MAN-POWER PRODUCTIVITY RATIO (MPR) ANALYSIS

Submitted by:

Y.DUSHYANT
GuruNanak Institutions Technical Campus-Hyderabad
S.NARESH
IIIT Pune

Carried out at:

SOUTH CENTRAL RAILWAY (Head Office: Rail Nilayam), SECUNDERABAD

CONTENTS

1 .Introduction

2 .Visual Analytics

2.1 Definition of visual analytics
2.2 State of research in visual analytics
2.3 State-of-the-art of tools and software

3 .Building blocks of visual analytics

3.1 Information visualization
3.1.1 Human perception
3.1.2 Data Graphics
3.1.3 Visualization techniques
3.2 Data mining
3.3 Integrated data sources

4. Literature survey

4.1 Man-power productivity ratio(MPR)

5. Design approach

5.1 Introduction to Power BI
5.2 Hardware Requirements
5.3 Software Requirements
5.4 Data under consideration

6.Analysis and visualizations

6.1 Visualizations for year 2017
6.2 Visualizations for year 2018
6.3 Inferences from the data

7. Conclusion

1. Introduction

Information visualization is defined as “The use of computer-supported, interactive, visual representations of abstract data to amplify cognition” [Card et al. 1999]. The goal is to improve understanding of the data with graphical presentations. The principle behind information visualization is to utilize the powerful image processing capabilities of the human brain. Visualizations increase the human cognitive resources. They extend the working memory; reduce the search of information and enhance the recognition of patterns. Information visualization has been an active research area since 1990. It has evolved from the use of computer graphics in scientific problems. The special issue of Computer Graphics on Visualization in Scientific Computing in 1987 is considered as the starting point. Since then there have been several conferences and workshops, co-sponsored by the IEEE Computer Society and ACM SIGGRAPH, devoted to the general topic, and special areas in the field. The “Visualization Time Line” is introduced in [Cook et al. 2007]. The driving forces of information visualization have been the growth of computing power, lowering computing costs, the development in user interaction technologies and the large amounts of information accumulated into databases. In the recent years internet technology, new rendering and user interface technologies, 3D and virtual environments, have activated the field. The recent advances and challenges of visualization are introduced in the report “Visualization Research Challenges” by U.S. National Science Foundation (NSF) and National Institutes of Health (NIH) [Johnson et al. 2006]. Information visualization has traditionally been divided into scientific visualization and visualization of abstract information, data visualization. Scientific visualization deals with large sets of scientific data to see phenomena in data. It handles mostly with physical data: human body, earth, molecules and with data that has a natural geometric structure, for instance wind flows. In scientific visualization the computer renders visible some properties of interest from the data. Data visualization handles abstract, non-physical information using abstract visualization structures like trees or graphs. It has applications with financial data, business information, document collections, web content and other abstract concepts. It renders visible properties of the objects of interest and can be combined with information access techniques.

Visual analytics is the most recent field of information visualization. It has been defined as “the science of analytical reasoning supported by the interactive visual

interface" [Thomas & Cook 2005]. It has origins in U.S. national security and is one approach to tackle the information overload problem caused by the improvement of data storage devices and means to collect data. It provides visual tools to support analytical reasoning and decision making from data with interactive visualizations, optimized for efficient human perception. It is a multi-disciplinary research area, combining information visualization science, data mining, mathematical and statistical methods, data management, user interface techniques, human perception and cognition research (Figure 1).

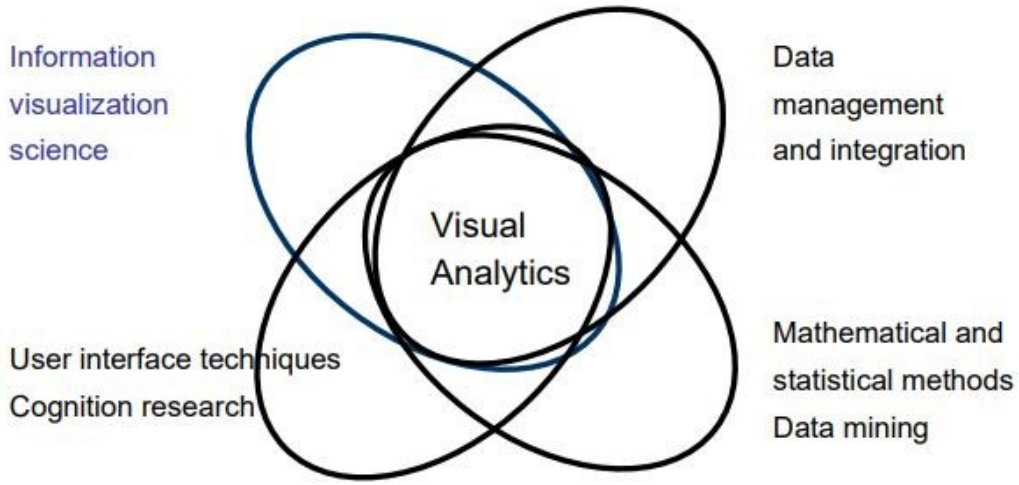


Figure 1. Building blocks of visual analytics.

The branches of information visualization have lots of overlapping goals and techniques. In this report we focus on visual analytics. We introduce the concept; review the state-of-the art of research and the tools and software on the markets. We introduce the main building blocks of visual analytics. We also introduce the visual analytics tool developed in the project and outline roadmaps for visual analytics in industrial applications and consumer applications.

2. Visual analytics

2.1 Definition of visual analytics

The basic idea of visual analytics is to combine the strengths of automatic data analysis with the visual perception and analysis capabilities of the human user. It uses visualizations, user interaction and data analysis techniques to find insight from complex, conflicting and dynamic information. Visual analytics is especially focused on situations where the huge amount of data and the complexity of the problem make automatic reasoning impossible without human interaction. The starting point for visual analytics is the report ”Illuminating the Path: the R&D Agenda for visual analytics” by U.S. National Science Foundation(NSF) and National Institutes of Health (NIH) [Thomas & Cook 2005]. It defines visual analytics as follows: “Visual analytics is a science of analytical reasoning facilitated by interactive visual interfaces.”

“People use visual analytics tools and techniques to synthesize information and derive insight from massive, dynamic, ambiguous, and often conflicting data; detect the expected and discover the unexpected; provide timely, defensible and understandable assessments; and communicate assessment effectively for action. “ The report leaves the concept to a quite general level and since then several publications has appeared introducing the idea and challenges [Thomas & Cook 2006], [Keim et al. 2006], [Cook et al. 2007], [Wong 2007], [Keim et al. 2008]. The objective of visual analytics is to develop visual and interactive tools and techniques for reasoning and decision making from large data sets. The principle of a visual analytics tool has been summed up in the visual analysis mantra by Daniel Keim [Keim et al. 2006]. “Analyze First – Show the important – Zoom, Filter and Analyze Further – Details on demand”. The visual analytics process is outlined in Figure 2 [Keim et al. 2008].

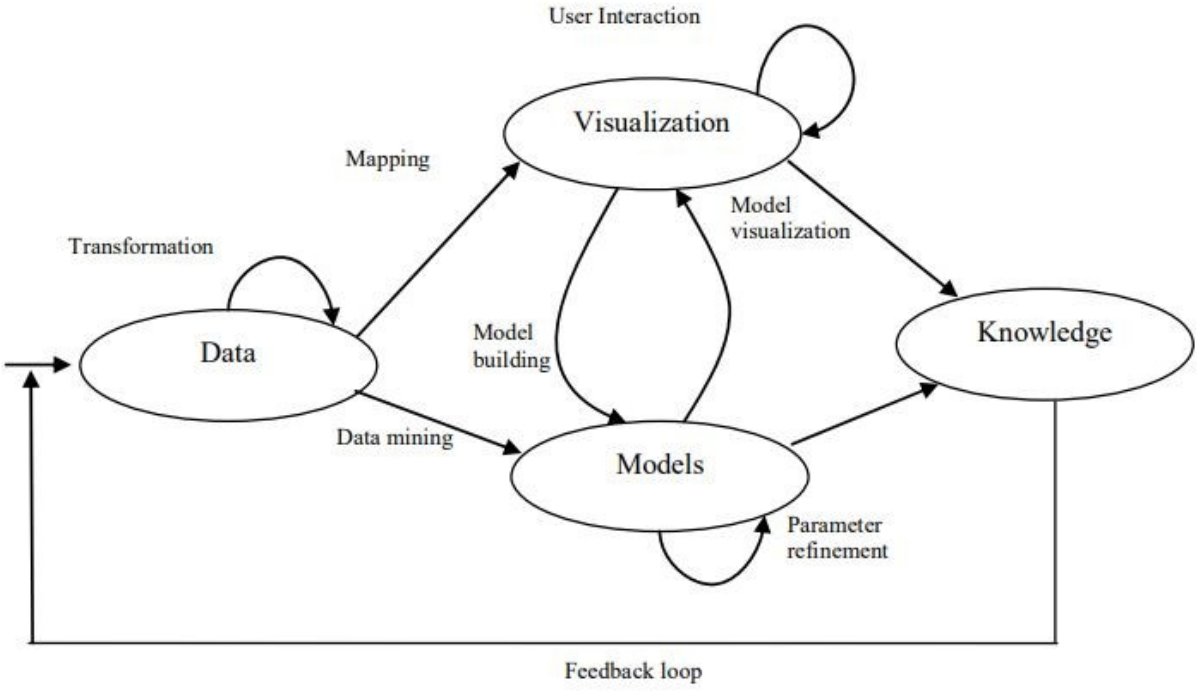


Figure 2. Visual analytics process.

Complex and heterogeneous data from different sources and various types and levels of quality need to be filtered, noise removed and transformed to be processed jointly. The data sources can range from well organized data bases to continuous input data streams. The data is processed and abstracted using mathematical, statistical and data mining algorithms and models. Visualizations highlight the important features, including commonalities and anomalies, making it easy for users to perceive new aspects of the data. Visualizations are optimized for efficient human perception taking into account the capabilities and limitations of the human visual system. Interactivity in visualizations allows users to explore the data and achieve new knowledge and insight.

A visual analytics tools supports:

+

- showing different views to data: from raw data to data abstractions
- representations of large quantities of information in small space
- finding pattern from data: similarities, anomalies, relationships and events
- simulation, prediction, testing hypothesis
- data retrieval, browsing and exploration
- information extraction and distillation.

Application areas for visual analytics are everywhere where is a need for decision making based on accumulated data. Although the origin is in safety and security, applications can be found as well in industry, engineering, business, media and consumer applications.

2.2 State of research in visual analytics

The increasing international importance of the topic is reflected by the strong presence of visual analytics at leading international conferences such as IEEE Visualization, IEEE Symposium on Information Visualization, and the Eurographics/IEEE-VGTC Symposium on Visualization. In October 2006, the first IEEE Symposium on Visual Analytics Science and Technology2 has been organized. Partners of this consortium have been actively involved in these conferences. Visual analytics also found its way into major journals in this area. Special issues on visual analytics have appeared in IEEE Transactions on Visualization and Computer Graphics, Computer & Graphics, the International Journal for Geographical Information Science, and SIGKDD Explorations.

The highly prospective utility of visual analytics research has led to significant national and international activities in this area. In the United States, the National Institute of Health (NIH) and the National Science Foundation (NSF) recently published the "Visualization Research Challenges" report [Johnson et al. 2006], which lists visual analytics as one of the most important challenges for future research. The US Department of Homeland Security (DHS) started a research initiative on visual analytics for homeland security. The "National Visualization and Analytics Center" (NVAC)3 , founded in 2004, coordinates these research efforts. The agenda for the US visual analytics research program is laid out in the book "Illuminating the Path" [Thomas & Cook 2005], which describes visual analytics research challenges focussing on security applications such as border security. In Canada, significant related research initiatives are under way and additional visual analytics programs supported by the Canadian government are prepared by Brian Fisher of Simon Fraser University4 . In Australia, Peter Eades of University of Sydney is currently leading an initiative to launch a national visual analytics research initiative. All of these international initiatives have stated their support for a European Coordination Action on visual analytics

In Europe, several national initiatives relating to visual analytics have been introduced or are in preparation. In Germany, the German Research Foundation DFG accepted a strategic research initiative on Scalable Visual Analytics5 in April 2007 proposed by a consortium of working groups6 led by Daniel Keim of University of Konstanz. The program is targeted to advance the national visual analytics research by supporting a number of individual research projects. In Switzerland, there are efforts underway to initiate a national centre for scientific visualization and visual analytics. In Austria, recently there has been introduced a research initiative on Visual Computing with specifically stated visual analytics relations within the Austrian FIT-IT program.

Data mining and machine learning research is in the focus of the KDNet7 and MLNet NoEs and the KDubiq8 CA. The visualization, analysis and management of mobility referenced to geography are in the focus of the FET-Open project GeoPKDD9 .

Disaster and emergency management applications are in the focus of the OASIS10 and ESS projects11. Visualization problems related to information retrieval have been addressed in the DELOS12 NoE as part of the user interface work done within that project. VisMaster Coordinating Action Project13 began on 2008 [Keim 2008]. The purpose of the VisMaster is to bring together a critical mass of interdisciplinary European researchers to scope the prospects for a European visual analytics initiative.

2.3 State-of-the-art of tools and software

Information visualizations are usually implemented with specialized software. Some of these have been released as open source software, having often origins in universities. There are also many proprietary software packages available.

They can be categorized e.g. in the following way:

- Office tools. The most familiar and used visualization tool is Excel with its bar charts and pie representations.
- Business intelligence tools offering visualizations of the business status and the future for enterprise management, often connected to the company enterprise management system.

- Statistical and mathematical tools. Statistical analysis has a long history of visualizing the results as time series, bar charts, plots and histograms. Examples of tools providing statistical and mathematical visualization are R14 and Mahtlab15 , tools for statistical computing and graphics.
- Visualization-related libraries and software packages. Prefuse16 visualization toolkit for creating rich interactive data visualizations, GGobi17, an open source visualization program for exploring high-dimensional data, XGVis18, an interactive multidimensional scaling (MDS) software.
- Algorithmic tools developed by the research communities based on some algorithm. An example is Graphviz19 for drawing graphs.
- Visual data mining tools. Visual data mining creates visualizations to reveal hidden patterns from data sets. The need of new methods in data analysis has the launched the field. Several products are on the markets, often focused on "Business intelligence" such as marketing, risk analysis, selling analyses and customer management. The field is closely related to visual analytics.
- Web tools and packages. An increasing amount of tools is available in the web, either open source packages for download or on-site use. With the tools users can create more or less fancy visualizations from data. See for example Many Eyes20 , an IBM application for social data analysis.
- Scientific visualization tools for modelling some complicated physical phenomenon.

3. Building blocks of visual analytics

3.1 Information visualization

This chapter represents a selection of topics of information visualization research to be considered in visual analytics: human perception, data graphics and data visualization techniques. The content is mainly based on the information visualization course held in Helsinki University of Technology (TKK) in spring 200721 by Kai Puolamäki, the books about Information Visualization by Colin Ware [Ware 2004] and the fundamental report of visual analytics "Illuminating the path" [Thomas & Cook 2005].

3.1.1 Human Perception

Designing effective visualizations require knowledge about the capabilities and limits of the human information and visual system. Visualizations, often designed by application developers without knowledge about human perceptual and cognitive principles or graphic design can lead to poor or misleading ad hoc solutions. If the information is presented in an inappropriate way it can in the worst case lead to incorrect decisions. On the opposite, good visualizations can improve the efficiency, effectiveness, and capabilities of decision makers and analysts. Human perception research studies the understanding of sensory information. There are several types of perception but the most important from in information visualization is the visual perception. Human visual system is a pattern seeker of enormous power. The human eye and brain form together an efficient parallel processor. On the other hand, the visual system has its limitations. Understanding the rules of the human visual perception helps us to display information effectively. We can present data in such a way that the important and informative patterns stand out. If we disobey these rules, our data will be incomprehensive of misleading.

The study of human perception has advanced enormously over the past decades and great deal of the results is relevant to information visualization. But much of the information is not yet accessible to information visualization designers. There is still a long way to convert the findings to design principles for the everyday use. An excellent source for more information is the book of Information Visualization by Colin Ware [Ware 2004].

Important aspects of human perception for the effectiveness of visualizations and user interaction are:

- Processing visual symbols. Some symbols are understood without learning and processing them is fast (sensory), others are learned and easy to forget (arbitrary). For example letters and numbers are arbitrary symbols, a line connecting two areas is a sensory symbol.
- Human perceptual processing. It proceeds in stages with different rates: starting with rapid parallel processing including extraction of features, orientation, colour, texture and movement, continues with pattern perception, and ends up to slow sequential goal-driven processing.

- Human eye properties including aquities (the ability to see detail), contrast sensitivity, colour vision, perception of shape or motion with colours.
- Visual attention, the process of seeking out visual stimuli and focusing on them. Some visual objects are processed pre-attentively, before conscious attention. These objects seem to "pop out". Bolding text is an example of preattentive features. Othes are line orientation, length, width, co linearity, size, curvature, spatial grouping, added marks, numerosity (up to four), colour (hue, intensity), blur motion (flicker, direction) and, spatial position (2d position, stereo depth, convex/concave shape of shading).
- Pattern perception. Pattern perception is summoned up in Gestalt laws that can be translated directly into design principles. They include similarity, continuation, proximity or nearness, symmetry, closure and relative size. Visual grammars like UML are applications of Gestalt laws.
- Perception of visual objects. There are two complementary approaches of object perception, image-based theory founded on efficient image recognition and structure-based theory suggesting that images are recognised as structures build from three dimensional primitives.

Neither theory can explain all cases of unit recognition.
? Perception of distance and size. There is no general theory of depth perception but several depth cues, principles that can be used to create distance and depth effects, have been recognized including occlusion, perspective, depth of focus, cast shadows, surface shading, surface contours, motion, stereoscopic depth.

? Visual interaction bringing a dialogue between the user and data. Interactive visualizations can be characterized by a feedback loop divided into three phases: 1) data manipulation, 2) view refinement and navigation (exploration and navigation), 3) problem solving loop. Each step has a time scale for the human action reflecting what the user is cognitively and perceptually capable of doing.

3.1.2 Data Graphics

The theory of data graphics introduces principles for efficient visual representations. Many of them are developed by Edward Tufte [Tufte 1982]. Tufte's principles were originally for graphics design but the principles are valid to computer based visualizations as well. They are not an exact theory but more a

collection of rules-ofthumb. Later, cognition research has confirmed the principles. In general, representations should match the task to be performed by the user. The visual representation should provide neither more nor less information than that is needed for the task at hand. Additional information may be distracting and makes the task more difficult. Also, the proportions of the visual representation should match the information being represented.

Attention to data: The main purpose of visualization is to show the data. Data-ink ratio is a measure for the effectiveness of the graphic of showing the data.

data-ink ratio = ink used to represent the data / total ink used to graphics

The larger the share of data-ink is, the more the focus is on data. If the ratio is low, some graphic can be removed without the loss of information.

Avoid chartjunk:hartjunk is the decoration of graphics that does not tell the viewer anything new. The purpose of chartjunk may be to make the graphics appear more scientific or lively (or to give the designer an opportunity to exercise artistic skills). Gridlines, decorations, facejunk, vibrations and redundancy belongs to chartjunk. In the worst cases the design overwhelms the data.

Care with multifunction graphics: Multifunction graphical elements can effectively display complex, multivariate data. A map that shows coordinate data and other properties with shading and colour is an example of multifunction graphics. The complexity of multifunction elements can easily turn data graphics into graphical puzzles. Lie factor. Graphics should be proportional to the numerical quantities. The lie factor is a measure for the proportion.

lie factor = size of effect shown in graphics / size of effect in data.

3.1.3 Information Visualization Techniques

Several visualization techniques have been developed for mapping data to visual presentations. The techniques differ depending on the properties of the data and the purpose of use. The dimensionality of the data, the data structure and the size of data sets are the main factors that determine the appropriate visualization technique. A selection of the huge variety of visualization techniques are

introduced here. A good source for visualization techniques is the book Readings in Information Visualization – Using Vision to Think" [Card et al. 1999].

Line graphs are used for displaying single-valued or continuous functions of one variable. Line graphs are applicable only with few dimensions.
Scatter plot is probably the most popular visual data mining tool. Scatter plots are good at finding outliers and seeing clusters and correlations. Scatter plots are inadequate for higher dimensions.
Bar charts are normally used for presentation purposes. Histograms are bar charts where the value for the bar represents the sum of data points. Multiple bar charts can be used effectively in data mining.
Glyphs are symbols that are used to describe multivariate discrete data. A single glyph corresponds to one sample in a data set. Data values are mapped to the visual properties of the glyph. Glyphs can be constructed so that they are perceived pre-attentively using visual features that "pop out". Examples of glyphs are Chernoff faces where data is mapped to facial expressions, star glyphs, where the dimensions are represented as angular spokes radiating from the center and stick figure icons where data dimensions are mapped to the rotation angles of the limbs. Clustering attempts to show possible clusters in data sets. Several techniques exist, both algorithmic solutions and visual methods.
In pixel oriented visualization techniques each attribute value is represented by one pixel. If the data has a hierarchical structure it can be represented as trees and graphs. There are two major approaches: node link diagrams and enclosure (for example treemaps). The most used are node-link diagrams, including the traditional vertical or horizontal hierarchies, H-trees, radial views, hyperbolic (star) trees and concept maps. There are plenty of drawing algorithms for node-link diagrams optimizing the graph layout.

With large data sets a special problems is how to represent, navigate and find details in them (Focus + context problem). The basic principle by Ben Shneiderman [Schneiderman 1996] is

"Overview first, zoom and filter, details on demand".

Several visualization techniques have been developed based on this principle:

- Ellison techniques. Part of the structure are hidden until they are needed (Furnas' Fisheye view, Cone Tree)
- Distortion techniques. Magnify regions of interest, decrease space to irrelevant regions (Table lens)

- Rapid zooming techniques. User zooms in and out of regions of interest (Pad++)
- Multiple windows. Some windows show overview and others content (Spiral Calendar)
- Micro-macro readings. A good static visualization supports focus.

Other approaches to visualize large data sets are dense layouts and 3D layouts, for instance cone tree and hyperbolic tree (star tree) in 3D and information pyramid.

Often, real world data is multidimensional consisting of many data items or without clean hierarchy. Dimension reduction aims at projecting data into low dimensional space (1D-3D) while maintaining the correct relations between the nodes. There are several methods with different optimization goals and complexities. Among them are :

- Multidimensional scaling (MDS) trying to preserve a measure of similarity (or dissimilarity or distance) between pairs of data points. It can be used as an explanatory visualization technique to find the structure of data and testing hypothesis. It has the roots in psychology .
- Principal component analysis (PCA). It tries to find a linear subspace that has maximal variance. It is based on matrix algebra. It is usually the first dimension reduction method to try.

Other methods to mention: Sammon's projection – a variation of the MDS, pays more attention to short distances, Isometric mapping of data manifolds (ISOMAP) – a graphbased method (of the MDS spirit), Curvilinear component analysis (CCA) – MDS-like method that tries to preserve distances in small neighborhoods, Maximum variance unfolding – maximizes variance with the constraint that the short distances are preserved, Self-organizing map (SOM) – a flexible and scalable method that tries a surface that passes through all data points (developed at TKK), Independent component analysis (ICA) – a fast linear method, suitable for some applications.

3.2 Data mining

Data mining22 has been be defined as “a science of extracting useful information from large data sets or databases”. It shares the goals and techniques with visual analytics and uses much abstract visualization to reveal hidden patterns from data sets. Data mining combines machine learning, artificial intelligence, pattern recognition, statistics, and data base systems..

The data mining is an iterative process starting with selecting the target data from the raw material, pre-processing and transforming it into a suitable form. After that the data is run by the data mining algorithm that creates patterns from the data. The user interpreters and evaluates the results and starts a new iteration with possible modifications on the raw data, algorithm and algorithm parameters. The process is shown in Figure 4

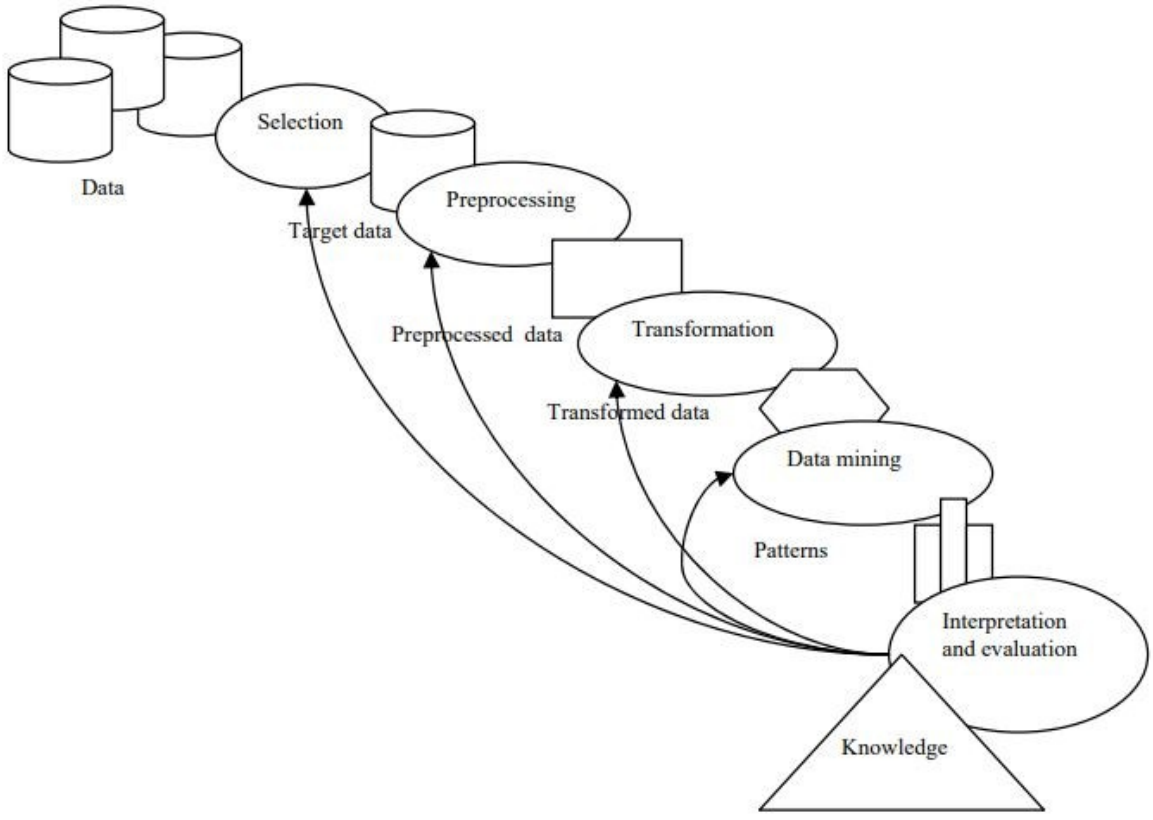


Figure 4. The data mining process (edited from original in [Tan et al. 2006]).

Examples of data sources used in data mining are web data, purchase data, e-commerce transactions, banking data, credit card transactions, sensor data, industry process and maintenance data, satellites, telescopes, gene data and scientific simulations. Data mining uses several data types: data base records, matrix data, documents, graphs, links, transaction data, transaction sequences, sequence data, genomes, spatiotemporal data. The quality of data causes problems. The data can contain noise, there are missing values and duplicate data. Often a data cleaning phase is required before using the data. Also other pre-processing is required such as data aggregation, sampling, dimensionality reduction, subset selection, feature creation, attribute transformation. A new challenging trend in data mining is combining data from diverse data sources.

The methods used in data mining are divided into :

- Prediction methods that use variables to predict unknown but predefined or selected object of interest

- Description methods for finding patterns that describe data in general.

The most important prediction methods are classification and regression.

In classification a model for a class attribute as a function of the values of other attributes (Training set) is created. Then unseen records are assigned to the class. The accuracy of the models is evaluated with a test set. Several techniques are used including decision tree based methods, rule based methods, memory based reasoning, neural networks, naive bayes and bayesian belief networks, support vector machines.

Classification is a much used method and also commercial applications are available. Examples of classification applications are direct marketing, credit card fraud detection, customer analysis and sky survey cataloguing. Regression predicts a value of a continuous variable based on other variables. It uses linear or non linear models.

The area is studied in statistics, and neural networks. It has examples in predicting sales based on advertising expenditure, predicting wind as a function of temperature or humidity and in stock market time series. Description methods

include clustering, association rule discovery, sequential pattern discovery and deviation detection.

In clustering a set of data points, each having a set of attributes and a similarity measure among is created and the algorithm finds clusters with more or less similar data to one another. The field has been an active research topic and lots of (enough) algorithms are available. Application examples of clustering are customer segmentation, document clustering and stock data observation. The goal in association rule discovery is prediction without a selected target. It predicts occurrences of some item based on occurrences of other items (buyers of milk and diapers also by beer) produce dependency rules.

The difficulty with the method is that it produces easily too many rules and it is difficult to find the important ones. Association rule discovery has introduced few successful applications. Some examples are marketing and sales promotion, supermarket self management, inventory management and sales prediction. Sequential pattern discovery concentrates for finding objects that are associated with their own timelines of events. It finds rules that predict strong sequential dependencies among different events. Several algorithms are available. Applications are in telecommunications alarm logs analysis, sales transaction sequences. A promising new application area could be process monitoring from sensor data. Deviation/anomaly detection searches significant exceptions from normal behaviour. The interest is in single observations. Application examples are in credit card fraud detection, network intrusion detection and process anomalies detection.

Data mining has been studied intensively and several algorithms exist. Lots of tools and commercial applications are available, some of which are highly competitive, for instance Customer Relationship Management (CRM). Sensor data and process industry could offer new potential application areas and needs. It has been recognized recently that visualization and interaction are highly beneficial in arriving at optimal results.

The role of information visualization is communicating the results of the automatic analysis. Visualizations represent the found patterns to the user. Another important feature is the ability to explain the patterns either by providing easy ways to explore and summarize data, or to demonstrate the reasoning logic behind.

The visualization, data analysis and user interaction has evolved separately in the past. The recent development has integrated the fields (present). In the future the

fields will integrate to visual analytics solutions (Figure 5). Other challenges lay in the following areas: multiple sources of data, dimensionality, complex and heterogeneous data, data quality, missing values, scalability of the algorithms, data ownership and distribution, privacy preservation and streaming data.

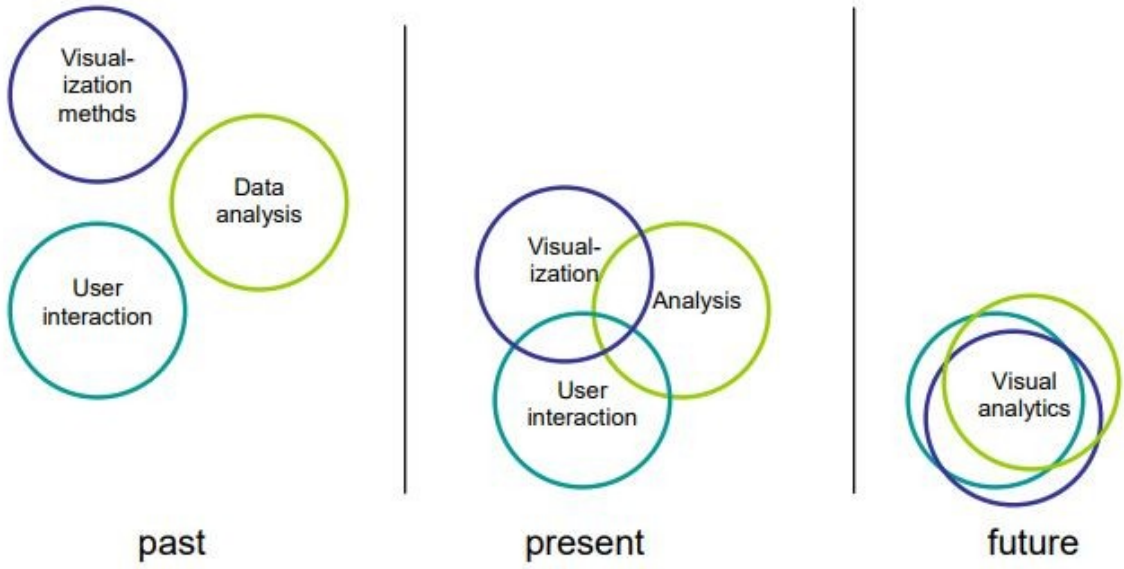


Figure 5. State-of-the art of visual data mining. Based on [Aigner et al. 2008].

3.3 Integrating data sources

Integrating data from different data sources is the foundation on which visual analytics applications are often built. In order to make visualizations, data must be represented in a format suitable for applying the analysis algorithms.

Data used in visualization can be characterized from multiple perspectives [Thomas & Cook 2005]:

- Data type: numeric, non-numeric, or both
- Level of structure: from completely structured, such as categorical data, to completely unstructured, such as narrative description on a web page.

Unstructured doesn’t always mean that there’s no structure, instead it means that the structure is only interpretable by human

- Geospatial characteristics: data are associated with a particular location or region
- Temporal characteristics: data of all types may have a temporal association, and this association may be either discrete or continuous
- Language data: language data can be processed without any acknowledgement of their linguistic structure because meaning is inherent in the communication of the originator
- Image and video data: one of the key challenges for visual analytics is to derive semantic content or meaning from images in real time.
- Traditional relational databases are usually well structured, but the structures and concepts used in the data bases are database-specific and require extensive integration work when used in e.g. visual analytics applications.

However, when applying visual analytics into existing industrial data, the data to be used is normally stored in legacy databases. \

Problems that are faced when integrating legacy databases include the following:

- Syntactic problems: differences in representation format of data
- Structural problems: the native model or structure to store data differs in the databases
- Semantic problems: differences in interpretation of the 'meaning' of data
- System problems: use of different operating system, software versions, or hardware platforms.

Syntactic problems can in many cases be avoided by using standard or de facto standard data formats (e.g. XML or Excel) when exporting applications from legacy databases.

These export formats are normally available on all modern applications, but using them does not solve structural or semantic problems: same information is often represented in different structures or named differently in two legacy systems. These problems are being tackled using semantic ontology based integration [Wache et al. 2001]

In practise, probably the biggest integration problems when trying to apply visual analytics are the problems related to the system problems.

Users of the visual analytics tools should be able to use familiar user interfaces, or at least switch easily between visual analytics tools and legacy applications.

This is not easily achieved by exporting data files between applications, but by direct integration using application programming interfaces.

Integration needs give requirements to the visual analytics tools. Building specific visualisation tools for every use case is not a durable solution in a long run. Instead visual analytics platforms, enabling integration of different data sources, analysis methods and visualizations are needed. Such a platform should enable implementation of reusable integration components for different purposes instead of making point to point integrations. (Figure 6).

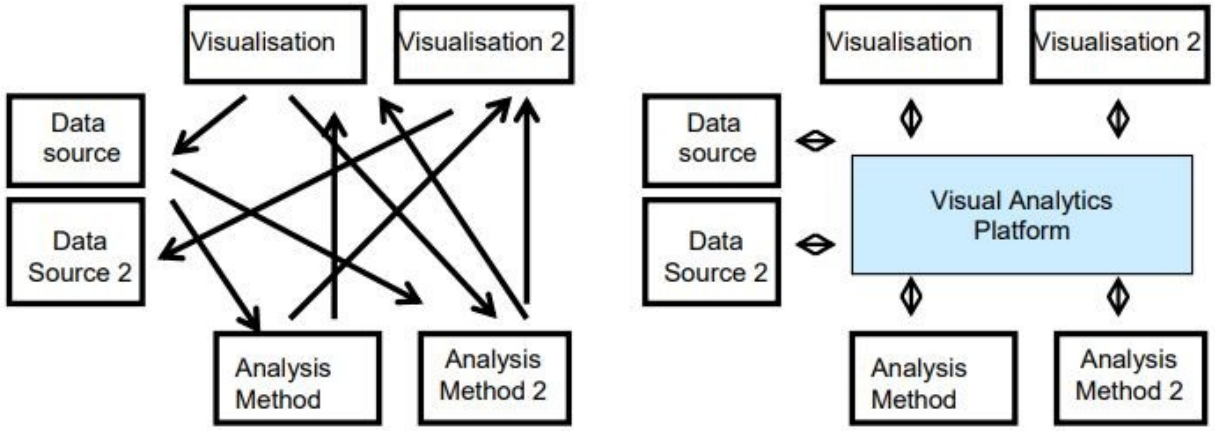


Figure 6. Point to point integrations vs. generic visual analytics platform.

The platform should contain tools for creating software and data modelling components for integrating different types of data sources, via file import or direct application interface programming. It should also include different visualization components, allowing different types of visualization to be used for different analysis methods.

4 . Literature survey

4.1 Man-power productivity ratio(MPR):

The manpower productivity ratio is a measure of how much value a business or an organisation can create with its workforce . You can view man power productivity from two different standpoints. A business views productivity as the degree to which an employee's efforts result in units of production. In other words, man power productivity depends on how much value was created by the employee per hour of his work, either by producing widgets, selling widgets, or providing some sort of service.

Now, let's ratchet this concept up a bit to the general economy. Economists will calculate the man power productivity of an entire country to determine the productivity of its economy. Governments will find this data useful as one metric to compare its relative economic strength to other countries. Of course, we can actually expand this concept to the entire global economy.

Equations for Determining Manpower Productivity:

Manpower Productivity Ratio=Total no. Men on roll/ Total units managed

Here the units are subjective to the branch or the department.

**Typical units are gates,ITKms, units consumed, Div staff strength etc

The Manpower Productivity Ratios (MPRs) of 15 selected activities listed below are presented graphically with a brief analysis as per the data received from Zonal Railways.

- Diesel Loco
- Sheds Electric
- Loco Sheds
- P. Way
- Coaching Depot
- DEMU and EMU Sheds
- Building
- Wagon ROH Depot
- TRD IOW(Bridges)
- Signal
- Elec. Power
- Accounts
- Telecom
- Medical
- Personnel

Benchmarking is based on dynamic and comparative analysis and thus is a very useful tool to manage efficient deployment of staff and monitor effects of improvement in working practices, use of new technologies and level of out sourcing. Board in successive meetings have directed all units to achieve average of the Indian Railway Benchmark. CRB vide DO letter no. E(MPP)2008/1/84 dated 13.012.08 has reiterated the instructions

5.Design Approach

5.1 Introduction to Power BI

Power BI is a collection of software services, apps, and connectors that work together to turn your unrelated sources of data into coherent, visually immersive, and interactive insights. Your data may be an Excel spreadsheet, or a collection of cloud-based and on-premises hybrid data warehouses. Power BI lets you easily connect to your data sources, visualize and discover what's important, and share that with anyone or everyone you want.

Power BI can be simple and fast, capable of creating quick insights from an Excel spreadsheet or a local database. But Power BI is also robust and enterprise-grade, ready for extensive modeling and real-time analytics, as well as custom development. It can be your personal report and visualization tool, and also serve as the analytics and decision engine for group projects, divisions, or entire corporations.

The parts of Power BI:

Power BI consists of:

- A Windows desktop application called Power BI Desktop
- An online SaaS (Software as a Service) service called the Power BI service
- Power BI mobile apps for Windows, iOS, and Android devices

These three elements—Power BI Desktop, the service, and the mobile apps—are designed to let people create, share, and consume business insights in the way that serves them, or their role, most effectively.
A fourth element, Power BI Report Server, allows you to publish Power BI reports to an on-premises report server, after creating them in Power BI Desktop

5.2 Hardware Requirements

Hardware Requirements:-

- intel i3 processor.
- 4 GB Ram
- 512 KB Cache Memory
- Hard disk 100 GB

5.3 Software Requirements

Software Requirements: -

- Operating System : Windows 10
- Softwares: Power BI , MS Excel

5.4 Data under consideration

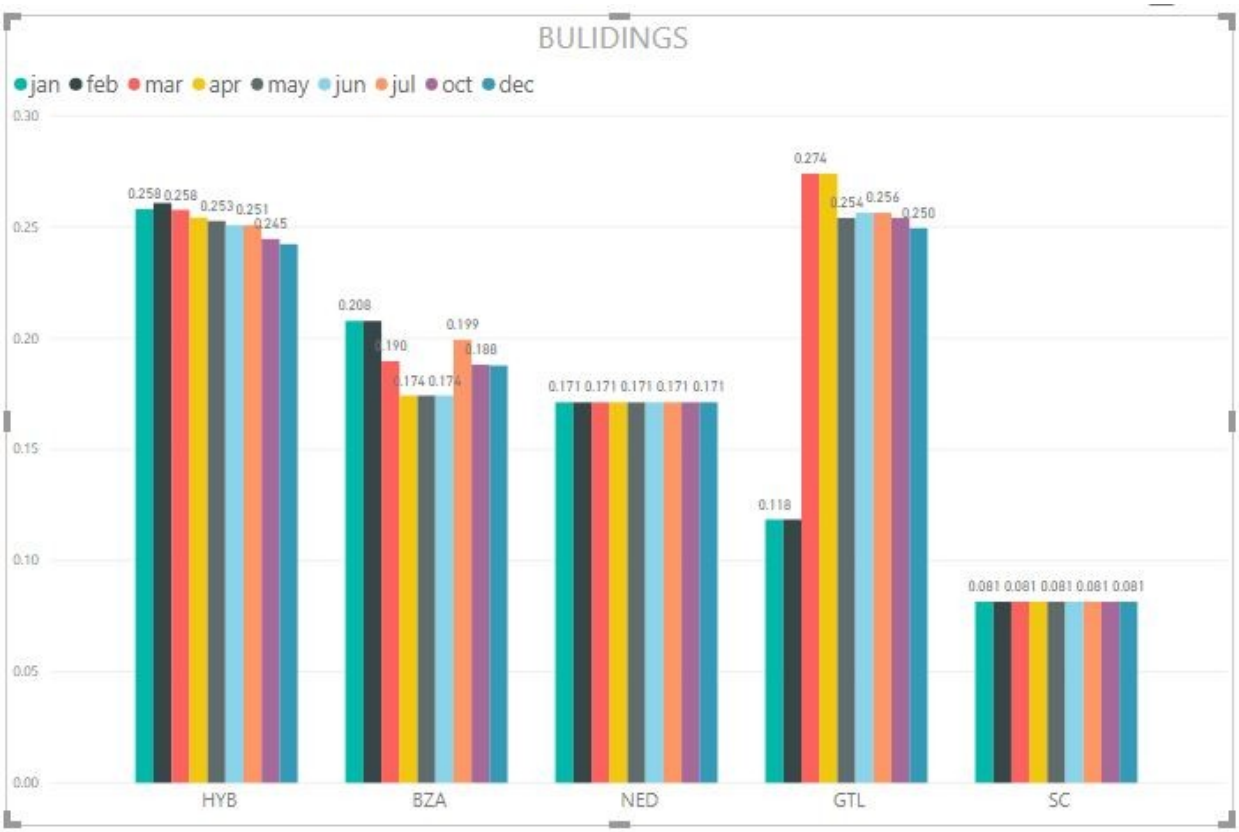
The data under consideration is taken from the senior officials of Railway department.

The data is in the form of excel sheet(.xls).

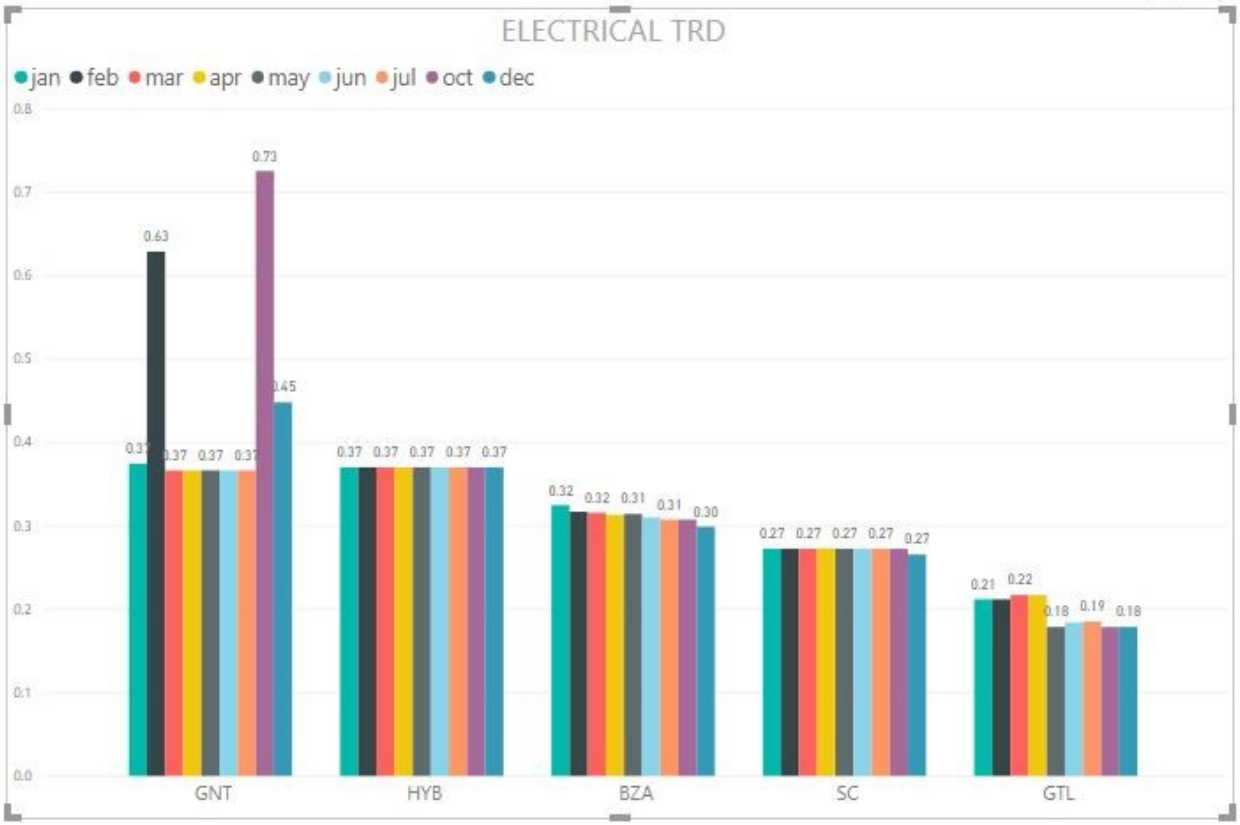
The data consisted of MPR details of different branches in a tabular format.

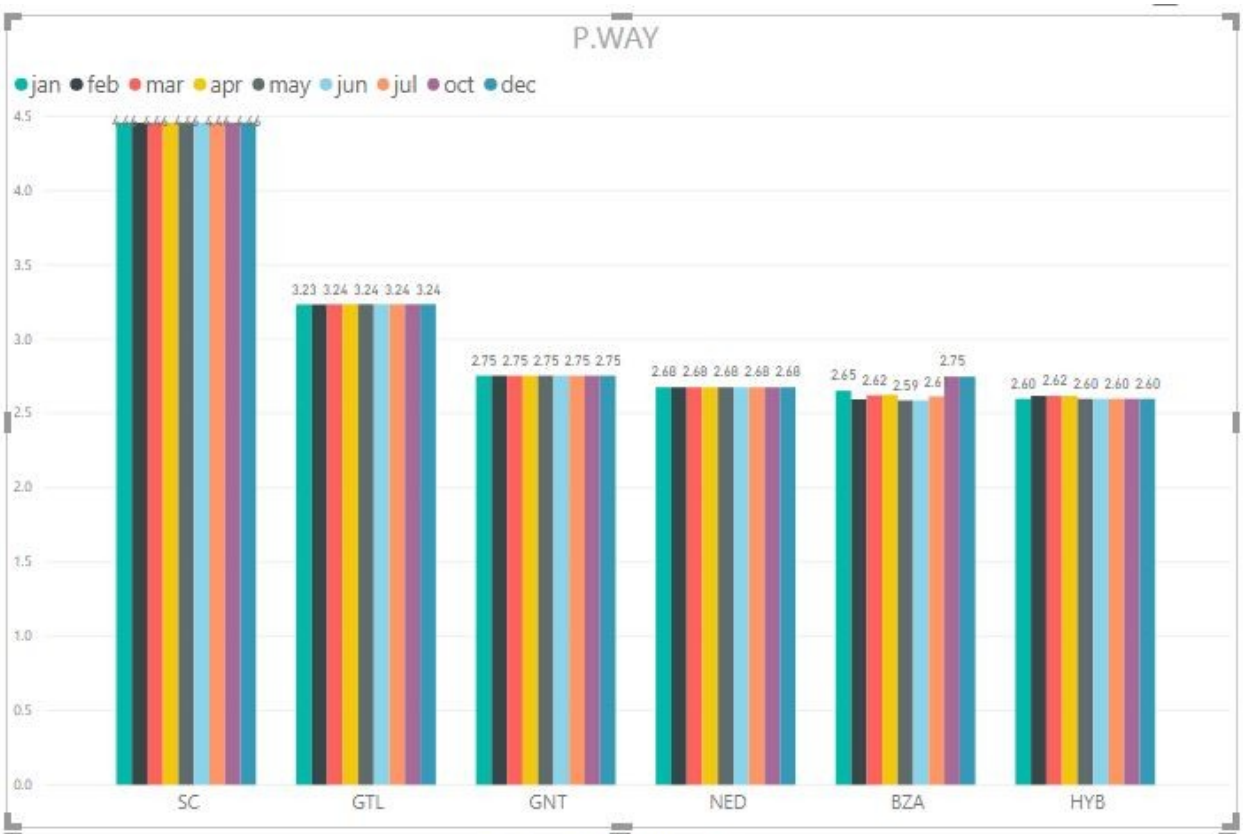
6.Analysis and visualizations

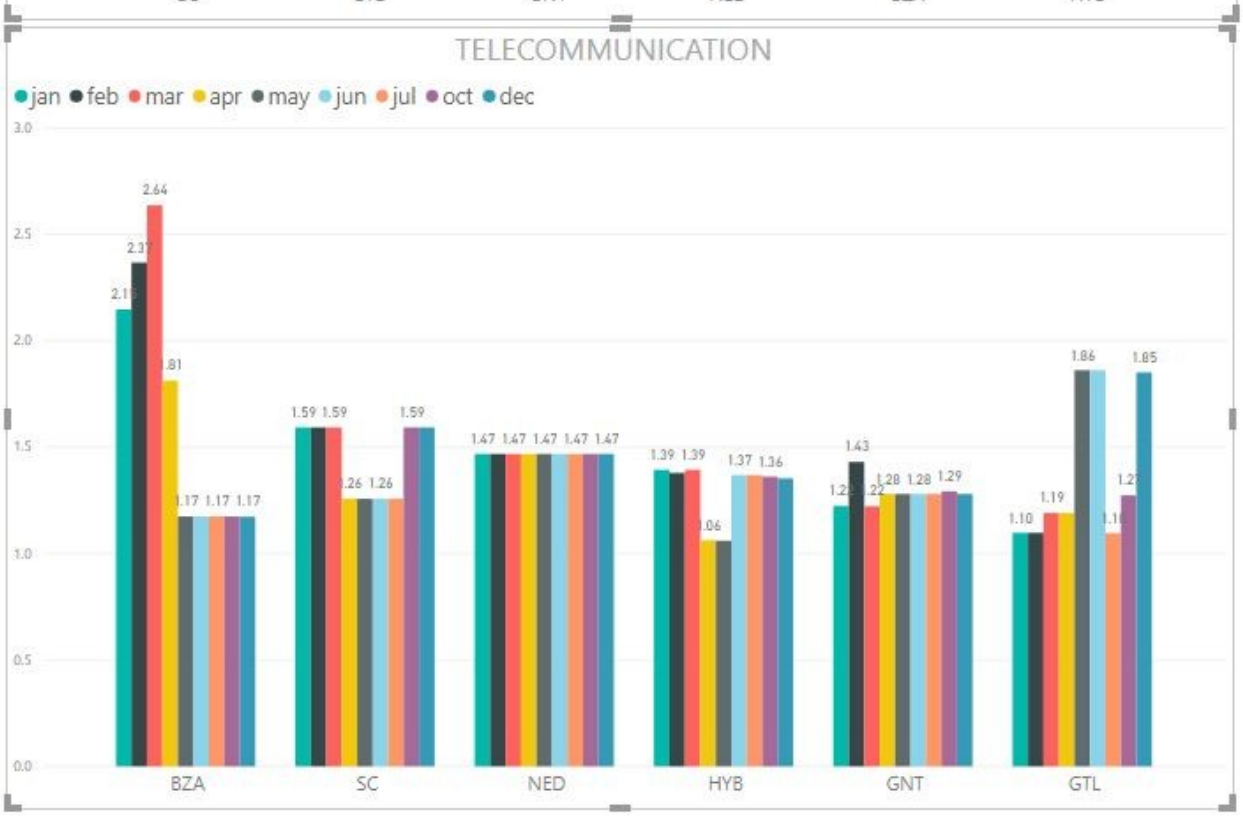
6.1 Visualizations for the year 2017:



DIESEL LOCO SHED
jan feb mar apr may jun jul oct dec
7
6
5
4
3
2
1
0
4.4 4.4 4.4 4.4 4.4 4.4 4.6 4.4 5.6
4.3 4.3 4.3 4.3 4.3 4.3 4.3 4.3 4.3
3.3 3.3 3.3 3.3 3.0 3.2 3.1 3.0 3.2
3.2 3.2 3.2 3.3 3.2 3.2 3.2 3.3 6.2
GTL
KZJ
GY
MLY

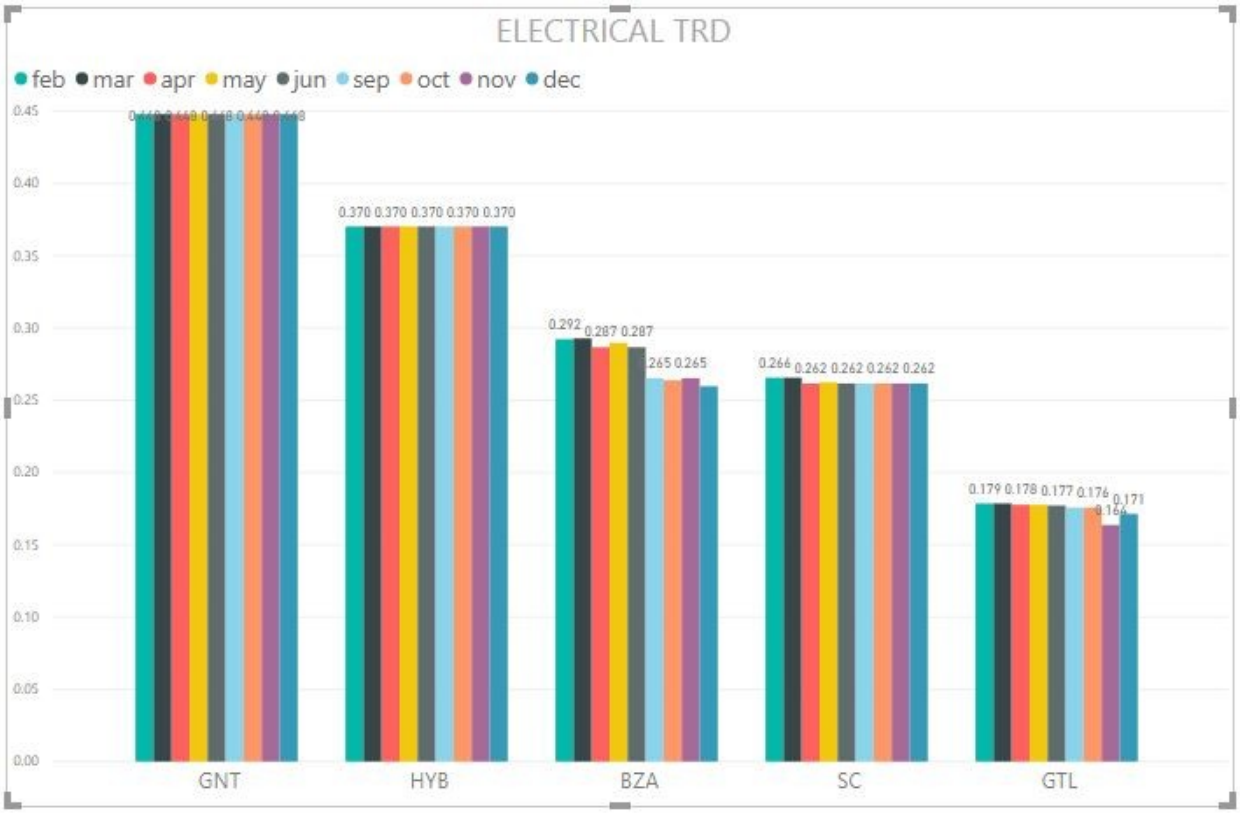

ELECTRICAL TRD
jan feb mar apr may jun jul oct dec
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
0.37 0.63 0.37 0.37 0.37 0.73 0.45
0.37 0.37 0.37 0.37 0.37
0.32 0.32 0.31 0.31 0.30
0.27 0.27 0.27 0.27 0.27
0.21 0.22 0.18 0.19 0.18
GNT
HYB
BZA
SC
GTL

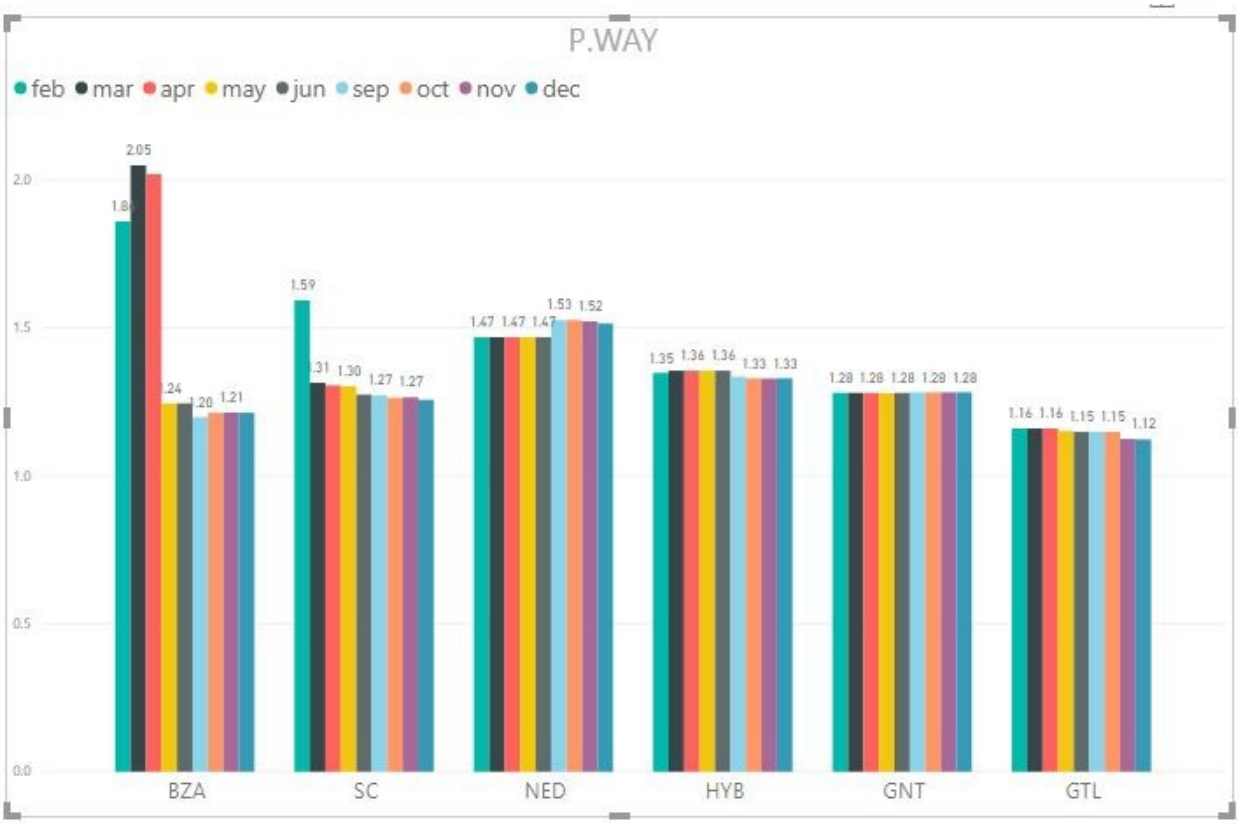

P.WAY
jan feb mar apr may jun jul oct dec
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
3.23 3.24 3.24 3.24 3.24
2.75 2.75 2.75 2.75 2.75
2.68 2.68 2.68 2.68 2.68
2.65 2.62 2.59 2.6 2.75
2.60 2.62 2.60 2.60 2.60
SC
GTL
GNT
NED
BZA
HYB

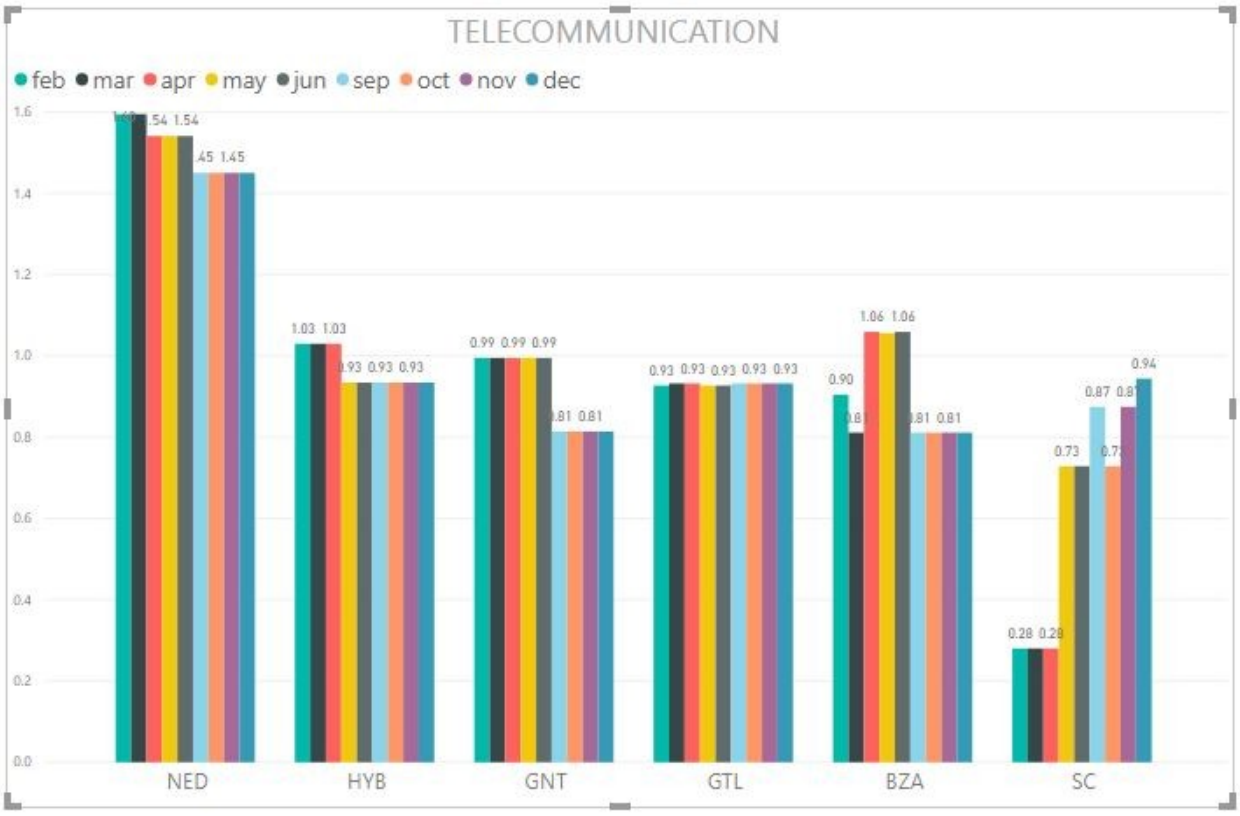

TELECOMMUNICATION
jan feb mar apr may jun jul oct dec
3.0
2.5
2.0
1.5
1.0
0.5
0.0
2.64
1.81
1.17 1.17 1.17
1.59 1.59
1.26 1.26
1.59
1.47 1.47 1.47 1.47 1.47
1.39 1.39
1.06
1.37 1.36
1.43
1.28 1.28 1.29
1.10
1.19
1.86
1.85
BZA
SC
NED
HYB
GNT
GTL

6.2 Visualizations for the year 2018:

DIESEL LOCO
feb mar apr may jun sep oct nov dec
5
4
3
2
1
0
4.4 4.4 4.4 4.6 4.5 4.5 4.5 4.7 4.5 4.1
4.3 4.3 3.8 3.8 3.8 3.8 3.8 3.8 3.8 3.8
4.0 3.9 3.9 3.9 3.9 3.9 3.8 3.8 4.0 4.0
3.1 3.6 3.0 3.0 3.0 3.6 3.6 3.5 3.5
3.0 3.0 3.0 3.0 3.0 2.9 2.9 2.9 2.7
GTL
KZJ
MLY
BZA
GY


ELECTRICAL TRD
feb mar apr may jun sep oct nov dec
0.45
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
0.370 0.370 0.370 0.370 0.370
0.292 0.287 0.287 0.265 0.265
0.266 0.262 0.262 0.262 0.262
0.179 0.178 0.177 0.176 0.164 0.171
GNT
HYB
BZA
SC
GTL


P.WAY
feb mar apr may jun sep oct nov dec
2.05
1.59
1.47 1.47 1.47 1.53 1.52
1.35 1.36 1.36 1.33 1.33
1.28 1.28 1.28 1.28 1.28
1.16 1.16 1.15 1.15 1.12
1.21
BZA
SC
NED
HYB
GNT
GTL


TELECOMMUNICATION
feb mar apr may jun sep oct nov dec
1.54 1.54
1.45 1.45
1.03 1.03
0.93 0.93 0.93
0.99 0.99 0.99
0.81 0.81
0.93 0.93 0.93 0.93 0.93
0.90
1.06 1.06
0.81 0.81
0.28 0.28
0.73
0.87
0.94
NED
HYB
GNT
GTL
BZA
SC

6.3 Inferences from the data

The following inferences are made from the visualizations of the data-ink:

1 . There is a steep decrease in MPR at P.Way department at BZA after april which continues till december for the year 2018.

2 . There is no drastic change in the MPR at Electric trd during the year 2018 across all branches.

3 . There is a steady increase in MPR in telecommunication department at SC during the year 2018

4 . There is an abrupt downfall of MPR at the NED region in the department of Buildings after 1st half of the year in 2018.

5 . There is a sudden onset of MPR in the months of feb and oct in the electric trd department at GNT in 2018.

6 . There is an abrupt onset of MPR at the GTL region in the department of Buildings after feb in 2017.

7 . There is an sudden uprise of MPR at the MLY region in the department of Diesel loco in december in 2017.

8. There is no drastic change in the MPR at P.Way during the year 2017 across all branches

9 . There is a steep decrease in MPR at Telecommunication department at BZA after april which continues till december for the year 2017.

10 . There is an abrupt onset of MPR at the GNT region in the department of Electric trd in feb and oct in 2017.

7. Conclusion

The package was designed in such a way that future modifications can be done easily. The following conclusions can be deduced from the development of the project.

- Visualization of the entire data improves the efficiency.
- It provides a friendly graphical user interface which proves to be better when compared to the existing system.
- It gives accurate and appropriate information to the users.
- It effectively overcomes the delay in communications.
- Updating of information becomes so easier on Power Bi module.
- The System has adequate scope for modification in future if it is necessary.
- Different visualizations like pie charts and histograms can also be used instead of Bar Charts.

THANK YOU