Retail_Sales

November 9, 2025

[2]:
```
import pandas as pd
from sqlalchemy import create_engine
import os
os.chdir(r'C:\Users\Muneer Khan\OneDrive\Desktop\Project')
print(os.getcwd())
```

```
C:\Users\Muneer Khan\OneDrive\Desktop\Project
```

[12]:
```
# 1. Load dataset
df = pd.read_csv(r"C:\Users\Muneer Khan\OneDrive\Desktop\Project\retail_sales.
 ↪csv")
(df.head())
```

[12]:

	Order_ID	Date	Customer_Name	Region	Category	Quantity	Sales	Profit
0	67a9c378-7c65-41e5-82e2-e662f728b4fa	2023-11-24	Steve Williams	North	Books	3	904.49	76.78
1	37ebdcd9-e87a-4613-a443-df789558867f	2023-12-23	Elizabeth Woods	South	Electronics	10	700.03	40.21
2	e8e5216a-fcbd-44c3-8021-2ef7cca5a5a1	2023-07-16	Charles Davis	South	Groceries	10	1520.14	89.53
3	d9b8a714-e61a-441c-92e0-c8b2bad640fb	2023-07-26	Dr. Stephanie Collins	West	Books	4	121.88	16.44
4	e9bb17bc-a3f2-49bf-9c63-16b950f24455	2023-09-23	Justin Gomez	North	Clothing	5	1579.85	175.42

[14]:
```
# 2. Clean and preprocess
df['Date'] = pd.to_datetime(df['Date'], errors='coerce')
df = df.dropna(subset=['Date', 'Sales', 'Profit'])
df['Month'] = df['Date'].dt.month_name()
```

[20]:
```
# 3. Derive new metric
df['Profit_Percentage'] = (df['Profit'] / df['Sales'])
```

[21]:
```
# 4. Connect to SQLite database
engine = create_engine(r'sqlite:///C:\Users\Muneer␣
 ↪Khan\OneDrive\Desktop\Project\retail_sales.db')
```

[22]:
```
# 5. Store cleaned data into SQL table
df.to_sql('sales_data', con=engine, if_exists='replace', index=False)

print("  Data cleaned and loaded successfully into SQL database!")
```

```
 Data cleaned and loaded successfully into SQL database!
```

[34]:
```
df.to_csv(r'C:\Users\Muneer Khan\OneDrive\Desktop\Project\Sales.
 ↪csv',index=False)
```

[ ]:
```

```