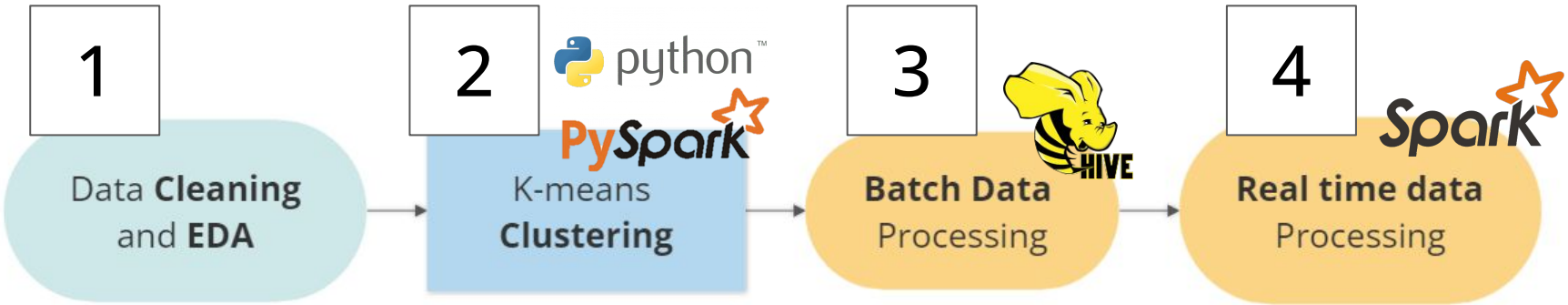
Stock Market Analysis

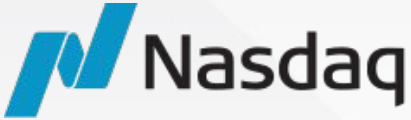
Module 6: Final Presentation

ALY6110, 2023 June 30th

Presented to Professor Behzad Ahmadi
Prepared by Shyamala Venkatakrishnan, Heejae Roh

Work Flow & Index





1 Data Cleaning & EDA

Stock Market Analysis

	A	B	C	D	E	F	G	H
1	Date	Open	High	Low	Close	Adj Close	Volume	Symbol
2	1980-12-12	0.128348	0.128906	0.128348	0.128348	0.099584	469033600	AAPL
3	1980-12-15	0.12221	0.12221	0.121652	0.121652	0.094388	175884800	AAPL
4	1980-12-16	0.113281	0.113281	0.112723	0.112723	0.087461	105728000	AAPL
5	1980-12-17	0.115513	0.116071	0.115513	0.115513	0.089625	86441600	AAPL
6	1980-12-18	0.118862	0.11942	0.118862	0.118862	0.092224	73449600	AAPL
201602	2023-06-02	98.309998	99.660004	98.190002	99.5	99.5	6220300	SBUX
201603	2023-06-05	100.04	100.87	99.5	99.93	99.93	6536500	SBUX
201604	2023-06-06	99.739998	100.33	97.529999	98.220001	98.220001	8657100	SBUX
201605	2023-06-07	98.550003	98.860001	97.589996	97.919998	97.919998	6285600	SBUX
201606	2023-06-08	97.650002	99.18	96.919998	99.150002	99.150002	6877800	SBUX

Stock data of Top 24 Nasdaq companies (by Market Cap, June 8th)
Num rows: 201,606 and Num columns: 8
Dataset describe: from Yahoo Finance, Top 24 Market Cap, Max period of each stock
Fields explanation: Date/ Prices (open, high, low, close, adj close)/ Volume/ Symbol (added)

Cleaning the Data

Type of Data & isnull				
#	Column	Non-Null Count	Dtype	isnull
0	Date	201611 non-null	object	0
1	Open	201611 non-null	float64	0
2	High	201611 non-null	float64	0
3	Low	201611 non-null	float64	0
4	Close	201611 non-null	float64	0
5	Adj Close	201611 non-null	float64	0
6	Volume	201611 non-null	int64	0
7	Symbol	201611 non-null	object	0

Drop: Adj Close (Adjusted Close), Checking NA values
After Drop: 201,606 Num of rows and 7 Columns

Descriptive Analysis of 24 stocks

	Open	High	Low	Close	Volume	Year
count	201611	201611	201611	201611	201611	201611
mean	47.75	48.44	47.23	47.86	4.27e+07	2003.7
std	81.85	82.80	80.74	81.80	1.15e+08	13.08
min	0.00	0.05	0.05	0.05	0.00	1962.00
25%	3.84	4.34	4.18	4.26	3.14	1994.00
50%	18.95	19.25	18.67	18.95	1.01	2006.00
75%	51.70	52.31	51.00	51.70	3.82	2015.00
max	835.12	921.78	798.80	812.73	7.42	2023.00

2

1

Min year: 1962 (HON) Mean price: 47.86

Focused Analysis on AAPL

	Open	High	Low	Close	Volume	Year
count	10718	10718	10718	10718	10718	10718
mean	17.99	18.20	17.80	18.00	3.24e+08	2001.68
std	37.90	38.36	37.49	37.94	3.37e+08	12.28
min	0.05	0.05	0.05	0.05	0.00e+00	1980
25%	0.29	0.30	0.28	0.29	1.18e+08	1991
50%	0.50	0.50	0.49	0.50	2.12e+08	2002
75%	17.80	17.94	17.61	17.80	4.04e+08	2012
max	186.73	186.99	194.27	186.01	7.42e+09	2023

1

2

3

4

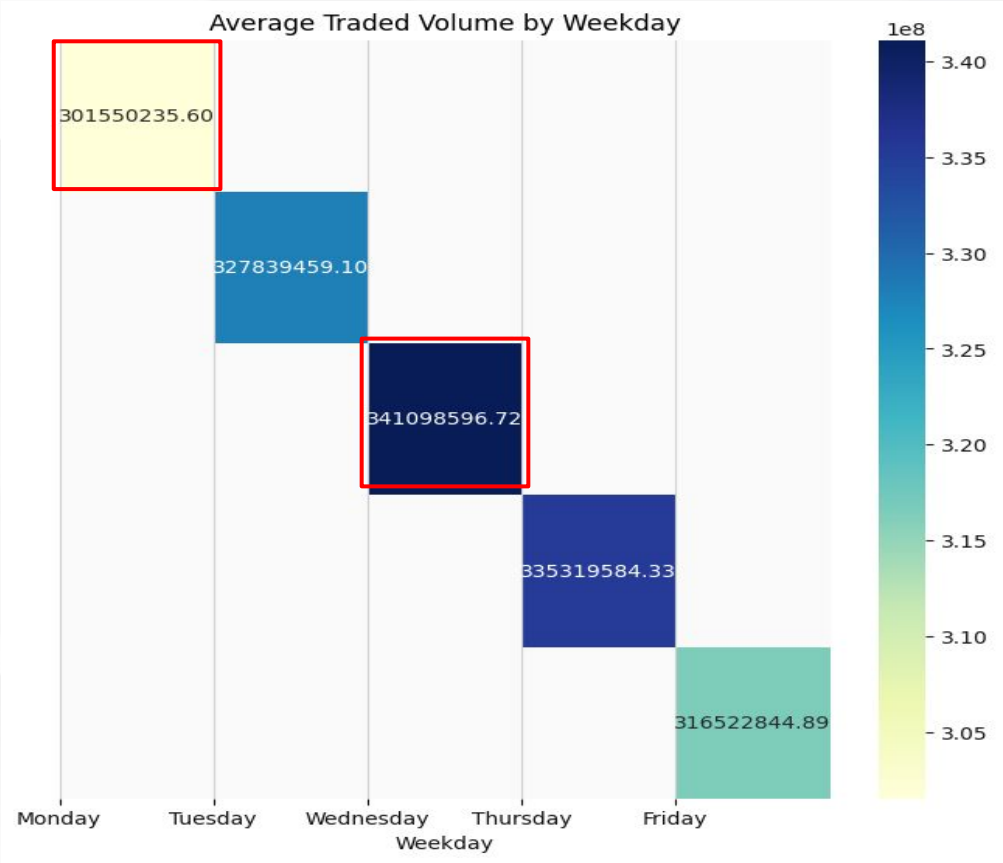
Min year: 1980(APPL)
Mean price: 18.00

Max year: 2023
Max Price: 186.73

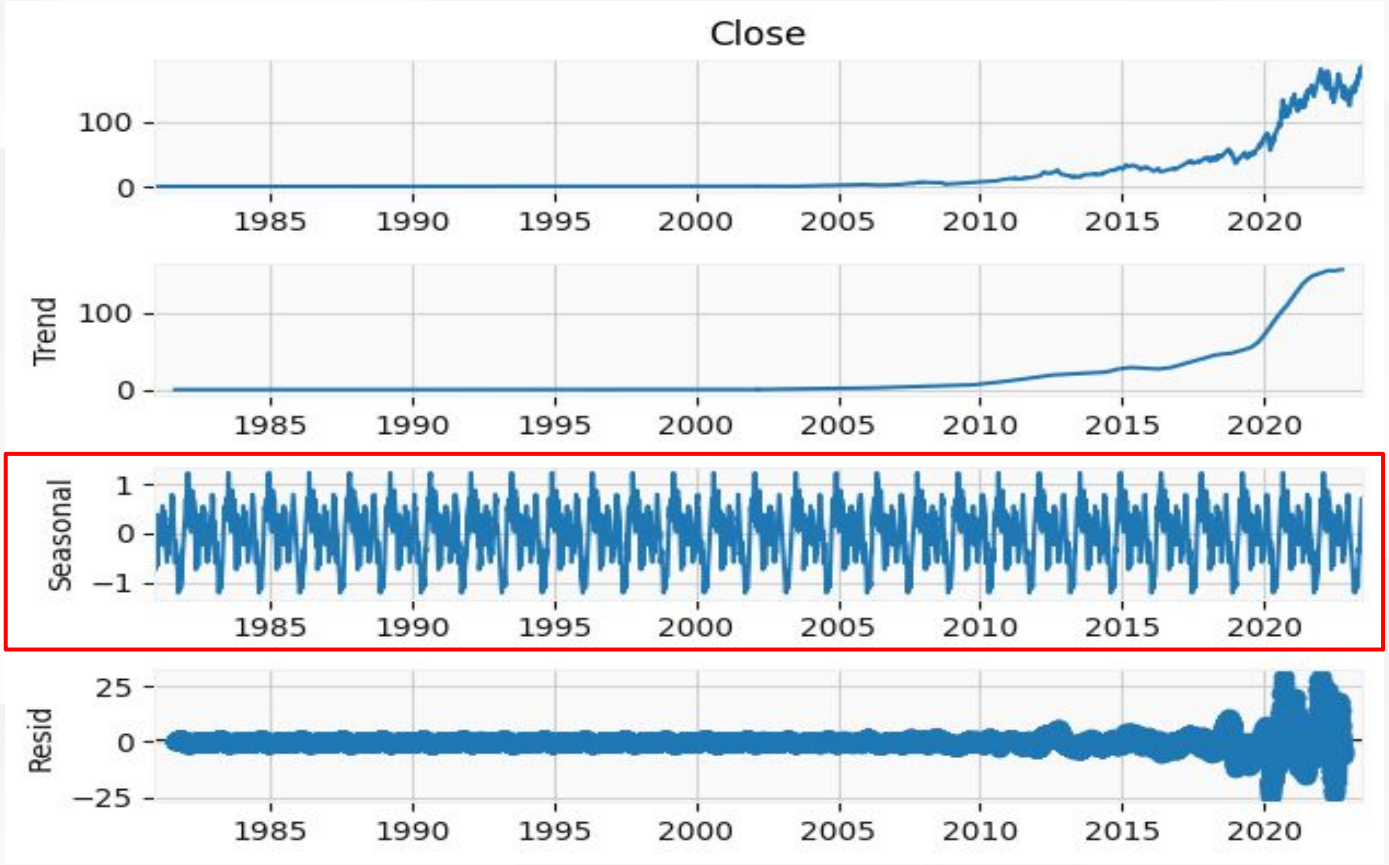
Volume min: 0.00
Volume max: 7.42e+09

Candle chart of AAPL for one month

Average Traded Volume by Weekday



Decomposing close price of Apple



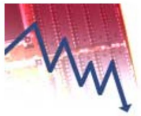
AAPL Daily Returns

CNN Money

ADVERTISEMENT

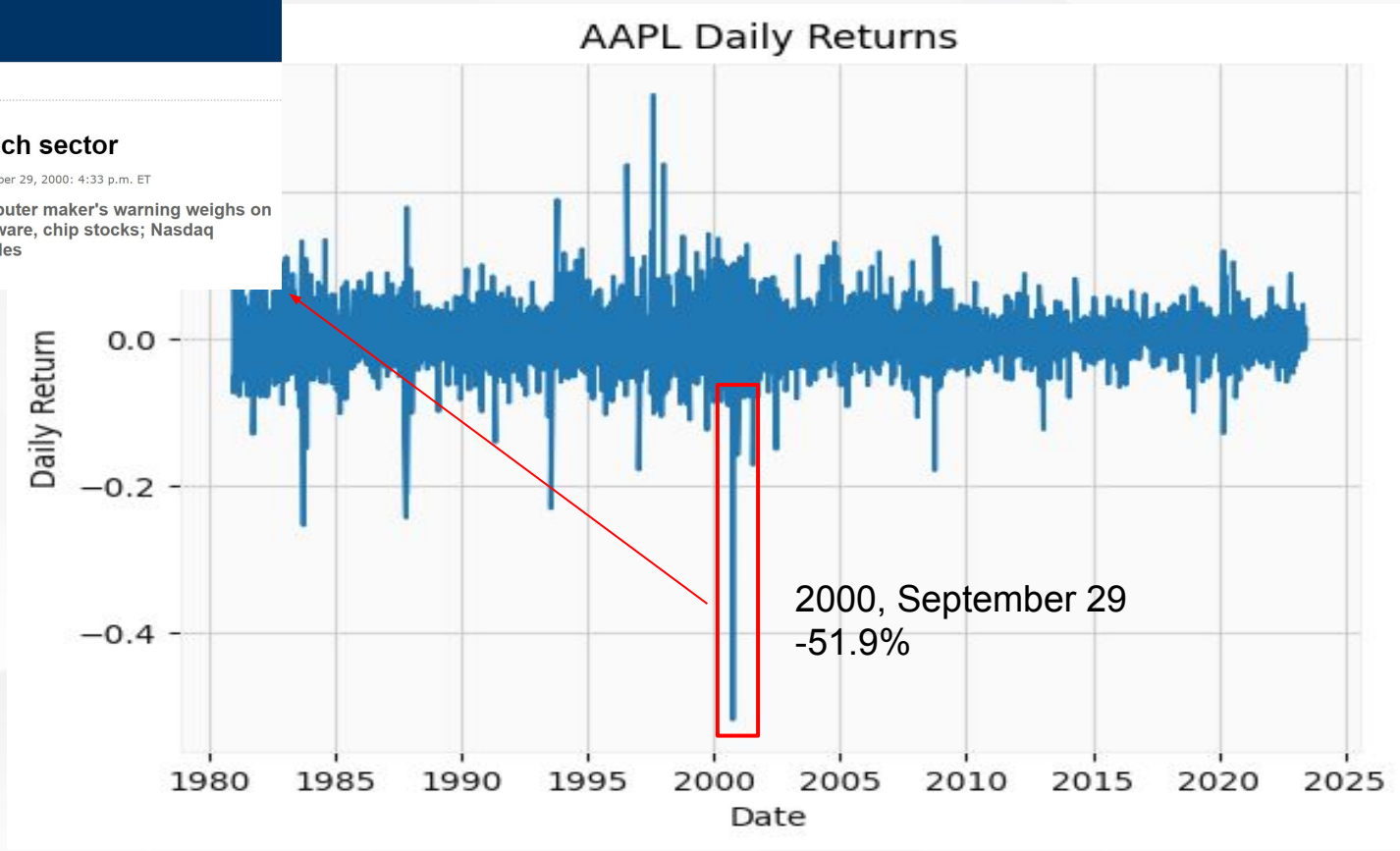
Markets & Stocks

Apple bruises tech sector

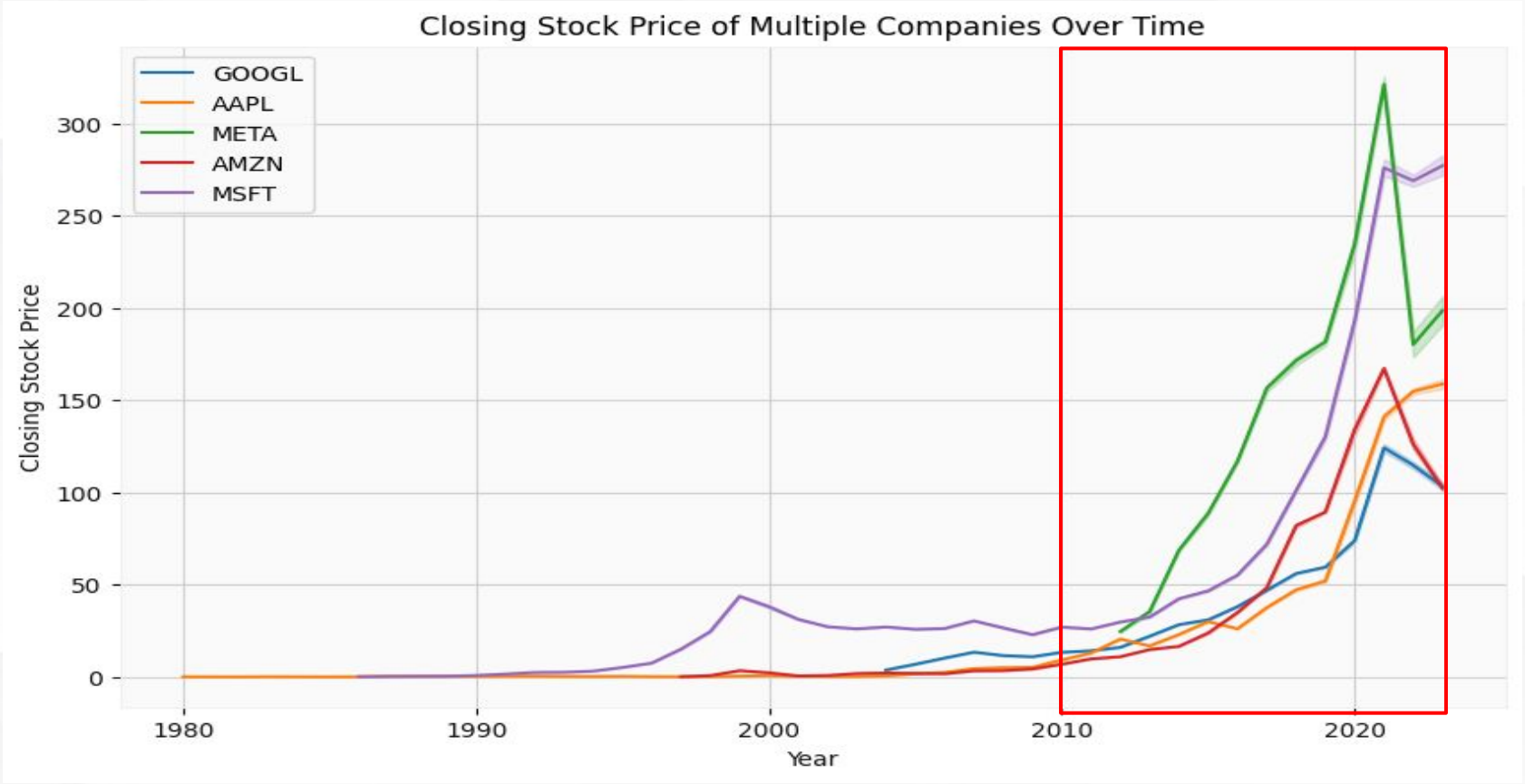


September 29, 2000: 4:33 p.m. ET

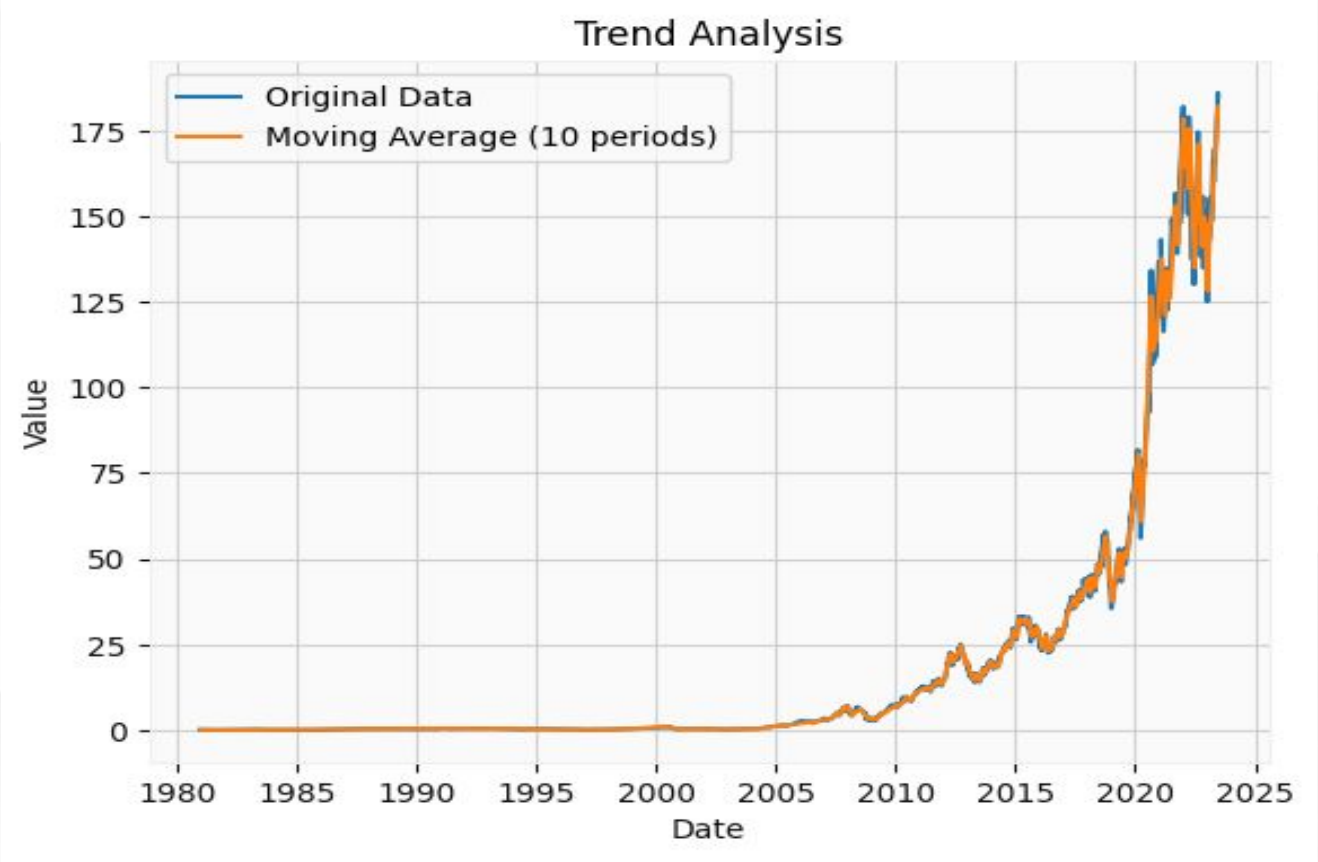
Computer maker's warning weighs on hardware, chip stocks; Nasdaq tumbles



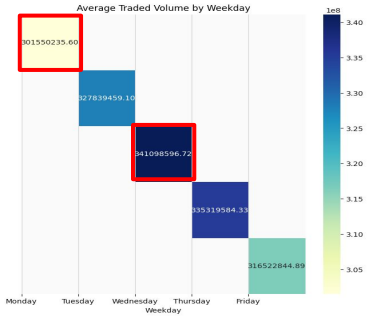
Closing Stock Price of 5 Companies

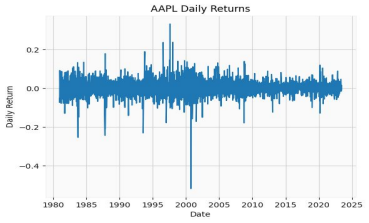


Moving Average (10 periods)



Summary of insights from EDA



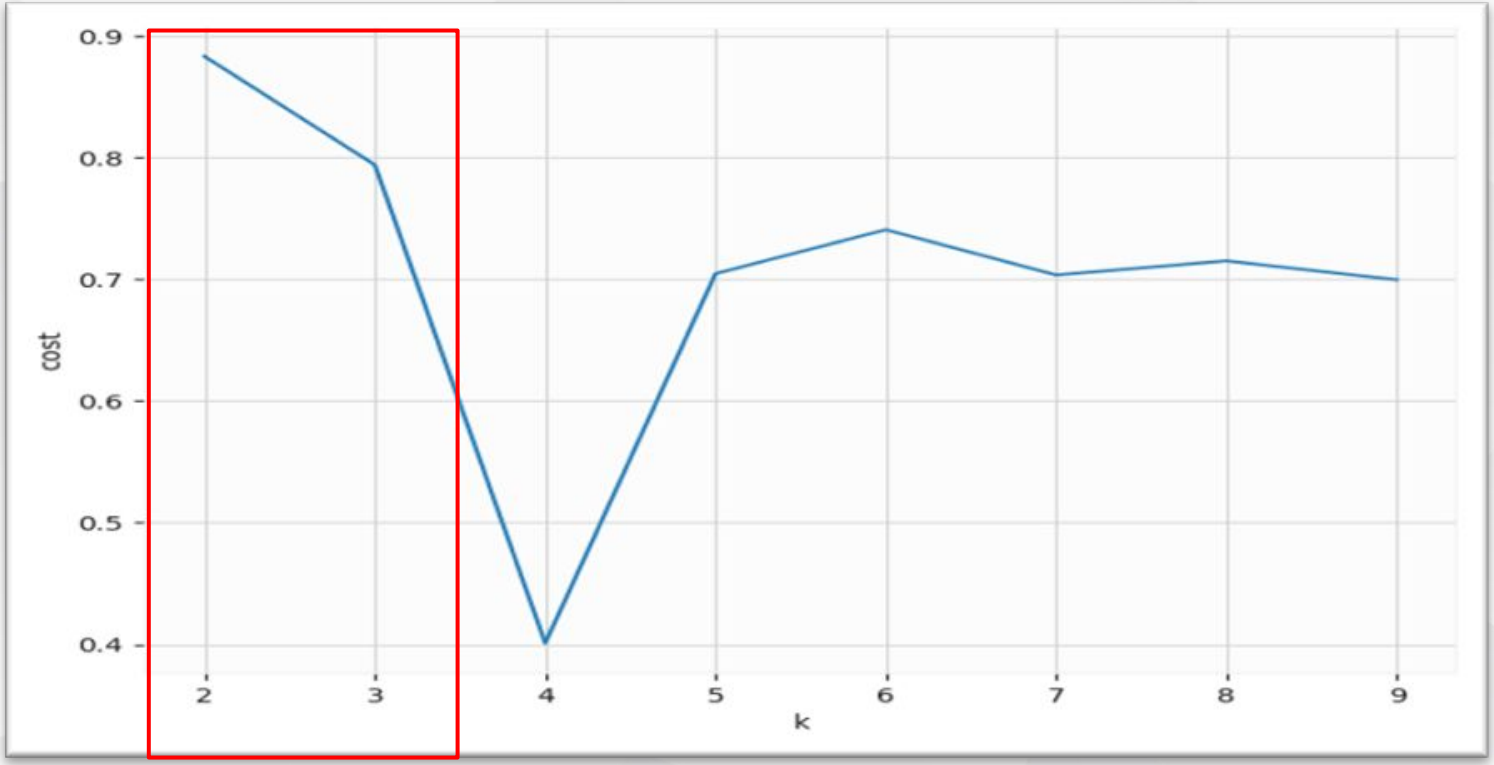


- Close Price over time & one year & one month
  1. AAPL have risen sharply since 2010 and did not exceed $20 until 2010.
  2. Within one year, affected monetary policy and interest rates
  3. Within one month, affected big issue which is Apple Vision pro announcement
- Average Trade Volume by Weekday
  - Lowest in Monday and Highest in Wednesday according to News and Events
- Decomposing close price trend, seasonal, and residual
  - a change or variations is uniform for seasonal component
- 5 Companies in one graph
  - Confirm when the company started to rise
- AAPL Daily Returns
  - Checking lowest return of Day in 2000, September
- Correlation between Stock Close price
  - Most Companies generally have positive correlations, because they are most successful companies by market cap
  - AAPL & GOOGL, and GOOGL & AMZN have highest correlations

2

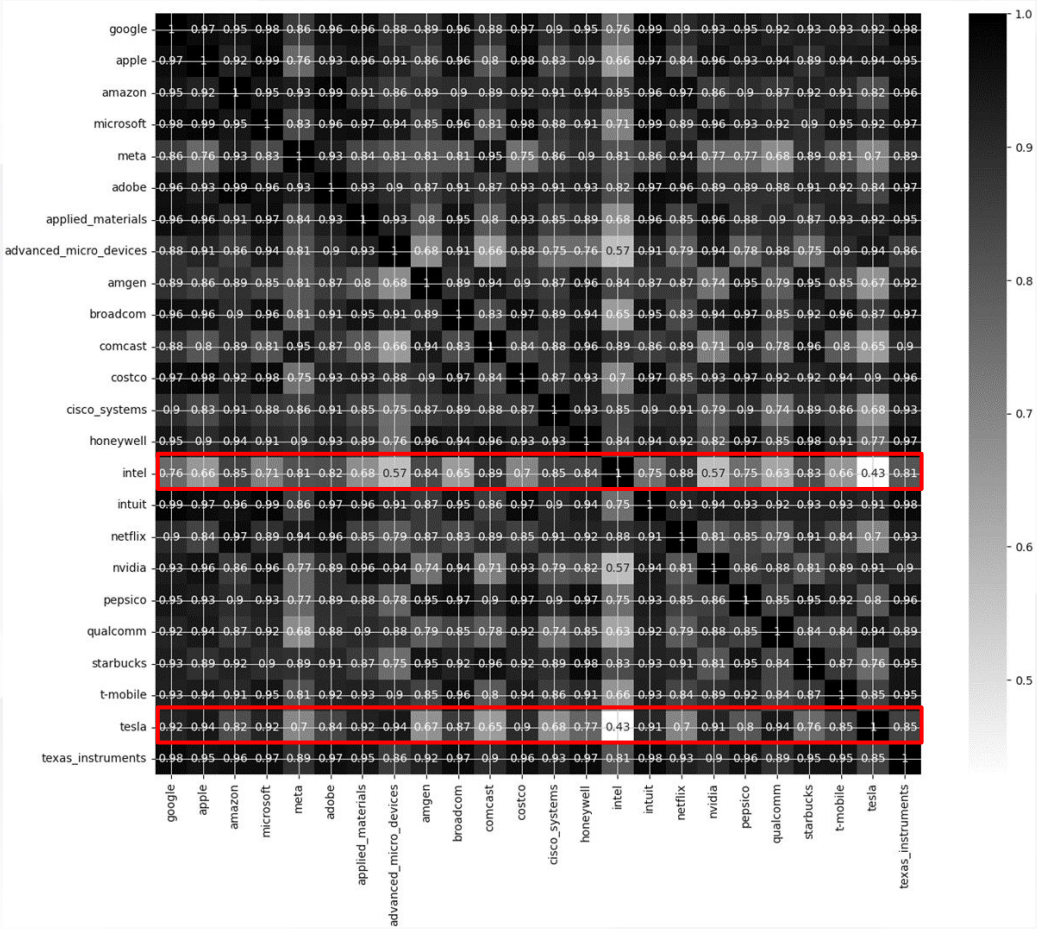
Diversified Portfolio with K-means Clustering

Silhouette Score

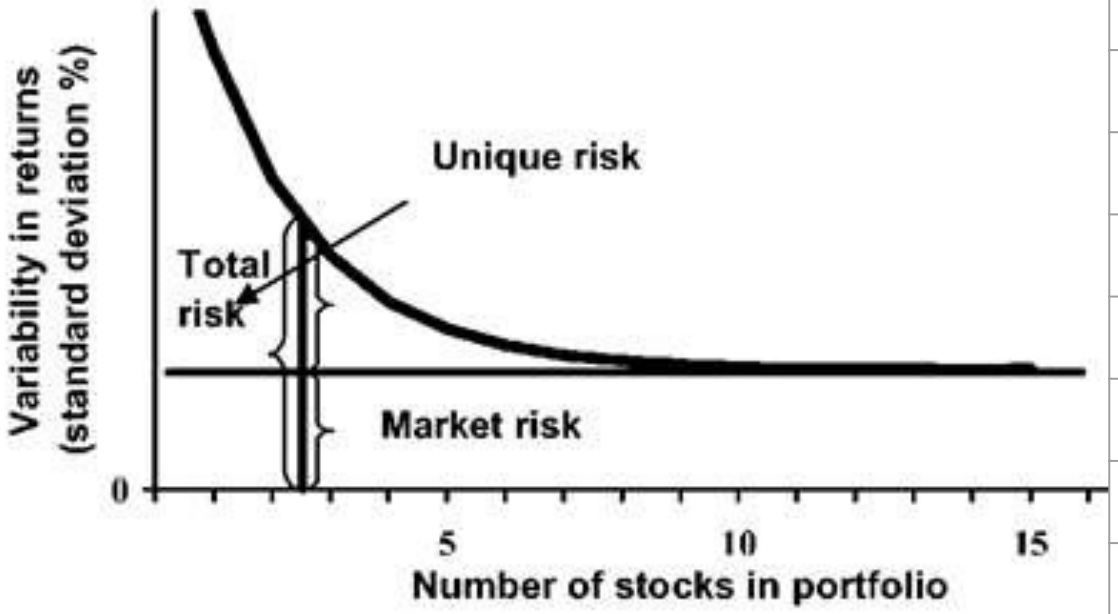


Silhouette Coefficient or silhouette score is close to 1, meaning that the cluster is well divided. This is extracted using pyspark.

Correlation Matrix of 24 companies' close price

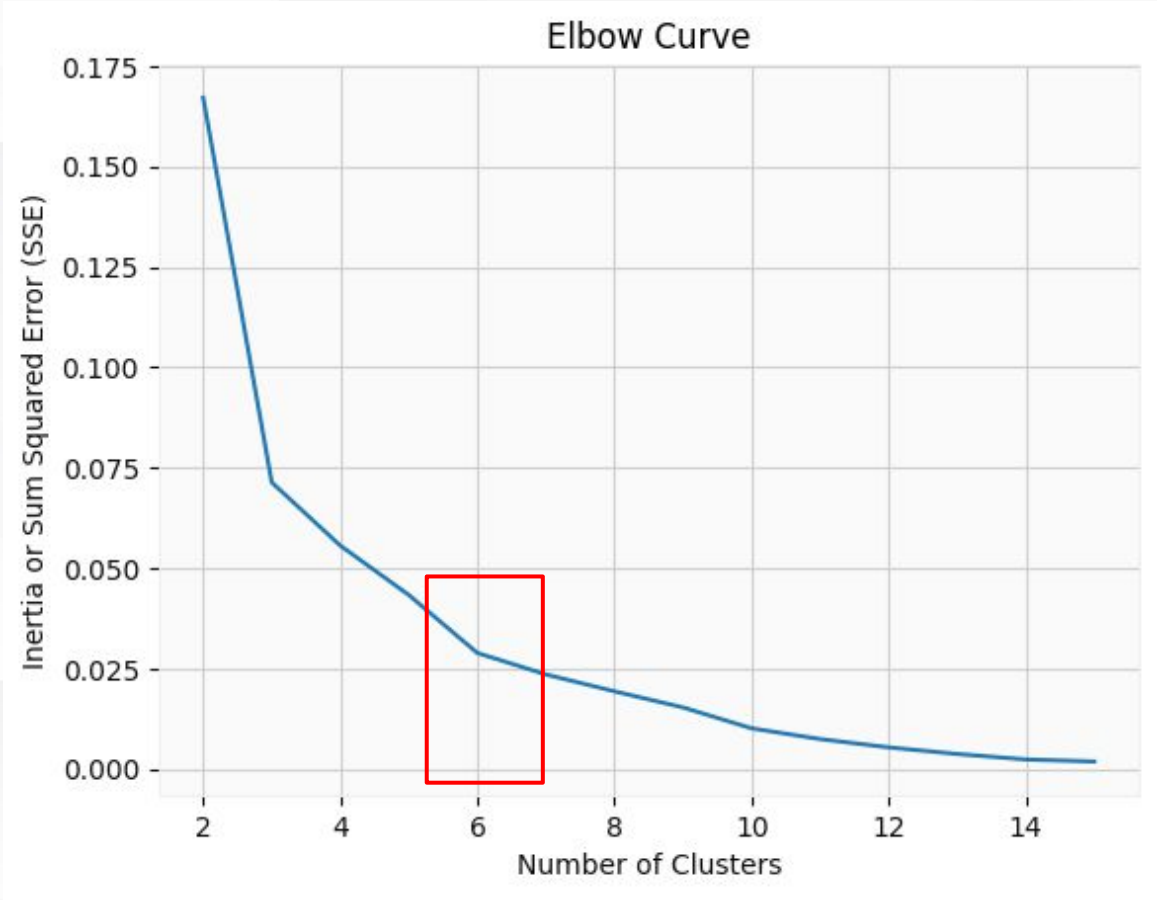


Return and Variance of Return

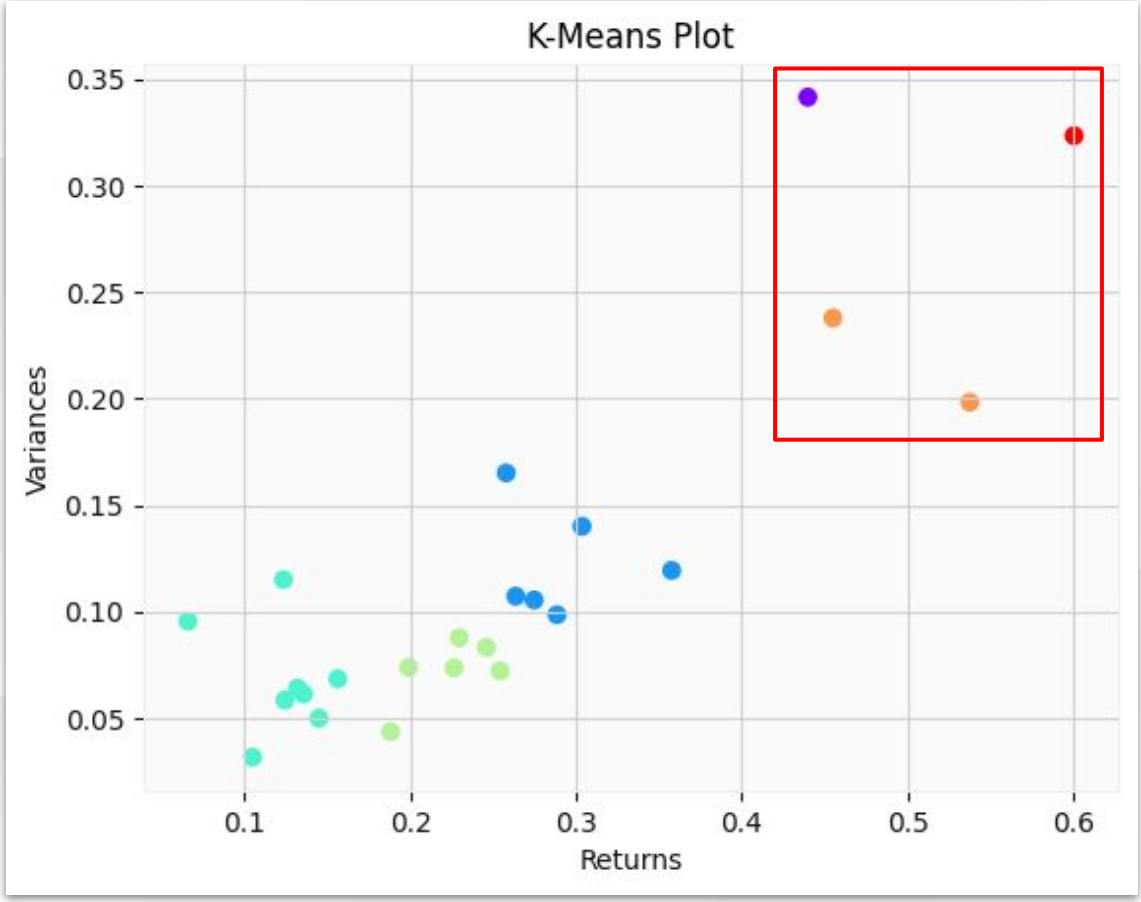


Company_name	Variances	Returns
google	0.0736	0.2265
apple	0.0831	0.2460
amazon	0.1054	0.2749
microsoft	0.0721	0.2543
meta	0.1650	0.2580
adobe	0.0984	0.2886
applied_materials	0.1400	0.3037
amd	0.3415	0.4399

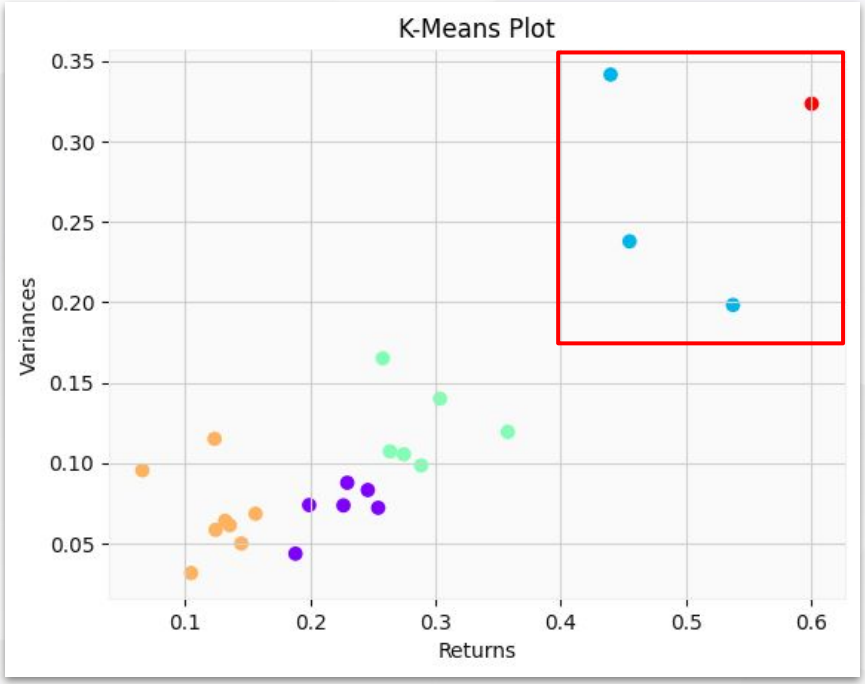
Elbow Curve for Clustering



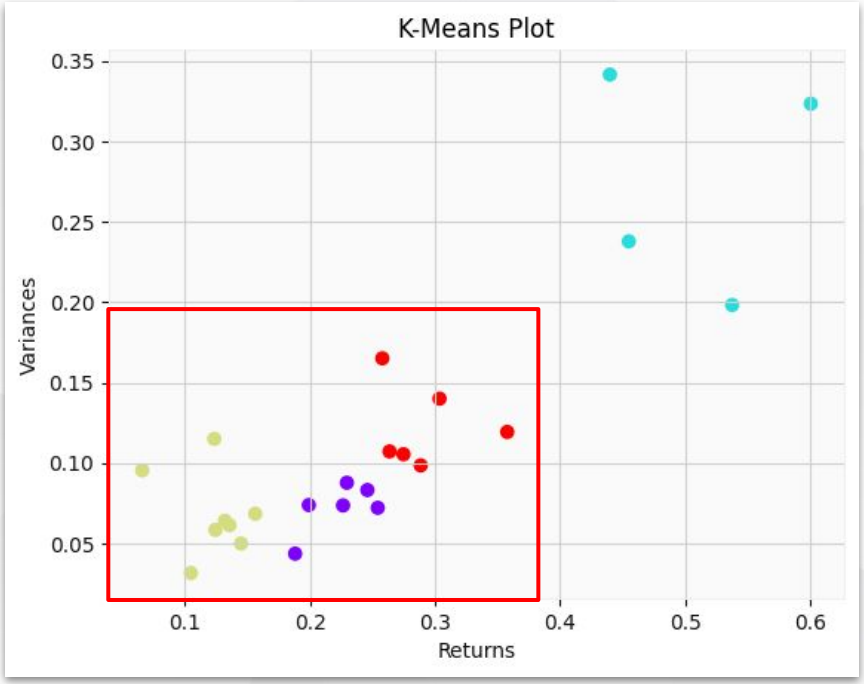
Cluster number 6



Cluster number 5 & 4

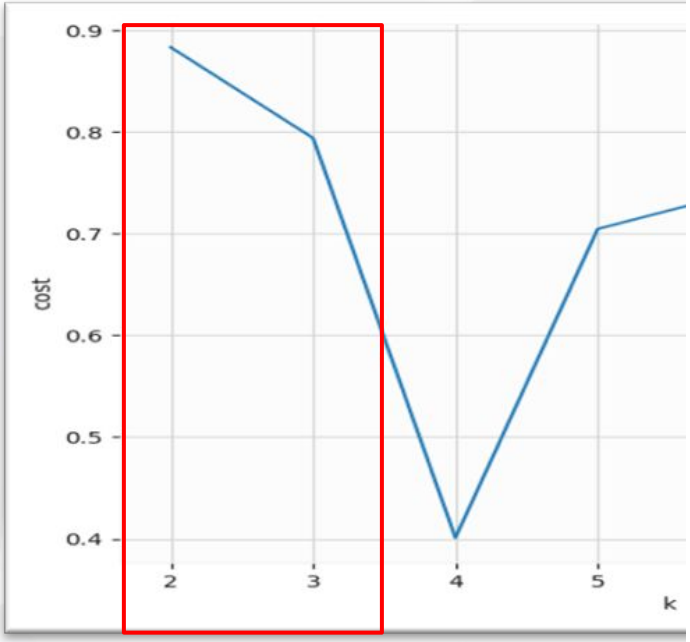


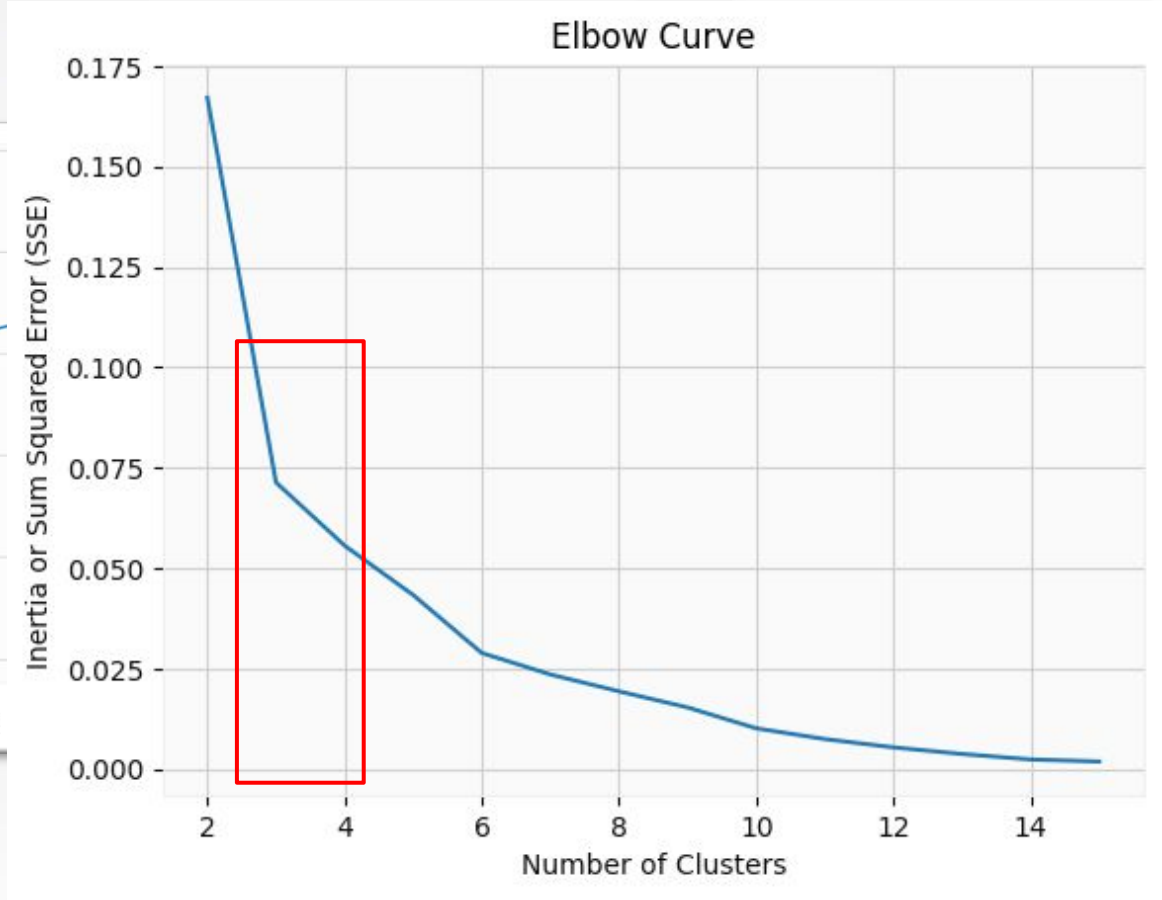
N = 5



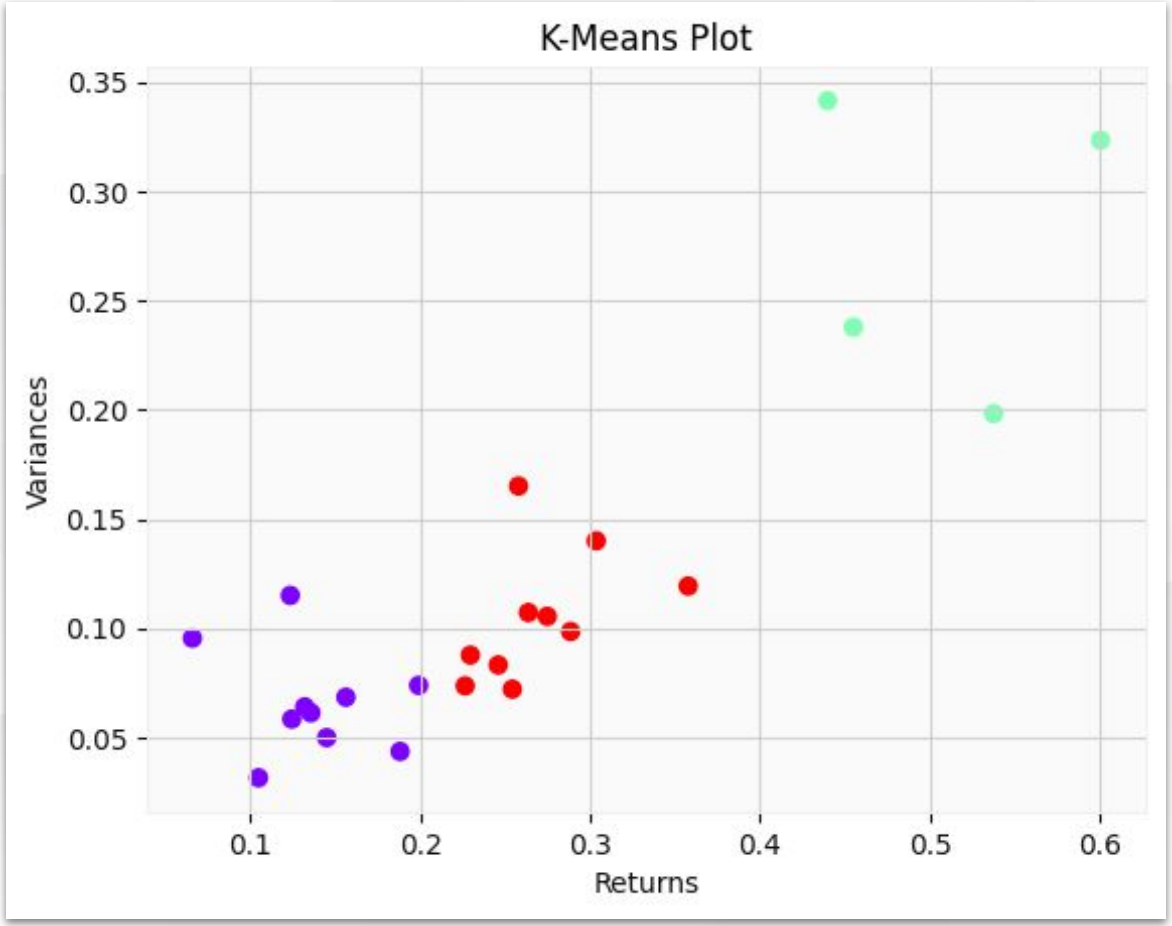
N = 4

Elbow Curve and Silhouette Score





Finally, Cluster number 3

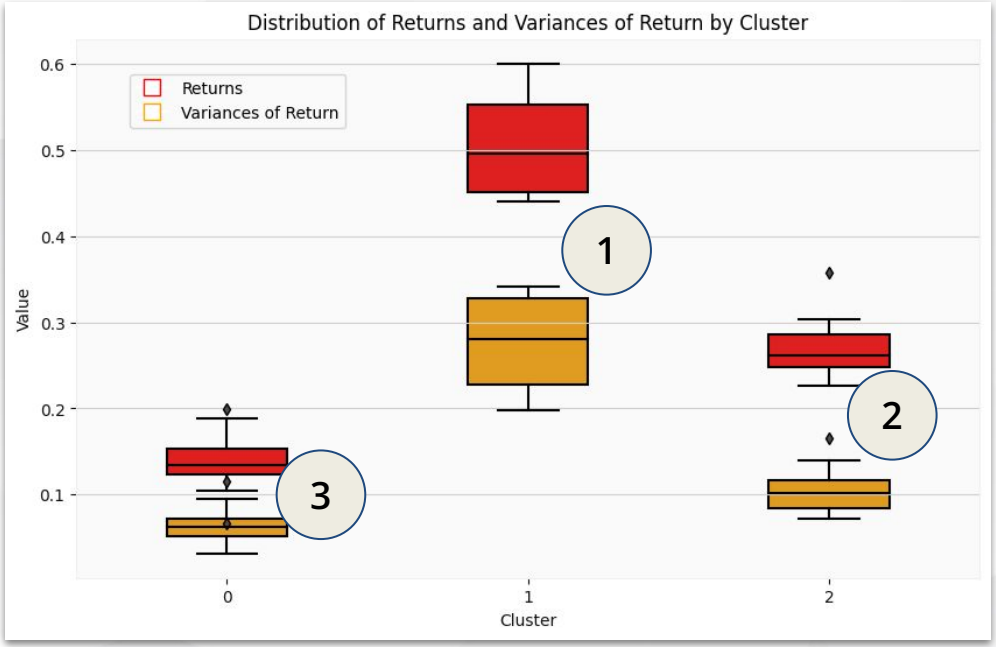


Finally, Cluster number 3

Company_name	Variances	Returns	Cluster
google	0.0736	0.2265	**2**
apple	0.0831	0.2460	**2**
amazon	0.1054	0.2749	**2**
microsoft	0.0721	0.2543	**2**
meta	0.1650	0.2580	**2**
adobe	0.0984	0.2886	**2**
applied_materials	0.1400	0.3037	**2**
amd	0.3415	0.4399	**1**
amgen	0.0612	0.1357	**0**

Company_name	Variances	Returns	Cluster
intuit	0.0876	0.2296	**2**
netfilx	0.2378	0.4551	**1**
nvidia	0.1982	0.5377	**1**
pepsico	0.0315	0.1049	**0**
qualcomm	0.1150	0.1236	**0**
starbucks	0.0684	0.1563	**0**
t-mobile	0.1071	0.2637	**2**
tesla	0.3234	0.6006	**1**
texas_instrument	0.0738	0.1992	**0**

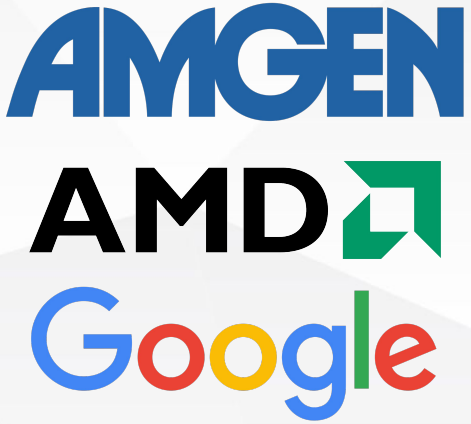
Visualization of Cluster



Avg_Variance	0.0661	0.2752	0.1052
Avg_Return	0.1375	0.5083	0.2703
Cluster	0	1	2

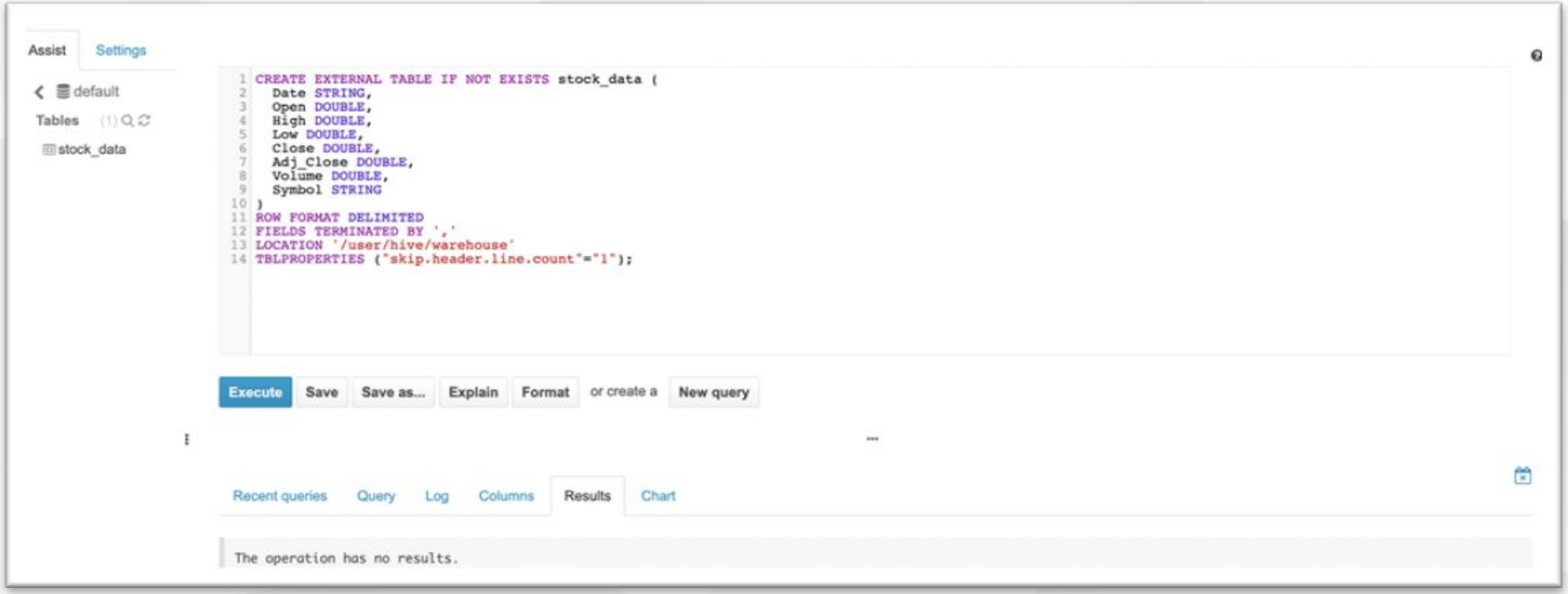
Diversified Portfolio with Cluster number = 3

index	Company_name	Cluster_Labels
8	amgen	0
0	advanced_micro_devices	1
3	google	2



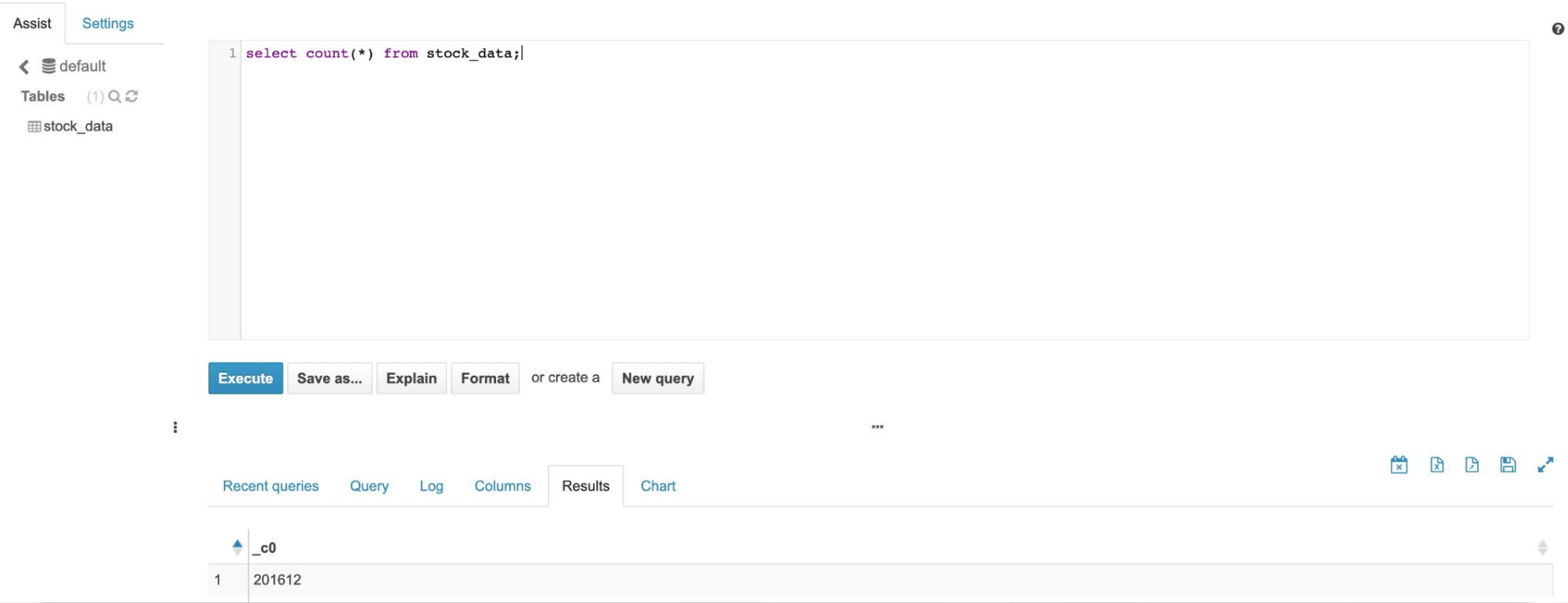
3 Batch Data Processing

Creation of HIVE table and loading data



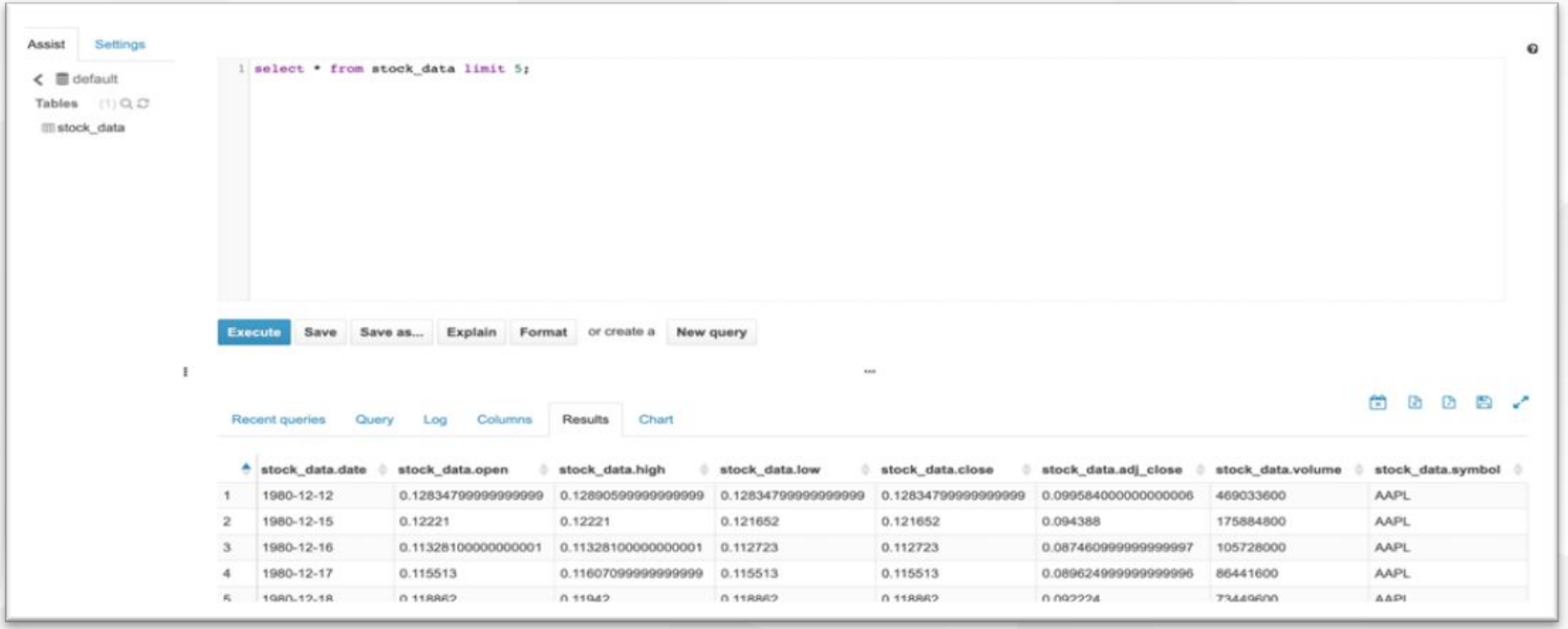
Initially, the stock market csv file was imported into HDFS from the local machine using Hive File browser. An external HIVE table was created using the file location and there are a total of 201,612 records in the table.

Creation of HIVE table and loading data



Data is Uploaded

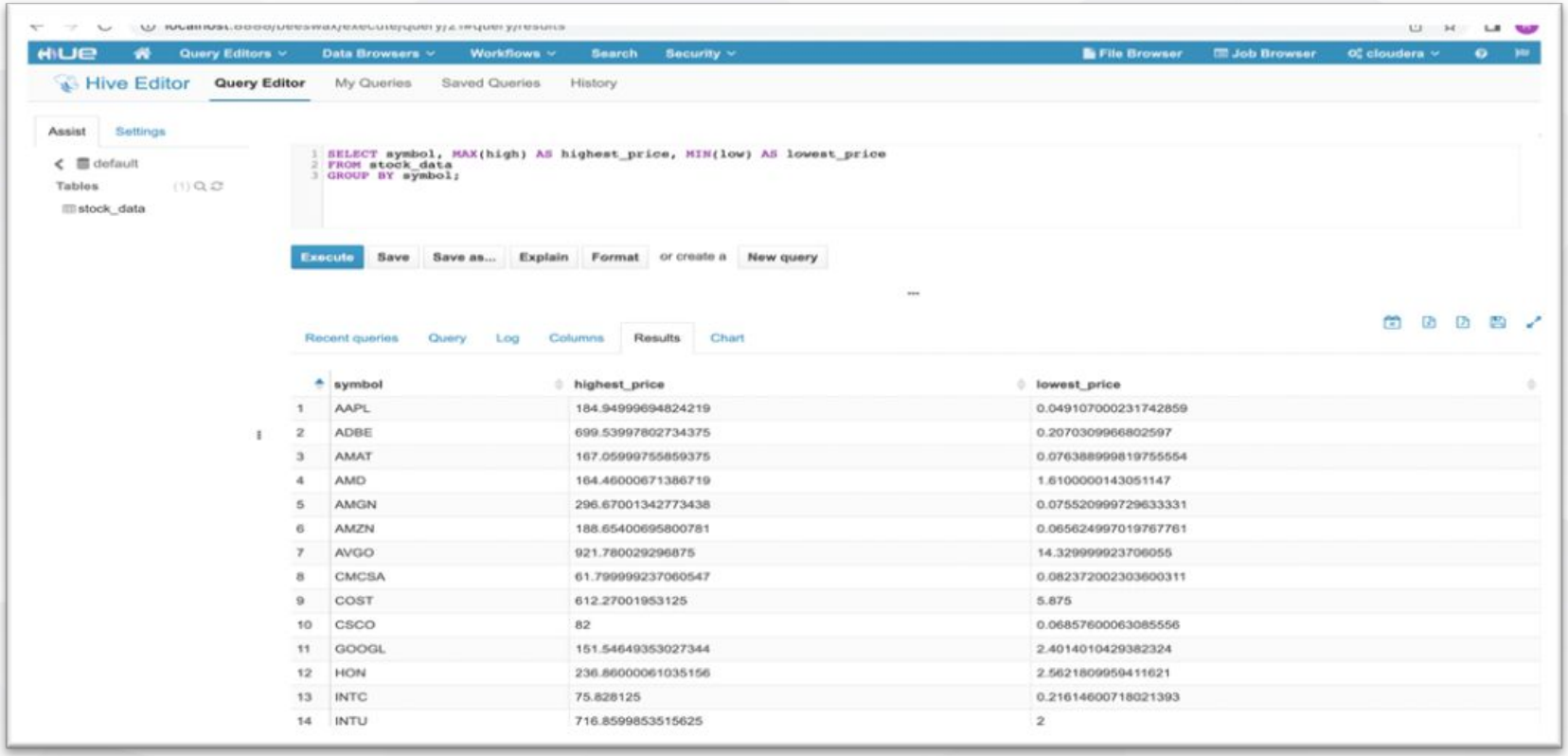
Creation of HIVE table and loading data



	stock_data.date	stock_data.open	stock_data.high	stock_data.low	stock_data.close	stock_data.adj_close	stock_data.volume	stock_data.symbol
1	1980-12-12	0.12834799999999999	0.12890599999999999	0.12834799999999999	0.12834799999999999	0.09958400000000006	469033600	AAPL
2	1980-12-15	0.12221	0.12221	0.121652	0.121652	0.094388	175884800	AAPL
3	1980-12-16	0.11328100000000001	0.11328100000000001	0.112723	0.112723	0.08746099999999997	105728000	AAPL
4	1980-12-17	0.115513	0.11607099999999999	0.115513	0.115513	0.08962499999999996	86441600	AAPL
5	1980-12-18	0.118862	0.11942	0.118862	0.118862	0.092224	73449600	AAPL

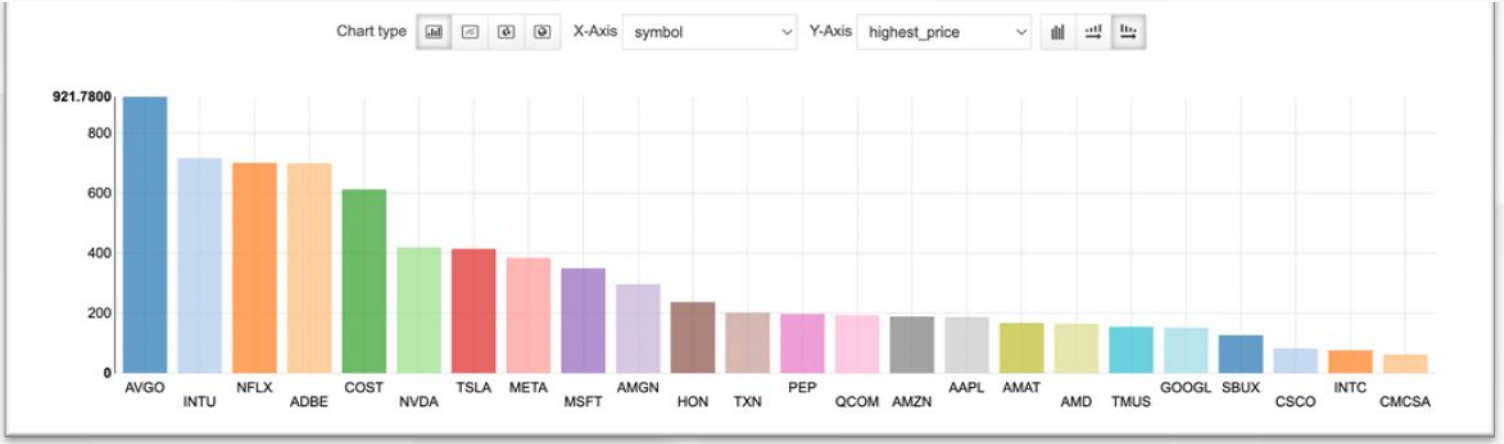
Data is Uploaded

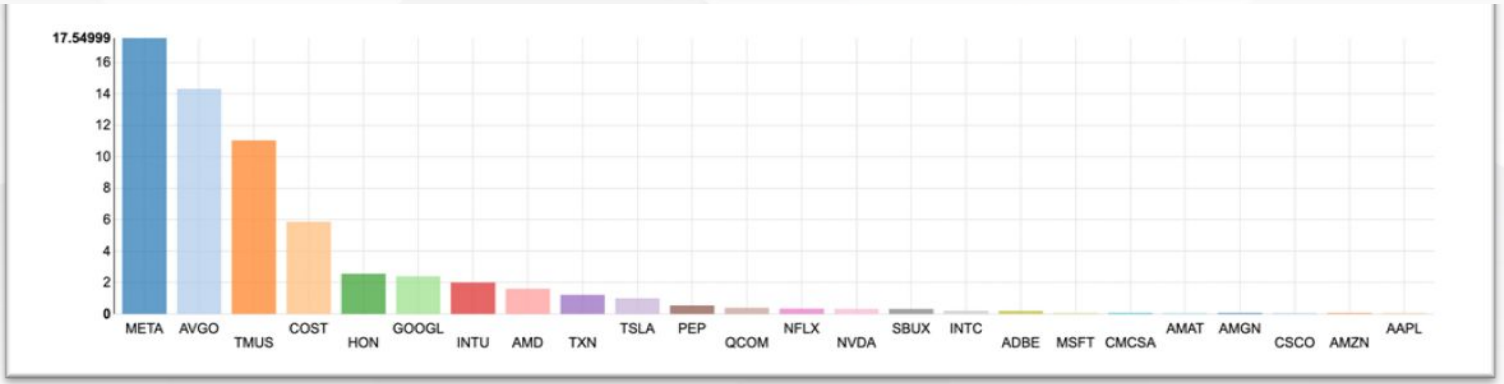
1. What is the highest and lowest price for each of the stock symbols?



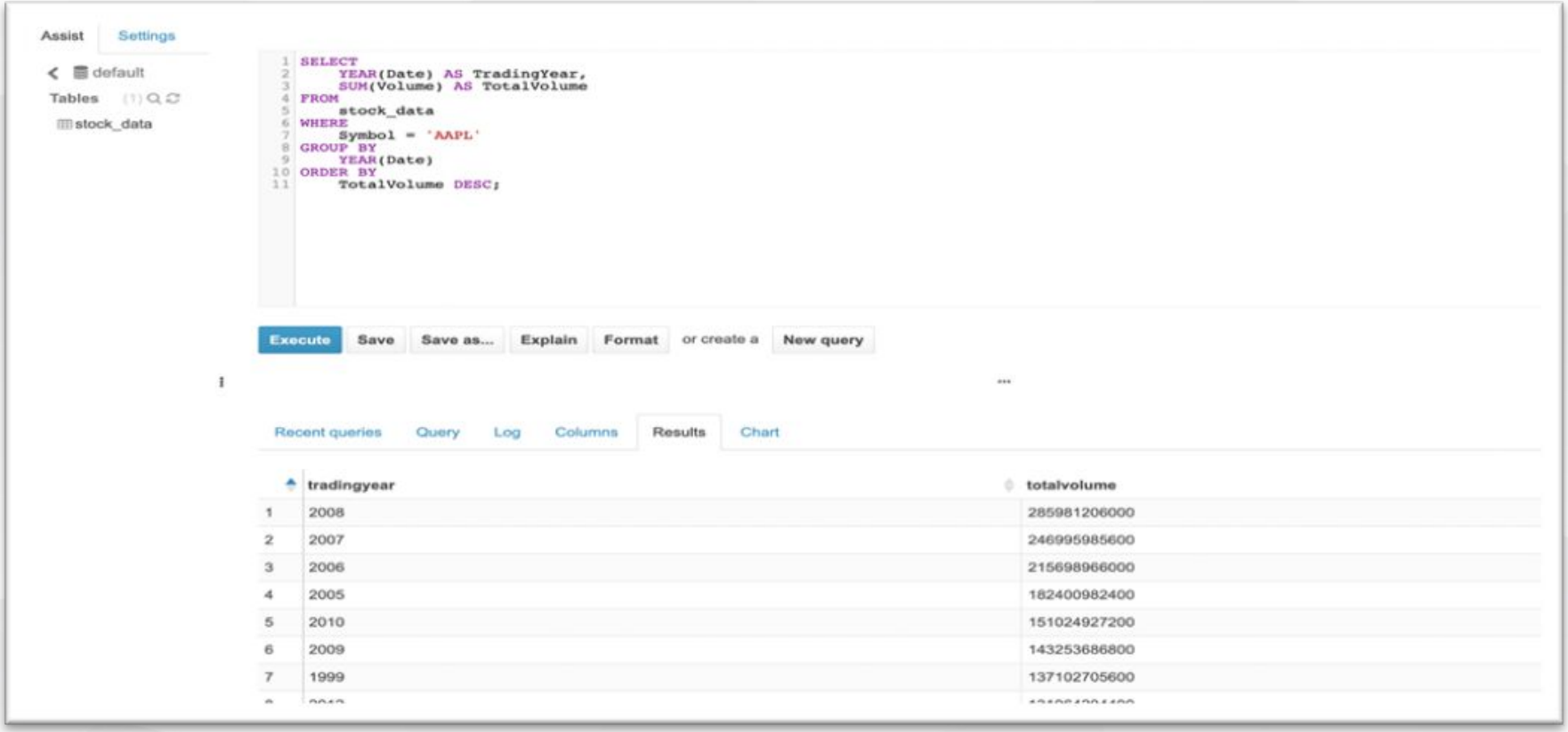
	symbol	highest_price	lowest_price
1	AAPL	184.94999694824219	0.049107000231742859
2	ADBE	699.53997802734375	0.2070309966802597
3	AMAT	167.05999755859375	0.076388999819755554
4	AMD	164.46000671386719	1.6100000143051147
5	AMGN	296.67001342773438	0.075520999729633331
6	AMZN	188.65400695800781	0.065624997019767761
7	AVGO	921.780029296875	14.329999923706055
8	CMCSA	61.799999237060547	0.082372002303600311
9	COST	612.27001953125	5.875
10	CSCO	82	0.06857600063085556
11	GOOGL	151.54649353027344	2.4014010429382324
12	HON	236.86000061035156	2.5621809959411621
13	INTC	75.828125	0.21614600718021393
14	INTU	716.8599853515625	2

1. What is the highest and lowest price for each of the stock symbols?

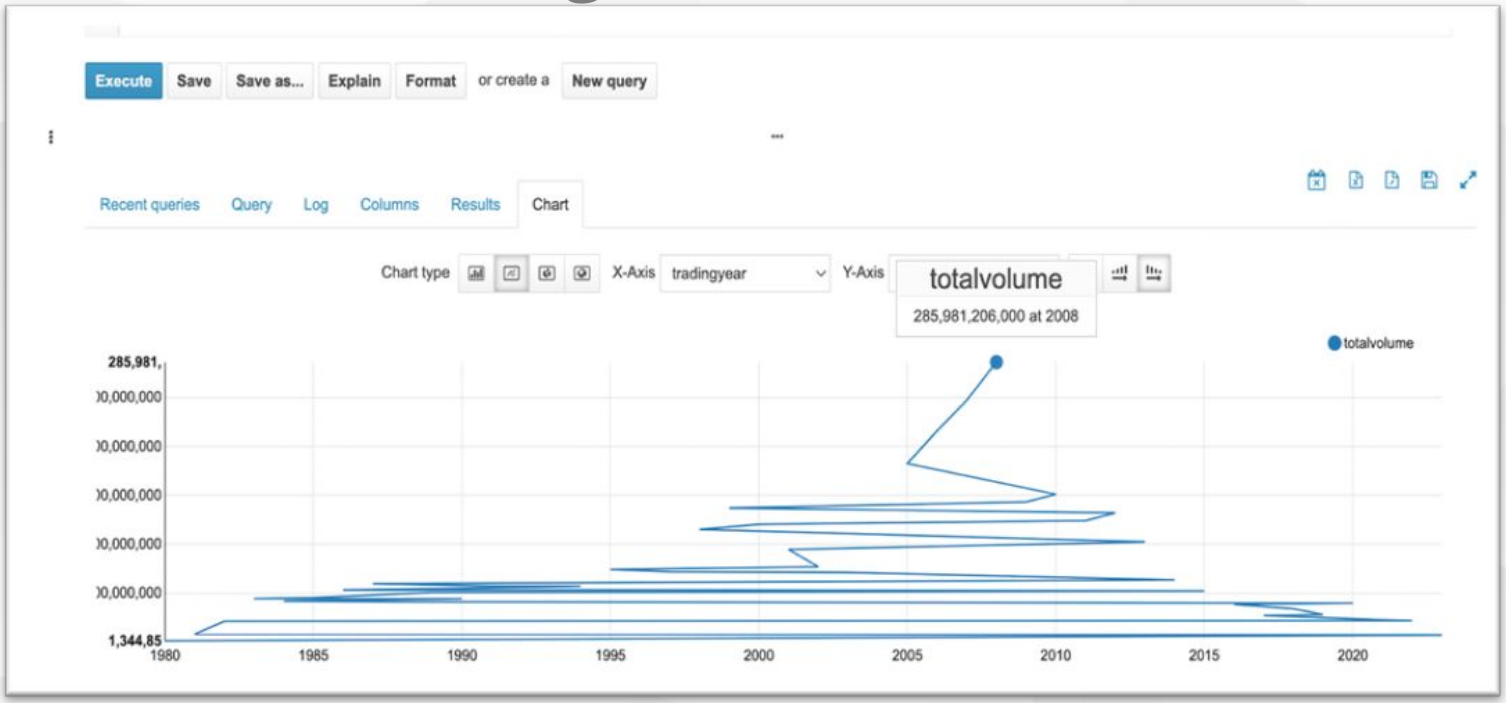




2. Which year has the greatest volume of trading for AAPL stock?

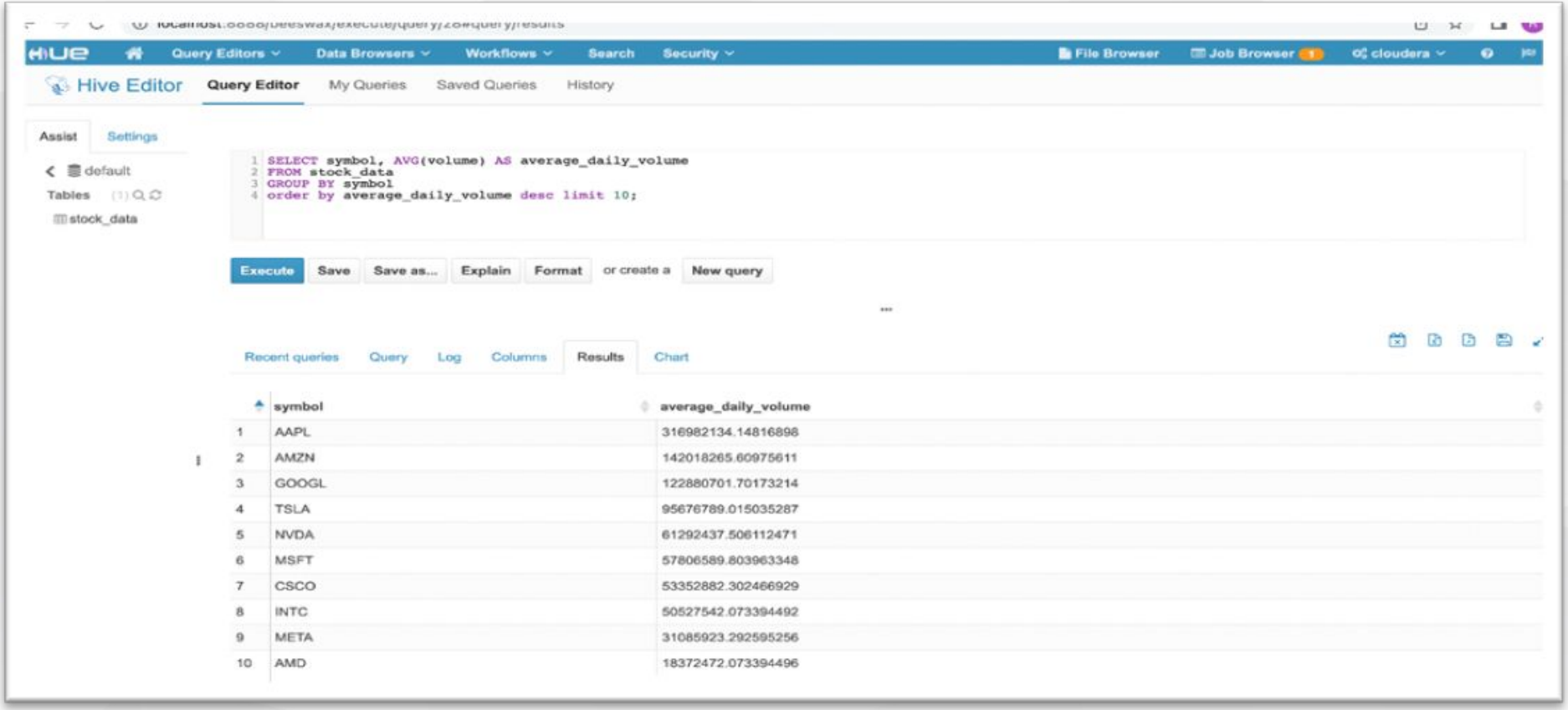


2. Which year has the greatest volume of trading for AAPL stock?



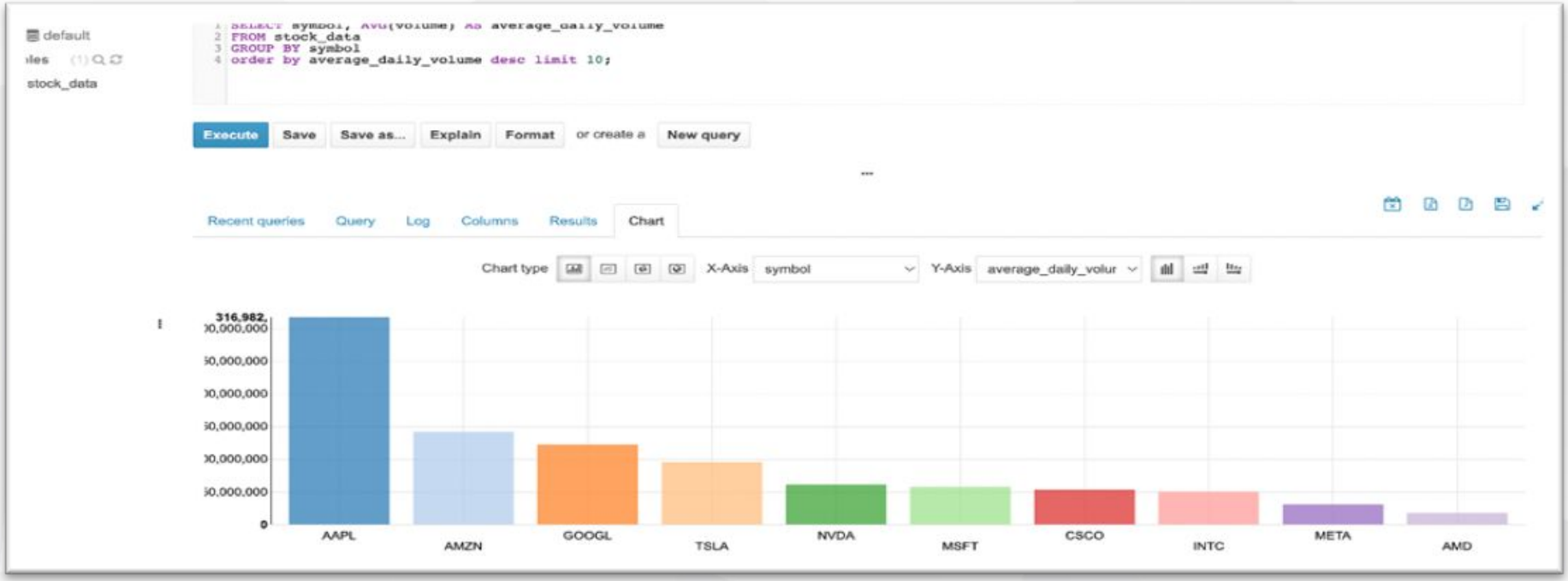
A line chart is used to visualize the total traded volume for AAPL stock every year and based on the above results, 2008 is the year in which maximum trading of stock units happened for AAPL stock

3. What are the top 10 stocks with high average daily volume of stocks traded?



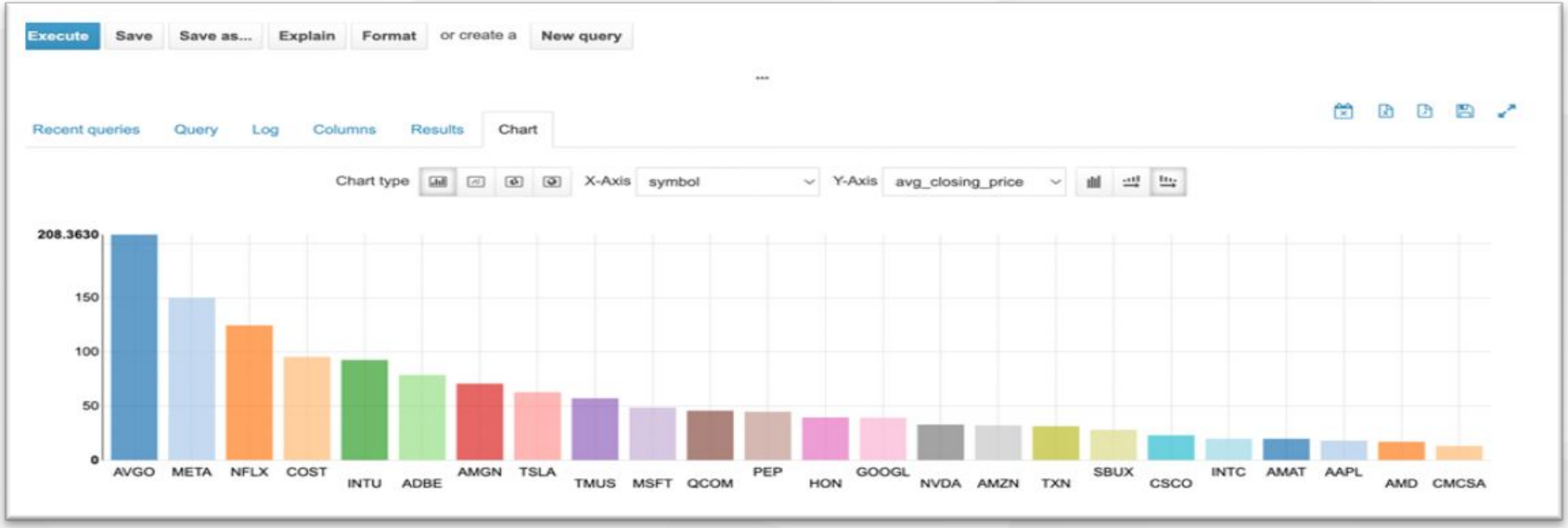
	symbol	average_daily_volume
1	AAPL	316982134.14816898
2	AMZN	142018265.60975611
3	GOOGL	122880701.70173214
4	TSLA	95676789.015035287
5	NVDA	61292437.506112471
6	MSFT	57806589.803963348
7	CSCO	53352882.302466929
8	INTC	50527542.073394492
9	META	31085923.292595256
10	AMD	18372472.073394496

3. What are the top 10 stocks with high average daily volume of stocks traded?



The average of the volume of stocks traded is calculated for each company and the top 10 stocks with highest average volume traded are displayed and a bar chart is used to visualize the result.

4. What is the average closing stock price of the companies?

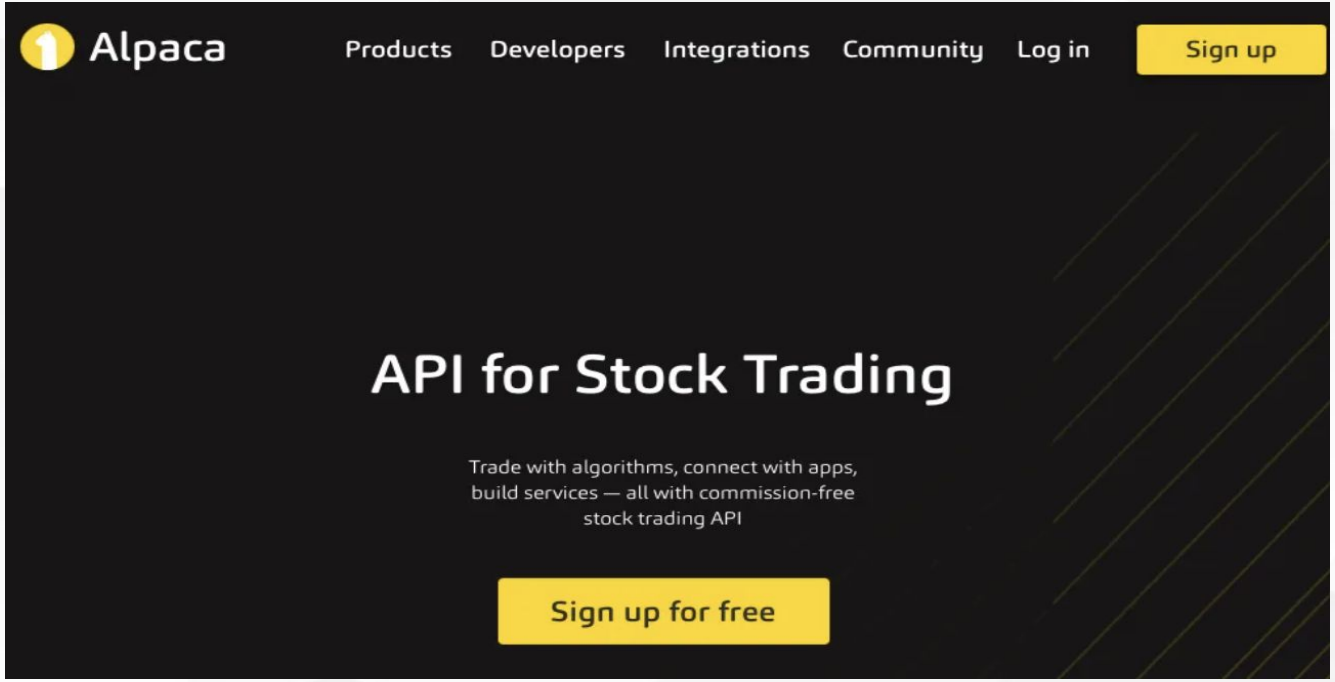


Alpaca

4

Real Time Data Processing

Data Source - Alpaca Markets

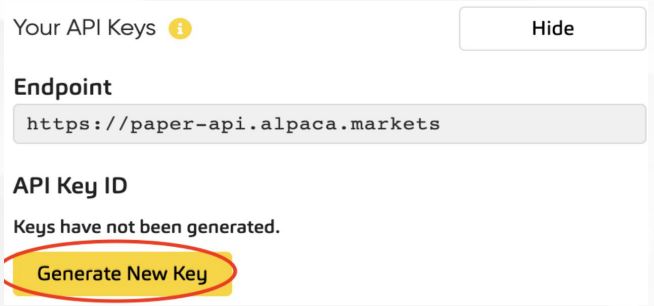


- U.S based stock brokerage firm that provides commission-free trading services.
- It offers a platform and API that allows developers and traders to build trading algorithms, access real-time market data, and manage their investment portfolios.

Accessing real time market data API

- Creating an account - free or paid subscription
- Generating API keys - API_KEY and API_SECRET_KEY

- Communication protocol - Websocket
- WebSocket endpoint - wss://stream.data.alpaca.markets/v2/iex



Client to Server

Input and output:

Server to Client

```
$ wscat -c wss://stream.data.alpaca.markets/v2/sip
connected (press CTRL+C to quit)
< [{"T":"success","msg":"connected"}]
> {"action": "auth", "key": "*****", "secret": "*****"}
< [{"T":"success","msg":"authenticated"}]
> {"action": "subscribe", "trades": ["AAPL"], "quotes": ["AMD", "CLDR"], "bars": [
< [{"T":"subscription","trades":["AAPL"],"quotes":["AMD","CLDR"],"bars":["*"],"upd
< [{"T":"q","S":"AMD","bx":"K","bp":91.95,"bs":2,"ax":"Q","ap":91.98,"as":1,"c":["
< [{"T":"t","S":"AAPL","i":628,"x":"K","p":162.92,"s":3,"c":["@","F","T","I"],"z"
< [{"T":"q","S":"AMD","bx":"P","bp":91.9,"bs":1,"ax":"Q","ap":91.98,"as":1,"c":["R
```

Spark Structured Streaming

- Spark 2.x release onwards
- Built on the Spark SQL library
- Streaming based on Dataframe and Dataset APIs
- Implementation using PySpark library - Python
- **Response Json:**


```
{
  "T": "b",
  "S": "SPY",
  "o": 388.985,
  "h": 389.13,
  "l": 388.975,
  "c": 389.12,
  "v": 49378,
  "t": "2021-02-22T19:15:00Z"
}
```


Defining Schema:

```
schema = StructType([
    StructField("T", StringType(), True),
    StructField("S", StringType(), True),
    StructField("o", DoubleType(), True),
    StructField("h", DoubleType(), True),
    StructField("l", DoubleType(), True),
    StructField("c", DoubleType(), True),
    StructField("v", LongType(), True),
    StructField("t", StringType(), True)
])
```

1. Identify the top 5 stocks with the highest closing price in the incoming batch of stock data.

Output

result

S	max(c)
COKE	646.39
LMT	447.83
NVDA	414.65
IGV	341.83
META	285.73

2. Find the stock with largest price variation in the incoming batch data.

T	S	o	h	l	c	v	t	n	vw
b	CVNA	27.42	27.46	27.41	27.46	503	2023-06-28T18:35:00Z	6	27.424911
b	SPGI	390.73	390.74	390.73	390.74	277	2023-06-28T18:35:00Z	7	390.738339
b	UPST	34.86	34.86	34.86	34.86	115	2023-06-28T18:35:00Z	2	34.858696
b	TAST	4.615	4.615	4.615	4.615	321	2023-06-28T18:35:00Z	3	4.615
b	ADMA	3.84	3.84	3.84	3.84	200	2023-06-28T18:35:00Z	2	3.84
b	EVLO	0.2545	0.2545	0.2545	0.2545	200	2023-06-28T18:35:00Z	2	0.2545
b	CMA	41.99	41.99	41.99	41.99	315	2023-06-28T18:35:00Z	5	41.99
b	PLAY	43.95	43.95	43.95	43.95	104	2023-06-28T18:35:00Z	3	43.949519
b	ARCC	18.62	18.62	18.62	18.62	100	2023-06-28T18:35:00Z	1	18.62
b	AMD	110.525	110.55	110.525	110.55	1274	2023-06-28T18:35:00Z	8	110.529443
b	PEB	13.745	13.745	13.745	13.745	200	2023-06-28T18:35:00Z	1	13.745
b	NU	7.88	7.88	7.88	7.88	1395	2023-06-28T18:35:00Z	14	7.88
b	MGM	42.475	42.475	42.475	42.475	475	2023-06-28T18:35:00Z	5	42.474611
b	RYAAY	110.025	110.025	110.025	110.025	364	2023-06-28T18:35:00Z	5	110.025481
b	PHYS	14.845	14.845	14.845	14.845	3700	2023-06-28T18:35:00Z	3	14.845
b	NFLX	432.33	432.59	432.33	432.59	454	2023-06-28T18:35:00Z	6	432.456784
b	BERY	63.75	63.76	63.75	63.75	400	2023-06-28T18:35:00Z	3	63.7525

Output

result

S	max(c)	min(o)	variation
NFLX	432.59	432.33	0.2599999999999909

Conclusions

EDA

- Close Price over time & one year & one month: AAPL did not exceed $20 until 2010.
- Average Trade Volume by Weekday: Lowest in Monday and Highest in Wednesday
- Decomposing close price trend: a change or variations is uniform for seasonal component
- AAPL Daily Returns: Checking lowest return of Day in 2000, September
- Correlation between Stock Close price: Most Companies generally have positive correlations

K-means Clustering

- Silhouette Score shows number of cluster 3 and Elbow Curve shows 6 is suitable
- After checking the visualization by applying cluster numbers from 6 to 3, 3 is decided
- Confirm that return and variance are in descending order in the order of cluster 1,2,0 through Visualization of Cluster
- The Simplest Diversified Portfolio is amgen from cluster 0, AMD from cluster 1, and Google from 2.

Hive and Spark

- AVGO is highest price stock. The lowest price of stocks, META was the highest among stocks at 17.
- Apple stock had the highest trading volume in 2008.
- Average daily volume was in order of Apple, Amazon, Google, Tesla, and NVDA.

Recommendations

- Analyze more Nasdaq stocks
  - Extract meaningful information by analyzing not only 24 companies but also companies of a wider spectrum
- Composition of a specific portfolio
  - Composing a specific portfolio including the proportion of investment by each company
  - Reflect the seasonal component in the investment period
- Establishment of clustering targeting more companies and earning specific criteria
  - Recorded the criterion how stocks are divided into cluster and stored historically.
- Dashboard composition through API data
  1. Real-time trade price tracking
  2. Find anomalies in volume
  3. Real-time Trend analysis
- Building an automated trading system
  - Establishment of a model for automatic trading based on API data and standards

Thanks!

Do you have any questions?

REFERENCE

yahoo finance. (2023). Retrieved from https://finance.yahoo.com/

yahoo!. (n.d.). What is the adjusted close?. Retrieved from https://help.yahoo.com/kb/SLN28256.html

Victor. (2013, April 21). Should I use regular or adjusted close for backtesting?. StackExchange. Retrieved from https://money.stackexchange.com/questions/21780/should-i-use-regular-or-adjusted-close-for-backtesting

CNNMoney. (2000, September 29). Apple bruises tech sector. Retrieved from https://money.cnn.com/2000/09/29/markets/techwrap/

marketchameleon. (n.d.). Stock seasonality example by market chameleon. Retrieved from https://marketchameleon.com/Learn/Stock-Seasonality

Florian, Follonier. (2020, June 1). Stock market prediction using multivariate time series and recurrent neural networks in python. relataly.com. Retrieved from https://www.relataly.com/stock-market-prediction-using-multivariate-time-series-in-python/1815/

vantage. (2023). Candlestick pattern complete guide: how to use it. Retrieved from https://www.vantagemarkets.com/academy/a-guide-to-candlestick-patterns/

REFERENCE

MuleSoft. (2023). What is an API?. Retrieved from https://www.mulesoft.com/resources/api/what-is-an-api

James, Chen. (2020, August 9). Cluster analysis. Investopedia. Retrieved from https://www.investopedia.com/terms/c/cluster_analysis.asp

James, Royal. (2023, June 16). Why is portfolio diversification important for investors?. Bankrate. Retrieved from https://www.bankrate.com/investing/diversification-is-important-in-investing/

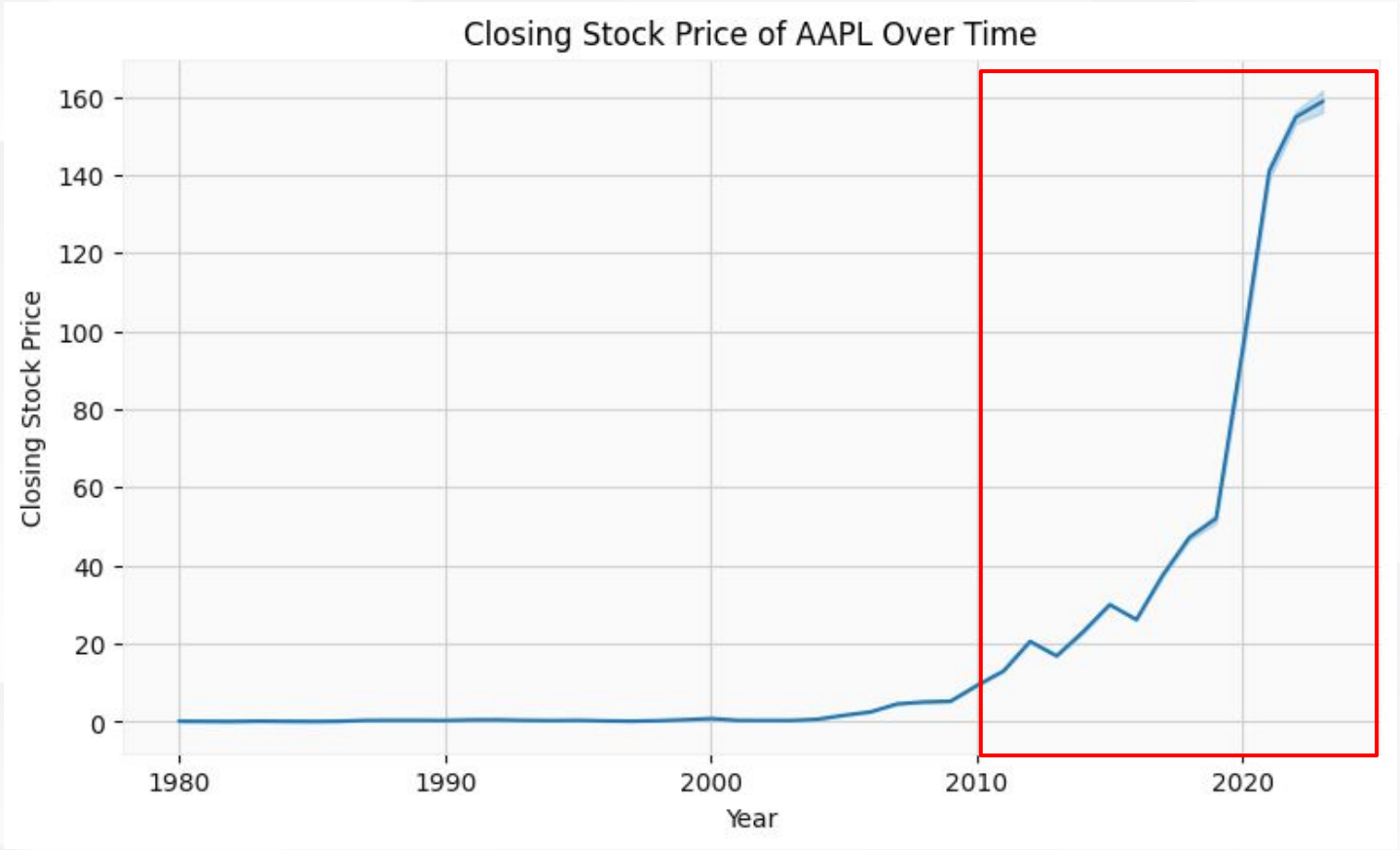
THIRDEYE. (n.d.). Why Spark SQL came into picture?. Retrieved from https://thirdeyedata.ai/spark-sql/

Wallstreetmojo Team. (n.d.). Closing price meaning. Retrieved from https://www.wallstreetmojo.com/closing-price/

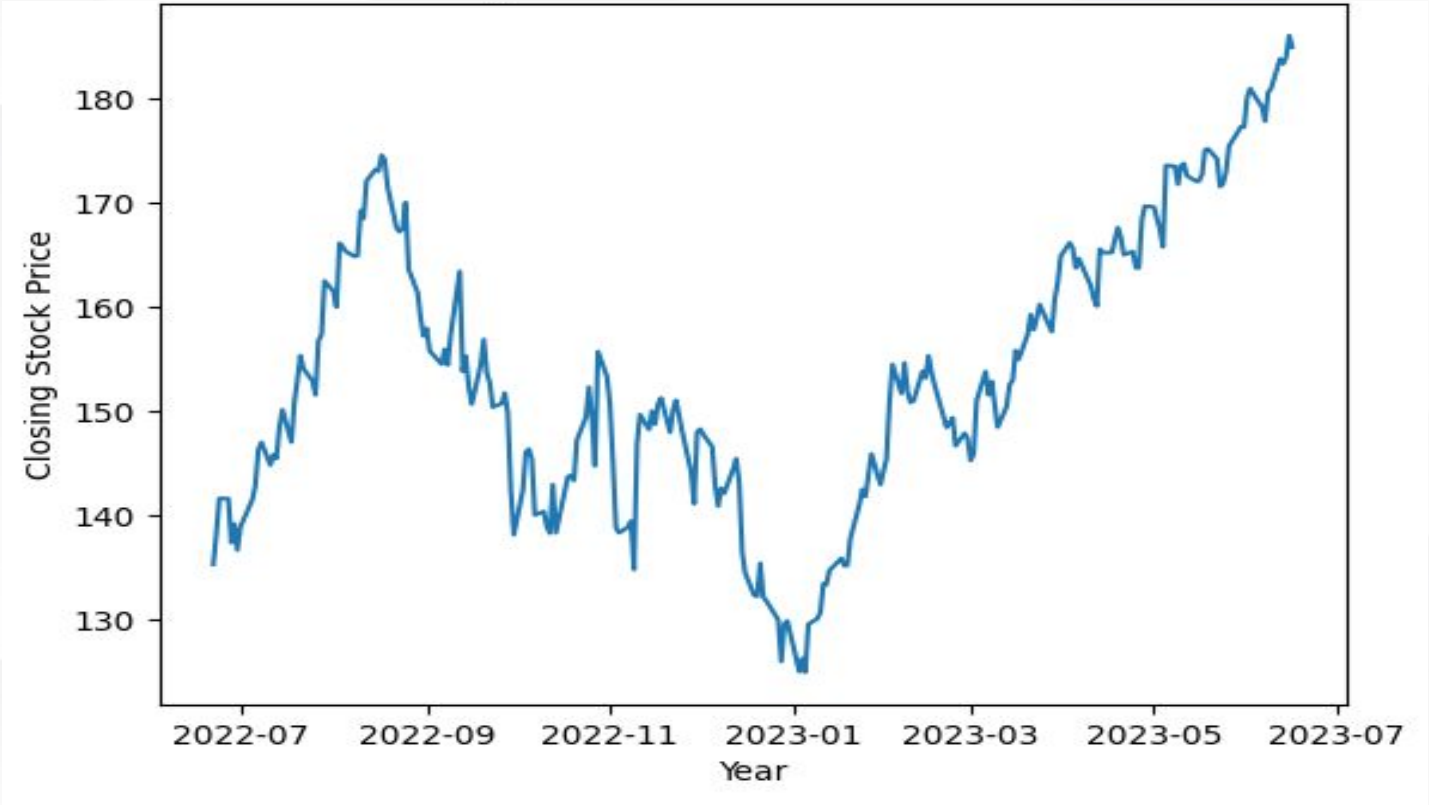
AngelOne. (n.d.). How is the adjusted closing price different from the closing price?. Retrieved from https://www.angelone.in/knowledge-center/share-market/difference-between-closing-price-and-adjusted-closing-price

APPENDIX
EDA

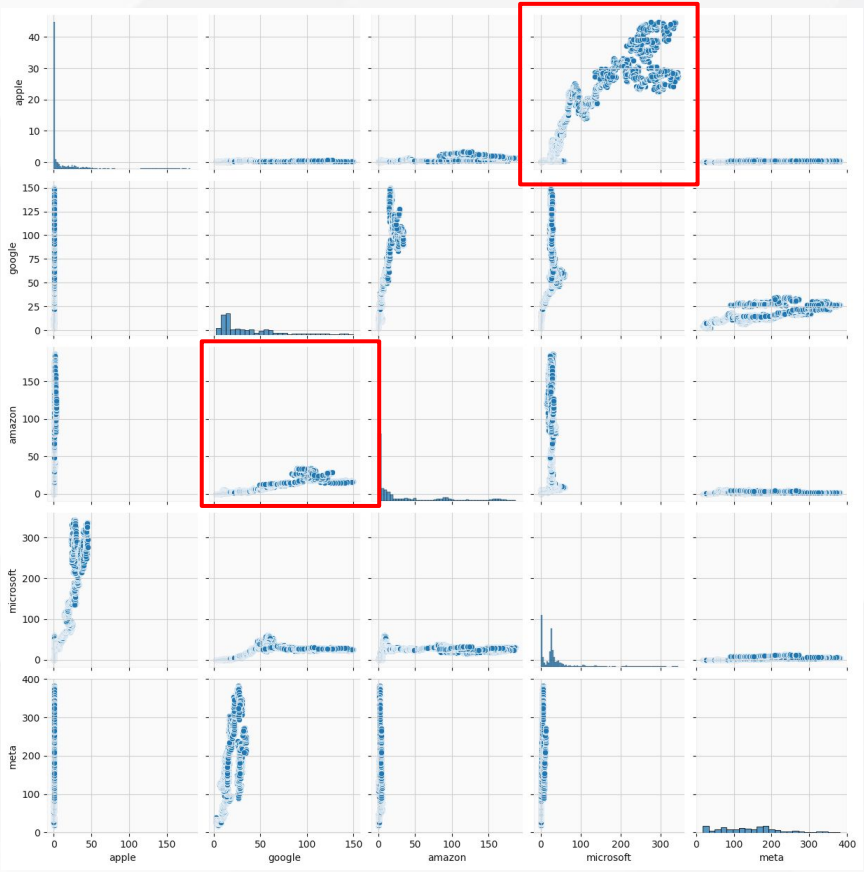
Closing Stock Price of AAPL over time



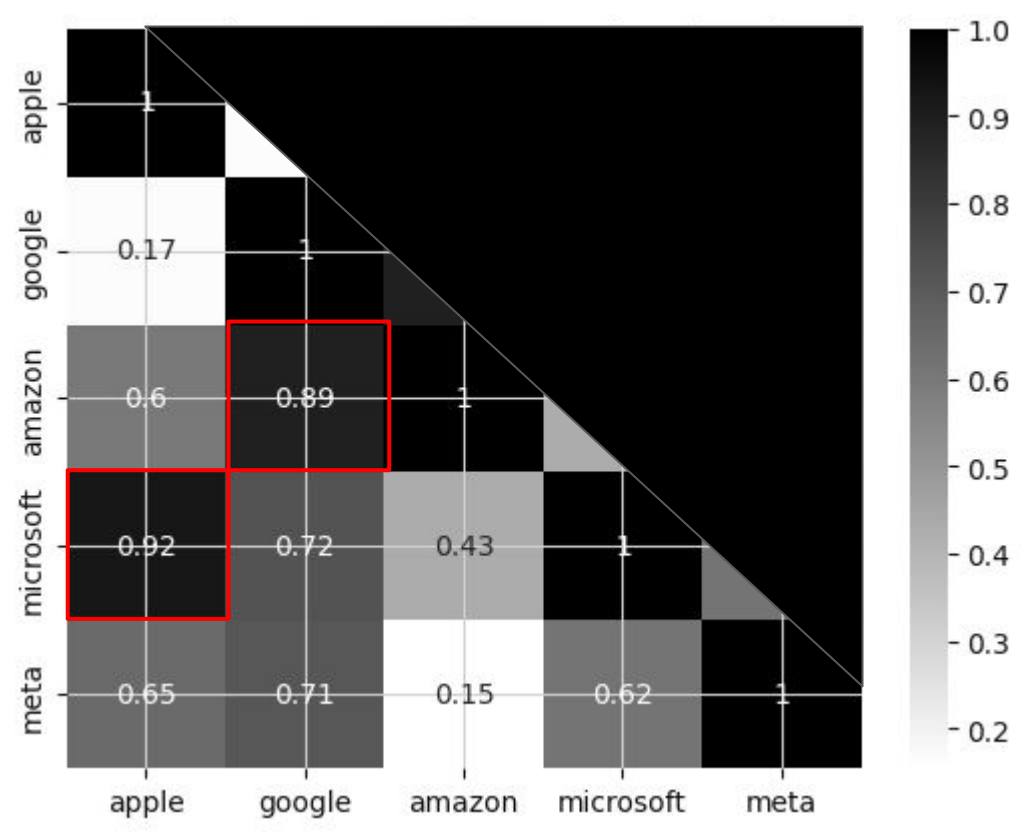
Closing Stock Price of AAPL for one year



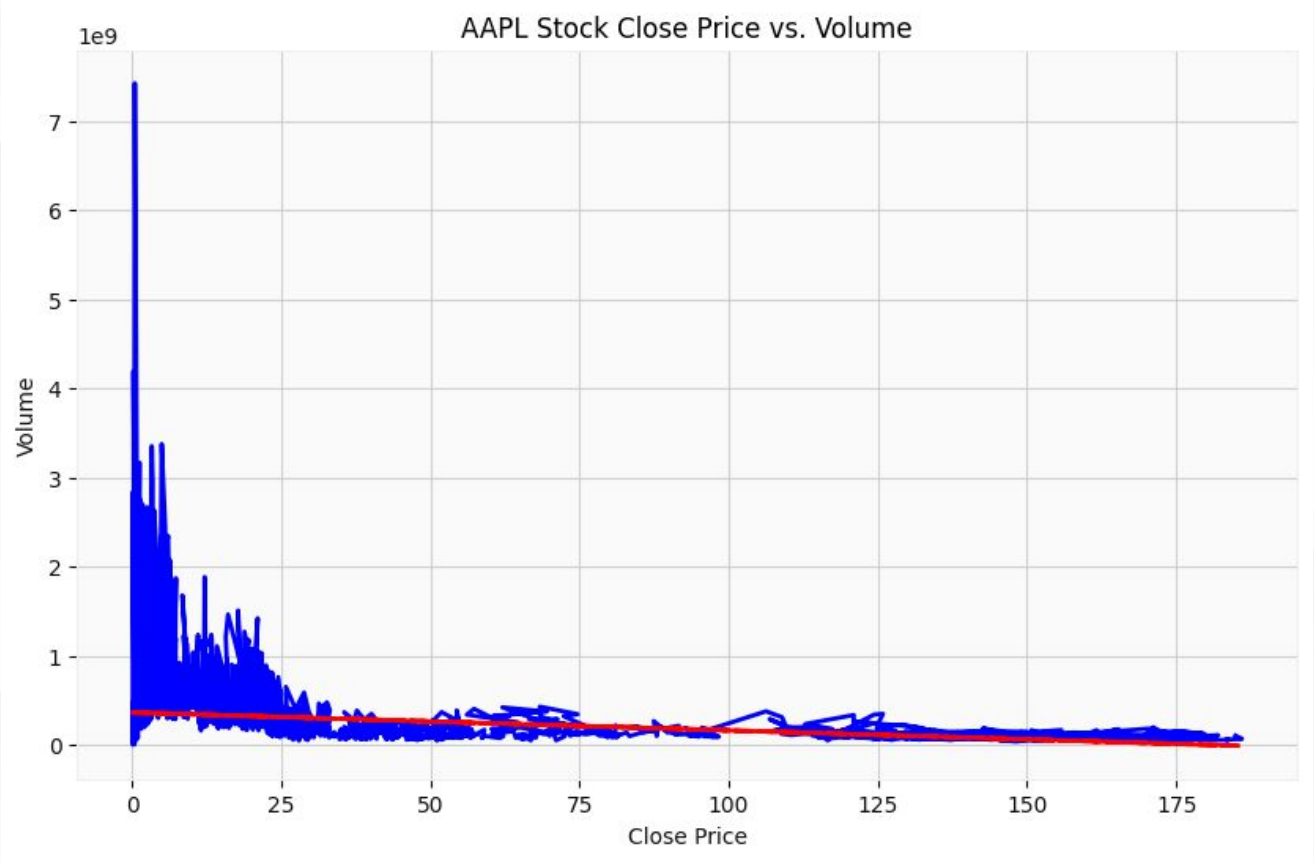
Pair Plot of 5 stocks(AAPL, GOOGL, AMZN, MSFT, META)



Heatmap of 5 stocks(AAPL, GOOGL, AMZN, MSFT, META)



AAPL Stock Close Price vs Volume



OHLC Price Chart of GOOGL

AAPL/ GOOGL/ AMZN/ MSFT date vs volume