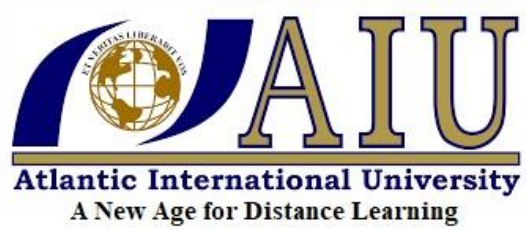


ANTHONY BABAJIDE BALOGUN
ID No: **UB73361SIN82521**

COURSE TOPIC:

DATA ANALYSIS: INTERPRETATION AND PRESENTATION

ATLANTIC INTERNATIONAL UNIVERSITY
JUNE 2022

TABLE OF CONTENTS

LIST OF FIGURES

1.0 INTRODUCTION

Many people use the words Data and Information very frequently because they somehow look similar but in actual sense of it, both have lots of differences between them. The word 'Data' which is the plural for 'datum' can simply be defined as plain facts. So, when data are processed, structured, organized, or presented in a given context in order to make them useful, then they are referred to as Information.

Data are known to be fairly useless until they are interpreted and processed to determine its true meaning, then they become useful and can therefore be known as an Information. Information is data that has been processed in such a manner to be meaningful to the person who receives it. However, data is anything that is communicated or being communicated.

Data may be new to people known as the beginners, but it is actually very interesting and simple to understand. Data is the name given to basic facts and entities; it can be anything and as simple as the name of a person, place or a number.

Data analysis can be defined as a method of cleaning and modeling data in order to find useful information most especially for business decision making.

It can further be defined as a process of extracting useful information from data gathered from different data source and then taking decisions based on the data analysis. Data analysis comprises of processes such as; Data gathering, Collection, Cleaning, Analysis, Interpretation and Visualization. Its types include; Text, Statistics, Diagnostic, Predictive and Predictive analysis.

A Data Warehouse can simply be described as a place where data can be stored for useful mining. It can be likened to a quick computer system with exceptionally huge data storage capacity. Various organizations stored their data in a Warehouse where queries or requests can be made against the warehouse storage of data. Data comes or converge into a data warehouse from different databases.

Data warehouse combines data from various sources to ensure the data quality, accuracy, and consistency. It also boosts system execution by extracting analytics processing from transnational databases. Data are sorted out into a pattern that depicts the format and types in a data warehouse by making use of a query tools which examine the data tables using a specific pattern.

Databases and Data warehouses both are relative data systems but they are made to serve different purposes and functions.

A data warehouse is designed to store a large amount of historical data which empowers fast requests over all the data by using a technique called an Online Analytical Processing (OLAP). A database on the other hand is designed to store current transactions which can allow quick access to specific transactions for ongoing business processes, this process is commonly known as Online Transaction Processing (OLTP).

Data Mining is the process of extracting information in order to identify patterns, trends and useful data that would enable the business to take the data-driven decision from bulk sets of data. In other words, we can simply say that Data Mining is the process of investigating hidden patterns of information which is collected and assembled in particular areas such as data warehouses, efficient analysis, data mining algorithm, helping decision making and other data requirement to eventually cut cost and generate revenue adequately.

Data mining is very powerful but it faces various types of challenges during its execution. Such challenges could be attributed to performance, data, methods, and techniques, etc. We can then say that, the process of data mining becomes successful and effective when all the challenges or problems encountered are correctly recognized and adequately solved.

Data visualization is a very important process in data mining because, it is the primary method that reveals the output to the user in a well presentable way. However, the extracted data should portray the exact meaning of what it intends to describe. Most times, showing the information to the end-user in a precise and easy way is a bit difficult. Furthermore, the input data and output information must be very efficient and successful data visualization processes need to be implemented in order to make it successful.

2.0 DATA, INFORMATION, DATABASES AND DATA ANALYSIS

2.1 Definitions

2.1.1 Data

Many people use the words Data and Information very frequently because they somehow look similar but in actual sense of it, both have lots of differences between them. The word 'Data' which is the plural for 'datum' can simply be defined as plain facts. So, when data are processed, structured, organized, or presented in a given context in order to make them useful, then they are referred to as Information.

Data are known to be fairly useless until they are interpreted and processed to determine its true meaning, then they become useful and can therefore be known as an Information. Information is data that has been processed in such a manner to be meaningful to the person who receives it. However, data anything that is communicated or being communicated.

Data may be new to people known as the beginners, but it is actually very interesting and simple to understand. Data is the name given to basic facts and entities; it can be anything and as simple as the name of a person, place or a number.

Other simple examples of data are; weights, prices, costs, numbers of items sold, employee names, product names, addresses, tax codes, registration marks to mention but a few. Data is also known as the raw material which can be processed by any computing machine represented in the form of numbers and words which can be stored in computer's language as, images, sounds, multimedia and animated data.

2.1.2 Information

Information is fairly described as the data that has been converted into a more useful or intelligible form. This is the collection of data that has been organized for direct usage of mankind as it helps human beings in their decision-making process. Examples of a piece of information are: time table, printed documents, pay slips, receipts, reports etc.

Information is derived at by assembling items of data into a more meaningful form. For example, the marks scored by students and their respective roll numbers forms data, thereby making the report card the information. Other forms of information include; pay-slips, schedules, reports, worksheet, bar charts, invoices and account return etc. However, information containing wisdom is known as knowledge because it may further be processed or manipulated to form knowledge.

Information is needed in order to:

(i) Have knowledge about the surroundings and every other thing that is happening in the society and universe at large

(ii) Keep the systems updated.

(iii) Have ideas on the rules and regulations of society, government, associations, clients etc. as we know that, ignorance is no bliss.

(iv) Arrive at a particular decision concerning the; planning, forming, running and protecting a process or system

2.1.3 Databases

A database is a collection of information or data typically organized and stored electronically in a computer system which is usually controlled by a database management system (DBMS). Therefore, the data, the DBMS and other applications associated with them are referred to as a database system often called a database.

A Database Management System (DBMS) is software designed to store, retrieve, define, and manage data in a database. Its primarily function is to serve as an intermediary between the end user and the database and at the same time managing the data, the database engine and the database schema.

What databases contains depends on the way it is classified. For example: document-text, statistical, or multimedia objects. Another way is by their application area such as accounting, music compositions, movies, banking, manufacturing, or insurance. A third way is by some technical aspect, such as the database structure or interface type.

2.1.4 Data analysis

Data analysis can be defined as a method of cleaning and modeling data in order to find useful information most especially for business decision making. It can further be defined as a process of extracting useful information from data gathered from different data source and then taking decisions based on the data analysis.

Data analysis comprises of processes such as; Data gathering, Collection, Cleaning, Analysis, Interpretation and Visualization. Its types include; Text, Statistics, Diagnostic, Predictive and Predictive analysis.

3.0 DATA ANALYSIS TOOLS

These tools give users the easiest ways to process, manipulate and analyze the correlations which exist between data sets. The tools also help to identify the ways and trends for interpreting data. Here is a complete list of tools used for data analysis in research. Such tools include;

- Text Analysis
- Statistical Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

3.1 Text Analysis

In Information Technology world, text analysis is also known and described as Data Mining which is the one of the methods of data analysis in order to discover a certain pattern that exists in large data sets by using databases or data mining tools. Text analysis is used to change raw data into a meaningful business information.

3.2 Statistical Analysis

Statistical Analysis actually reveals the answer to the question: What happen? by making use of past data. This type of analysis includes; the collection, Analysis, interpretation, presentation and ultimately the modeling of data. The common types of this analysis are; Descriptive Analysis which analyses a sample of summarized numerical data and Inferential Analysis which deals with finding the different summaries from the same data by selecting different samples.

3.3 Diagnostic Analysis

This type of Analysis shows an answer to the question: Why did it happen? by finding the root cause from the insight detected in the Statistical Analysis. Moreover, this Analysis is used in order to find out how data behaves. For example, any new problem found in the business process can be analysed in order to find similar patterns of that problem which may provide the chances of using a similar prescription for such new problem.

3.4 Predictive Analysis

Predictive Analysis reveals an answer to the question: what is likely to happen? by using previous data collection. This type of analysis predicts the future outcomes which is based on either the current or the past data gathered. However, predicting or forecasting is simply defined as an estimate which its accuracy depends on how much information in your possession.

3.5 Prescriptive Analysis

This combines the observations found in all previous analysis in order to determine which action to take to solve any arising problem. Furthermore, it has been observed that; both predictive and descriptive analysis are not enough in order to enhance the performance of data. This type of analysis analyse data based on the current issues and problems in order to make constructive decisions.

4.0 DATA ANALYSIS PHASES

Data analysis consists of the following phases;

- Data Requirement Gathering
- Data Collection
- Data Cleaning
- Data Analysis
- Data Interpretation
- Data Visualization

4.1 Data Requirement Gathering

The need to do data analysis depends on its necessity, that is, you need to ask yourself why do you want to do this data analysis?. Secondly, you have to determine the type of data analysis you wanted to carry out. So, in this phase of data analysis, you need to decide what type of data to be analyzed and also to have a clear understanding of why you are investigating the measures you want to engage in order to do this Analysis.

4.2 Data Collection

You will certainly have a clear idea about what to be measured and what should be the outcome of your findings after a thorough data requirement gathering.

So, data collection can then commence based on the requirement gathering phase which must be processed or organized for analysis. Moreover, as data is being collected from different sources, a log of the collection date and the source must be kept for future references.

4.3 Data Cleaning

It is pertinent to mention here that, most of the time, the data collected may not be as useful or irrelevant to your primary reason for an analysis, hence, the data needs to be cleaned. The reason for this is that, the type of data collected may contain duplicate records, spaces errors etc.

Therefore, it is very essential for the data to be cleaned and free of errors. However, this phase must be implemented before any analysis because, you will be closer to getting your expected outcome based on the data cleaning process.

4.4 Data Analysis

After the data has been collected, cleaned up and processed, then it is ready for analysis. As you interact with the collected data, you may find out that, you have collected the exact information you need, or it may give you an insight into the possibility of collecting more data. Furthermore, you may use data analysis tools and software which can assist you to understand, interpret, and to get a conclusive outcome based on the requirements.

4.5 Data Interpretation

In this phase, the results of the analysis are presented after a careful and thorough analyzing the data has been done, that is, it is the time to interpret your results.

You can do this by choosing the way and manner in which to express or communicate the data analysis either by expressing your results in plain words or present it in a kind of table or chart which will allow you to use the results of your data analysis process to determine your best course of action to be taken.

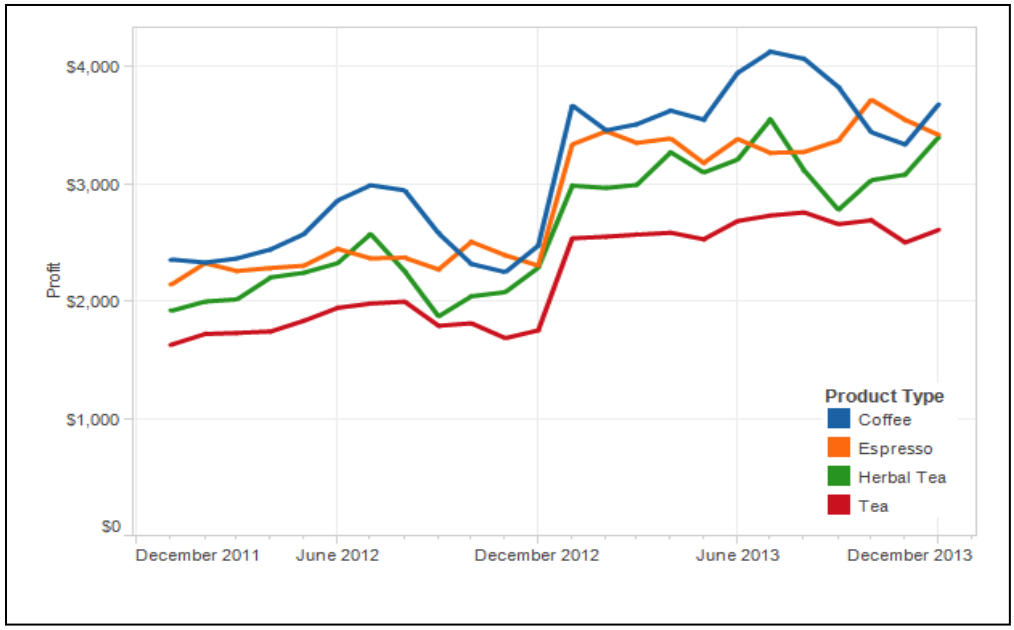
4.6 Data Visualization

This is a phase that is very common in our day-to-day activities. Visualization often appears in the form of charts and graphs, in other words, data is shown or represented in a graphical form in order for the human brain to understand and processed.

Furthermore, data visualization is often used to discover some unknown facts and trends; and by observing the relationships and datasets comparison, you can get a way to pull out a meaningful information.

(Petersen, 2014)

Figure 1: showing different types of data visualization



(Petersen, 2014)

Figure 2: showing another form of data visualization

5.0 DATA ANALYSIS SOFTWARE

The following software has been observed and noted to support data analysis:

5.1 DevInfo

This is a database system which has been endorsed by the United Nations Development Group. It is used to monitor and analyze the human development

5.2 ELKI

This is a data mining framework embedded in Java with data mining visualization functions

5.3 KNIME

This stands for; the Konstanz Information Miner which is a user-friendly data analytics framework

5.4 Orange

Orange has been noted to be a visual programming tool with interactive data visualization methods for statistical data analysis and data mining

5.5 Pandas

This a library embedded in Python programming language commonly used for data analysis

5.6 PAW

This framework was developed in FORTRAN/C programming language at CERN for data analysis

5.7 R

Another notable programming language developed for statistical computing with graphics combability

5.8 ROOT (C++)

C++ is another notable and widely used data analysis framework developed at CERN

5.9 SciPy

SciPy is another library function in Python programming language for data analysis

5.10 Julia

This is a Julia programming language used for numerical analysis alongside computational science

6.0 TECHNIQUES FOR ANALYZING QUANTITATIVE DATA

There are several techniques use for analyzing quantitative data, they include;

(i) Check raw data for anomalies before performing an analysis

(ii) Re-perform important calculations

(iii) Confirm that the main totals are actually the sum of subtotals

(iv) Check the relationships between numbers

(v) Normalize numbers to make comparisons easier

(vi) Break problems into component parts

However, for the variables under examination, data analysts must get the descriptive statistics such as; the mean (average), median, and standard deviation for them, such as the mean (average), median, and standard deviation.

Analysts can also use a well-designed and a robust statistical measurement such as a hypothesis to solve certain analytical problems Furthermore, regression analysis can be used in a scenario where a data analyst is trying to determine the extent to which independent variable X affects dependent variable Y. Necessary condition analysis (NCA) can be used in a situation where the data analyst is trying to determine the extent to which independent variable X allows variable Y.

7.0 BARRIERS TO EFFECTIVE DATA ANALYSIS

Barriers to effective data analysis have been noted to exist between the data analysts that is performing the data analysis. Differentiating the fact from opinion and innumeracy have all become a challenge to data analysis. The different barriers to an effective data analysis are highlighted below:

7.1 Confusing fact and opinion

The idea of, you are entitled to your own opinion but you are not entitled to your own facts is one of the greatest barriers in getting an effective data analysis.

7.2 Cognitive biases

Various type of cognitive biases that can adversely affect analysis are; confirmation bias which is the tendency to search for an information in a way that confirms one's preconceptions. Moreover, individuals may not give credits to the information that does not support their own views.

So, in order to overcome these biases, data analysts must be trained specifically to be aware of and how to overcome them.

7.3 Innumeracy

In order to arrive or get an effective data analysis, data analysts must be adept with a variety of numerical techniques. However, audiences may not have such privilege to literacy with numbers because they are regarded to be innumerate. Furthermore, those communicating the data may also attempt to mislead intentionally by using bad numerical techniques.

It is possible for analysts to analyze data under different assumptions when performing financial statement analysis. For example, analysts can recast the financial statements under different scenarios in order to arrive at an estimate of future cash flow.

8.0 DATA WAREHOUSE I

A Data Warehouse can simply be described as a place where data can be stored for useful mining. It can be likened to a quick computer system with exceptionally huge data storage capacity. Various organizations stored their data in a Warehouse where queries or requests can be made against the warehouse storage of data. Data comes or converge into a data warehouse from different databases.

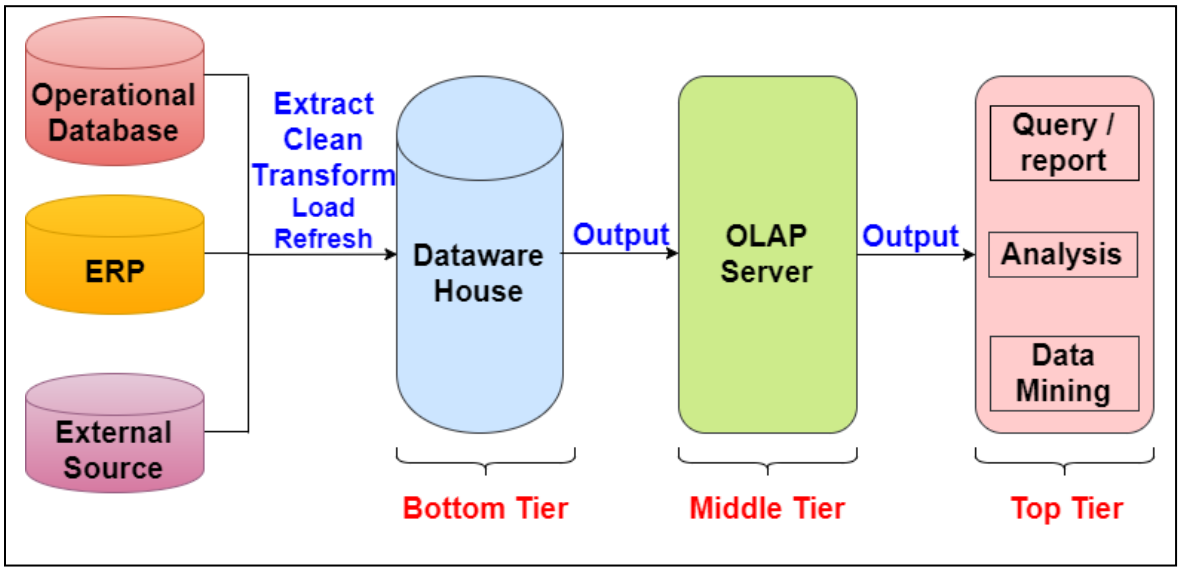
Data warehouse combines data from various sources to ensure the data quality, accuracy, and consistency. It also boosts system execution by extracting analytics processing from transnational databases. Data are sorted out into a pattern that depicts the format and types in a data warehouse by making use of a query tools which examine the data tables using a specific pattern.

Databases and Data warehouses both are relative data systems but they are made to serve different purposes and functions. A data warehouse is designed to store a large amount of historical data which empowers fast requests over all the data by using a technique called an Online Analytical Processing (OLAP).

A database on the other hand is designed to store current transactions which can allow quick access to specific transactions for ongoing business processes, this process is commonly known as Online Transaction Processing (OLTP).

8.1 Three-Tier Data Warehouse Architecture

Figure 3



(Lastnightstudy, n.d.)

Figure 3: Three-Tier Data Warehouse Architecture

According to figure 1 above, we could see clearly that, data warehouses often adopt a three–tier architecture such as; Bottom tier, Middle tier and Top tier.

(i) Bottom tier

The bottom tier as shown in the architecture is known as the data warehouse database server. This is a relational database system in which data is feed into bottom tier by some back-end tools and utilities. These back-end tools and utilities performs such functions as: Data Extraction, Data Cleaning, Data Transformation, Data Load and Data Refresh

(ii) Middle tier

This type of Online Analytical Processing (OLAP) server present user with a multidimensional data from data warehouse. This type of server can be implemented using either the relational OLAP (ROLAP) model, an extended relational database management system, which maps the operations on multidimensional data to standard relational operations or use the multidimensional OLAP (MOLAP) model which directly implements multidimensional data and operations.

(iii) Top tier

The top tier architecture is a front-end client layer which holds following tools: Query and Reporting tools (used for production reporting), Analysis tools (used to prepare charts based on analysis) and Data mining tools (use to discover the hidden knowledge and pattern).

8.2 Types of Data Warehouse architecture

There are three different types of Data Warehouse architecture such as; Enterprise, Operational and Data Mart.

(i) Enterprise Data Warehouse

The function of an enterprise is to provide a central repository database for decision support throughout the enterprise. It has been identified as a central place where all business information from various sources and applications are made available. This data once stored, can be used for analysis and also be useful to the people across the organization. The ultimate enterprise goal is to produce a complete overview of any particular object in the data model.

(ii) Operational Data Warehouse

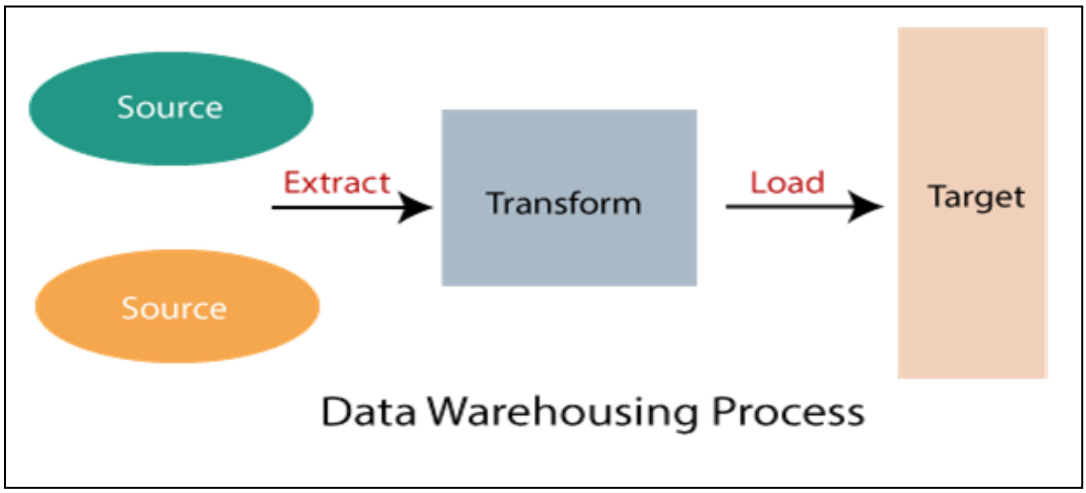
In this warehouse, data is refreshed in real-time and used for routine commercial activity. It also assists in fetching data directly from the database, which also aids data transaction processing. However, the data saved in the Operational data store can be scrubbed and whatever the duplication present can be reviewed and later fixed by examining the corresponding market rules.

(iii) Data Mart

Data Mart may be classified as a subset of knowledge warehouse which supports a specific region, business unit or even a business function. Its main purpose is to focus on storing data for a particular functional area which contains a subset of data saved in a memory. It also helps in advancing user responses by reducing the volume of data for data analysis thereby making it more comfortable to go forward with the report.

9.0 DATA WAREHOUSE II

Figure 4



(Javatpoint, n.d.)

Figure 4: Data Warehousing Process

A Data Warehouse is a technology that collects the data from different sources within an organization in order to produce a meaningful business insight. The bulk amount of the data comes from various places such as; Marketing and Finance. Such extracted data is used for analytical purposes and in addition, it also helps in the decision-making for a business organization.

However, the data warehouse is designed mainly for the analysis of data rather than transaction processing.

9.1 Data Repositories

Data Repository simply refers to a location for data storage. However, many IT professionals uses the term more clearly to describe a specific setup within an IT structure such as; a group of databases where an organization has stored various types of information.

9.2 Object-Relational Database

Object-Relational Model is described as a combination of an object-oriented database model and relational database model which supports Classes, Objects, Inheritance, etc. One of its notable primary objectives is to close the gap between the Relational database and the object-oriented model practices frequently utilized in various programming languages such as; C++, Java, C# and many more.

9.3 Transactional Database

A transactional database which is also known and refers to as a database management system (DBMS), has the ability to undo a database transaction if it is not adequately and appropriately performed. Although, this was a noticeable and unique capability long ago, but today, most of the relational database systems now support transactional database activities.

9.4 Important Features of Data Warehouse

The notable important features of Data Warehouse are described below:

(i) Subject Oriented

A data warehouse is known to be subject-oriented because it provides useful data about a subject instead of the organization's ongoing operations. Such subjects can be customers, suppliers, marketing, product, promotion, etc. A data warehouse as we know, usually pay more attention on modeling and analysis of data that can help the organization business to make data-driven decisions.

(ii) Time-Variant

The various data which is stored and present in the data warehouse only provides information for a specific period of time.

(iii) Integrated

A data warehouse is described by combining data from heterogeneous sources such as; social databases, level documents etc.

(iv) Non-Volatile

This simply means that, data which is committed into the warehouse cannot be changed.

9.5 Advantages of Data Warehouse

(i) More accurate data access

(ii) Improved productivity and performance

(iii) Cost-efficient

(v) Consistent and quality data

10.0 DATA MINING

This is the process of extracting information in order to identify patterns, trends and useful data that would enable the business to take the data-driven decision from bulk sets of data. In other words, we can simply say that Data Mining is the process of investigating hidden patterns of information which is collected and assembled in particular areas such as data warehouses, efficient analysis, data mining algorithm, helping decision making and other data requirement to eventually cut cost and generate revenue adequately.

10.1 Data Mining Process

Figure 5

(Javatpoint, n.d.)

Figure 5: Data Mining Process

In the process of data mining, the computer analyzes the data and extract all the useful and meaningful information from it. It then looks for hidden patterns within the data set in order to use it to predict future behavior. The primary use of Data mining is to discover and to indicate relationships amongst the data sets.

It also aims to enable business organization's view business behaviors and trends relationships that will allow the business to make data-driven decisions. Furthermore, Data mining is also known as knowledge Discover in Database (KDD). Moreover, Data mining tools makes use of Artificial Intelligence, statistics, databases and machine learning systems to uncover the relationship between the data.

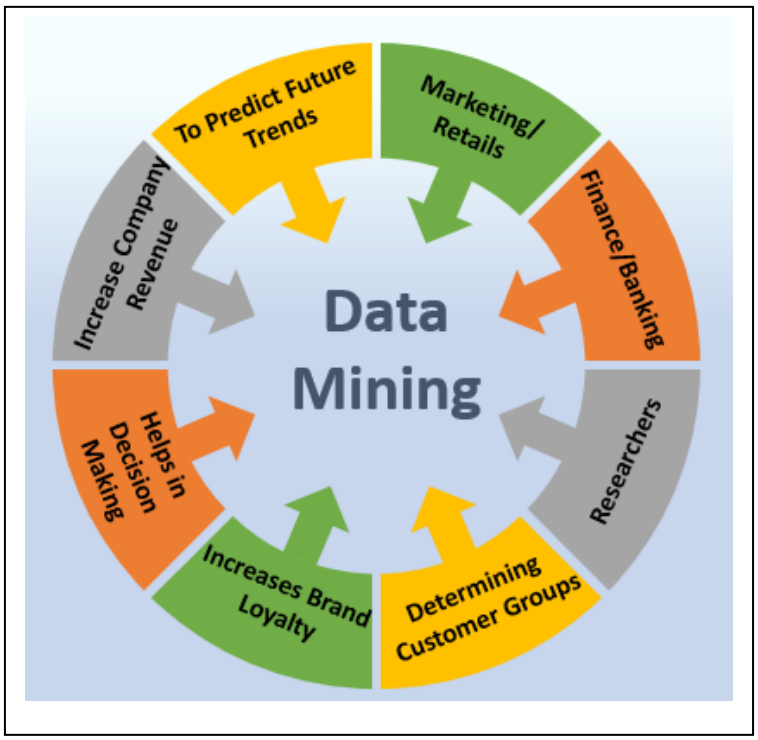
10.2 Important features of Data Mining

The most important features of Data Mining are listed below:

- It utilizes the Automated discovery of patterns
- It predicts the expected results
- It focuses on large data sets and databases
- It creates actionable information

10.3 Advantages of Data Mining

Figure 6



(Tawde, n.d.)

Figure 6: Advantages of Data Mining

(i) Market Analysis

Data Mining can be use used to predict the market that helps the business to make the decision. For example, it can predict an entity who is keen to buy a type of a particular products.

(ii) Fraud detection

Fraud is on the trend lately and with the help of data mining, frauds in cellular phone calls, insurance claims, credit or debit card purchases
can be uncovered.

(iii) Financial Market Analysis

These techniques are notable and widely used to assist Model Financial Market

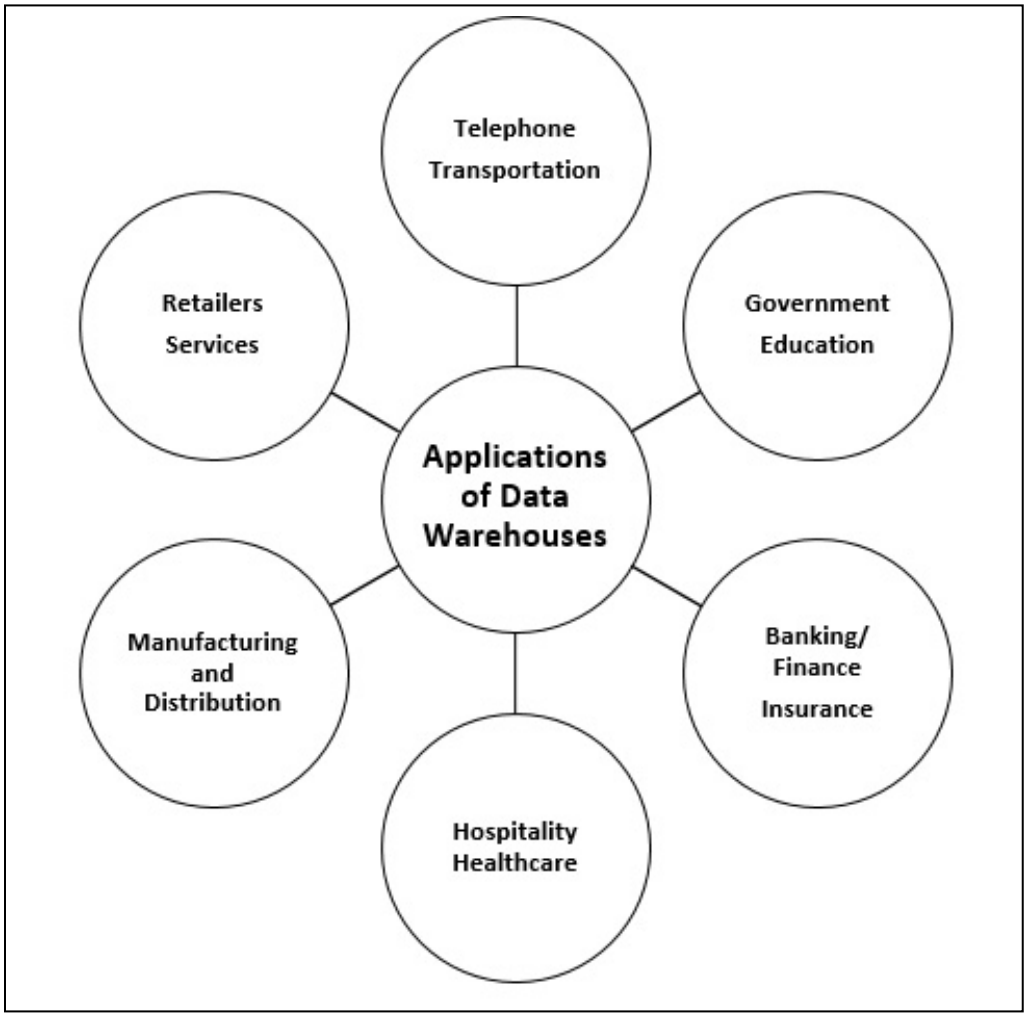
(iv) Trend Analysis

Another great advantage of data mining is in the area of trend analysis in the market place. Analyzing the current and existing trend in the marketplace is a strategic benefit because it assists in cost reduction and manufacturing process regarding market demand.

11.0 DATA MINING APPLICATIONS

Data Mining is primarily used by organizations with intense consumer demands and enables a retailer to use the point-of-sale records of customer purchases to develop products and promotions that will help the organizations to attract their customers.

Figure 7



(Chegg, n.d.)

Figure 7: Applications of Data Warehouses

These are a few areas where data mining is widely used such as; healthcare, market analysis, education, engineering, manufacturing, customer relationship management (CRM), forensic investigation and financial banking

(i) Healthcare

Data mining uses data and analytics for better insights and to identify best practices that will enhance health care services and reduce costs. Analysts utilizes data mining processes such as Machine learning, Multi-dimensional database, Data visualization, Soft computing, and statistics. Data Mining can also be used to forecast patients in each category thereby ensuring that, patients get intensive care needed at the right place and at the right time. Furthermore, data mining allows healthcare insurers to recognize fraud and abuse.

(ii) Market Analysis

Market analysis is a modeling approach that is based on a hypothesis. For example, if you purchase a specific group of products, then it is more likely for you to buy another set of products. This method allows the retailer to know or understand the purchase behavior of a buyer.

(iii) Education

This has been noted to be a newly emerging field concerned with developing techniques that explore knowledge from the data generated from educational Environments. Its objectives are recognized as; confirming student's future learning behavior, studying the impact of educational support and promoting learning science. Therefore, an organization can use data mining to make a precise decision of the student and with such results, the school can concentrate on what to teach and how to go about the teaching.

(iv) Manufacturing & Engineering

Data mining tools can be useful to finding patterns in a complex manufacturing process as we know that, knowledge is the best asset possessed by a manufacturing company.

Data mining can also be useful in system-level designing in order to obtain the relationships between product architecture, product portfolio and data needs of the customers. On the other hand, it can also be used to forecast the product development period, cost and expectations among other various tasks.

(v) Customer Relationship Management

The purpose of Customer Relationship Management (CRM) is all about obtaining and holding Customers, also enhancing customer loyalty and implementing customer-oriented strategies. To get a decent relationship with the customer, a business organization needs to collect data and analyze the data. With data mining technologies, the collected data can be used for analytics.

(vi) Forensic & Fraud Detection

A lot of money has been lost to fraudulent activities and the traditional methods of fraud detection are a bit time consuming and sophisticated. Data mining has helped a lot in providing meaningful patterns and then turning the data into useful information in order to detect fraud. Moreover, a good and effective fraud detection system should protect user's data. An adequate and well supervised method consist of a collection of sample records, and these records are either classified as fraudulent or non-fraudulent.

A model is constructed in the process using this data which the technique is made to identify whether the document is fraudulent or not.

(vii) Financial Banking:

The computerization of the banking process is meant to generate an enormous amount of data with every new transaction.

The data mining technique is very helpful to the banking industry in such that, it can help bankers to solve business related problems by identifying trends, casualties, and relations in business information and market costs that are not immediately known to managers or executives because, the data size is too large or they are hurriedly produced on the screen by experts. However, the manager can find these data useful for better targeting, acquiring and maintaining a profitable customer.

12.0 CHALLENGES IN DATA MINING

Data mining is very powerful but it faces various types of challenges during its execution. Such challenges could be attributed to performance, data, methods, and techniques, etc. We can then say that, the process of data mining becomes successful and effective when all the challenges or problems encountered are correctly recognized and adequately solved. Some of the challenges in data mining are described below:

(i) Incomplete and noisy data

Data mining is the process of extracting useful data from large volumes, the data extracted in the real-world is heterogeneous, incomplete and noisy. Data in large volume will usually and always be inaccurate or even unreliable. These may be as a result of bad data measuring instrument or because of human errors. For example, some customers may not be willing to disclose their complete phone numbers, which results in incomplete data or perhaps the data could be changed due to human or system error. These mentioned instances make data mining challenging.

(ii) Data Distribution

In real world, data is usually stored on various platforms in a distributed computing environment. It could be in a database, individual systems, or even on the internet (cloud computing).

It is quite a difficult task to store all the data to a centralized data repository, due to organizational and technical concerns. For example, different regional offices may have their own servers to store their data. It is not advisable and feasible to store the data from all the offices on a central server. Therefore, data mining needs the development of tools and algorithms that will allow the mining of the distributed data.

(iii) Complex Data

Complex data is known to be heterogeneous, that is, it could be multimedia data, audio and video, images, spatial data, time series and so on. So, managing and extracting useful information from these types of data is typically a huge task. In most cases, new technologies, new tools and methodologies would have to be engaged in order to obtain a specific information.

(iv) Performance

The data mining methods and performance depends primarily on the efficiency of algorithms and techniques used. If the algorithm and techniques that are used are not up to the task, then the effectiveness and the efficiency of the data mining process will be adversely affected.

(v) Data Privacy and Security

Data mining in most cases usually leads to a crucial issue when it comes to data security, governance and privacy. For example, if a customer analyzes the detailed information of the items bought, then it will reveal data about the retailer buying habits and preferences without their permission.

(vi) Data Visualization

Data visualization is a very important process in data mining because, it is the primary method that reveals the output to the user in a well presentable way. However, the extracted data should portray the exact meaning of what it intends to describe. Most times, showing the information to the end-user in a precise and easy way is a bit difficult. Furthermore, the input data and output information must be very efficient and successful data visualization processes need to be implemented in order to make it successful.

13.0 DATA WAREHOUSE AND DATA MINING SUMMARY

Data Warehouse	Data Mining
✓ The process of compiling and organizing data into one common database	✓ The process of extracting useful data from the databases
✓ Data is stored periodically	✓ Data is analyzed repeatedly
✓ It allows easier reporting	✓ The process of determining data patterns
✓ It updates frequently	✓ Uses pattern recognition techniques to identify patterns
✓ They are created to support management systems	✓ Cost-efficient as compared to other statistical data applications
✓ It is a database system designed for analytics	✓ Business entrepreneurs utilizes data mining with the help of engineers
✓ The process of combining all the relevant data	✓ Detection and identification of the unwanted errors in the system
✓ It simplifies every type of business data	✓ Companies benefits from this analytical tool by equipping suitable and accessible knowledge-based data.
✓ It stores a large amount of historical data that helps users to analyze different trends to make future predictions	✓ The techniques are not 100 percent accurate; it may lead to serious consequences in a certain situation

14.0 CONCLUSION

Data analysis can be defined as a method of cleaning and modeling data in order to find useful information most especially for business decision making.

It can further be defined as a process of extracting useful information from data gathered from different data source and then taking decisions based on the data analysis. Data analysis comprises of processes such as; Data gathering, Collection, Cleaning, Analysis, Interpretation and Visualization. Its types include; Text, Statistics, Diagnostic, Predictive and Predictive analysis.

Analysts can also use a well-designed and a robust statistical measurement such as a hypothesis to solve certain analytical problems Furthermore, regression analysis can be used in a scenario where a data analyst is trying to determine the extent to which independent variable X affects dependent variable Y. Necessary condition analysis (NCA) can be used in a situation where the data analyst is trying to determine the extent to which independent variable X allows variable Y.

Barriers to effective data analysis have been noted to exist between the data analysts that is performing the data analysis. Differentiating the fact from opinion and innumeracy have all become a challenge to data analysis.

The term Data Warehousing is not a new phenomenon because all large organizations already have data warehouses, but they are just not managing them efficiently. Data warehouses are an integrated part of data that can be used to support business analysis and reporting. Although, a few organizations have actually implemented multiple data warehouses in order to support different locations or functions within their organization.

Data warehouse helps to enable data integration in an organization manageable by providing a centralized hub of data to be used for reporting and analysis. Therefore, every consumer who want to access the data can simply get it from a single point rather than having to go to different operational applications directly.

In today's real world, data warehouse plays an important role in order to perform important operations. Different indexing techniques has been engaged and analyzed using various types of queries on different size of databases in the data warehouse in order to perform operation in an efficient manner. Furthermore, data warehousing has been noted to be the most trustworthy technology used today by corporatist for scheduling, forecasting and management.

In the next coming years, the growth of data warehousing is going to be big with new products and technologies coming out frequently and in order to get the most out of this period, it is important that, data warehouse planners and developers must have a better understanding of what they want and then opt for strategies and methods that will provide them with performance today and flexibility for the future.

Nevertheless, the awareness of data warehouse and data mining in the organization should take into consideration many aspects regardless of what industries. Such aspects include; support of the top management, understanding of the data needed by the organization, governance and policy, the right design of the data warehouse and the right tools or techniques for data mining.

15.0 BIBLIOGRAPHY

Chegg. (n.d.). *Data Warehouse Application Examples*. Retrieved from Data Warehouse Application Examples: https://www.chegg.com/homework-help/questions-and-answers/data-warehouse-application-examples-project-required-write-shell-script-create-edit-view-d-q28566160

Javatpoint. (n.d.). *Data Mining Vs Data Warehousing*. Retrieved from Data Warehouse: https://www.javatpoint.com/data-mining-cluster-vs-data-warehousing

Lastnightstudy. (n.d.). *Data Warehouse Architecture*. Retrieved from Three-Tier Data Warehouse Architecture: http://www.lastnightstudy.com/Show?id=48/Data-Warehouse-Architecture

Petersen, R. (2014, August 17). *8 guidelines for great data visualization (with examples)*. Retrieved from 8 guidelines for great data visualization (with examples): https://barnraisersllc.com/2014/08/17/critical-components-great-data-visualization/

Tawde, S. (n.d.). *Advantages of Data Mining*. Retrieved from Introduction to Data Mining: https://www.educba.com/advantages-of-data-mining/