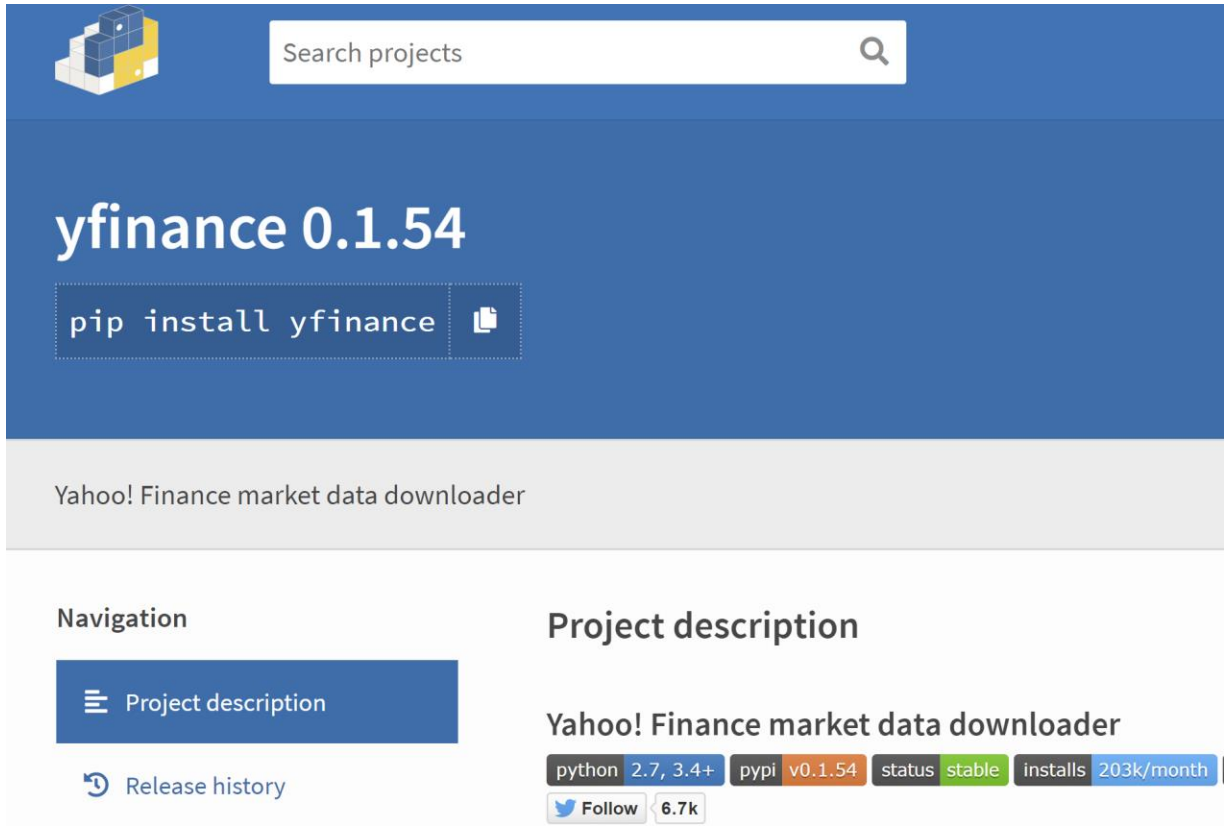
S&P500 STOCK ANALYSIS AND PREDICTION

- *Allisa Dao & Long Bui* -

What we have done

1. Data Gathering

1) yFinance: open-source library to acess the financial database on Yahoo Finance



2) Data that is retreivable:
+ stock, bonds, currencies, cryptocurrencies
+ market news, reports, analysis
+ company information

Data that we collected:

+ real-time and historical od S&P500 bond data

+ information about S&P500 companies

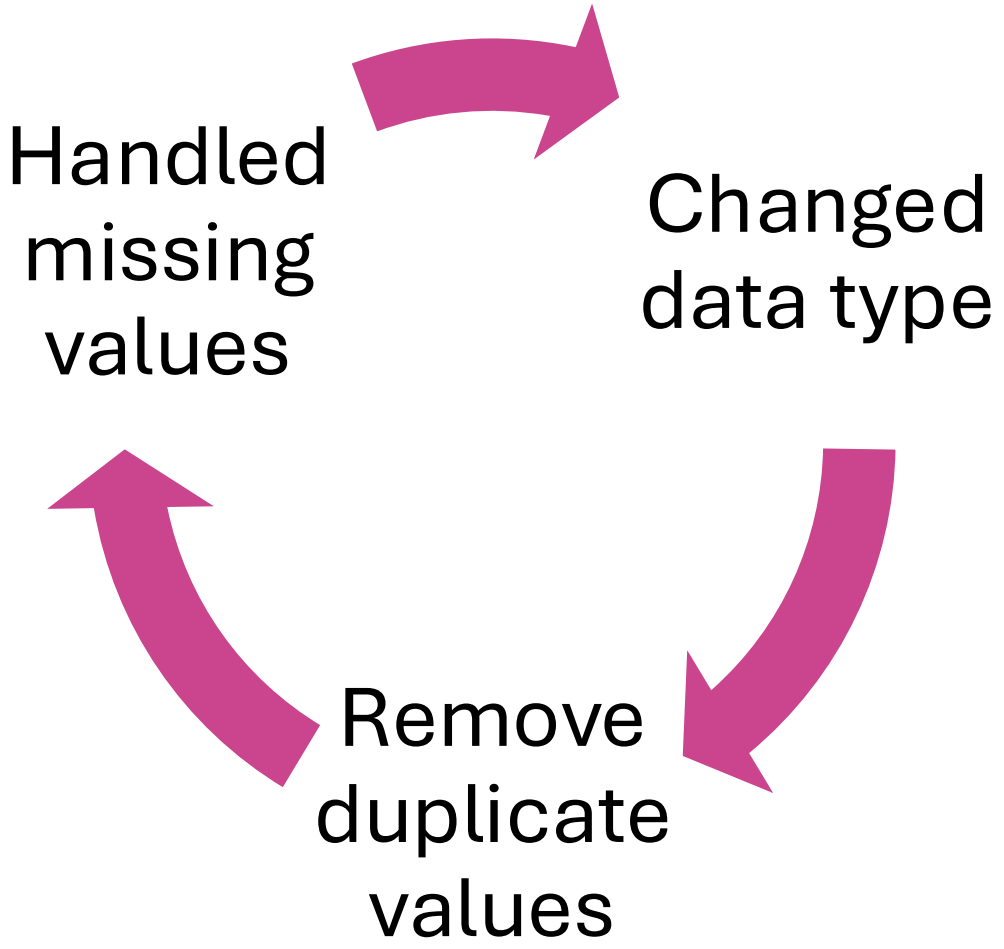
+ financial data by quarter of S&P500 companies including balance sheet statements, income statements, cash-flow statements

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
		Date	Open	High	Low	Close	Volume	Dividends	Stock Split	Ticker	SMA_20	EMA_20	RSI	MACD	Signal	Middle_Ba	Upper_Bar	Lower_Band	
0		2022-11-0	94.373138	94.76689	93.366022	94.410995	2289503	0	0	MMM		94.41099548339844		0	0				
1		2022-11-0	94.925923	95.895174	93.92638	94.918350	2182461	0	0	MMM		94.45931498209636		0.040473	0.008095				
2		2022-11-0	94.872905	95.168224	93.684062	93.76736	2131750	0	0	MMM		94.39341420975943		-0.0201	0.002456				
3		2022-11-1	96.66753	97.879093	96.069327	97.7958	3779001	0	0	MMM		94.71745088077006		0.2540357	0.052772				
4		2022-11-1	98.128992	101.08217	97.795808	100.69599	4101443	0	0	MMM		95.28684		0.6972708	0.18167198156119643				
5		2022-11-1	100.37038	101.40021	99.151250	99.219398	3127899	0	0	MMM		95.66136549321716		0.9187978	0.3290971491776602				
6		2022-11-1	100.34765	101.07459	98.26528	99.363265	3008060	0	0	MMM		96.01392744540705		1.0933648	0.4819506827974669				
7		2022-11-1	99.537426	99.537426	97.235454	97.485343	3837127	0	0	MMM		96.15406234899737		1.0678680	0.5991341635127206				
8		2022-11-1	96.642628	97.462352	95.50881	97.408721	2689804	1.245819	0	MMM		96.27355		1.0296102	0.6852293870572465				
9		2022-11-1	98.059909	98.320379	96.451108	97.109954	3798616	0	0	MMM		96.35321098440106		0.9640694	0.7409974085973589				
10		2022-11-2	97.018026	97.96798	96.703926	97.776458	2575586	0	0	MMM		96.48875838971851		0.9549017	0.7837782731882187				
11		2022-11-2	98.17482	98.573192	97.538968	98.519569	2101492	0	0	MMM		96.68216896183795		0.9961164	0.8262459030835687				
12		2022-11-2	98.511906	98.91794	97.47768	98.05991	2810241	0	0	MMM		96.81338		0.980388	0.8570742594427982				
13		2022-11-2	98.734059	99.45419	98.458266	98.856636	1055350	0	0	MMM		97.00798	64.694635	1.0204482	0.889749				
14		2022-11-2	98.144189	98.504251	95.087477	95.485847	3317106	0	0	MMM		96.863013	52.905392	0.7713109	0.8660614443791989				
15		2022-11-2	95.05683	96.788196	95.003201	96.62732	2228268	0	0	MMM		96.840566	54.466339	0.658386	0.8245262757323107				
16		2022-11-3	96.451107	96.719245	92.973049	96.50474	6973398	0	0	MMM		96.80858	57.560472	0.5526297	0.7701469746903885				
17		2022-12-0	97.523636	98.726406	95.999111	96.520057	3156005	0	0	MMM		96.781104	45.472921	0.4646969	0.7090569712215037				
18		2022-12-0	95.577768	97.485340	95.1794	97.28616	2178992	0	0	MMM		96.829204	35.740093	0.4516212	0.6575698306548988				
19		2022-12-0	95.815249	96.114027	95.202372	95.47817	2563148	0	0	MMM	97.164502	96.70053	34.776186	0.2920035	0.5844565	97.164502	100.68740	93.6416	
20		2022-12-0	95.876542	96.320881	94.466933	95.439872	2295722	0	0	MMM	97.215946	96.580471	33.896557	0.1605639	0.4996780	97.215946	100.59673	93.83515795228479	
21		2022-12-0	95.248350	97.53131	95.072150	96.795860	3230994	0	0	MMM	97.30982	96.600984	47.043358	0.1639243	0.4325273	97.30982	100.5221	94.09759	
22		2022-12-0	97.293818	99.163082	96.10637	96.527725	4527219	0	0	MMM	97.447840	96.594007	46.283161	0.1432994	0.3746817	97.447840	100.22724	94.66843759386924	
23		2022-12-0	97.125273	97.699843	96.274911	96.343864	2792540	0	0	MMM	97.375243	96.570184	46.736296	0.1108402	0.3219134	97.375243	100.19197	94.55851128949924	
24		2022-12-1	96.58902	97.209555	95.14876	97.17891	4637849	0	0	MMM	97.199389	96.628158	47.490355	0.1507594	0.2876826	97.199389	99.542535	94.85624298475348	
25		2022-12-1	99.170754	99.607424	96.8878	97.51599	3895252	0	0	MMM	97.11422	96.712713	45.636248	0.2072068	0.2715874	97.11422	99.264068	94.96436904000896	
26		2022-12-1	96.910772	98.419974	95.63906	96.596672	3702457	0	0	MMM	96.975889	96.701662	43.882101	0.1757346	0.2524169	96.975889	98.855460	95.09631767692585	

A	B	C	D	E	F	G	H
Symbol	Security	GICS Sector	GICS Sub-Industry	Headquarters Locatic	Date added	CIK	Founded
MMM	3M	Industrials	Industrial Conglomer	Saint Paul, Minneso	1957-03-04	66740	1902
AOS	A. O. Smith	Industrials	Building Products	Milwaukee, Wisconsi	2017-07-26	91142	1916
ABT	Abbott Laboratories	Health Care	Health Care Equipm	North Chicago, Illino	1957-03-04	1800	1888
ABBV	AbbVie	Health Care	Biotechnology	North Chicago, Illino	2012-12-31	1551152	2013 (1888)
ACN	Accenture	Information Technol	IT Consulting & Othe	Dublin, Ireland	2011-07-06	1467373	1989
ADBE	Adobe Inc.	Information Technol	Application Software	San Jose, California	1997-05-05	796343	1982
AMD	Advanced Micro Dev	Information Technol	Semiconductors	Santa Clara, Californ	2017-03-20	2488	1969
AES	AES Corporation	Utilities	Independent Power I	Arlington, Virginia	1998-10-02	874761	1981
AFL	Aflac	Financials	Life & Health Insurar	Columbus, Georgia	1999-05-28	4977	1955
A	Agilent Technologie	Health Care	Life Sciences Tools	Santa Clara, Californ	2000-06-05	1090872	1999
APD	Air Products	Materials	Industrial Gases	Upper Macungie Tov	1985-04-30	2969	1940
ABNB	Airbnb	Consumer Discretion	Hotels, Resorts & Cru	San Francisco, Calif	2023-09-18	1559720	2008
AKAM	Akamai Technologie	Information Technol	Internet Services & In	Cambridge, Massach	2007-07-12	1086222	1998
ALB	Albemarle Corporati	Materials	Specialty Chemicals	Charlotte, North Carc	2016-07-01	915913	1994

cker	Amortization	Amortization Of Intar	Average Dilution Ear	Basic Average Share	Basic EPS	Cost Of Revenue	Depletion Income St	Depreciation Amorti	Depreciation And Ar	Depreciation Income	Diluted Average Sha	Diluted EPS	Diluted NI Availto Cc	EBIT	EBITDA	Earnings From Equit	Earnings From Equit	Excise Taxes	Gain On Sale Of Bu	Gain
				294000000	4.22	3368000000					296000000	4.19	1240000000	1434000000	1705000000					
APL						210352000000							93736000000	123216000000	134661000000					
BBV				1765940733	2.753773	20415000000					1773000000	2.72	4820000000	8474000000	17172000000					
BNB				637000000	7.52	1703000000					662000000	7.24	4792000000	2185000000	2229000000					
BT	1966000000	1966000000		1734076358	3.300316	17975000000		1966000000	1966000000		1749000000	3.26	5723000000	7301000000	10544000000					
CGL	95000000			368700000	11.94				95000000		378800000	11.62	4403000000	3518000000			184000000			
CN			7198000	627852613	11.57	43734147000					635940044	11.44	7271985000	9758292000	11188334000				0	
CBE	168000000	168000000		457000000	11.87	2354000000		168000000	168000000		459000000	11.82	5428000000	6912000000	7784000000					
CI	959618000	959618000		502232000	6.6	4428321000		959618000	959618000		505959000	6.55	3314579000	3872644000	6165747000					
CM				541000000	6.44	86422000000					542000000	6.43	3483000000	4941000000	6000000000	551000000				
CP				410600000	9.14	10476700000					412200000	9.1	3752000000	5233700000	5795600000					
CSK				214000000	4.23						216000000	4.19								
EE				262800000	4.39	4033000000		1387000000	1387000000		263400000	4.38	1152000000	1906000000	3406000000	1000000				
EP				518903682	4.26	7854600000		3090400000	3090400000		520206258	4.24	2208100000	4015600000	7202600000		58500000		0	
ES				669000000	0.37	10164000000					712000000	0.35	249000000	1423000000	2551000000		-32000000		134000000	
FL				596173000	7.81						598745000	7.78	4659000000	5457000000						
G				719506291	5.02						725233068	4.98	3614000000	4994000000						
Z				53455139	12.02						53783069	11.95	642500000	914800000						
IG	531300000	531300000		214900000	4.51	5826600000		696500000	696500000	165200000	219300000	4.42	969500000	1481800000	2178300000					
AM	66751000	66751000		152510000	3.59	1511063000		132568000	132568000	65817000	155397000	3.52	547629000	670236000	1241012000		1475000			
B				117317000	13.41	8431294000					117766000	13.36	1573476000	362810000	792754000		1854082000			

2. Data Cleaning

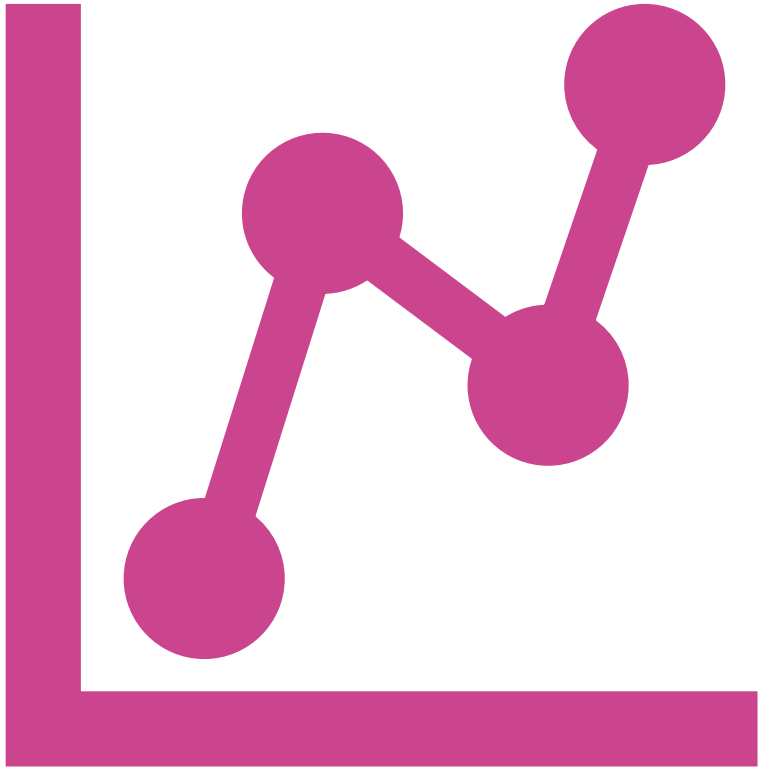


Model and data analysis planning

EDA questions: Using Apple Stock as a specific case for stock analysis

Stock Market Data

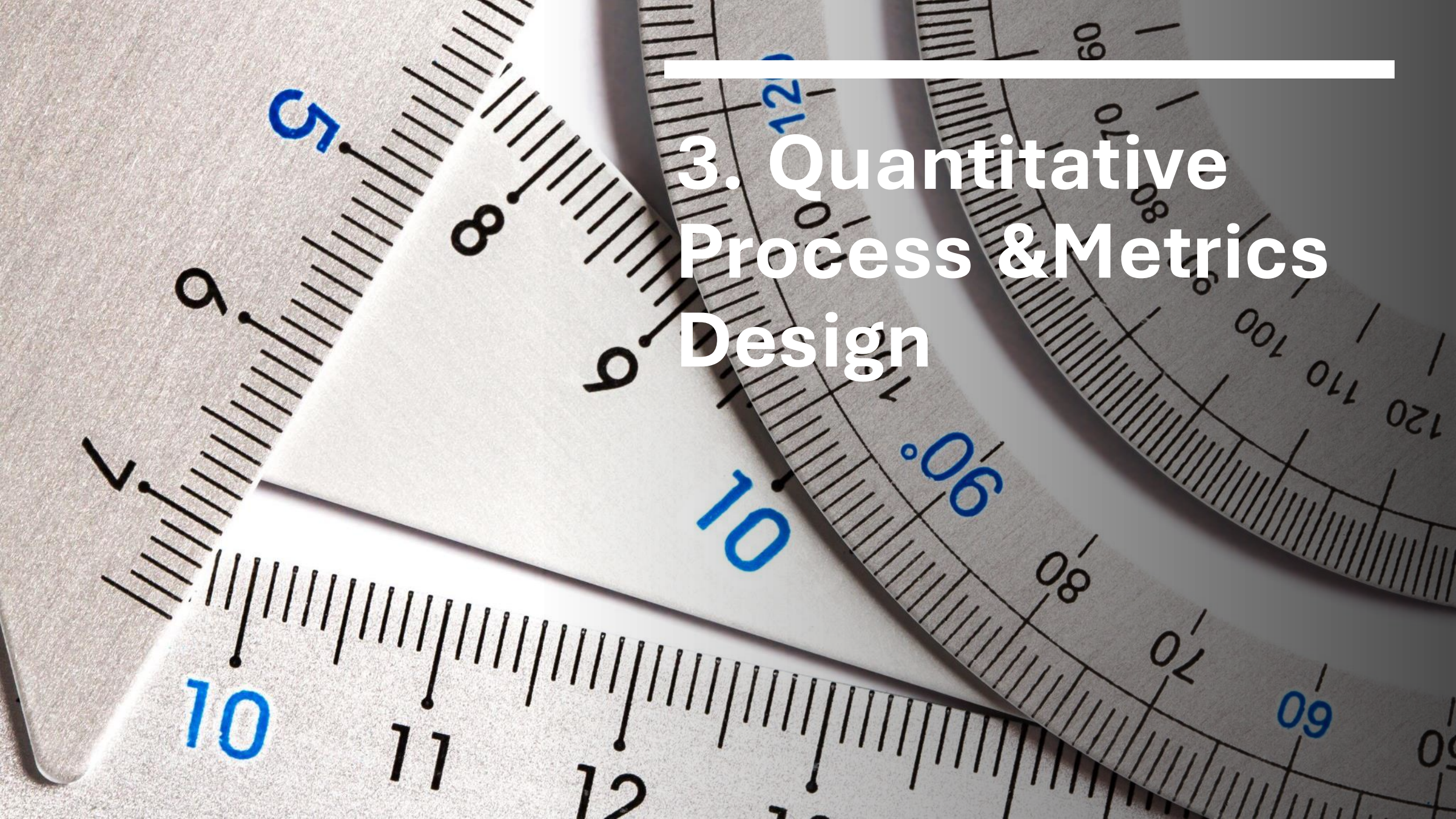
- How have stock prices (open, high, low, close) varied for different companies over the past two years?
- What are the average daily trading volumes for companies, and how do they differ across sectors?
- Are there any seasonal trends in stock prices across the two years of historical data?
- How often do dividends or stock splits occur in the dataset, and which companies have had the most frequent changes?



EDA questions

Comparative Questions

- How do companies with high Net Income compare in stock performance to those with lower Net Income?
- Is there a correlation between financial metrics (e.g., EBITDA, Total Revenue) and stock performance (e.g., average closing price)?
- How does the volatility in stock prices (difference between high and low prices) differ by sector?



3. Quantitative Process &Metrics Design

Moving Average

Simple Moving Average (SMA)

The Simple Moving Average (SMA) is calculated by taking the arithmetic mean of a given set of values over a specified period. For a series of closing prices, the formula is:

$$SMA = \frac{\sum_{i=1}^{n} P_i}{n}$$

Where:

- P_i : is the price of an asset at period i
- n : is the number of periods

Exponential Moving Average (EMA)

The Exponential Moving Average (EMA) gives more weight to recent prices. The formula for EMA is:

$$EMA_t = \alpha \times P_t + (1 - \alpha) \times EMA_{t-1}$$

Where:

- EMA_t is the EMA value at time t
- P_t is the price at time t
- α is the smoothing factor, calculated as $\frac{2}{n+1}$
- n is the number of periods

For the initial EMA calculation, you can use the SMA of the first n periods as the starting point.

RSI and MACD charts

Creating an RSI Chart: A Step-by-Step Guide

The Relative Strength Index (RSI) is a popular momentum oscillator used in technical analysis.

Basic RSI Formula

The RSI is calculated using the following formula:

$$RSI = 100 - \frac{100}{1+RS}$$

Where: $RS = \frac{\text{Average Gain}}{\text{Average Loss}}$

Detailed Calculation Steps

1. **Calculate Average Gain and Average Loss**: For the first 14 periods: $\text{First Average Gain} = \frac{\text{Sum of Gains over the past 14 perio}}{14}$
   $\text{First Average Loss} = \frac{\text{Sum of Losses over the past 14 periods}}{14}$
2. **Subsequent Calculations**: For the 15th period onward: $\text{Average Gain} = \frac{(\text{Previous Average Gain} \times 13 + \text{Current Gain})}{14}$
   $\text{Average Loss} = \frac{(\text{Previous Average Loss} \times 13 + \text{Current Loss})}{14}$
3. **Calculate RS**: $RS = \frac{\text{Average Gain}}{\text{Average Loss}}$
4. **Calculate RSI**: $RSI = 100 - \frac{100}{1+RS}$

Interpreting RSI

- **RSI > 70**: Generally considered overbought
- **RSI < 30**: Generally considered oversold
- **RSI = 50**: Neutral

tutorial will guide you through calculating and interpreting the MACD chart.

Calculating MACD

The MACD consists of three components:

1. **MACD Line**: The difference between two exponential moving averages (EMAs)
2. **Signal Line**: An EMA of the MACD Line
3. **MACD Histogram**: The difference between the MACD Line and Signal Line

Formula

$$MACDLine = EMA_{12period} - EMA_{26period}$$

$$SignalLine = EMA_{9-period} of MACDLine$$

$$MACDHistogram = MACDLine - SignalLine$$

$$EMA = (Close - EMA_{previous}) \times Multiplier + EMA_{previous}$$

Where:

- **Multiplier** = $2/(number of periods + 1)$
- **For 12-period EMA: $Multiplier = 2 / (12 + 1) = 0.1538**
- **For 26-period EMA: Multiplier = 2 / (26 + 1) = 0.0741**

Interpretation

- **MACD Line > 0: Short-term momentum is bullish**

Boillinger Graph

1. Middle Band: A simple moving average (SMA)
2. Upper Band: Middle Band + (Standard Deviation × 2)
3. Lower Band: Middle Band - (Standard Deviation × 2)

Formula The Middle Band is a simple moving average (SMA), typically using a 20-day period:

$$MB = SMA(TP, n)$$

Where:

- TP = Typical Price = (High + Low + Close) / 3
- n = Number of days (usually 20)

Upper and Lower Bands

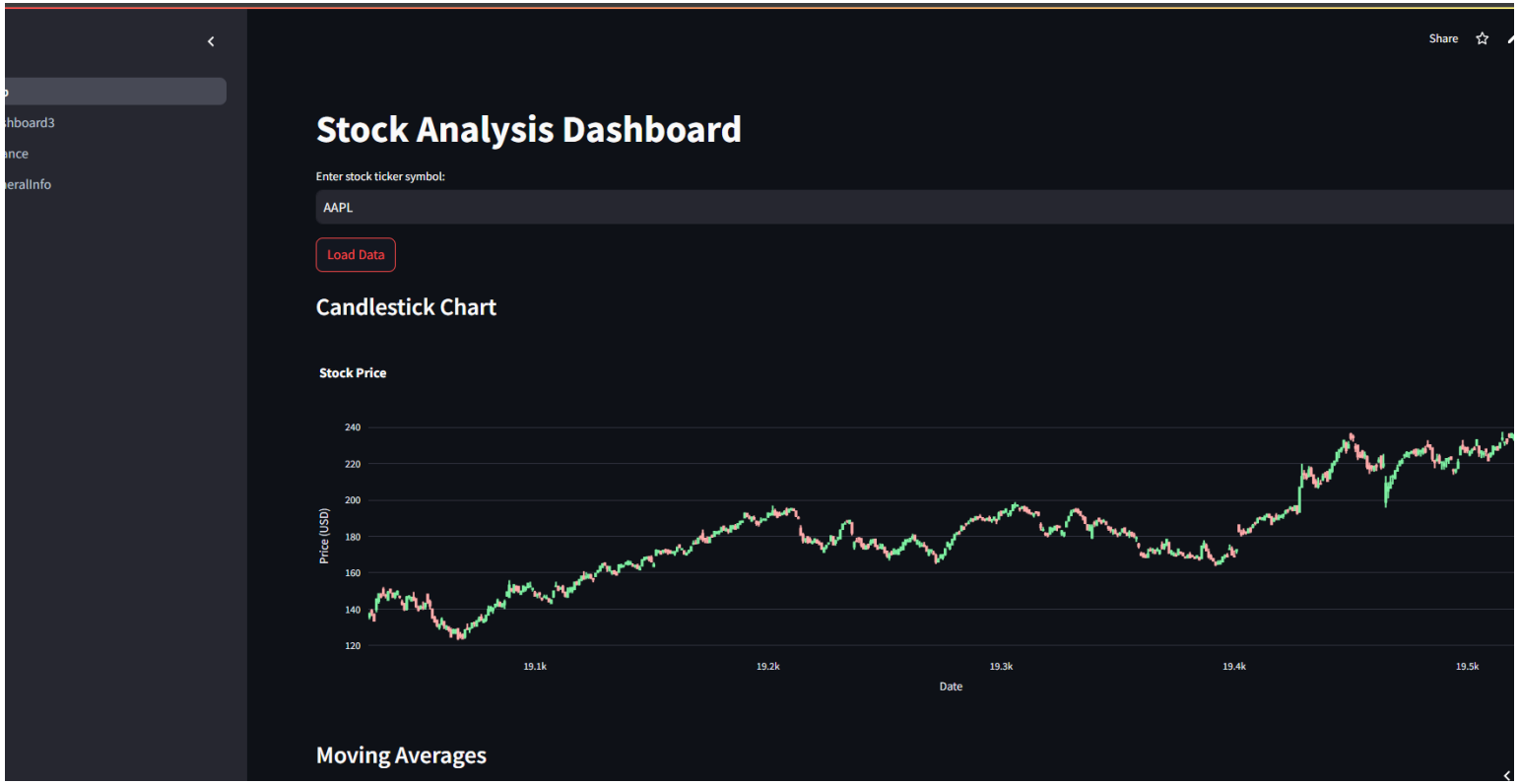
The Upper and Lower Bands are calculated as follows:

$$UB = MB + (m \times \sigma) \; LB = MB - (m \times \sigma)$$

Where:

4. Analytic Dashboards

Utilize Streamlit to host dashboard



5. Machine Learning Model

STEPS:

Test Stationary

Applying different models to see which is the most optimized

Using the models to forecast future stock price

Test Stationary

```python
# Load the data (replace with your own dataset)
# Calculate moving averages for a specific stock, e.g., 'AAPL'
df = pd.read_csv('all_stock_data.csv')
df['Date'] = pd.to_datetime(df['Date'])

# We'll calculate 20-day and 50-day moving averages
apple_data = df[df['Ticker'] == 'AAPL'].copy()
apple_data.set_index('Date', inplace=True)

df = pd.read_csv('all_stock_data.csv')
# For this example, we'll use a simple sine wave with some noise
np.random.seed(0)
dates = pd.date_range(start='2020-01-01', end='2022-12-31', freq='D')
y = np.sin(np.arange(len(dates)) * 2 * np.pi / 365) + np.random.normal(0, 0.1, len(dates))
df = pd.DataFrame({'date': dates, 'value': y})
df.set_index('date', inplace=True)

# Function to check stationarity
def check_stationarity(timeseries):
    result = adfuller(timeseries, autolag='AIC')
    print('ADF Statistic:', result[0])
    print('p-value:', result[1])
    print('Critical Values:', result[4])
    if result[1] <= 0.05:
        print("Strong evidence against the null hypothesis")
        print("Reject the null hypothesis")
        print("Data is stationary")
    else:
        print("Weak evidence against null hypothesis")
        print("Fail to reject the null hypothesis")
        print("Data is non-stationary")

# Check stationarity
check_stationarity(df['value'])
```

Purpose:

•consistency allows for more reliable predictions, as the future behavior of the series is expected to follow similar pattern --> data transformation

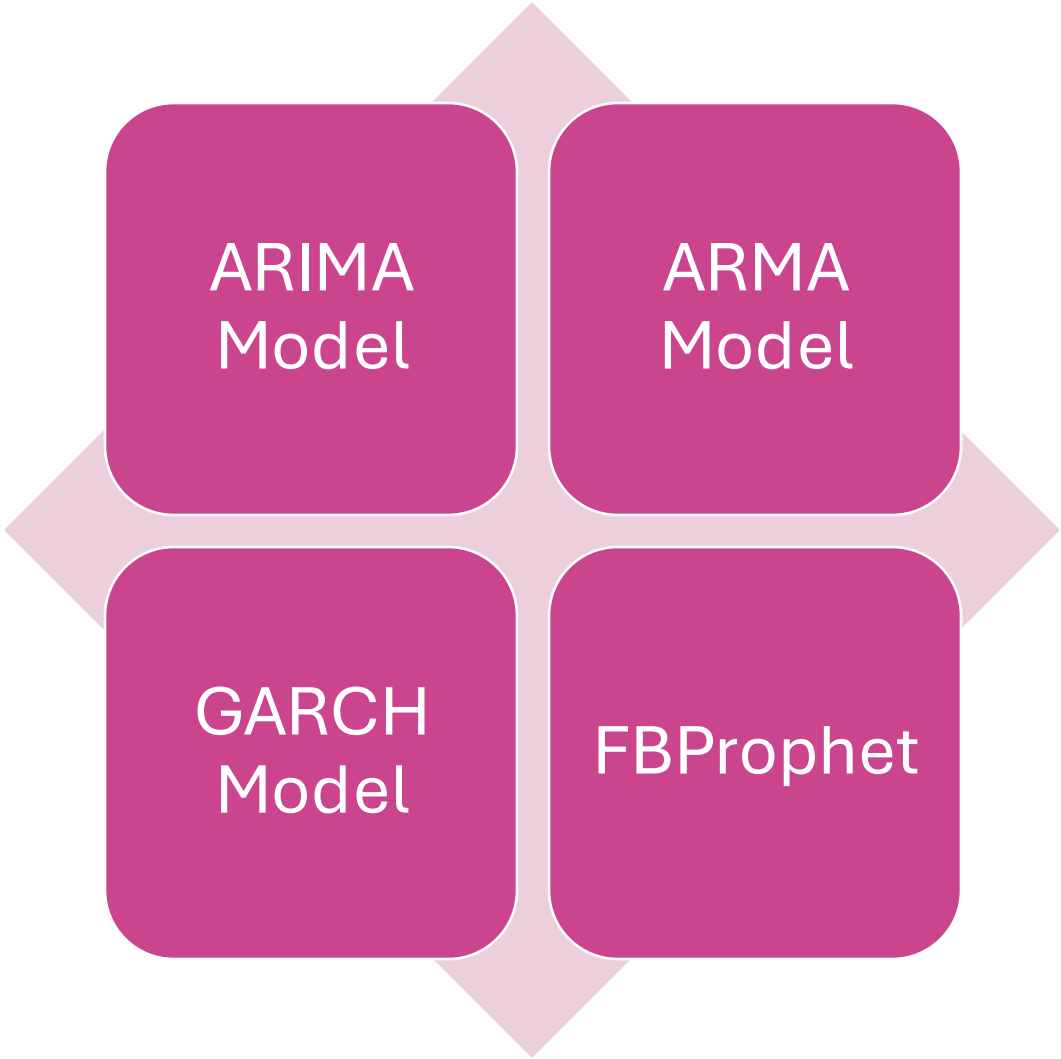
Method:

•Apply statistical tests like the Augmented Dickey-Fuller (ADF) and KPSS tests to assess stationarity.

Post-Test Actions:

•If non-stationary, apply transformations (differencing, detrending) and re-test for stationarity.

- **Models implemented**



THANK YOU FOR LISTENNING!