– PROJECT REPORT –

AAPL (Apple) Stock Data

~ Group Members ~

Andrey Zhuravlev

Khushnud Ahmed

Momna Ali

Steven Luc

Table of Contents

a. Project Definition

Background:

The stock market is a critical component of the global economy, reflecting business growth and investor sentiment. With Apple's prominence in the technology sector, its stock data offers invaluable insights into market trends and investment patterns. This project utilizes historical Apple stock data (1980–2024) to explore key financial trends, trading volume behavior, and long-term growth, providing actionable intelligence for stakeholders.

Objective:

- Analyze AAPL stock data to extract trends, patterns, and insights.
- Answer critical questions, such as:
    - What are the yearly trends in trading volume and prices?
    - How do price adjustments impact trading behavior?
    - What are the most volatile periods in Apple's stock history?

By leveraging tools from the Hadoop ecosystem, the project will process, query, and visualize insights effectively.

b. Data Description

i. Attributes Description

Dataset 1: aapl_raw_data (main dataset)
(Source: Kaggle)

This dataset contains unadjusted historical stock prices for Apple Inc.

Attributes	Description
date	Trading session date
open	Open stock price
high	Highest stock price during the session
low	Lowest stock price during the session
close	Closing stock price
volume	Number of shares traded during the session

Dataset 2: aapl_split_adjusted (main dataset)
(Source: Kaggle)

Attribute	Description
date	Trading session date
open	Adjusted opening stock price
close	Adjusted closing stock price
adjusted_close	Fully adjusted closing price considering all splits
volume	Adjusted trading volume

Dataset 3: Apple_products_release_dates.csv (additional)
(Source: Wikipedia)

Attribute	Description
released	The date the product was first released
model	The specific model name of the product
family	The product family or category (e.g., iPhone, iPad, Mac)

discontinued	The date the product was discontinued (if applicable)

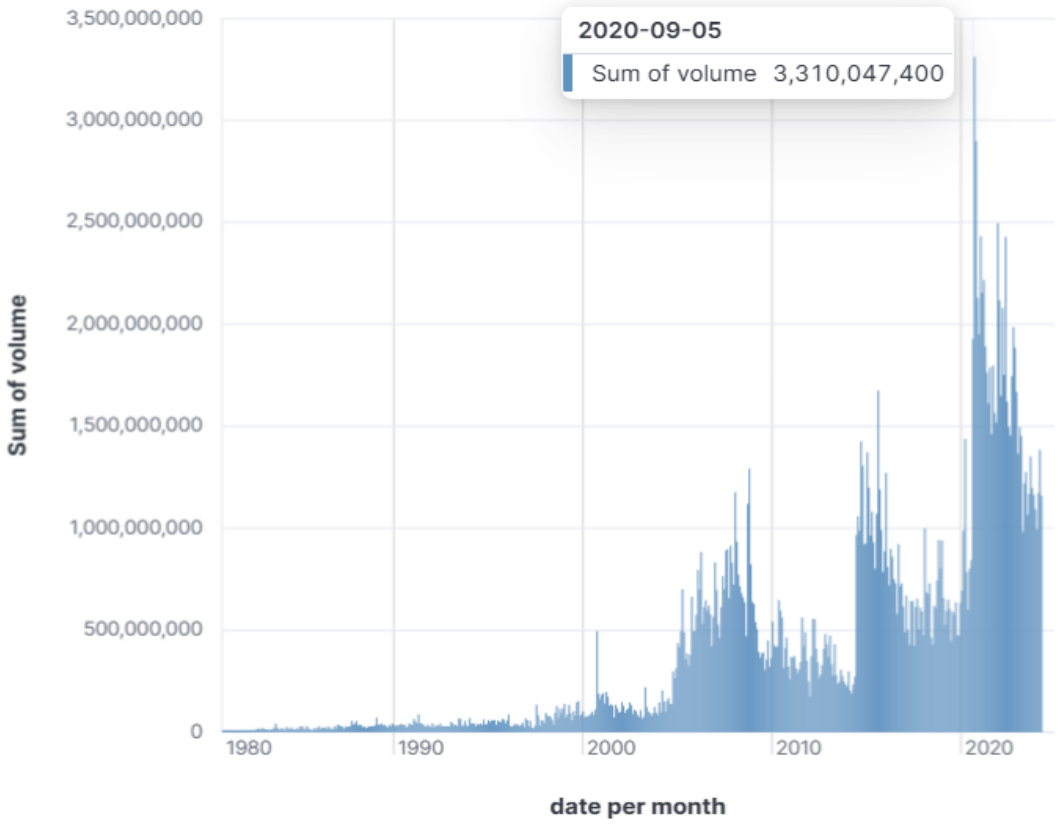
Dataset 4: aapl_EDGAR_all_filings.csv (additional)
(Source: U.S. Securities and Exchange Commission)

Attribute	Description
released	The date the product was first released
model	The specific model name of the product
family	The product family or category (e.g., iPhone, iPad, Mac)
discontinued	The date the product was discontinued (if applicable)

ii. Statistics of the Data

The datasets were preprocessed, queried, and analyzed using tools like Hive, Spark, MapReduce, and ElasticSearch, resulting in the following insights:

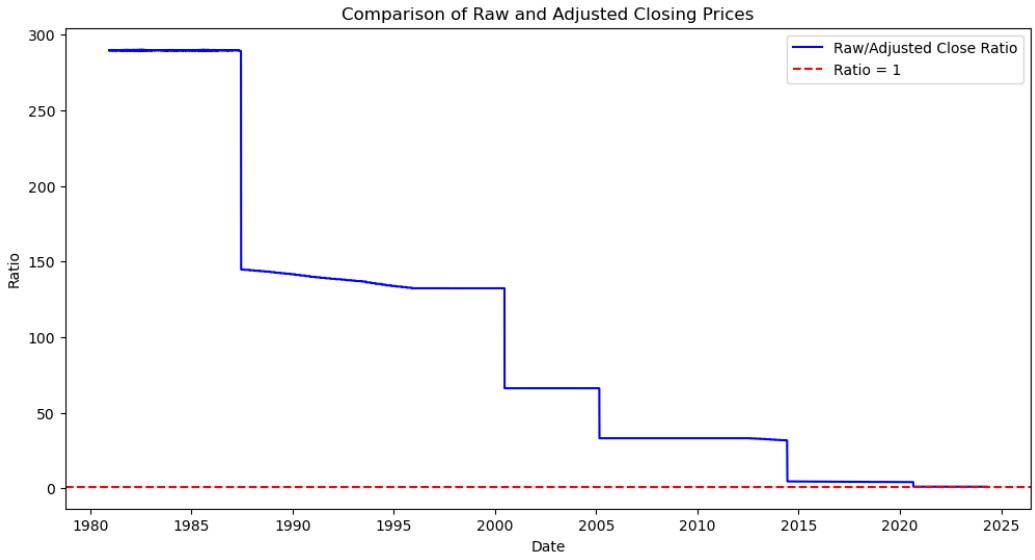
Insight 1.



Peak activity observed during the March 2020 represents the start of the COVID 19 pandemic as demonstrated by the clear representation in the bar chart that depicts monthly trading volumes. It's a direct reflection of the rise of market volatility off of global uncertainty — if I had taken this moment a few weeks ago to do this writeup, I wouldn't have this surge in trading.

Insight 2.

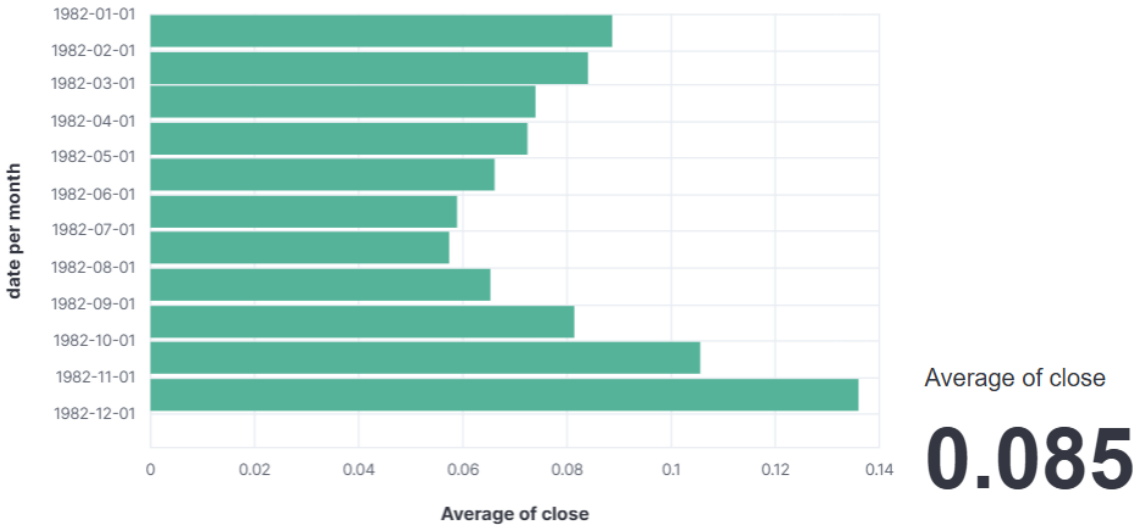
Analysis on how stock splits have influenced stock pricing over the years



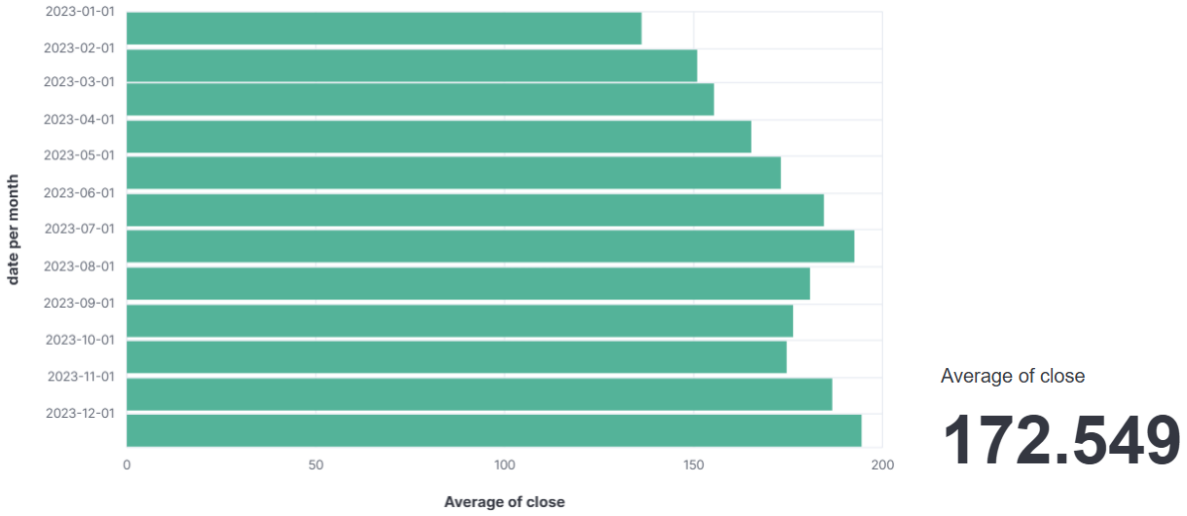
The effect of stock splits and adjustments on raw_close and adjusted_close prices are shown in a linear chart. Corporate actions distort daily prices and adjust prices smooth out these irregularities so that the stock's value is viewed more closely; this is known as 'adjusting' prices.

Insight 3.

Difference between average close prices over time



Average Closing Price for 1982: 0.085, a time when Apple was still a small computer company operating in a small market in California.



Average Closing Price for 2023: 172.549, which shows Apple's influence in the technology market.

Insight 4.

Impact of Product Release on Stock Price by Product Type.

Top 10 product categories (families) that, on average, had the best impact on the stock price and the top 10 with the worst impact for those families that pass the threshold of at least 3 different releases:

68000	3	14.228188375874481
Printers	9	9.750150025174664
AirPort Express	3	6.79244363613175
iPod Shuffle	5	5.66147805653108
Quadra	7	4.831862819571727
Display	7	4.750477940590318
PowerBook Duo	5	4.498568057954815
Macintosh Server	5	4.4329376218376195
Apple II	10	4.285343175886791
iPod Classic	9	3.6918897736630236

Modems	4	-13.295102075972714
MacBook	10	-4.5047819056035205
PowerBook G3	4	-4.304842451578907
PowerBook 100	8	-3.563522780730759
Workgroup Server	5	-3.3391906390511212
MacBook Air	17	-3.3354880299183822
Keyboards	7	-2.7514171533025102
Mac Mini	13	-2.237675184183968
Speakers	9	-1.6654844906239905
iPhone	30	-1.415226059154507

How product releases impact volatility in the first week after product releases:

10 highest volatility:

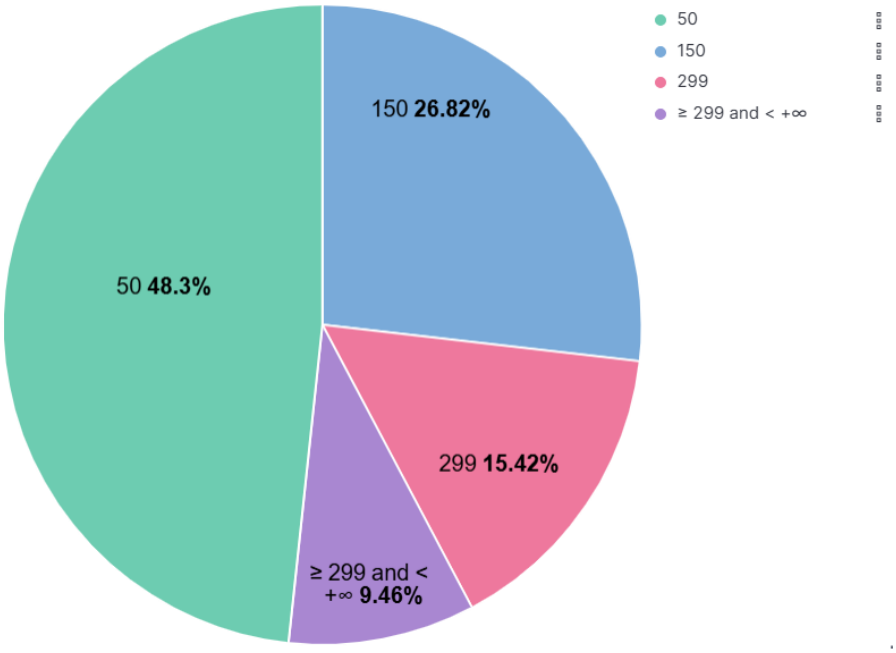
1997-08-05	LaserWriter 8500, Apple ColorSync/AppleVision 750 Display	20.69
1990-10-15	Macintosh LC, Macintosh Classic, Macintosh IIsi	9.05
1983-01-19	Apple Lisa[2]	8.59
1999-10-13	Power Mac G4 Graphite	8.44
2001-01-09	PowerBook G4 Titanium, Power Mac G4 Digital Audio, Apple Pro Speakers (minijack)	7.37
2000-09-13	iBook (FireWire)	7.08
2000-07-19	Power Mac G4 Cube, Cinema Display (22") (ADC), Apple Pro Speakers (USB)	6.88
1994-07-18	Quadra 630, PowerBook 150, Apple Multiple Scan 15 Display	6.81
1999-01-05	Power Macintosh G3 (Blue & White), Macintosh Server G3 (Blue & White)	5.82
2022-10-26	iPad (10th generation), iPad Pro (11-inch) (4th generation), iPad Pro (12.9-inch) (6th generation)	5.74

10 lowest volatility:

1999-10-05	iMac (slot loading)	0.06
1996-04-01	Power Macintosh 7600	0.12
2018-07-12	MacBook Pro (13-inch, 2018, Four Thunderbolt 3 ports), MacBook Pro (15-inch, 2018)	0.26
2011-02-10	iPhone 4 (CDMA) (16 & 32 GB)	0.26
2003-11-18	iMac G4 20"	0.26
2014-06-26	iPod Touch 16GB (5th generation, Mid 2013)	0.28
1989-09-20	Macintosh IIci, Macintosh Portable	0.31
2014-04-29	MacBook Air (Early 2014)	0.34
2013-05-30	iPod Touch (5th generation) (16 GB)	0.34
2017-11-03	iPhone X	0.36

Insight 5.

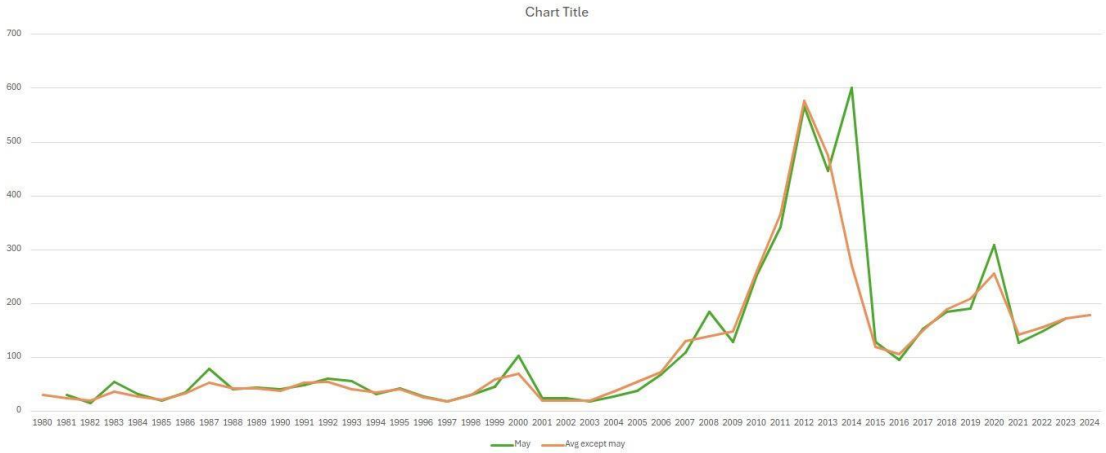
Distribution of price ranges



On the pie chart, Apple's adjusted closing prices are separated into given ranges to visualize how stock prices cycle over time. What this does is help us understand how often the stock traded within given price bands, and show important points of growth or stability.

Insight 6.

Seasonal Volatility



The green line in this line chart depicts the average stock prices in May, while the orange line shows the average stock prices for all months other than May, of the years from 1980 through 2024. On the horizontal axis we have the years, on the vertical axis the stock prices.

Key Observations:

- Steady Growth: Growing since both lines, the products innovations spikes post 2000.
- Volatility in May: In the early 2000s and 2012–2015 the green line is much more volatile, the orange line (other months) is smoother.
- Major Peaks: After Apple's product successes, it peaks in 2012–2015; it dips thereafter.
- May vs. Non-May: In years like 2014 and 2020, May usually underperforms, but in some years, May often outperforms other months.

Insight 7.

Filing types ranked by their average impact on stock returns (left) and ranked by their average volatility (right):

NT 10-Q	1	7.404131328922311
25-NSE	2	7.364050252049568
PX14A6N	1	6.432746780273339
25	1	6.084830240780215
10-K405	2	5.105239520045617
CERT	2	2.6821310735795114
IRANNOTICE	2	2.1177869463452708
CERTNYS	4	1.4789439161716695
424B2	56	1.4716412388620346
SD	11	1.395124467774746
8-K	210	1.3165359268932026
FWP	28	1.298218048287302
8-A12B	6	1.2893221826264296
CORRESP	20	1.2093700248910146
PX14A6G	38	1.1171344711653537
144	13	1.0448388197899605
10-Q	90	0.8713677658883314
UPLOAD	28	0.7702679690139693
SC TO-I	2	0.7401480042102266
S-3ASR	4	0.7243570952551484
SC 13G	21	0.6601932797014202
4	1243	0.6199043431513676
DEFA14A	20	0.11719312412784669
S-8	29	0.05328170838707768
3	34	-0.03023085737074426
10-K	28	-0.09815999818078448
S-8 POS	3	-0.23492189660571552
DEF 14A	30	-0.4019918306917345
DFAN14A	1	-0.7318158079391958
NO ACT	17	-0.8937875998758936
PRE 14A	6	-1.077303838159431
424B3	6	-1.7923842987108618
424B5	2	-2.7793437673413144
S-3	2	-3.248999748103661
SC 13D	1	-3.606748396803263
S-4	1	-4.344457028417287
DEFR14A	2	-5.459823621292452
NT 10-K	1	-7.6620678996120795

PX14A6N	1	3.572286868245634
25-NSE	2	3.0797903684094665
SC 13D	1	3.065925579535839
S-4	1	2.967743246936032
10-K405	2	2.365272224372913
8-K	205	2.3206599819307634
DEFR14A	2	2.26903166643613
SC 13G	20	2.2687613757392024
NO ACT	16	2.121084426781071
S-8	26	2.1056544883994395
SC TO-I	2	2.0542070075328716
PRE 14A	6	1.967734982206389
10-Q	90	1.961155207690532
NT 10-K	1	1.8635807627768066
CORRESP	20	1.8275730498803628
10-K	28	1.8173877104619909
NT 10-Q	1	1.7360587822361764
S-8 POS	2	1.695779526946302
DEF 14A	30	1.687638373232248
3	28	1.5794836554354066
424B5	2	1.5469866464513737
DEFA14A	20	1.50258097742314
S-3ASR	4	1.4691933380942264
4	568	1.415391346051501
UPLOAD	28	1.3939612195444013
PX14A6G	29	1.3495263579680294
DFAN14A	1	1.317054779841815
25	1	1.1345042884437944
CERTNYS	4	1.105136590491358
8-A12B	6	1.0979227223166161
S-3	2	1.0285300058689754
424B3	6	0.9928107177490073
CERT	2	0.9823424873170953
FWP	28	0.9454592271508357
424B2	56	0.8843112489268067
144	9	0.789013652201175
SD	11	0.7398316344979817
IRANNOTICE	2	0.583962683807224

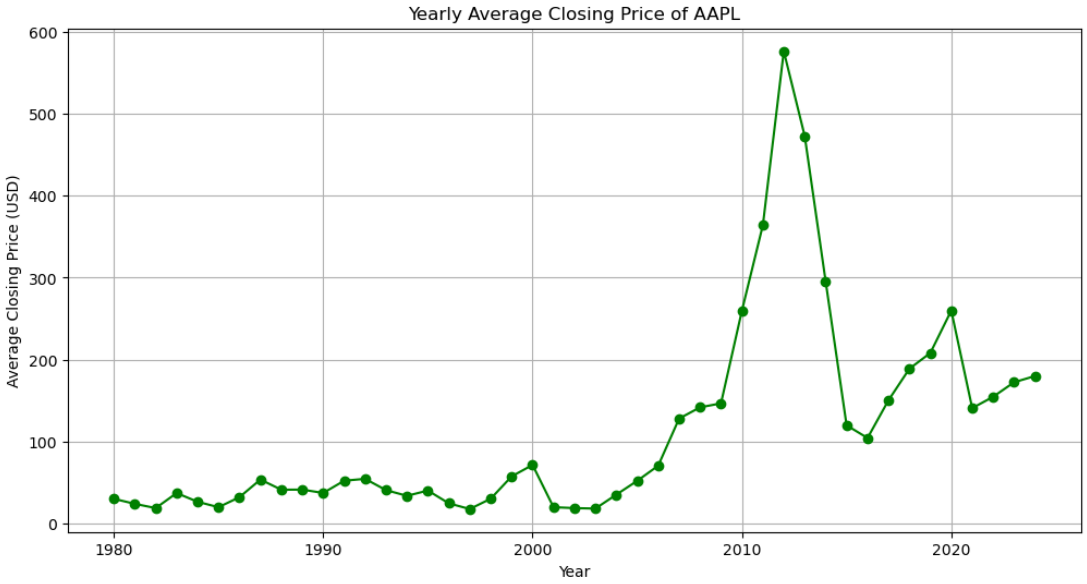
Filing types with only one or two occurrences (like NT 10-Q or PX14A6N), even though quite interesting, might skew results due to small sample sizes, so here are the filing types that have at least 5 filings:

424B2	56	1.4716412388620346
SD	11	1.395124467774746
8-K	210	1.3165359268932026
FWP	28	1.298218048287302
8-A12B	6	1.2893221826264296
CORRESP	20	1.2093700248910146
PX14A6G	38	1.1171344711653537
144	13	1.0448388197899605
10-Q	90	0.8713677658883314
UPLOAD	28	0.7702679690139693
SC 13G	21	0.6601932797014202
4	1243	0.6199043431513676
DEFA14A	20	0.11719312412784669
S-8	29	0.05328170838707768
3	34	-0.03023085737074426
10-K	28	-0.09815999818078448
DEF 14A	30	-0.4019918306917345
NO ACT	17	-0.8937875998758936
PRE 14A	6	-1.077303838159431
424B3	6	-1.7923842987108618

8-K	205	2.3206599819307634
SC 13G	20	2.2687613757392024
NO ACT	16	2.121084426781071
S-8	26	2.1056544883994395
PRE 14A	6	1.967734982206389
10-Q	90	1.961155207690532
CORRESP	20	1.8275730498803628
10-K	28	1.8173877104619909
DEF 14A	30	1.687638373232248
3	28	1.5794836554354066
DEFA14A	20	1.50258097742314
4	568	1.415391346051501
UPLOAD	28	1.3939612195444013
PX14A6G	29	1.3495263579680294
8-A12B	6	1.0979227223166161
424B3	6	0.9928107177490073
FWP	28	0.9454592271508357
424B2	56	0.8843112489268067
144	9	0.789013652201175
SD	11	0.7398316344979817

Insight 8.

Historical Growth



This line chart shows the yearly average close of Apple Inc. (AAPL) stock dating back to the early days of trading in the 1980s, to 2024. Finally, the horizontal (years) axis is recorded on the vertical (average closing price in USD) axis.

c. Work Distribution

Team Members Roles and Responsibilities:

Andrey Zhuravlev: analyzed the impact of product releases and SEC filings on stock prices and volatility with hive, alongside generating key insights into price-volume relationships.

Khushnud Ahmed: designed and implemented intuitive visualizations using ElasticSearch and Kibana to effectively communicate trends and patterns.

Momna Ali: managed the ingestion of raw datasets into Hadoop HDFS and performed data aggregation using MapReduce to identify trends, utilizing Python to generate visual insights from the results.

Steven Luc: optimized high-performance data processing with PySpark and focused on seasonal volatility.

d. Solution Description

i. Tools Used

1. **Apache Hadoop:** A framework for distributed storage and parallel processing of large datasets, scalable, fault tolerant and facilitates handling of large data on clusters.
2. **MapReduce:** A programming model where expensive tasks are broken into small pieces, executed in parallel and aggregated to obtain results efficiently for large scale data computation.
3. **Hive:** A Hadoop based data warehousing tool that allows accessing large datasets in an SQL like syntax with some additional syntax for building tables and constructing queries.
4. **Spark:** A high-performance analytics engine for real time data processing, high performance machine learning model integration, fast data handling, and the ability for iterative processing.
5. **ElasticSearch and Kibana:** Kibana is an interactive visualizations tool with Elasticsearch (which is a search and analytics engine). Together they allow for fast queries, real time data analysis, and straightforward exploration of trends and patterns in exciting, intuitive dashboards.

ii. How They Were Used

- *Hadoop*
    - **Distributed Data Storage:** To store large data sets on a vast distributed network of multiple nodes, Hadoop was used. This approach stored data

in such a way that it could also be accessed from anywhere within the system, while making it efficient for scalability and fault tolerance.

- **Data Processing with MapReduce:** The data was processed using the

```
[root@sandbox-hdp aapl_stock_data]# hadoop fs -put mapper_preprocess.py /user/root/aapl_stock_data
[root@sandbox-hdp aapl_stock_data]# hadoop fs -put reducer_preprocess.py /user/root/aapl_stock_data
[root@sandbox-hdp aapl_stock_data]# hadoop fs -put aapl_raw_data.csv /user/root/aapl_stock_data
[root@sandbox-hdp aapl_stock_data]# hadoop fs -put aapl_split_adjusted.csv /user/root/aapl_stock_data
```

MapReduce framework on Hadoop. Key parameters such as average prices and trading volumes were calculated with this method. It was able to

```
24/11/23 08:25:04 INFO mapreduce.Job:  map 0% reduce 0%
24/11/23 08:25:09 INFO mapreduce.Job:  map 100% reduce 0%
24/11/23 08:25:15 INFO mapreduce.Job:  map 100% reduce 100%
24/11/23 08:25:15 INFO mapreduce.Job: Job job_1732334798004_0011 completed successfully
```

handle massive amounts of data quickly and accurately broken down, allowing it to answer multiple smaller tasks to get useful insights from large datasets.

- *Hive*
  - **SQL-based Data Querying:** Hive was used to query structured data stored in Hadoop, leveraging its SQL-like interface to simplify complex data analysis tasks. By managing and integrating multiple tables, Hive facilitated efficient data exploration, summarization, and transformation. This approach enabled the creation of derived tables and allowed for easier identification of patterns and trends within the data.
  - **Query:**

```
CREATE TABLE apple_stock_data (
  `date` DATE,
  open FLOAT,
  high FLOAT,
  low FLOAT,
  close FLOAT,
  volume BIGINT,
  raw_close FLOAT,
  change_percent FLOAT,
  avg_vol_20d FLOAT
)
ROW FORMAT DELIMITED
FIELDS TERMINATED BY ','
STORED AS TEXTFILE
TBLPROPERTIES ("skip.header.line.count"="1");

LOAD DATA INPATH '/user/root/project/aapl_split_adjusted.csv' INTO TABLE apple_stock_data;
```

```
    > LOAD DATA INPATH '/user/root/project/aapl_split_adjusted.csv' INTO TABLE apple_stock_data;
Loading data to table default.apple_stock_data
Table default.apple_stock_data stats: [numFiles=4, numRows=0, totalSize=3872292, rawDataSize=0]
OK
Time taken: 1.2 seconds
hive> CREATE TABLE sec_filings_10k_10q (
    >   form_type STRING,
    >   form_description STRING,
    >   filing_date DATE,
    >   reporting_date DATE
    > )
    > ROW FORMAT DELIMITED
    > FIELDS TERMINATED BY '\t'
    > STORED AS TEXTFILE;
OK
Time taken: 0.582 seconds
```

- *Spark*
    - **Efficient Data Processing:** Millions of records were processed quickly using Apache Spark to help take advantage of Spark's in memory computing capabilities. With this we were able to analyze huge datasets with high performance and small delay.
    - **Query:**

```
spark.sql("""
SELECT min(date), max(date)
FROM aapl
""").show(500)

--total price range
spark.sql("""
SELECT min(open),min(close),min(high),min(low)
FROM aapl
""").show(500)
```

- *ElasticSearch and Kibana*
    - **Visualization and Reporting:** Using ElasticSearch and Kibana, such as price trend reports, trading activity heat maps, and volatility reports were created to be interactive and impactful reports. They were able to provide actionable insights through these visualizations by showing data through a dynamic, user-friendly format.
    - **Real-time Data Analysis:** With elastic search, queries ran quickly even with large datasets, and with the real time functionality of Kibana, the users could track price moves as well as trading patterns. The combination made for an easy and fast decision policy relying on real time data.
    - **Query:**

```
GET aapl_raw/_search
{
    "query": {
        "script_score": {
            "query": { "match_all": {} },
            "script": {
                "source": "Math.abs((doc['close'].value - doc
                    ['open'].value) / doc['open'].value) * 100"
            }
        }
    },
    "_source": ["date", "open", "close"]
}
```

iii. Why These Tools Were Chosen

- **Hadoop:** The scalable architecture of Hadoop and its fault tolerance turn out to be the best choice for handling big data. As an example, its distributed computing model makes it a great choice of a data processing engine, as it can store huge data sets across numerous nodes, and process it in parallel. For instance, Hadoop can handle large scale data needed to do things like market trend analysis or even process historical financial data without sacrificing performance of reliability.
- **MapReduce:** Was selected because it reduces the computation overhead from large data extensively. It allows efficient processing of complex tasks by breaking such tasks down into smaller more manageable units (called “map” tasks) and aggregating result of those tasks in parallel in the “reduce” phase. MapReduce is especially valuable when application needs to quickly process and analyze lots of data in the distributed environment, e.g. financial transactions or real time processing some large data sets.
- **Hive:** Hive was chosen for its SQL-like query language, which simplifies working with large datasets stored in Hadoop by providing a familiar syntax for analysts and data scientists. This enables faster analysis and easier interaction with complex datasets. Additionally, Hive excels at managing and merging multiple tables to create derived tables, streamlining the process of summarization, filtering, and pattern identification. By abstracting away the complexity of Hadoop's lower-level APIs, Hive allows users to perform these high-level tasks efficiently and focus on deriving insights.
- **Spark:** Apache Spark choice was due to its fantastic performance and capability to handle big data processing in large scale effectively. Compared to traditional forms of batch processing like Hadoop, Spark lets you process data in real time and perform streaming analytics, for example, monitoring stock prices or performing financial transactions, where data has to be processed fast. Thanks to its speed of in memory processing Spark makes processing much faster and its potential to perform complex tasks like machine learning and pattern recognition.
- **ElasticSearch and Kibana:** We chose these tools because they allow us to quickly analyze data and to see the data in an intuitive and interactive visualization. Real time analytics requires high-speed searching and querying capabilities, and that’s what ElasticSearch provides you. ElasticSearch is complemented by Kibana, the rich and user-friendly dashboards and reports that allow users to quickly see data. Then these tools are really good for generating interactive reports like trading activity heat maps orients and price

trend visualizations, for fast visiting insights, and fast decision making on the part of a client.

e. Describing Insights Gathered From Data

1. Insight 1 (Trading Volume Analysis)

- The average trading volume of 23.7 million shares per session indicates consistent market participation by investors.
- The spike in 2020 demonstrates how external global events directly impact market dynamics.

2. Insight 4 (Price Trends)

- The top-performing product families (e.g., 68000, Printers, AirPort Express) are largely from Apple's earlier years. The 68000, a family of products tied to Apple's early innovations with Motorola processors, shows a particularly strong average return of 14.23%, outperforming all others, while recent products like iPhone appear to have a lower impact.
- The volatility results further support the idea that Apple's early product launches were met with stronger market reactions, with only one release date after 2001 making to the top 10 of highest volatility weeks.
- **Possible reasons:**
    - Early innovations might have been perceived as groundbreaking, leading to stronger market reactions.
    - Apple's stock was less established, meaning that positive news or innovation could create larger relative changes in stock prices.
    - By the time modern products are released, the market might have already priced in expectations, reducing the impact of the launch itself, which was not the case in the earlier years.
    - Earlier products launched into less saturated markets, allowing them to make a bigger splash. As Apple grew into a more mature company, the stock's sensitivity to individual product launches might have decreased due to its diversified revenue streams and larger market capitalization.
- *The Modems family shows up as the worst-performing product category (average return -13.30% over 4 releases).*
- ***Possible reasons:***

- *During this period, Apple faced intense competition from companies specializing in networking equipment like 3Com and Hayes*
- *Modem releases are seen as supplementary, commodity-like products and are not exciting to releases*
- *The releases of modems might have coincided with broader company struggles in the early 90s.*

3. Insight 5 (Correlational Analysis)

Observations from the Pie Chart:

- **Majority of Prices in Lower Ranges (1980–2000):**
  - A significant portion of the data lies in the lowest price range ($0–$50). This aligns with Apple's early years, where its stock was less valuable due to its smaller market size and limited product portfolio.
  - The high frequency of low prices also reflects stock splits that kept shares accessible to retail investors.
- **Growth into Higher Price Ranges Post-2000:**
  - As Apple expanded its product line (e.g., iPods, iPhones, and MacBooks), the stock moved to higher ranges ($50–$150).
  - This growth period highlights Apple's successful market positioning and increased investor confidence during the 2000s.
- **Recent Domination in Premium Price Ranges (2010–2024):**
  - A substantial share of stock prices now falls in the $150–$300 and $300+ categories, demonstrating Apple's evolution into a dominant global technology leader.
  - This range reflects its current valuation, driven by sustained innovation, global demand, and market trust.

4. Insight 6 (Volatility Identification)

- **Seasonal Trends:**
  - May often shows significant activity, which could be tied to Apple's quarterly reports, new product announcements, or investor sentiment during that period.
  - This insight might suggest an opportunity for short-term investors to capitalize on May-specific trends.

- **Event-Driven Peaks:**
    - Years with large differences between May and other months (e.g., 2012) could be linked to significant announcements or global market events.
- **Overall Growth:**
    - Despite some fluctuations, the chart demonstrates Apple's robust growth trajectory, with May occasionally showing accelerated growth or volatility.

5. Insight 7 (Impact of SEC Filing Type on Stock Price and Volatility)

- The 25-NSE filings (debt securities that have matured or been redeemed) have happened only twice, but on average have one of the highest positive returns (+7.36%). The reason for this can be that successfully retiring debt obligations on schedule can be seen as a positive sign of a company's financial health and ability to meet its commitments.
- However, looking at filings with small sample sizes without historical context doesn't provide strong results. For example, 0-K405 is a historical form type that was used to indicate that a company had failed to file required insider trading reports, but in the data it shows+5.10% average return. Therefore it might be more productive to look at the filings which meet a sample threshold.
- From the ones that meet the threshold, we can see that routine filings (10-Q and 10-K which represent quarterly and annual reports) hover near neutral and don't have a strong impact on the stock.
- The most positive and stable market reactions consistently come from debt-related filings like 424B2 (Prospectus Supplements - documents detailing specific terms of new debt or security offerings) and FWP (Free Writing Prospectus - additional marketing materials for security offerings), showing returns above 1.3% with notably low volatility (below 1%). This suggests the market consistently rewards Apple's debt management and communication strategy.
- In contrast, filings that introduce uncertainty see negative returns despite their different purposes: NO ACT (No Action Letters - formal requests to SEC asking if certain actions would violate securities laws) at -0.89%, PRE 14A (Preliminary Proxy Statements - early drafts of shareholder voting materials) at -1.08%, and 424B3 (Prospectus Updates - amendments announcing material changes to previously filed offering documents) at -1.79%. The higher volatility in these filings (ranging from 1-2%) reinforces that the market dislikes uncertainty in regulatory interactions and changes to existing plans.

- Another interesting pattern is in 8-K filings (Current Reports - announcements of significant company events like executive changes, major contracts, or acquisitions), which show the highest volatility (2.32%) while generating third-highest returns (1.32%). This suggests that while material company events generally drive Apple's stock price up,

they also create significant market uncertainty, unlike the debt-related filings that inspire more consistent market confidence.

6. Insight 8 (Historical Growth)

- **Clear Representation of Apple's Growth Journey:**
    - The chart highlights Apple's historical growth trajectory, showcasing periods of stability, rapid growth, and correction phases. It provides a comprehensive view of the company's market performance over decades.
    - It reveals how Apple evolved from a relatively stable stock in its early years to a technology giant driving investor confidence and generating substantial returns.
- **Major Milestones in Apple's History:**
    - **1980s and 1990s:** The chart shows a relatively flat growth pattern during these decades, indicating Apple's early developmental phase and market uncertainties.
    - **2000–2010:** A clear rise starts during the early 2000s, coinciding with the release of game-changing products like the **iPod** (2001) and **iPhone** (2007), which redefined Apple's position in the market.
    - **2010–2015:** A significant peak in this period aligns with Apple's dominance in the smartphone market and record-breaking sales, especially after the release of multiple iPhone iterations.
    - **Post-2015:** A slight dip followed by recovery, likely due to market corrections, global economic factors, and competition, reflects Apple's ability to sustain and regain its market position.
- **Demonstration of Resilience:**
    - Despite corrections (notable dips post-2010 and around 2020), the chart demonstrates Apple's long-term resilience, with the stock eventually rebounding and maintaining an upward trajectory.

f. Future Work

- **Real-Time Data Streaming:** an API will have to be developed to support live data streaming to improve the system's capability. With this, the system will be able to process data as generated, and do real time analysis. This will increase responsiveness and users will be able to respond quickly to data by incorporating live stock prices or live market events, delivered in real time. Real time analytics can also have features such as automated alerts, dynamic visualizations, instant reporting.
- **Predictive Modeling:** such as ARIMA (AutoRegressive Integrated Moving Average) and LSTM (Long Short-Term Memory) networks should be added to improve forecasting accuracy. These models will help predicted future prices on based of historical data, patterns, trend, and other factors that determines market behavior. By using these sophisticated techniques, the system will provide more reliable predictions with which investors and analysts can assess when the market is about to shift into a more volatile regime.
- **Incorporating Macroeconomic Indicators:** are critical to further refine the analysis by being ingested into the data models. The system can gain a better understanding of economic forces behind market trends by including factors, which includes inflation, interest rates and GDP data. The addition of these indicators will provide a more comprehensive window into the overall economic canvas that can offer a more complete view of what's happening in the relationship between the economy and the financial markets. Once added, you have an additional layer of information that can help make a prediction more accurate and give you additional context for your investment strategy.

g. References

Kaggle AAPL Dataset
Apache Hadoop Documentation
Apache Hive
Apache Spark
ElasticSearch Guide
EDGAR (Electronic Data Gathering, Analysis, and Retrieval) - database system operated by SEC (the U.S. Securities and Exchange Commission).
Wikipedia (https://en.wikipedia.org/wiki/List_of_Apple_products)