```
[9]
```

0s

```
df.duplicated().sum()  
np.int64(0)
```

```
[10]
```

0s

```
df.info()  
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 149116 entries, 0 to 149115  
Data columns (total 11 columns):  
#   Column                Non-Null Count  Dtype    
---  ---                  
0   transaction_id         149116 non-null int64    
1   transaction_date       149116 non-null datetime64[ns]  
2   transaction_time       149116 non-null object    
3   transaction_qty        149116 non-null int64    
4   store_id              149116 non-null int64    
5   store_location         149116 non-null object    
6   product_id            149116 non-null int64    
7   unit_price            149116 non-null float64   
8   product_category      149116 non-null object    
9   product_type          149116 non-null object    
10  product_detail         149116 non-null object    
dtypes: datetime64[ns](1), float64(1), int64(4), object(5)  
memory usage: 12.5+ MB
```

```
[11]
```

0s

```
df.describe()
```

Observation: -Min number of transactions per customer is 1 and max is 8.

- data was first collected on the 01/01/2023 and last date was 30/06/2023
- cheapest product is R0.80 and maximum is R45

```
[14]
```

0s

```
df.shape  
(149116, 11)
```

The data set has 149 116 rows and 11 columns

```
[15]
```

0s

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

```
[ ]
```

keyboard\_arrow\_down

Insight : There are no null values in the data set

Calculating total amount

[16]

0s

```
df['total_amount']= df ['transaction_qty'] * df['unit_price']
```

[20]

0s

```
display (df)
```

[46]

0s

```
grouped= df.groupby ('store_location')['total_amount'] .sum()
```

[47]

0s

```
display (grouped)
```

Total Sales by Store Location

[43]

0s

```
groupedcat= df.groupby ('product_category')['total_amount'] .sum()
```

[45]

0s

```
display(groupedcat)
```

Product which generates the highest revenue top 3 products at Bright coffee are  
1) Coffee 2)Tea 3) Bakery

[48]

0s

```
df['month_id'] = df['transaction_date'].dt.month
```

[49]

0s

```
display(df)
```

keyboard\_arrow\_down

Adding Month ID to the data set

[51]

0s

```
groupedmnth= df.groupby ('month_id')['total amount'].sum()
```

[52]

0s

```
display(groupedmnth)
```