**Table of Contents**

1. **INTRODUCTION**
2. **Data Sources**
3. **Data Warehouse Data Model**
4. **Logical Data Mapping**
5. **Queries**
6. **Visualization**
7. **Conclusion**

**1. Introduction**

**1.1. Introduction about the Project**

The S&P 500 tracks the stock performance and market capitalization of the roughly 500 companies included in the index,

measuring the value of the stock of those companies.

**1.2. Business Requirements**

* Reporting stock market daily statistics.
* Reporting top and lowest performers.
* Tracking stock market changes.

**2. Data Sources**

**2.1. Wikipedia List of S&P 500 companies**

**Link:** [List of S&P 500 companies - Wikipedia](https://en.wikipedia.org/wiki/List_of_S%26P_500_companies)

**Descriptions: this data set includes the 503 companies’ info**

**(symbol, name, sector, industry)**

**ETL:** extracted the file with python as csv file (web scrubbing)

**2.2. performance of S&P 500 companies**

**Link:** [**https://www.tradingview.com/symbols/SPX/components/?exchange=SP**](https://www.tradingview.com/symbols/SPX/components/?exchange=SP)

**Descriptions: this data set include the companies’ performance over different periods of time**

**2.3. S&P 500 index**

**Link:** [**https://www.kaggle.com/datasets/paultimothymooney/stock-market-data**](https://www.kaggle.com/datasets/paultimothymooney/stock-market-data)

**Data collected date: 12/7/2022**

**Descriptions: this data set includes S&P index performance**

**2.4. S&P stock data**

**Link:** <https://www.kaggle.com/camnugent/sandp500>

**Data collected date: 12/7/2022**

**Description: this data set includes daily changes for the stock for the past 5 years for each company**

**2.5. constituents financials**

**Link:** <https://datahub.io/core/s-and-p-500-companies-financials#resource-constituents-financials>

**Data collected date: 2022**

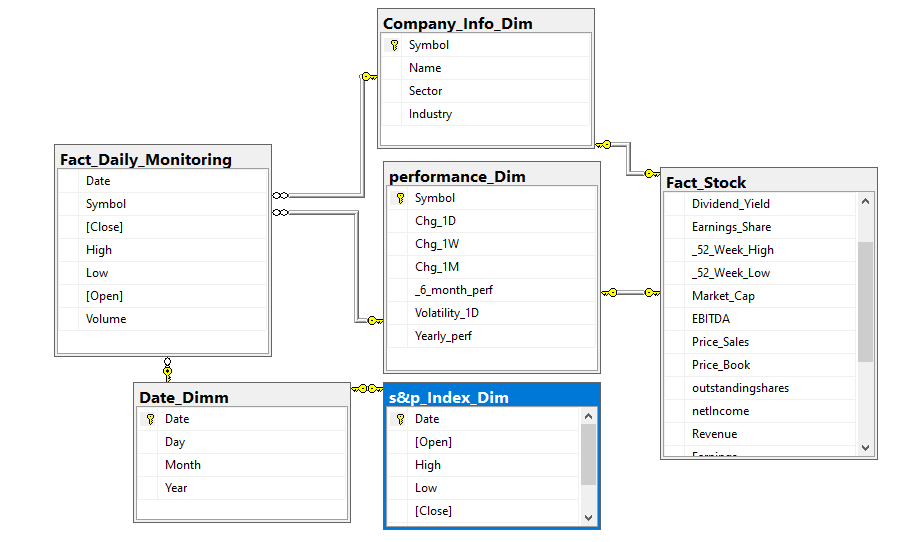
**Description: this data set includes financials data for each company**

**3. Data Warehouse Data Model**

3.1. Galaxy Schema

we used the galaxy schema as we have more than a fact table, and multiple dimensions connected with other dimensions so the most suitable schema to build such a complex model is the galaxy schema.

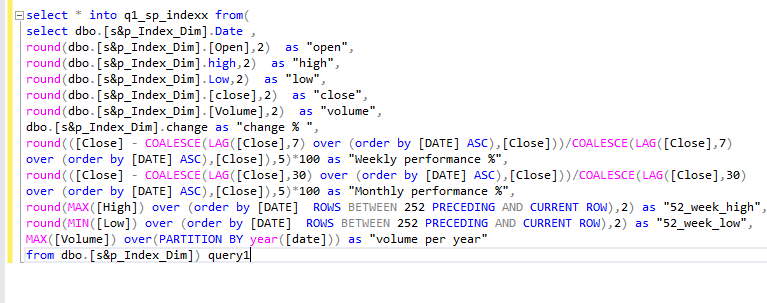
3.2. Dimensional Model

****

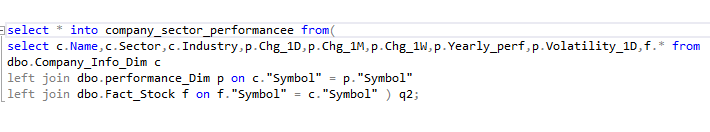
**4. Logical Data Mapping**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Source Table Name | Column | Data Type | PK | Table Type | Data Source | Transformation | Target Table Name |
| fact daily monitoring | **Symbol**  **date** | **Varchar**  **date** | (no)  (no) | Fact |  |  | **- Company info dimension**  **- Performance dim**  **-Date dim** |
| Fact stock | **symbol** | **Varchar** | (yes) | fact |  |  | **- Company info dimension**  **- Performance dim** |
| Company info dim | **symbol** | **Varchar** | (yes) | dimension |  |  | **The 2 fact tables** |
| Performance dim | **symbol** | **Varchar** | (yes) | dimension |  |  | **The 2 fact tables** |
| S&P index dim | **DATE** | **date** | (yes) | dimension |  |  | **Date dim** |
| date dim | **date** | **date** | (yes) | dimension |  |  | **-Fact daily monitoring**  **-S&P index dim** |

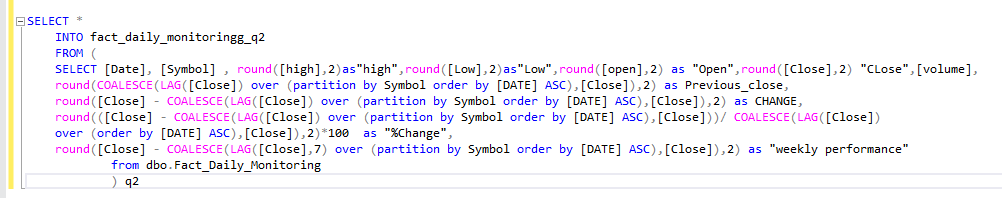
**5. Queries**

****

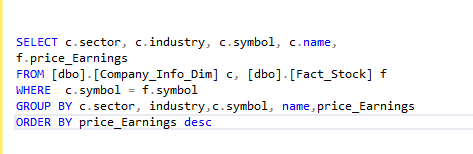
**Calculating s&p index performance**

****

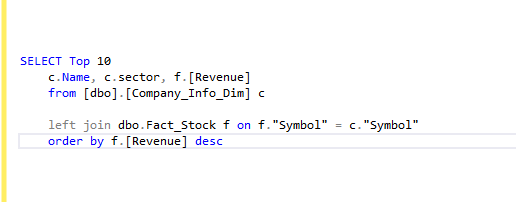
**Retrieving daily tracking data for a company**

****

**Calculating daily and weekly performance for a company**

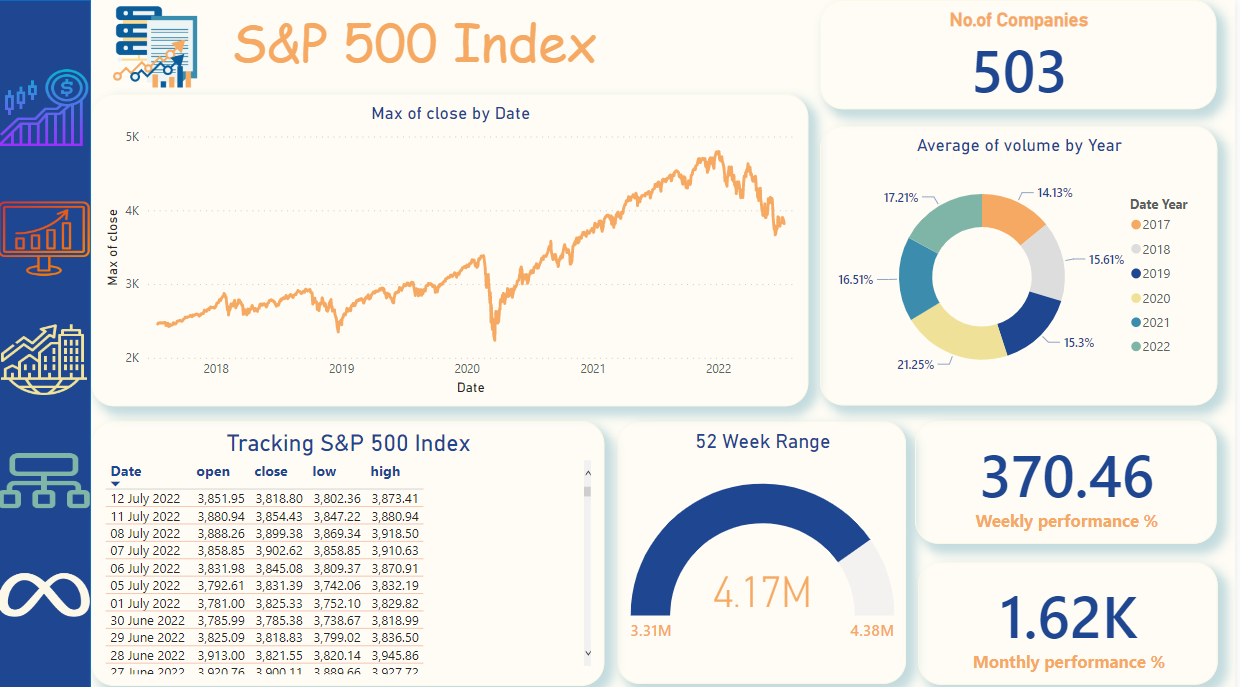
****

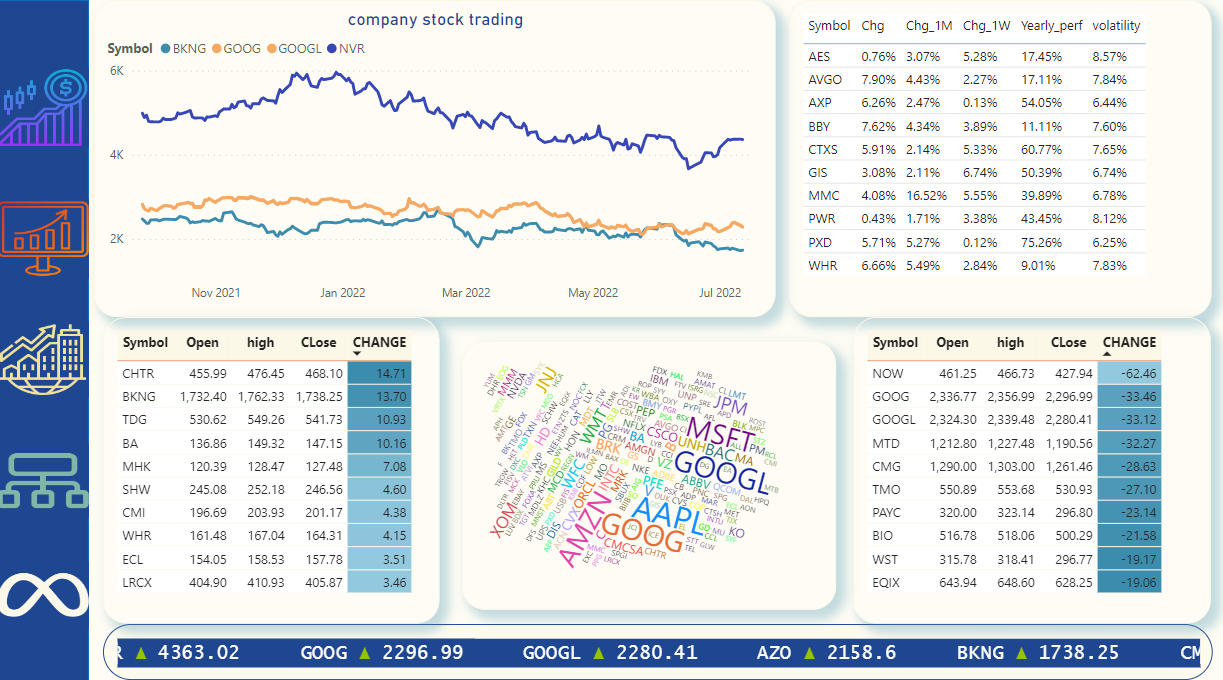
**Comparing different companies in same industry to p/e value**

****

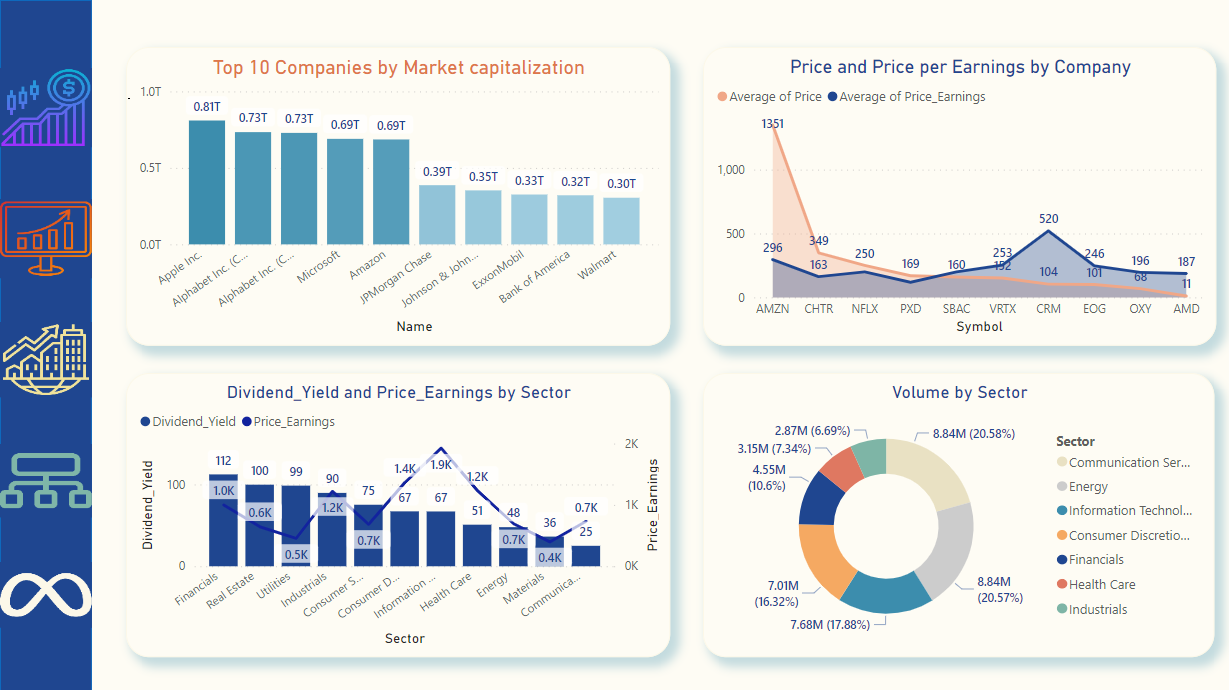
**Top 10 companies by revenue**

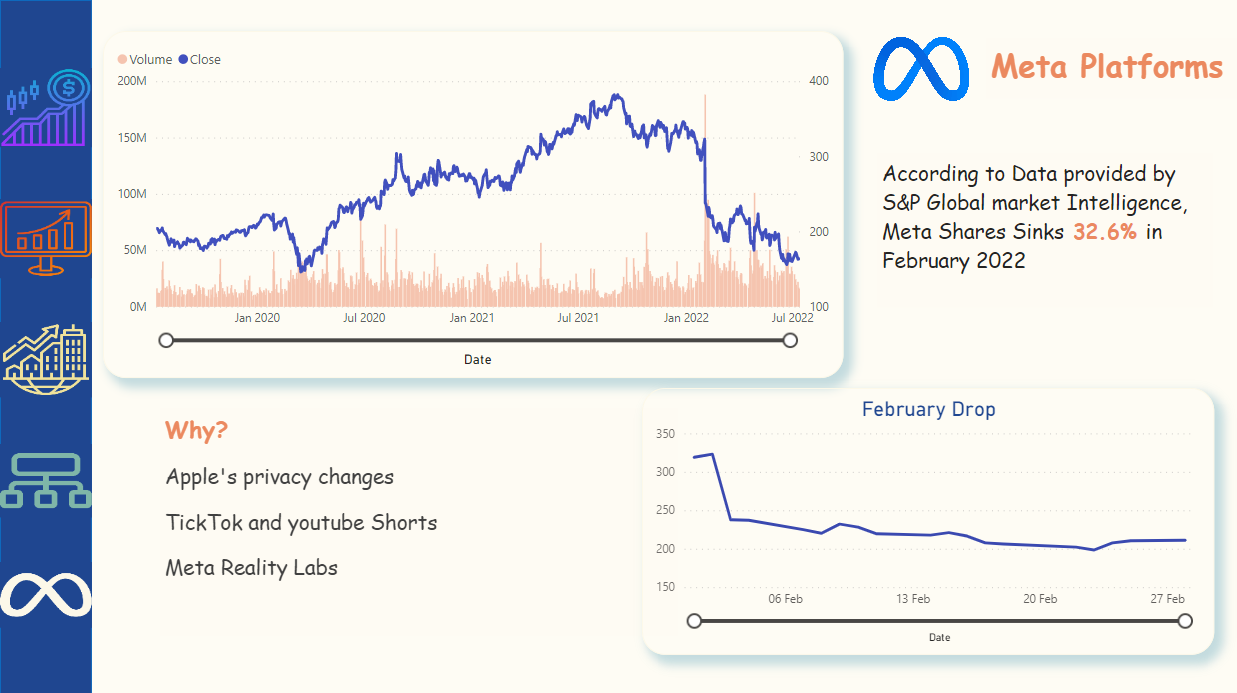
**6. Visualization**

****

****

****

****

****

**7. Conclusion**

**Project walkthrough:**

1. **Data understanding**
2. **Data gathering**
3. **Building logical model**
4. **Uploading data on database and managing relationships**
5. **Building physical model**
6. **Uploading data queries on power bi and building the dashboard**

Technologies used in the project

1. Python programing language (to extract data from web pages)
2. Excel (data cleaning)
3. SQL server data base (SSMS)
4. Power bi

**Future plan**

**Make the dashboard updated till the last second.**

**Predict the stock price.**

**Represent more KPIs.**