In [ ]:
```
print("Hello world i am roushan")
```

```
Hello world i am roushan
```

In [ ]:
```
import pandas as pd
```

In [ ]:
```
print(pd.__version__)
```

```
2.2.3
```

In [ ]:
```
df=pd.read_csv('C:\\Users\\caree\\OneDrive\\Desktop\\Project_walmart\\Walmart.cs
```

In [ ]:
```

```

In [ ]:
```
df.shape
```

Out[ ]:
```
(9969, 11)
```

In [ ]:
```
df.dropna(inplace=True)
```

In [ ]:
```
df.head(5)
```

Out[ ]:

	invoice_id	Branch	City	category	unit_price	quantity	date	time
0	1	WALM003	San Antonio	Health and beauty	74.69	7.0	05/01/19	13:08:00
1	2	WALM048	Harlingen	Electronic accessories	15.28	5.0	08/03/19	10:29:00
2	3	WALM067	Haltom City	Home and lifestyle	46.33	7.0	03/03/19	13:23:00
3	4	WALM064	Bedford	Health and beauty	58.22	8.0	27/01/19	20:33:00
4	5	WALM013	Irving	Sports and travel	86.31	7.0	08/02/19	10:37:00

In [ ]:
```
df.describe()
```

Out[ ]:

	invoice_id	quantity	rating	profit_margin
count	10051.000000	10020.000000	10051.000000	10051.000000
mean	5025.741220	2.353493	5.825659	0.393791
std	2901.174372	1.602658	1.763991	0.090669
min	1.000000	1.000000	3.000000	0.180000
25%	2513.500000	1.000000	4.000000	0.330000
50%	5026.000000	2.000000	6.000000	0.330000
75%	7538.500000	3.000000	7.000000	0.480000
max	10000.000000	10.000000	10.000000	0.570000

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

```
<class 'pandas.core.frame.DataFrame'>
Index: 9969 entries, 0 to 9999
Data columns (total 12 columns):
 #   Column          Non-Null Count  Dtype  
---  ------          --------------  -----  
 0   invoice_id      9969 non-null   int64  
 1   Branch          9969 non-null   object 
 2   City            9969 non-null   object 
 3   category        9969 non-null   object 
 4   unit_price      9969 non-null   float64
 5   quantity        9969 non-null   float64
 6   date            9969 non-null   object 
 7   time            9969 non-null   object 
 8   payment_method  9969 non-null   object 
 9   rating          9969 non-null   float64
 10  profit_margin   9969 non-null   float64
 11  Total           9969 non-null   float64
dtypes: float64(5), int64(1), object(6)
memory usage: 1012.5+ KB
```

In [ ]:
```
df.drop_duplicates(inplace=True)
```

In [ ]:
```
df['unit_price']=df['unit_price'].astype(float)
```

In [ ]:
```
df['Total']=df['unit_price']*df['quantity']
```

In [ ]:
```
import pymysql
from sqlalchemy import create_engine
```

In [ ]:
```
engine_mysql=create_engine("mysql+pymysql://root@localhost:3306/walmart_db")
try:
    engine_mysql
    print("connected")
except:
    print("Not Connected")
```

```
connected
```

In [ ]:
```
df.to_sql(name='Walmart',con=engine_mysql,if_exists='append',index=False)
```

```
C:\Users\caree\AppData\Local\Temp\ipykernel_10532\3120032506.py:1: UserWarning: T
he provided table name 'Walmart' is not found exactly as such in the database aft
er writing the table, possibly due to case sensitivity issues. Consider using low
er case table names.
  df.to_sql(name='Walmart',con=engine_mysql,if_exists='append',index=False)
```

Out[ ]: 9969

In [ ]:
```
import tex
```