*International Journal of Latest Trends in Engineering and Technology*
*Vol.(8)Issue(3), pp.051-058*
*DOI: http://dx.doi.org/10.21172/1.83.007*
*e-ISSN:2278-621X*

# AN EFFICIENT APPROACH FOR DATA VISUALIZATION USING DATA MINING

S.Sai Rama[1],G.Mownica[2], K.Ramanjaneyulu[3],M.Prasad[4],K.Madhavi[5]

**Abstract— Every successful enterprise uses data and make use of scientific tools to make business decisions.One of the most recent weapon for decision making,marketing effectiveness,operational efficiency is the advanced analytics.With the assistance of analytics,we can foresee,know and affect the customer behavior.We are first in analytics and thus has provided various services to perform the functions of a business.Our customers will benefit through the skill in visualizing the data as well as performing pattern mining on big business and outsider information.**
*Keywords—Data,Analytics,Patterns;*

## I. INTRODUCTION

Real time data is the data that is used by many organizations.So in order to have easy understanding of the data even without the knowledge about the data visualization is used.In static visualization the data and their corresponding queries are fixed while in dynamic visualization the data is not fixed and the queries can be written accordingly when required.

### *A.Data Visualization*

The graphical representation of data is data visualization,which is used to provide the customer to understand the information content.Generally bargraphs,line chart,pie charts and maps have been in use for decades.Informationgraphics,ScientificVisualization are closely related to data visualization.

### *B.Why we Visualize Data*

To meet a very basic need we visualize information which is nothing but telling a story.It has its origin in the early 30,000 B.C and is one of the most primitive forms of communication.The single most faculty to communicate information is vision.To communicate more information than a table in much smaller space we use visual.Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space said by a data visualization expert,Edward Tufte.The main goals of data visualization are the Exploratory and Explanatory,sometimes they can overlap.To visualize the data we have used Fusion Charts.With the help of it we can the chart in 15minutes and can add additional features further when required.

### *C.Data Analysis*

The process of viewing and giving a brief description of the main points in the data with a goal to get the useful information and also to provide conclusions. It has a close relation with data mining,but data mining generally focus on larger data setsfor effective building of computation on cube in the business trending data analysis has a crucial role.There are types in data analysis like Exploratory data analysis and Qualitative data analysis.

[1] *Information Technology,V R Siddhartha Engineering college,India*
[2] *Information Technology,V R Siddhartha Engineering college,India*
[3] *Information Technology,V R Siddhartha Engineering college,India*
[4] *Information Technology,V R Siddhartha Engineering college,India*
[5] *Information Technology,V R Siddhartha Engineering college,India*

D.*Data Governance*

Incorporating all the requirements like the people,processes,innovation required to make a standard, enterprise view of an organization's data is data governance.It helps to increase the consistency and the confidence in decision making.Data security can be improved,the range of regulatory fines can be decreased,the pay era capability of the data can be increased.

*E.Data Management*

The development and execution of policies,practices and procedures,architectures that properly manage the full data lifecycle needs of an enterprise is data management.This definition is generally related to the professions that may not have direct contact technically with the low-level aspects of data management such as RDBMS.It is very essential for the higher authorities in the business.

*F.Data Mining*

The way toward sorting through a lot of data and taking up just the pertinent data is characterized as data mining. Data mining can be utilized as a part of different fields like business knowledge organizations financial examination and once in a while utilizes as a part of sciences to get the helpful data from a lot of datasets created by the modern experiments and observational methods. Powerful PCs are utilized to filter through huge volumes of general store scanner data and analyze market research reports for a considerable length of time in companies.Continuous developments have drastically expanding the exactness of analysis by driving down the cost.

*G.Knowledge Discovery*

The procedure of naturally scanning tremendous volumes of data for the purpose of patterns which is considered as knowledge about the data.Knowledge discovery is frequently named as getting knowledge from the data givenThere are mostly two perspectives that are considered.One among them is the thing that what sort of data is looked and the other is in what frame the aftereffect of the search is represented. Knowledge Discovery is one of the well-known branch of data mining also termed as Knowledge discovery in data bases(KDD).Sometimes the knowledge obtained through the procedure may turn into extra data that can be utilized for further use and discovery.Some of the areas where this can be used are software modernization,weakness discovery.

*H.Applications of DataMining*

- Financial Data Analysis
- Retail Industry
- Telecommunication Industry
- Intrusion Detection
- Other Scientific Applications

**II.EXISTING SYSTEM**

Existing system is the Online Transactions System where it doesn't bolster operations on enormous amount of data as the time of retrieval of the data is slow,additionally the data couldn't be appeared on the dash-boardalso the decision making cannot be made.The purpose of data is to control and run fundamental business task and it reveals the snapshot of ongoing of ongoing processes of a business.Simple queries like insert,delete,Update are performed.

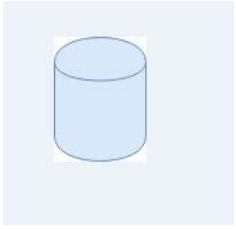

Fig. 1.Online Transaction Processing

*A.Online Transaction System*

Day-Day transactions are being dealt by databases or Online Transaction Systems. Generally databases maintain integrity and the data can be accessed fastly. Normalized data is present in databases.

*B.Drawback*

Only predined reports are visualized.

**III.PROPOSED SYSTEM**

Online Analytical Processing is the proposed system for Online Transactions System where it bolsters taking decisions and also the information can be viewed Statiscally.

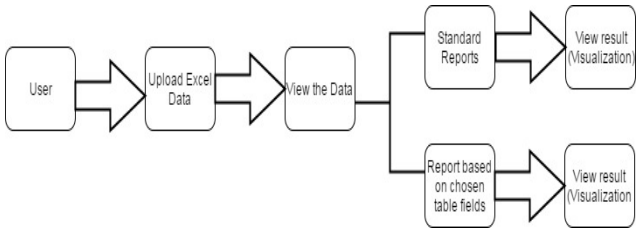


Fig.2.Online Analytical Processing

*A.OLAP*

Historical data is being dealt by OLAP for analysis and basing on that decisions can be taken.The data can be in any format like normalized or de-normalized.

*B.Dataset*

*Supermarket Data:* Consists of data with twenty three attributeslikeorderdate,orderpriority,orderquantity,sales,discount,unitcost,shipmode,profit,shippingcost,customer name,province,region,customersegment,product category,product sub category, product container,product base margin,Ship date.

*D.Hardware and software specifications*

Processor : INTEL P4 PENTIUM 1.8Ghz
Size of RAM :1 GB RAM
Size of hard Disk :80 GB HD
Operating System : Windows XP
Languages : Java
Development Kit :JDK 1.6.0
Front End : Jsp
Database :MY SQL

*E.Algorithm*

The task of finding frequent patterns in large databases is very important in data mining and has been studiedin large scale in the past few years.When a large number of patterns exist,the computational work is expensive,so Han proposed FP-Growth algorithm which is efficient and scalable method for mining the complete set of frequent patterns.It is also the alternative way to find the frequent itemsets without using candidate key,thus improving performance.Generally divide-and-conquer strategy and a special data structure named frequent-pattern tree is used which retains the itemset association information.

*F.Advantages*

- We can visualize any type of information dynamically.
- Reports are generated with user friendly environment.
- We can provide hassle free UI for users to visualize their data.
- We can also find the frequent patterns in the data set given.

IV.RESULTS & OBSERVATIONS

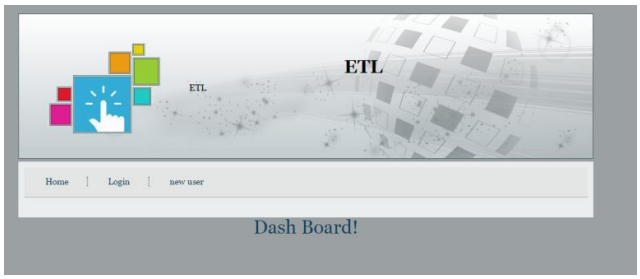


Fig.3 Webpage

Once we run the executable file the above webpage will be opened.It consists of Home,Login,Register.If the user is new then register option will be chosen otherwise the login option.

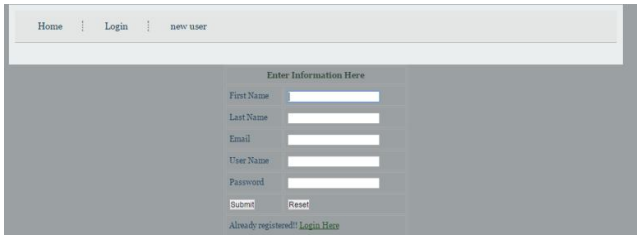


Fig.4.User Registration form

It consists of various fields to be filled by the user like firstname,lastname,email,username and password.Once the details are filled the user clicks the submit button.


Fig.5.Directing to Login

Once the submit button is clicked the user is directed to login page via Go To Login option.

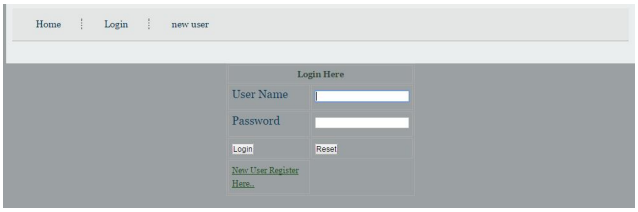


Fig.6.Login page

When the user clicks the go to login option then the login form will be appeared.The required details like username and password should be given and if they both matches with the data present in the database,the user login is said to be done successfully.Also, if the user is already registered then the user chooses the login option

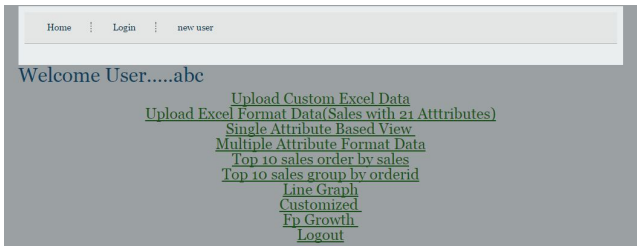


Fig.7.After Successful Login

When the user login is said to be done successfully ,the webpage with several options like Upload Custom Excel data,Upload Excel Format Data (Sales with 21 attributes),Single attribute view,Multiple attribute view,Top 10 sales orderby sales,Top 10 sales group by orderid,Line graph, Customized,F-p growth will be appeared.


Fig.8.Upload Excel format data(sales)

When the user chooses to upload the sales data then Upload Excel format data(sales with 21 attributes) will be chosen.By clicking an option select your file or upload excel format data the data can be inserted.After the data has been insertedthe required option can be chosen based on the requirement.

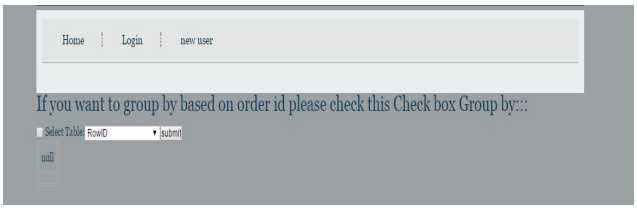


Fig.9.Single attribute based view

When the user chooses for single attribute based view the above page will be appeared.The user has a privilege to choose any of the 21 attributes from the drop down box where the row id will be given as default.The user can select the check box provided when wants to group by based on order id.


Fig.10.selection of attribute and display

If the user selects the order-id for single attribute view then only the order id's in the data uploaded will be displayed.

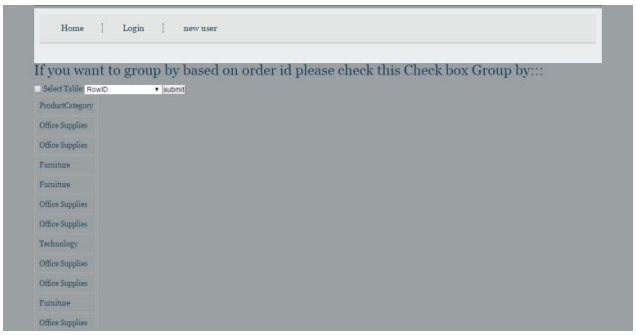

Fig.11.selection of attribute and display

When the user chooses product category then that respective details will be appeared.


Fig.12.selection of multiple attributes

When the user chooses the multiple attribute view then all the attributes will be displayed.Checkboxes are provided tochoose the required attributes so that the view can be given based on the selected attributes.


Fig.13.Multiple attribute view

The above figure is the view based on the selected attributes order-date,Order-id,profit,discount,order-category,order-subcategory, Product-name,ship-date.

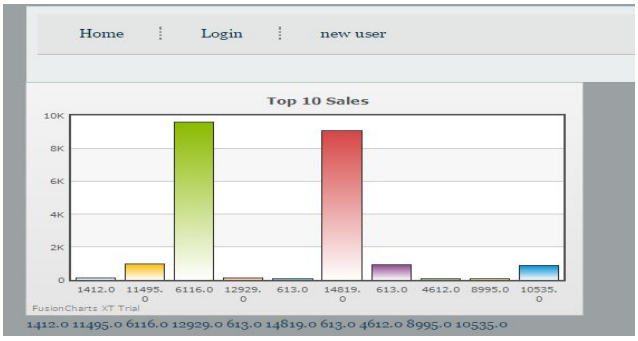


Fig.14.Bar Graph

When the user selects Top 10 sales order by sales then the above graph appears.On horizontal axis order-id is plotted and on vertical axis the sales are plotted.

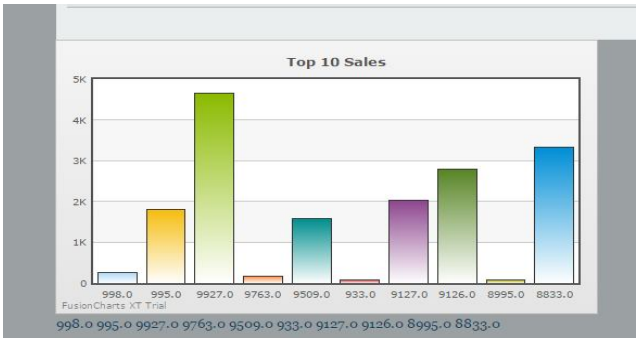


Fig.15.Bar Graph

When the user selects Top 10 sales group by orderid then the above graph appears which means the group by function gets executed and aggregates the sales when the order id repeats and then display on the graph.

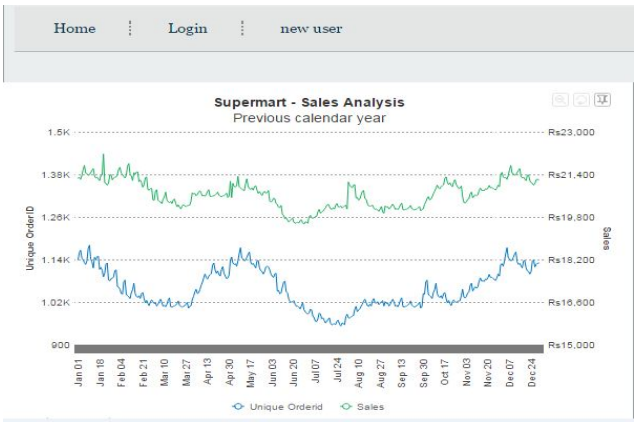


Fig.16.Line chart

The line chart gives the visualization for a calendar year having sales on the horizontal axis and their unique id's on the vertical axis.The line chart is useful for the supermarket sales analysis.


Fig.17.Frequent patterns

The above gives the frequent patterns for the data given.

- Our project not only supports the sales data but any excel file can be uploaded using custom excel file on the webpage.The data uploaded can be given single and multipleattribute view as well as we can also visualize the data using customized on the webpage.

**V.CONCLUSION**

RDBMS have a problem that the data is only represented in the form of tables.If the data want to be visualized pictorial representations have to be used.In excel the dashboard concept is not available.Based on the representations the customer can easily visualize their business and can make changes accordingly to improve their business.The data can be visualized according to the needs the customer like daily,monthly,yearly,half year,quarterly.Based on the visualization performance and decision making can be done efficiently to improve the business.

## VI.FUTURE WORK

Day by day the data is growing enormously and the storage becomes difficult.So one of the solution for this is using hadoop.

## REFERENCES

[1] C. Stolte et al. Polaris: a system for query, analysis, and visualization of multidimensional databases. IEEETVCG, 2002.

[2] V. Friedman, "Data Visualization: Modern Approaches," Internet: http://www.smashingmagazine.com/2007/08/02/data-visualization-modern-approaches, Aug. 2, 2007[Mar. 12, 2012].

[3] D. Keim "Visual Techniques for Exploring Databases," International Conference on Knowledge Discovery in Databases (KDD '97), California, USA, August 1997.

[4] J. Han, H. Pei, and Y. Yin. Mining Frequent Patterns without Candidate Generation. In: Proc. Conf. on the Management of Data (SIGMOD'00, Dallas, TX). ACM Press, New York, NY, USA 2000.

[5] *Judd, Charles and, McCleland, Gary (1989).* Data Analysis. *Harcourt Brace Jovanovich.* ISBN 0-15-516765-0.

.