Data Analysis and Visualization by extracting insights for Efficient Project Planning

Syed Tameem Ansari[1*], Azra Nasreen[2]

[1,2]Dept. of Computer Science and Engineering, R V College of Engineering, Bengaluru, India

*Corresponding Author: syedtameema.scs18@rvce.edu.in, Tel.: +91-8123816082

Abstract—Data analysis and Visualization is a new and trending field in computer science. Visualizations represent a means to communicate data and analysis results. It uses graphic effects to reveal patterns, trends and relationships out of data. It involves in representation of data in pictorial or graphical form which makes the information easy to understand. Visualization and data analysis can be performed on many tools such as Tableau, Apache spark, Excel spread sheets. Visualization and data analysis using Power BI is affordable, quite simple and easy to develop and maintain as it supports real time data processing. Multiple data sources can be used to imported data for analysis into Power BI for developing custom visualizations that helps in-detail understanding of resources and budget used in development and management of a project. M-query and DAX are the functional languages used to extract insights, merge or append queries, create new measures and new custom columns for developing attractive and self-explanatory dashboards about resources in developing and maintaining projects. Visualizations and data analysis help in better prediction and forecasting for the future projects in the organization, these visualizations and dashboards can be shared on cloud based BI environment at managerial level and respective lower levels with row level security.

Keywords— Data Analysis, DAX, M-query, Power BI, Visualization

I. INTRODUCTION

Data visualizations are the graphical representations of data or information where visual elements such as graphs, charts, and maps form the data visualizations that provide the users with an easy and simple way of understanding the represented information [1][2]. With Big-Data, visualization allows decision-makers of every enterprise or an industry to look into analytics and reports and understand concept that might otherwise be very difficult to achieve.

Basically, Data analysis and visualizations is all about taking up data and recognizing patterns which is an advantage for planning of future projects systematically and helps in making data-driven decisions and predictions. It is hard to imagine of professional industries that does not benefit from making data more understandable. Every STEM (science, Technology, Engineering and mathematics) field takes benefits from understanding data and so do fields in government, marketing, history, goods, consumer service industries, education, sports and so on. It is highly valuable for professionals to find patterns in the data and make decisions and use visualizations to tell stories of what data informs, how, where, who and when, while the conventional method commonly draw a distinct line between technical analysis and creative storytelling, the new or modern professional also merits those who can cross between the two: data visualizations comes in the midst of storytelling and analysis. Visualizations consists of variety

of tools and that will help the user to fetch knowledge from complex data and communicate efficiently. Analytics using visualization is multi-disciplinary field starting from Database, Business, Risk management and also security [3].

Data analysis is process of modeling, cleaning, and transforming data with the aim of discovery of patterns which provides useful information, conclusion and support for making data-driven decisions. Data analysis has multiple forms and approaches, including various methods of working under various names, and it is used in variety of businesses, science and technology, and other domains. In today's competitive world, data analysis plays a vital role in making business decisions more scientific and helping businesses operate more efficiently [4].

Data mining is a process of a particular data analyzing technique that has a goal of statistical modeling and discovery of knowledge for prediction rather than given description, while business intelligence (BI) covers data analysis technique that depends more on focusing, aggregation mostly on business information. In statistical applications, data analysis can be divided into descriptive statistics, exploratory data analysis (EDA), and confirmatory data analysis (CDA). EDA aims on discovery of new features in the data, while CDA aims on utilizing statistical tools like significance, confidence and falsifying or confirming existing present hypothesis, it also depends on regression analysis and hypotheses testing. All the above are forms of data visualization and analysis.
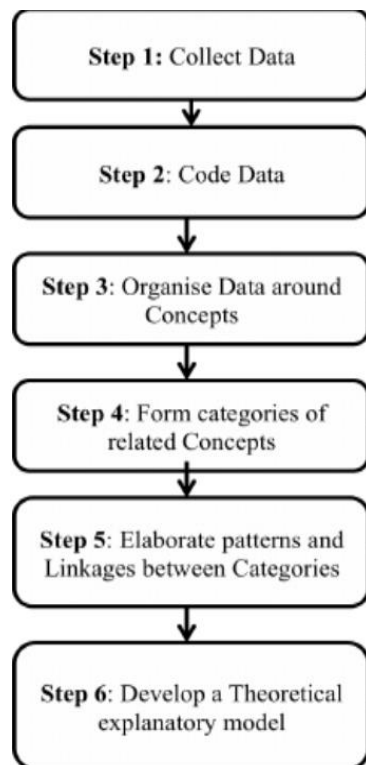
Figure 1: Common steps for Data analysis

As in figure. 1, the entire data analysis is split into multiple layers like data collection, organizing, preserving relevant data and generating theoretical reports. The collection data is done by importing the data from multiple data sources that may or may not be organized. Organizing the data is responsible for cleaning and the performing data pruning operation on the collected data. Categorization of data is necessary to form groups of related concepts. So here, the complete software application is deployed as whole, i.e., all or nothing and the application runs as a single process.

Excel is one of the best data analysis tools for accountants, but most accountants will not be aware about the of Excel data analysis functions and features like visual basic applications. Many accountants will be able to perform data analysis using PivotTables, but quickly realized about the constraints and disadvantages of Excel spreadsheets, which will lead to complex calculations for simple tasks.

Microsoft released a Power Pivot tool initially to remove the weakness of PivotTables, this entire suit of product is called Power BI (Business Intelligence). Power BI is a collection of connectors, software services and application to work jointly to turn unrelated data into logical, visually immersive, and extract interactive insights. The data may be in an excel file format or a on-premises hybrid data warehouse and collection of cloud-based data.[5].

Similarly, just like other BI tools and Business analytics methods, visualizations must include effective data management methods and strategies so as to standardize, unify and integrate data incoming from multiple sources. This approach mainly focuses on collecting data from

multiple sources and performing data analysis and developing visualization reports using Power BI desktop and publishing it on organizational clouds for project managers and employees understanding patterns of resources used in the development and maintenance projects which is of great help to overcome wastage of resouces.

This paper emphases on Data Analysis and Visualization by Extracting insights for Efficient Project planning, where section I contains brief introduction on Data Analysis, section II contains description of related work which refers two different papers, section III consists of Theory related to Data analysis using Power BI tool, sections IV contains Methodology of the study carried out, sections V and VI contains Discussions and brief results of the study, and this paper concludes with the section VII that contains the Conclusion and Future scope of this research.

II.   RELATED WORK

Related work is carried out to analyze different research works, research gaps and discuss their results that are already done in the field data analysis.

Various performance metrics and visualizations are considered to prioritize and identify the locations of the projects. A company has a concern for business process to identify threats and opportunities, saving of cost, and ambiguities in hundreds of projects in transportation system. This gives a description of supporting data analytics, and prioritization to help in the planning of improvement in projects. This approach includes the algorithm to extract relevant insights from multiple data sources into a single file. The results aid in planning of transportation at multiple levels at the organization to make data-driven decisions in planning projects efficiently [6]. This approach provides feature such as data security, how to confirm interfaces are flexible to detect missing or inaccurate data.

The data mining becomes a gold mine only if intelligent analytics and tricky algorithms are used over data, simultaneously analytics results are visualized in efficient, impressive manner, this paper [7] gives an idea of most used and spread techniques and tools of visualization for large datasets, and presents the main functional and nonfunctional characteristics of tools which are surveyed.

Visual analytics techniques build an interactive dashboard of huge data which helps in identifying the patterns. This [7 from ijcse] approach focuses on providing an overview of visual analytics, scope, techniques used, tools used and the most significant technical research challenges in the area. Visual analytics combines computer science, perceptual science, social science, and graphic design. Visual analytics allows the human -information, communication by bring together theory-based and computational tools with new interactive visual representation.

III. THEORY

This section gives a theoretical explanation of performing Data analysis using BI tool and its salient features.

III.I   Motivation

Lately there has been rapid changes in the field of data analytics tools and software landscape with introduction of some of the technologies like support of structured and unstructured data formats, Cloud-based BI environment, support of real-time data processing. Similarly, Power BI is one such tool that can be deployed on-cloud or on-premise within the organization. It can also import data from multiple sources such as local database, simple excel files, hybrid sources or cloud-based data sources. Power BI seems to be better among many other BI tools since it is efficient and user-friendly as it supports formula language called DAX (Data Analysis Expression) and M-query for data analysis. It qualifies to consolidate data from many sources, find patterns in the data, make interactive dashboards, develop reports and share it with other users which is of great help for efficient project's planning and to make business related decisions.

III.II   Introduction to Power BI

Power BI is one of the Business Intelligence and data visualizations tool which transforms data from multiple sources to good interactive dashboards and BI reports. Power BI desktop application is used to create reports and Power BI service i.e., software as a service is used to publish interactive reports on cloud based BI environment within the organisation with row level security [9].

Visualizations or the visual representation of huge data plays an important role in this BI tool. It provides a various range of attractive and detailed visualizations. Reports and dashboards can be created using some simple or as complex visualizations as necessary [10].

III.III   Basic Power BI architecture

The Power BI service architecture is mainly based on two clusters Web Front end (WFE) the Back-end cluster. The WFE cluster maintains the initial connection and authentication to Power Bi service and acts as an interface between client and back end, the Back-end maintains all subsequent interactions such as visualizations storage, reports, data refreshing with Power BI. In the Back-end the client has two points of interaction Gateway role and API management which is responsible for load balancing, routing, and authorization.

Visualizations are used to present the data constructively, helps in giving life to huge data and are basic building blocks of any business Intelligence tool.
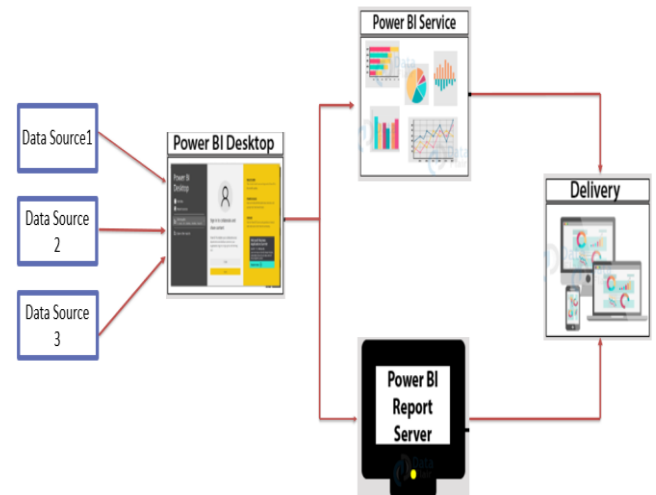

Figure 2: Power BI architecture

Each independent service is responsible for preserving its data, control, and external states unlike traditional models where a separate layer manages. These services communicate between each other using well defined APIs (REST, HTTP). Power BI also supports security used by cloud that includes Multitenant Environment security, networking security and Active group security.

Besides these, there are some other common components in Power BI service architecture for performing data analysis and building visualizations as shown in figure.2 such as

  i. **Data sources:** Power BI can import from multiple data sources simultaneously and refresh the required data as and when necessary.
 ii. **Power BI desktop:** It is client-side tool knows authoring tool which provides software as a service and gives a platform to organize, create business metrics, define hierarchies, develop visuals and publish reports.
iii. **Power BI Service:** It is web-based platform where reports, dashboards are shared and collaborate with other users.
 iv. **Power BI report Server** It is similar to the power service, but the only difference is that it is an on-premise platform i.e., it is only used within the organization and also follows the security protocols.
  v. **Power BI gateway:** This is used to connect and access on-premises in a secured environment where gateway helps to extract data through safe channels for analysis and reporting [8].

IV. METHODOLOGY

This section is about methodology for Data analysis using Power BI using the data which is being imported from an Engineering Data Base and other data sources.

IV.I. Importing Data

Data is collected from multiple sources such as Engineering Data Base and other organizational databases and imported to spread sheets. Since the data is imported from data sources, it may or may not be organized in

spread sheets. Data cleaning is performed which includes removal of corrupted values, organizing, adding missing values and preserving relevant data and it is imported to Power BI.

IV.II. Data in Power BI
Data in Power BI is again organized according to the user requirement. Time to time interactions with Business analysts as they suggest how and which data should be preserved for making data-driven decisions.

i. **DAX** – Data analysis expressions was created by Microsoft for specifically handling data models which uses formulas and expressions. DAX is scalable, extensive, easy for developer and end-user, efficiently handles data regardless of volume and variety [11]. DAX expressions include aggregate functions, logical functions, statistical functions, date and time functions, table manipulation functions, text functions, relationship management functions and many other dynamic library functions. All these functions are used to determine patterns and relationships within the data. DAX enables the data analyst to extract complex insights. DAX is used to develop calculated columns by combining two or more entities of the existing data. Calculations can also be applied on an existing column to create a new measure or metric.

ii. **M-query** – It is also knowns as Power-Query which enables to easily understand and connect to data from corporate and public data sources, this includes new data search to merge and transform data from multiple sources. M-query is used to extract in-detail insights from complex data from multiple files that may not be understandable.

In this approach DAX functions written to perform virtual build analysis for every project. In virtual build analysis the total count of modules used in a single project is extracted. The details like man hours required to develop the module, department responsible for developing the module for a project, if the module is in development loop or not etc.

IV.III. Import query vs Direct query
Direct query method restricts the option of data manipulation and data remains in database. In direct query method there is no need to schedule a refresh. Import method enables to perform data manipulation and transformation. When the data is published on PBI service it can be refreshed 8 times a day or schedule can setup.

Advantages of using Direct query method is, it can be used to build visualizations for large datasets, which is not possible through Import method. Since the data is huge for the managing the projects efficiently Direct query is the best suitable method.

IV.IV Visualizations and Reports
Visualizations are the basic building blocks of any business intelligence tool which is developed to give life to the data and present it effectively [12]. There are many default visualizations components in Power BI which can present simple and complex data from multiple sources in a clear way. Visualizations are developed to make data-driven decision which further helps the business analysts to give predictions and forecasting for the future projects.

In this approach we have visualizations that present the details of every individual, number of man hours required, version of the modules used in every project.

IV.V Power BI service or Cloud based BI environment
Cloud business intelligence apps run on virtual network [13], they are used to provide access to the organizations to BI related data such as KPIs, dashboards, and other business analytics [14]. Therefore, in this approach Dashboards and reports are published on to Power BI report server and are accordingly shared with respective individuals such as at managerial level, business analysts and project managers within the organization with row level security. Basically, row level security is process to filter content based on user's role.

V. DISCUSSIONS

Before beginning to collecting data, data analyst should start by identifying the business questions that you want to answer to achieve the organizational goals. The organization missions are to understand the amount to resources used to manage existing projects or the projects in Development Loop (DL) with the help of Data Visualization.

The resources used the projects include budgets used for project development, man hours, contribution of every department in the organization and other respective details in an efficient manner using Power BI tools which gives quick results unlike finding for answers using conventional method of spreadsheets. But converting data into relevant format using DAX functions or M-query in BI tool consumes time according to the number of rows. However, DAX function takes more time to compute data from multiple sources when compared to M-query.

VI. RESULTS

As discussed, data is imported from multiple sources and analysed. In Power BI, performance of filtering the data using DAX functions according to the visualizations in the report is better when compared to pre-processing the data using spreadsheets as shown below in table.1.

Table 1:Time for processing data using DAX functions

No. of Rows processed using DAX function	Size in KB	Total Time duration in milliseconds
2,520,000	96,871	13,115ms

Filtering huge amount data using DAX functions and M-query on Power BI is quick and fast which consumed 13,115ms to process 2,520,000 rows.

This approach helps business analysts in understanding and performing business decisions or data-driven decisions for predicting and forecasting the usage of resources used for maintaining existing projects or the resources used for future projects in an efficient, quick and robust manner.

VII. CONCLUSION AND FUTURE SCOPE

The development of various dashboards using project's raw data and the data related to resources used in every project. The organisation is able to visualize data from several different applications which facilitates efficient planning for future projects. The visualizations provide graphical displays that allow users to easily identify patterns in data and resources used during a project that may be in progress or completed. Hence, this approach of data analysis is user-friendly and more focused which enables the business analysts and project managers to perform data-driven decisions, better predictions and forecasting of future projects.

In this approach initially the data are being imported into spreadsheets where data pre-processing (cleaning and transforming) is performed which requires considerable amount and is then uploaded to BI tool. In future advancements the data may be directly staged in BI tool which may increase the performance and may require more complicated functions for data pre-processing.

REFERENCES

[1]. Stephen Few, "Selecting the Right Graph for Your Message", Perceptual Edge, **2004.**
[2]. Manuela Aparicio and Carlos J. Costa "Data visualization". Communication Design Quarterly Review. **3** (1): **pp 7–11. 2014.**
[3]. Shilpa M Patil, Dr. H S Guruprasad," A Study on Visual Analytics" International Journal of Computer Science and Information Technology Research, Vol. **3**, Issue **1**, **pp: (303-308)**, Month: January - March 2015
[4]. Xia, B. S., & Gong, P. "Review of business intelligence through data analysis." Benchmarking, An International Journal 21(2), **pp-300-311, 2015.**
[5]. "Power BI desktop" Published by Microsoft, 2018
[6]. Lily X.Chen, Ahsan A Chowdhury, Charles M. Loulakis, Micheal A.ownes. "Visualization of large data sets for project planning and prioritization on transportation corridors", Published in Systems and Information Engineering Design Symposium in University of Virginia. DOI: 10.1109/SIEDS.**2015**.
[7]. Enrico Giacinto Caldarola and Antonio Maria Rinaldi "Big Data Visualization Tools: A Survey - The New Paradigms, Methodologies and Tools for Large Data Sets Visualization", 6th international conference on Data science, technology and applications, University of Naples, Federico II, 80134 via Toledo, 402, **Napoli, Italy,** July **2017**.
[8]. R. Shankar, S. Duraisamy, "Visual Analytics: Need, Process, Scope, Tools & Techniques and Challenges" International Journal of Computer Science and Information Technology Research, Vol. **6**, Issue **10**, oct. **2018.**
[9]. Michael Hart, "Quick Insights with Power BI", Published by Microsoft, **2017**.
[10]. Ms.Komal "A Review Paper on Big Data Analytics Tools" International Journal of Technical Innovation in Modern Engineering & Science, Vol.**4, Issue 5, May-2018.**
[11]. Alberto Ferrari "The Definitive Guide to DAX The: Business intelligence for Microsoft Power BI, SQL Server Analysis", Published by Microsoft, **2019**.
[12]. Ana Lavalle, Alejandro Maté, Juan Trujillo "Requirements-Driven Visualizations for Big Data Analytics: A Model-Driven Approach" oct **2019**
[13]. A Venkatesh, Marrynal S Eastaf "A Study of Data Storage Security Issues in Cloud Computing" International Journal of Scientific Research in Computer Science, Engineering and Information Technology **2018**, Vol.**3**, Issue **1**.
[14]. Antonio M. Rinaldi "A Document Visualization Strategy based on Semantic Multimedia Big Data", Springer Conference paper, University of Naples Federico II, The 16th International Symposium on Pervasive Systems, Algorithms and Networks, **Napoli, Italy** September **2019**.

AUTHORS PROFILE

Mr.Syed Tameem Ansari is pursuing Master of Technology in the field of Computer science of Engineering from R V College of Engineering, Bengaluru, India from 2018 batch.

Dr. Azra Nasreen is currently working as an Assistant Professor at R V College of Engineering, Bengaluru, India and has teaching experience of 14 years, 1 year of experience in the industry and also has done 1 National Conference Publication, 10 International Journals, and 6 International Conferences. Area of interest are Video Analytics , High Performance Computing.