Using Python

1.Data Cleaning and Preprocessing

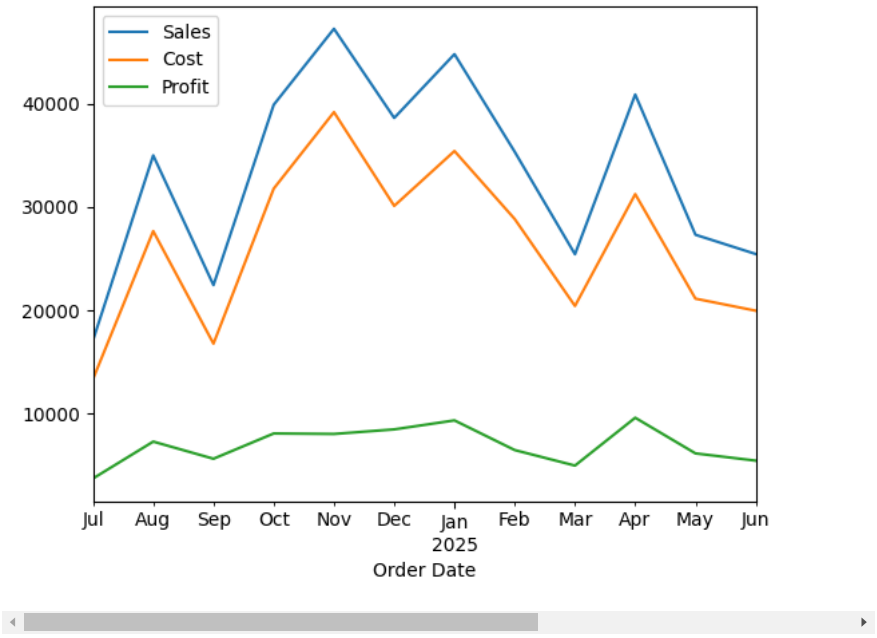
```
import pandas as pd

df = pd.read_csv("/sample_retail_data.csv")
df['Order Date'] = pd.to_datetime(df['Order Date'])

# Example: create Profit column
df['Profit'] = df['Sales'] - df['Cost']
```

2.Exploratory Data Analysis (EDA)

a. Sales, Cost, and Profit Over Time

```
df.set_index('Order Date').resample('M')[['Sales', 'Cost', 'Profit']].sum().plot()
```

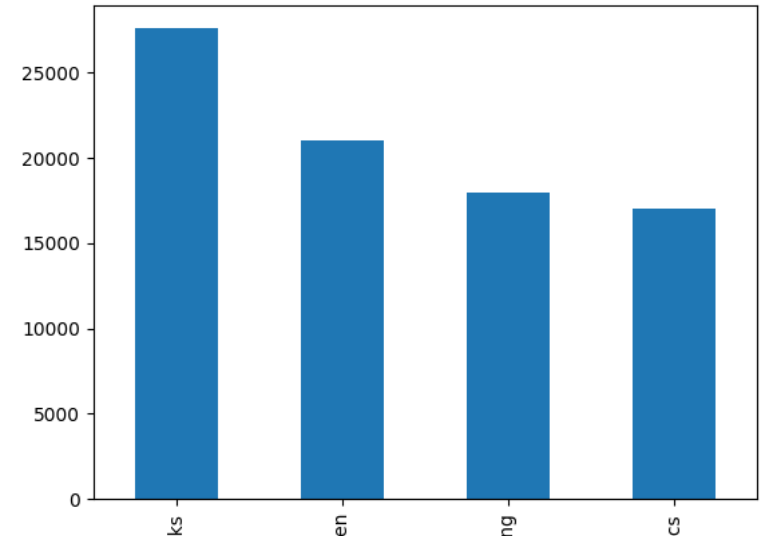
```
/tmp/ipython-input-3-1633758727.py:1: FutureWarning: 'M' is deprecated and will be re
  df.set_index('Order Date').resample('M')[['Sales', 'Cost', 'Profit']].sum().plot()
<Axes: xlabel='Order Date'>
```



b. Category-wise Profitability

```
df.groupby('Category')['Profit'].sum().sort_values(ascending=False).plot(kind='bar')
```

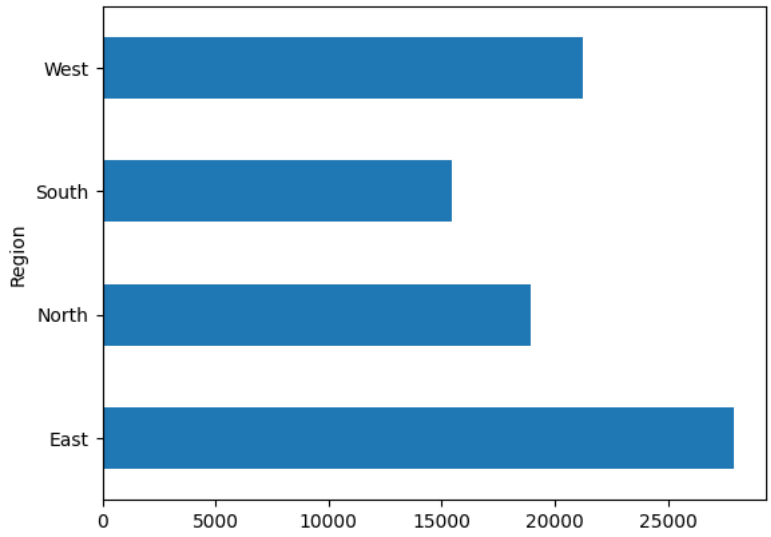
```
<Axes: xlabel='Category'>
```



c. Region-wise or Segment-wise Analysis

```
df.groupby('Region')['Profit'].sum().plot(kind='barh')
```

```
<Axes: ylabel='Region'>
```



d. Top 10 Most Profitable Products

```
df.groupby('Product')['Profit'].sum().sort_values(ascending=False).head(10)
```

Product	Profit
Scene	2127.40
Heavy	2089.00
True	1960.68
Again	1930.46
Western	1862.37
Meeting	1799.90
Easy	1765.84
Day	1747.70
Yes	1580.85
Term	1535.50

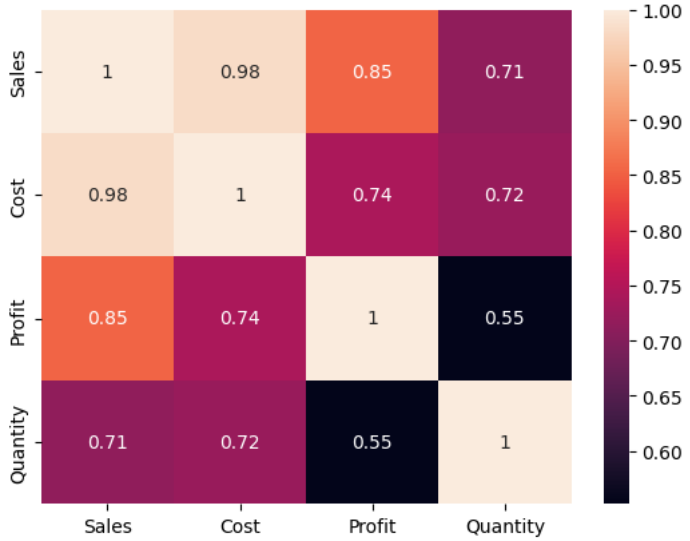
dtype: float64

3.Correlation and Trend Insights

Correlation matrix for Sales, Cost, Profit, Quantity, etc.

```
import seaborn as sns
sns.heatmap(df[['Sales', 'Cost', 'Profit', 'Quantity']].corr(), annot=True)
```

```
<Axes: >
```



Profit trends over quarters/seasons:

```
df['Quarter'] = df['Order Date'].dt.to_period('Q')
df.groupby('Quarter')['Profit'].sum().plot()
```

```
<Axes: xlabel='Quarter'>
```