Faculty of
Computer & Informatics Eng.
Department of Software &
Information Systems Eng.

كلية الهندسة المعلوماتية
هندسة البرمجيات و نظم
المعلومات

Building a Business Intelligence Solution for Stocks Market

بناء نظام ذكاء الأعمال لسوق الأسهم

Prepared by

Mohammad Orabi

Naser Alouka

Supervised by
Dr.Mouhib Alnoukari

Eng.Anas Abdulazez

Academic Year

2024-2023

شكر و تقدير Acknowledgements

الى أساتذتي الدكتور المهندس مهيب النقري و الدكتور المهندس وسام الخطيب و الدكتور المهندس اكرم مسوح و الدكتور المهندس فادي ابراهيم و الدكتورة المهندسة ثراء اصلان و الدكتور المهندس وسيم احمد و الدكتور المهندس بسيم برهوم و الدكتورة المهندسة كرستين زينة و والدكتورة المهندسة روز المحمد

أتوجه إليكم بأخلص الشكر والتقدير على الدور الهام الذي لعبتموه في رحلتي الأكاديمية في كلية هندسة المعلوماتية. بمناسبة تخرجي، أود أن أعبر عن امتناني العميق لكم جميعًا وللجهود التي بذلتموها لمساعدتي وتوجيهي خلال هذه السنوات.

شكرًا لكم على الاهتمام الذي أبديتموه بتعليمي وتطوير مهاراتي في مجال هندسة المعلوماتية. كنتم أساتذةً ملهمين ومتفانين، وقد كان لكم دور فعّال في توسيع مداركي وتعزيز فهمي للمفاهيم الهامة والتطورات الحديثة في المجال.

أشكركم على توفير بيئة تعليمية حافلة بالتحديات والفرص، حيث تمكنت من تطوير مهاراتي العلمية والتقنية بشكل كبير. كانت رحلتي التعليمية مثمرة وممتعة بفضل الدروس الملهمة والمشاريع العملية التي قمتم بتوجيهي فيها.

أتعهد بأنني سأستغل التعليم الرائع الذي تلقيته بأفضل طريقة ممكنة، وسأعمل جاهدًا لتحقيق النجاح والتفوق في مسيرتي المهنية. سأحمل القيم والمعرفة التي أكتسبتها منكم وأسعى لتحقيق التفوق والابتكار في مجال هندسة المعلوماتية.

أتوجه بالشكر الخاص لكم جميعًا على التفاني والإلهام الذي أظهرتموه في توجيهي ومساعدتي على تحقيق أهدافي الأكاديمية. لقد كنتم أكثر من مجرد أساتذة، بل كنتم أصدقاء ومرشدين لي في مسيرتي الأكاديمية.

أتمنى لكم جميعًا التوفيق والنجاح في مسيراتكم الأكاديمية والمهنية. أعرف أن تأثيركم الإيجابي سيظل حاضرًا في حياتي، وأنني سأحمل ذكراكم بامتنان عميق طوال حياتي المهنية.

مرة أخرى، أشكركم من القلب على كل ما قدمتموه لي كطلابكم، وأتمنى لكم كل خير في المستقبل.

مع خالص التقدير والاحترام

محمد ابراهيم عرابي　　　　ناصر العكة

إهــــــــــداء Dedication

أبي.

أهديك أجمل التحيات وأعذب الكلمات في رسالة الاهداء هذه. أنت الأب الحنون والمثال الذي أعتز به في حياتي. لقد كنت الدعامة القوية والمصدر الأول للحب والدعم. شكرًا لك على كل العناية والتضحيات التي قدمتها لي. أنا فخور بكوني ابنك وأتمنى أن أكون دائمًا عند حسن ظنك.

أمي.

أهديك أعمق الشكر وأعذب العبارات في رسالة الاهداء هذه. أنت الأم المحنونة والقوة العظيمة في حياتي. بفضلك ورعايتك الدائمة، أصبحت الشخص الذي أنا عليه اليوم. أنا ممتن جدًا لحبك العميق وتضحياتك الغير محدودة. أنت ملاكي الحارس ومصدر إلهامي. أشكرك على كل شيء.

أخي سامر.

أهديك أحر التهاني وأطيب التمنيات في رسالة الاهداء هذه. أنت الأخ العزيز والصديق الوفي الذي لا يمكنني الاستغناء عنه. شكرًا لك على الوقت الممتع والذكريات الجميلة التي قضيناها معًا. أنا ممتن لكونك دعامة قوية في حياتي وشريكًا في كل مغامراتي. أتمنى لك النجاح والسعادة الدائمة.

أخواتي سمر وسحر ودعاء ورهام،

أهديكن أطيب التحيات وأجمل الأماني في رسالة الاهداء هذه. أنتن المؤنسات الغاليات ,أنتن أشخاص رائعات وأخوات رائعات، وأنا فخور بكوني أخوكن. شكرًا لكن على الحب والدعم الذي تقدمنه لي دائمًا. أنتن أصدقائي المقربات ورفيقاتي في كل رحلة. أتمنى لكن كل السعادة والنجاح في الحياة.

أصدقائي الأعزاء الدكتور إبراهيم و المهندس إياد والدكتور تميم والمهندس حسن والمهندس شاهر والمهندس يمان والدكتور محمد والأستاذ مناف .

أهديكم التحية والامتنان العميق في رسالة الاهداء هذه. أنتم الأصدقاء المميزون الذين جعلوا حياتي مليئة بالمرح والمغامرات. شكرًا لكم على الوقت الجميل الذي قضيناه معًا وعلى الدعم المتواصل. أنتم دعمي الحقيقي ومصدر إلهامي. أتمنى لكم السعادة والتوفيق الدائم.

أصدقائي الأعزاء المهندس أنس و المهندس جود و المهندس زيد و الدكتور محمود والمهندس بشير والمهندسة انتصار و والمهندس أحمد و المهندسة آية و المهندس أحمد يونس.

أهديكم أطيب التحيات وأجمل الأماني في رسالة الاهداء هذه. أنتم أصدقائي الرائعين الذين جعلوا فترة جامعتنا مليئة بالمرح والنجاح. شكرًا لكم على الذكريات الثمينة والدعم المستمر الذي قدمتموه. أنتم رفاق دربي وزملائي المحبوبين. أتمنى لكم مستقبلًا مشرقًا ونجاحًا لا يعد ولا يحصى.

مع خالص الحب والتقدير،

محمد ابراهيم عرابي.

SUPERVISION CERTIFICATION

I certify that the preparation of this project entitled **Building Business Intelligence Solution for Stocks Market**, prepared by **Mohammad Ibrahim Orabi**, and **Naser Aloka** was made under my supervision at *Department of Software & Information System Eng. – Faculty of Computer & Informatics Engineering* in partial fulfillment of the Requirements for the Degree of *Bachelors of Software & Information System Engineering*.

Dr. Mouhib Alnoukari
Supervisor
4/6/2023

Abstract:

Due to the rapid growth of new technologies, the Business Intelligence (BI) markets growing as well. Adoption of Business Intelligence system has become one of the most important technological and organizational innovations in modern organizations, and cornerstone of decision-making processes.

This project consists of implementing a business intelligence solution to stocks market data, which increases decisions accuracy by providing accurate knowledge and easy-to-read reports to decision-makers.

This is accomplished by building a data warehouse relying on the extracted data, then analysis and business intelligence tools are applied to acquire the maximum benefits.

****

ملخص:

نظرًا للنمو السريع للتقنيات الجديدة ، فإن سوق ذكاء الأعمال (BI) ينمو أيضًا. أصبح اعتماد نظام ذكاء الأعمال أحد أهم الابتكارات التكنولوجية والتنظيمية في المنظمات الحديثة ، وحجر الزاوية في عمليات صنع القرار.

يتكون هذا المشروع من تنفيذ حل ذكاء الأعمال لبيانات الأسهم ، مما يزيد من دقة القرارات من خلال توفير معرفة دقيقة وتقارير سهلة القراءة لصناع القرار.

يتم تحقيق ذلك من خلال بناء مستودع بيانات يعتمد على البيانات المستخرجة ، ثم يتم تطبيق أدوات التحليل وذكاء الأعمال للحصول على أقصى الفوائد.

Table of contents

Contents

Table of Figure

Abbreviation table

BI	Business intelligence
DW	Data warehouse
ETL	Extract, transform, and load
ETFs	Exchange-Traded Funds

0.1 Introduction

Within this chapter, we will present an overview of the problem statement, the objectives of the project, and outline the proposed solution.

Problem Statement

The US stock market faces several challenges that business intelligence aims to address. These challenges can be summarized as follows:

1. Complex Data: The US stock market generates a vast amount of complex data from various sources, including stock prices, trading volumes, financial statements, and market news. This complexity makes it difficult for traders and investors to handle and process the data effectively.

2. Unstructured Information: The market is flooded with a plethora of information, which often lacks structure and is not easily understandable. This information overload can overwhelm market participants and hinder their ability to make informed decisions.

3. Lack of Predictive Analytics: Traditionally, there has been a scarcity of robust predictive analytics in the US stock market. Predictive analytics involves using historical data and advanced algorithms to forecast future market trends, stock prices, and other relevant indicators. The absence of such analytics hampers traders and investors in making accurate predictions and proactive decisions.

Proposed Solution

To address the challenges identified in the project regarding the US stock market, several solutions can be implemented using business intelligence techniques. These solutions include:

1. Advanced Data Processing: Business intelligence tools provide advanced data processing capabilities, allowing for efficient handling of complex data. Techniques such as data integration, data cleansing, and data transformation can be applied to consolidate and standardize the diverse data sources from the stock market. This ensures that the data is accurate, consistent, and ready for analysis.

2. Data Visualization and Reporting: Business intelligence tools offer robust data visualization and reporting features. Visualizations such as charts, graphs, and interactive dashboards can be used to present complex stock market information in a visually appealing and understandable manner. Users can explore the data, detect patterns, and gain insights through interactive visualizations, making it easier to comprehend the vast amount of information available.

3. Predictive Analytics: Business intelligence tools incorporate predictive analytics capabilities to address the lack of predictive analysis in the stock market. These tools can leverage historical data and apply advanced algorithms to generate predictive models. By forecasting future market trends, stock prices, and other indicators, traders and investors can make more accurate predictions and informed decisions. Predictive analytics helps identify potential investment opportunities, assess risks, and

optimize trading strategies.

4. Data-driven Insights: Business intelligence tools empower users to derive data-driven insights from the vast amount of information available. By leveraging advanced analytics techniques, such as data mining and machine learning, these tools can uncover hidden patterns, correlations, and trends within the stock market data. These insights can guide traders and investors in making informed decisions, identifying potential investment opportunities, and managing risks effectively.

The Goal of the Project

1. Finding common relationships between technical indicators and stock prices
2. Improving decision-making: The business intelligence system helps analyze available data and extract important patterns and trends in the US stock market. This enables investors and traders to make informed investment decisions and achieve better results.
3. Risk Reduction: E-commerce financial data, market capitalization estimates, and other market factors over time
4. Providing insights to decision makers (investors) through reports showing stock performance in each field
5. Provide forecasts for the most common stock prices

The Importance Of The Project:

The importance of the research lies in the following point work on important and relative lyre cent concepts, which are business intelligence systems in decision support:

- Improving Decision Making: The Business Intelligence Project helps analyze data on the US stock market and turn it into valuable information. These decisions can be used in more investment decisions

- Monitoring and Control: The Business Intelligence Project makes it possible to monitor fraud in the US stock market. Analyze historical and current stock data and track patterns and changes in the market, which helps to identify opportunities and avoid risks.
- Providing Strategic Insights: The Business Intelligence Project provides signals and insights for investing in the US stock market. Analyzing the influences of influences affecting equality and expectations
- Improved risk management: Advanced analytics can be used to manage risks in the stock market. Financial data, forecasts and economic factors can be analyzed to identify potential risks and develop strategies to deal with them and reduce investors' exposure to financial risks.
- Improving Decision Making: The Business Intelligence Project helps analyze data on the US stock market and turn it into valuable information. These decisions can be used in more investment decisions
- Monitoring and Control: The Business Intelligence Project makes it possible to monitor fraud in the US stock market. Analyze historical and current stock data and track patterns and changes in the market, which helps to identify opportunities and avoid risks.

Chapter 1
Theoretical Study

1.1 Introduction

The world is becoming more and more data-driven, with endless amounts of data available to work with. Big companies like Google ,Meta ,and Microsoft use data to make decisions, but they're not the only ones.

The importance of data analytics in any sector is compounded, creating enormous quantities of knowledge that can provide useful insights into the field. In the last ten years, this has led to a surge in the data market. In order to gain decision-making insights, the compilation of data can be supplemented by its analysis. Data analytics help organizations and businesses gain insight into the enormous amount of knowledge they need for further production and growth.

"Information is the oil of the 21st century, and analytics is the combustion engine." – Peter Syndergaard, senior vice president, Gartner Research.

Business intelligence (BI) software can help by combining online analytical processing (OLAP), location intelligence, enterprise reporting, and more. BI software offers enterprise businesses the opportunity to connect disparate data sources into one unified source, collate and structure the data, and offer an interface for end-users to extract reports and dashboards that can drive more informed business decisions.

1.2 Concepts

1.2.1 Business Intelligence:

Business intelligence (BI) is a technology-driven process for analyzing data and delivering actionable information that helps executives, managers and workers make informed business decisions. As part of the BI process, organizations collect data from internal IT systems and external source

prepare it for analysis, run queries against the data and create data visualizations, BI dashboards and reports to make the analytics results available to business users for operational decision making and strategic planning.[1]

The ultimate goal of BI initiatives is to drive better business decisions that enable organizations to increase revenue, improve operational efficiency and gain competitive advantages over business rivals. To achieve that goal, BI incorporates a combination of analytics, data management and reporting tools, plus various methodologies for managing and analyzing data.[1]

BI data can include historical information and real-time data gathered from source systems as it's generated, enabling BI tools to support both strategic and tactical decision-making processes. Before it's used in BI applications, raw data from different source systems generally must be integrated, consolidated and cleansed using data integration and data quality management tools to ensure that BI teams and business users are analyzing accurate and consistent information

1.2.2 Business Intelligence Environment

Business intelligence environment Business intelligence (BI) is all about converting large amounts of corporate data into useful information, thereby triggering some profitable business action with the help of knowledge acquired through BI analysis.[3]

Implements BI is a long process and it requires a lot of analysis and investment. A typical BI environment involves business models, data models, data sources, ETL, tools needed to transform and organize the data into useful information, target data warehouse, data marts, OLAP analysis and reporting tools. Setting up a BI environment not only rely on tools, techniques and processes, it also requires skilled business people to carefully drive these in the right direction. Care should be taken in understanding the business requirements, setting up the targets, analyzing and defining the various processes associated with these, determining what kind of data needed to be analyzed, determining the source and target for that data, defining how to integrate that data for BI analysis and determining and gathering the tools to achieve this goal.[1]

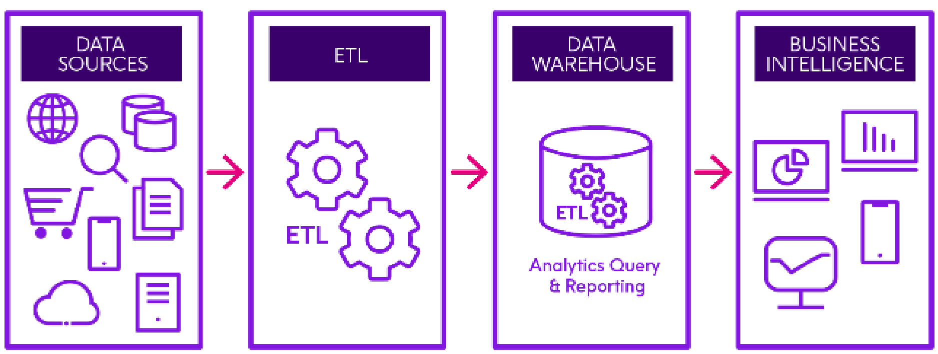


Figure 1: BI environment [9]

1.2.3 Benefits Of Business Intelligence

A successful BI program produces a variety of business benefits in an organization. For example, BI enables C-suite executives and department managers to monitor business performance on an ongoing basis so they can act quickly when issues or opportunities arise. Analyzing customer data helps make marketing, sales and customer service efforts more effective. Supply chain, manufacturing and distribution bottlenecks can be detected before they cause financial harm. HR managers are better able to monitor employee productivity, labor costs and other workforce data.[1]

Overall, the key benefits that businesses can get from BI applications include the ability to:

- speed up and improve decision-making;
- optimize internal business processes;
- increase operational efficiency and productivity;
- spot business problems that need to be addressed;
- identify emerging business and market trends;
- develop stronger business strategies;
- drive higher sales and new revenues;

BI initiatives also provide narrower business benefits -- among them, making it easier for project managers to track the status of business projects and for organizations to gather competitive intelligence on their rivals. In addition, BI, data management and IT teams themselves benefit from business intelligence, using it to analyze various aspects of technology and analytics operations.

1.4.4 Data Warehouse (DW)

Data warehousing and on-line analytical processing (OLAP) are essential elements of decision support, which has increasingly become a focus of the database industry. Many commercial products and services are now available, and all of the principal database management system vendors now have offerings in these areas. Decision support places some rather different requirements on database technology compared to traditional on- line transaction processing applications.

This paper provides an overview of data warehousing and OLAP technologies, with an emphasis on their new requirements. [8] We describe back-end tools for extracting, cleaning and loading data into a data warehouse; multidimensional data models typical of OLAP;

front end client tools for querying and data analysis; server extensions for efficient query processing; and tools for metadata management and for managing the warehouse. In addition to surveying the state of the art, this paper also identifies some promising research issues, some of which are related to problems that the database research community has worked on for years, but others are only just beginning to be addressed. This overview is based on a tutorial that the authors presented at the VLDB Conference, 1996. A data warehouse is a "subject-oriented, integrated, time varying, nonvolatile collection of data that is used
primarily in organizational decision making." 1 Typically, the data warehouse is maintained separately from the organization's operational databases. There are many reasons for doing this. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases.[4]

Data warehouses, in contrast, are targeted for decision support. Historical, summarized and consolidated data is more important than detailed, individual records. Since data warehouses contain consolidated data, perhaps from several operational databases, over potentially long periods of time, they tend to be orders of magnitude larger than operational databases; enterprise data warehouses are projected to be hundreds of gigabytes to terabytes in size. The workloads are query intensive with mostly ad hoc, complex queries that can access millions of records and perform a lot of scans, joins, and aggregates. Query throughput and response times are more important than transaction throughput.[8]

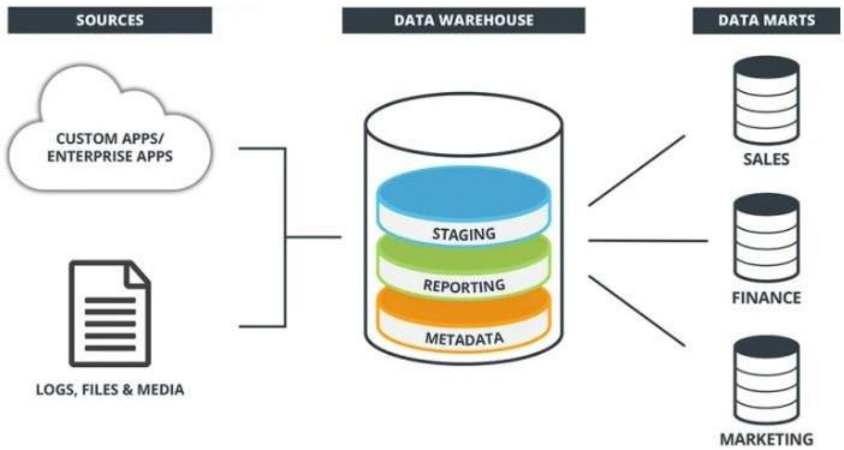


Figure 2: Data warehouse [8]

1.4.4.1 *Why Star Schema*

A star schema is a database organizational structure optimized for use in a data warehouse or business intelligence that uses a single large fact table to store transactional or measured data, and one or more smaller dimensional tables that store attributes about the data.

It is called a star schema because the fact table sits at the center of the logical diagram, and the small dimensional tables branch off to form the points of the star.

A fact table sits at the center of a star schema database, and each star schema database only has a single fact table. The fact table contains the specific measurable (or quantifiable) primary data to be analyzed, such as sales records, logged performance data or financial data.

It may be transactional -- in that, rows are added as events happen -- or it may be a snapshot of historical data up to a point in time.

1.4.4.2 How Star Schema Works [2]

The fact table stores two types of information: numeric values and dimension attribute values. Using a sales database as an example: Numeric value cells are unique to each row or data point and do not correlate or relate to data stored in other rows. These might be facts about a transaction, such as an order ID, total amount, net profit, order quantity or exact time The dimension attribute values do not directly store data, but they store the foreign key value for a row in a related dimensional table. Many rows in the fact table will reference this type of information. So, for example, it might store the sales employee ID, a date value, a product ID or a branch office ID.

Dimension tables store supporting information to the fact table. Each star schema database has at least one dimension table, but will often have many. Each dimension table will relate to a column in the fact table with a dimension value, and will store additional information about that value. Constructing a start schema should be carefully done. Each table should have either fact data or dimension data, and avoid mixing the two. Consider the total number of dimension tables to maximize performance. Also, consider the granularity of the data captured to optimize for the types of queries that will be run.

Optimized for querying large data sets, data warehouses and data marts, star schemas support online analytical processing (OLAP) cubes, analytic application, ad hoc queries and business intelligence (BI). They also support count, sum, average and other rapid aggregations of many fact records. Users can filter and group (sliced and diced) these aggregations by dimensions.

After the data was loaded into the data warehouse, we needed to connect the SQL server DW to the power BI environment to build the dashboards.

1.4.5 Architecture And End-To-End Process:

It includes tools for extracting data from multiple operational databases and external sources; for cleaning, transforming and integrating this data; for loading data into the data warehouse; and for periodically refreshing the warehouse to reflect updates at the sources and to purge data from the warehouse, perhaps onto slower archival storage.

In addition to the main warehouse, there may be several departmental data marts. Data in the warehouse and data marts is stored and managed by one or more warehouse servers, which present multidimensional

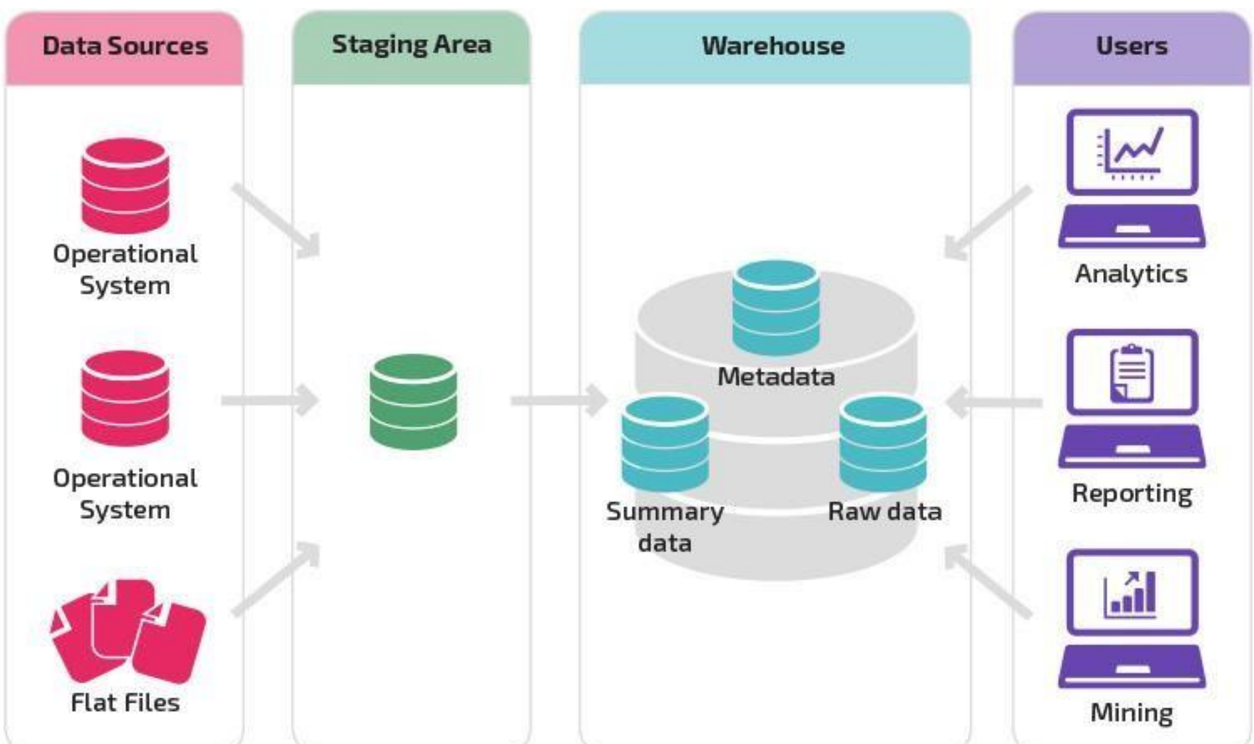


Figure 3: Data warehouse architecture [11]

views of data to a variety of front-end tools: query tools, report writers, analysis tools, and data mining tools.

Finally, there is a repository for storing and managing metadata, and tools for monitoring and administering the warehousing system.[8]

Designing and rolling out a data warehouse is a complex process, consisting of the following activities.

- Define the architecture, do capacity planning, and select the storage servers, database and OLAP servers, and tools.
- Integrate the servers, storage, and client tools.
- Design the warehouse schema and views.
- Define the physical warehouse organization, data placement, partitioning, and access methods.
- Connect the sources using gateways, ODBC drivers, or other wrappers.
- Design and implement scripts for data extraction, cleaning, transformation, load, and refresh.
- Populate the repository with the schema and view definitions, scripts, and other metadata.
- Design and implement end-user applications.
- Roll out the warehouse and applications.[4]

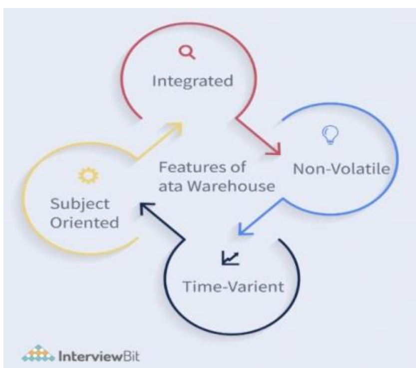


Figure 4: Data warehouse attributes.[8]

1.4.6 Visualization

Data Visualization is a process of taking raw data and transforming it into graphical or pictorial representations such as charts, graphs, diagrams, pictures, and videos which explain the data and allow you to gain insights from it. So, users can quickly analyze the data and prepare reports to make business decisions effectively.

1.4.7 *The Importance of Data Visualization*

We are inherently in the visual world where pictures or images speak more than words. So, it is easy to visualize a large amount of data using graphs and charts than depending on reports or spreadsheets.
Data visualization is a quick and easy way to convey concepts to the end-users, and you can do experiments with different scenarios by making slight changes.

1.4.8 *Extract, Transform, and Load (ETL)*

An ETL tool is used to execute an integration project, and it includes three steps: extraction, transformation, and loading.
Thus, an ETL tool extracts data from disparate sources, transforms it to make it compatible with the destination system, and then loads it into the destination system. This destination could be a data warehouse, data lake, database, or any other application system
A properly designed ETL system extracts data from the source systems, enforces data quality and consistency standards, conforms data so that separate sources can be used together, and finally delivers data in a presentation-ready [9]

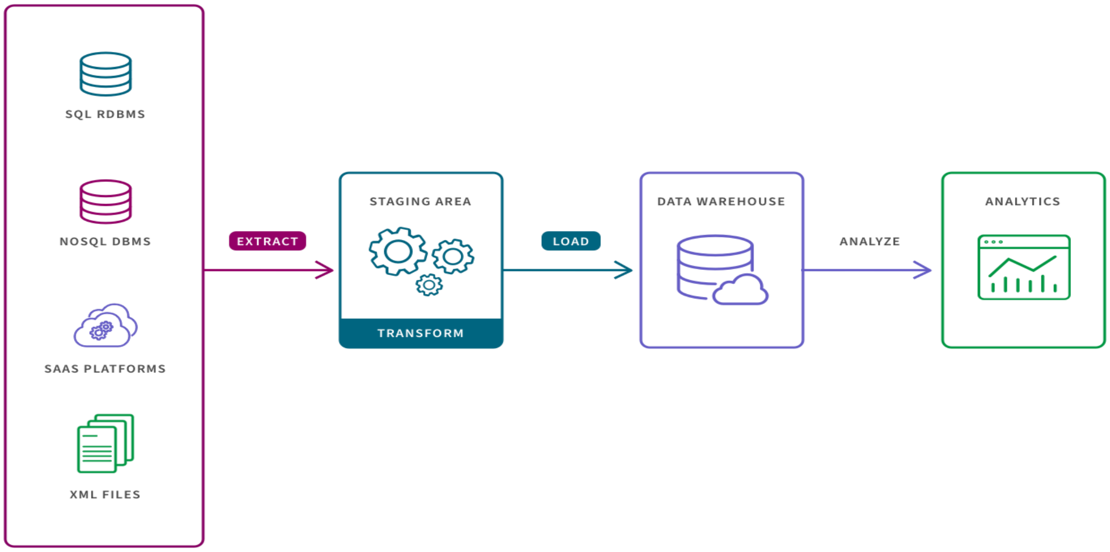


Figure 5 : Extract, transform, and load) ETL) [7]

format so that application developers can build applications and end users can make decisions.

Since the data extraction takes time, it is common to execute the three phases in pipeline. While the data is being extracted, another transformation process executes while processing the data already received and prepares it for loading while the data loading begins without waiting for the completion of the previous phases.

1.4.8.1 Extract

In this step of ETL architecture, data is extracted from the source system into the staging area. Transformations if any are done in staging area so that performance of source system in not degraded. Also, if corrupted data is copied directly from the source into Data warehouse database, rollback will be a challenge. Staging area gives an opportunity to validate extracted data before it moves into the Data warehouse.

Data warehouse needs to integrate systems that have different DBMS, Hardware, Operating Systems and Communication Protocols. Sources could include legacy applications like Mainframes, customized applications, point of contact devices like ATM, Call switches, text files, spreadsheets, ERP, data from vendors, partners amongst others. [9]

Hence one needs a logical data map before data is extracted and loaded physically. This data map describes the relationship between sources and target data.

Three Data Extraction methods:

1. Full Extraction.
2. Partial Extraction- without update notification.
3. Partial Extraction- with update notification. Irrespective of the method used, extraction should not affect performance and response time of the source systems. These source systems are live production databases. Any slow down or locking could affect company's bottom line.[9]

1.4.8.2 Transform

Data extracted from source server is raw and not usable in its original form. Therefore, it needs to be cleansed, mapped and transformed. In fact, this is the key step where ETL process adds value and changes data such that insightful BI reports can be generated.

It is one of the important ETL concepts where you apply a set of functions on extracted data. Data that does not require any transformation is called as direct move or pass-through data.

In transformation step, you can perform customized operations on data. For instance, if the user wants sum-of-sales revenue which is not in the

database One of the used tools is Python NumPy package:
NumPy is the fundamental package for scientific computing in Python. It is a Python library that provides a multidimensional array object, various derived objects (such as masked arrays and matrices), and an assortment of routines for fast operations on arrays, including mathematical, logical, shape manipulation, sorting, selecting, I/O, discrete Fourier transforms, basic linear algebra, basic statistical operations, random simulation and much more.

At the core of the NumPy package, is the array object. This encapsulates dimensional arrays of homogeneous data types, with many operations being performed in compiled code for performance. There are several important differences between NumPy arrays and the standard Python sequences:
NumPy arrays have a fixed size at creation, unlike Python lists (which can grow dynamically). Changing the size of an array will create a new array and delete the original.
The elements in a NumPy array are all required to be of the same data type, and thus will be the same size in memory. The exception: one can have arrays of (Python, including NumPy) objects, thereby allowing for arrays of different sized elements.

NumPy arrays facilitate advanced mathematical and other types of operations on large numbers of data. Typically, such operations are executed more efficiently and with less code than is possible using Python's built-in sequences.

A growing plethora of scientific and mathematical Python-based packages are using NumPy arrays; though these typically support Python-sequence input, they convert such input to NumPy arrays prior to processing, and they often output NumPy arrays. In other words, in order to efficiently use much (perhaps even most) of today's scientific/mathematical Python-based software, just knowing how to use Python's built-in sequence types is insufficient - one also needs to know how to use NumPy arrays.

Another one is Python Pandas:

pandas is a software library written for the Python programming language

for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free software released under the three-clause BSD license. The name is derived from the term "panel data", an econometrics term for data sets that include observations over multiple time periods for the same individuals. Https Its name is a play on the phrase "Python data analysis" itself. Wes McKinney started building what would become pandas at AQR Capital while he was a researcher there from 2007 to 2010.

1.4.8.3 Load

Loading data into the target Datawarehouse database is the last step of the ETL process. In a typical Data warehouse, huge volume of data needs to be loaded in a relatively short period (nights). Hence, load process should be optimized for performance.

In case of load failure, recover mechanisms should be configured to restart from the point of failure without data integrity loss.

Data Warehouse admins need to monitor, resume, cancel loads as per prevailing server performance.

Types of Loading:

- **Initial Load** : populating all the Data Warehouse tables
- **Incremental Load** : applying ongoing changes as when needed periodically.
- **Full Refresh** : erasing the contents of one or more tables and reloading with fresh data. [9]

Figure 6: Load Process in Power bi [2]

1.4.9 Data mining

1.4.9.1 Introduction

Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Using a broad range of techniques, you can use this information to increase revenues, cut costs, improve customer relationships, reduce risks and more. The process of digging through data to discover hidden connections and predict future trends has a long history. Sometimes referred to as "knowledge discovery in databases," the term "data mining" wasn't coined until the 1990s. But its foundation comprises three intertwined scientific disciplines: statistics (the numeric study of data

relationships), artificial intelligence (human-like intelligence displayed by software and/or machines) and machine learning (algorithms that can learn from data to make predictions). What was old is new again, as data mining technology keeps evolving to keep pace with the limitless potential of big data and affordable computing power. Over the last decade, advances in processing power and speed have enabled us to move beyond manual, tedious and time-consuming practices to quick, easy and automated data analysis. The more complex the data sets collected, the more potential there is to uncover relevant insights. Retailers, banks, manufacturers, telecommunications providers and insurers, among others, are using data mining to discover relationships among everything from price optimization, promotions and demographics to how the economy, risk, competition and social media are affecting their business models, revenues, operations and customer relationships. Businesses that utilize data mining are able to have a competitive advantage, better understanding of their customers, good oversight of business operations, improved customer acquisition, and new business opportunities. Different industries will have different benefits from their data analytics. Some industries are looking for the best ways to get new customers, others are looking for new marketing techniques, and others are working to improve their systems. The data mining process is what gives businesses the opportunities and understanding for how to make their decisions, analyze their information, and move forward. Today, stocks market operate in a highly competitive and complex environment. With rapid technology development and cheaper IT equipment, the amount of data stored in educational databases increases rapidly, but if this data is not further analyzed, it remains only huge amounts of data. Data mining tools, methods and techniques, allow us to analyze this data and find hidden patterns and information. Data mining is used to detect patterns and relationships in data to improve decision-making processes. It is an interdisciplinary area that brings together techniques from statistics,

artificial intelligence, neural networks, database systems, machine learning, pattern recognition, data visualization, knowledge acquisition and information theory (Sumathi and Sivananda, 2013). The application of data mining is wide and various. It is used in finance for analyzing customer behavior data to increase customer loyalty, it also helps in finding hidden correlations between various financial indicators to detect suspicious activities. By collecting historical data and turning them into useful and valid information it can detect fraudulent and non-fraudulent actions. Using data mining in healthcare can help in discovering the relationships between diseases and the effectiveness of treatments. It also supports healthcare insurers in detecting fraud. It is used in crime agencies for finding patterns related to money laundering, narcotics trafficking, etc. A common use of data mining in telecommunication is in analyzing customer data to improve profitability by providing customized services and also to reduce customer churn by understanding demographic characteristics and predicting customer behavior. The results of the data mining process can be used to develop appropriate marketing campaigns and pricing strategies. In marketing and sales, data mining techniques are used to find the hidden patterns from historical purchasing data. Results of data mining provide information on combinations of products purchased together in market basket analysis and are used to identify customer's behavior buying patterns. It is also used for the prediction of future trends and customer purchase habits. The banking industry usually uses data mining methods to predict customer churn, as well as in fraud and bankruptcy detection.

There are also disadvantages of data mining, namely in user privacy and security. It has to be clear how and with whom the information will be used and shared. Data mining tools and techniques work with very big amounts of data, so there is great cost at the implementation stage. It requires great IT experts for preprocessing data and finding the best model and technique for analysis.

The techniques of data mining are not 100% accurate, so it may cause serious consequences and expenses. This work is based on special use of data mining algorithms, techniques and concepts in the educational environment, called educational data mining (EDM). The remaining of the paper is organized as follows. After a summary of the history and definition of educational data mining, the process is presented in "Educational data mining process" section, by detailing the data pre-processing and the knowledge extraction phase, and by describing all phases. After that, selected methods and techniques, as well as its use in the educational sector are described in” Methods and techniques” section. In the next section, related work is covered and in the most relevant section, benefits and applications of educational data mining are presented and discussed, along with relevant research in the application of educational data mining. Final remarks conclude into “Conclusion” section.

1.4.9.2 classification analysis

Classification analysis plays a crucial role in data mining for the US stock market. Here are some key reasons for its importance:

- Pattern Recognition: Classification analysis helps in identifying and recognizing patterns in the stock market data. By classifying stocks into different categories based on their attributes and historical data, it becomes easier to identify trends, patterns, and potential opportunities.

- Decision Making: Classification analysis provides valuable insights for decision making. It helps investors and traders in making informed decisions regarding buying, selling, or holding specific stocks based on their classification into different classes, such as high-growth stocks, dividend stocks, or value stocks.

- Risk Assessment: Classification analysis aids in assessing the risk associated with different stocks. By categorizing stocks based on their risk levels, such as low risk, moderate risk, or high risk, investors can make more calculated decisions and manage their portfolio accordingly.

- Predictive Modeling: Classification analysis is often used in building predictive models for stock market forecasting. By training the model on historical data and classifying stocks based on various factors like company fundamentals, market conditions, and technical indicators, it becomes possible to predict future stock performance and make investment predictions.

- Portfolio Optimization: Classification analysis assists in optimizing investment portfolios. By classifying stocks into different asset classes or sectors, investors can diversify their portfolios and allocate their investments strategically, aiming for a balanced and well-diversified portfolio.

1.4.9.3 *CLUSTERING ANALYSIS:*

The importance of clustering analysis in data mining for the US stock market lies in its ability to reveal hidden patterns and groupings within the stock data. Here are some key reasons why clustering analysis is significant:

- Pattern Discovery: Clustering analysis helps in identifying similar patterns or behaviors among stocks. By grouping stocks based on their similarities in terms of price movements, trading volumes, or other relevant factors, clustering analysis can uncover patterns that may not be evident through traditional analysis methods.

- Portfolio Management: Clustering analysis assists in portfolio management by providing insights into diversification strategies. By clustering stocks with similar risk and return characteristics, investors can create well-diversified portfolios to mitigate risk and maximize returns.

- Risk Assessment: Clustering analysis enables risk assessment by identifying groups of stocks with similar risk profiles. This information helps investors and analysts in assessing the risk associated with specific stock clusters and adjusting their investment strategies accordingly.

- Market Segmentation: Clustering analysis helps in segmenting the stock market based on different criteria, such as industry sectors or market capitalization. This segmentation allows for targeted analysis and decision-making, as different segments may exhibit distinct behaviors and trends.

- Trading Strategies: Clustering analysis can be used to develop trading strategies based on the identified clusters. By understanding the behavior of stocks within each cluster, traders can create specific trading rules or algorithms to exploit the patterns and generate better trading outcomes.

1.4.9.4 Regression analysis

Regression analysis is of great importance in data mining for the US stock market. It is a statistical technique that helps in understanding and modeling the relationship between a dependent variable (such as stock price) and one or more independent variables (such as market trends, financial indicators, or economic factors). Here are some key reasons why regression analysis is valuable in data mining for stocks

- Prediction: Regression analysis allows us to make predictions about the future values of the dependent variable based on the observed values of the independent variables. By analyzing historical data, regression models can be used to forecast stock prices, identify potential trends, and make informed investment decisions.

- Risk Assessment: Regression analysis helps in assessing the risk associated with stock investments. By analyzing the relationship between independent variables and stock prices, regression models can identify factors that influence stock volatility, enabling investors to evaluate and manage risks effectively.

- Performance Evaluation: Regression analysis enables the evaluation of the performance of stocks or investment portfolios. By analyzing the relationship between independent variables and stock returns, regression models can assess the effectiveness of investment strategies, identify factors contributing to returns, and compare the performance of different stocks or portfolios.

- Factor Analysis: Regression analysis helps in identifying the factors that significantly impact stock prices. By examining the regression coefficients, significance levels, and effect sizes, data miners can determine the relative importance of different variables in explaining variations in stock prices. This information can be valuable for understanding market dynamics and developing effective trading strategies.

➢ Decision Support: Regression analysis provides a quantitative basis for decision-making in the stock market. By analyzing the relationship between independent variables and stock prices, regression models can support investment decisions, asset allocation, and portfolio optimization strategies.

Figure 30: Data MINING[8]

Chapter 2
Design and Analytic Study

2.1 Introduction

Business intelligence (BI) comprises the strategies and technologies used by enterprises for the data analysis of business information. BI systems leverage software and services beginning with collecting data, cleansing data, integration of data, and then using data visualization to transform data into actionable insights that might be helpful while decision making process. [1]

Implementing a business intelligence system requires careful planning to assure that it meets expectations. These are the basic steps:

- ✓ Identify End-User Requirements
- ✓ Identify the Data Sources (Data sources: which are reports in excel forms.)
- ✓ ETL (Extract, Transform, and Load): in this area we managed gather them, perform cleaning operations, and then visualize the data in a better form.
- ✓ Design the Data Mode
- ✓ The Data Warehouse: Which contains all the relevant data, and can be described as an enterprise database.
- ✓ The OLAP environment.
- ✓ The BI application.

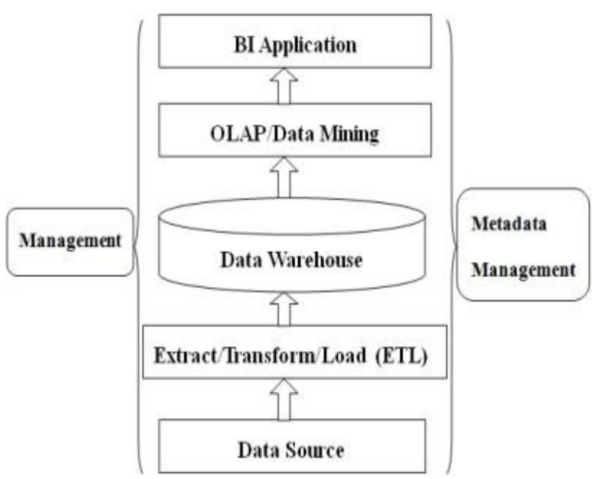


Figure (32): BI application architecture[10

2.2. Identify End-User Requirements:

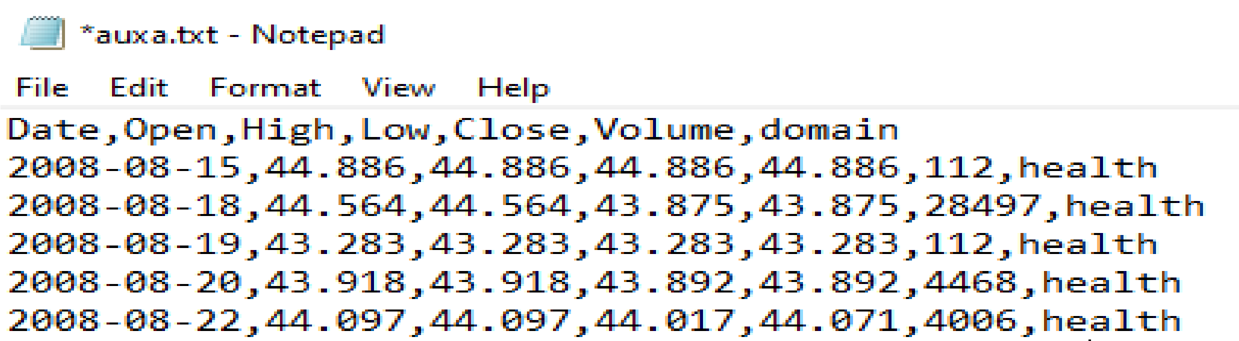
1. What is the safest area for domain?
2. What is the most risky for domain?
3. What is the most profitable for stocks?
4. What is the biggest loser in terms of stocks?
5. What are the most common shares?
6. What is the impact of the global financial crisis that occurred in 2008 on the stock market?
7. What is the impact of political and administrative decisions on the stock market?
8. What are the best times to invest?

2.3 Identify The Data Sources:

We obtained the stock data from the (Kaggle) website, and the data consisted of 8,742 text files, where the initial data size was 0.8 gigabytes.

We used the Python language and merged these files into Excel files, which contained 16 million records

And this data contained (date, open, close, high, low, volume, range)


*auxa.txt - Notepad

File Edit Format View Help

```
Date,Open,High,Low,Close,Volume,domain
2008-08-15,44.886,44.886,44.886,44.886,112,health
2008-08-18,44.564,44.564,43.875,43.875,28497,health
2008-08-19,43.283,43.283,43.283,43.283,112,health
2008-08-20,43.918,43.918,43.892,43.892,4468,health
2008-08-22,44.097,44.097,44.017,44.071,4006,health
```

Figure8: Data obtained from kaggle

2.4 Extract,Transform,Load (ETL)

The first challenge we faced was converting the text files into excel files
with the addition of ID and the addition of the stock name and type
This process was done using the Python language

id	Date	Open	High	Low	Close	Volume	domain	name stocks	type
2	1999-11-18	30.713	33.754	27.002	29.702	66277506	health	a.us	stocks
3	1999-11-19	28.986	29.027	26.872	27.257	16142920	health	a.us	stocks
4	1999-11-22	27.886	29.702	27.044	29.702	6970266	health	a.us	stocks
5	1999-11-23	28.688	29.446	27.002	27.002	6332082	health	a.us	stocks
6	1999-11-24	27.083	28.309	27.002	27.717	5132147	health	a.us	stocks
7	1999-11-26	27.594	28.012	27.509	27.807	1832635	health	a.us	stocks
8	1999-11-29	27.676	28.65	27.38	28.432	4317826	health	a.us	stocks
9	1999-11-30	28.35	28.986	27.634	28.48	4567146	health	a.us	stocks
10	1999-12-01	28.48	29.324	28.273	28.986	3133746	health	a.us	stocks

Figure9: transportation data

2.4.1 Data Cleansing Phase:

At this stage, we did not encounter many difficulties, because the data you are somewhat clean
We found some empty values that were thousands and we filled them in by calculating the arithmetic mean of the stock

2.4.2 Add data

We have added a lot of new data to our database Where we have added all the technical indicators that will help us achieve a deeper and more accurate analysis. We have reviewed some references that helped us understand the calculation of these technical indicators From this point of view, we have used the Python language to apply these calculations to the data that we have

2.5 Design the Data Model:

The Data warehouse was built using mySQL and we used the star schema to model the DW. Star Schema in data warehouse, in which the center of the star can have one fact table and a number of associated dimension tables. It is known as star schema as its structure resembles a star. The Star Schema data model is the simplest type of Data Warehouse schema. It is also known as Star Join Schema and is optimized for querying large data sets. The design one fact table, which is connected further with multiple dimension tables. The fact table contains all the facts, while the dimension table contains the objects

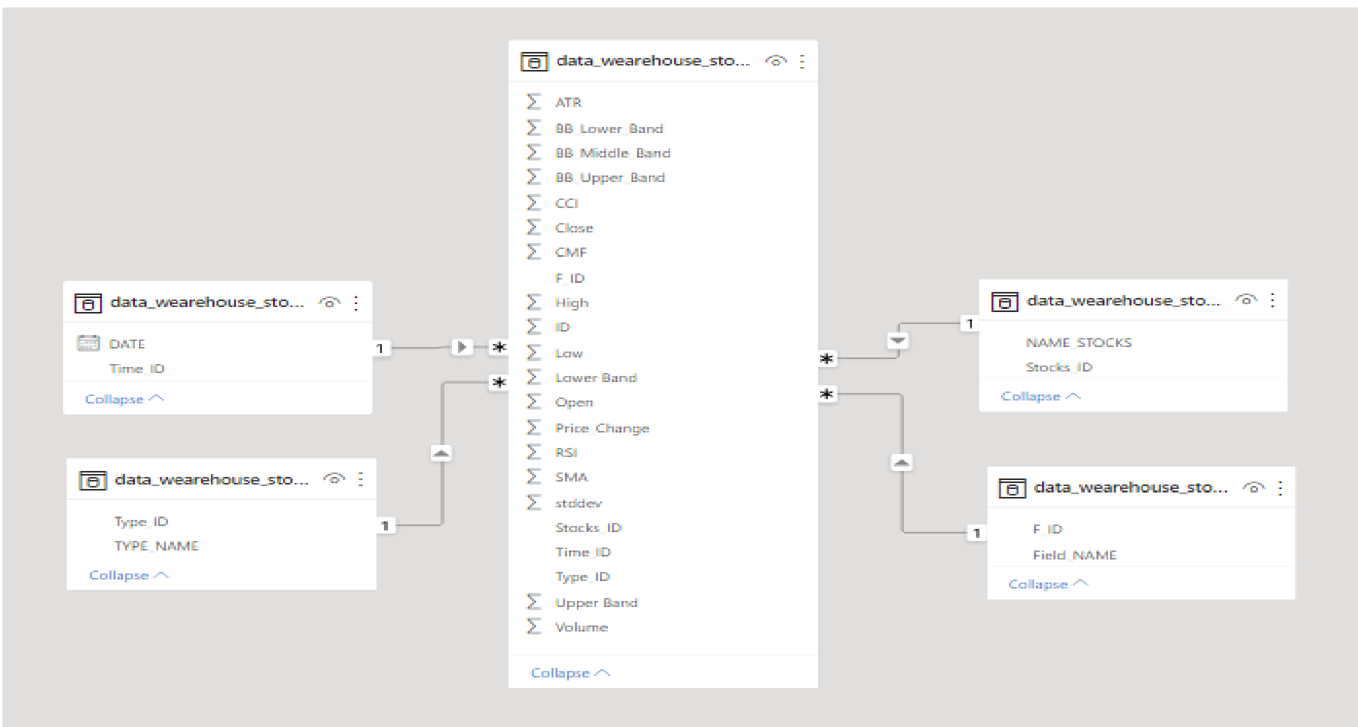


Figure 10 : Data warehouse architecture-2-

2.6 *Tools Used in My Project for Data Analysis and Visualization*

The key tools employed in project include Python programming language for data cleaning and transformation, SQL for building the data warehouse, and Power BI for visualization, analysis, and predictive modeling.

- Python:

Python played a vital role in my project as it provided a versatile and powerful programming language for data processing tasks. I utilized various libraries in Python, such as pandas, NumPy, and scikit-learn, to clean and transform the data. Python's rich ecosystem of packages allowed me to handle data manipulation, feature engineering, and preprocessing efficiently. Additionally, I used Python to load the prepared data into the data warehouse.

- SQL:

SQL (Structured Query Language) was instrumental in building the data warehouse for my project. I leveraged SQL to create the necessary database tables, define relationships, and perform data integration and aggregation. With SQL, I could write complex queries to extract relevant data from different sources and load it into the data warehouse. SQL's querying capabilities enabled efficient data retrieval and analysis.

- Power BI:

Power BI served as a comprehensive tool for visualizing and analyzing the data stored in the data warehouse. I utilized Power BI's interactive dashboards and visualizations to gain insights into the stock market data. With Power BI, I could create informative charts, graphs, and reports that helped in identifying patterns, trends, and correlations. Moreover, Power BI's advanced analytics capabilities allowed me to apply prediction algorithms and generate forecasts based on historical data.

- ✓ The combination of Python, SQL, and Power BI proved to be a powerful toolkit for my project on stock market analysis. Python facilitated data cleaning, transformation, and loading, while SQL enabled the construction of a robust data warehouse. Power BI provided interactive and visually appealing representations of the analyzed data, along with predictive modeling capabilities. By utilizing these tools effectively, I was able to gain valuable insights and make informed decisions based on the analyzed stock market data.

Chapter 3
Project Implementation

3.1 Samples of Our Analysis and Visualizing

3.1.1 General report

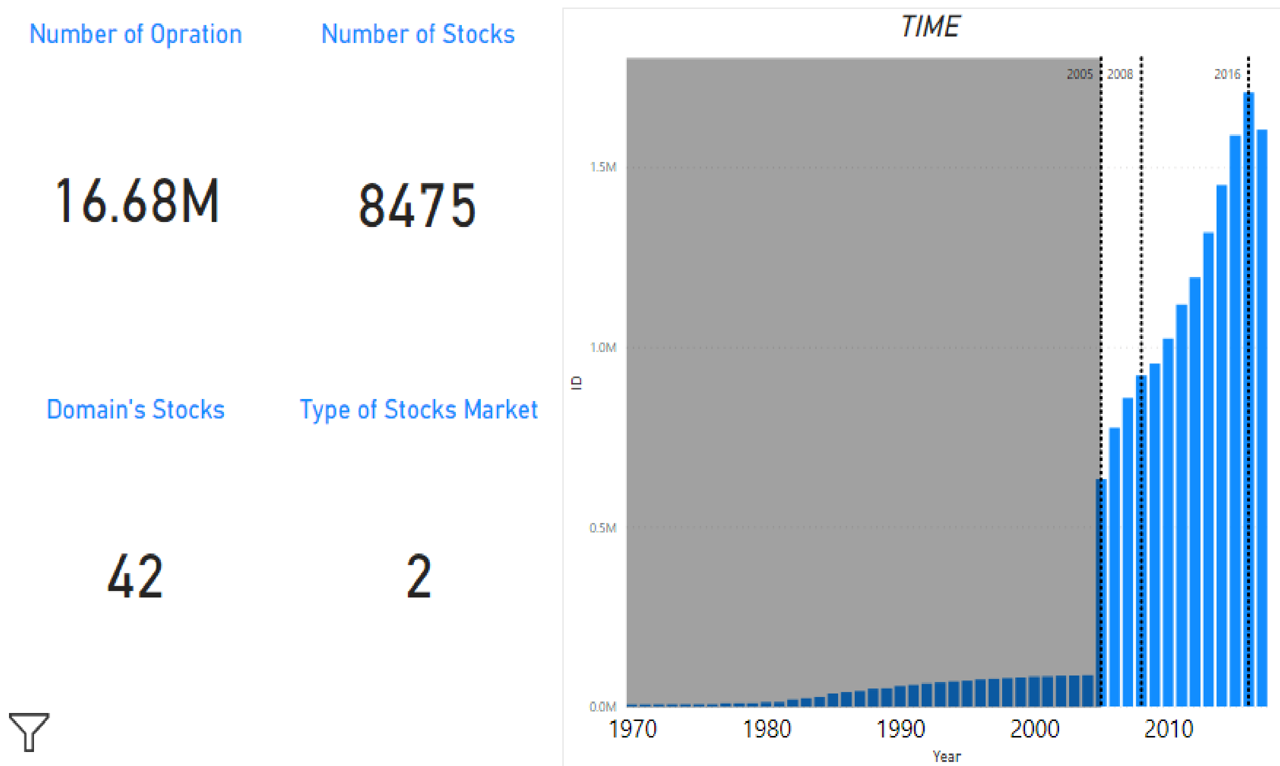


Figure 12: The main report

The main report shows all the main important information, such as: the number of stocks studied, the types of stocks, and the areas in which the stocks were invested.

There is a chart showing the state of stock changes throughout history, as we found bullish indicators in 2005 and 2016, and both cases will be discussed in the upcoming reports.

3.1.2 The types on which the analysis was conducted in the stock market

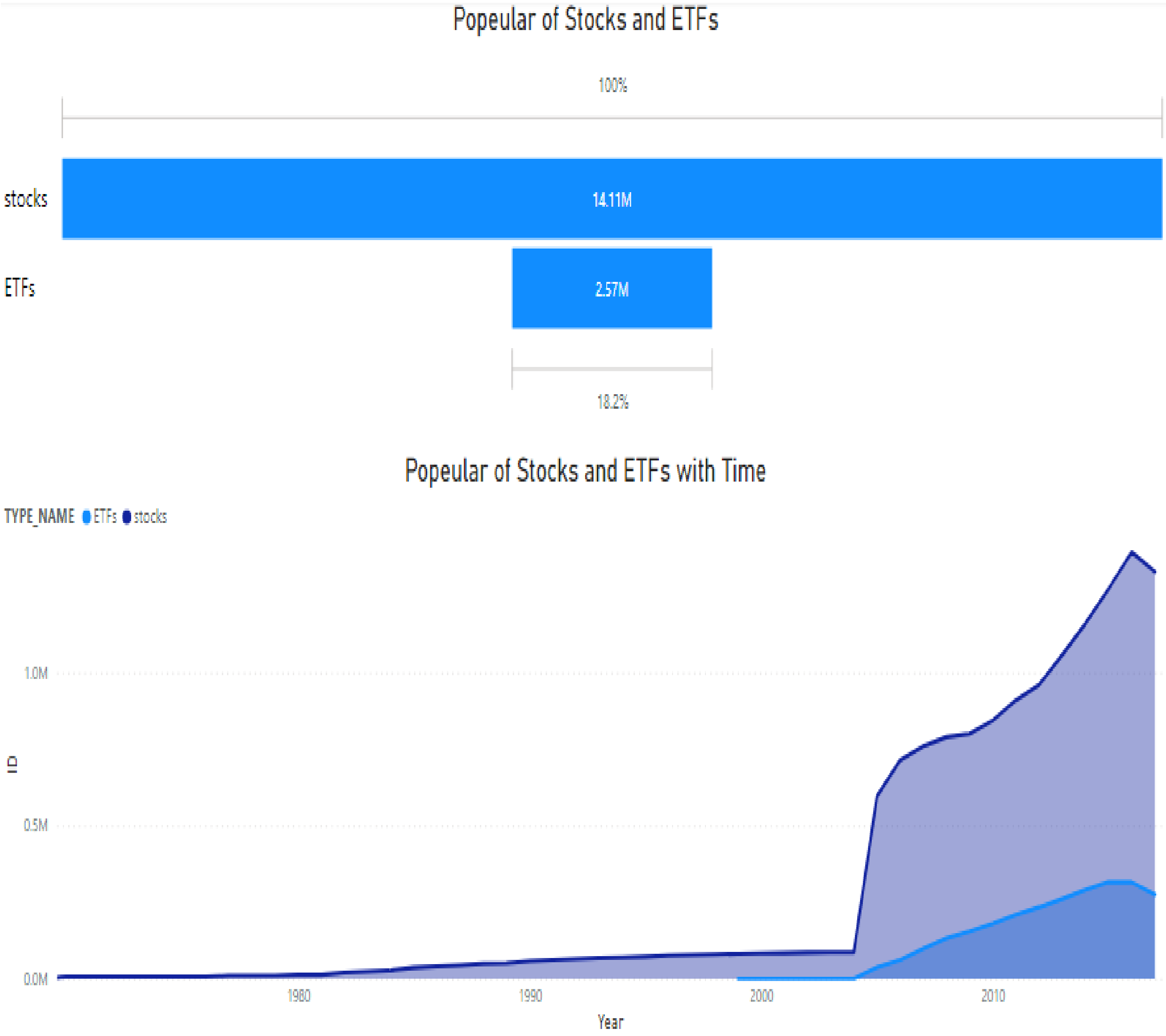


Figure 12: stocks VS ETFs

This report presents the difference between the two types of shares stocks and ETFs and reflects the popularity of each of them among investors in the stock market, where type stocks was the most common

3.1.3 What are the most invested areas in the stock market??

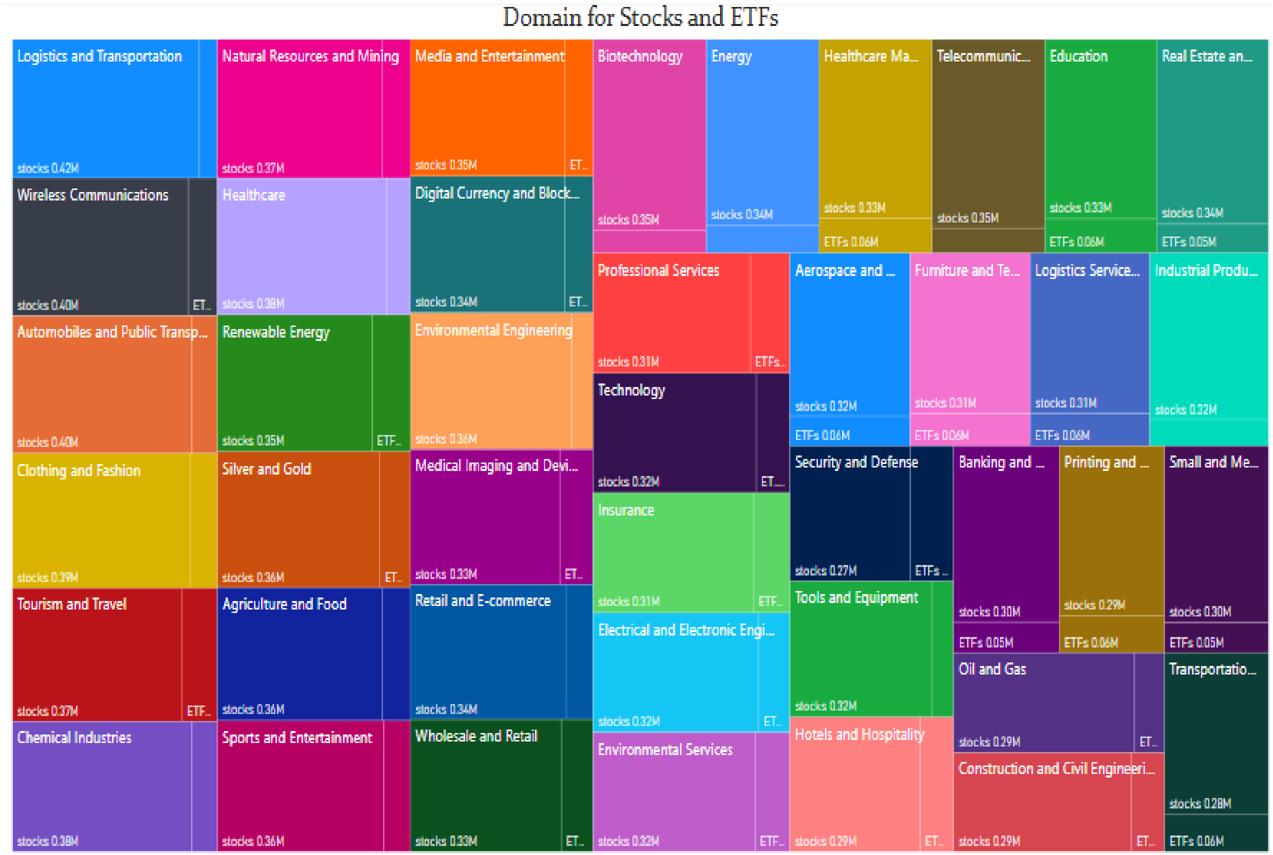


Figure 13: What are the most invested areas in the stock market?

In this report, we can see the most common areas in the US stock market, as each field was divided into types of US stocks stock and ETFs
This report shows the fields that are invested in by Stocks, ETFs,
where the most popular fields were the medical,
educational and agricultural fields.
The report also shows the percentage of investment types of shares (ETFs) in each of the fields
Whereas, shares are an investment in a specific company, and ETFS is an investment in several shares

3.2.4 What happened in 2005 ?

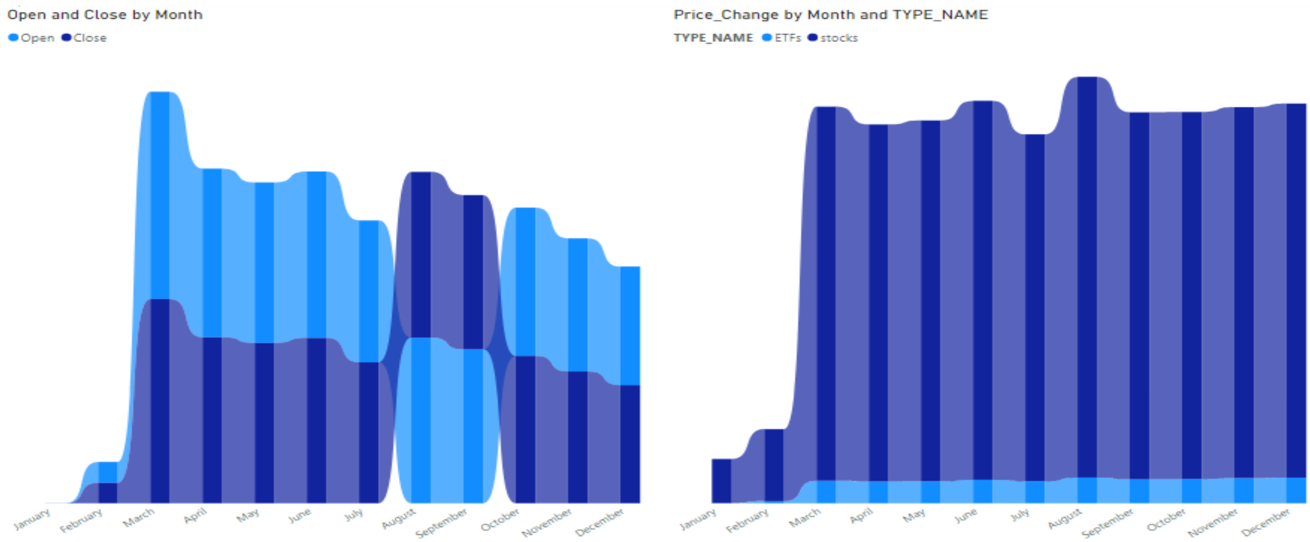


Figure 14: What happened in 2005

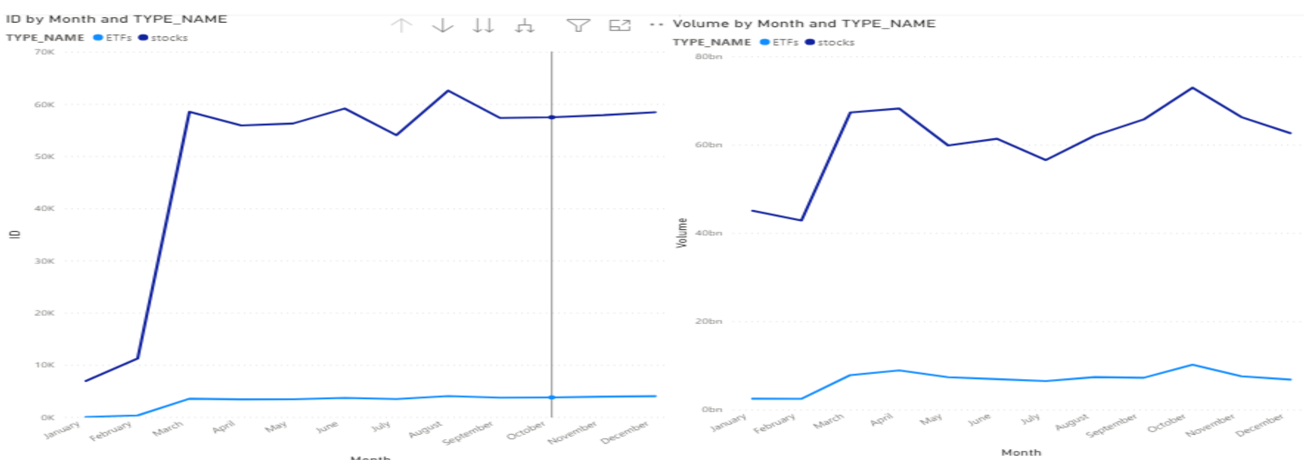


Figure 14: What happened in 2005

In an important year 2005, the US stock market experienced an evolution. There are factors to increase the volume of trading, one of the most important reasons for which was political factors and political decisions that affect the volume of trading in the US stock market

- Interest policy: In 2005, the US Federal Reserve (the US central bank) made a decision to gradually increase the interest rate. This decision affects the cost of borrowing and can affect the behavior of investors and the volume of trading in the market.

- Tax policy: There may be political decisions and adjustments in tax policy that affect companies and investors. Changes in tax rates or tax exemptions can affect investment decisions and thus affect the volume of trading in the market.

3.1.4 What happened in 2008 ?

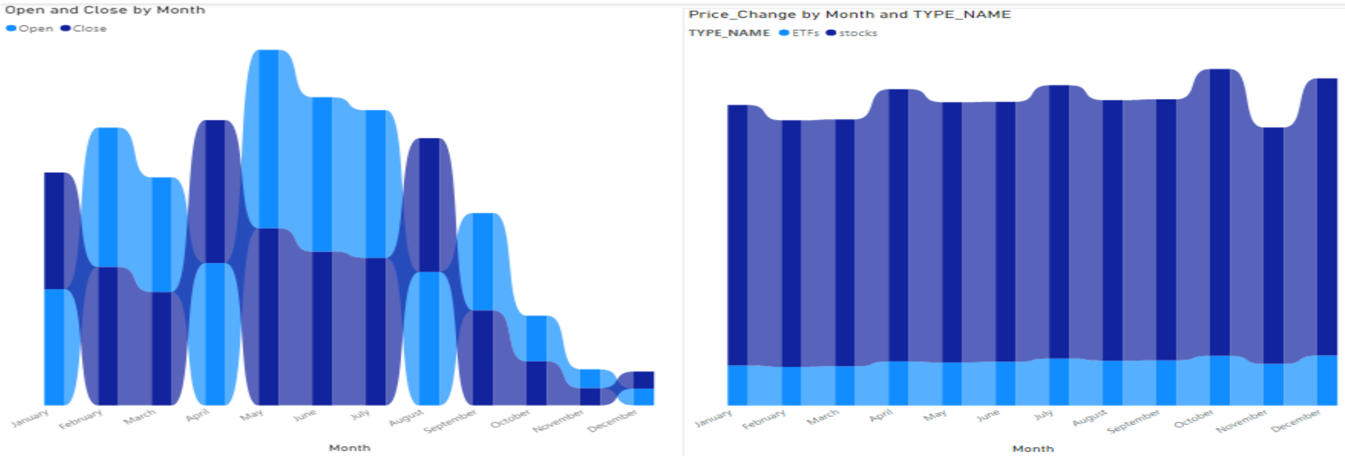


Figure 15: *What happened in 2008*

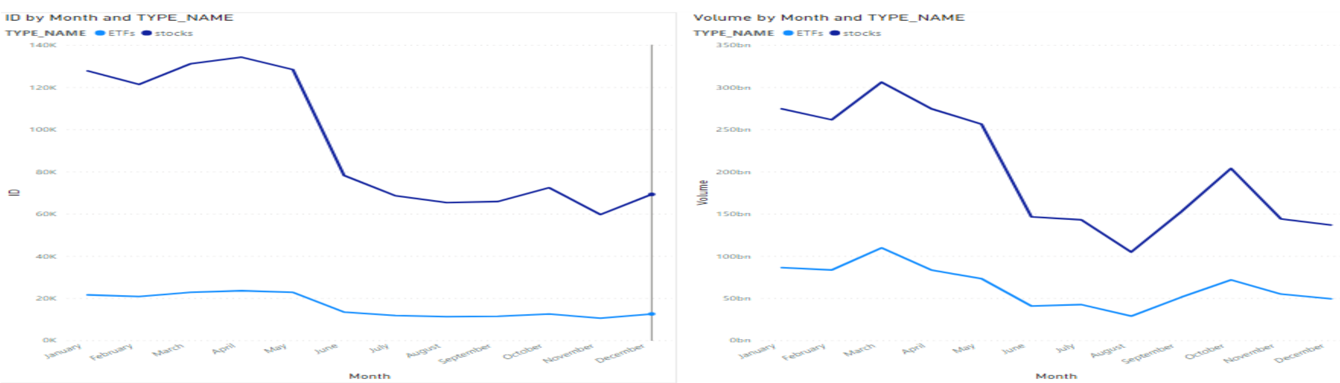


Figure 15: *What happened in 2008*

In 2008, the US stock market experienced one of the worst financial crises in history known as the "Global Subprime Mortgage Crisis" or "Global Financial Crisis". This crisis began in the mortgage sector in the United States and spread to affect the global financial markets.

In 2008, the bad mortgage problem worsened as many subprime mortgages became non-payment, and this led to the collapse of the real estate market in the United States. Banks and financial institutions were carrying a lot of these bad loans in mortgage-related debt packages

3.1.5 *Trading volume varies over the years*

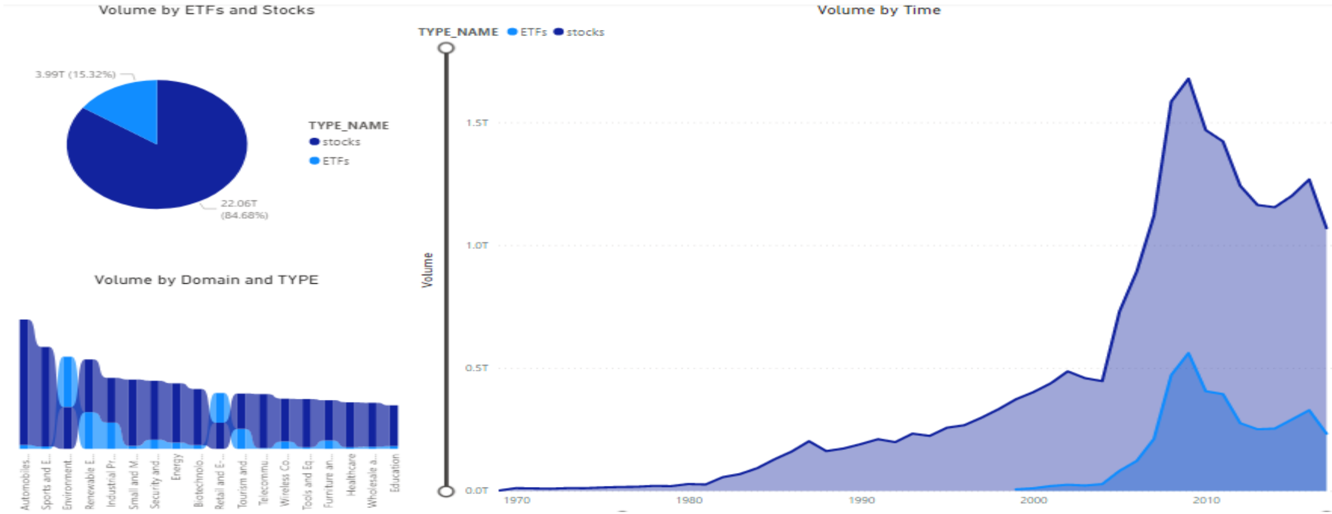


Figure 16: Trading volume varies over the years

The difference in trading volume over the past 50 years, where a suspicious Increase was observed in 2008. This report also displays the difference in the Stocks and ETFs ratio, and the trading volume in relation to the invested areas in the stock market

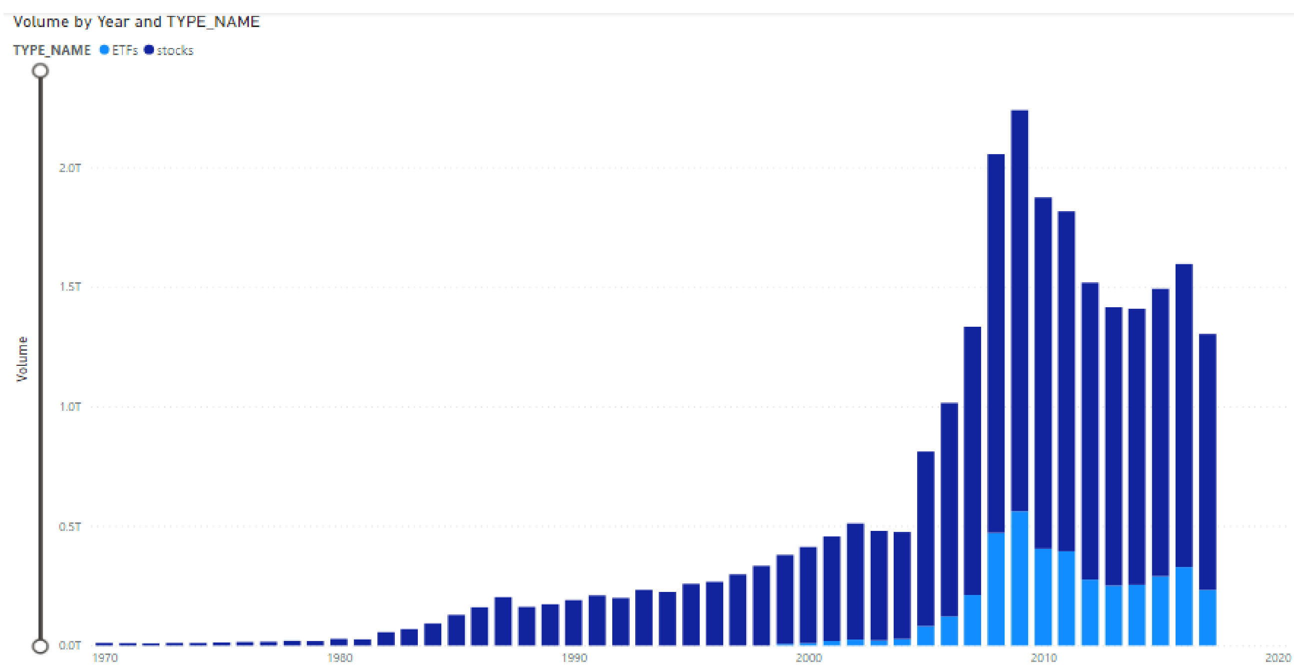


Figure 16: Trading volume varies over the years

3.1.6 What are the safest and most risky areas??

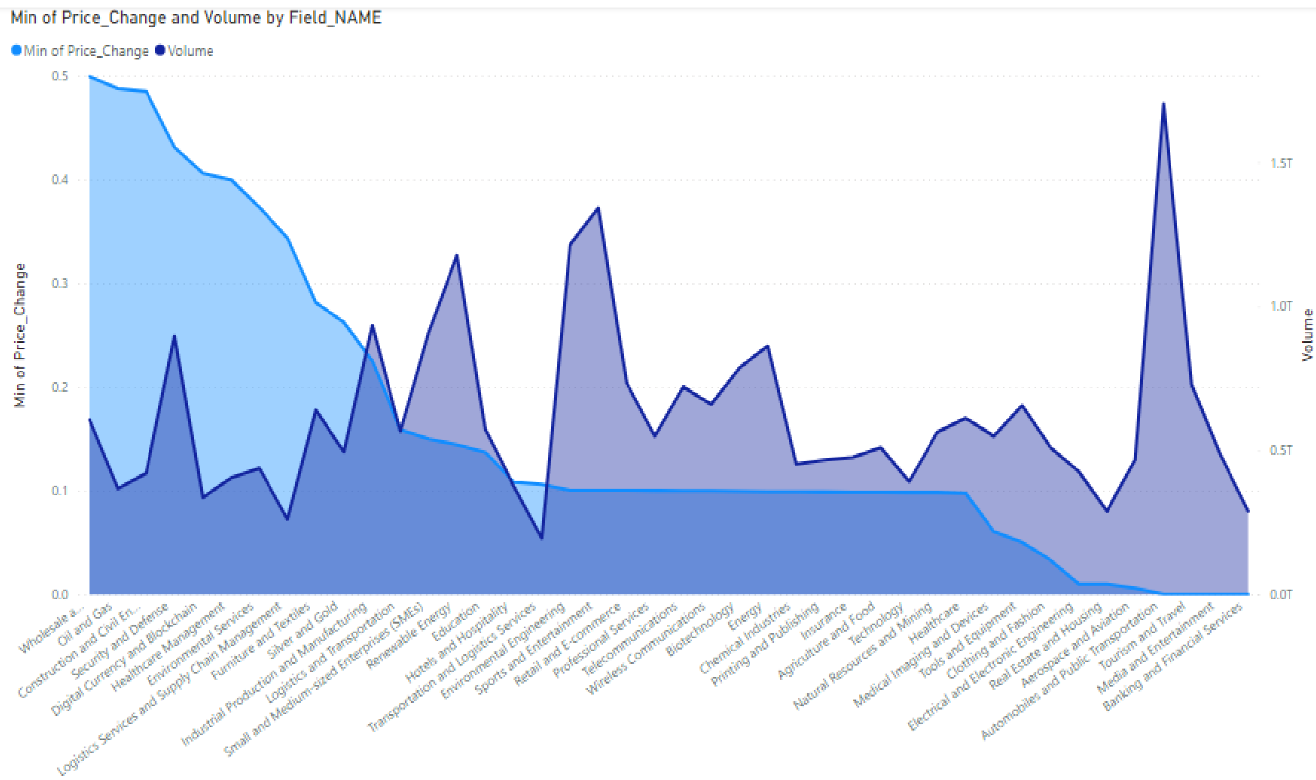


Figure 17: What are the safest and most risky areas?

This chart shows US investment areas and divides them into safe areas and dangerous areas in terms of trading. The chart aims to clarify the diversity and disparity in the various investment fields and the level of risk associated with them..

Different investment areas are represented. These areas can include real estate, stocks, bonds, commodities, currencies, exchange-traded funds (ETFs), and more.

The level of security and risk associated with each investment area is represented. The chart can use various metrics to measure the level of security, such as the expected rate of return or the rate of extreme declines. Through this chart, you can identify the investment areas that are considered safe and stable and that can provide expected returns in a sustainable manner. On the other hand, you can also learn about investment areas that are considered risky and volatile, and that can provide opportunities for high returns but come with greater risks.

In general, this chart can be used to highlight the importance of diversification in an investment portfolio and the importance of assessing the level of risk before making investment decisions. It can also be used to guide investors in choosing investment areas that are compatible with their objectives and risk tolerance level.

3.3.7 The most common shares

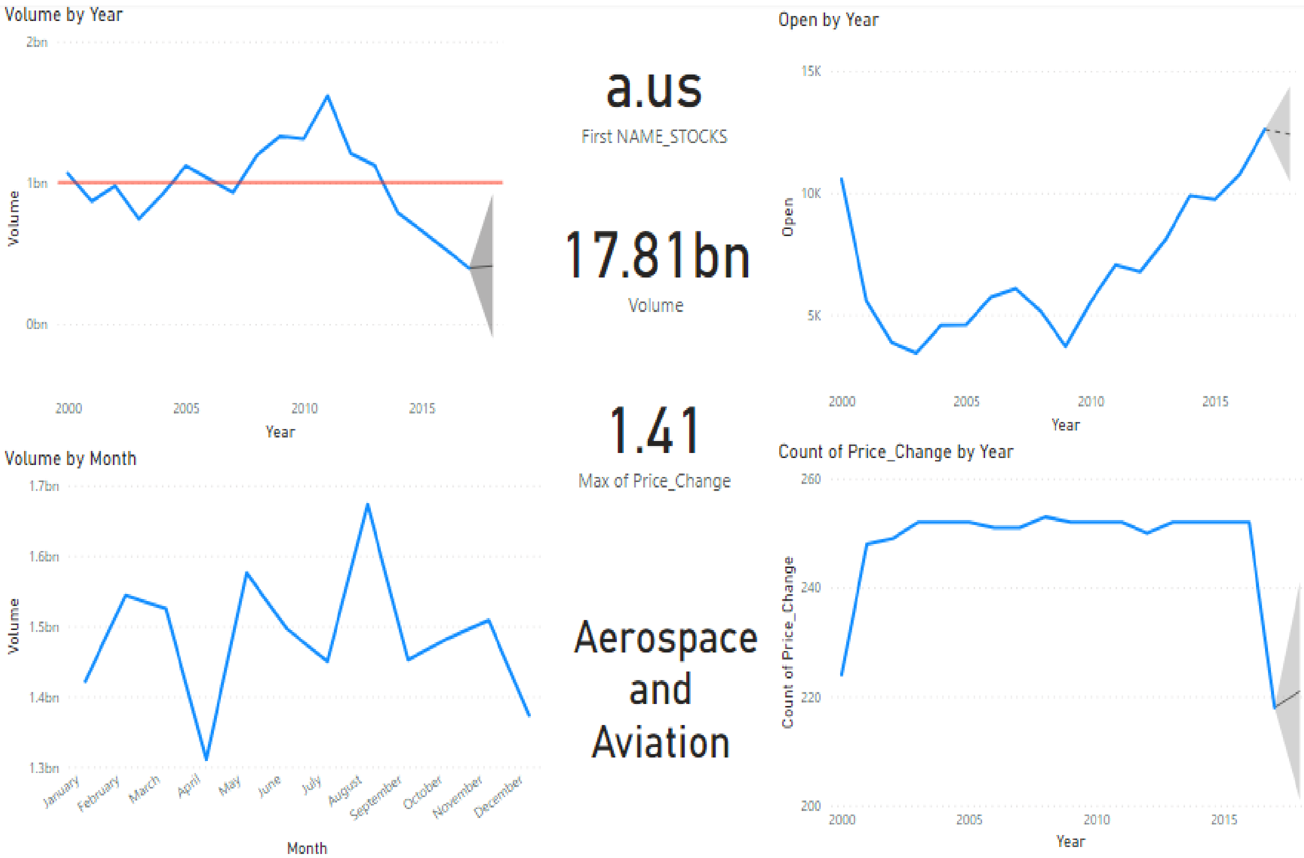


Figure 18: The most common shares stocks a.us

This report presents what the final projected system looks like
Where some of the trading volume is in addition to the share price,
 the amount of risk in the investment, and the field in which it is invested,
with expectations of the price of the price at the level of the day, month, or year

3.3.8 The relationship between opening price and trading volume

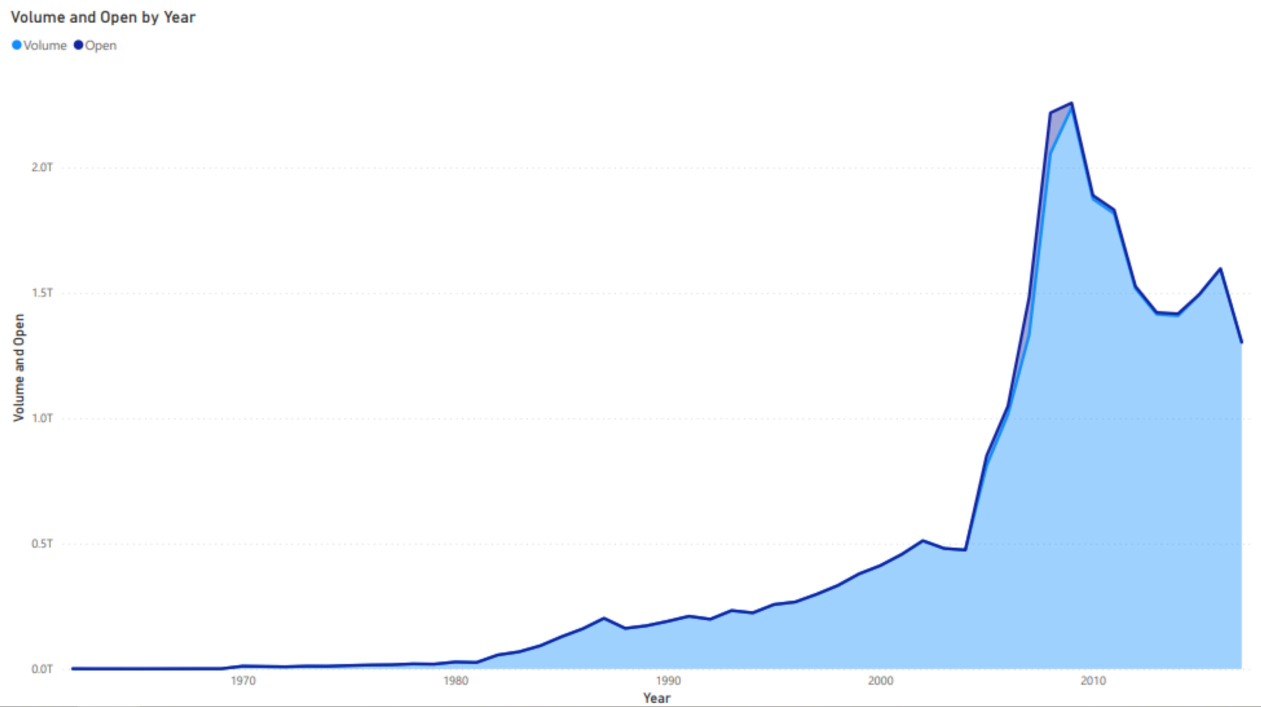


Figure 19: The relationship between opening price and trading volume

There is a relatively identical correlation between the opening price and the trading volume, which means that an increase in the opening price may lead to an increase in the trading volume, and vice versa is also true. This could indicate that there is a lot of interest by investors in these assets when the opening price is high, which leads to increased trading volume.

On the other hand, if there is a decrease in the opening price, there may be a decrease in trading volume, which indicates less interest from investors in these assets.

This tool can be used to analyze financial assets and make investment decisions. When studying specific assets, the chart can be used to assess the extent of interest and reaction in the market towards those assets based on the opening price and trading volume.

3.3.9 The relationship between opening price and trading volume

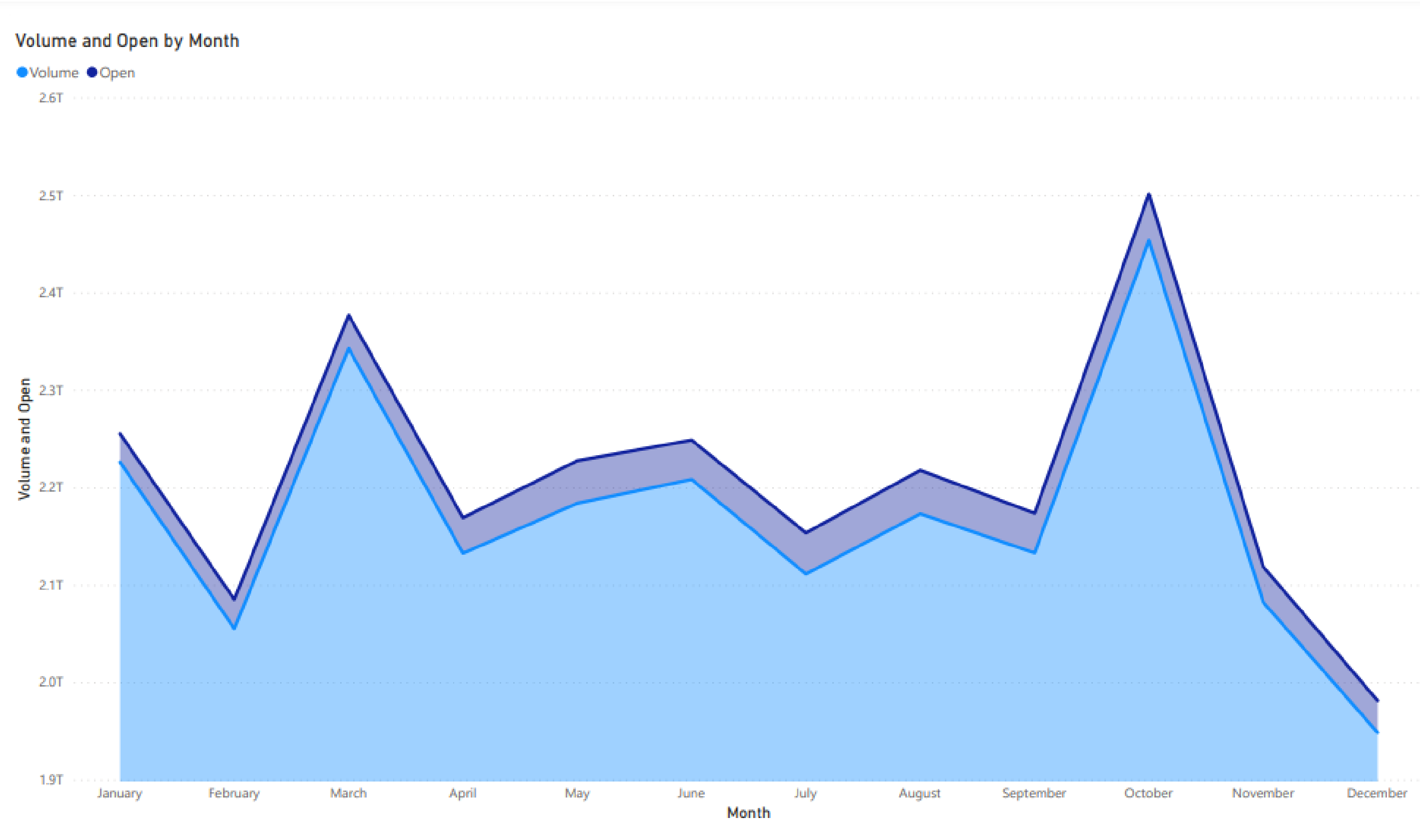


Figure 20: The relationship between opening price and trading volume

There is a relatively identical correlation between the opening price and the trading volume, which means that an increase in the opening price may lead to an increase in the trading volume, and vice versa is also true. This could indicate that there is a lot of interest by investors in these assets when the opening price is high, which leads to increased trading volume.

On the other hand, if there is a decrease in the opening price, there may be a decrease in trading volume, which indicates less interest from investors in these assets.
This tool can be used to analyze financial assets and make investment decisions. When studying specific assets, the chart can be used to assess the extent of interest and reaction in the market towards those assets based on the opening price and trading volume

3.3.10 Expected trading volume for the year 2008

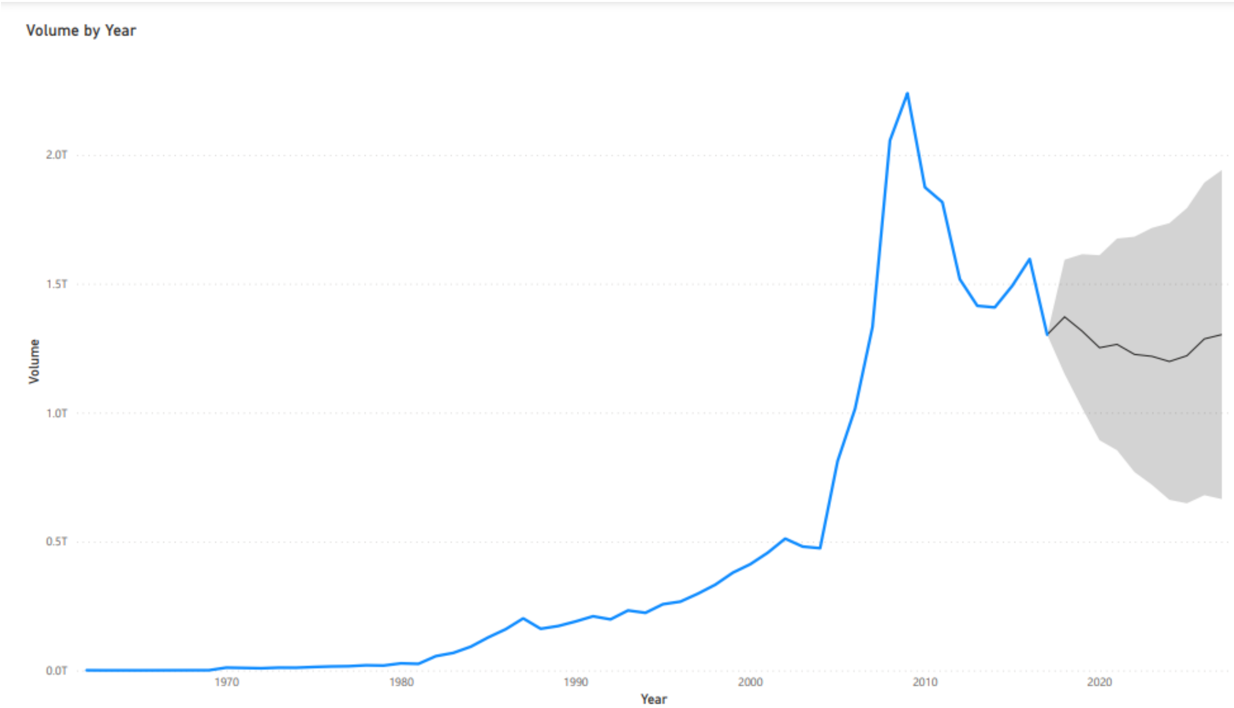


Figure 21: Expected trading volume for the year 2018

The chart displays a forecast of trading volume for the year 2018.
The forecast is based on the use of a prediction algorithm to analyze historical data.
Historical data of trading volume from previous years was used to generate forecasts.
The forecasts generated are represented in the chart by a line or curve that runs along the numerology axis.
The chart helps clarify expected trends and changes in trading volume for 2018.

3.3.10 Expected trading high for the year 2018 for each Stocks

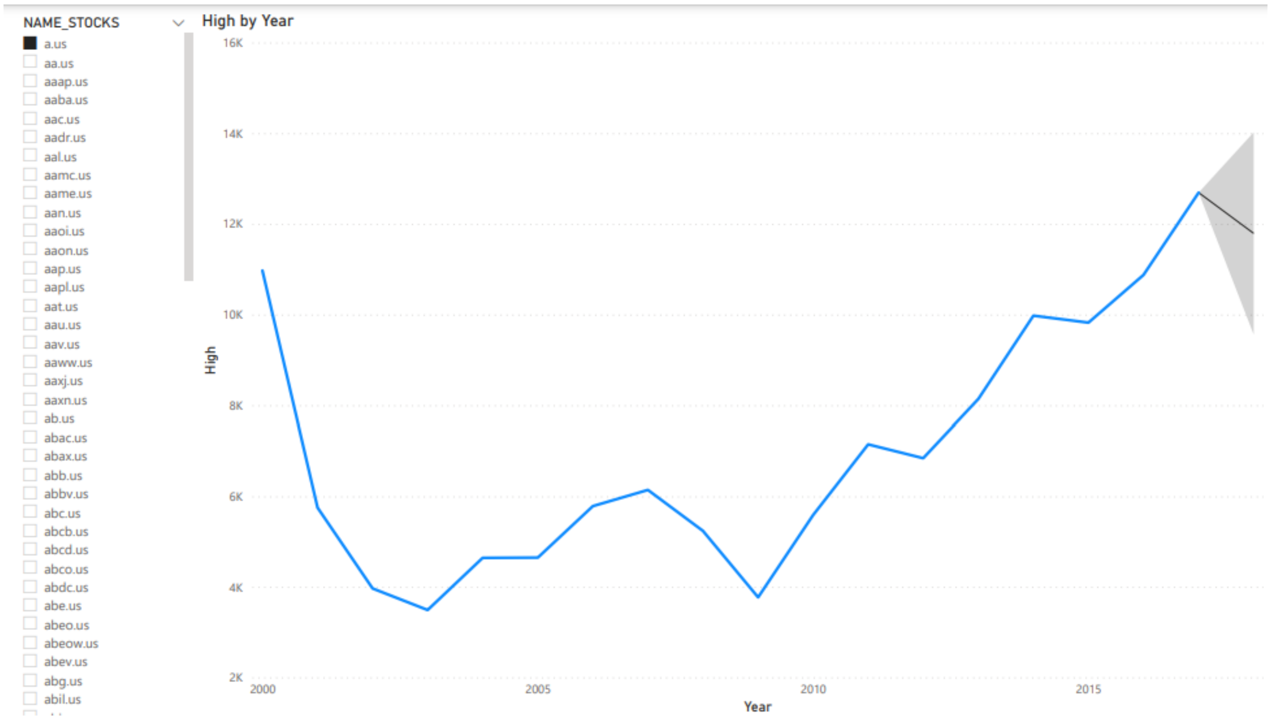


Figure 22: Expected trading high for the year 2018 for each Stocks

Horizontal axis (X axis): This axis will represent the time period for 2018. The time could be broken down into months, weeks, or any other unit of time that fits the data you're using in the forecast.
Vertical axis (Y axis): This axis will represent the forecast for the highest price in the specified time period. The prediction algorithm will analyze historical data and statistical models to predict the highest price for each time period.
Forecast Curve: Forecasts based on the prediction algorithm will be represented as a curve on the chart. This curve can show the general trend of the highest price and the expected changes during the year.
Historical data: It is important to have historical price data for a previous period of time (such as 2017) to train the prediction algorithm and provide forecasts for 2018. Historical data can include past price levels and potential influencing factors.
Prediction information: The chart should show whether the prediction algorithm is based on a mathematical model or a specific technology and the variables used in the prediction. The explanation can include a reference to the accuracy of the forecast and any assumptions or limitations that were made.

3.3.11 Expected trading open for the year 2018 for each Stocks

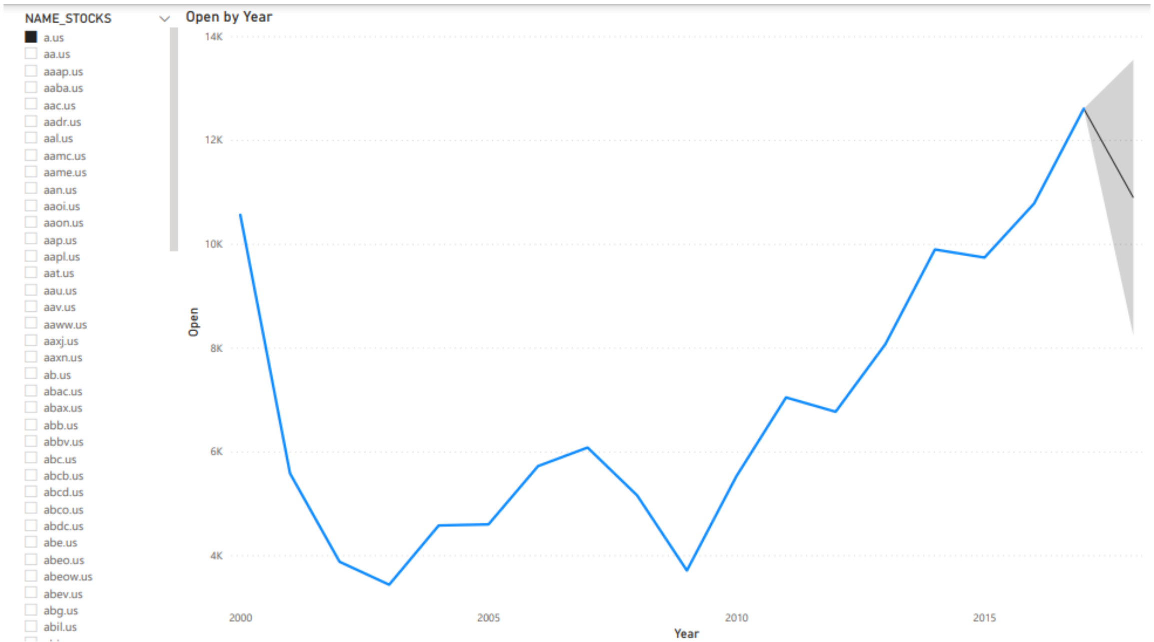


Figure 23: Expected trading open for the year 2018 for each Stocks

This chart is a powerful tool for analyzing financial markets and predicting future events. Forecasting is based on inference from historical data and analysis of past patterns and trends in the opening price of the stock.
On the horizontal axis, the time periods for the year 2018 are represented, and these periods can be daily, weekly, or as per your choice. On the vertical axis, the opening price forecast for each time period is represented.
The prediction algorithm is based on a deep analysis of past data and extraction of significant patterns and factors affecting the opening price. Advanced techniques are used to train predictive models, such as artificial neural networks and statistical analysis.
By analyzing the chart, you can monitor the opening price forecast for different days in 2018 and compare it to the actual market values. This can help you make informed investment decisions and effective strategic planning.
The project report should include a detailed explanation of the algorithm that was used in the prediction and the indicators and variables that were considered in the analysis process. This can help readers understand the methodology used and the reliability of the forecasts provided.

3.3.12 The relationship of technical indicators with the main indicators

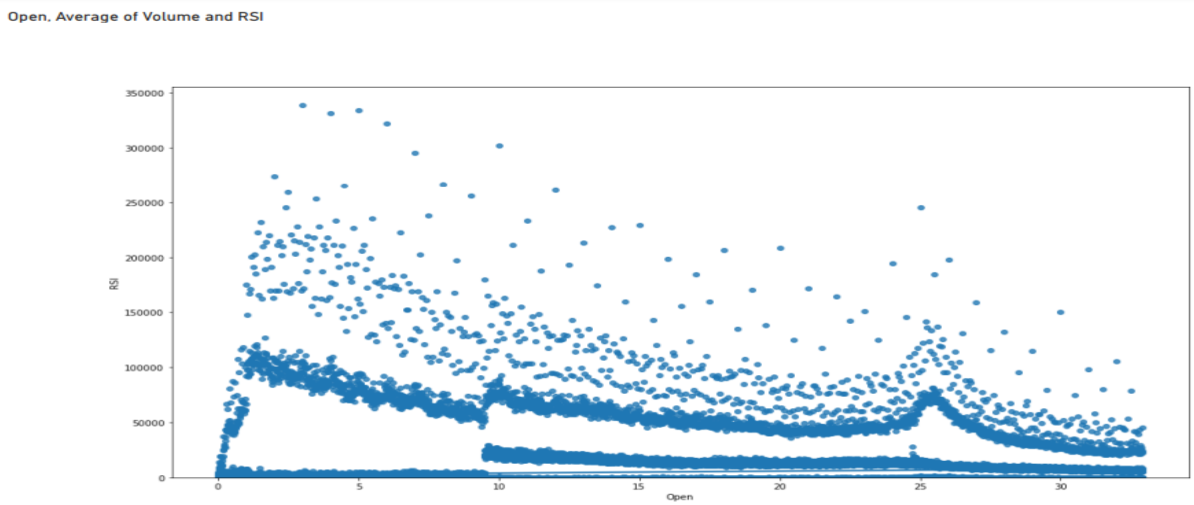


Figure 24: The relationship of technical indicators with the main indicators

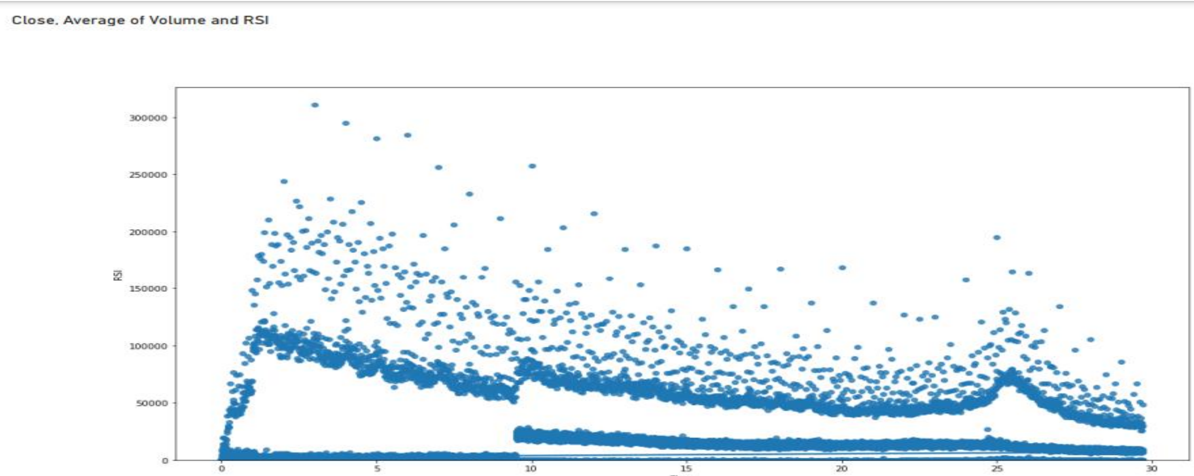


Figure 24: The relationship of technical indicators with the main indicators

This chart highlights the relationship between the opening and closing price of a stock, and uses trading volume and the RSI indicator as indicators of trend strength and reversal confirmation. This chart will help you understand the relationship between these variables and analyze stock performance.

On the horizontal axis, you will find the opening price values and on the vertical axis, you will find the closing price values. Points are drawn on the chart to represent each point in time, giving you an immediate view of the relationship between the two values.

Trading volume is an important factor in analyzing stocks, as high volume can indicate high interest and activity in trading. If there is a direct relationship between the opening price and the trading volume, this may indicate the influence of demand and supply on the opening price.

The RSI (Relative Strength Index) is a technical indicator used to measure the strength of a trend and to determine whether a stock is oversold or oversold. If there is a positive relationship between the opening price and the RSI value, this may indicate an opportunity for a trend reversal trade.

When analyzing the chart, you can notice the points that are spread out uniformly on the chart and assess the nature of the relationship between the opening and closing price. This can give you a general framework for possible trends and forecasts for stock behavior.

The project report should include a detailed explanation of the RSI indicator and how to use it in stock analysis, in addition to explaining the meaning of the direct relationship and the effect of trading volume on the opening price. A practical example should be provided to explain the chart and to clarify the appropriate reading of the points on it.

3.3.13 The relationship of technical indicators with the main indicators

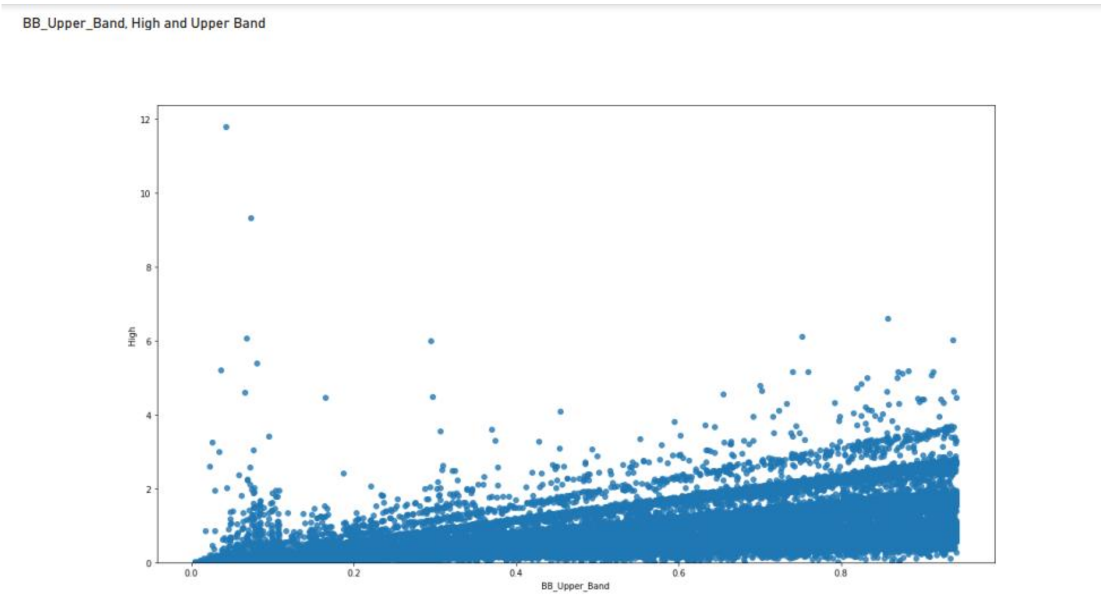


Figure 25: The relationship of technical indicators with the main indicators

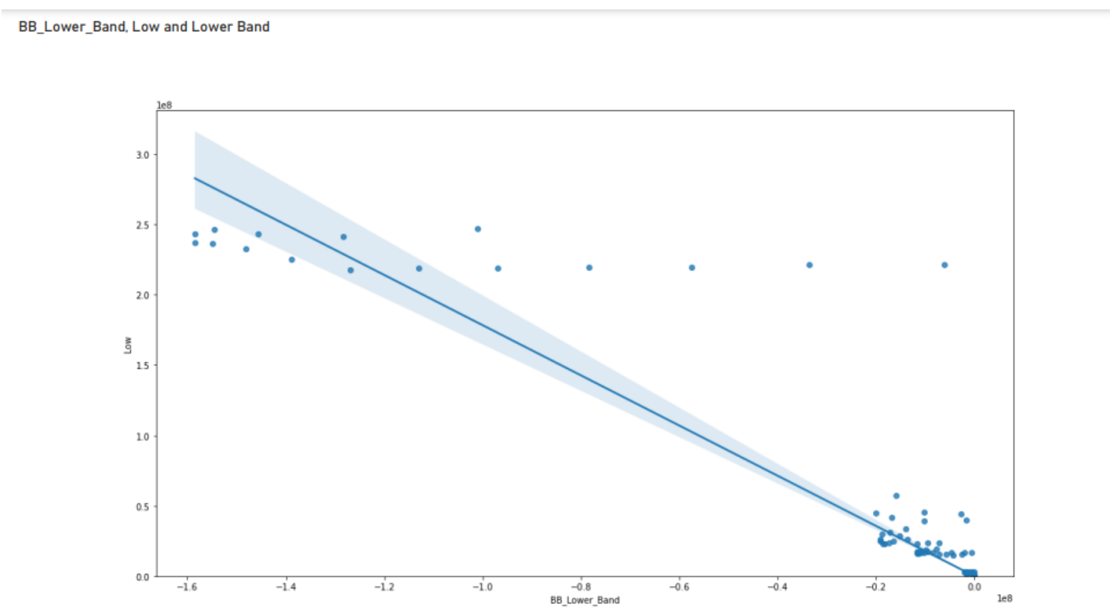


Figure 25: The relationship of technical indicators with the main indicators

This chart provides a comprehensive view of the relationship between the highest price and the lowest price of a particular financial instrument, such as a stock or financial market. The BB_LOWER_BAND indicator is a technical indicator that identifies the lower border of the price channel, while the LOWER BAND indicator is considered to provide the lowest price level in a specific time period.

On the horizontal axis, you will find the time period selected for the analysis, whether it is daily, weekly or monthly. On the vertical axis, the values of the highest and lowest price of the financial instrument are represented.

The direct relationship between the BB_LOWER_BAND indicator and the LOWER BAND indicator shows the frequency of achieving the lowest price in the range of the price channel specified by the BB_LOWER_BAND indicator. Similarly, it shows the achievement of the highest price in the range of the price channel specified by the LOWER BAND indicator.

By analyzing the chart, you can clearly identify the periods in which the direct relationship between the highest price and the lowest price is repeated. This relationship can be used to guide investment decisions, such as determining the right times to buy or sell.

When you include this chart in your project report, you can explain what indicators are used and how BB_LOWER_BAND and LOWER BAND are calculated. It should also explain the time period analyzed and how the direct correlation is used to guide decisions.

Chapter 4
Conclusion And Future Improvements

4.1 Conclusion

In conclusion, the project focused on building a Business Intelligence system for the US stock market, aiming to provide valuable insights and support decision-making processes in the realm of stock investments. Throughout the project, several key findings and outcomes have emerged.

Firstly, the data warehousing process was successfully implemented, integrating various dimensions such as stocks, types, time, and domains. This allowed for the consolidation of relevant data sources and facilitated efficient data analysis.

Secondly, the analysis of the stock market data revealed significant patterns and relationships. Techniques such as association analysis, clustering analysis, classification analysis, and regression analysis were employed to extract meaningful insights. These analyses provided valuable information about market trends, risk assessment, performance evaluation, and factor analysis.

Moreover, the implementation of the Power BI platform proved to be instrumental in visualizing and presenting the analyzed data. The interactive dashboards and visualizations enabled stakeholders to gain a comprehensive understanding of the market dynamics, identify investment opportunities, and make data-driven decisions.
The project also highlighted the importance of continuous monitoring and updating of the Business Intelligence system.

As the stock market is highly dynamic and influenced by numerous factors, it is essential to regularly update the data warehouse, incorporate new data sources, and refine the analytical models. This ensures the system's relevance and accuracy over time.

In summary, the project successfully achieved its objective of developing a robust Business Intelligence system for the US stock market. The system demonstrated its capability to provide actionable insights, support investment decisions, and enhance overall performance in the stock market. The findings and outcomes of the project contribute to the understanding of stock market dynamics and offer valuable guidance for investors and stakeholders.

However, it is important to acknowledge that the system's effectiveness and success depend on the availability and quality of data, the accuracy of analytical models, and the expertise of the users. Therefore, further enhancements and refinements should be considered to continuously improve the system's capabilities and ensure its relevance in the ever-evolving stock market landscape.

Overall, the project's findings and the developed Business Intelligence system hold immense potential for empowering investors, analysts, and decision-makers in navigating the complexities of the US stock market and maximizing their returns on investment.

Future Prospects

1. a) Advanced Data Analysis: The utilization of big data in the project enables us to conduct more comprehensive and sophisticated data analysis. With larger datasets and more diverse variables, we can apply advanced techniques such as machine learning algorithms, natural language processing, and sentiment analysis to gain deeper insights into market trends, investor sentiment, and other relevant factors impacting stock prices.

2. b) Real-time Analytics: Incorporating real-time data processing and analytics into the Business Intelligence system opens up opportunities for timely decision-making. By capturing and analyzing streaming data from various sources, including news feeds, social media platforms, and financial news websites, we can provide up-to-date information and insights to investors, enabling them to react quickly to market changes and optimize their investment strategies.

3. c) Predictive Modeling: By leveraging the power of big data and machine learning algorithms, we can develop predictive models that forecast future stock prices with higher accuracy. These models can help investors in making informed decisions, identifying potential investment opportunities, and managing risks effectively.

4. d) Personalized Recommendations: With the integration of big data analytics, the Business Intelligence system can offer personalized recommendations to individual investors. By analyzing historical trading patterns, risk profiles, and investment preferences, the system can provide tailored investment suggestions, asset allocation strategies, and portfolio

diversification recommendations.

5. e) Data Visualization and Reporting: Big data technologies enable us to visualize complex data sets and present them in intuitive and interactive dashboards. Through visualizations and reports, the Business Intelligence system can provide a comprehensive overview of the stock market,

displaying key performance indicators, market trends, and comparative analyses, aiding investors in making data-driven decisions.

- In conclusion, by embarking on this project using a big data infrastructure, I anticipate a bright future for the Business Intelligence system in the US stock market. The ability to leverage big data analytics, real-time insights, predictive modeling, personalized recommendations, and impactful data visualizations will empower investors to navigate the stock market with greater confidence and efficiency.

.

4.2 References:

Business intelligence19891 .... [1]

EMC Education Services (2014), Data Science and Big Data Analytics..[2]

thecustomizewindows.com .... [3]

www.datawarehouse4u.info. .... [4]

www.geeksfogeeks.com .... [5]

www.mobcoder.com .... [6]

www.researchgate.net .... [7]

What Is a Data Warehouse? https://www.oracle.com/database/what-is-a-data-warehouse/..[8]

what is ETL https://www.ibm.com/topics/etl .... [9]

www.mobcoder.com .... [10]

https://www.astera.com..[11]