Enhance Performance of Data using Visualization Tool: A Systemic Review

Archana Acharya[1], Harshita Sharma[2], Aisha Dhakre[3], Nitesh K. Rathore[4], Satyam Agarwal[5]

[1]Archana Acharya
Assistant Professor
Institute of Technology and Management, Gwalior-MP, India
Email Id: archanaacharya.it@itmgoi.in
[2]Harshita Sharma, [3]Aisha Dhakre, [4]Nitesh K. Rathore, [5]Satyam Agrawal
Student CSE-Data Science 3rd year
Institute of Technology and Management, Gwalior-MP, India
Email Id:harshitasharma000000@gmail.com

Abstract: Big volumes of data are being produced every day in the world in which we live. Data visualization is a crucial component of the data-driven decisions that corporations make based on their data. Through the use of graphics and science, data visualization is a relatively new and exciting area of computer science that tells tales through data. Patterns, trends, and correlations are extracted from databases using computer graphic effects. Tableau, Google Charts, Data Wrapper, Infogram, and other tools are among the many available for data visualization on the market. With a focus on their significance in a number of domains for efficient data transfer and analysis, the study attempts to give a thorough review of data visualization tools and approaches. Furthermore, we'll serve as an example project for data visualization using tableau desktop tool. Because it makes it easier for data scientists to convey insights, make decisions, tell stories, and study data, data visualization is essential to data science too. It would be difficult to draw conclusions from large, complicated data sets and make wise judgments without data visualization. This paper focuses on the importance of data visualization, primary tools, and software for data visualization, and theoretical architectural framework for data visualization. The article's last section will address mitigating strategies for the major problems with data visualization.

Keyword: visualization, mitigation, Tableau, data,etc.

1. INTRODUCTION

The graphical depiction of data is called data visualization, or simply Data Viz. Usually, data visualization is used to either explore or explaindata. When data exploration is employed, the creator aims to investigate and comprehend patterns inside their own study data. When it is used to explain data, the person who created the visualization does so because they are already aware of the patterns in the data and want to help others comprehend them. The proverb "a picture is worth a thousand words" comes to mind. A data-driven discovery may be effectively communicated with the help of data visualization. Sometimes the visualization is so compelling that more analysis is not necessary.

Many enterprises, academic institutions, and government missions now depend more heavily on data visualizations due to the proliferation of useful datasets and software tools. A graphical depiction of data is called data visualization. In order to facilitate the identification of patterns and the comprehension of complex ideas, data is presented as a picture or graphic. With the use of technology, users may change the parameters of the data to view more information and get fresh insights.

Here are several key aspects of data visualization:

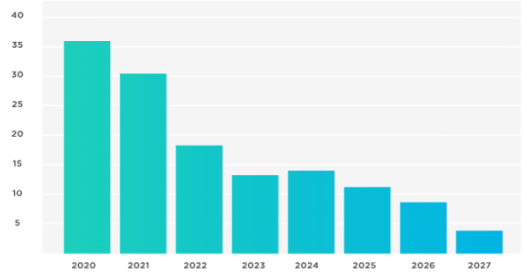
1. Communication of information
2. Storytelling with data
3. Importance in decision making
4. Data cleaning and preparation
5. Exploratory and explanatory visualization

2. TYPES OF DATA VISUALIZATION

There are several types of data visualizations:

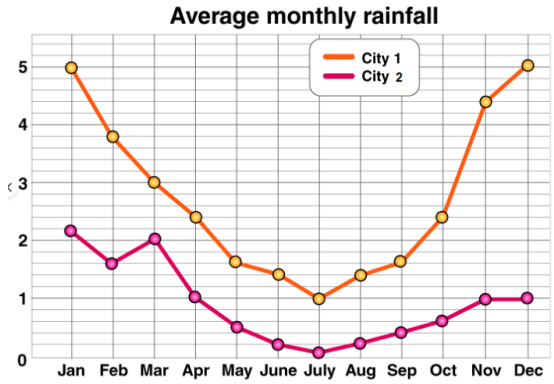
2.1 Column chart

A column chart is a method of displaying the provided data using vertical bar charts, which are represented by rectangles. This method makes it simple to compare several items and analyze trends. When given in textual or tabular style, numbers and figures are generally hard to interpret.



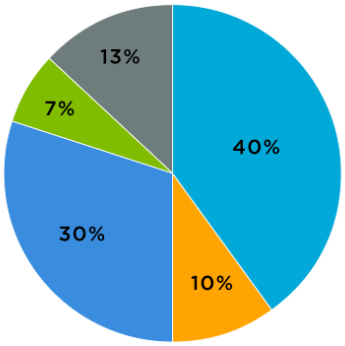
2.2 Line graph

A graph that uses lines to show change over time is called a line graph. It is a graph that shows a line connecting a sequence of subsequent data points to illustrate quantitative data between two changing variables.



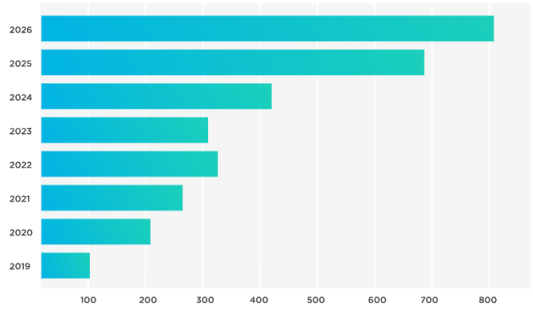
2.3 Pie chart

A pie chart is a statistical visual that is circular and has been sliced into slices to show numerical proportions. Every section represents a commensurate amount of the whole.



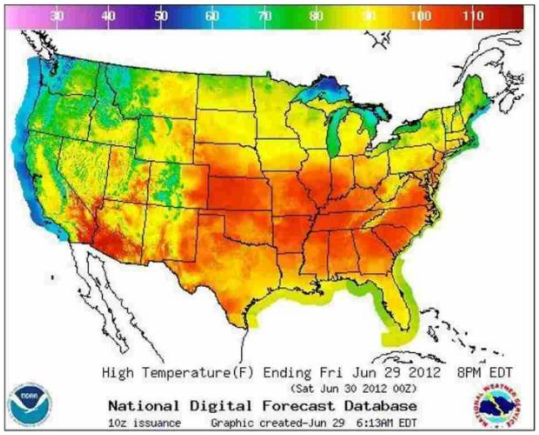
2.4 Bar chart

A bar chart is a graphical representation that uses rectangular bars to show and compare discrete data categories. Each bar's height or length corresponds to the frequency or value of the associated category.



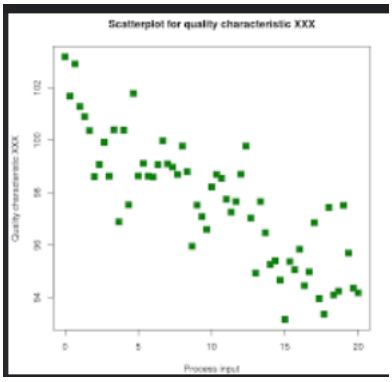
2.5 Heat maps

A heat map is a two-dimensional data representation where different colours correspond to different values. Although they may be used for many types of analytics, heat maps are most frequently employed to display user behaviour on certain webpages or webpage layouts. Heat maps can be used to show users the locations of their clicks on a website, the amount of scrolling they have done, or the outcomes of eye tracking experiments.



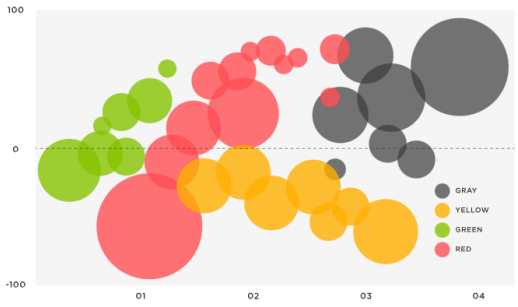
2.6 Scatter plot

In a scatter plot, the values of two distinct numerical variables are represented by dots. The values for each individual data point are shown by the position of each dot on the horizontal and vertical axes. Relationships between variables are observed through the use of scatter plots.



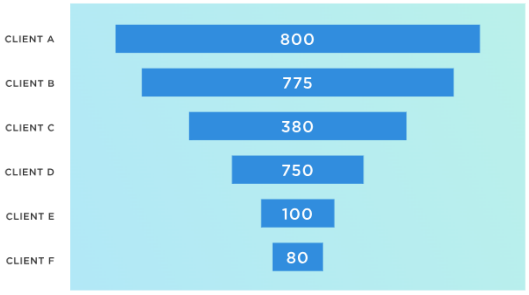
2.7 Bubble chart

A type of scatter plot in which the data points are shown as dots, and are further indicated by the size and color of the bubbles that represent the data points.



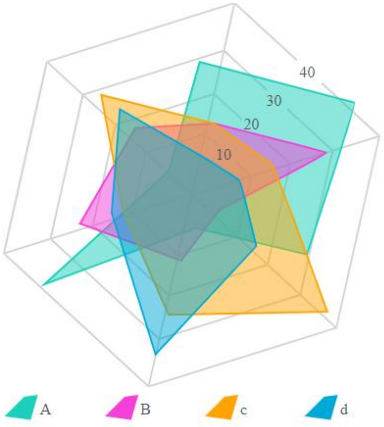
2.8 Funnel chart

A special kind of data visualization called a funnel chart shows data as it moves through several phases of a process visually. The name of this unusual chart comes from its design, which is a funnel with portions that get smaller and smaller one after the other.



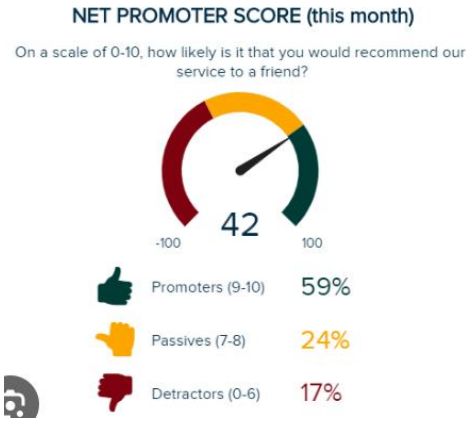
2.9 Radar chart

Multivariate data stacked at an axis with the same center point is shown on a radar chart. Three or more quantitative variables, referred to as radii, are displayed on the chart for comparison. The map is also known as a spider chart because of its resemblance to a spider web.



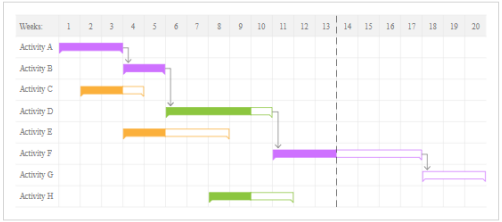
2.10Gauge

A gauge chart, often called a dial chart or a speedometer chart, is a style of data visualization where a single value of data is shown quantitatively. Visualizations of business intelligence frequently employ gauge charts. They provide information in an understandable manner, especially to non-specialists or people unfamiliar with the industry. They are frequently shown as a reminder of ongoing critical performance metrics on company dashboards.



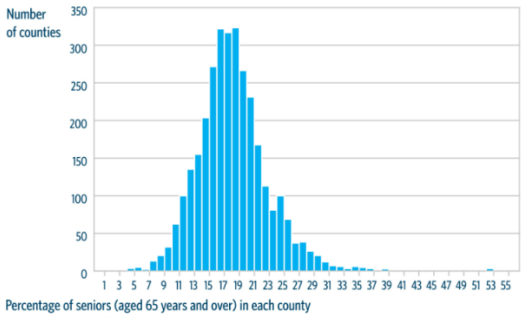
2.11Gantt chart

As a project management tool, a Gantt chart shows the amount of work that has been finished over time in comparison to the amount of time that was originally scheduled.



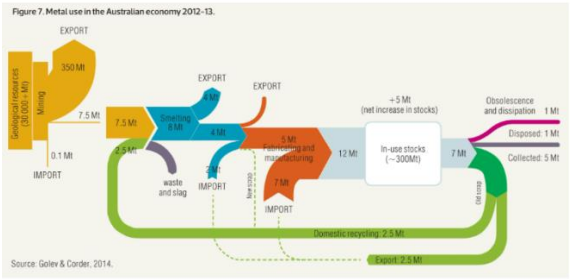
2.12Histogram

A figure with rectangles in it whose width equals the class interval and whose size is proportionate to the frequency of a variable is called a histogram. A figure with rectangles in it whose width equals the class interval and whose size is proportionate to the frequency of a variable is called a histogram.



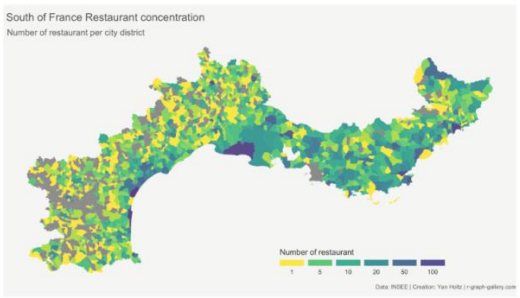
2.13Sankey diagrams

A Sankey diagram represents a flow of values from one set to another visually. Nodes are the objects that are connected, and links are the connections between them.



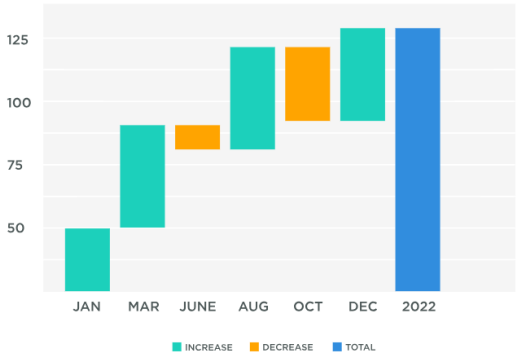
2.14Choropleth Map

Choropleth maps are themed maps that are used to represent statistical data. They are made using the colour mapping symbology approach. It displays parts or areas that are geographically divided and coloured, shaded, or patterned based on enumeration units, which are data variables.



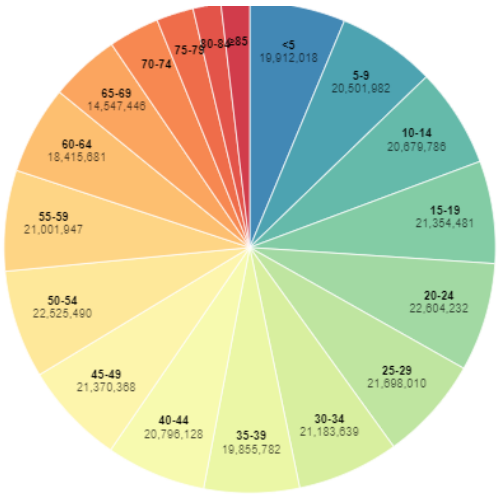
2.15Waterfall chart

An initial beginning point is the combined impact of consecutive positive and negative factors, represented graphically by a waterfall chart. Essentially, it provides an organized and clear way to illustrate the incremental changes, stressing the complex process by which various elements contribute to an outcome and their overall effect.



2.16Radial wheel

Each bar in a radial representation is surrounded by a circle, with longer bars denoting higher values. To view a bar's specific value and other data, hover your cursor over it. Beginning at noon, each bar moves clockwise for positive values and counterclockwise for negative ones.

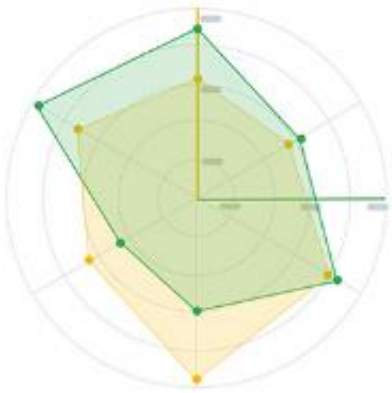


2.17Percentage bar

A horizontal bar with proportionate segments is used in percentage bars, a kind of data visualization, to display figures as percentages of the total and the relative sizes of each category.

2.18Polar chart

In essence, polar charts are a variant of the conventional Cartesian coordinate system. They plot data points on a circle using a radial method as opposed to x and y coordinates. One variable is represented by the angle (θ) formed between the data point and the circle's centre, while another variable is represented by the distance (r) between the two. Apart from the circle's centre, a distinct spoke represents each category or data item.



2.19Cone chart

Cone charts are column (or bar chart) that use conical shaped items to show data. Although cone charts do not add any

additional data, sometimes using this shape allows to achieve a better visual appearance of your data.

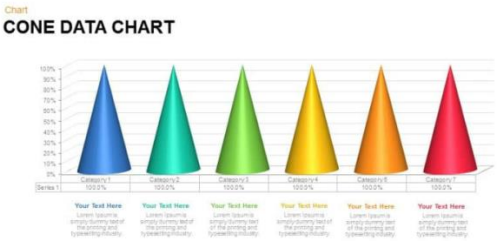


Table 1: Summary Table for Various Visualization Methods and their applications

Visualization Method	Description	Best For	Advantages	Disadvantages
Bar Chart	Displays data using rectangular bars	Comparing categories	Easy to understand, good for comparing quantities	Can become cluttered with too many categories
Line Chart	Shows data trends over time with lines connecting data points	Time series data, trends	Good for showing changes over time	Not suitable for categorical data
Histogram	Displays the distribution of a dataset	Frequency distribution	Shows data distribution clearly	Not suitable for comparing categories
Scatter Plot	Plots data points on a Cartesian coordinate system	Correlation and relationships between variables	Shows relationship between two variables	Can be hard to read with too many data points
Box Plot	Shows data distribution with quartiles and outliers	Statistical distribution, spotting outliers	Displays spread and skewness of data	Not easily understood by non-statisticians
Heatmap	Uses color to represent data values in a matrix	Correlation matrices, displaying data patterns	Good for showing patterns and correlations	Can be misleading if not properly scaled

3. PRINCIPLES OF DATA VISUALIZATION:

The data visualization principles can be summarized with the DESIGN acronym that stands for the following six data viz imperatives:

D eclutter

E mphasize

S torify **I**nvolve

G ive meaning

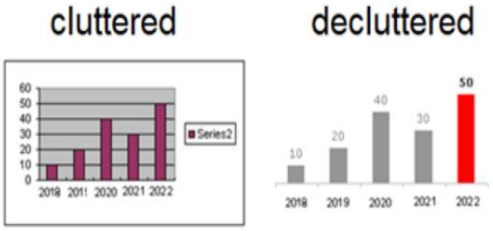
N o distortions

The logic behind this set of principles is a simple one:

A great data visualization presents a captivating tale (with a "so-what"), is free of clutter and distraction, takes into account and engages its audience, and graphically accentuates its primary insight. It prevents distortions, pointless attempts, and misunderstandings while providing statistics with a clear meaning.

3.1 Declutter

Decluttering a chart means removing everything that takes attention away from your data. Thus, eliminate borders, prominent grid lines, extraneous features (like decimals), 3D effects, (overuse of) colors, tones, and other decorative elements.



3.2 Emphasize

Two things are meant by emphasizing: first, it implies selecting the visual style that effectively communicates your main idea or the goal of the chart. Secondly, use visual cues to highlight the most significant feature in the chart, such circling it or using a different color.

3.3 Storify

Presenting a chart (or a succession of them) in a way that allows you to narrate an engaging story about the information you are displaying is known as "storifying data."

3.4 Involving

When making and presenting a chart, you must consider your intended audiences in order to include people in the context of the data. This may be achieved by offering your users easy options to offer comments on the chart and by allowing them to click on the chart to delve deeper into their inquiry.

You may include your data users in an interactive chart by giving them the option to choose regions of interest, zoom in for additional information, examine various data characteristics, remove items, alter the presentation, or link the chart to fresh data.

3.5 Give design

Data visualization is all about making data *meaningful* for your audience.

There are 8 ways of giving a data meaning:

(a) Linking data directly to possible *actions* or responses is one way to make it more meaningful.

(b) Giving the chart an *action title* that expresses its so-what is another one.

(c) Adding self-*explanatory labels* and axes descriptions to a chart helps to make it more meaningful even for hurried viewers.

(d) Carefully adding *symbols* to a chart can help in its interpretation, for example a £, $, or € symbol for line charts with currency comparisons over time.

(e) Explaining the reasons behind *outliers* or other strange data patterns

(f) Making data meaningful in a dashboard can be achieved by providing a *reference point* that shows whether a value is actually good or bad.

(g) You can also give meaning to numbers by showing them in their *development* over time.

(h) Last but not least, you can make any number more relatable by comparing it to a phenomenon that the audience is *familiar* with.

3.6 No Distortions

Regarding the last point, the general guideline is to avoid using visual representations that make data hard to read or easily misinterpreted. Pie, donut, and arch charts are examples of sub-optimal formats because they are difficult to compare and perceptually inefficient. Other sub-optimal formats include stacked bar and area charts because of their changing baselines, charts that mix units, and charts with two separate y axes in one picture (or a y-axis that does not begin at the value 0). Additionally, you should steer clear of line charts with a lot of crossing lines (instead, use so-called tiny multiples) as these are particularly difficult to read because of all the crossovers and overlaps.

4. SIGNIFICANCE OF DATA VISUALIZATION IN DATA SCIENCE

For a number of reasons, data visualization is crucial in the field of data science. It is essential at every stage of the data science lifecycle, from conveying findings to conducting exploratory data analysis. The following are some salient features that underscore the importance of data visualization in data science:

4.1 Exploratory data Analysis

A useful tool for examining and comprehending the structure of data is data visualization. Data scientists may more easily create hypotheses and direct more research when they can see patterns, trends, and outliers in the data using

visual representations like scatter plots, histograms, and box plots.

4.2 Sharing insights

Data scientists must share their discoveries with a range of stakeholders, including non-technical audiences. Through the simplification and increased accessibility of complicated data, visualizations facilitate the successful sharing of important discoveries, trends, and insights.

4.3 Pattern recognition

Data scientists can find connections and patterns in the data that might not be immediately visible in raw datasets by using visualizations. This is necessary in order to extract useful insights from the data and make well-informed judgments.

4.4 Model Validation and Evaluation

Visualizations are helpful in assessing the performance of machine learning models as they are being built. Visual techniques like as ROC curves, confusion matrices, and precision-recall curves are employed to evaluate the precision and efficacy of models.

4.5 Feature selection and validation

By pointing out the most important factors for modelling, visualizations help in feature selection. They also support data scientists in feature engineering, which involves helping them develop fresh, educational features based on patterns seen in the data's visual representations.

4.6 Time series and analysis

Data that is dependent on time is common across a number of fields. Data scientists can analyze temporal patterns, seasonality, and anomalies in time series data with the use of visualizations like heatmaps and line charts.

4.7 Preprocessing and data cleaning

Missing values, outliers, and inconsistent data may all be found and fixed with the use of visualizations. They direct the data cleaning and preparation stages by offering insights into the distribution and properties of the data.

4.8 Decision support

Visualizations are tools that help with data-driven decision-making. They let stakeholders to comprehend the ramifications of various situations and make defensible decisions by providing a visual representation of the facts.

4.9 Reporting and storytelling

Data scientists frequently have to provide their findings in an understandable and engaging way. By arranging data in a narrative fashion, visualizations help communicate data-driven stories by making it simpler for others to follow and comprehend.

4.10Application user interface design

Visualizations are essential to the user interface design of data-driven applications. They improve the user experience by displaying data in an aesthetically pleasing and intuitive manner.

5. Data Visualization Tools Used In Data Science

There are many data visualization tools for creating plots and charts. Here are some most popular:

5.1 Python libraries

Python is a popular programming language for data science, and there are many libraries available for creating plots and charts. Matplotlib, Seaborn, and Plotly are often used for data science visualization.

5.2 R packages

R is another popular programming language for data science. It also has many packages for creating plots and charts. Popular ones include ggplot2, lattice, and ggvis.

5.3 Tableau

Tableau is a powerful data visualization tool that lets users create interactive dashboards and reports. It supports a range of data sources and allows users to easily create and share visualizations.

5.4 Excel

Excel is a widely used spreadsheet program that includes basic charting functionality. It's often employed for quick data exploration and visualization.

5.5 Power BI

Power BI is a data visualization tool from Microsoft that lets users create interactive dashboards and reports. It supports a range of data sources and allows users to easily create and share visualizations.

5.6 D3.js

D3.js is a JavaScript library for creating dynamic and interactive data visualizations in web browsers. It's often used for creating custom visualizations that are impossible to do with other tools

6. USE CASES OF DATA VISUALIZATION

6.1 Business

Data science visualization is an essential tool in business for comprehending and sharing data-driven insights. Sales teams may monitor industry trends and pinpoint opportunities for development. Marketing departments are able to track the effectiveness of advertising efforts and keep an eye on social media involvement. Business executives may keep an eye on revenue, profit margins, and client retention rates by tracking key performance indicators (KPIs) in the interim. In the corporate world, Tableau and Power BI are especially well-liked because they let users generate interactive dashboards and reports that can be sent to stakeholders.

6.2 Healthcare

In the medical field, visualization is used to track disease outbreaks, assess medical imaging data, and monitor patient outcomes. Epidemiologists are able to monitor the transmission of infectious illnesses and pinpoint possible contributing variables. Radiologists are able to identify anomalies in medical imaging by analyzing them.

Enhancing patient outcomes and keeping an eye on hospital performance are two other uses for data visualization. In healthcare data analysis, Tableau, R packages, and Python libraries are frequently utilized.

6.3 Finance

Data visualization is used in finance to manage investments, assess stock market patterns, and keep an eye on financial performance. For instance, finance teams may monitor financial performance parameters like sales, profit, and costs, and financial analysts can spot investment opportunities.

Excel, Python libraries, and Tableau are widely used in financial data analysis.

6.4 Government

Data visualization is used to track economic indicators, keep an eye on public health and safety, and assist in decision-making. Public health experts, for example, can monitor the transmission of infectious illnesses and pinpoint possible risk factors.

Government agencies also utilize data visualization to help with decision-making processes including determining where to invest and gauging the success of policy efforts. Tableau, R packages, and Python libraries are frequently used.

6.5 Marketing

Data science visualization is used in marketing to watch consumer behaviour, evaluate advertising efforts, and keep an eye on social media activity. Email campaign performance, consumer preferences, and website traffic may all be tracked by marketers. Social media analysts may follow the success of social media efforts and keep an eye on social media participation in the interim.

Tableau and Power BI are particularly popular in marketing data analysis.

7. FUTURE SCOPE OF DATA VISUALIZATION

The future scope of data visualization is quite promising as it continues to play a crucial role in various fields. Here are some trends and areas where data visualization is expected to have a significant impact in the future:

7.1 Advanced Analytics and AI Integration

Data visualization will be combined with artificial intelligence (AI) and machine learning (ML) technologies as they develop to offer more sophisticated and predictive insights. This could entail sophisticated analytics, pattern recognition, and automated data finding.

7.2 Augmented and Virtual Reality

The integration of virtual reality (VR) and augmented reality (AR) into data visualization will improve immersion and offer fresh approaches to working with intricate datasets. This can be especially helpful in sectors like gaming, healthcare, and education.

7.3 Storytelling with Data

Using data visualization to effectively convey stories will remain a major area of attention. Tools and methods that let people utilize data to create engaging tales that are more widely accessible and intelligible

7.4 Data Democratization

Non-technical people will continue to benefit from the trend of making data available. Tools for data visualization that are easy to use and don't require much technical knowledge will enable more people in more fields to gain insights from data.

7.5 Customization and Personalization

One of the main areas of development will be adjusting data visualizations to the tastes and requirements of specific users. This may entail user-driven analytics, customizable dashboards, and flexible visualizations.

7.6 Integration with Business Intelligence (BI) systems

Data visualization tools will keep on smoothly integrating with BI systems, offering businesses a more complete way to examine and comprehend their data.

7.7 Cross-platform and Mobile Visualization

Data visualization solutions must be designed for cross-platform capabilities in order to provide a consistent and useful user experience across various platforms, given the growing popularity of mobile devices.

7.8 Sustainability and Environment Data Visualization

There will be a greater emphasis on environmental data visualization as worries about sustainability and climate change intensify. This might involve showcasing additional environmental data, charting developments in renewable energy, and mapping carbon footprints.

8. CONCLUSION

Finally, in the field of information analysis and decision-making, data visualization tools have become essential. Their capacity to convert intricate datasets into aesthetically pleasing representations enables users in a variety of fields to derive important insights and effectively convey discoveries. The potential for data visualization tools in the future seem quite promising.

Data visualization will become more powerful when advanced analytics, AI, and machine learning are combined, giving users the ability to extract more insightful and foresighted information from their data. In an age of perpetual information flow, real-time and streaming data visualization will help with decision-making quickly, while augmented and virtual reality will provide immersive dataset exploration a whole new level.

Data will continue to become more accessible to a wider range of people because to user-friendly technologies that enable decision-making based on data for people from a variety of backgrounds. Data visualization ethics will become more important, guaranteeing that the visualizations are truthful and compliant with privacy and fairness laws. Individual tastes will be satisfied by the personalization and customisation of data visualizations, and businesses will receive complete solutions through the smooth integration of these platforms with business intelligence systems. Anytime, anyplace access to data visualization will be possible thanks to cross-platform and mobile optimization.

Moreover, the growing significance of using data-driven insights to address global concerns is reflected in the increased emphasis on environmental and sustainability data visualization. Data visualization tools will become increasingly important as they develop, influencing how we see complicated problems and bringing about constructive change.

Fundamentally, the confluence of state-of-the-art technology, improved accessibility, and a dedication to ethical principles will determine the direction of data visualization tools in the future. These technologies will enable people and businesses to successfully traverse the complex data landscape as they develop further, transforming information into knowledge that can be put to use.

References:

[1] https://rafalab.dfci.harvard.edu/dsbook/introduction-to-datavisualization.htmlhttps://guides.library.georgetown.edu/datavisualization

[2] https://www.simplilearn.com/types-of-data-visualization-article

[3] https://www.linkedin.com/pulse/data-design-six-must-know-visualization-principles-everyone-eppler/

[4] https://www.cell.com/patterns/pdf/S2666-3899(20)30189-6.pdf

[5] https://www.perceptualedge.com/articles/Whitepapers/Data_Visualization.pdf

[6] https://doi.org/10.22362/ijcert/2023/v10/i01/v10i0102

[7] https://scholar.google.co.in/scholar?q=data+visualization+research+papers&hl=en&as_sdt=0&as_vis=1&oi=scholart#d=gs_qabs&t=1728298976213&u=%23p%3DQpf3EltSv74J

[8] https://www.researchgate.net/publication/370593444_An_Introduction_to_Data_Visualization_Tools_and_Techniques_in_Various_Domains

[9] https://www.tableau.com/learn/articles/data-visualization

[10] Tableau-1, "The history of data visualizations - From cave drawings to tableau," 1 April 2020. [Online]. Available: https://www.tableau.com/whitepapers/designin g-great-visualizations. [Accessed 4 Dec 2022].

[11] Data.org, "Introduction to data visualization," data.org, 22 Dec 2022. [Online]. Available: https://data.org/resources/introduction-to-data visualization/. [Accessed 22 Dec 2022].

[12] M. Sharapa, "Data visualization: Principles, tools, and useful tricks," Medium, 10 July 2020. [Online]. Available: https://towardsdatascience.com/data visualization-principles-tools-and-useful-tricks b68d9c138a86. [Accessed 20 Dec 2022].