LEVEL 1: BASIC TASKS

Task 1: Data Cleaning and Preprocessing (Stock Price Data & Cleaning Iris)

In [9]:
```
import pandas as pd
import warnings
warnings.filterwarnings('ignore')

# Load the dataset
df = pd.read_csv("stock_prices_cleaned.csv")  # Replace with your actual filename

# Convert 'date' to datetime
df['date'] = pd.to_datetime(df['date'])

# Check for missing values
print(df.isnull().sum())

# Drop duplicates
df = df.drop_duplicates()

# Save cleaned file
df.to_csv("stock_prices_cleaned.csv", index=False)
```

```
symbol    0
date      0
open      0
high      0
low       0
close     0
volume    0
dtype: int64
```

In [10]:
```
df.head(5)
```

Out[10]:

	symbol	date	open	high	low	close	volume
0	AAL	2014-01-02	25.07	25.82	25.06	25.36	8998943
1	AEE	2014-01-02	36.05	36.11	35.47	35.53	1394264
2	ITW	2014-01-02	83.97	84.10	83.07	83.19	1390787
3	AAP	2014-01-02	110.36	111.88	109.29	109.74	542711
4	ABBV	2014-01-02	52.12	52.33	51.52	51.98	4569061

In [17]:
```
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.datasets import load_iris
```

```
import pandas as pd

# Load Iris
iris = load_iris(as_frame=True)
df = iris.frame

# Check for missing values
df.isna().sum()

# Drop duplicates
df = df.drop_duplicates()

# Save cleaned file
df.to_csv("iris.csv", index=False)
```

In [18]:
```
df.head(5)
```

Out[18]:

	sepal length (cm)	sepal width (cm)	petal length (cm)	petal width (cm)	target
0	5.1	3.5	1.4	0.2	0
1	4.9	3.0	1.4	0.2	0
2	4.7	3.2	1.3	0.2	0
3	4.6	3.1	1.5	0.2	0
4	5.0	3.6	1.4	0.2	0