Data Warehouse – Stock Market Analysis

GROUP 1 – PANDA

Sai Yarra

Ipsa Bhatt

Payal Kaur

Saroni Sinha

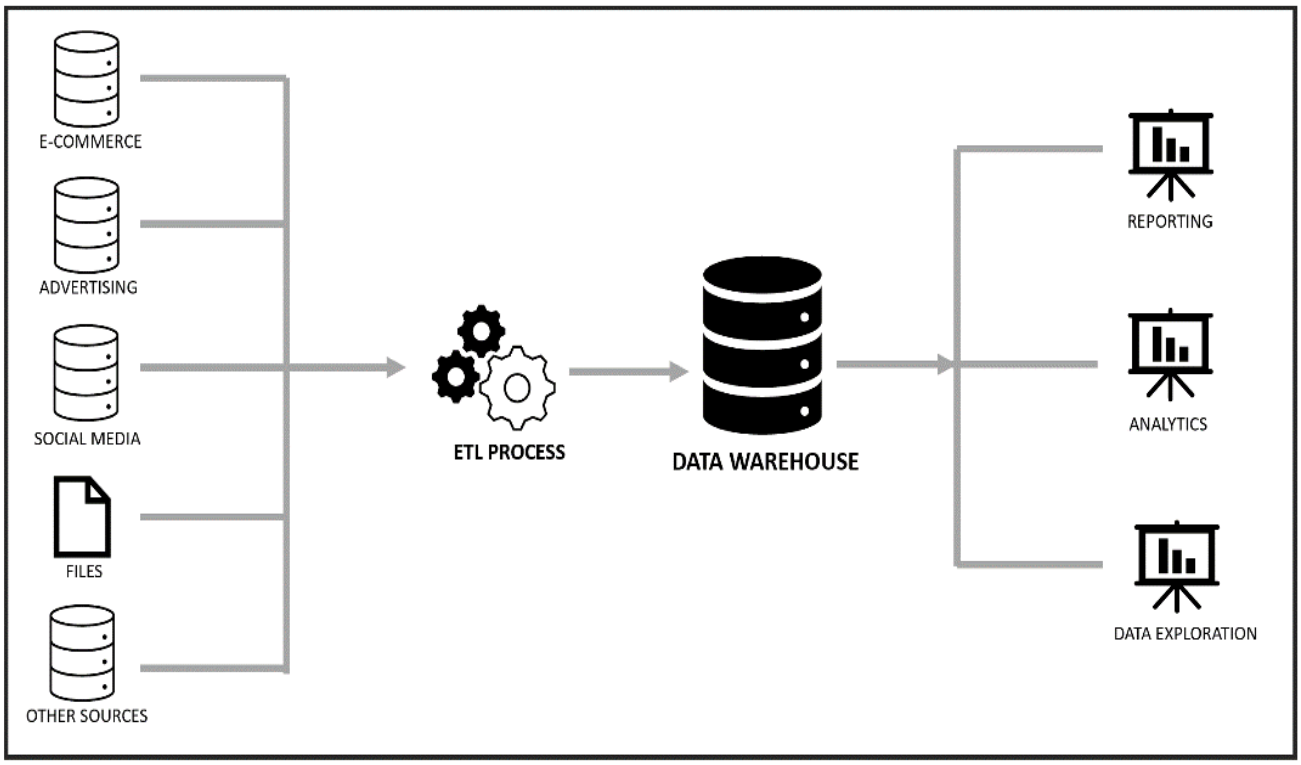
Purva Khandelwal

CONTENTS

EXECUTIVE SUMMARY

Data Warehouse is a subject oriented, integrated, nonvolatile, time variant collection of data in support of managements decisions [1]. It is used by businesses in their business intelligence systems as well as for their decisions support systems. It organizes, stores, and analyzes huge amount of transactional day to day data in a systematic way which is easy to feed into analytical platforms that help in deriving insights and recommendations. Data is usually stored in de-normalized form for higher performance and easy retrieval. Data Warehouse uses Extract-Transform-Load (ETL) process for data cleaning and integrating it from various databases and from online analytical processing (OLAP). OLAP gives ability to analyses business data from different(multi) angles and provides capability for complex calculations, summarization, consolidation, aggregation, and trend analysis. OLAP enables end users to perform ad hoc analysis of data in multiple forms thus providing them with meaningful insights for making better decisions.

Stock market is where shares or stocks of various listed companies are traded. Share prices of stocks vary during the day and throughout time depending on different factors. Few people foresee rising stock value so they buy them at current price to sell it in future gaining profit whereas few people assume price will fall so they sell at current price and in future may buy at lower prices. There are many factors that impact the prices of stock, the basic amongst them is balance between supply and demand that determines its prices on daily basis. This project aims at building warehouse model and analyze various parameters of stock market exchange data of various companies. It inculcates the information about stock trading date, volume, active & inactive stocks. It also throws light on the GDP per sector contribution and stock all time high, low, and open close values.



PROBLEM STATEMENT

Analysis of stock market has taken a greater significance in over the last decade. Our trading markets have changed substantially in the past two decades and are far different from the markets of a generation ago. Today's markets are dominated by electronic trading and sophisticated technologies. Policy makers and investors are aware of the usefulness of the stock market in mobilizing the resources [2].

With the increased participation in the stock market both at, the institutional and individual, domestic and foreign level has brought the Stock Market at new heights. Volatility assumes a greater significance when it comes to investments in stock market. Volatility in stock prices adversely affects the individual earnings and economy's health. It creates the environment of uncertainties and thus hampers the investments [3].

Intelligent investing also implies having the sufficient knowledge of the market as well as the stocks trading in the market. Various categories of stocks from different sectors are listed and traded in the market. Where in the market large cap stocks from different sectors like banking, power, infrastructure, Fast Moving Consumer Goods, Automobile etc. are trading, at the same time there is a market for mid cap and the small cap stocks too from the same sectors [2]. Not all the stocks of all the sectors portray the same behavior in terms of risk and return. Performance of different sector stocks varies and depends on several factors both common and specific to that sector. There is a need of proper planning so that the investments not only withstand the adverse impacts but could utilize such situations for value appreciation. Risk can be managed efficiently only when the investors understand the volatility behavior of individual stocks and the interrelationship of the assets [2].

With the data used in this case study, a Machine Learning/Data Scientist can study the data and implement several algorithms that can extract patterns and provide meaningful outcome from the historical data[3] The current research attempts to apply the various steps to see the open, close, high, low, max value, average of stocks in day-to-day basis.

LITERATURE REVIEW

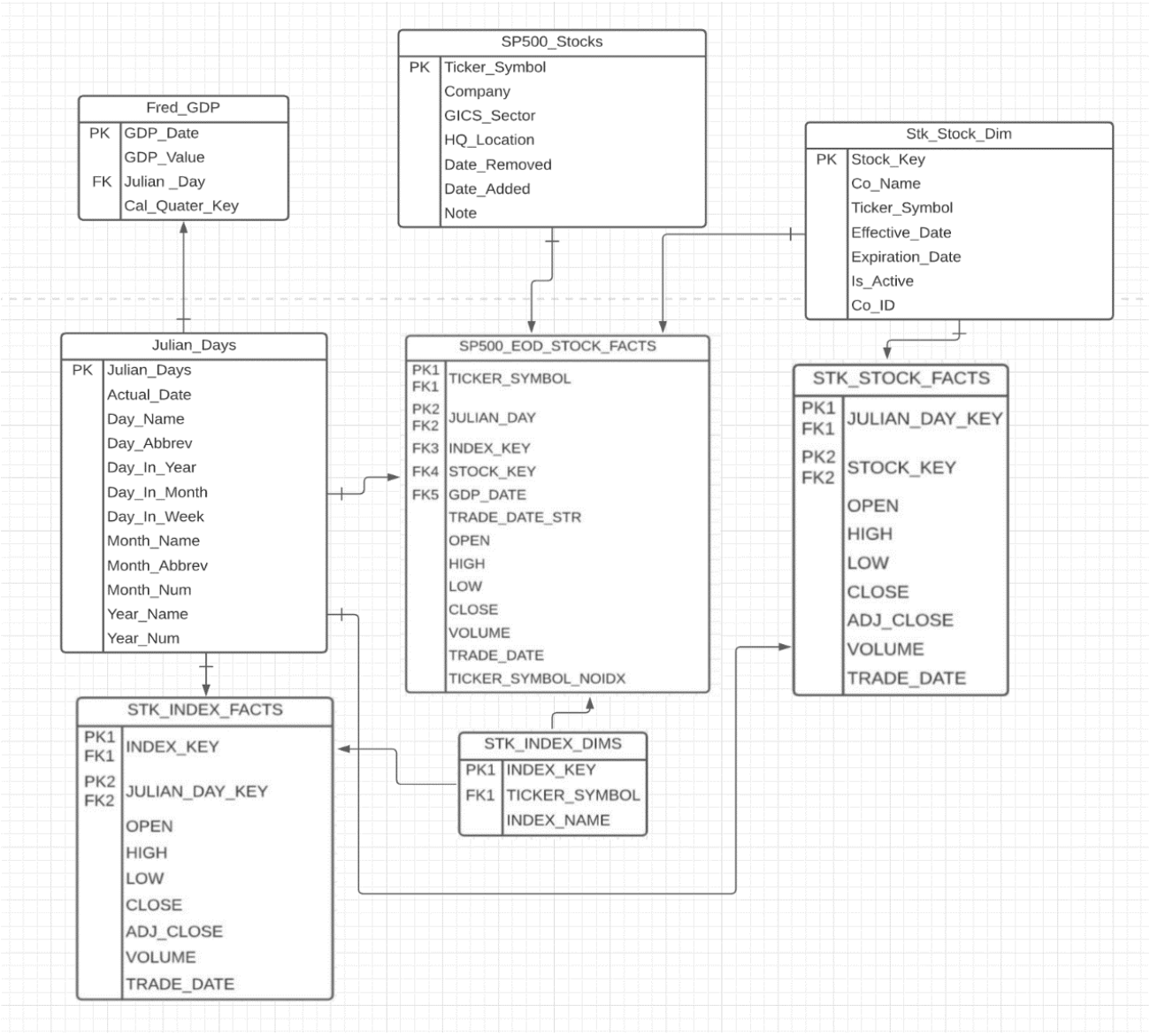
The project aims at building warehouse model and analyzes various parameters of stock market exchange data of various companies. It specifies the information about stock trading date, volume, active & inactive stocks. It also throws light on the GDP per sector contribution and stock high, low and open close values as well. We referred research paper "***A Data Warehouse based Modelling Technique for Stock Market Analysis***", where the author discusses about identifying a warehouse model to build analytical framework and analyze different parameters which directly impact the changes of share market. It also identified different applications of the analytical model for forecasting information help in decision making [4]. The paper helped us in understanding some basic parameters which impact stock market. Apart from this research paper we also read "***The Data Warehouse Toolkit***" book and course material for understanding the concepts used in building the warehouse and performing analysis.

DATA COLLECTION AND PREPARATION

The data used in our project has been collected from various tables in the **FIN** schema. Firstly, the data from the table has been exported to excel files. Next the data has been cleaned in excel and made ready for the purpose of visualization. As the dataset contained null values, while importing data into the visualization tool Tableau, we have further omitted null values from the dataset. We have performed joins between tables on keys and common fields to build proper visualizations of the dataset spanning multiple tables, that will help us address the business queries covered later in the report.

DATABASE DESIGN

ER Diagram shows relationships among various entity sets that are stored in the database FIN.



TABLES USED

1. **JULIAN_DAYS**

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. **STK_DATE_DIMSv1**

NAME	DATATYPE	KEY CONSTRAINT
JULIAN DAY KEY	NUMBER(9)	PRIMARY KEY
CAL DATE	DATE	
CAL TEXT	VARCHAR(10)	
DAY NAME	VARCHAR(10)	
DAY ABBREV	VARCHAR(3)	
DAY IN YEAR	NUMBER(3)	

DAY IN MONTH	NUMBER(3)	
DAY IN WEEK	NUMBER(3)	
MONTH NAME	VARCHAR(10)	
MONTH ABBREV	VARCHAR(3)	
MONTH NBR	NUMBER(2)	
YEAR NAME	VARCHAR(40)	
YEAR NUM	NUMBER(4)	

4. 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)	

5. SP500_EOD_STOCK_FACTSv1

NAME	DATATYPE	KEY CONSTRAINT
TRADE DATE STR	VARCHAR(10)	COMPOSITE PRIMARY KEY
TICKER SYMBOL	VARCHAR(10)	
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)	
JULIAN DAYS JULIAN DAY	NUMBER(12)	COMPOSITE PRIMARY KEY

6. 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

7. 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)	

8. STK_INDEX_ FACTS

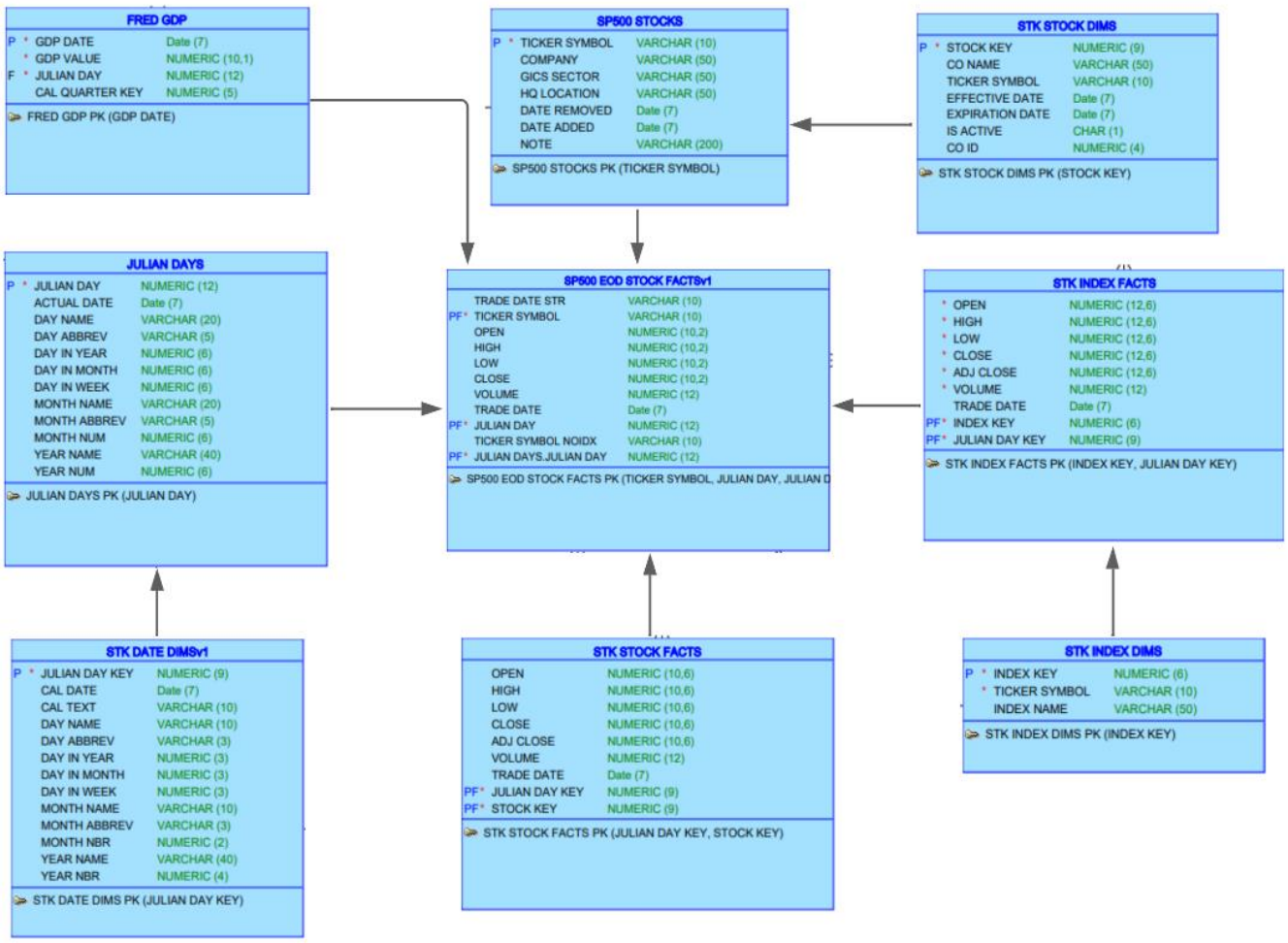
NAME	DATATYPE	KEY CONSTRAINT
OPEN	NUMBER(12,6)	
HIGH	NUMBER(12,6)	
LOW	NUMBER(12,6)	
CLOSE	NUMBER(12,6)	
ADJ CLOSE	NUMBER(12,6)	
VOLUME	NUMBER(12)	
TRADE DATE	DATE	
INDEX KEY	NUMBER(6)	COMPOSITE PRIMARY KEY
JULIAN DAY KEY	NUMBER(9)	COMPOSITE PRIMARY KEY

9. STK INDEX DIMS

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

DIMENSIONAL MODEL

The Dimensional model is a logical design technique used for building data warehouses. Dimensional models identify the data needed for analysis and decision making by the businesses. Every model comprises of one or more table called Fact table and a set of smaller tables called dimension tables. Each dimension table has a single part primary key that relates to exactly one unit of the multi-part key in the fact table.



FACT TABLE – Fact table store important business measurements at the intersection of all radiating dimensions. We have used following fact tables:

SP500 EOD STOCK FACTSv1 – consists of every changing stock price value and their respective ticker symbols of 500 largest companies listed on stock exchanges in the United States. Every grain of the table includes trade date, stock opening, high low and open close values, their volume traded along with date and time dimension details.

STK STOCK FACTS - consists of every changing stock price value and their respective ticker symbols. Every grain of the table includes trade date, stock opening, high, low and open close values, their volume traded along with date and time dimension along with their respective stock primary key.

STK INDEX FACTS - consists of every changing stock price value and their respective ticker symbols. Every grain of the table includes trade date, stock opening, high, low and adjusted closing values, their volume traded along with date and time dimension details.

FRED GDP - consist of GDP value and its date details.

DIMENSION TABLE – Dimension table store discrete categories that identify the granularity of the measurements in fact table. We have following tables:

JULIAN DAYS – Every grain of this table includes various date and time dimension details of every stock change.

STK INDEX DIMS - Every grain of this table includes index number of a particular stock along with its ticker symbol.

STK STOCK DIMS - Every grain of this table includes company name, its ticker symbol, its date of being effective in the market along with its expiration date.

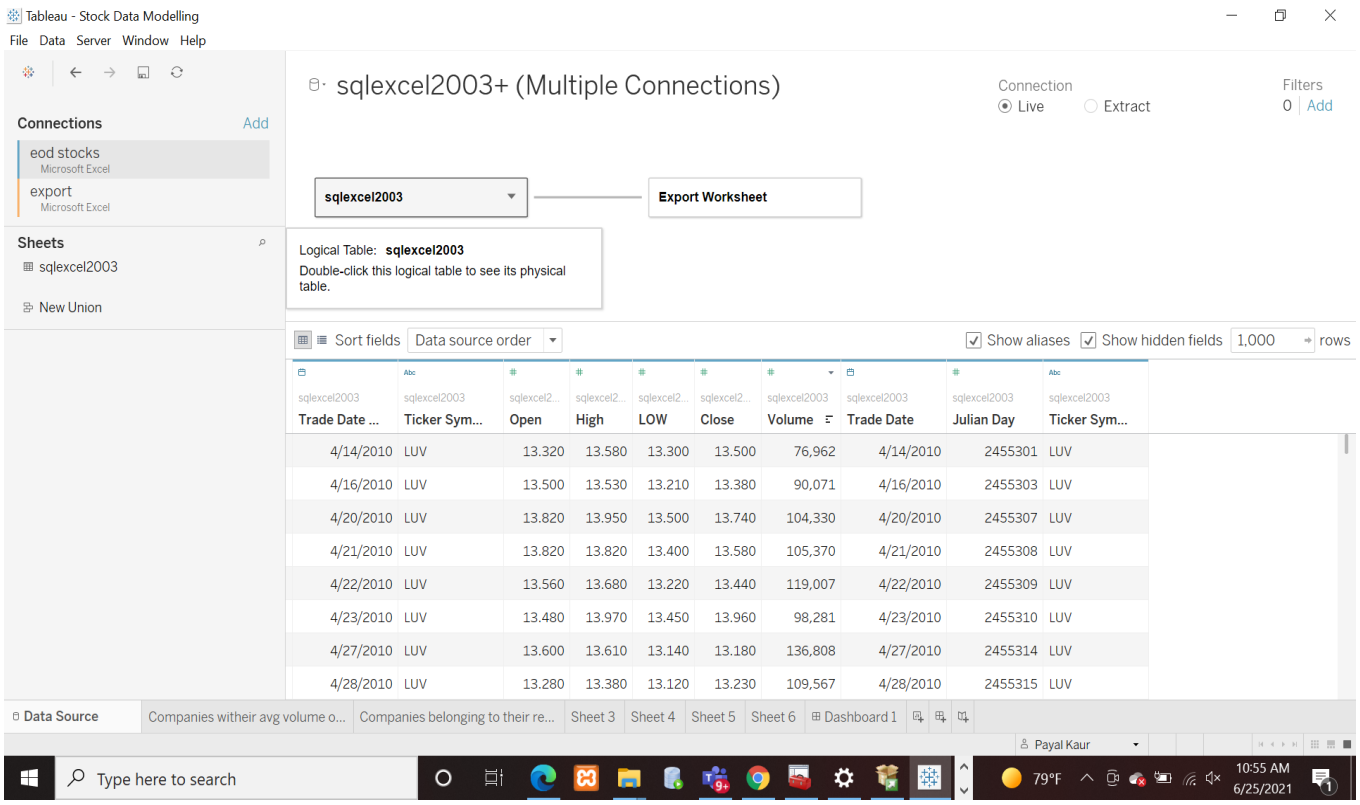
STK DATE DIMSv1 - Every grain of this table includes date and time dimension details.

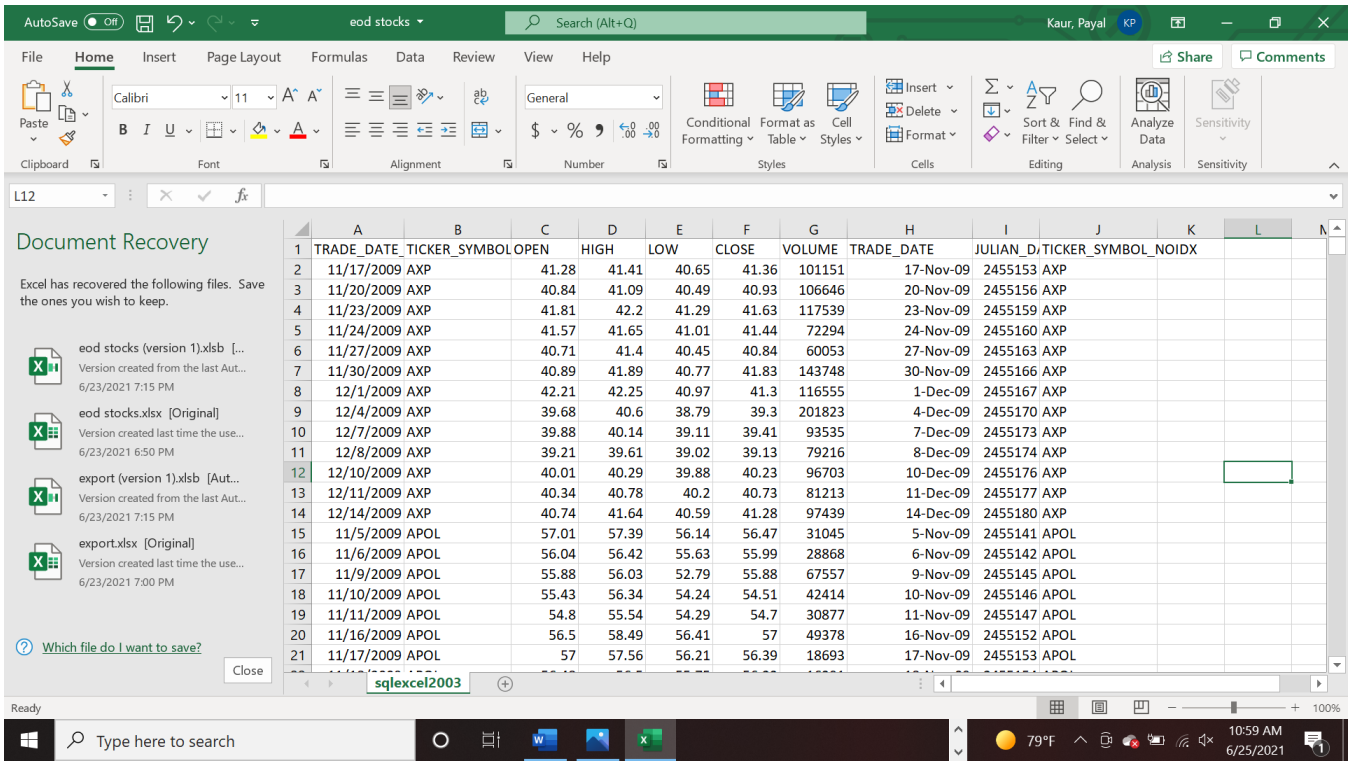
REPORTING, MODELLING AND STORYTELLING

Tableau connects and extracts the data stored in various places. It can pull data from any platform like an excel, pdf, to a complex database like Oracle. When Tableau is launched, ready data connectors are available which allows you to connect to any database. The pulled data can be either connected live or extracted to the Tableau's data engine, Tableau Desktop. On this the data analyst, data engineer work with the data that was pulled up and develop visualizations. The created dashboards are shared with the users as a static file. The data from the Tableau Desktop can be published to the Tableau server. This is an enterprise platform where collaboration, distribution, governance, security model, automation features are supported. With the Tableau server, the end users have a better experience in accessing the files from all locations be it a desktop, mobile or email.

Connection:

As we have fetched our data from oracle FIN schema, we have exported files from oracle database to excel workbook. From excel we imported all the different sheets of data include sp500 stocks data, particular trade dates, stock index facts dataset including stock's particular days high and low price , market opening and closing price , their effective date of market existence and date of expiration from market, stocks index dimension data including indexing of every company's stock with their respective ticker symbols , with company's internal information like the volume of stocks being traded its current day price.





	A	B	C	D	E	F	G	H	I	J	K	L
1	TRADE_DATE	TICKER_SYMBOL	OPEN	HIGH	LOW	CLOSE	VOLUME	TRADE_DATE	JULIAN_D	TICKER_SYMBOL_NO	IDX	
2	11/17/2009	AXP	41.28	41.41	40.65	41.36	101151	17-Nov-09	2455153	AXP		
3	11/20/2009	AXP	40.84	41.09	40.49	40.93	106646	20-Nov-09	2455156	AXP		
4	11/23/2009	AXP	41.81	42.2	41.29	41.63	117539	23-Nov-09	2455159	AXP		
5	11/24/2009	AXP	41.57	41.65	41.01	41.44	72294	24-Nov-09	2455160	AXP		
6	11/27/2009	AXP	40.71	41.4	40.45	40.84	60053	27-Nov-09	2455163	AXP		
7	11/30/2009	AXP	40.89	41.89	40.77	41.83	143748	30-Nov-09	2455166	AXP		
8	12/1/2009	AXP	42.21	42.25	40.97	41.3	116555	1-Dec-09	2455167	AXP		
9	12/4/2009	AXP	39.68	40.6	38.79	39.3	201823	4-Dec-09	2455170	AXP		
10	12/7/2009	AXP	39.88	40.14	39.11	39.41	93535	7-Dec-09	2455173	AXP		
11	12/8/2009	AXP	39.21	39.61	39.02	39.13	79216	8-Dec-09	2455174	AXP		
12	12/10/2009	AXP	40.01	40.29	39.88	40.23	96703	10-Dec-09	2455176	AXP		
13	12/11/2009	AXP	40.34	40.78	40.2	40.73	81213	11-Dec-09	2455177	AXP		
14	12/14/2009	AXP	40.74	41.64	40.59	41.28	97439	14-Dec-09	2455180	AXP		
15	11/5/2009	APOL	57.01	57.39	56.14	56.47	31045	5-Nov-09	2455141	APOL		
16	11/6/2009	APOL	56.04	56.42	55.63	55.99	28868	6-Nov-09	2455142	APOL		
17	11/9/2009	APOL	55.88	56.03	52.79	55.88	67557	9-Nov-09	2455145	APOL		
18	11/10/2009	APOL	55.43	56.34	54.24	54.51	42414	10-Nov-09	2455146	APOL		
19	11/11/2009	APOL	54.8	55.54	54.29	54.7	30877	11-Nov-09	2455147	APOL		
20	11/16/2009	APOL	56.5	58.49	56.41	57	49378	16-Nov-09	2455152	APOL		
21	11/17/2009	APOL	57	57.56	56.21	56.39	18693	17-Nov-09	2455153	APOL		

ANALYTICAL QUERIES WITH INSIGHTFUL DATA CHARTS

Query 1

Demonstrating the list of the companies with average volume of their stocks being traded in the market displayed in descending order.

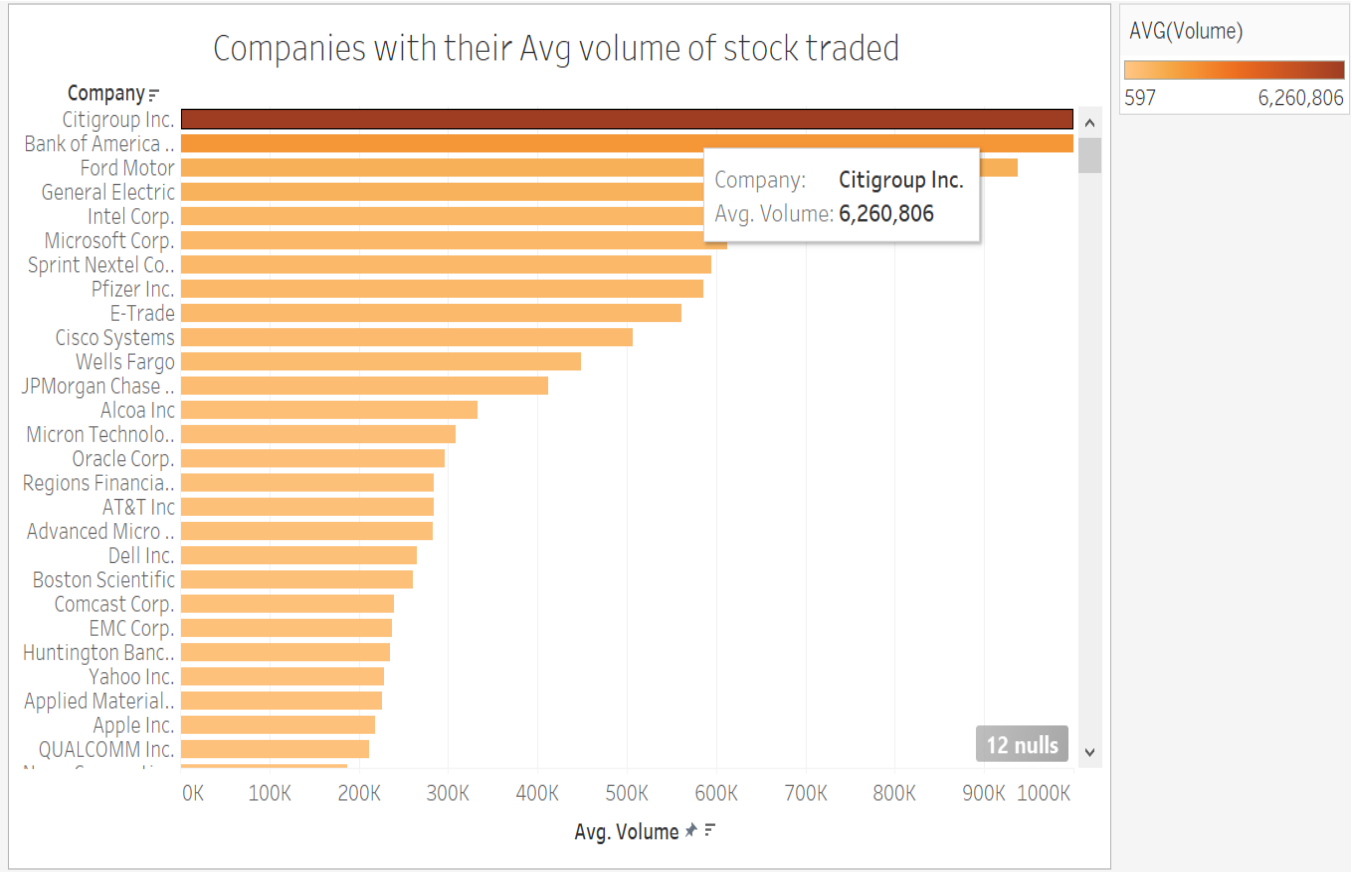
```
SELECT "Export Worksheet"."COMPANY" AS "COMPANY"
FROM "TableauTemp"."'Export Worksheet$'" "Export Worksheet"
WHERE (CASE WHEN (("Export Worksheet"."COMPANY" IN ('Cablevision Systems Corp.', 'Chipotle
Mexican Grill', 'Covidien plc', 'Dentsply International', 'Stanley Black & Decker', 'Supervalu Inc.', 'Target Corp.',
'Teradyne Inc.', 'Tesoro Petroleum Co.', 'Total System Services', 'Tyco International', 'United States Steel Corp.',
'Verisign Inc.', 'Verizon Communications', 'Viacom Inc.', 'Visa Inc.', 'Vornado Realty Trust', 'Wal-Mart Stores',
'Walgreen Co.', 'Waste Management Inc.', 'Whirlpool Corp.', 'Whole Foods Market', 'Williams Cos.', 'Xilinx Inc',
'Zimmer Holdings')) OR ("Export Worksheet"."COMPANY" IS NULL)) THEN FALSE ELSE TRUE END)
GROUP BY 1

SELECT "Export Worksheet"."COMPANY" AS "COMPANY"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
LEFT JOIN "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet" ON
("sqlexcel2003"."TICKER_SYMBOL" = "Export Worksheet"."TICKER_SYMBOL")
GROUP BY 1
LIMIT 64

SELECT "t0"."COMPANY" AS "COMPANY",
AVG(CAST("sqlexcel2003"."VOLUME" AS DOUBLE PRECISION OR NULL)) AS "avg:VOLUME:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
INNER JOIN (
   SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL","Export Worksheet"."COMPANY"
AS "COMPANY"
   FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
   LEFT JOIN "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet" ON
("sqlexcel2003"."TICKER_SYMBOL" = "Export Worksheet"."TICKER_SYMBOL")
   WHERE (CASE WHEN (("Export Worksheet"."COMPANY" IN ('Cablevision Systems Corp.', 'Chipotle
Mexican Grill', 'Covidien plc', 'Dentsply International', 'Stanley Black & Decker', 'Supervalu Inc.', 'Target Corp.',
'Teradyne Inc.', 'Tesoro Petroleum Co.', 'Total System Services', 'Tyco International', 'United States Steel Corp.',
'Verisign Inc.', 'Verizon Communications', 'Viacom Inc.', 'Visa Inc.', 'Vornado Realty Trust', 'Wal-Mart Stores',
'Walgreen Co.', 'Waste Management Inc.', 'Whirlpool Corp.', 'Whole Foods Market', 'Williams Cos.', 'Xilinx Inc',
'Zimmer Holdings')) OR ("Export Worksheet"."COMPANY" IS NULL)) THEN FALSE ELSE TRUE END)
    GROUP BY 2,1) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" IS NOT DISTINCT FROM
"t0"."TICKER_SYMBOL")
GROUP BY 1
```

EXPLANATION:

The chart displayed below shows list of companies with their volume of stock traded in a particular year mostly 209 and 2010.The bar graph below is color coded. The darker the shade of orange indicated the higher the volume

of stock of that company is traded. Based on the analysis made Citi group Inc. has the maximum volume of stocks being traded whose average value is more than 6 million.

CHART



QUERY 2:

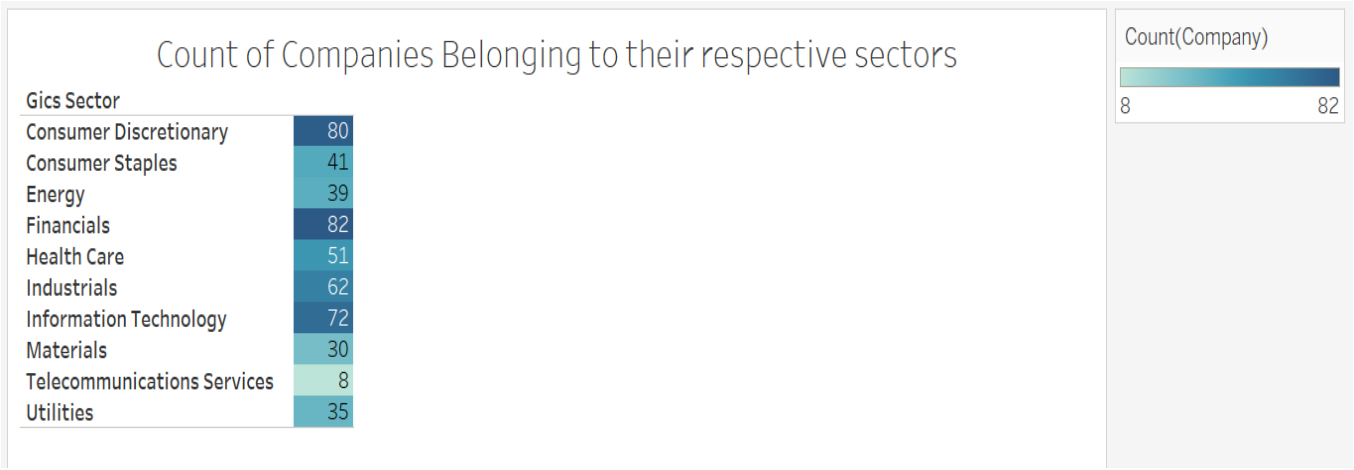
Demonstrating the distinct count of companies belonging to their respective GDP/GIS sector in alphabetical order.

```
SELECT "Export Worksheet"."GICS_SECTOR" AS "GICS_SECTOR",
COUNT(DISTINCT "Export Worksheet"."COMPANY") AS "ctd:COMPANY:ok"
FROM "TableauTemp"."'Export Worksheet$'" "Export Worksheet"
WHERE (NOT ("Export Worksheet"."GICS_SECTOR" IS NULL))
GROUP BY 1
```

EXPLANATION:

The table below show the count of the companies with their respective Global Industry Classification Standard (GICS) sectors indicating that most of the companies that we have in our dataset belong to financial sector.

CHART:



QUERY 3:

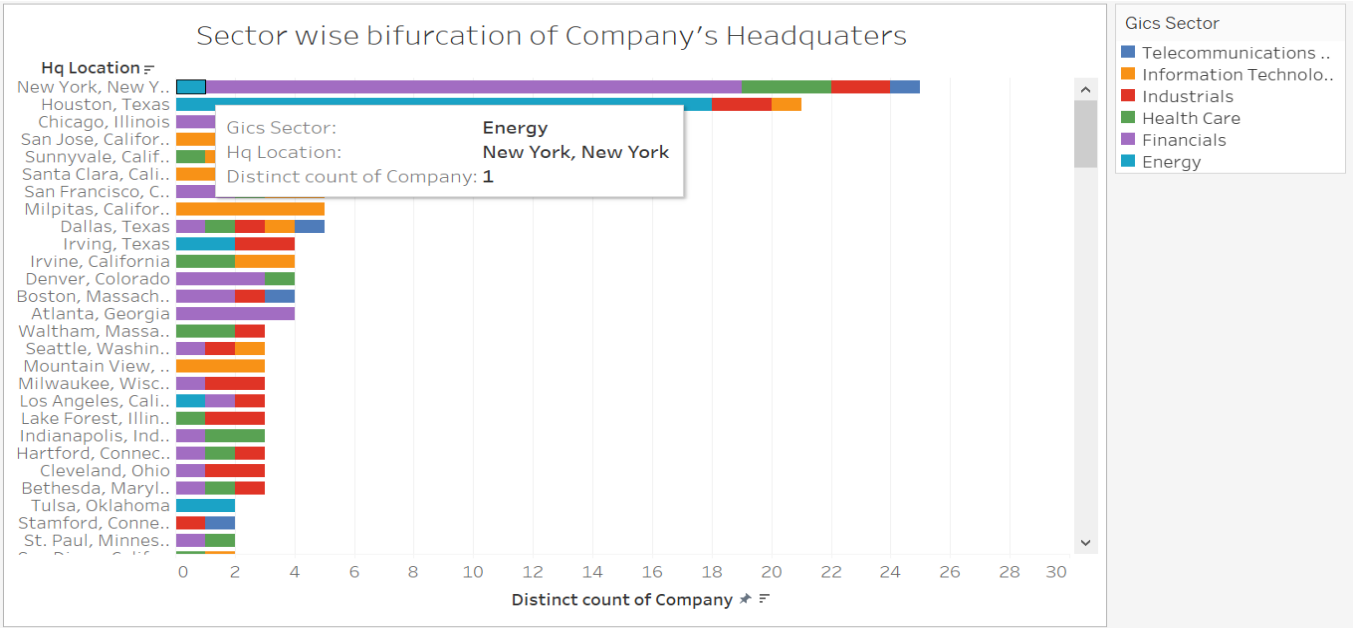
Demonstrating the sector wise and city wise bifurcation of companies with their respective headquarters

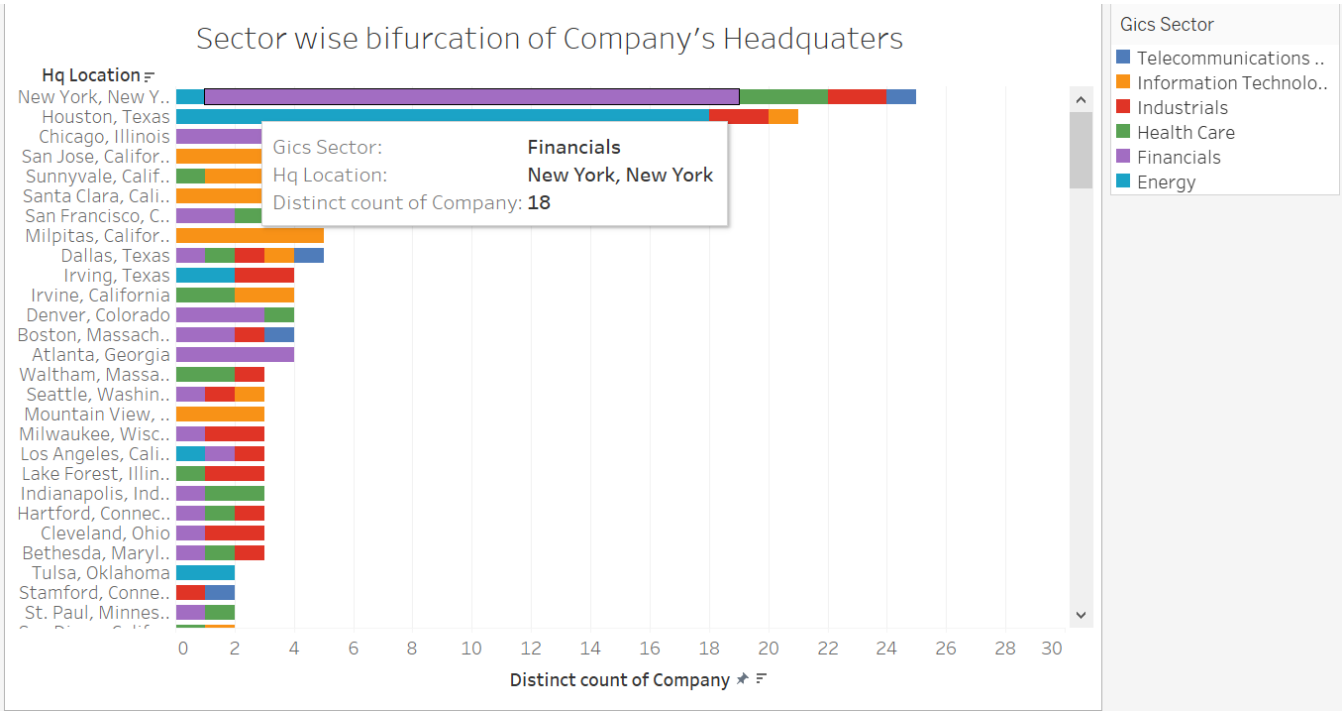
```
SELECT "Export Worksheet"."GICS_SECTOR" AS "GICS_SECTOR","Export Worksheet"."HQ_LOCATION"
AS "HQ_LOCATION",
COUNT(DISTINCT "Export Worksheet"."COMPANY") AS "ctd:COMPANY:ok"
FROM "TableauTemp"."'Export Worksheet$'" "Export Worksheet"
WHERE ((CASE WHEN ("Export Worksheet"."GICS_SECTOR" IN ('Consumer Discretionary', 'Consumer
Staples', 'Materials', 'Utilities')) THEN FALSE ELSE TRUE END) AND (NOT ("Export
Worksheet"."HQ_LOCATION" IS NULL)))
GROUP BY 1,2
```

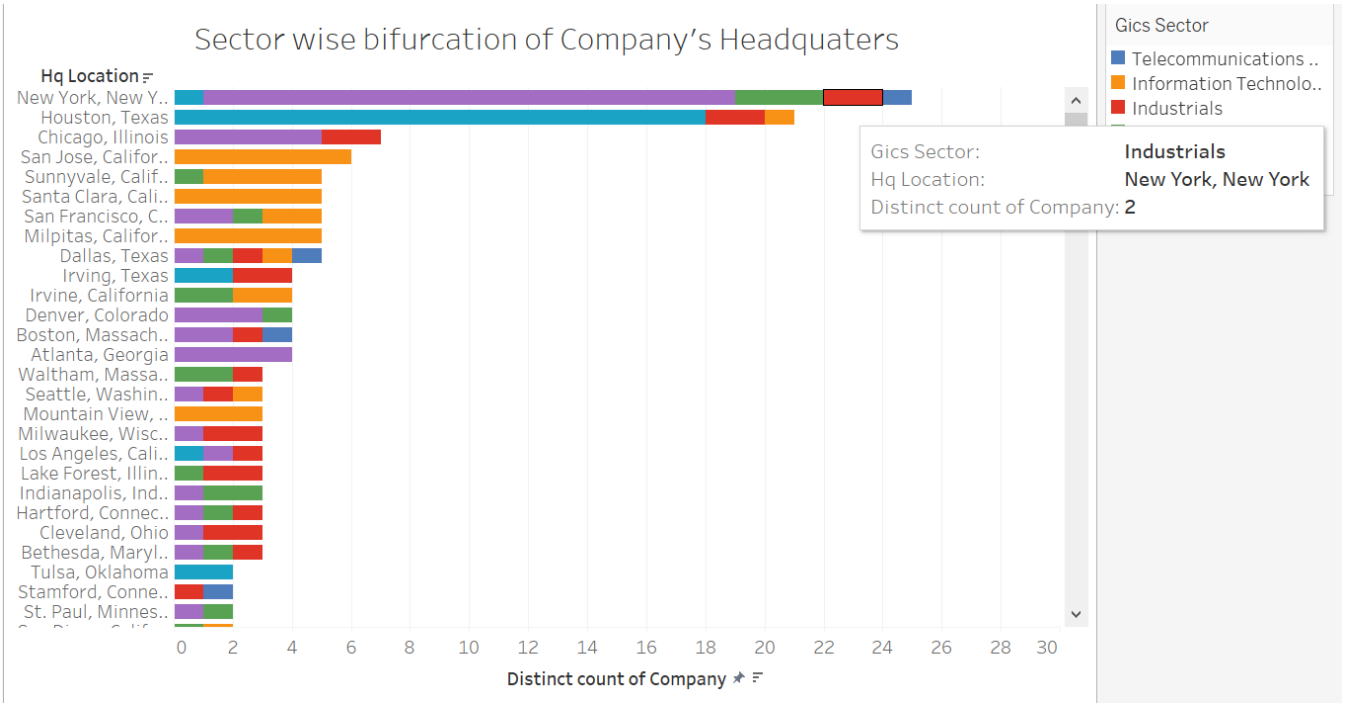
EXPLANATION:

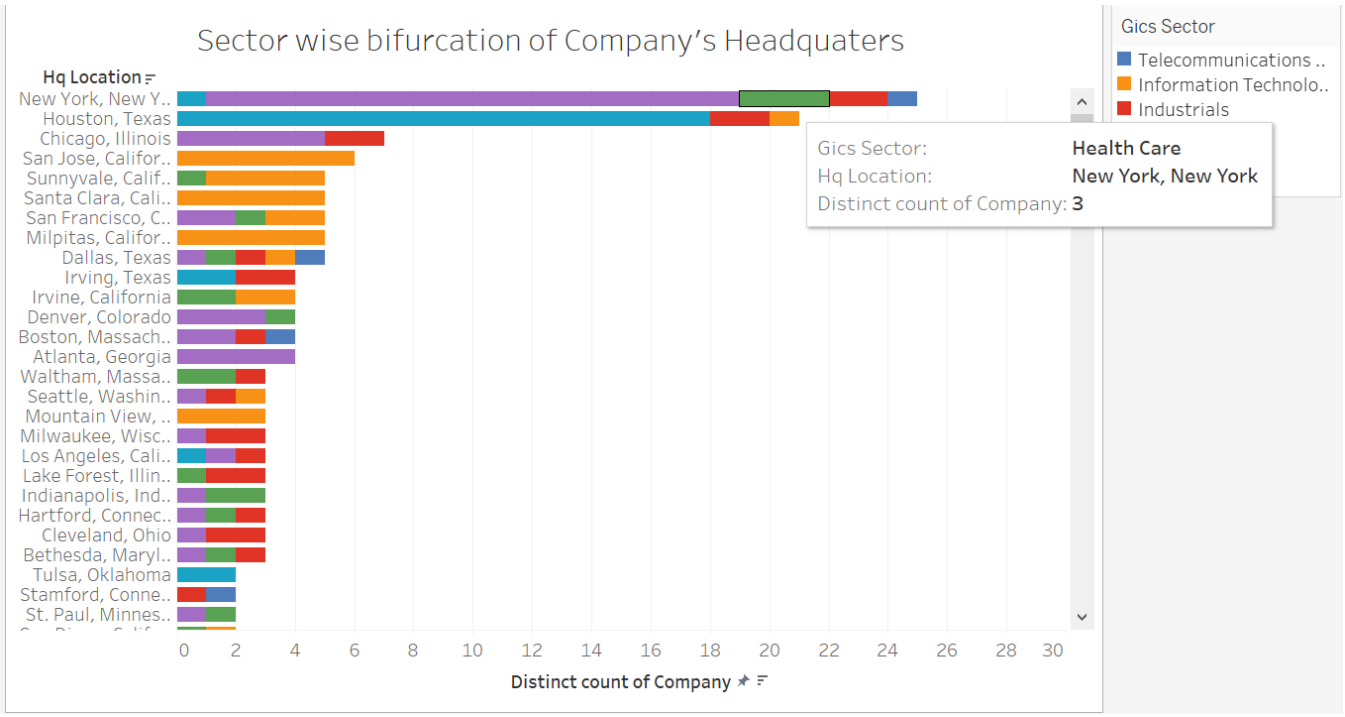
In the charts below, the city of New York is the house of 1 energy sector ,18 financial sector ,2 industrial sector and 1 telecommunication sector company head quarter. We can deduce that New York is the financial hub among all the other cities present in our data set. Similarly, San Jose has 6 company headquarters all of which belong of Information Technology sector.

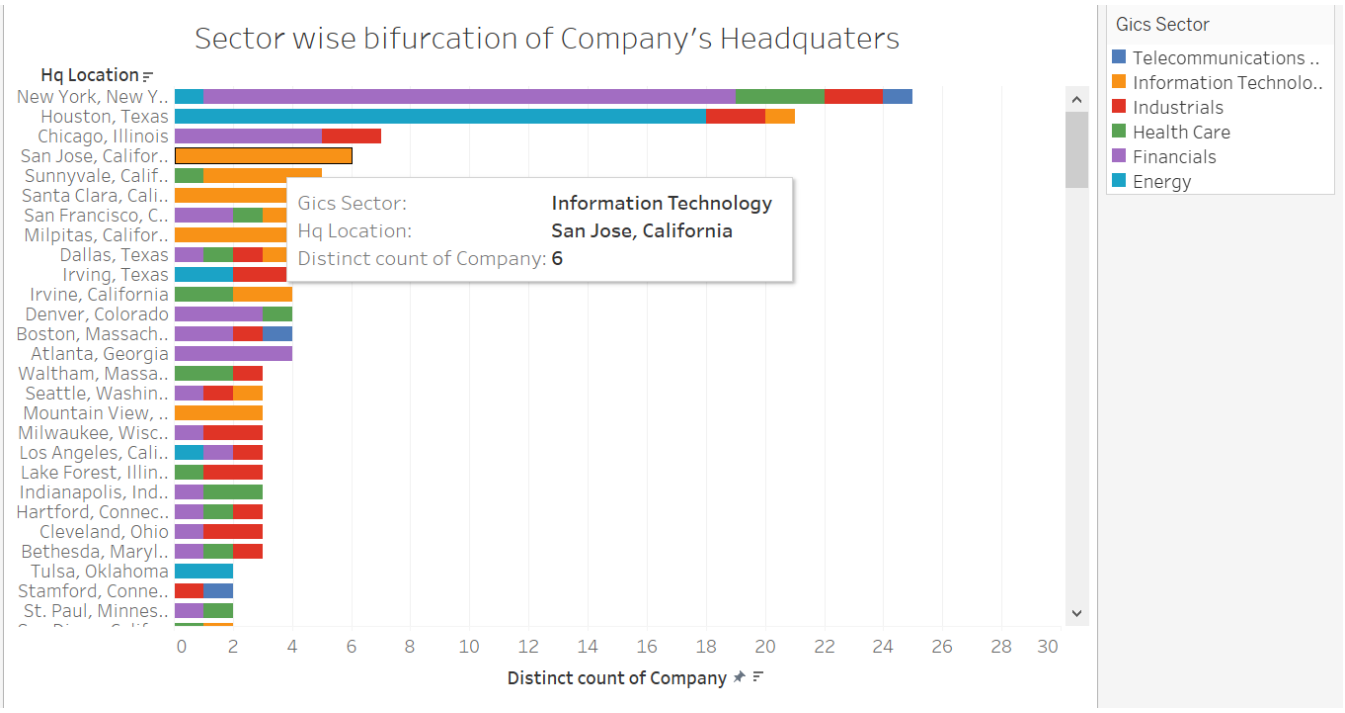
CHART:


Sector wise bifurcation of Company's Headquaters
Hq Location
New York, New Y..
Houston, Texas
Chicago, Illinois
San Jose, Califor..
Sunnyvale, Calif..
Santa Clara, Cali..
San Francisco, C..
Milpitas, Califor..
Dallas, Texas
Irving, Texas
Irvine, California
Denver, Colorado
Boston, Massach..
Atlanta, Georgia
Waltham, Massa..
Seattle, Washin..
Mountain View, ..
Milwaukee, Wisc..
Los Angeles, Cali..
Lake Forest, Illin..
Indianapolis, Ind..
Hartford, Connec..
Cleveland, Ohio
Bethesda, Maryl..
Tulsa, Oklahoma
Stamford, Conne..
St. Paul, Minnes..
Gics Sector: Energy
Hq Location: New York, New York
Distinct count of Company: 1
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Distinct count of Company
Gics Sector
Telecommunications ..
Information Technolo..
Industrials
Health Care
Financials
Energy


Sector wise bifurcation of Company's Headquaters
Hq Location
New York, New Y..
Houston, Texas
Chicago, Illinois
San Jose, Califor..
Sunnyvale, Calif..
Santa Clara, Cali..
San Francisco, C..
Milpitas, Califor..
Dallas, Texas
Irving, Texas
Irvine, California
Denver, Colorado
Boston, Massach..
Atlanta, Georgia
Waltham, Massa..
Seattle, Washin..
Mountain View, ..
Milwaukee, Wisc..
Los Angeles, Cali..
Lake Forest, Illin..
Indianapolis, Ind..
Hartford, Connec..
Cleveland, Ohio
Bethesda, Maryl..
Tulsa, Oklahoma
Stamford, Conne..
St. Paul, Minnes..
Gics Sector: Financials
Hq Location: New York, New York
Distinct count of Company: 18
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Distinct count of Company
Gics Sector
Telecommunications ..
Information Technolo..
Industrials
Health Care
Financials
Energy


Sector wise bifurcation of Company's Headquaters
Hq Location
New York, New Y..
Houston, Texas
Chicago, Illinois
San Jose, Califor..
Sunnyvale, Calif..
Santa Clara, Cali..
San Francisco, C..
Milpitas, Califor..
Dallas, Texas
Irving, Texas
Irvine, California
Denver, Colorado
Boston, Massach..
Atlanta, Georgia
Waltham, Massa..
Seattle, Washin..
Mountain View, ..
Milwaukee, Wisc..
Los Angeles, Cali..
Lake Forest, Illin..
Indianapolis, Ind..
Hartford, Connec..
Cleveland, Ohio
Bethesda, Maryl..
Tulsa, Oklahoma
Stamford, Conne..
St. Paul, Minnes..
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Distinct count of Company
Gics Sector
Telecommunications ..
Information Technolo..
Industrials
Gics Sector: Industrials
Hq Location: New York, New York
Distinct count of Company: 2


Sector wise bifurcation of Company's Headquaters
Hq Location
New York, New Y..
Houston, Texas
Chicago, Illinois
San Jose, Califor..
Sunnyvale, Calif..
Santa Clara, Cali..
San Francisco, C..
Milpitas, Califor..
Dallas, Texas
Irving, Texas
Irvine, California
Denver, Colorado
Boston, Massach..
Atlanta, Georgia
Waltham, Massa..
Seattle, Washin..
Mountain View, ..
Milwaukee, Wisc..
Los Angeles, Cali..
Lake Forest, Illin..
Indianapolis, Ind..
Hartford, Connec..
Cleveland, Ohio
Bethesda, Maryl..
Tulsa, Oklahoma
Stamford, Conne..
St. Paul, Minnes..
0 2 4 6 8 10 12 14 16 18 20 22 24 26 28 30
Distinct count of Company
Gics Sector
Telecommunications ..
Information Technolo..
Industrials
Gics Sector: Health Care
Hq Location: New York, New York
Distinct count of Company: 3



QUERY 4:

Demonstrating the open and close stock price trade value of companies in the year 2009 and 2010 .

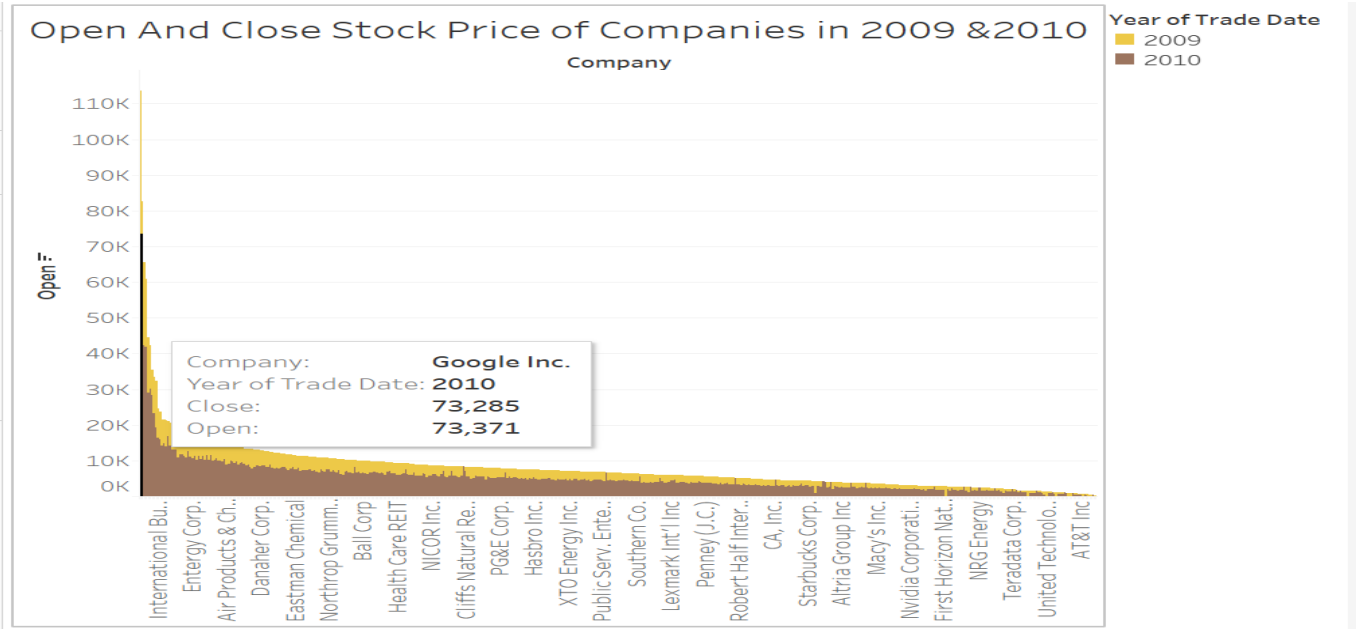
```
SELECT "t0"."COMPANY" AS "COMPANY",CAST(TRUNC(EXTRACT(YEAR FROM
"sqlexcel2003"."TRADE_DATE")) AS BIGINT OR NULL) AS "yr:TRADE_DATE:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "Export Worksheet"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "Export Worksheet"."COMPANY" AS "COMPANY"
  FROM "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet"
  WHERE (NOT ("Export Worksheet"."COMPANY" IS NULL))
) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t0"."TICKER_SYMBOL")
GROUP BY 1, 2

SELECT "t1"."COMPANY" AS "COMPANY",
  SUM("sqlexcel2003"."CLOSE") AS "sum:CLOSE:ok",
  SUM("sqlexcel2003"."OPEN") AS "sum:OPEN:ok",
  CAST(TRUNC(EXTRACT(YEAR FROM "sqlexcel2003"."TRADE_DATE")) AS BIGINT OR NULL) AS
"yr:TRADE_DATE:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "t0"."COMPANY" AS "COMPANY"
  FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
    INNER JOIN (
    SELECT "Export Worksheet"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
```

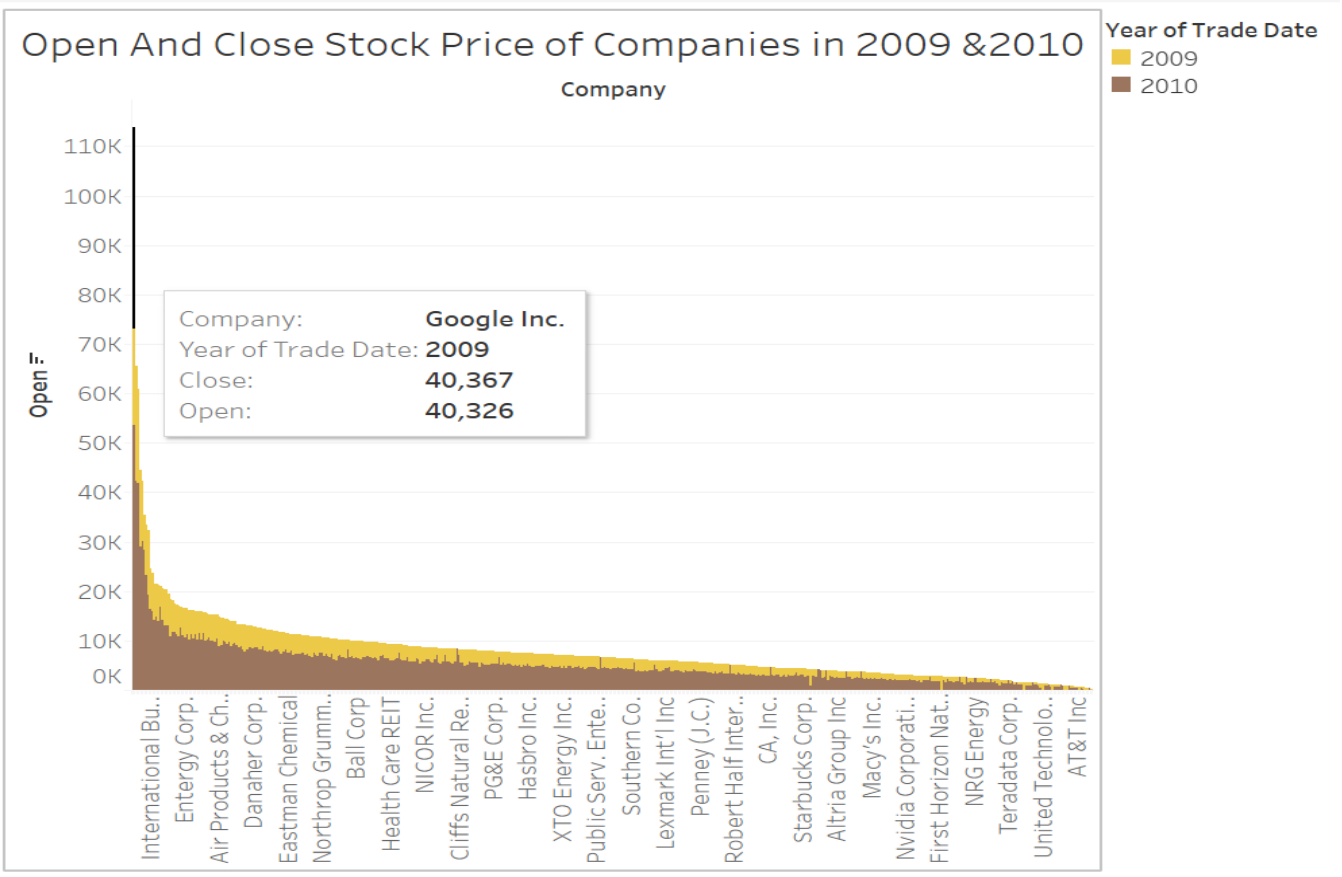
```
    "Export Worksheet"."COMPANY" AS "COMPANY"
   FROM "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet"
   WHERE (NOT ("Export Worksheet"."COMPANY" IS NULL))
  ) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t0"."TICKER_SYMBOL")
  GROUP BY 2, 1
) "t1" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t1"."TICKER_SYMBOL")
GROUP BY 1, 4

SELECT "t1"."COMPANY" AS "COMPANY",
  SUM("sqlexcel2003"."OPEN") AS "sum:OPEN:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "t0"."COMPANY" AS "COMPANY"
  FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
    INNER JOIN (
    SELECT "Export Worksheet"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
      "Export Worksheet"."COMPANY" AS "COMPANY"
    FROM "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet"
    WHERE (NOT ("Export Worksheet"."COMPANY" IS NULL))
  ) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t0"."TICKER_SYMBOL")
  GROUP BY 2,1
) "t1" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t1"."TICKER_SYMBOL")
GROUP BY 1
```

EXPLANATION:

The Charts below demonstrate the aggregate of open and close stock prices of companies in the year 2009 and 2010. If we compare, Google has highest opening and closing stock price in both consecutive years.

CHART:





QUERY 5:

Demonstrating the high and low stock price trade value of companies in the year 2009 and 2010 .

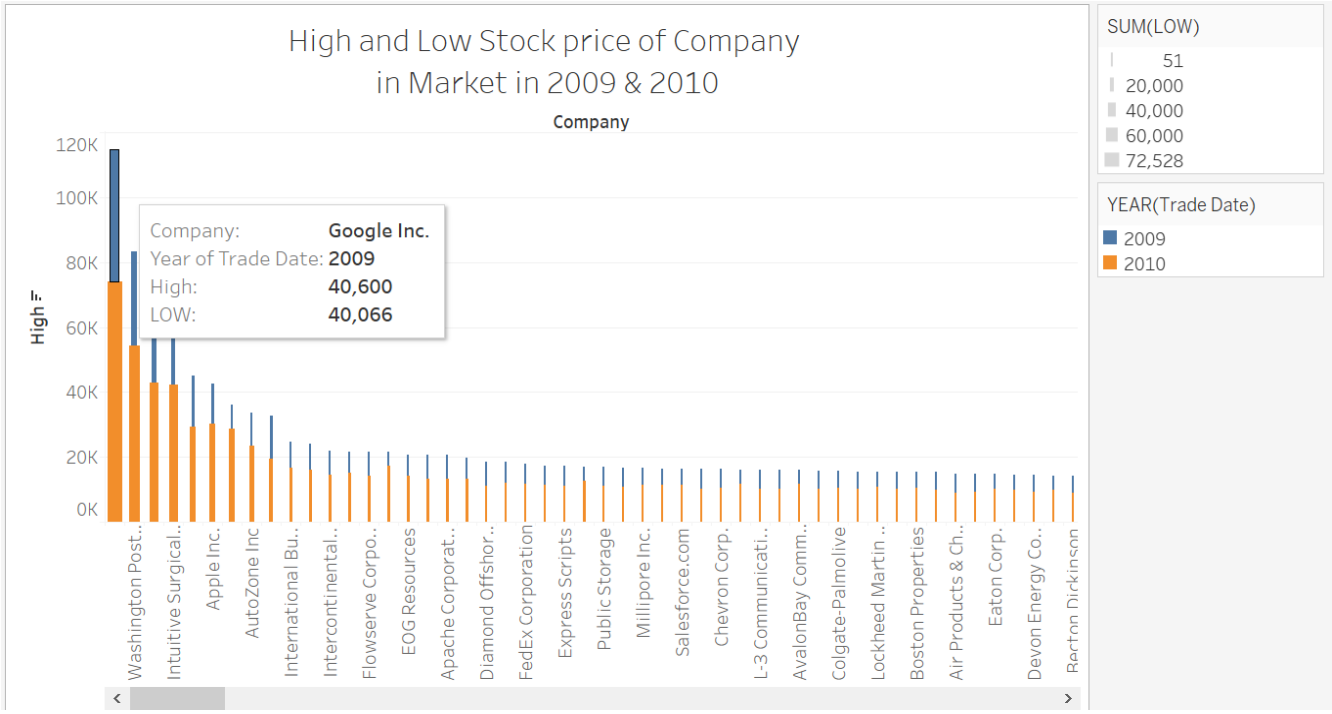
```
SELECT "t0"."COMPANY" AS "COMPANY",
  TABLEAU.TO_DATETIME(DATE_TRUNC('YEAR',
TABLEAU.NORMALIZE_DATETIME("sqlexcel2003"."TRADE_DATE")), "sqlexcel2003"."TRADE_DATE")
AS "tyr:TRADE_DATE:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "Export Worksheet"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "Export Worksheet"."COMPANY" AS "COMPANY"
  FROM "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet"
  WHERE (NOT ("Export Worksheet"."COMPANY" IS NULL))
) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t0"."TICKER_SYMBOL")
GROUP BY 1,2

SELECT "t1"."COMPANY" AS "COMPANY",
  SUM("sqlexcel2003"."HIGH") AS "sum:HIGH:ok",
  SUM("sqlexcel2003"."LOW") AS "sum:LOW:ok",
  TABLEAU.TO_DATETIME(DATE_TRUNC('YEAR',
TABLEAU.NORMALIZE_DATETIME("sqlexcel2003"."TRADE_DATE")), "sqlexcel2003"."TRADE_DATE")
AS "tyr:TRADE_DATE:ok"
```

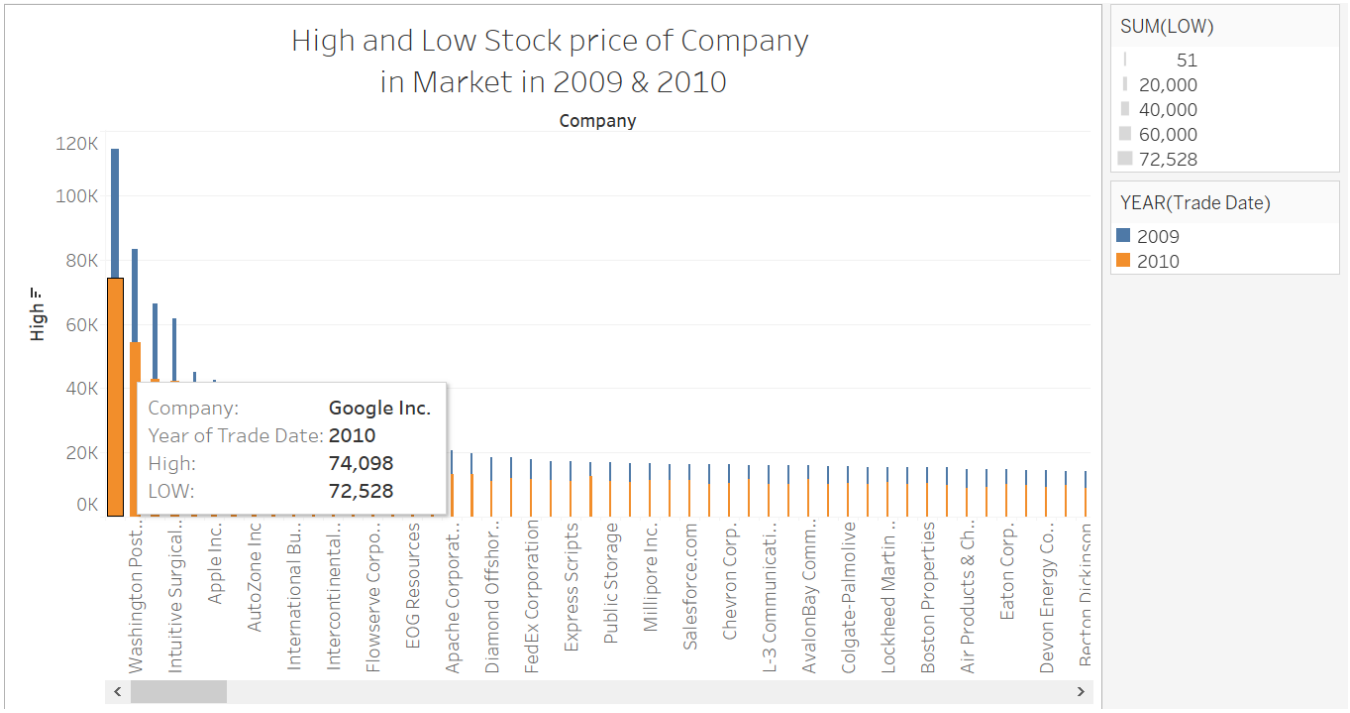
```
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "t0"."COMPANY" AS "COMPANY"
  FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
    INNER JOIN (
    SELECT "Export Worksheet"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
      "Export Worksheet"."COMPANY" AS "COMPANY"
    FROM "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet"
    WHERE (NOT ("Export Worksheet"."COMPANY" IS NULL))
  ) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t0"."TICKER_SYMBOL")
  GROUP BY 2, 1
) "t1" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t1"."TICKER_SYMBOL")
GROUP BY 1, 4

SELECT "t1"."COMPANY" AS "COMPANY",
  SUM("sqlexcel2003"."HIGH") AS "sum:HIGH:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "t0"."COMPANY" AS "COMPANY"
  FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
    INNER JOIN (
    SELECT "Export Worksheet"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
      "Export Worksheet"."COMPANY" AS "COMPANY"
    FROM "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet"
    WHERE (NOT ("Export Worksheet"."COMPANY" IS NULL))
  ) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t0"."TICKER_SYMBOL")
  GROUP BY 2, 1
) "t1" ON ("sqlexcel2003"."TICKER_SYMBOL" = "t1"."TICKER_SYMBOL")
GROUP BY 1
```

EXPLANATION:

The Charts below demonstrate the aggregate of high and low stock prices of companies in the year 2009 and 2010. If we compare, Google has highest opening and closing stock price in both consecutive years .

CHART:


High and Low Stock price of Company
in Market in 2009 & 2010
Company
120K
100K
80K
60K
40K
20K
0K
Company: Google Inc.
Year of Trade Date: 2009
High: 40,600
LOW: 40,066
Washington Post..
Intuitive Surgical..
Apple Inc.
AutoZone Inc
International Bu..
Intercontinental..
Flowserve Corpo..
EOG Resources
Apache Corporat..
Diamond Offshor..
FedEx Corporation
Express Scripts
Public Storage
Millipore Inc.
Salesforce.com
Chevron Corp.
L-3 Communicati..
AvalonBay Comm..
Colgate-Palmolive
Lockheed Martin ..
Boston Properties
Air Products & Ch..
Eaton Corp.
Devon Energy Co..
SUM(LOW)
51
20,000
40,000
60,000
72,528
YEAR(Trade Date)
2009
2010


High and Low Stock price of Company
in Market in 2009 & 2010
Company
120K
100K
80K
60K
40K
20K
0K
Company: Google Inc.
Year of Trade Date: 2010
High: 74,098
LOW: 72,528
Washington Post..
Intuitive Surgical..
Apple Inc.
AutoZone Inc
International Bu..
Intercontinental..
Flowserve Corpo..
EOG Resources
Apache Corporat..
Diamond Offshor..
FedEx Corporation
Express Scripts
Public Storage
Millipore Inc.
Salesforce.com
Chevron Corp.
L-3 Communicati..
AvalonBay Comm..
Colgate-Palmolive
Lockheed Martin ..
Boston Properties
Air Products & Ch..
Eaton Corp.
Devon Energy Co..
SUM(LOW)
51
20,000
40,000
60,000
72,528
YEAR(Trade Date)
2009
2010

QUERY 6:

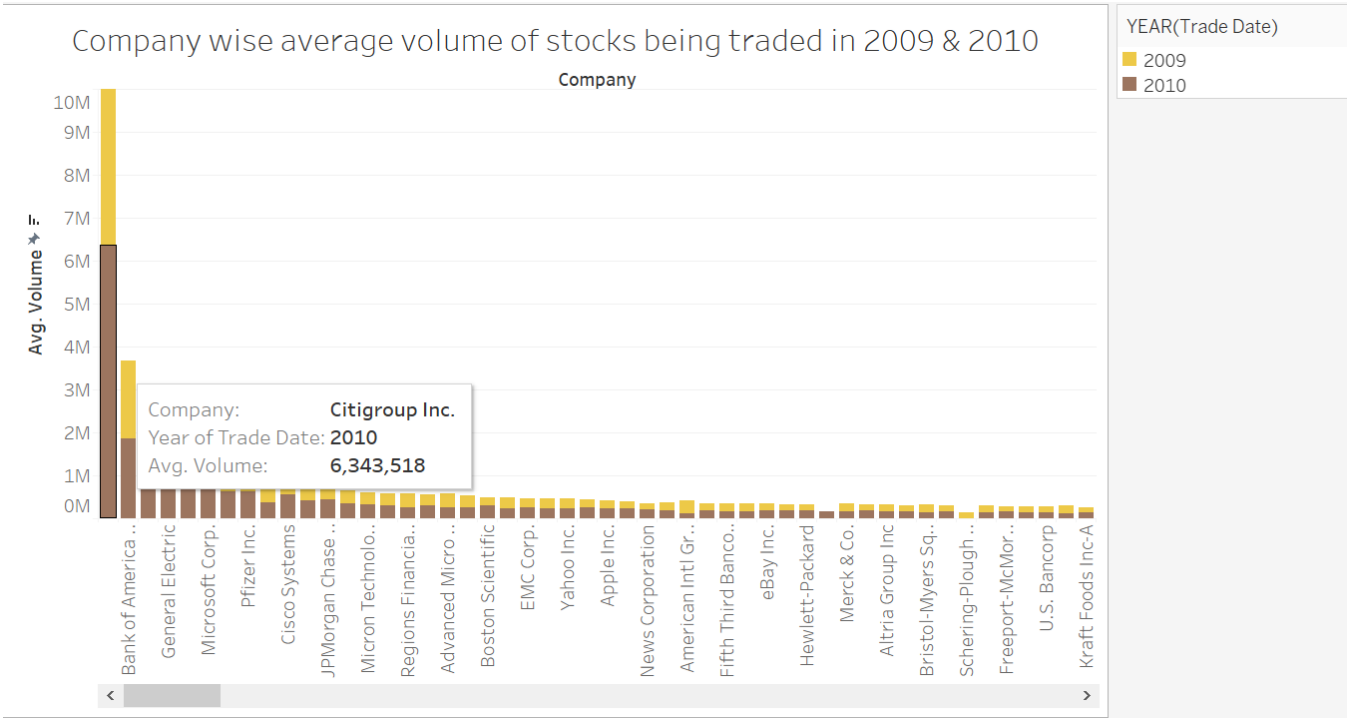
Demonstrating the list of the companies with average volume of their stocks being traded in the market displayed in descending order in year 2009 and 2010.

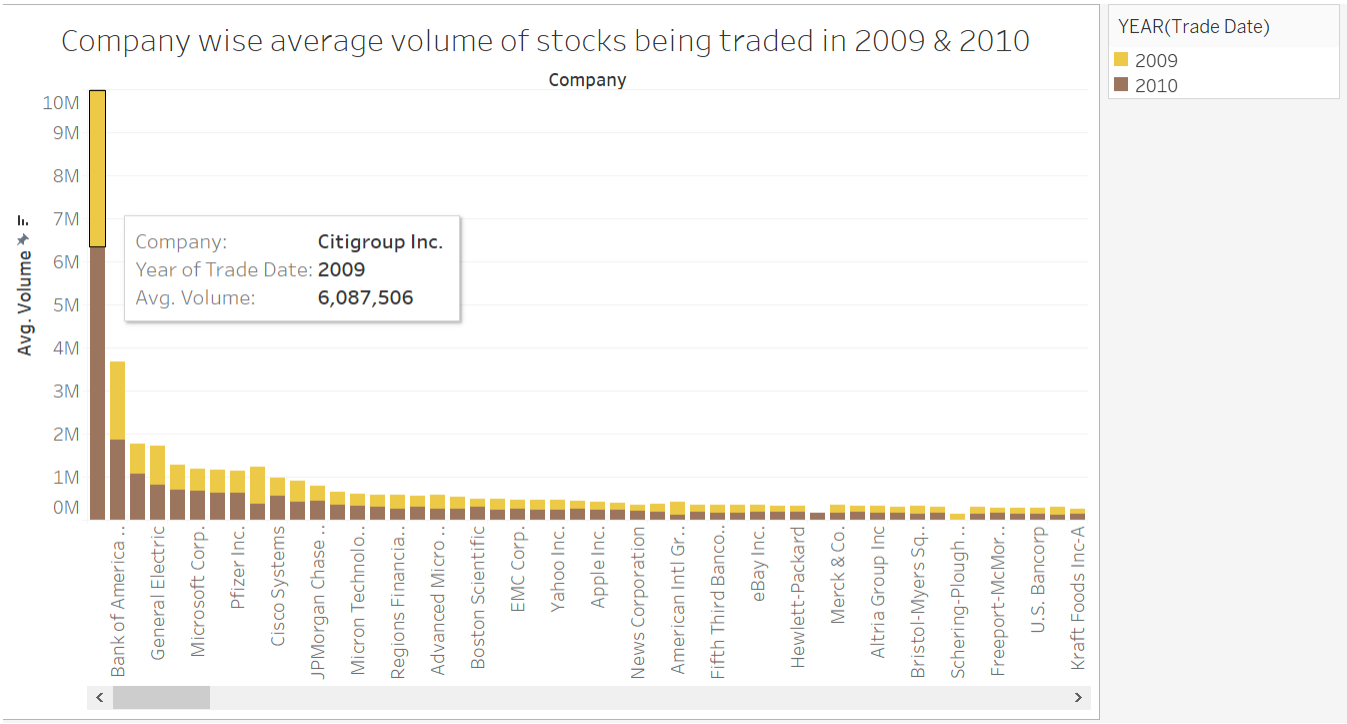
```
SELECT "Export Worksheet"."COMPANY" AS "COMPANY",CAST(TRUNC(EXTRACT(YEAR FROM
"sqlexcel2003"."TRADE_DATE")) AS BIGINT OR NULL) AS "yr:TRADE_DATE:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet" ON
("sqlexcel2003"."TICKER_SYMBOL" = "Export Worksheet"."TICKER_SYMBOL")
GROUP BY 1,2

SELECT "t0"."COMPANY" AS "COMPANY",
  AVG(CAST("sqlexcel2003"."VOLUME" AS DOUBLE PRECISION OR NULL)) AS "avg:VOLUME:ok",
  CAST(TRUNC(EXTRACT(YEAR FROM "sqlexcel2003"."TRADE_DATE")) AS BIGINT OR NULL) AS
"yr:TRADE_DATE:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "Export Worksheet"."COMPANY" AS "COMPANY"
  FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
    LEFT JOIN "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet" ON
("sqlexcel2003"."TICKER_SYMBOL" = "Export Worksheet"."TICKER_SYMBOL")
  GROUP BY 2,1
) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" IS NOT DISTINCT FROM "t0"."TICKER_SYMBOL")
GROUP BY 1, 3

SELECT "t0"."COMPANY" AS "COMPANY",
  AVG(CAST("sqlexcel2003"."VOLUME" AS DOUBLE PRECISION OR NULL)) AS "avg:VOLUME:ok"
FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
  INNER JOIN (
  SELECT "sqlexcel2003"."TICKER_SYMBOL" AS "TICKER_SYMBOL",
    "Export Worksheet"."COMPANY" AS "COMPANY"
  FROM "db1001"."TableauTemp"."sqlexcel2003$" "sqlexcel2003"
    LEFT JOIN "db1002"."TableauTemp"."'Export Worksheet$'" "Export Worksheet" ON
("sqlexcel2003"."TICKER_SYMBOL" = "Export Worksheet"."TICKER_SYMBOL")
  GROUP BY 2,1
) "t0" ON ("sqlexcel2003"."TICKER_SYMBOL" IS NOT DISTINCT FROM "t0"."TICKER_SYMBOL")
GROUP BY 1
```

EXPLANATION:

The Charts below demonstrate the average volume stock traded of companies in the year 2009 and 2010. If we compare, Citi Group Inc. has highest volume of stock traded in both consecutive years.

CHART:


Company wise average volume of stocks being traded in 2009 & 2010
Company
Avg. Volume
10M
9M
8M
7M
6M
5M
4M
3M
2M
1M
0M
Company: Citigroup Inc.
Year of Trade Date: 2010
Avg. Volume: 6,343,518
Bank of America ..
General Electric
Microsoft Corp.
Pfizer Inc.
Cisco Systems
JPMorgan Chase ..
Micron Technolo..
Regions Financia..
Advanced Micro ..
Boston Scientific
EMC Corp.
Yahoo Inc.
Apple Inc.
News Corporation
American Intl Gr..
Fifth Third Banco..
eBay Inc.
Hewlett-Packard
Merck & Co.
Altria Group Inc
Bristol-Myers Sq..
Schering-Plough ..
Freeport-McMor..
U.S. Bancorp
Kraft Foods Inc-A
YEAR(Trade Date)
2009
2010


Company wise average volume of stocks being traded in 2009 & 2010
Company
Avg. Volume
10M
9M
8M
7M
6M
5M
4M
3M
2M
1M
0M
Company: Citigroup Inc.
Year of Trade Date: 2009
Avg. Volume: 6,087,506
Bank of America ..
General Electric
Microsoft Corp.
Pfizer Inc.
Cisco Systems
JPMorgan Chase ..
Micron Technolo..
Regions Financia..
Advanced Micro ..
Boston Scientific
EMC Corp.
Yahoo Inc.
Apple Inc.
News Corporation
American Intl Gr..
Fifth Third Banco..
eBay Inc.
Hewlett-Packard
Merck & Co.
Altria Group Inc
Bristol-Myers Sq..
Schering-Plough ..
Freeport-McMor..
U.S. Bancorp
Kraft Foods Inc-A
YEAR(Trade Date)
2009
2010

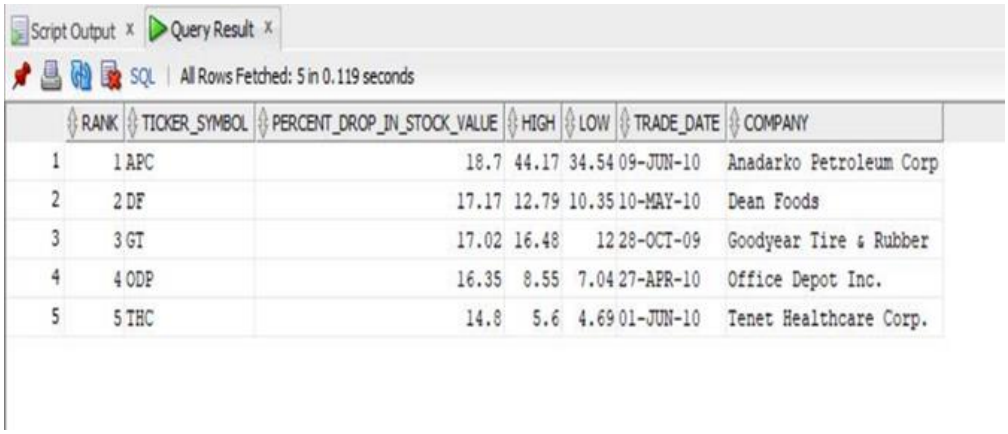
OTHER INTERESTING QUERIES

1. **Displaying the list top 5 companies whose stock value decreased majorly on a particular 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<6) a
 inner join FIN.sp500_stocks b on b.ticker_symbol=a.ticker_symbol
 order by a.stock_value_change desc;
```

OUTPUT-


Script Output | Query Result

SQL | All Rows Fetched: 5 in 0.119 seconds

	RANK	TICKER_SYMBOL	PERCENT_DROP_IN_STOCK_VALUE	HIGH	LOW	TRADE_DATE	COMPANY
1	1	APC	18.7	44.17	34.54	09-JUN-10	Anadarko Petroleum Corp
2	2	DF	17.17	12.79	10.35	10-MAY-10	Dean Foods
3	3	GT	17.02	16.48	12	28-OCT-09	Goodyear Tire & Rubber
4	4	ODP	16.35	8.55	7.04	27-APR-10	Office Depot Inc.
5	5	THC	14.8	5.6	4.69	01-JUN-10	Tenet Healthcare Corp.

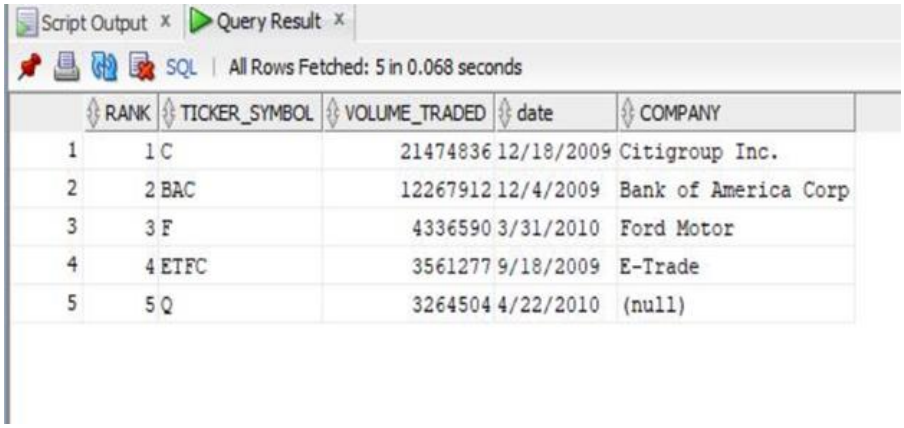
2. **Top 5 companies whose stocks have maximum volume traded on a single day.**

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<6) 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

SQL | All Rows Fetched: 5 in 0.068 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)

3. The average high, low, volume of a particular company(Walmart) 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=100016
and dd.year_nbr=2018
GROUP BY sd.co_name,dd.year_nbr;
```

OUTPUT-

Query Result

SQL | All Rows Fetched: 47 in 0.081 seconds

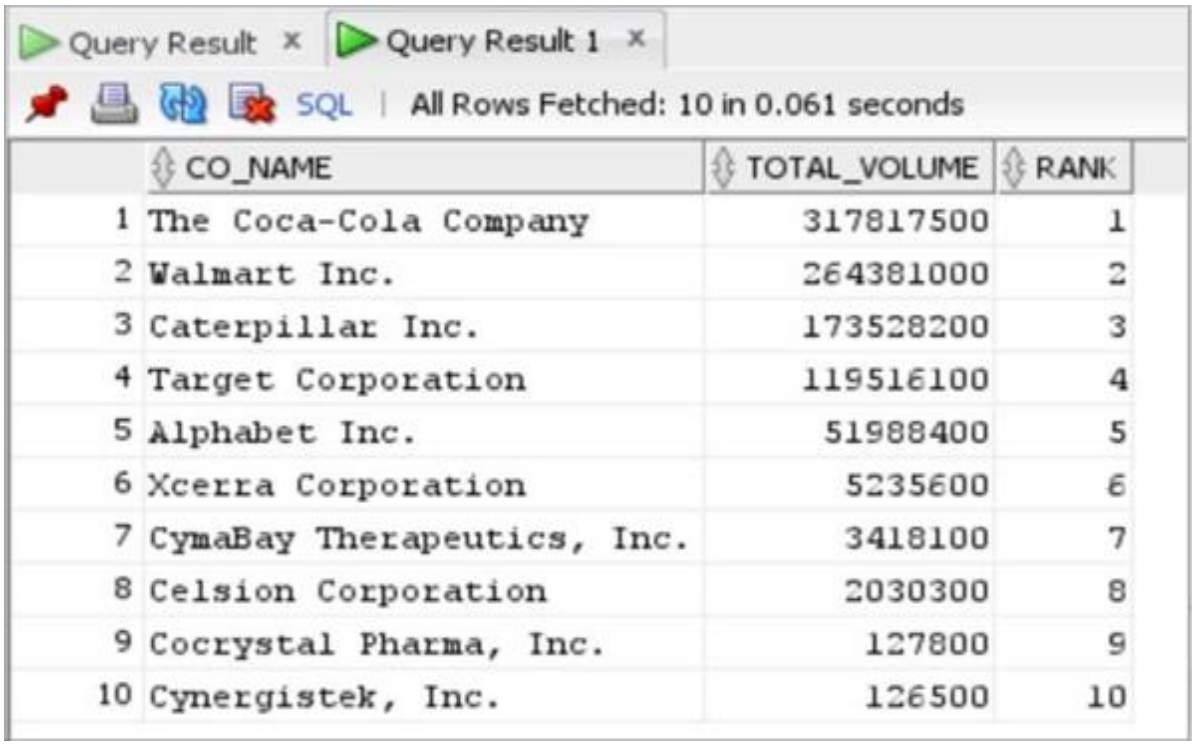
	CO_NAME	YEAR_NBR	AVG_HIGH	AVG_LOW	AVG_VOLUME
1	Walmart Inc.	2018	91.26	89.58	10324282.84
2	Walmart Inc.	2017	79.44	78.46	8902196.02
3	Walmart Inc.	2016	70.02	69.02	9362808.33
4	Walmart Inc.	2015	73.06	72.03	9028227.78
5	Walmart Inc.	2014	77.74	76.86	6515612.3
6	Walmart Inc.	2013	75.73	74.84	6951496.03
7	Walmart Inc.	2012	67.6	66.79	9239015.2
8	Walmart Inc.	2011	54.73	53.97	12127132.14
9	Walmart Inc.	2010	53.35	52.66	13405130.95
10	Walmart Inc.	2009	51.03	50.08	19253366.67
11	Walmart Inc.	2008	55.82	54.21	23798944.27
12	Walmart Inc.	2007	47.34	46.54	17458298.8
13	Walmart Inc.	2006	46.99	46.27	13382808.76
14	Walmart Inc.	2005	49.12	48.43	13152660.71
15	Walmart Inc.	2004	55.37	54.48	10004678.17
16	Walmart Inc.	2003	54.84	53.8	8346963.49
17	Walmart Inc.	2002	56.12	54.6	9083488.49
18	Walmart Inc.	2001	52.71	51.2	7525891.53
19	Walmart Inc.	2000	54.81	52.55	8115940.48

4. **Rank-wise list of companies based on the total volume of stock traded in a particular 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 '%JANUARY%'
 GROUP BY sd.co_name,
 dd.month_name;
```

OUTPUT



Query Result | Query Result 1

SQL | All Rows Fetched: 10 in 0.061 seconds

	CO_NAME	TOTAL_VOLUME	RANK
1	The Coca-Cola Company	317817500	1
2	Walmart Inc.	264381000	2
3	Caterpillar Inc.	173528200	3
4	Target Corporation	119516100	4
5	Alphabet Inc.	51988400	5
6	Xcerra Corporation	5235600	6
7	CymaBay Therapeutics, Inc.	3418100	7
8	Celsion Corporation	2030300	8
9	Cocrystal Pharma, Inc.	127800	9
10	Cynergistek, Inc.	126500	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_name,
 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=2017
 AND dd.MONTH_NAME LIKE '%JUNE%'
 GROUP BY sd.co_name,dd.year_nbr,dd.month_name
 ORDER BY sd.co_name;
```

OUTPUT

Query Result

All Rows Fetched: 10 in 0.046 seconds

	CO_NAME	YEAR_NBR	MONTH_NAME	MAX_FLUCTUATION_HIGH	MAX_FLUCTUATION_LOW
1	Alphabet Inc.	2017	JUNE	63.609985	67.020019
2	Caterpillar Inc.	2017	JUNE	3.910003	4.619995
3	Celsion Corporation	2017	JUNE	2.85	1.16
4	Cocrystal Pharma, Inc.	2017	JUNE	1.8	2.4
5	CymaBay Therapeutics, Inc.	2017	JUNE	1.22	1.21
6	Cynergistek, Inc.	2017	JUNE	0.85	0.9
7	Target Corporation	2017	JUNE	7.410004	8.539997
8	The Coca-Cola Company	2017	JUNE	0.95	1.129997
9	Walmart Inc.	2017	JUNE	4.970001	6.190002
10	Xcerra Corporation	2017	JUNE	0.16	0.2

CONCLUSION

Data Warehouses have become an integral part in storing data from multiple heterogeneous sources and utilizing the data for online analytical processing. In our project we have implemented the data warehouse design for stock market analysis where there's a single/multiple fact tables and multiple dimensional tables around it. After the structure is formed and the relations are made, we have developed analytical queries to fetch the data to address business questions. To explain the results in a more elaborate and intuitive way, we have used the visualization tool tableau to represent the results of our business questions where we found answers to some interesting questions. Our project is a proof of concept of a real time perspective of how businesses collect and store their data in a warehouse and utilize it for online analytical processing to bring out insights and utilize it for making important and data driven decisions. Data warehouses are very critical especially with the current trend of increase in data generation and increase in usage of data to drive decisions.

REFRENCES

[1] Inmon
[2] Adverity https://blog.adverity.com/data-warehouse-marketing-better-faster-insights
[3] https://shodhgangotri.inflibnet.ac.in/bitstream/123456789/5543/2/02_synopsis.pdf
[4]https://www.sec.gov/news/public-statement/statement-chairman-clayton-2018-10-16
[5]https://www.sciencedirect.com/topics/computer-science/dimensional-model

https://olap.com/olap-definition/