

Exploratory Data Analysis (EDA)

Objective:

Perform Exploratory Data Analysis (EDA) on the Titanic Dataset to extract meaningful insights, identify trends, and find patterns that could help understand survival outcomes.

Tools and Technologies Used

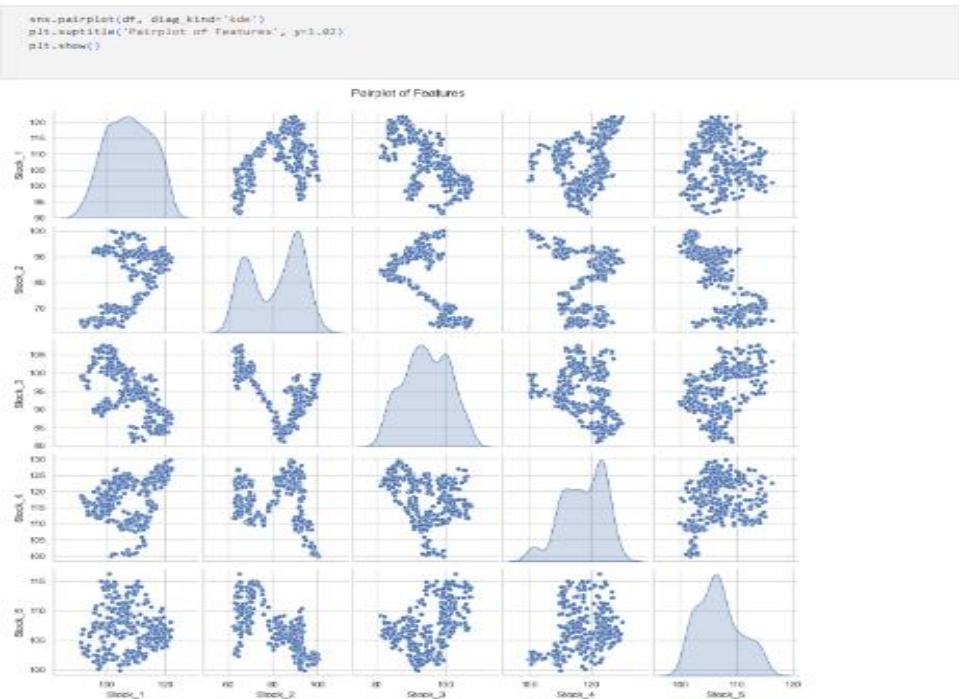
- **Python** (Programming Language)
- **Jupyter Notebook** (Development Environment)
- **Pandas** (Data manipulation and analysis)
- **Matplotlib** (Data visualization)
- **Seaborn** (Statistical data visualization)

Dataset:

Stock_data:

Here's what the dataset looks like:

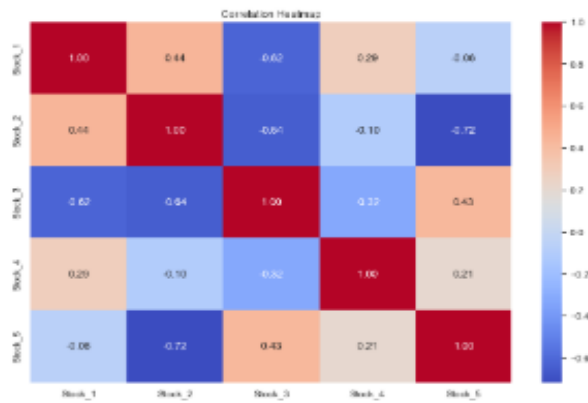
Date (Unnamed: 0)	Stock_1	Stock_2	Stock_3	Stock_4	Stock_5
2020-01-01	101.76	100.16	99.49	99.91	101.76
2020-01-02	102.17	99.97	98.68	100.64	102.53
2020-01-03	103.17	99.58	98.18	100.57	101.89
...



```

plt.figure(figsize=(12,8))
# Select only numeric columns
sns.heatmap(df.select_dtypes(include=[np.number]).corr(), annot=True, cmap='coolwarm', fmt=".2f")
plt.title("Correlation Heatmap")
plt.show()

```

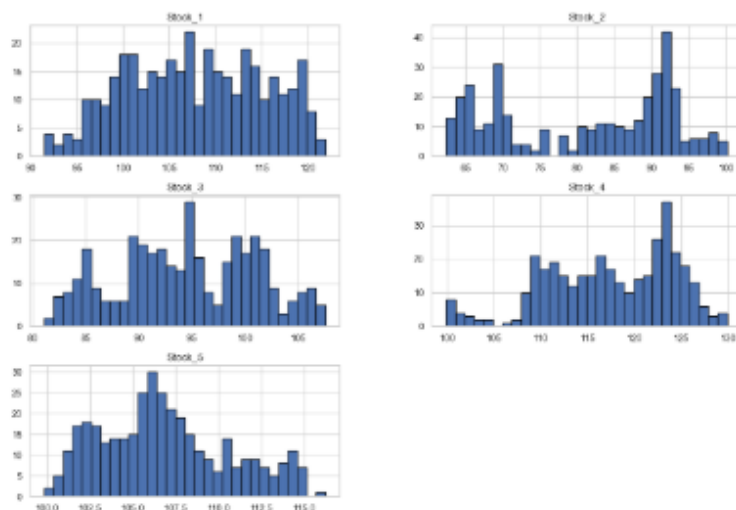


```

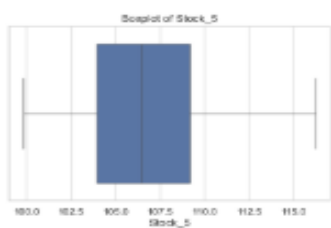
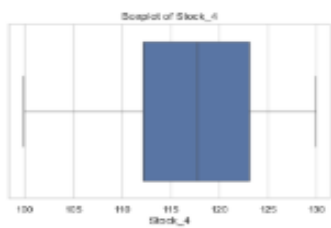
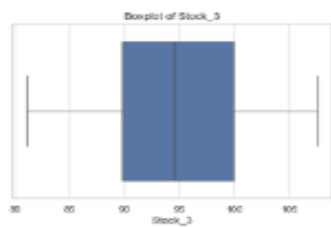
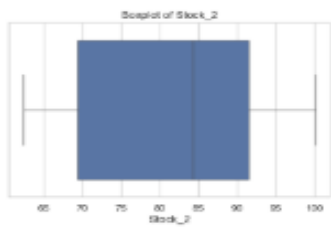
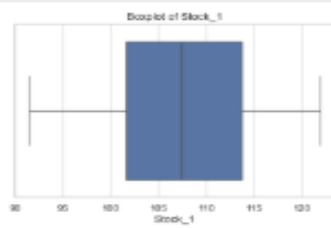
df.hist(figsize=(15,10), bins=10, edgecolor='black')
plt.suptitle("Histograms of Numerical Features", y=1.02)
plt.show()

```

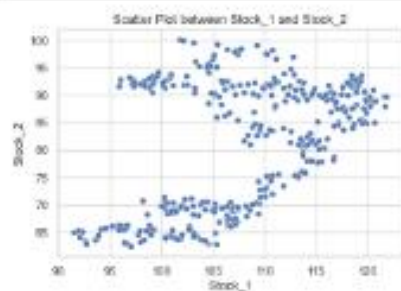
Histograms of Numerical Features



```
# Draw boxplots for numerical columns
for col in df.select_dtypes(include=['float64', 'int64']).columns:
    plt.figure(figsize=(6,4))
    sns.boxplot(x=df[col])
    plt.title(f'Boxplot of {col}')
    plt.show()
```



```
num_cols = df.select_dtypes(include=[np.number]).columns
plt.figure(figsize=(6,6))
sns.scatterplot(x=num_cols[0], y=num_cols[1], data=df)
plt.title(f'Scatter Plot between {num_cols[0]} and {num_cols[1]}')
plt.show()
```



Observations to Write

When you plot:

- **Pairplot:** Check which stocks move together.
- **Heatmap:** See strong correlations (positive or negative).
- **Histograms:** See the distribution (normal, skewed, etc.).
- **Boxplots:** Identify outliers.
- **Scatterplots:** Find relationships between stock pairs.

✓ After each plot, note 1–2 key points (e.g., “Stock_1 and Stock_2 are highly positively correlated.”)

5. Summary of Findings Example:

At the end, you can write something like:

Summary:

- Stock_1 and Stock_2 show a strong positive correlation (~ 0.9).
- Stock_3 has slightly more variance compared to others (seen in boxplot).
- Stock_5 prices are more stable (narrow histogram).
- No severe outliers detected except slight spikes in Stock_3.