Final Project
Stock Market Analysis

Group Members

1. Meghanjali Chennupati
2. Shivani Reddy Donthi reddy
3. Sri Venkata Likhitha Duggi
4. Surendra Maddipati
5. Sai Sruthi Arepalli

Table of Contents

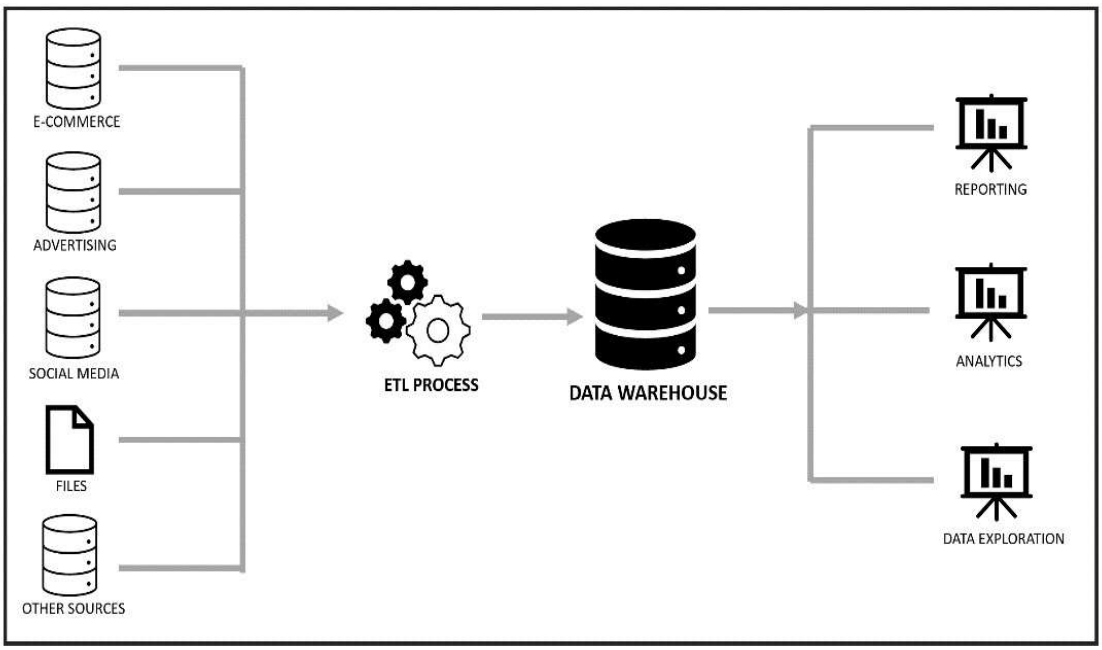
EXECUTIVE SUMMARY

A data warehouse is a collection of data that is subject-oriented, integrated, nonvolatile, and time-variant, designed to support management decisions. It is widely used by businesses for their business intelligence and decision support systems. The data warehouse efficiently organizes, stores, and analyzes large amounts of transactional data in a structured manner, making it easy to utilize in analytical platforms for deriving valuable insights and recommendations. To achieve high performance and easy retrieval, data is typically stored in a denormalized format. The data warehouse employs the Extract-Transform-Load (ETL) process for cleaning and integrating data from various databases and online analytical processing (OLAP). OLAP allows businesses to analyze their data from multiple perspectives, enabling complex calculations, summarization, consolidation, aggregation, and trend analysis. By utilizing OLAP, end users can perform ad hoc analysis of data in various formats, gaining meaningful insights to make informed decisions.

The Stock Market is a platform where stocks or shares of different listed companies are bought and sold. Throughout the day and over time, stock prices fluctuate due to various factors. Some individuals anticipate an increase in stock value and purchase stocks at the current price with the intention of selling. them in the future to make a profit. On the other hand, some people expect the price to decline, so they sell stocks at the current price and may repurchase them at lower prices later.

Several factors influence stock prices, with the most fundamental being the balance between supply and demand, which determines daily price fluctuations. This project aims to construct a warehouse model and analyze various parameters of stock market exchange data for different companies. The model incorporates information such as the date of stock trading, trading volume, and active and inactive stocks.

Additionally, it sheds light on the contribution of GDP per sector and provides details on the all-time high, low, open, and close values of stocks.



PROBLEM STATEMENT

Over the past decade, the analysis of the stock market has gained significant importance. Our trading markets have undergone substantial changes compared to those of previous generations, now being dominated by electronic trading and advanced technologies. Policymakers and investors recognize the stock market's role in mobilizing resources. The increased participation in the stock market, both from institutional and individual, domestic, and foreign investors, has propelled it to new heights. Volatility is a crucial factor when it comes to investing in the stock market. Fluctuations in stock prices have negative effects on individual earnings and the overall health of the economy. They create an environment of uncertainty that hampers investments.

Intelligent investing entails having sufficient knowledge about the market and the stocks traded within it. Various categories of stocks from different sectors are listed and traded. Large-cap stocks from sectors like banking, power, infrastructure, fast-moving consumer goods, and automobiles are traded, alongside mid-cap and small-cap stocks from the same sectors. However, not all stocks within sectors exhibit the same risk and return behavior. The performance of stocks varies depending on both common and sector-specific factors. Proper planning is necessary to withstand adverse impacts and capitalize on value appreciation opportunities. Efficient risk management requires investors to understand the volatility patterns of individual stocks and the interrelationships of assets.

In this case study, the data can be utilized by machine learning/data scientists to study and apply various algorithms that extract patterns and yield meaningful insights from historical data. The current research aims to implement steps that analyze the open, close, high, low, maximum value, and average of stocks on a day-to-day basis.

LITERATURE REVIEW

The objective of this project is to construct a warehouse model and analyze various parameters of stock market exchange data for multiple companies. The model includes information such as stock trading dates, trading volume, and active and inactive stocks. Additionally, it provides insights into the contribution of GDP per sector and the high, low, open, and close values of stocks.

To guide our work, we consulted a research paper titled "**A Data Warehouse-based Modelling Technique for Stock Market Analysis**," where the author discusses the identification of a warehouse model for building an analytical framework.

The paper explores different parameters that directly influence stock market fluctuations and highlights the applications of the analytical model in forecasting information. for decision-making purposes. This research paper greatly contributed to our understanding of the key parameters that impact the stock market.

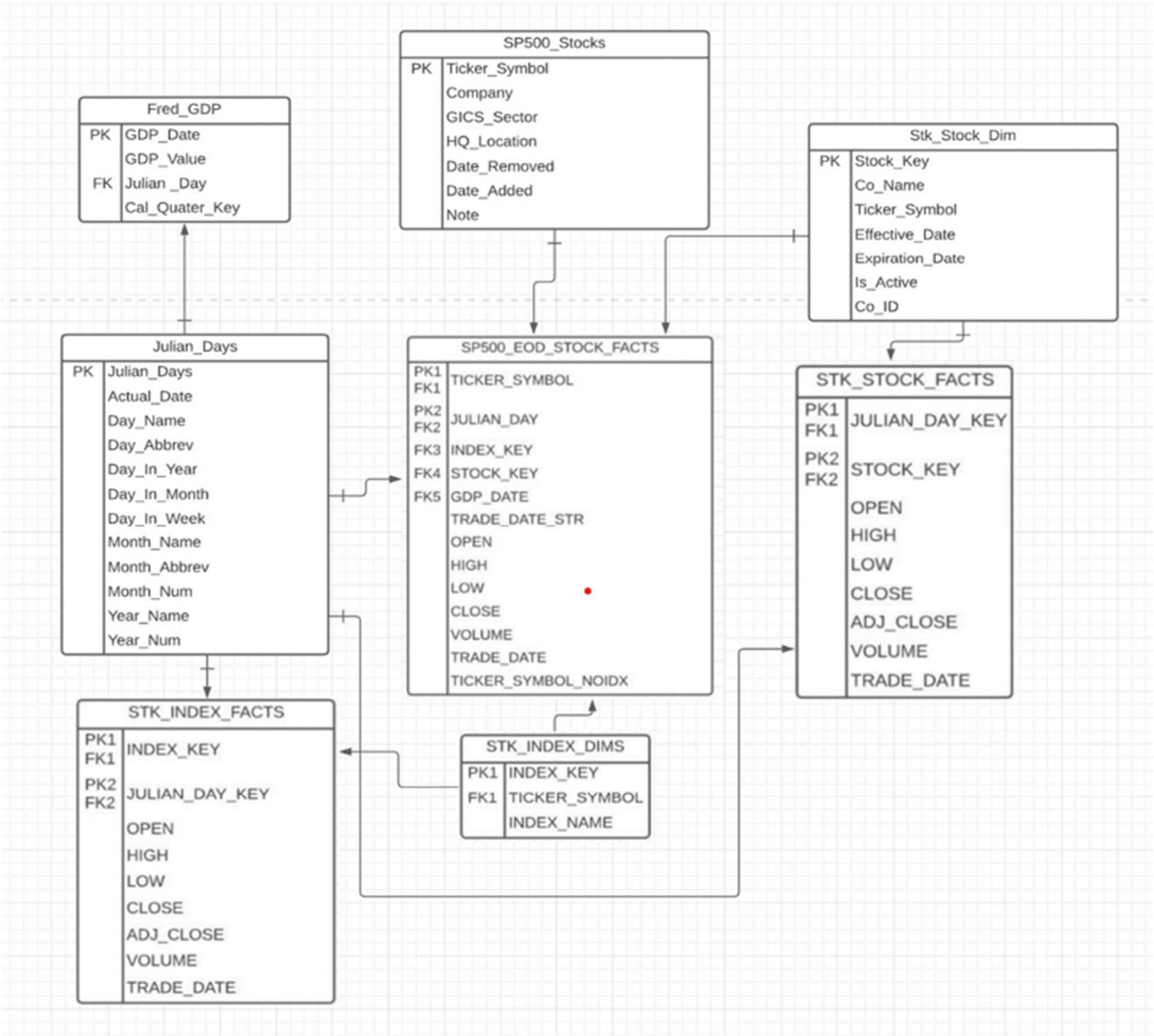
In addition to this paper, we also referred to "**The Data Warehouse Toolkit**" book and course materials to grasp the concepts involved in building the warehouse and conducting the analysis.

DATA COLLECTION AND PREPARATION

The data utilized in our project has been sourced from different tables in the **FIN** schema. To begin with, the data was exported to Excel files. We then proceeded to clean the data within Excel, ensuring its suitability for visualization purposes. During the import of the dataset into the Tableau visualization tool, we excluded any null values present in the data. By performing joins between tables based on key and shared fields, we constructed comprehensive visualizations that span multiple tables. These visualizations will play a crucial role in addressing the business queries covered in the subsequent sections of the report.

DATABASE DESIGN

ER Diagram shows relationships among different entity sets that are stored in the **FIN** schema database.



TABLES USED

NAME	DATATYPE	KEY CONSTRAINT
JULIAN DAY	NUMBER(12)	PRIMARY KEY
ACTUAL DATE	DATE	
DAY NAME	VARCHAR(20)	
DAY ABBREV	VARCHAR(5)	
DAY IN YEAR	NUMBER(6)	
DAY IN MONTH	NUMBER(6)	
DAY IN WEEK	NUMBER(6)	
MONTH NAME	VARCHAR(20)	
MONTH ABBREV	VARCHAR(5)	
MONTH NUM	NUMBER(6)	
YEAR NAME	VARCHAR(40)	
YEAR NUM	NUMBER(6)	

2. SP500_STOCKS

NAME	DATATYPE	KEY CONSTRAINT
TICKER SYMBOL	VARCHAR(10)	PRIMARY KEY
COMPANY	VARCHAR(50)	
GICS SECTOR	VARCHAR(50)	
HQ LOCATION	VARCHAR(50)	
DATE REMOVED	DATE	
DATE ADDED	DATE	
NOTE	VARCHAR(200)	

3. FRED_GDP

NAME	DATATYPE	KEY CONSTRAINT
GDP DATE	DATE	PRIMARY KEY
GDP VALUE	NUMBER(10,1)	
JULIAN DAY	NUMBER(12)	FOREIGN KEY
CAL QUARTER KEY	NUMBER(5)	

3. STK INDEX DIMS

NAME	DATATYPE	KEY CONSTRAINT
INDEX KEY	NUMBER(6)	PRIMARY KEY
TICKER SYMBOL	VARCHAR(10)	
INDEX NAME	VARCHAR(50)	

4. SP500_EOD_STOCK_FACTS

NAME	DATATYPE	KEY CONSTRAINT
TRADE DATE STR	VARCHAR(10)	
TICKER SYMBOL	VARCHAR(10)	COMPOSITE PRIMARY KEY
OPEN	NUMBER(10,2)	
HIGH	NUMBER(10,2)	
LOW	NUMBER(10,2)	
CLOSE	NUMBER(10,2)	
VOLUME	NUMBER(12)	
TRADE DATE	DATE	
JULIAN DAY	NUMBER(12)	COMPOSITE PRIMARY KEY
TICKER SYMBOL NOIDX	VARCHAR(10)	
GDP_DATE	DATE	FOREIGN KEY
INDEX_KEY	NUMBER(10,2)	FOREIGN KEY
STOCK_KEY	NUMBER(10,2)	FOREIGN KEY

6. STK_INDEX_ FACTS

NAME	DATATYPE	KEY CONSTRAINT
OPEN	NUMBER(12,6)	
HIGH	NUMBER(12,6)	
LOW	NUMBER(12,6)	
CLOSE	NUMBER(12,6)	
ADJ CLOSE	NUMBER(12,6)	
VOLUME	NUMBER(12)	
TRADE DATE	DATE	
INDEX KEY	NUMBER(6)	COMPOSITE PRIMARY KEY
JULIAN DAY KEY	NUMBER(9)	COMPOSITE PRIMARY KEY

7. STK_STOCK_FACTS

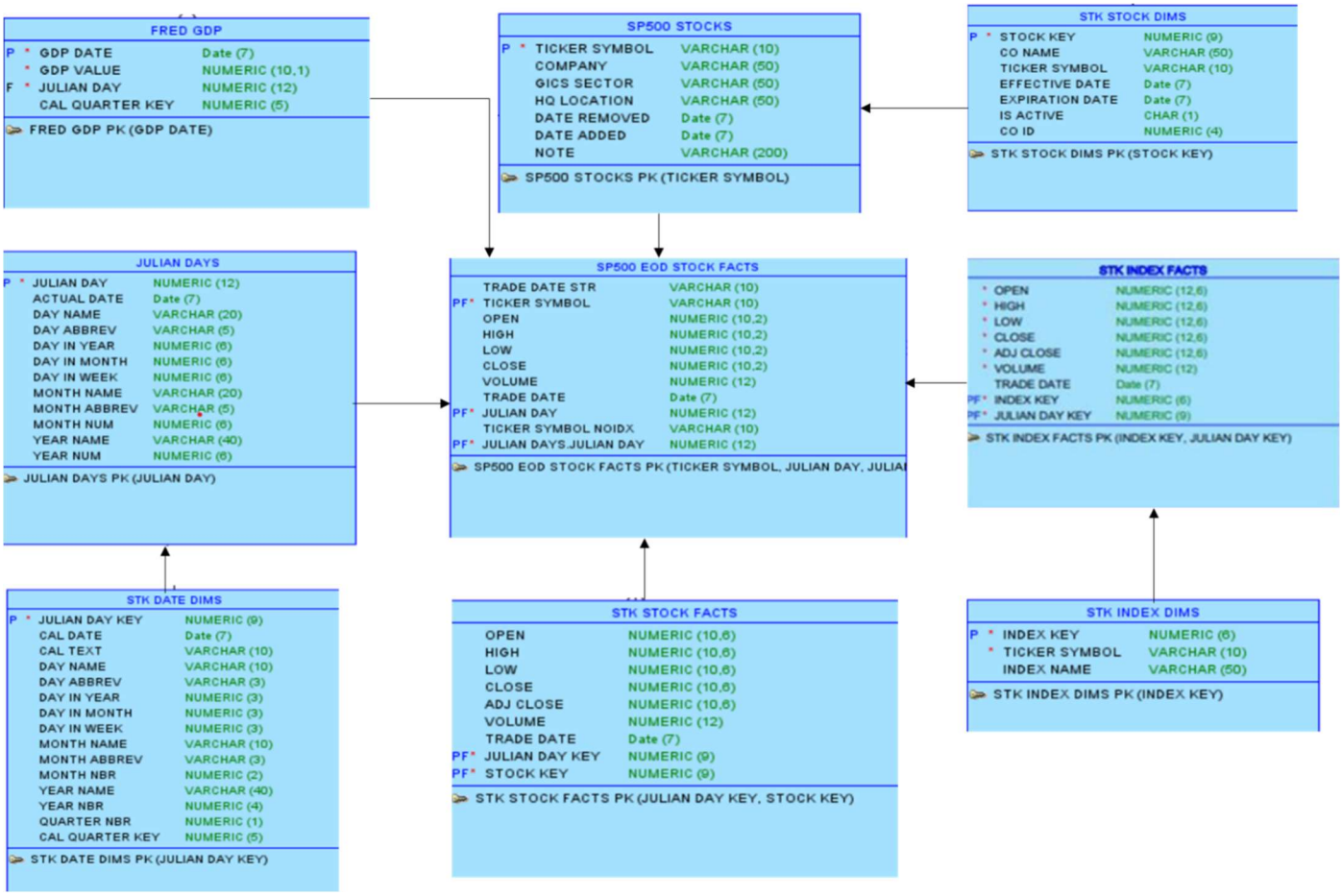
NAME	DATATYPE	KEY CONSTRAINT
OPEN	NUMBER(10,6)	
HIGH	NUMBER(10,6)	
LOW	NUMBER(10,6)	
CLOSE	NUMBER(10,6)	
ADJ CLOSE	NUMBER(10,6)	
VOLUME	NUMBER(12)	
TRADE DATE	DATE	
JULIAN DAY KEY	NUMBER(9)	COMPOSITE PRIMARY KEY
STOCK KEY	NUMBER(9)	COMPOSITE PRIMARY KEY

8. STK_STOCK_DIMS

NAME	DATATYPE	KEY CONSTRAINT
STOCK KEY	NUMBER(9)	PRIMARY KEY
CO NAME	VARCHAR(50)	
TICKER SYMBOL	VARCHAR(10)	
EFFECTIVE DATE	DATE	
EXPIRATION DATE	DATE	
IS ACTIVE	VARCHAR(1)	
CO ID	NUMBER(4)	

DIMENSIONAL MODEL

The Dimensional model is a design approach utilized in constructing data warehouses, aimed at facilitating analysis and decision-making for businesses. It involves the identification of data required for these purposes. A Dimensional model consists of a Fact table, which is typically a large table, and several smaller tables known as dimension tables. Each dimension table contains a primary key that corresponds to a specific unit of the multi-part key found in the fact table.



FACT TABLE – Fact tables store crucial business metrics at the intersection of various related dimensions, effectively capturing essential information. The fact tables utilized in our system include the following:

SP500 EOD STOCK FACTS – The fact table comprises the constantly fluctuating stock prices and corresponding ticker symbols of the top 500 companies listed on US stock exchanges. Each entry in the table contains details such as the trading date, opening and closing values, high and low values, volume traded, as well as the specific date and time dimensions associated with the trades.

STK STOCK FACTS -The fact table encompasses the dynamic stock price values and their corresponding ticker symbols. Each granular entry within the table includes essential details such as the trade date, stock opening value, highest value reached, lowest value reached, closing value, the volume of trades conducted, and associated date and time dimensions. Additionally, the fact table includes the primary key that uniquely identifies each stock.

STK INDEX FACTS - The fact table comprises the ever-changing stock price values and their corresponding ticker symbols. Each individual entry within the table captures essential information such as the trade date, stock opening value, highest value attained, the lowest value reached, adjusted closing value, volume of trades conducted, and associated date and time dimensions.

FRED GDP - The fact table consists of the GDP value along with the date details.

DIMENSION TABLE – Dimension tables store distinct categories that define the level of detail in the measurements found in the fact table. The following tables have been utilized for this purpose:

JULIAN DAYS – Each entry (grain) in this table encompasses diverse to date and time-dimension details related to every change in stock values.

STK INDEX DIMS - Every individual entry in this table comprises the index number and corresponding ticker symbol of a specific stock.

STK STOCK DIMS - Each entry in this table contains the company name, its corresponding ticker symbol, the date when it became effective in the market, and its expiration date.

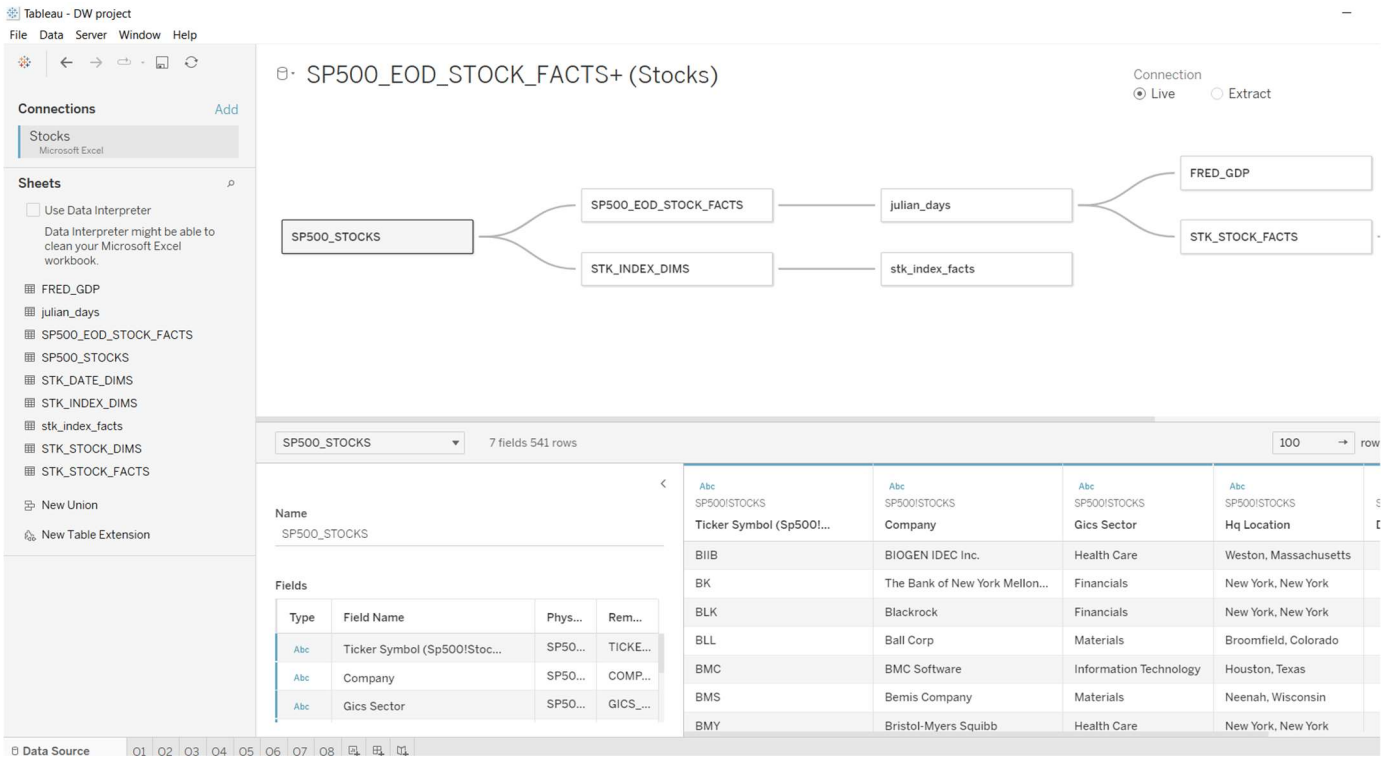
STK DATE DIMS - Each record in this table encompasses specific date and time dimension details.

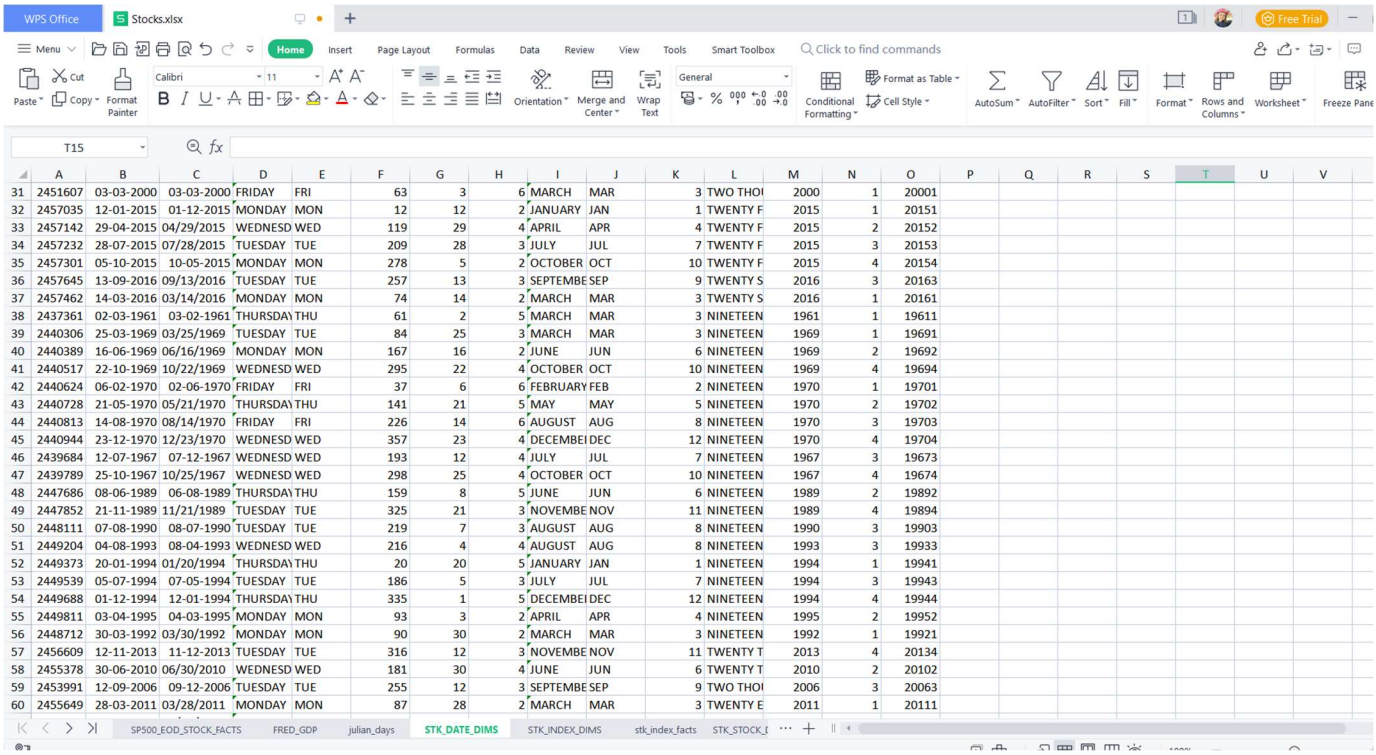
REPORTING, MODELLING AND STORYTELLING

Tableau is a versatile and user-friendly data visualization tool that allows users to connect and extract data from various sources. It provides powerful data modeling and analytics capabilities, enabling users to create insightful visualizations and dashboards. These dashboards can be shared as static files or published to Tableau Server, a secure and collaborative platform. Tableau Server supports automated data refreshes, scheduled reports, and mobile access, enhancing data accessibility and empowering organizations to make data-driven decisions. With its broad range of features, Tableau is a popular choice for effective data visualization and analysis.

Connection:

After extracting the data from the Oracle FIN schema, we saved the files as an Excel workbook. Each table in the schema was represented by a separate sheet in the workbook. The sheets included SP500 stocks data, SP500_EOD_STOCK_FACTS, Julian days, and FRED_GDP. The connection between these sheets was established based on a common field called "Julian_day". Additionally, there were other sheets like SP500_STOCKS, STK_INDEX_DIMS, and STK_INDEX_FACTS, which were connected through a common field called "ticker symbol".Furthermore, the Excel workbook also encompassed crucial data regarding the stock index dimensions. This information comprised the systematic categorization of each company's stock





	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
31	2451607	03-03-2000	03-03-2000	FRIDAY	FRI	63	3	6	MARCH	MAR	3	TWO THOI	2000	1	20001							
32	2457035	12-01-2015	01-12-2015	MONDAY	MON	12	12	2	JANUARY	JAN	1	TWENTY F	2015	1	20151							
33	2457142	29-04-2015	04/29/2015	WEDNESD	WED	119	29	4	APRIL	APR	4	TWENTY F	2015	2	20152							
34	2457232	28-07-2015	07/28/2015	TUESDAY	TUE	209	28	3	JULY	JUL	7	TWENTY F	2015	3	20153							
35	2457301	05-10-2015	10-05-2015	MONDAY	MON	278	5	2	OCTOBER	OCT	10	TWENTY F	2015	4	20154							
36	2457645	13-09-2016	09/13/2016	TUESDAY	TUE	257	13	3	SEPTEMBE	SEP	9	TWENTY S	2016	3	20163							
37	2457462	14-03-2016	03/14/2016	MONDAY	MON	74	14	2	MARCH	MAR	3	TWENTY S	2016	1	20161							
38	2437361	02-03-1961	03-02-1961	THURSDA	THU	61	2	5	MARCH	MAR	3	NINETEEN	1961	1	19611							
39	2440306	25-03-1969	03/25/1969	TUESDAY	TUE	84	25	3	MARCH	MAR	3	NINETEEN	1969	1	19691							
40	2440389	16-06-1969	06/16/1969	MONDAY	MON	167	16	2	JUNE	JUN	6	NINETEEN	1969	2	19692							
41	2440517	22-10-1969	10/22/1969	WEDNESD	WED	295	22	4	OCTOBER	OCT	10	NINETEEN	1969	4	19694							
42	2440624	06-02-1970	02-06-1970	FRIDAY	FRI	37	6	6	FEBRUARY	FEB	2	NINETEEN	1970	1	19701							
43	2440728	21-05-1970	05/21/1970	THURSDA	THU	141	21	5	MAY	MAY	5	NINETEEN	1970	2	19702							
44	2440813	14-08-1970	08/14/1970	FRIDAY	FRI	226	14	6	AUGUST	AUG	8	NINETEEN	1970	3	19703							
45	2440944	23-12-1970	12/23/1970	WEDNESD	WED	357	23	4	DECEMBEI	DEC	12	NINETEEN	1970	4	19704							
46	2439684	12-07-1967	07-12-1967	WEDNESD	WED	193	12	4	JULY	JUL	7	NINETEEN	1967	3	19673							
47	2439789	25-10-1967	10/25/1967	WEDNESD	WED	298	25	4	OCTOBER	OCT	10	NINETEEN	1967	4	19674							
48	2447686	08-06-1989	06-08-1989	THURSDA	THU	159	8	5	JUNE	JUN	6	NINETEEN	1989	2	19892							
49	2447852	21-11-1989	11/21/1989	TUESDAY	TUE	325	21	3	NOVEMBE	NOV	11	NINETEEN	1989	4	19894							
50	2448111	07-08-1990	08-07-1990	TUESDAY	TUE	219	7	3	AUGUST	AUG	8	NINETEEN	1990	3	19903							
51	2449204	04-08-1993	08-04-1993	WEDNESD	WED	216	4	4	AUGUST	AUG	8	NINETEEN	1993	3	19933							
52	2449373	20-01-1994	01/20/1994	THURSDA	THU	20	20	5	JANUARY	JAN	1	NINETEEN	1994	1	19941							
53	2449539	05-07-1994	07-05-1994	TUESDAY	TUE	186	5	3	JULY	JUL	7	NINETEEN	1994	3	19943							
54	2449688	01-12-1994	12-01-1994	THURSDA	THU	335	1	5	DECEMBEI	DEC	12	NINETEEN	1994	4	19944							
55	2449811	03-04-1995	04-03-1995	MONDAY	MON	93	3	2	APRIL	APR	4	NINETEEN	1995	2	19952							
56	2448712	30-03-1992	03/30/1992	MONDAY	MON	90	30	2	MARCH	MAR	3	NINETEEN	1992	1	19921							
57	2456609	12-11-2013	11-12-2013	TUESDAY	TUE	316	12	3	NOVEMBE	NOV	11	TWENTY T	2013	4	20134							
58	2455378	30-06-2010	06/30/2010	WEDNESD	WED	181	30	4	JUNE	JUN	6	TWENTY T	2010	2	20102							
59	2453991	12-09-2006	09-12-2006	TUESDAY	TUE	255	12	3	SEPTEMBE	SEP	9	TWO THOI	2006	3	20063							
60	2455649	28-03-2011	03/28/2011	MONDAY	MON	87	28	2	MARCH	MAR	3	TWENTY E	2011	1	20111							

ANALYTICAL QUERIES WITH INFORMATIVE DATA CHARTS

Query 1

1] "What is the average GDP and how many unique ticker symbols are associated with each company in each sector?"

```
SELECT
   s.GICS_SECTOR,
   s.COMPANY,
   AVG(g.GDP_VALUE) AS average_gdp,
   COUNT(DISTINCT s.TICKER_SYMBOL) AS ticker_symbol_count
FROM
   fin.SP500_STOCKS s
JOIN
   FRED_GDP g ON s.GICS_SECTOR = g.SECTOR
GROUP BY
   s.GICS_SECTOR,
   s.COMPANY;
```

EXPLANATION:

The Chart explains the correlation between average GDP and the count of unique ticker symbols, providing insights into the economic performance and stock market diversity across different sectors.The bar graph displayed utilizes a color-coded scheme for visual representation based on the company. Based on the analysis we can clearly see that the financial sector has the highest performance.

CHART

Query 2

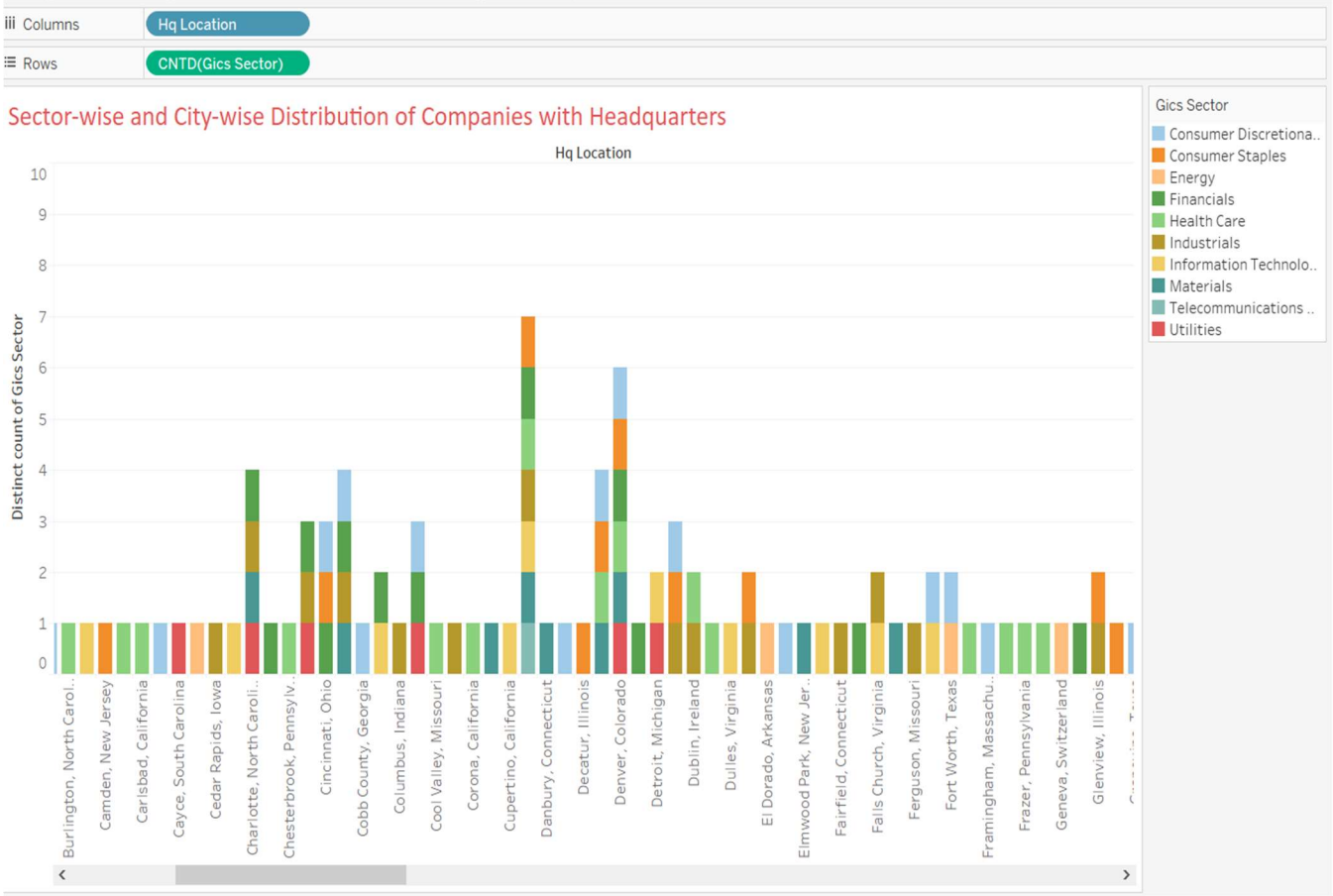
2]What are the sector-wise distributions of companies in the S&P 500 index, and which cities have the highest concentration of companies in each sector?"

```
SELECT
   s.GICS_SECTOR,
   COUNT(s.TICKER_SYMBOL) AS company_count,
   c.HQ_LOCATION
FROM
   SP500_STOCKS s
JOIN
   SP500_EOD_STOCK_FACTS f ON s.TICKER_SYMBOL = f.TICKER_SYMBOL
JOIN
   COMPANY_LOCATIONS c ON s.COMPANY = c.COMPANY
GROUP BY
   s.GICS_SECTOR,
   c.HQ_LOCATION
HAVING
   COUNT(s.TICKER_SYMBOL) > 0
ORDER BY
   s.GICS_SECTOR,
   company_count DESC;
```

EXPLANATION:

The bar chart visually represents the sector-wise distribution of companies, along with the cities where these companies have their headquarters. Each bar corresponds to a specific sector, and its height indicates the number of companies within that sector. The bars are further color-coded to represent different cities, showcasing the concentration of companies in each city within a particular sector. This visualization provides insights into the geographic distribution of companies across sectors, allowing for a quick understanding of the business landscape in different cities and sectors

CHART:



QUERY 3:

3] What are the average closing prices, highest highs, lowest lows,for technology companies for a specific location?

```
SELECT
   s.COMPANY,
   f.TRADE_DATE_STR,
   AVG(f.CLOSE) AS AVERAGE_CLOSE,AVG(f.HIGH) ASAVG_HIGH,AVG(f.LOW) AS
AVG_LOW,
   FROM fin.SP500_STOCKS s
JOIN
  fin.SP500_EOD_STOCK_FACTS f ON s.TICKER_SYMBOL = f.TICKER_SYMBOL
WHERE
    s.GICS_SECTOR = 'Technology'
GROUP BY
   s.COMPANY,
   f.TRADE_DATE_STR
HAVING
   COUNT(f.TICKER_SYMBOL) >= 1
```

```
ORDER BY
    AVERAGE_CLOSE DESC;
```

EXPLANATION:

This chart depicts the average closing prices, highest highs and lowest lows for technology companies. It joins the "SP500_STOCKS" and "SP500_EOD_STOCK_FACTS" tables based on the ticker symbol and filters the results for companies in the technology sector. The data is grouped by company and trade date, with a condition to include only companies with at least one ticker symbol. In this graph from the dropdown I have selected a location named Baltimore,Maryland and the stock performance was compared using average of high,low and close of stocks in the two consecutive years.

CHART:

Query 4:

4]What was the average range of high and low stock prices for companies in the years 2009 and 2010? How did the stock price volatility vary among different companies during this period?

```
SELECT
    s.COMPANY,
    AVG(f.HIGH) AS AVERAGE_HIGH,
    AVG(f.LOW) AS AVERAGE_LOW,
    (AVG(f.HIGH) - AVG(f.LOW)) AS PRICE_VOLATILITY
FROM
```

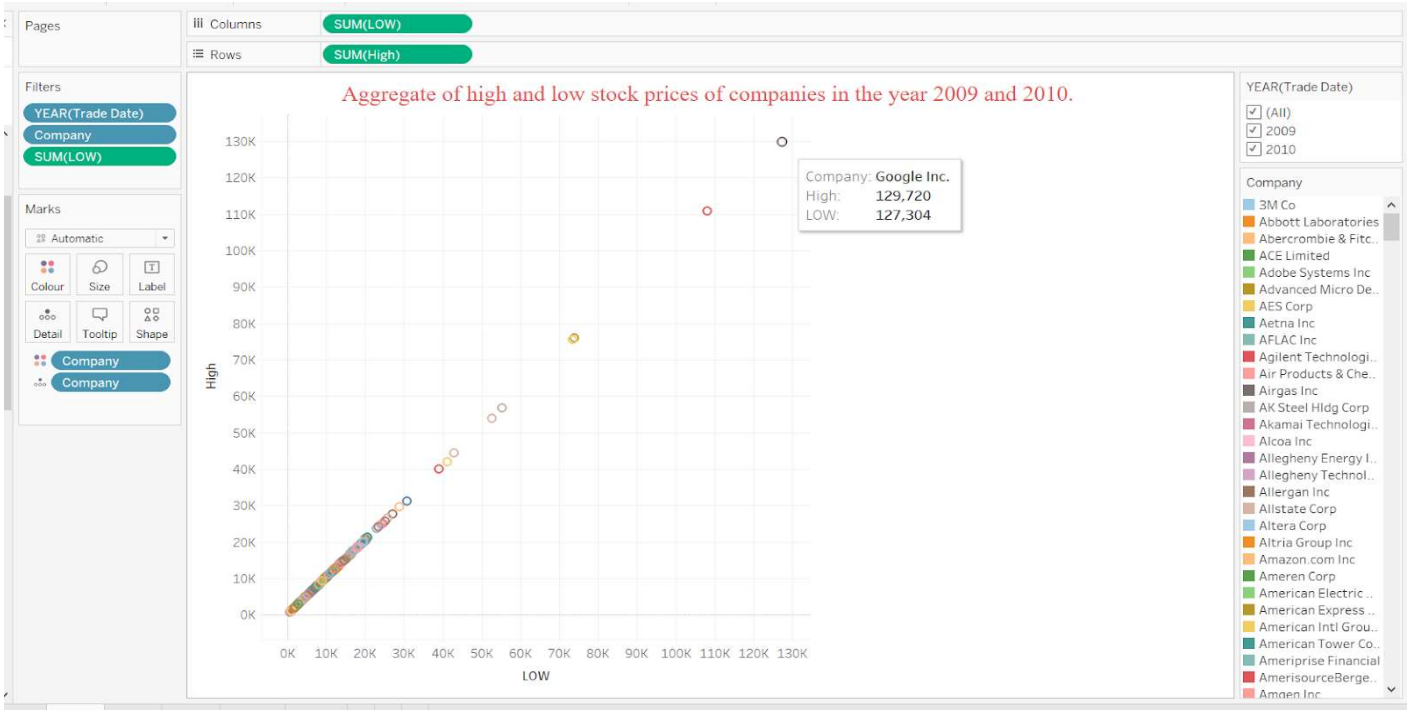
```
    fin.SP500_STOCKS s
JOIN
   fin.SP500_EOD_STOCK_FACTS f ON s.TICKER_SYMBOL = f.TICKER_SYMBOL
WHERE
    YEAR(f.TRADE_DATE) IN (2009, 2010)


GROUP BY
    s.COMPANY
ORDER BY
    PRICE_VOLATILITY DESC;
```

EXPLANATION:

The scatter plot presented below showcases the aggregated high and low stock prices of companies for the years 2009 and 2010. Notably, Google stands out with the highest opening and closing stock prices across both consecutive years.

CHART:



QUERY 5:

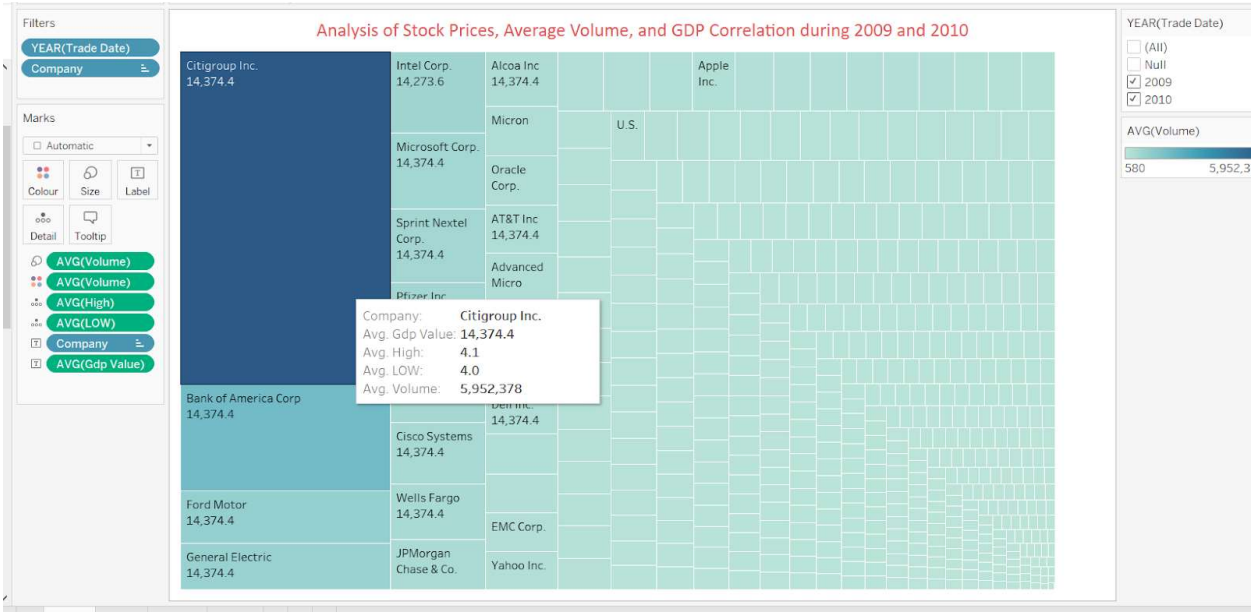
What is the average volume, high, low, and close prices for companies in the years 2009 and 2010?What was their average volume? Additionally, how did the GDP value correlate with the stock market performance during these years?

```
SELECT
    YEAR(f.TRADE_DATE_STR) AS trade_year,
    s.COMPANY,
    COUNT(DISTINCT f.TICKER_SYMBOL) AS unique_tickers,
    AVG(f.VOLUME) AS average_volume,
    AVG(f.HIGH) AS average_high,
    AVG(f.LOW) AS average_low,
    AVG(f.CLOSE) AS average_close,
    g.GDP_VALUE
FROM
    fin.SP500_EOD_STOCK_FACTS f
JOIN
    fin.FRED_GDP g ON f.TRADE_DATE_STR = g.GDP_DATE
JOIN
    fin.SP500_STOCKS s ON f.TICKER_SYMBOL = s.TICKER_SYMBOL
WHERE
    YEAR(f.TRADE_DATE_STR) IN (2009, 2010)
GROUP BY
    YEAR(f.TRADE_DATE_STR),
    s.COMPANY,
    g.GDP_VALUE
ORDER BY
    trade_year,
    average_volume DESC;
```

EXPLANATION:

The tree map visualization illustrates the distribution of stock trading volume among different companies during 2009 and 2010. Notably, Citigroup Inc. emerges as the company with the highest volume with average high as 4.1 and average low as 4.0, indicating its significant market activity during this period.

CHART:



QUERY 6:

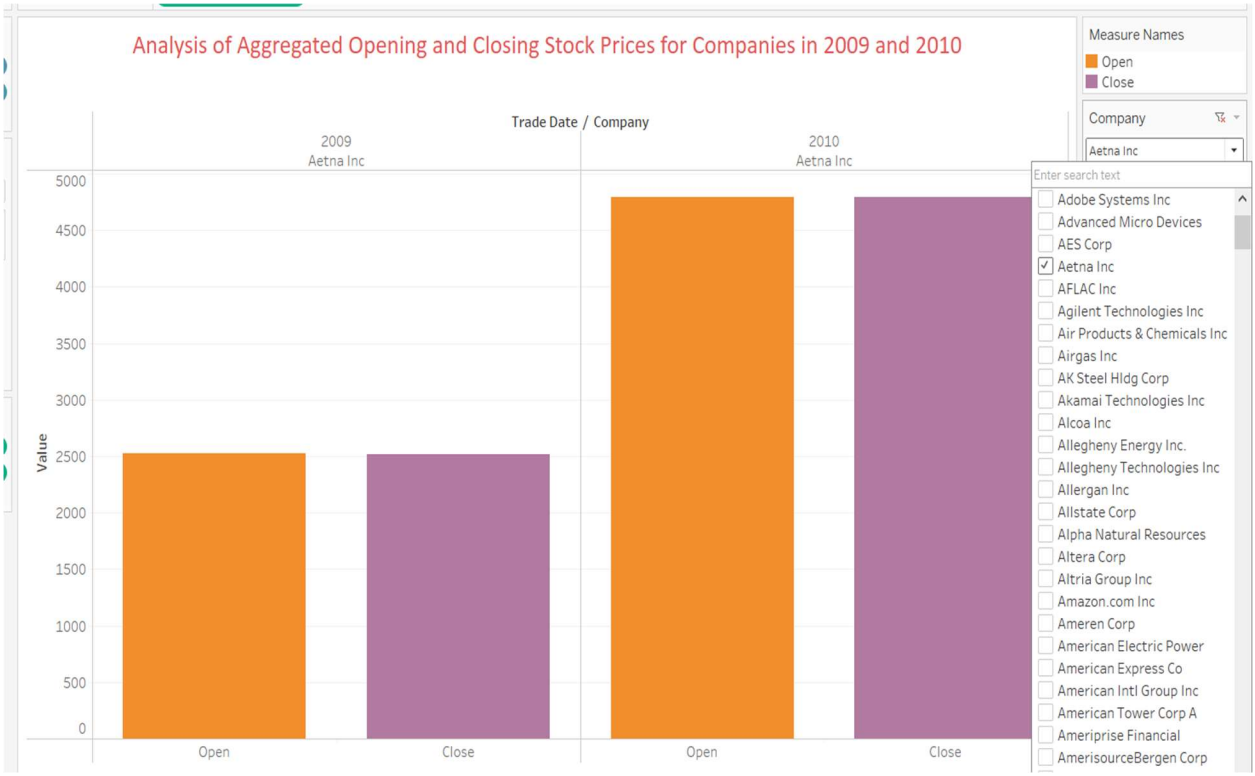
6] What were the total values of open and close stock prices aggregated for companies during the years 2009 and 2010? How did these aggregate values differ from 2009 to 2010 for a particular company?

```
SELECT
   s.COMPANY,
   SUM(f.OPEN) AS AGGREGATE_OPEN,
   SUM(f.CLOSE) AS AGGREGATE_CLOSE
FROM
   SP500_STOCKS s
JOIN
   SP500_EOD_STOCK_FACTS f ON s.TICKER_SYMBOL = f.TICKER_SYMBOL
WHERE
   YEAR(f.TRADE_DATE) IN (2009, 2010)
GROUP BY
   s.COMPANY;
```

EXPLANATION:

The bar chart analysis illustrates the aggregated opening and closing stock prices for a specific company during the years 2009 and 2010. It provides a visual comparison of the average prices for both opening and closing values, allowing for an examination of the company's stock performance over the two-year period. In this chart I have selected a particular company named Aetna.inc and the stock performance has increased from the year 2009 to 2010.

CHART:



QUERY 7:

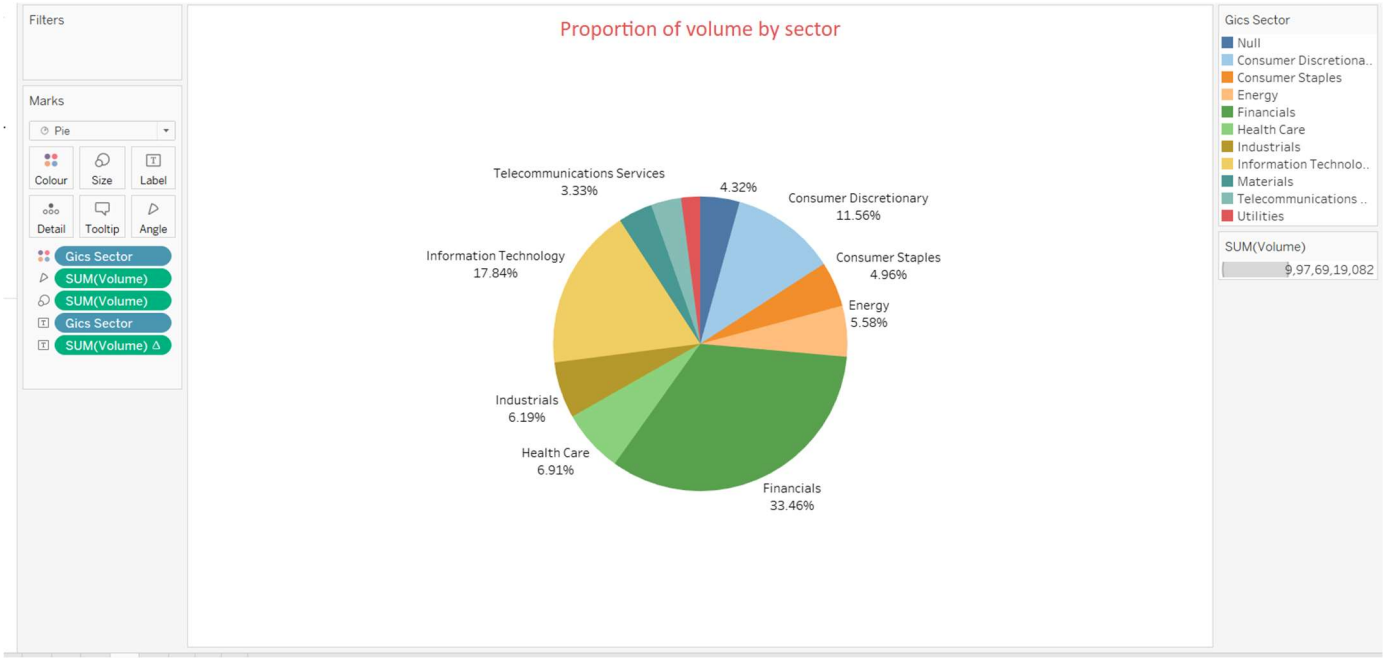
7]What is the distribution of trading volume among different sectors in the stock market? How does the proportion of trading volume vary across sectors, and which sectors exhibit the highest volume proportion?

```
SELECT
   s.GICS_SECTOR,
   SUM(f.VOLUME) AS TOTAL_VOLUME,
   (SUM(f.VOLUME) / (SELECT SUM(VOLUME) FROM SP500_EOD_STOCK_FACTS)) AS
VOLUME_PROPORTION
FROM
   SP500_STOCKS s
JOIN
   SP500_EOD_STOCK_FACTS f ON s.TICKER_SYMBOL = f.TICKER_SYMBOL
GROUP BY
   s.GICS_SECTOR;
```

EXPLANATION:

The pie chart visualization illustrates the distribution of trading volume among different sectors in the stock market. It highlights the proportion of volume allocated to each sector, providing insights into the relative importance of sectors in terms of trading activity. From the chart, we can observe that financial sector have the highest volume proportion followed by information technology etc indicating their significance and potential impact on market dynamics.

CHART



QUERY 8:

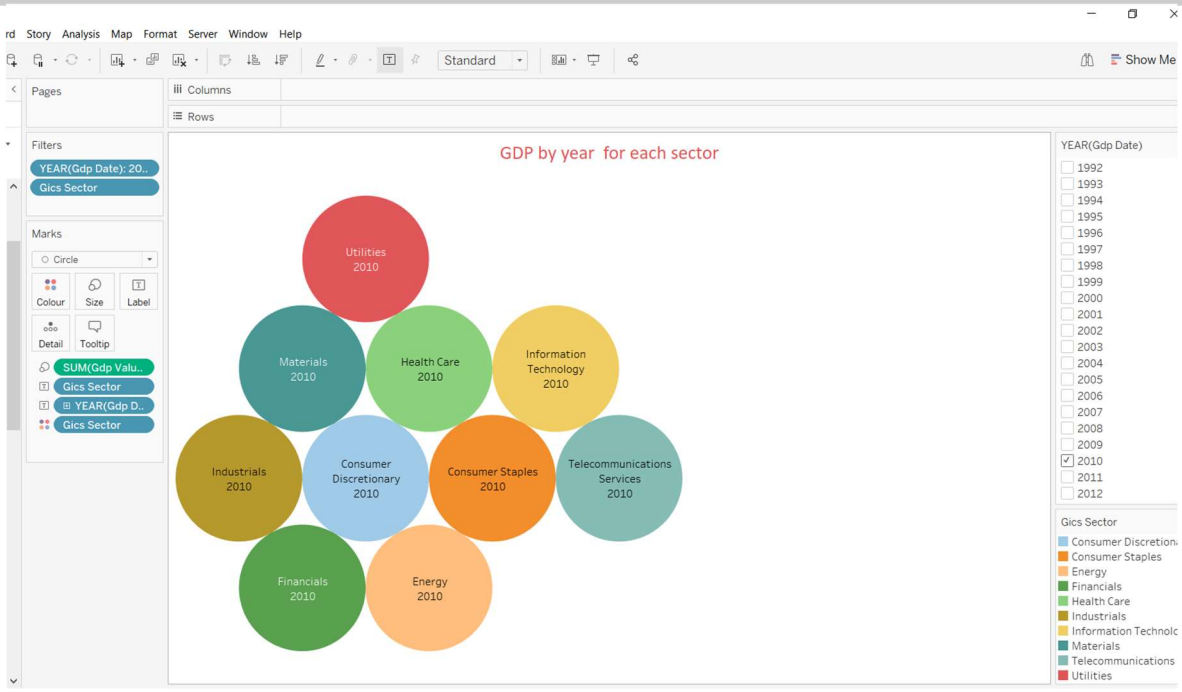
8]What is the distribution of GDP across different sectors over the years? How does the GDP vary among sectors, and which sectors demonstrate the highest contributions to the overall GDP?

```
SELECT
   g.CAL_QUARTER_KEY AS YEAR,
   s.GICS_SECTOR,
   SUM(g.GDP_VALUE) AS TOTAL_GDP
FROM
   FRED_GDP g
JOIN
   fin.SP500_STOCKS s ON g.JULIAN_DAY = s.JULIAN_DAY
GROUP BY
   g.CAL_QUARTER_KEY,
   s.GICS_SECTOR
ORDER BY
   g.CAL_QUARTER_KEY;
```

EXPLANATION:

The bubble chart visualization depicts the distribution of GDP across different sectors over the years. It showcases the relative size of bubbles representing each sector's GDP contribution, allowing for a visual comparison of sector-wise economic performance. The chart provides insights into sectors that exhibit larger bubbles, indicating their significant contributions to the overall GDP.

CHART



QUERY 9:

9] List the current, prior and next closing prices for a specific stock in the S&P 500 index chronologically, along with the elapsed hours between each closing price.

```
SELECT
  x.TRADE_DATE_STR,
  x.CLOSE AS current_closing_price,
  LAG(x.CLOSE) OVER (ORDER BY TO_DATE(x.TRADE_DATE_STR, 'MM/DD/YYYY'))
AS prior_closing_price,
  LEAD(x.CLOSE) OVER (ORDER BY TO_DATE(x.TRADE_DATE_STR, 'MM/DD/YYYY'))
AS next_closing_price,
  ROUND((LEAD(TO_DATE(x.TRADE_DATE_STR, 'MM/DD/YYYY')) OVER (ORDER BY
TO_DATE(x.TRADE_DATE_STR, 'MM/DD/YYYY')) - TO_DATE(x.TRADE_DATE_STR,
'MM/DD/YYYY')) * 24, 1) AS elapsed_hours
FROM
  FIN.SP500_EOD_STOCK_FACTS x
WHERE
  x.TICKER_SYMBOL = 'AXP'
ORDER BY
  TO_DATE(x.TRADE_DATE_STR, 'MM/DD/YYYY');
```

EXPLANATION:

The bar graph visualization displays the current, prior, and next closing prices of a specific stock in the S&P 500 index over time, with the elapsed hours between each closing price. It provides a clear comparison of the stock's performance and allows for easy identification of price trends and the time intervals between them.

CHART:

QUERY 10:

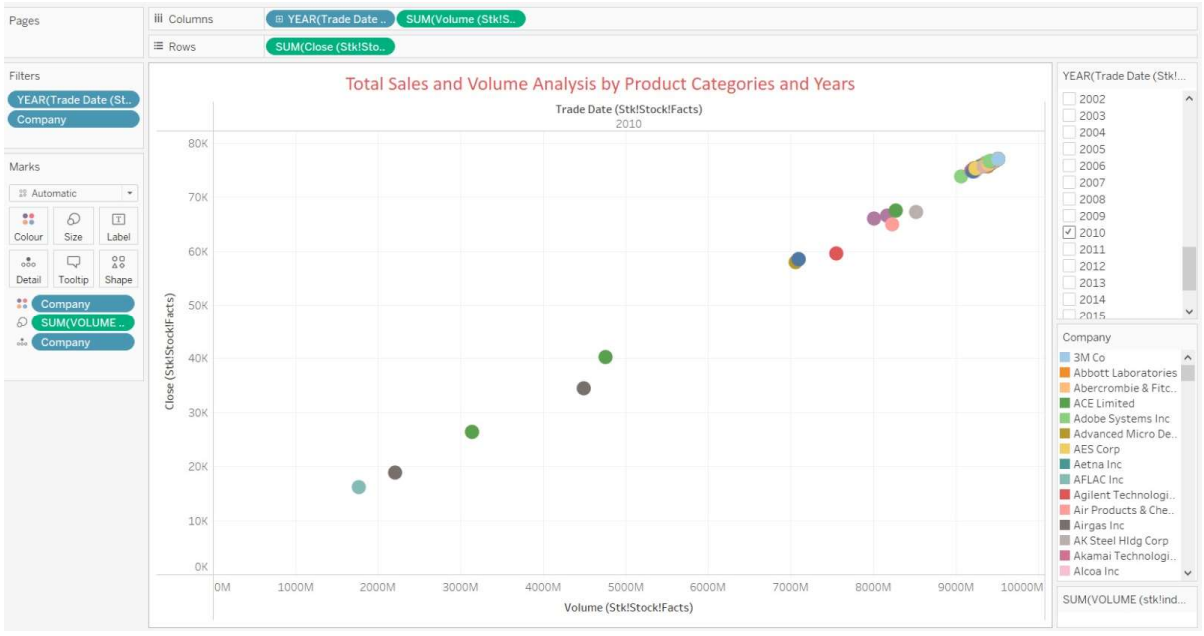
10] Use ROLLUP function to analyze the total sales and profits for different levels of product categories and years in the financial database.

```
SELECT
  sdim.CO_NAME AS company_name,
  ddim.YEAR_NAME AS year,
  SUM(sfact.CLOSE) AS total_sales,
  SUM(sfact.VOLUME) AS total_volume
FROM
  FIN.STK_STOCK_FACTS sfact
  INNER JOIN FIN.STK_STOCK_DIMS sdim ON sfact.STOCK_KEY = sdim.STOCK_KEY
  INNER JOIN FIN.STK_DATE_DIMS ddim ON sfact.JULIAN_DAY_KEY =
ddim.JULIAN_DAY_KEY
GROUP BY ROLLUP (sdim.CO_NAME, ddim.YEAR_NAME)
ORDER BY sdim.CO_NAME, ddim.YEAR_NAME;
```

EXPLANATION:

The scatter plot visualization displays the relationship between two variables, where each data point represents a combination of product categories and years. The x-axis represents total sales, and the y-axis represents total volume, allowing for the analysis of how sales and volume are distributed across different categories and years

CHART:



QUERY 11:

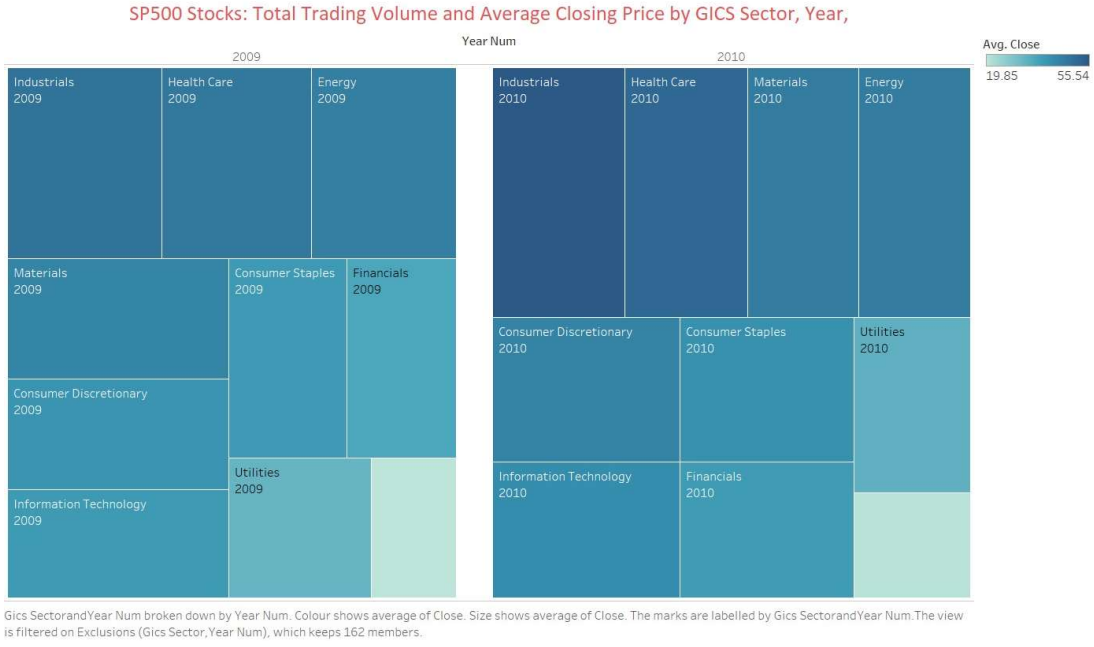
11] What is the total trading volume and average closing price for stocks in the SP500_STOCKS table, considering all possible combinations of GICS sector, year, and month, including subtotals?

```
SELECT
  NVL(s.GICS_SECTOR, 'All Sectors') AS sector,
  EXTRACT(YEAR FROM s.DATE_ADDED) AS year,
  EXTRACT(MONTH FROM s.DATE_ADDED) AS month,
  SUM(f.VOLUME) AS total_volume,
  ROUND(AVG(f.CLOSE), 2) AS avg_closing_price
FROM
  FIN.SP500_STOCKS s
  INNER JOIN FIN.SP500_EOD_STOCK_FACTS f ON s.TICKER_SYMBOL = f.TICKER_SYMBOL
GROUP BY CUBE(s.GICS_SECTOR, EXTRACT(YEAR FROM s.DATE_ADDED), EXTRACT(MONTH FROM s.DATE_ADDED))
ORDER BY sector, year, month;
```

EXPLANATION:

The tree map visualization displays the distribution of total trading volume for SP500 stocks across different GICS sectors, years. The size of each rectangle represents the trading volume, allowing for a quick comparison of volume proportions within and across sectors, years. In both the years 2009 and 2010 the industrial sector had the highest average close.

CHART:



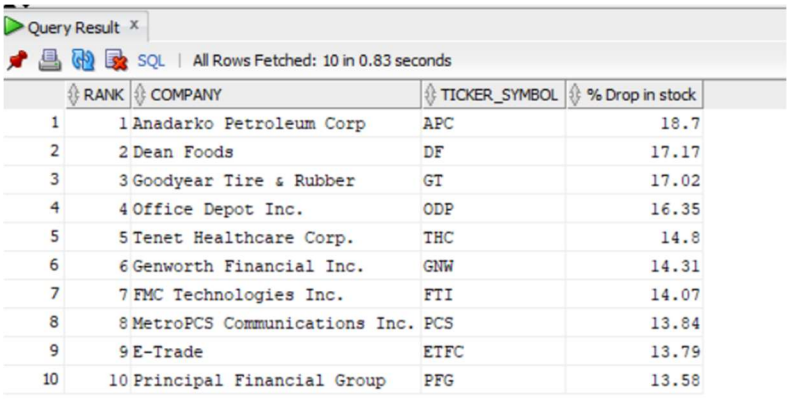
OTHER INTERESTING QUERIES

1. **Get the top 10 businesses whose stock value significantly dropped on a given day.**

QUERY 1

```
Select a.rank,b.company,a.ticker_symbol,a.stock_value_change "% Drop
 in stock" From
 (Select rownum rank,a.* From
 (Select ticker_symbol, round(max(((open - close)/open)*100),2) as
 stock_value_change from fin.sp500_eod_stock_facts SE
 group by ticker_symbol
 order by stock_value_change
 desc) a where rownum<11) a
 inner join FIN.sp500_stocks b on
 b.ticker_symbol=a.ticker_symbol order by
 a.stock_value_change desc;
```

OUTPUT-



Query Result | SQL | All Rows Fetched: 10 in 0.83 seconds

	RANK	COMPANY	TICKER_SYMBOL	% Drop in stock
1	1	Anadarko Petroleum Corp	APC	18.7
2	2	Dean Foods	DF	17.17
3	3	Goodyear Tire & Rubber	GT	17.02
4	4	Office Depot Inc.	ODP	16.35
5	5	Tenet Healthcare Corp.	THC	14.8
6	6	Genworth Financial Inc.	GNW	14.31
7	7	FMC Technologies Inc.	FTI	14.07
8	8	MetroPCS Communications Inc.	PCS	13.84
9	9	E-Trade	ETFC	13.79
10	10	Principal Financial Group	PFG	13.58

2. **Provide a list of the top 10 companies with the highest trading volume of stocks on a single day.**

QUERY 2

```
Select a.*,b.TRADE_DATE_STR
 "date",c.company from
 (select rownum rank,a.* from
 (select ticker_symbol, max(volume)
 volume_traded from
 fin.sp500_eod_stock_facts
 group by
 ticker_symbol order
 by 2 desc) a
```

```
 where rownum<11) a
 inner join fin.sp500_eod_stock_facts b on a.volume_traded=b.volume
inner join FIN.sp500_stocks c on
a.ticker_symbol=c.ticker_symbol order by rank
```

OUTPUT-

Script Output | Query Result | Query Result 1

SQL | All Rows Fetched: 10 in 0.257 seconds

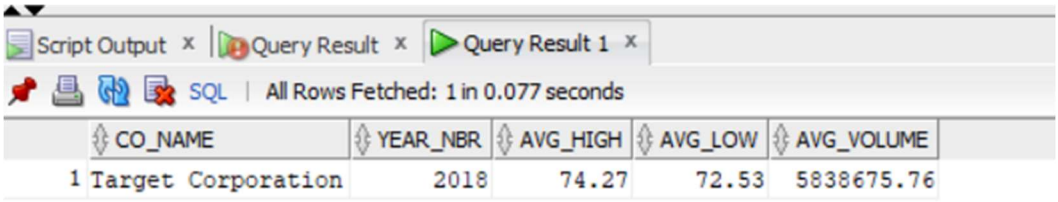
	RANK	TICKER_SYMBOL	VOLUME_TRADED	date	COMPANY
1	1	C	21474836	12/18/2009	Citigroup Inc.
2	2	BAC	12267912	12/4/2009	Bank of America Corp
3	3	F	4336590	3/31/2010	Ford Motor
4	4	ETFC	3561277	9/18/2009	E-Trade
5	5	Q	3264504	4/22/2010	(null)
6	6	MSFT	2817609	10/23/2009	Microsoft Corp.
7	7	GE	2689680	9/16/2009	General Electric
8	8	WFC	2545757	12/15/2009	Wells Fargo
9	9	S	2511273	2/10/2010	Sprint Nextel Corp.
10	10	BSX	2446275	3/15/2010	Boston Scientific

3. The average high, low, volume of a Target corporation over a period of time ?

QUERY 3

```
SELECT
sd.co_name,
dd.year_nbr,
ROUND(AVG(sf.high),2) AS AVG_HIGH,
ROUND(AVG(sf.low),2) AS AVG_LOW,
ROUND(AVG(sf.volume),2) AS
AVG_VOLUME FROM FIN.stk_stock_dims
sd
INNER JOIN
FIN.stk_stock_facts
sf
ON sf.stock_key=sd.stock_key
INNER JOIN FIN.stk_date_dims dd
ON dd.julian_day_key=sf.julian_day_key
WHERE
sf.stock_key=100015 and
dd.year_nbr=2018
GROUP BY
sd.co_name,dd.year_nb
```

OUTPUT-



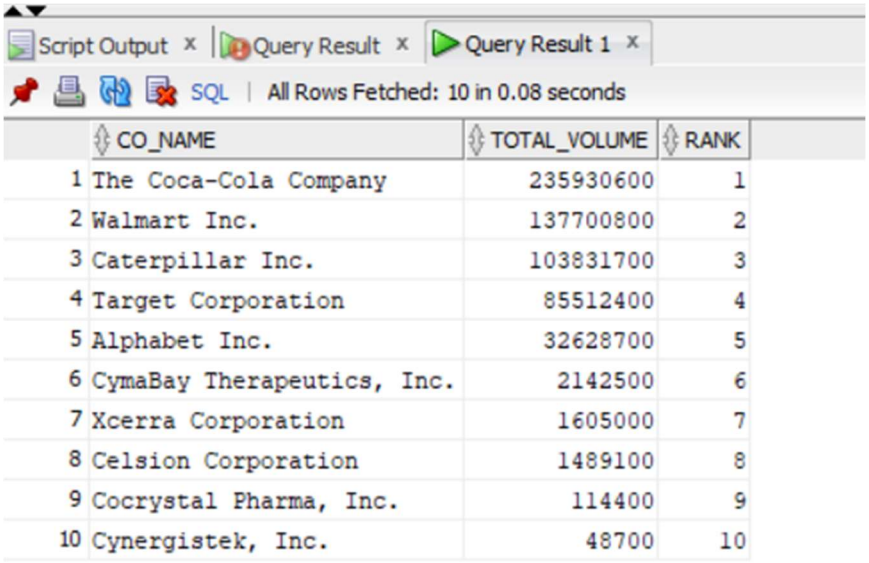
	CO_NAME	YEAR_NBR	AVG_HIGH	AVG_LOW	AVG_VOLUME
1	Target Corporation	2018	74.27	72.53	5838675.76

4. **Rank the companies according to the total volume of stock traded over a given year and month.**

QUERY 4

```
SELECT
 sd.co_name,
 SUM(sf.volume) as total_volume,
 ROW_NUMBER() OVER (ORDER BY SUM(sf.volume) DESC) AS
 RANK FROM
 FIN.stk_stock_dims
 sd INNER JOIN
 FIN.stk_stock_facts
 sf
 ON sf.stock_key=sd.stock_key
 INNER JOIN FIN.stk_date_dims dd
 ON dd.julian_day_key=sf.julian_day_key
 WHERE
 dd.year_nbr=2016
 and dd.month_name like '%JULY%'
 GROUP BY sd.co_name,
 dd.month_name;
```

OUTPUT



	CO_NAME	TOTAL_VOLUME	RANK
1	The Coca-Cola Company	235930600	1
2	Walmart Inc.	137700800	2
3	Caterpillar Inc.	103831700	3
4	Target Corporation	85512400	4
5	Alphabet Inc.	32628700	5
6	CymaBay Therapeutics, Inc.	2142500	6
7	Xcerra Corporation	1605000	7
8	Celsion Corporation	1489100	8
9	Cocrystal Pharma, Inc.	114400	9
10	Cynergistek, Inc.	48700	10

5. **The maximum fluctuation in the high and low value of stock of companies in a particular year and month.**

QUERY 5

```
SELECT
 sd.co_name,
 dd.year_nbr,
 dd.month_nam
 e,
 MAX(sf.HIGH)-MIN(sf.HIGH) AS
 MAX_FLUCTUATION_HIGH, MAX(sf.LOW)-
 MIN(sf.LOW) AS MAX_FLUCTUATION_LOW FROM
 FIN.stk_stock_dims
 sd INNER JOIN
 FIN.stk_stock_facts
 sf
 ON sf.stock_key=sd.stock_key
 INNER JOIN FIN.stk_date_dims dd
 ON dd.julian_day_key=sf.julian_day_key
 WHERE dd.YEAR_NBR=2018
 AND dd.MONTH_NAME LIKE '%MAY%'
 GROUP BY sd.co_name,dd.year_nbr,dd.month_name
 ORDER BY sd.co_name;
```

OUTPUT

Script Output | Query Result | Query Result 1

SQL | All Rows Fetched: 10 in 0.086 seconds

	CO_NAME	YEAR_NBR	MONTH_NAME	MAX_FLUCTUATION_HIGH	MAX_FLUCTUATION_LOW
1	Alphabet Inc.	2018	MAY	87.460083	96.910034
2	Caterpillar Inc.	2018	MAY	17.170013	20.550003
3	Celsion Corporation	2018	MAY	0.49	0.52
4	Cocrystal Pharma, Inc.	2018	MAY	0.73	0.73
5	CymaBay Therapeutics, Inc.	2018	MAY	2.83	2.21
6	Cynergistek, Inc.	2018	MAY	1.09	1.25
7	Target Corporation	2018	MAY	7.619995	7.970001
8	The Coca-Cola Company	2018	MAY	1.360001	1.25
9	Walmart Inc.	2018	MAY	5.580001	4.569999
10	Xcerra Corporation	2018	MAY	2.06	2

6. What is the correlation between the daily trading volume and closing price of stocks in the SP500_EOD_STOCK_FACTS table?

   QUERY 6

```
SELECT
  CORR(VOLUME, CLOSE) AS correlation
FROM
  FIN.SP500_EOD_STOCK_FACTS;
```

OUTPUT

Query Result

SQL | All Rows Fetched: 1 in 0.413 seconds

	CORRELATION
1	-0.089909800699576153239580458284303628187 5

CONCLUSION:

For storing data from various sources and utilizing it for online analytical processing, data warehouses have emerged as a crucial component. We used a data warehouse design specifically for stock market analysis in our project. One or more fact tables, as well as several dimensional tables revolving around them, are included in this design. We have created analytical queries to extract the data and answer various business questions once the structure and relationships have been established. We used the visualization tool Tableau to improve the presentation of the results and offer a more thorough and intuitive understanding. We have visualized the results of our business questions using Tableau, gaining insights and responding to intriguing questions. Our project shows how businesses gather and use data as a proof of concept. Our project serves as a proof of concept, demonstrating how businesses gather and warehouse their data before using it for online analytical processing to produce important insights and guide decision-making.

Data warehouses are very important given the current trend of data proliferation and the increasing reliance on data-driven decision-making. They are essential in managing the enormous amount of data generated and effectively utilizing it to support crucial business decisions.

REFERENCES:

[1]https://www.researchgate.net/publication/326701422_A_Data_Warehouse_Based_Modelling_Technique_for_Stock_Market_Analysis
[2] https://www.kimballgroup.com/data-warehouse-business-intelligence-resources/kimball-techniques/dimensional-modeling-techniques/