

Exploratory Data Analysis (EDA)

Wednesday 10th December, 2025



Description: This notebook we conduct EDA on the financial data.

```
#installing the relevant libraries and dataset
import pandas as pd
import matplotlib.pyplot as plt
import os

df = pd.read_csv("../data/sp500.csv", index_col = "Date", parse_dates = True)

#First
plt.figure(figsize=(10, 6))
df["Close"].plot(title="S&P 500 Close Price Over Time")
plt.ylabel("Price")
plt.grid(True)
plt.savefig("../images/close_price.png")
plt.show()
```

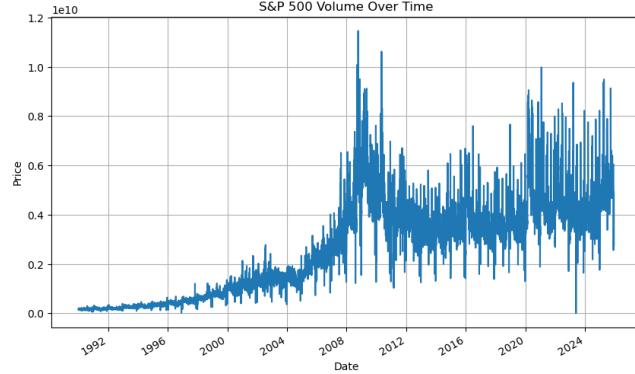


```
#Second
plt.figure(figsize=(10, 6))
df["Volume"].plot(title="S&P 500 Volume Over Time")
plt.ylabel("Price")
```

```

plt.grid(True)
plt.savefig("../images/volume.png")
plt.show()

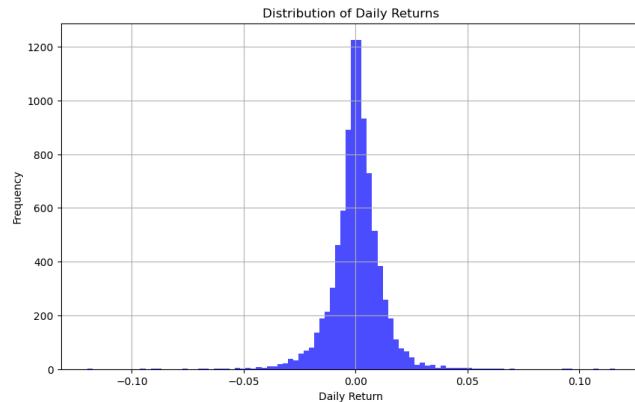
```



```

#Third
plt.figure(figsize=(10, 6))
df[["Return"]].hist(bins=100, color= "blue", alpha=0.7)
plt.title("Distribution of Daily Returns")
plt.xlabel("Daily Return")
plt.ylabel("Frequency")
plt.savefig("../images/returns.png")
plt.show()

```



```

print("Basic stats of daily returns:")
print(df[["Return"]].describe())

```

```

Basic stats of daily returns:
count    9051.000000
mean      0.000390
std       0.011377

```

```

min      -0.119841
25%     -0.004429
50%      0.000603
75%      0.005710
max      0.115800
Name: Return, dtype: float64

#Fourth
corr = df[["Close", "Return", "LogReturn", "MA10", "MA50", "EMA10", "EMA50", "Volatility20",

plt.imshow(corr, cmap="coolwarm", vmin= -1, vmax=1)
plt.colorbar()
plt.xticks(range(len(corr.columns)), corr.columns, rotation=90)
plt.yticks(range(len(corr.columns)), corr.columns)
plt.title("Correlation Matrix of Indicators")
plt.tight_layout()
plt.savefig("../images/correlation_matrix.png")
plt.show()

```

