International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 01, 2021

Automated Visual Assessment From Optical Data Sets To Enhance The Accuracy Of Data Analysis

Dr.Syed Khasim[1],Dr.Neelamadhab Padhy[2],Dr. Sarita V Balshetwar[3], Dr. G. Sivakumar[4]
Dr.Shaik Shakeer Basha[5],Jeipratha P N[6]

[1]Professor, Department of Computer Science and Engineering, Dr.Samuel George Institute of Engineering & Technology, Markapur, Prakasam Dt, Andhra Pradesh, 523316.
[2]Associate Professor, Department of Computer Science and Engineering, GIET University, Gunupur, Odisha 765022.
[3]Associate Professor, Department of Computer Science and Engineering, AGTIs DACOE, Karad, Satara, Maharashtra-415002.
[4]Professor, Department of Computer Science and Engineering, Erode Sengunthar Engineering College,Erode, Tamilnadu- 638057, India
[5]Assistant Professor, Department of Computer Science and Engineering, Avanthi Institute of Engineering and Technology, Gunthapally, Abdullahpurmet Mandal-501512, Hyderabad, Telangana.
[6]Assistant Professor, Department of Computer Science and Engineering,St.Joseph's College of Engineering, OMR Chennai 600119.

Abstract: Data visualization is a technique for extracting information from massive amounts of data. For their assumption, software engineers constantly generate multiple visualizations from datasets. Evaluating databases with a large number of characteristics may be time-consuming and error-prone. The objective of this study is to use optimal datasets from several sources to automatically propose attractive visualization patterns. It helps you save time by reducing the amount of time you spend on low-value visualizations and displaying suggested pattern.

Keywords: Query optimizer; Big Data Analyst; Data science; Visualization; Data analyst

1. INTRODUCTION

Data scientists and analysts are increasingly using data visualization technologies. They load alternative datasets and use visualization tools to test their theory; this step is followed numerous times until they discover an obvious pattern. Data scientists must use this time-consuming trial and error technique to gain insights. The primary objective of this study is to discover intriguing patterns in huge datasets from various sources [1].
In identify diverse patterns and abnormalities, data scientists must create various representations from fresh datasets. Finding numerous patterns in a dataset with a large dimensionality becomes a time-consuming process. For data analysis, connections between characteristics and their subsets must be determined. If the mapped information differs significantly from the reference points or historical information, important observations seem to be likely to occur. The variety of visualizations that may be created is enormous, even for a

tiny dataset. The visualizations should be shown at an engaging speed, with a faster reaction speed to the consumers.
Data is now stored in a variety of databases with storage models tailored to their specific requirements. When information needs to be evaluated, it must be retrieved from a variety of sources. To benefit from the performance advantages of native systems particularly built to handle them, organized data is saved in relational databases while unorganized information is stored in NoSQL relational databases.
MIMIC-III is a database that stores information about clients who have been admitted to hospitals. It includes vital sign information, medical equipment data, doctor comments, and patient admission information. After the patient information has been de-identified, the data is made available to researchers [2]. To create a graphical suggestion tool, data federation across these many databases is necessary. This would make it easier for data scientists and analysts to test their hypotheses using various datasets. Currently, this procedure is manual, the user is required to collect relevant data and then walk through all of the visualizations, which is a time-consuming effort [3].
De-identified patient data obtained from various sources can be displayed in graphs. It contains information such as the patient's admittance date, identity, physician remarks, and medical equipment time - series. This becomes complicated to get rid of all the visualizations created. Users are drawn to visualizations in which the training set differs significantly from the related data. They can concentrate on their duties since they have a federation SQL architecture that fetches data rapidly across systems [4,5].

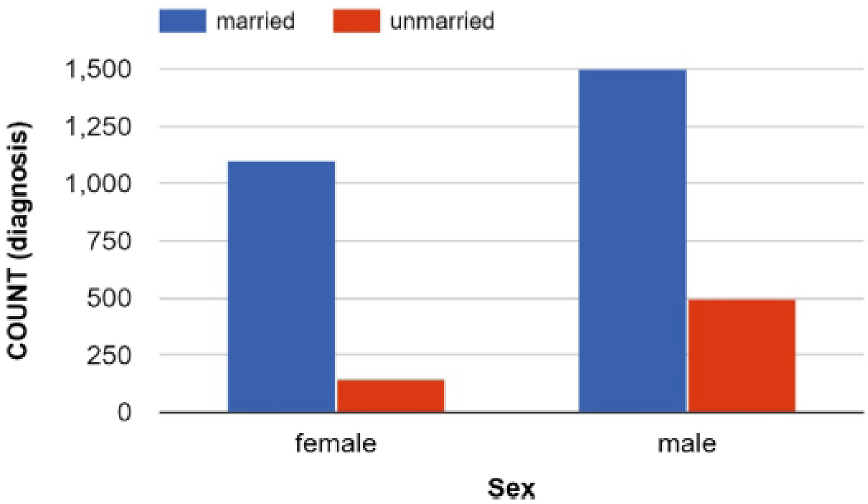

Figure 1 depicts the chart for hospitalized married and unmarried patients with cardiac issues. Figure 1 demonstrates how deviating target data from standard points aids in finding anomalies and abnormalities in the data, which can then be investigated subsequently.

2. RELATED WORKS

Data analysts examine information acquired from a variety of sources. The facts are represented by functional aspects, while the observed characteristics are obtained using conditional formatting. These two characteristics are utilized in visualization software [6]. This review concentrates on SeeDB [7], which suggests visualizations for a query based on a high utilitarian calculus that shows greater variances by collecting data from several

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 01, 2021

databases. Healthcare and medical equipment produce data in a variety of forms, which must be kept in several systems [8].

The customized SQL architecture searches data from any of the identified databases, and the classification process, where applicable, replace optimal data and gets the data fast, which is then utilized to construct the needed suggested visualization. The query's group-by attributes are represented as dimensional attributes D [9-11]. The measurement characteristics M and a collection of accumulation functions A are used to quantify the dimensional attributes. These searches are run against a collection of S databases that have been registered. Dimension characteristics D can be grouped and aggregated depending on measurement characteristics M.

This creates a two-dimensional table that may be used to visualize the data. Suggested

Q (target) = SELECT d; a (m) FROM T(S) GROUP BY d

Q (reference) = SELECT d; a (m) FROM R(S) GROUP BY d

Utilitarian manufacturing is derived from the perspectives of Q (goal) and Q (source) (reference). The discrepancy in views is used to determine which visualizations should be presented (see Figs. 2 and 3).

```
SELECT sex, count (diagnosis) FROM admission_married GROUP BY sex;
SELECT sex, count (diagnosis) FROM admission_unmarried GROUP BY sex;
```

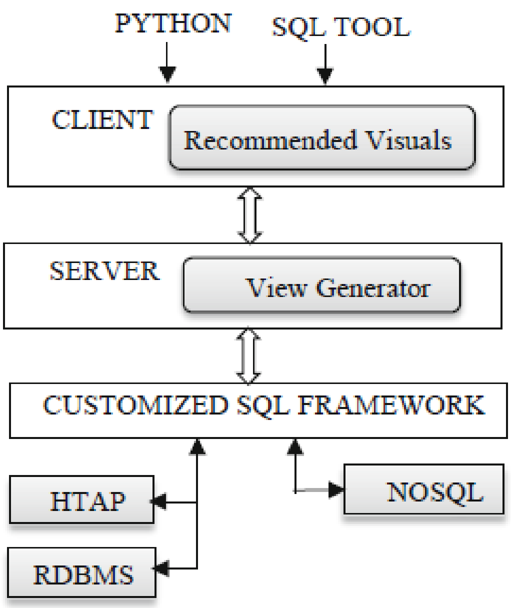

Figure 2 shows the visual suggestion application's architecture.

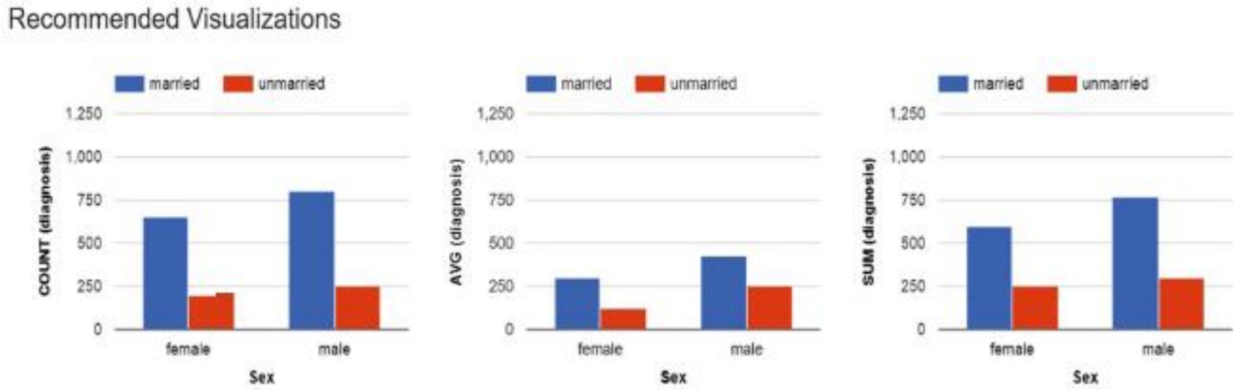

Fig. 3. Visualization chart

3. PROPOSED SYSTEMS

To get the data, it interacts with a bespoke SQL architecture that functions as a federated query layer. The technology revises the inquiry to the optimized copy and returns the result set at runtime. The relational node is transformed by the query optimizer by replacing it with whole or partial rules that fit the description. The various registered data is kept in a catalog, which the scheduling algorithm uses during runtime. It gives insights into the query's total execution price, the data size in tables, and CPU and Memory consumption.

Several group-by are merged when aggregate queries with the same group-by characteristics are aggregated into a single display. This leads to a reduction in query latency and performance improvement.

4. EVALUATION

The visualizations with the highest usefulness factor amongst these top perspectives, as well as improved precision and faster response times, were evaluated and shown in Figure 4.

International Journal of Aquatic Science
ISSN: 2008-8019
Vol 12, Issue 01, 2021

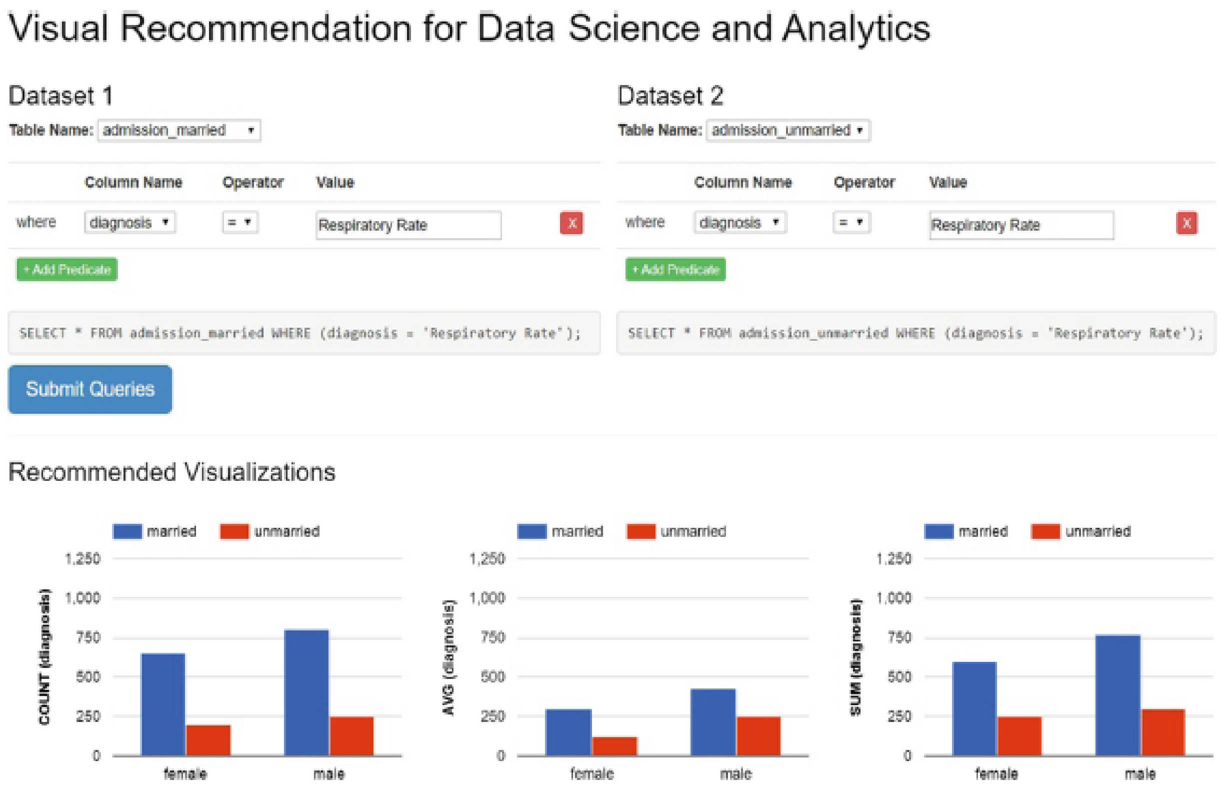

Fig. 4: Presentation of suggested visualizations for hospitalized treatment of respiratory problems.

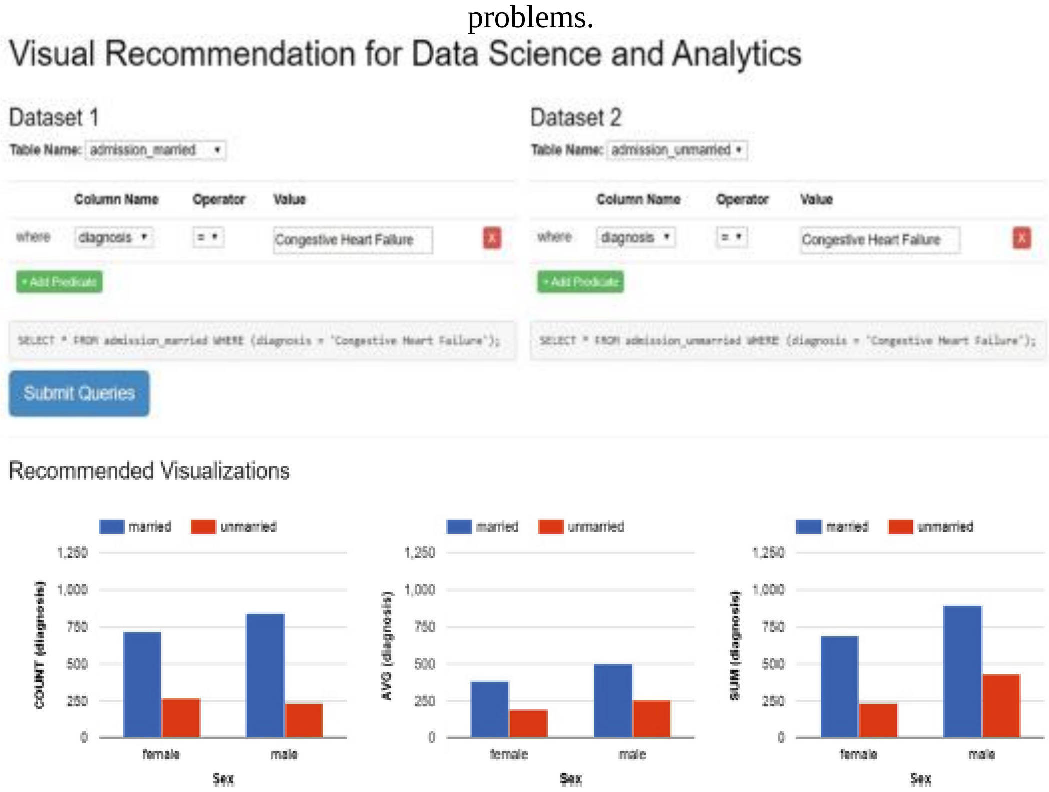

Fig. 5: Presentation of suggested visualizations for patients with heart-related problems.

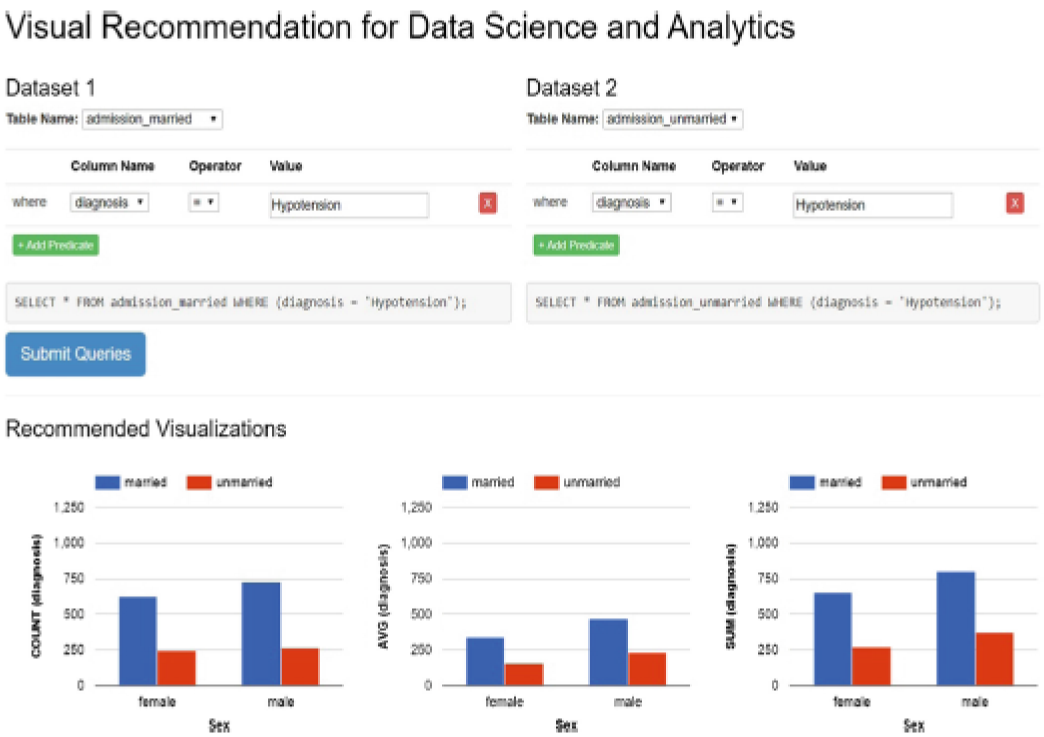

Fig. 6: Display of recommended visualizations for patients with hypotension-related problems.

All tests were run on a 32-bit Linux system with a 3.25 GHz Intel Xeon CPU and 8 GB of RAM. To store patient-related data, PostgresSQL was utilized. Splice Machine was used to store medical equipment data, while MongoDB was used to store text notes. The objective data is dataset1 and the comparison dataset is dataset2, as illustrated in Figs. 4, 5, and 6. The suggested visualizations are those that have a high usefulness factor and deviate from the reference points a most.

5. CONCLUSION

This research uses a visual analytical model in conjunction with an optimization to suggest intriguing visualizations from a variety of datasets on its own. This effort aids data scientists and analysts in their interactive information collection. Integration using data centers will be a significant expansion of this research.

6. REFERENCES

[1] Vartak, M., Madden, S., Parameswaran, A., Polyzotis, N.: SeeDB: automatically generating query visualizations. Proc. VLDB Endow. 7(13), 1581–1584 (2014)
[2] Dr.P.Sivakumar. “Trailing Mobile Sinks to Data Coverage Protocol for Wireless Sensor Networks”, International Journal of Innovations & Advancement in Computer Science (IJIACS), Vol.5, Issue No: 6, June 2016.
[3] Waller, M.A., Fawcett, S.E.: Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. J. Bus. Logistics 34(2), 77–84 (2013)

[4] Muniswamaiah, M., Agerwala, T., Tappert, C.C.: Context-aware query performance optimization for big data analytics in healthcare. In: 2019 IEEE High Performance Extreme Computing Conference (HPEC-2019), pp. 1–7 (2019)
[5] Latchoumi, T. P., Reddy, M. S., & Balamurugan, K. (2020). Applied Machine Learning Predictive Analytics to SQL Injection Attack Detection and Prevention. European Journal of Molecular & Clinical Medicine, 7(02), 2020.
[6] Fisher, D., et al.: Interactions with big data analytics. interactions 19(3), 50–59 (2012)
[7] Latchoumi, T. P., Ezhilarasi, T. P., & Balamurugan, K. (2019). Bio-inspired weighed quantum particle swarm optimization and smooth support vector machine ensembles for identification of abnormalities in medical data. SN Applied Sciences, 1(10), 1-10.
[8] Sridharan, K., & Sivakumar, P. (2018). A systematic review on techniques of feature selection and classification for text mining. International Journal of Business Information Systems, 28(4), 504-518.
[9] Perry, D.B., et al.: VizDeck: streamlining exploratory visual analytics of scientific data (2013)
[10] Latchoumi, T. P., Balamurugan, K., Dinesh, K., & Ezhilarasi, T. P. (2019). Particle swarm optimization approach for waterjet cavitation peening. Measurement, 141, 184-189.
[11] Wang, L., Wang, G., Alexander, C.A.: Big data and visualization: methods, challenges and technology progress. Digit. Technol. 1(1), 33–38 (2015)