Analyzing Data with Different Charts and Visualizations in Power BI

Dr. R. Sangeetha[1], Dr. D. Elantamilan[2], Dr. A. Indrapandi[3]

[1]*Assistant Professor, Department of BCA, Vallal P.T.Lee CNASC, India*
[2]*Principal, Vallal P.T.Lee CNASC, India*
[3]*Assistant Professor, Department of BCA,*
The American College,India.
[1]*Email: geethasanmca@gmail.com*
[2]Elantamilan.ds@gmail.com
[3]*indrapandiaaa@gmail.com*

Abstract: This paper explores the process of Extract, Transform, and Load (ETL) and its application in data analysis using Power BI. ETL is a critical step in the data integration pipeline, allowing organizations to collect, clean, and structure data from various sources for insightful analysis. The study delves into the stages of ETL, detailing how data is extracted from multiple sources, transformed through data cleaning and aggregation techniques, and loaded into a data model for reporting. Additionally, the paper highlights how Power BI, a powerful business intelligence tool, leverages this prepared data for creating interactive visualizations, reports, and dashboards. By combining ETL processes with Power BI's visualization capabilities, businesses can effectively analyze large datasets, uncover trends, and make data-driven decisions. The paper also discusses challenges related to data quality, performance optimization, and the integration of real-time data for enhanced analytics.The objectives of this paper are to gain an understanding of the ETL process and to demonstrate the visualization of a dashboard in Power BI. Overall, it demonstrates the synergy between ETL and Power BI in enabling comprehensive data analysis for organizations.

Keywords: Data Analyze, Visualizations, ETL and Power BI

1. Introduction

Power BI is a powerful business intelligence (BI) tool developed by Microsoft, designed to help organizations transform raw data into meaningful insights through interactive reports, dashboards, and visualizations. With its user-friendly interface and robust capabilities, Power BI enables business analysts, data professionals, and decision-makers to make data-driven decisions by analyzing data from a variety of sources, including databases, spreadsheets, and cloud services.At the heart of Power BI's data preparation and analysis process is the Extract, Transform, Load (ETL) process. ETL is a crucial step in the data integration pipeline, which involves extracting data from various sources, transforming it into a suitable format for analysis, and loading it into a data model or destination system. In Power BI, the ETL process is largely facilitated by Power Query Editor, an intuitive tool that simplifies data extraction and transformation(6).The Power BI architecture diagram highlights the core components of the Power BI ecosystem. It includes data sources, Power BI Desktop, Power BI Service, and Power BI Gateway. Data is extracted from various sources, transformed in Power BI Desktop, and published to the Power BI Service for sharing and collaboration. Power BI Gateway facilitates secure data transfer between on-premises data sources and the cloud.Figure 1 displays the architecture diagram of Power BI[1].

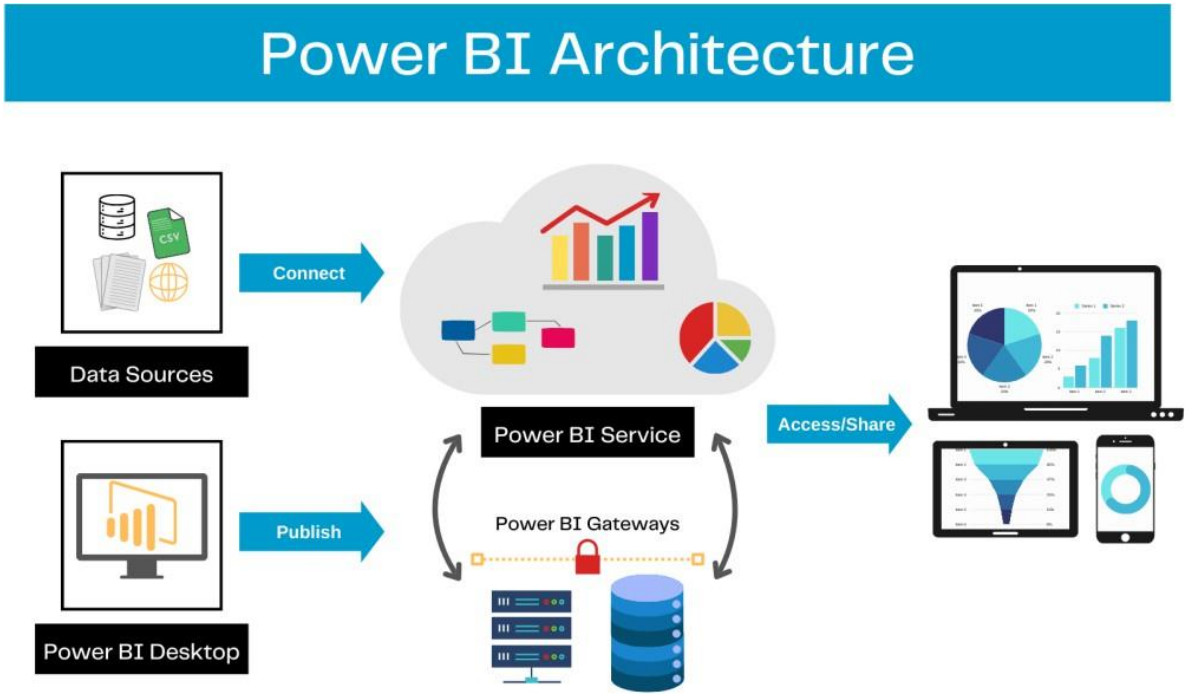

Fig 1: Power BI Architecture Diagram

2. LITERATURE REVIEW

Satkaur and Anuj Mehta assist the reviewer in understanding the ETL process. In their paper, they discuss ETL tools and the ETL process [1]. Sagar Bhujbal, Dhanesh Gite, Yadnesh Kadam, and Bhushan Narkhede, in their paper, explain ETL tools, the ETL process, and data warehouse ETL models [2]. Preeti Dhanda and Neetu Sharma, through their paper, help clarify the ETL processing involved and describe an approach for migrating both historical and current organizational data to a data warehouse product [4]. Qin Halnin, Jin Xianzhen, and Zhang Xianrong explore key ETL technologies, including data extraction, transformation, incremental loading, and breakpoint transmission [5]. This paper also discusses the best ETL tool: Informatica [3]. Priyanshu Gupta focuses on explaining data warehousing.

3. RESEARCH METHOD

The training dataset, provided by MRC Private Limited, is utilized for the research paper. This dataset includes basic and intermediate training across various teams and job levels, along with associated costs. The working process of Power BI involves several key steps that help users transform raw data into meaningful visualizations and reports.

Data Collection: Power BI connects to various data sources, including databases, Excel files, cloud services, APIs, and more, allowing users to import data from these sources into the platform.

Data Transformation: After data collection, Power BI provides tools like Power Query to clean, transform, and shape the data. This step includes filtering, merging, and modifying data types to make the data ready for analysis.

Data Modeling: Once the data is cleaned, Power BI enables users to create data models by establishing relationships between datasets. This includes defining calculated columns, measures, and hierarchies to support efficient analysis.

Data Visualization: Power BI offers a wide variety of visualization tools (such as charts, graphs, tables, and maps) to help users create interactive, dynamic reports and dashboards. These visuals can be customized to present data in an intuitive and meaningful way.

4. DATA ANALYSIS AND VISUALIZATION

First, Power BI collects data from various sources, such as databases, Excel files, and more. It then transforms the raw data into a structured and clean format for analysis.

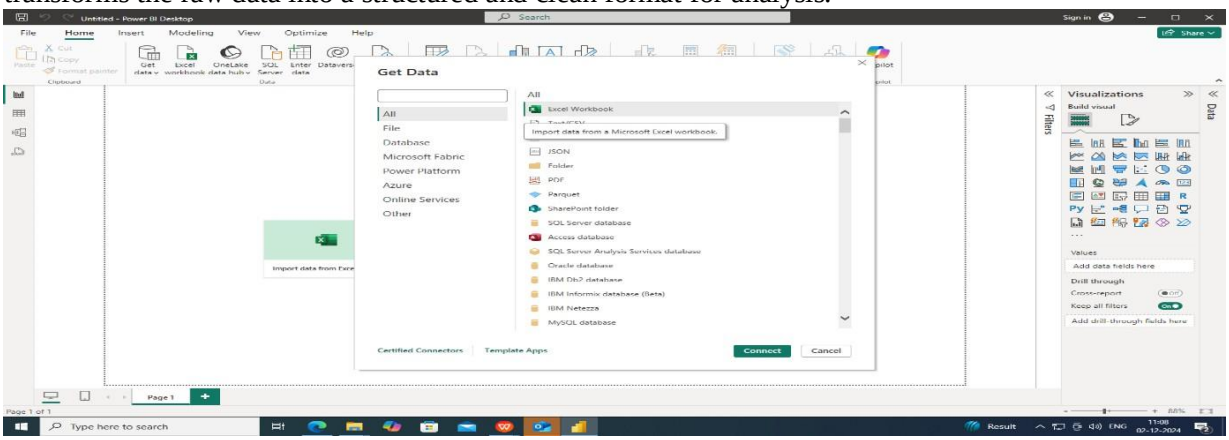

Fig2. Connect data sources

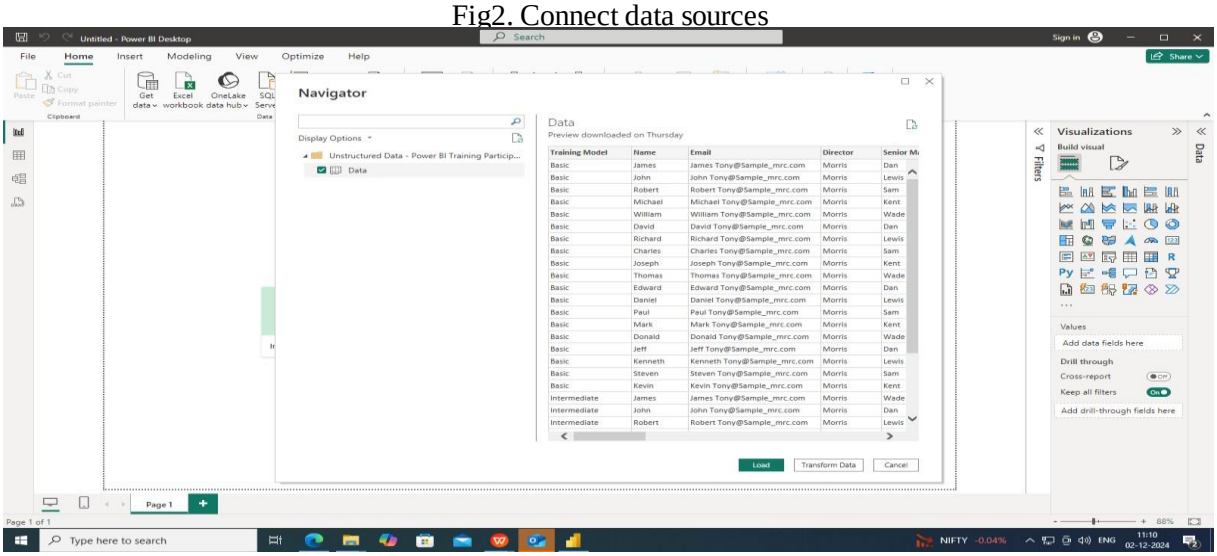

Fig 3. Load and Transform data

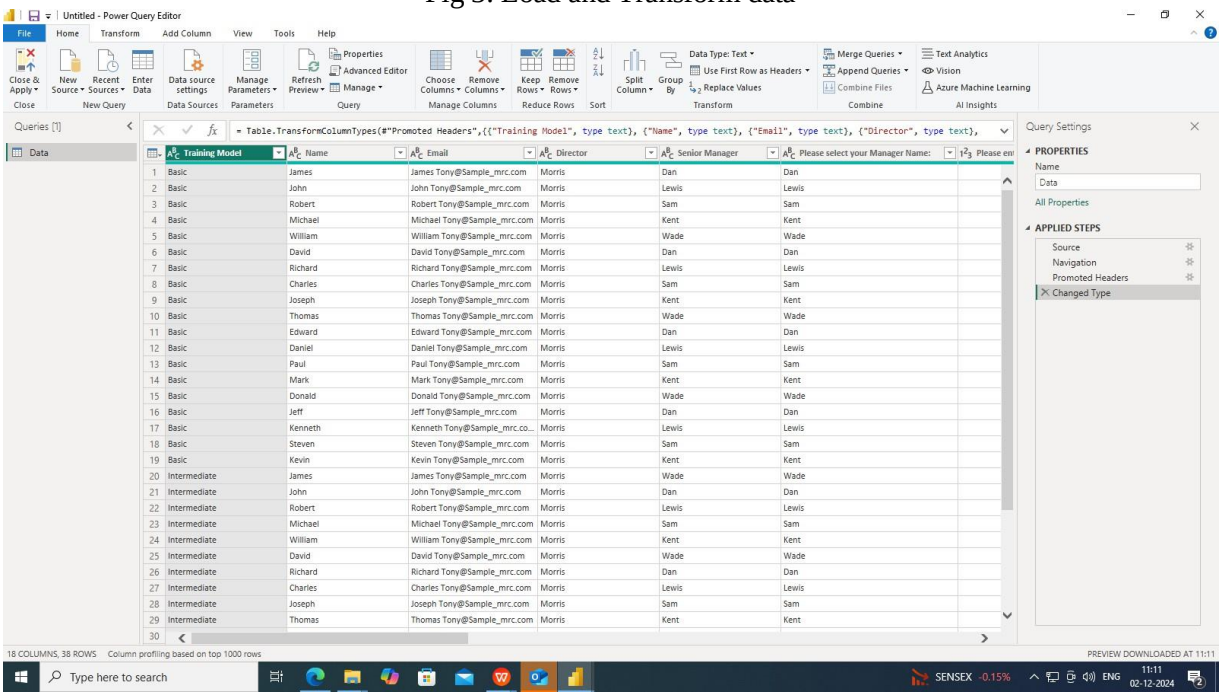

Fig 4. Shape and Transform the data

In this step, we can add new columns and remove duplicate columns. Additionally, modifications can be made based on the user's requirements.
Once the transformations are complete, go to the File tab in Power Query Editor and select "Close and Apply" to load the data. Visualizations are then created to analyze the data.Here are a few sample visualizations shown for reference.First, using this training data, Power BI visualized the data by team and manager using a column chart.

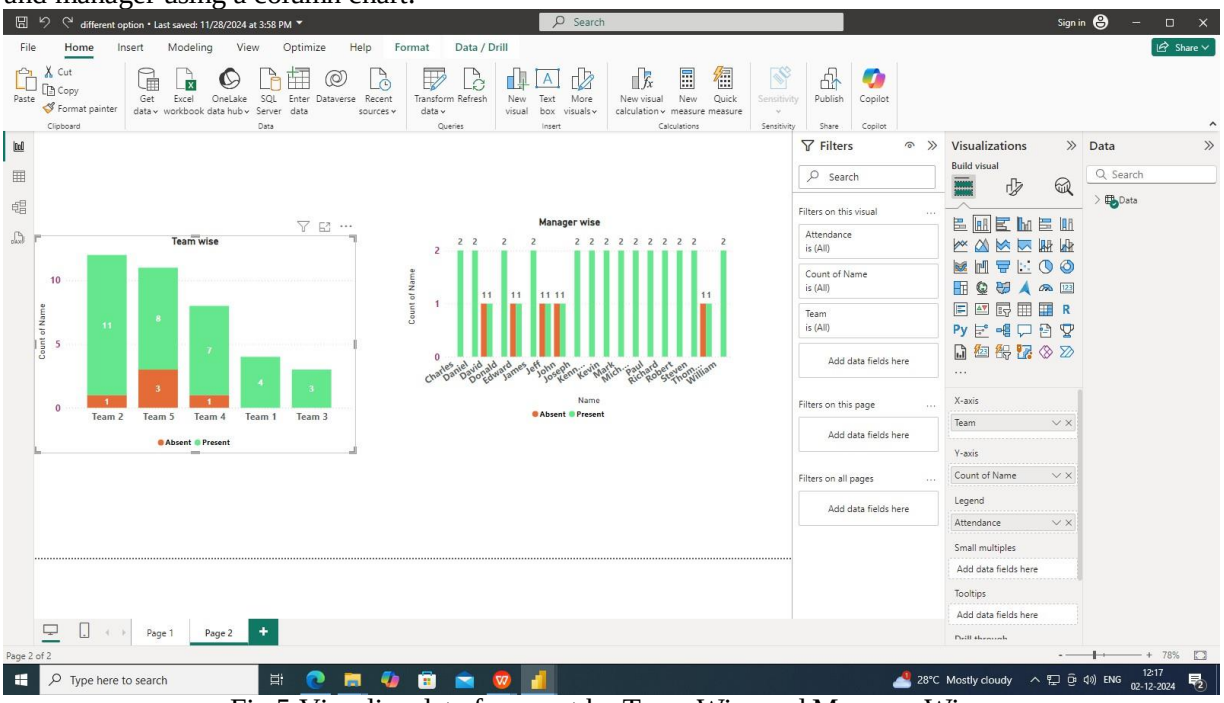

Fig 5.Visualize data for count by Team Wise and Manager Wise

In the same way, I have created a pie chart, treemap, and various cards to count the data for decision-making, all of which are displayed in a single visual.Figure 6 displays various visuals used to count the data for decision-making.It is shown in Fig 6.

5. FINDINGS

Training data is as follows:

- Gender – Male,Female
- Team Wise– Present,Absent
- Manager Wise-Present,Absent
- Total No of Enrolled-38
- Total No of Present & Absent-38 &5
- Average-86.84%

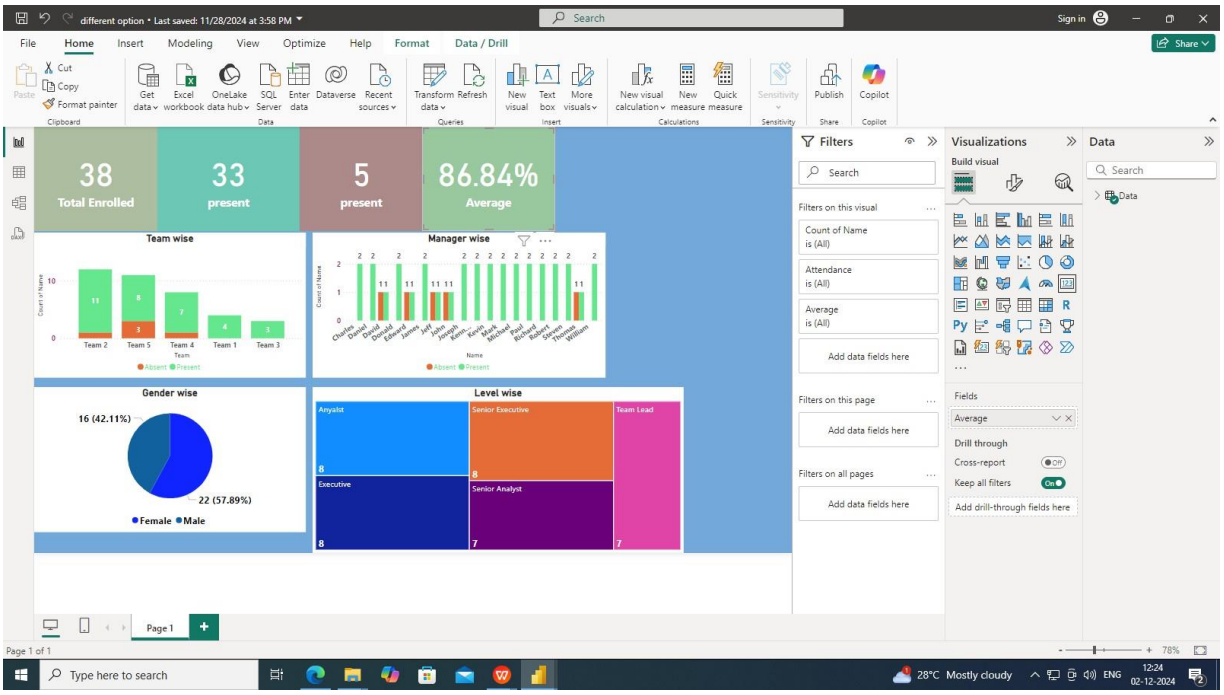

Fig 6. Different Visualized Data

Gender Distribution Analysis:

Pie Chart: Display the proportion of male vs. female participants. This visualization helps identify if there is a gender imbalance within the enrolled individuals, which can assist in ensuring diversity and inclusion in training programs.

Decision: If one gender has a significantly lower enrollment rate, it may indicate the need to tailor marketing or outreach efforts to improve gender representation.

Team-wise Attendance:

Column or Bar Chart: Visualize the attendance status (Present vs. Absent) for each team. This can highlight which teams have better attendance rates and which might require attention.

Decision: Teams with higher absenteeism may need additional support or engagement strategies, such as flexible scheduling or team-building activities to boost participation.

Manager-wise Attendance:

Manager Dashboard: Visualize how each manager's team performs in terms of attendance. A stacked bar or column chart can show the number of team members present and absent under each manager's supervision.

Decision: Managers with teams showing lower attendance could benefit from leadership training or better communication strategies to encourage attendance.

Total Number of Enrolled:

Card Visual: Display the total number of participants (38). This gives a quick snapshot of overall enrollment and can be used to monitor against set enrollment goals or historical trends.

Decision: If the number of enrolled participants is lower than expected, outreach efforts may need to be reassessed to increase enrollment.

Total Number of Present & Absent:

Card and KPI Visualization: Show the total number of participants present (38) and absent (5). This helps monitor the effectiveness of the training program and attendance rates.

Decision: If absenteeism is higher than expected, it may indicate issues such as poor engagement, lack of interest, or external factors impacting attendance.

Average Attendance Percentage:

KPI Visual: Display the average attendance rate of 86.84%. This provides an overall picture of participation and can serve as a benchmark for future sessions.

Decision: If the average attendance rate is lower than desired, additional measures such as incentives for perfect attendance or reminders could be implemented.

Trends Over Time:

Line Chart: Use a line chart to track attendance over time, which could show trends in present and absent figures over multiple training sessions.

Decision: If attendance drops at certain times or sessions, targeted interventions can be designed to boost attendance in future sessions.
Identify High and Low Performers:
Treemap : Visualize performance or attendance in a hierarchical format, categorizing team or manager performance. You can identify which teams or managers have the highest or lowest attendance.
Decision: This can help make decisions on where to allocate resources, such as additional training, team support, or motivational initiatives for underperforming areas.
Resource Allocation:
Bar/Column Chart: Use visualizations to see how resource allocation correlates with attendance. For example, teams or managers with higher attendance may be benefiting from better resource management or engagement strategies.
Decision: If specific teams or managers have higher attendance, it may be a sign that they require fewer resources, whereas those with lower attendance may need additional support.
Comparison of Enrolled vs. Present:
Stacked Bar Chart: Display a comparison between total enrolled vs. present participants for each team or manager.
Decision: If enrollment is high but attendance is low, strategies to improve attendance, such as reminders or incentives, may be considered.
Based on these visuals, decisions can be made regarding resource allocation, attendance tracking, or identifying areas for improvement in team management and employee engagement.

6. CONCLUSION AND FUTURE WORK

This paper exhibits the effectiveness of Power BI in analyzing and visualizing training data, providing actionable insights into participant attendance, gender distribution, and performance across teams and managers. By utilizing various Power BI visualizations such as column charts, pie charts, treemaps, and cards, the analysis reveals an overall attendance rate of 86.84%, with a small number of absentees. These insights enable decision-makers to identify trends, allocate resources more effectively, and implement targeted strategies to improve engagement and attendance. The ability to track performance at the team and manager levels provides valuable information for improving leadership practices and optimizing training programs. Ultimately, Power BI proves to be a powerful tool for data-driven decision-making, offering a clear view of key metrics that drive organizational success.Future work will focus on incorporating predictive analytics, enhancing data granularity, and integrating the tool with other organizational systems to further refine decision-making processes.

References

1. SatKaur, Anuj Mehta, A Review Paper on Scope of ETL in retail domain: International Journal of Advance Research in Computer Science and Software Engineering. Volume 3, Issue 5, May 2013.
2. Sagar Bhujbal, Dhanesh Gite, Yadnesh Kadam, Bhushan Narkhede A New Database for Retail Domain using ETL System: International Journal of Advance Research in Computer Science and Mobile Computing. Volume 4, Issue 5, May 2015.
3. Preeti Dhanda, Neetu Sharma, Extract Transform Load Data with ETL Tools: International Journal of Advance Research in Computer Science, Volume 7, No. 3, May-June 2016.
4. Priyanshu Gupta, Data warehousing and ETL processes: An Explanatory Reserch: International Journal of Engineering Development and Research. Volume 4, Issue 4, 2016.
5. Qin Halnin, Jin Xianzhen, Zhang Xianrong, Research on Extract, Transform and Load (ETL) in Land and Resources star schema Data Warehouse: 2012 fifth International Symposium on Computational Intelligence and Design 2012.
6. Miss. Gandhali Joshi &Dr. Jyoti Kharade,An International Bilingul peer reviewed refereed research journal, Vol 6. Issue 26 , April-june 2020, Page Nos.34-39,ISSN-2348-2397.