UNIT-V

Predictive Analytics and Visualizations: Predictive Analytics, Simple linear regression, Multiple linear regression, Interpretation of regression coefficients, Visualizations, Visual data analysis techniques, interaction techniques, Systems and application

Predictive Analytics, Simple linear regression, Multiple linear regression, Interpretation of regression coefficients:

1. Explain in briefly Predictive Analytics. **[Remembering]**
2. How Does a Linear Regression Work? Write some examples of linear regression? **[7M-R20-SET-2-July 2023] [Analysis]**
3. Write brief note on Predictive Analytics? **[7M-R20-SET-2-July 2023] [Analysis]**
4. What is multiple linear regressions? How is it different from linear regression? What type of data is used for Multiple Linear regression analysis? **[14M-R20-SET-2-July 2023][Remembering]**
5. Write the advantages of Predictive Analytics. And various challenges faced by it. **[7M-R20-SET-3-July 2023] [Create]**
6. How to interpret the regression coefficients? Explain with linear regression. **[7M-R20-SET-4-July 2023][Analysis]**
7. Difference between Big Data analytics and Predictive Analytics? **[Create]**

Visualizations, Visual data analysis techniques, interaction techniques, Systems and application:

8. What are the various data visualization techniques used in big data analytics? Illustrate any five . **[14M-R20-SET-1-July 2023] [Remembering]**
9. What are the Visual data analysis techniques and Systems application? **[7M-R20-SET-2-July 2023] [Remembering]**
10. Explain about the Interaction Techniques ? **[Understanding]**
11. Explain about the Systems and Applications? **[Understanding]**

Predictive Analytics

Definition of Predictive Analytics

"Predictive Analytics refers to the field that applies various quantitative methods on data to make real-time predictions."

It provides a method of approaching and solving problems using various technologies. Predictive Analytics often makes use of machine learning algorithms and techniques to build models that make predictions.

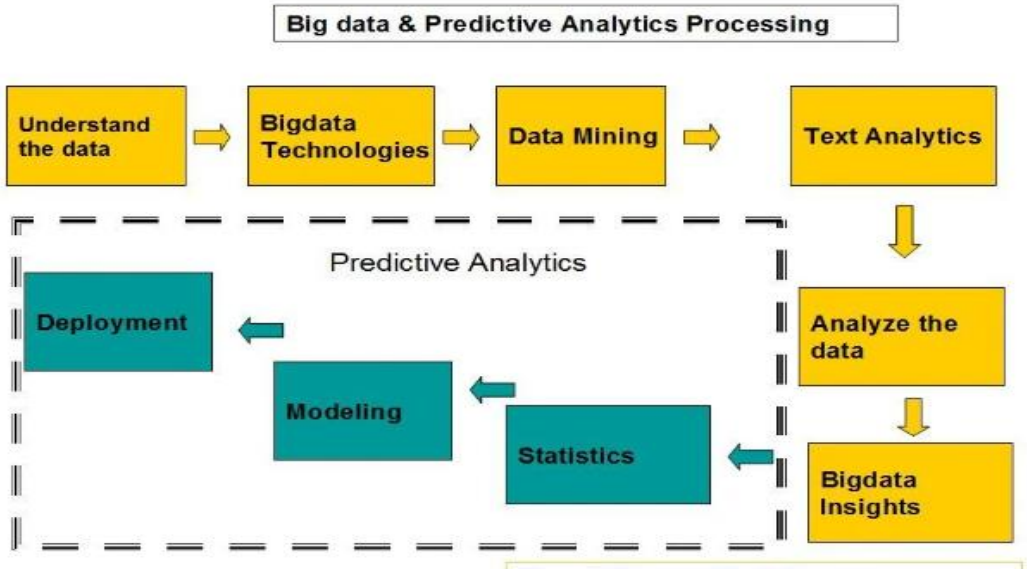
Big Data is huge, large or voluminous data, information, or the relevant statistics acquired by the large organizations and ventures. Many software and data storage is created and prepared as it is difficult to compute the big data manually. It is used to discover patterns and trends and make decisions related to human behavior and interaction technology.

Predictive Analytics encompasses making predictions about future outcomes by studying current and past data trends. It utilizes data modeling, data mining, machine learning, and deep learning algorithms to extract the required information from data and project behavioral patterns for the future.

Some industry tools used for Predictive analytics are Periscope Data, Google AI Platform, SAP Predictive Analytics, Anaconda, Microsoft Azure, Rapid Insight Veera, and KNIME Analytics Platform.

Predictive analytics models may be able to identify correlations between sensor readings. For example, if the temperature reading on a machine correlates to the length of time it runs on high power, those two combined readings may put the machine at risk of downtime. Predict future state using sensor values.

One predictive analytics tool is regression analysis, which can determine the relationship between two variables (single linear regression) or three or more variables (multiple regressions). The relationships between variables are written as a mathematical equation that can help predict the outcome should one variable change.



Difference between Big Data and Predictive Analytics

SR.NO	Big Data	Predictive Analytics
1.	Big Data is group of technologies. It is a collection of huge data which is multiplying continuously.	Predictive analytics is the process by which raw data is first processed into structured data and then patterns are identified to predict future events.
2.	It deals with the quantity of data, typically in the range of .5 terabytes or more.	It deals with the application of statistical models to existing data to forecast.
3.	It's a best practice for enormous data.	It's a best practice for data for future prediction.
4.	It has a vast backend technology imports for Dashboards and Visualizations like D3js and some paid ones like Spotfire a TIBCO tool for reporting.	It has tool with built-in integrations of the reporting tools like Microsoft BI tools. So, no need to fetch it from source or from some outside vendors.
5.	Its engines like Spark and Hadoop comes with built-in Machine Learning libraries but the incorporation with AI is still an R&D task for the Data Engineers.	It deals with the platform based on the probability and mathematical calculation.

6.	It has high level of advancement, its engines have eventually upgraded themselves throughout the development processes and level of cross-platform compatibility.	It has medium level of advancement, has a limited change of algorithmic patterns as they are giving them better score from the start with respect to their field and domain-specific work analysis.
7.	It is used to make data driven decisions.	It is used for risk evaluation and prediction of future outcomes.

Regression is a tool that allows you to estimate how the dependent variable changes as the independent variable(s) change.

Regression models describe the relationship between variables by fitting a line to the observed data. Linear regression models use a straight line, while logistic and nonlinear regression models use a curved line.

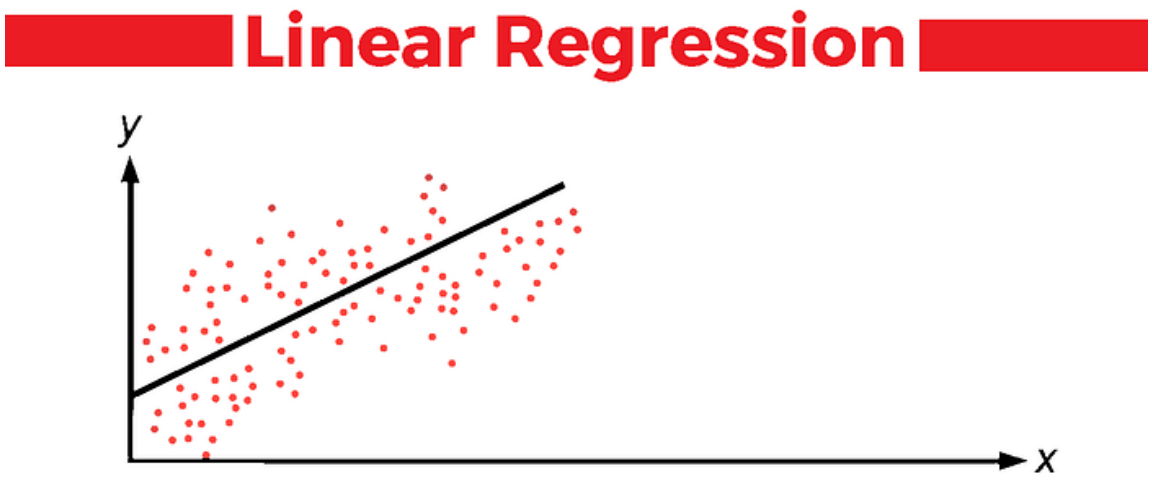
Types of Linear Regression

- Simple Linear Regression :
- Multiple Linear Regression:

Simple Linear Regression

What is Linear Regression?

Linear Regression is a statistical supervised learning technique to predict the quantitative variable by forming a linear relationship with one or more independent features.



How does a linear Regression Work?

The whole idea of the linear Regression is to find the best fit line,which has very low error(cost function).
This line is also called Least Square Regression Line(LSRL).

The line of best fit is described with the help of the formula ***y=mx+b.***
where,m is the Slope and
b is the intercept.

Regression models can be used for many purposes:
Evaluating the effect of an independent variable on a dependent variable.
Forecasting future values of the dependent variable based on prior observations of both variables.

What Is Simple Linear Regression?

$y = \beta 0 + \beta 1x + \varepsilon$ is the formula used for simple linear regression.
y is the predicted value of the dependent variable (y) for any given value of the independent variable (x).
B0 is the intercept, the predicted value of y when the x is 0.
B1 is the regression coefficient – how much we expect y to change as x increases.
x is the independent variable (the variable we expect is influencing y).
e is the error of the estimate, or how much variation there is in our regression coefficient estimate.

- Simple linear regression establishes a line that fits your data, but it does not guarantee that the line is good enough. For example, if your data points have an upward trend and are very far apart, then simple linear regression will give you a downward-sloping line, which will not match your data.

Assumptions of Linear Regression:

The *Independent variables* should be linearly related to the *dependent variables*.
This can be examined with the help of several visualization techniques like: Scatter plot or maybe you can use Heatmap or pairplot(to visualize every features in the data in one particular plot).

Every feature in the data is Normally Distributed.
This again can be checked with the help of different visualization Techniques,such as Q-Q plot,histogram and much more.

There should be little or no multi-collinearity in the data.
The best way to check the prescence of multi-collinearity is to perform VIF(Variance Inflation Factor).

The mean of the residual is zero.
A **residual** is the difference between the observed y-value and the predicted y-value.However, Having residuals closer to zero means the model is doing great.

Residuals obatined should be normally distributed.
This can be verified using the Q-Q Plot on the residuals.

Variance of the residual throughout the data should be same.This is known as homoscedasticity.
This can be checked with the help of residual vs fitted plot.

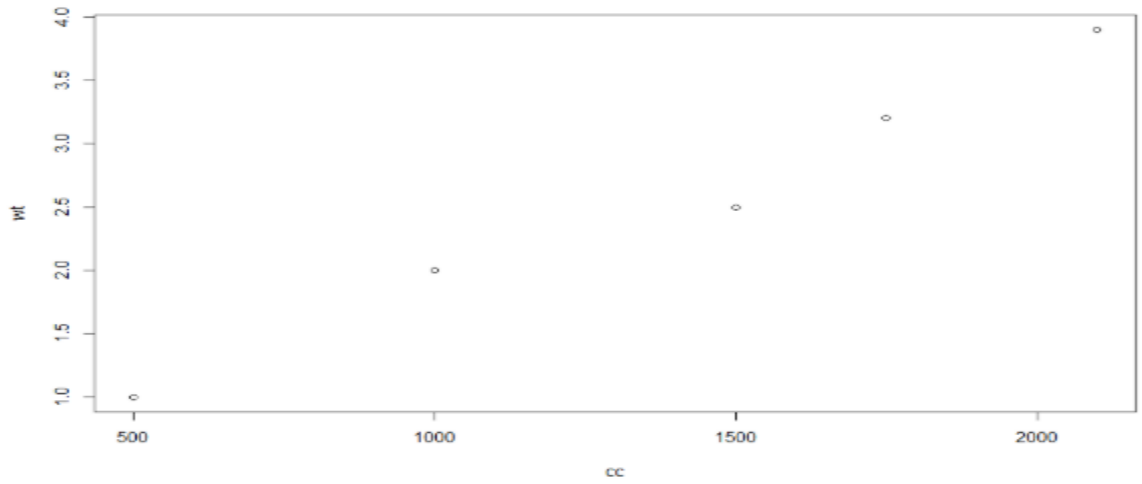
There should be little or no Auto-Correlation is the data.
Auto-Correlation Occurs when the residuals are not independent of each other.This usally takes place in time series analysis.
You can perform Durbin-Watson test or plot ACF plot to check for the autocorrelation.If the value of Durbin-Watson test is 2 then that means no autocorrelation,If value < 2 then there is positive correlation and if the value is between >2 to 4 then there is negative autocorrelation.

Example:
Let us understand simple linear regression by considering an example. Consider we want to predict the weight gain based upon calories consumed only based on the below given data.

Calories Consumed	Weight Gain
500	1
1000	2
1500	2.5
1750	3.2
2100	3.9

Now, if we want to predict weight gain when you consume 2500 calories.

Firstly, we need to visualize data by drawing a scatter plot of the data to conclude that calories consumed is the best independent variable X to predict dependent variable Y.

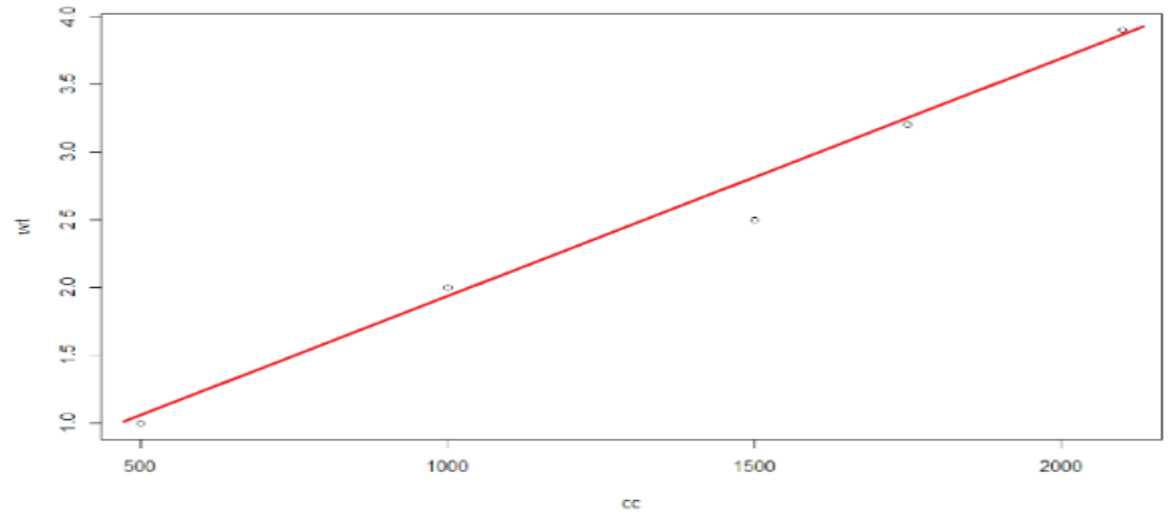


We can also calculate "r" as follows:

```
> cor(cc,wt)
[1] 0.9910422
```

As, r = 0.9910422 which is greater than 0.85, we shall consider calories consumed as the best independent variable(X) and weight gain(Y) as the predict dependent variable.

Now, try to imagine a straight line drawn in a way that should be close to every data point in the scatter diagram.

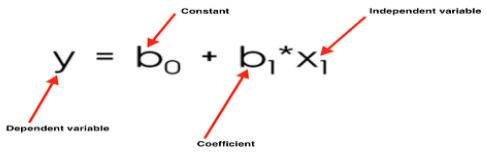


To predict the weight gain for consumption of 2500 calories, you can simply extend the straight line further to the y-axis at a value of 2,500 on x-axis . This projected value of y-axis gives you the rough weight gain. This straight line is a regression **line**.

Simple linear regression is a statistical method for establishing the relationship between two variables using a straight line. The line is drawn by finding the slope and intercept, which define the line and minimize regression errors.

The simplest form of simple linear regression has only one x variable and one y variable. The x variable is the independent variable because it is independent of what you try to predict the dependent variable. The y variable is the dependent variable because it depends on what you try to predict.

Simple Linear Regression helps to find the linear relationship between two continuous variables,One independent and one dependent feature.Formula can be represented as ***y=mx+b*** or,



Advantages and Disadvantages of Linear Regression

Advantages	Disadvantages
Linear Regression is simple to implement and easier to interpret the output coefficients.	On the other hand in linear regression technique outliers can have huge effects on the regression and boundaries are linear in this technique.
When you know the relationship between the independent and dependent variable have a linear relationship, this algorithm is the best to use because of it's less complexity compared to other algorithms.	Diversely, linear regression assumes a linear relationship between dependent and independent variables. That means it assumes that there is a straight-line relationship between them. It assumes independence between attributes.
Linear Regression is susceptible to over-fitting but it can be avoided using some dimensionality reduction techniques, regularization (L1 and L2) techniques and cross-validation.	But then linear regression also looks at a relationship between the mean of the dependent variables and the independent variables. Just as the mean is not a complete description of a single variable, linear regression is not a complete description of relationships among variables.

Multiple Linear Regression

Linear regression is a model that predicts one variable's values based on another's importance. It's one of the most popular and widely-used models in machine learning, and it's also one of the first things you should learn as you explore machine learning.

Linear regression is so popular because it's so simple: all it does is try to predict values based on past data, which makes it easy to get started with and understand.

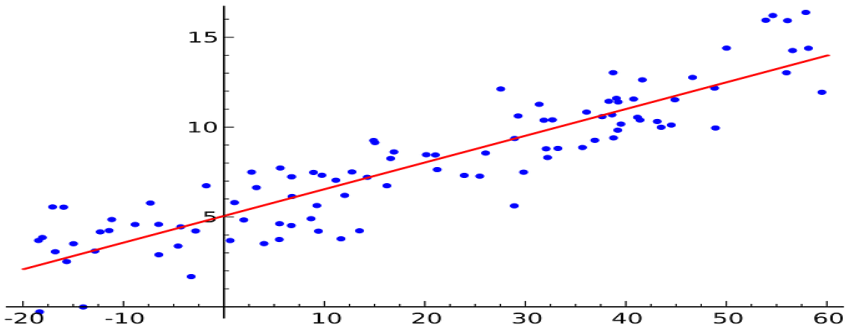
There are two types of linear regression algorithms:

- Simple - deals with two features.
- Multiple - deals with more than two features.

What is Multiple Linear Regression

One of the most common types of predictive analysis is multiple linear regression. This type of analysis allows you to understand the relationship between a continuous dependent variable and two or more independent variables.

The independent variables can be either continuous (like age and height) or categorical (like gender and occupation). It's important to note that if your dependent variable is categorical, you should dummy code it before running the analysis.



Several circumstances that influence the dependent variable simultaneously can be controlled through multiple regression analysis. Regression analysis is a method of analyzing the relationship between independent variables and dependent variables.

Let k represent the number of variables denoted by x1, x2, x3, ……, xk.

For this method, we assume that we have k independent variables x1, . . . , xk that we can set, then they probabilistically determine an outcome Y.

Furthermore, we assume that Y is linearly dependent on the factors according to

$$Y = \beta 0 + \beta 1x1 + \beta 2x2 + \cdots + \beta kxk + \varepsilon$$

The variable yi is dependent or predicted

The slope of y depends on the y-intercept, that is, when xi and x2 are both zero, y will be β0.

The regression coefficients β1 and β2 represent the change in y as a result of one-unit changes in xi1 and xi2.

βp refers to the slope coefficient of all independent variables

ε term describes the random error (residual) in the model.

Where ε is a standard error, this is just like we had for simple linear regression, except k doesn't have to be 1.

We have n observations, n typically being much more than k.

For i th observation, we set the independent variables to the values xi1, xi2 . . . , xik and measure a value yi for the random variable Yi.

Thus, the model can be described by the equations.

$$Yi = \beta 0 + \beta 1 xi1 + \beta 2 xi2 + \cdots + \beta kxik + i \text{ for } i = 1, 2, \ldots, n,$$

Where the errors i are independent standard variables, each with mean 0 and the same unknown variance σ2.

Altogether the model for multiple linear regression has k + 2 unknown parameters:

$$\beta 0, \beta 1, \ldots, \beta k, \text{ and } \sigma 2.$$

When k was equal to 1, we found the least squares line $y = \hat{\beta} 0 + \hat{\beta} 1x$.

It was a line in the plane R 2.

Now, with $k \geq 1$, we'll have a least squares hyperplane.

$y = \hat{\beta} 0 + \hat{\beta} 1x1 + \hat{\beta} 2x2 + \cdots + \hat{\beta} kxk$ in Rk+1.

The way to find the estimators $\hat{\beta} 0, \hat{\beta} 1, \ldots$, and $\hat{\beta} k$ is the same.

Take the partial derivatives of the squared error.

$Q = Xn\ i=1\ (yi - (\beta 0 + \beta 1 xi1 + \beta 2 xi2 + \cdots + \beta kxik))2$

When that system is solved we have fitted values

$\hat{y}i = \hat{\beta} 0 + \hat{\beta} 1xi1 + \hat{\beta} 2xi2 + \cdots + \hat{\beta} kxik$ for i = 1, . . . , n that should be close to the actual values yi.

Assumptions of Multiple Linear Regression

In multiple linear regression, the dependent variable is the outcome or result from you're trying to predict. The independent variables are the things that explain your dependent variable. You can use them to build a model that accurately predicts your dependent variable from the independent variables.

For your model to be reliable and valid, there are some essential requirements:

The independent and dependent variables are linearly related.

There is no strong correlation between the independent variables.

Residuals have a constant variance.

Observations should be independent of one another.

It is important that all variables follow multivariate normality.

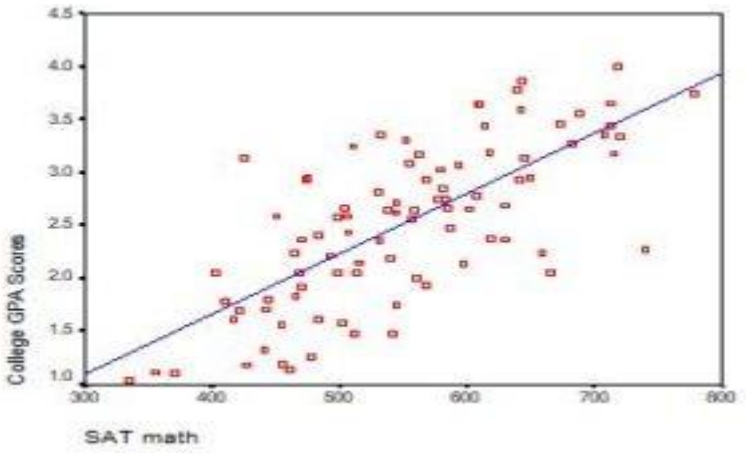
Simple Linear Regression vs Multiple Linear Regressions

- predicting a complex process's outcome, it's best to use multiple linear regression instead of simple linear regression. But it is not necessary to use complex algorithms for simple problems.
- A simple linear regression can accurately capture the relationship between two variables in simple relationships. But when dealing with more complex interactions that require more thought, you need to switch from simple to multiple regression.

- A multiple regression model uses more than one independent variable. It does not suffer from the same limitations as the simple regression equation, and it is thus able to fit curved and non-linear relationships.

Interpretation of regression coefficients

So let's interpret the coefficients in a model with two predictors: a continuous and a categorical variable. The example here is a linear regression model. But this works the same way for interpreting coefficients from any regression model without interactions.



A linear regression model with two predictor variables results in the following equation:

$Y_i = B_0 + B_1*X_{1i} + B_2*X_{2i} + e_i$.

The variables in the model are:

Y, the response variable;

X_1, the first predictor variable;

X_2, the second predictor variable; and

e, the residual error, which is an unmeasured variable.

The parameters in the model are:

B_0, the Y-intercept;

B_1, the first regression coefficient; and

B_2, the second regression coefficient.

One example would be a model of the height of a shrub (Y) based on the amount of bacteria in the soil (X_1) and whether the plant is located in partial or full sun (X_2).

Height is measured in cm. Bacteria is measured in thousand per ml of soil. And type of sun = 0 if the plant is in partial sun and type of sun = 1 if the plant is in full sun.

Let's say it turned out that the regression equation was estimated as follows:

$Y = 42 + 2.3*X_1 + 11*X_2$

Interpreting the Intercept

B_0, the Y-intercept, can be interpreted as the value you would predict for Y if both $X_1 = 0$ and $X_2 = 0$.

We would expect an average height of 42 cm for shrubs in partial sun with no bacteria in the soil. However, this is only a meaningful interpretation if it is reasonable that both X_1 and X_2 can be 0, and if the data set actually included values for X_1 and X_2 that were near 0.

If neither of these conditions are true, then B0 really has no meaningful interpretation. It just anchors the regression line in the right place. In our case, it is easy to see that X_2 sometimes is 0, but if X_1, our bacteria level, never comes close to 0, then our intercept has no real interpretation.

Interpreting Coefficients of Continuous Predictor Variables

Since X_1 is a continuous variable, B_1 represents the difference in the predicted value of Y for each one-unit difference in X_1, if X_2 remains constant.

This means that if X_1 differed by one unit (and X_2 did not differ) Y will differ by B_1 units, on average.

In our example, shrubs with a 5000/ml bacteria count would, on average, be 2.3 cm taller than those with a 4000/ml bacteria count. They likewise would be about 2.3 cm taller than those with 3000/ml bacteria, as long as they were in the same type of sun.

(Don't forget that since the measurement unit for bacteria count is 1000 per ml of soil, 1000 bacteria represent one unit of X_1).

Interpreting Coefficients of Categorical Predictor Variables

Similarly, B_2 is interpreted as the difference in the predicted value in Y for each one-unit difference in X_2 if X_1 remains constant. However, since X_2 is a categorical variable coded as 0 or 1, a one unit difference represents switching from one category to the other.

B_2 is then the average difference in Y between the category for which $X_2 = 0$ (the reference group) and the category for which $X_2 = 1$ (the comparison group).

So compared to shrubs that were in partial sun, we would expect shrubs in full sun to be 11 cm taller, on average, at the same level of soil bacteria.

Interpreting Coefficients when Predictor Variables are Correlated

Don't forget that each coefficient is influenced by the other variables in a regression model. Because predictor variables are nearly always associated, two or more variables may explain some of the same variation in Y.

Therefore, each coefficient does not measure the total effect on Y of its corresponding variable. It would if it were the only predictor variable in the model. Or if the predictors were independent of each other.

Rather, each coefficient represents the *additional* effect of adding that variable to the model, *if the effects of all other variables in the model are already accounted for*.

This means that adding or removing variables from the model will change the coefficients. This is not a problem, as long as you understand why and interpret accordingly.

Interpreting Other Specific Coefficients

I've given you the basics here. But interpretation gets a bit trickier for more complicated models, for example, when the model contains quadratic or interaction terms. There are also ways to rescale predictor variables to make interpretation easier.

Big Data Visualization

A perfect visualization tool will generate an efficient visual diagram which will lead to a correct decision. Insufficient visualization will lead to a loss for the organization.

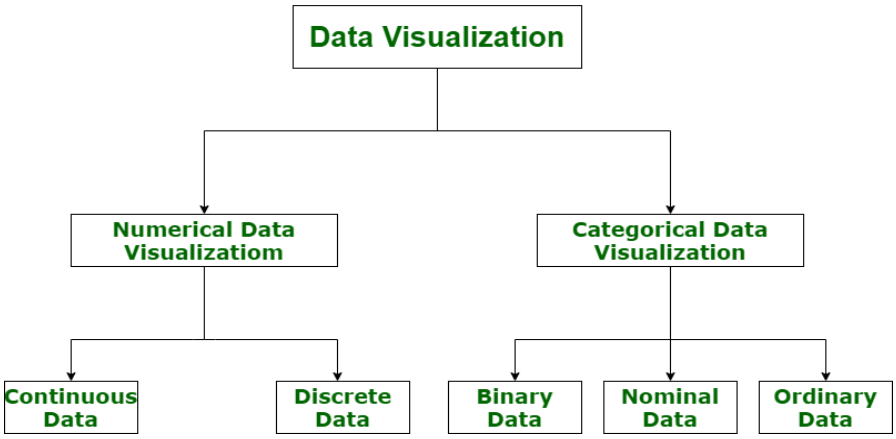
On Facebook, 4 petabytes of data are uploaded per day that contains different information like video, images, or textual information. Without visualizing those, it's hard to understand patterns and other relevant information.

The selection of efficient big-data visualization tools will help change complex and extensive volume data into simple and human-readable visual diagrams.

Categories of Data Visualization ;

Data visualization is very critical to market research where both numerical and

categorical data can be visualized that helps in an increase in impacts of insights and also helps in reducing risk of analysis paralysis. So, data visualization is categorized into following categories:



Numerical Data :

Numerical data is also known as Quantitative data. Numerical data is any data where data generally represents amount such as height, weight, age of a person, etc. Numerical data visualization is easiest way to visualize data. It is generally used for helping others to digest large data sets and raw numbers in a way that makes it easier to interpret into action. Numerical data is categorized into two categories :

Continuous Data :

It can be narrowed or categorized (Example: Height measurements).

Discrete Data :

This type of data is not “continuous” (Example: Number of cars or children’s a household has).

The type of visualization techniques that are used to represent numerical data visualization is Charts and Numerical Values. Examples are Pie Charts, Bar Charts, Averages, Scorecards, etc.

Categorical Data :

Categorical data is also known as Qualitative data. Categorical data is any data where data generally represents groups. It simply consists of categorical variables that are used to represent characteristics such as a person’s ranking, a person’s gender, etc. Categorical data visualization is all about depicting key themes, establishing connections, and lending context. Categorical data is classified into three categories:

Binary Data

In this, classification is based on positioning (Example: Agrees or Disagrees).

Nominal Data

In this, classification is based on attributes (Example: Male or Female).

Ordinal Data

In this, classification is based on ordering of information (Example: Timeline or processes).

Other benefits of data visualization include the following:

- the ability to absorb information quickly, improve insights and make faster decisions;
- an increased understanding of the next steps that must be taken to improve the organization;
- an improved ability to maintain the audience's interest with information they can understand;
- an easy distribution of information that increases the opportunity to share insights with everyone involved;
- It shows or visualizes data very clearly in an understandable manner.
- It encourages viewers to compare different pieces of data.
- It closely integrates statistical and verbal descriptions of data set.

Big Data Visualization Tools

Nowadays, there are many tools. Some of them are:

- Google Chart
- Tableau
- Microsoft Power BI
- D3 (Data-Driven Documents)

Some other popular techniques are as follows:

Line charts. This is one of the most basic and common techniques used. Line charts display how variables can change over time.

Area charts. This visualization method is a variation of a line chart; it displays multiple values in a time series -- or a sequence of data collected at consecutive, equally spaced points in time.

Scatter plots. This technique displays the relationship between two variables. A scatter plot takes the form of an x- and y-axis with dots to represent data points.

Treemaps. This method shows hierarchical data in a nested format. The size of the rectangles used for each category is proportional to its percentage of the whole. Treemaps are best used when multiple categories are present, and the goal is to compare different parts of a whole.

Population pyramids. This technique uses a stacked bar graph to display the complex social narrative of a population. It is best used when trying to display the distribution of a population.

Visual data analysis techniques

Data display represents data in a systematic manner, including information unit characteristics and variables. Data discovery techniques based on visualization enable company consumers to generate customized analytical opinions using disparate information sources.

Advanced analytics can be incorporated into techniques for the development on desktop and laptop or mobile devices like tablets and smartphones of interactive and animated Graphics.

Big data are large volumes, elevated speed and/or high-speed information sets that involve fresh types of handling to optimize processes, discover understanding and make choices. Data capture, storage, evaluation, sharing, searches and visualization face great challenges for big data. Visualization could be considered as "large information front end. There's no data visualization myth.

- It is important to visualize only excellent information: an easy and fast view can show something incorrect with information just like it helps to detect exciting patterns.
- Visualization always manifests the correct choice or intervention: visualization is not a substitute for critical thinking.
- Visualization brings assurance: data are displayed, not showing an exact image of what is essential. Visualization with various impacts can be manipulated.

Tables, diagrams, pictures and other intuitive display methods to represent the information are created using visualization methods.

Problems in Visualising Big Data

In the visual analysis, scalability and dynamics are two main difficulties. The visualization of big data (structured or unstructured) with diversity and heterogeneity is a big difficulty.

Big information does not make it simple to design a fresh visualization tool with effective indexing.

Unstructured information formats such as charts, lists, text, trees, and other information must be used by visualization schemes. Often large information has unstructured formats.

Due to the constraints on bandwidth and power consumption, visualization should step nearer to the data to effectively obtain significant information.

There are also the following problems for big data visualization:

Visual noise: Most items on the dataset are too related to each other. There are also the following issues when viewing large-scale information. Users can not split them on the display as distinct items.

Info loss: Visible data sets may be reduced, but information loss may occur.

Broad perception of images: data display techniques are restricted not only by aspect ratio and device resolution but also by physical perception limitations.

The elevated pace of changes in the picture: users view information and are unable to respond to the amount of changes in information or its intensity.

High-performance requirements: In static visualization it is hard to notice because of reduced demands for display velocity— high performance demands.

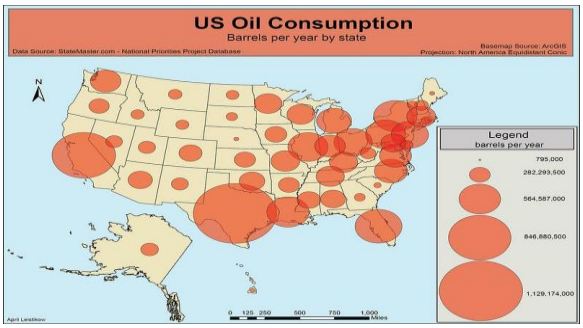
Visualization Techniques for Big Data

1. Word Clouds

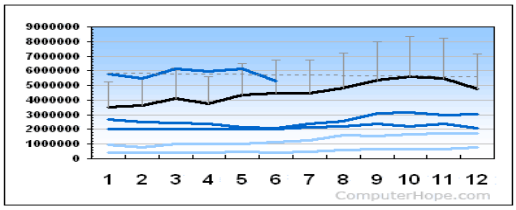
Word clouds work easy: the larger and bolder the word is in the term cloud the more a particular word is displayed in a source of text information (such as a lecture, newspaper post or database).

2. Symbol Maps

Symbol maps are merely maps shown over a certain length and latitude. You can rapidly create a strong visual with the “Marks” card at Tableau, which tells customers of their place information.



3. Line charts

Alternatively known as a row graph, a row graph is a graph of the information shown using a number of rows. Line diagrams show rows horizontally through the diagram, with the scores axis on the left hand of the diagram. 

4. Pie charts

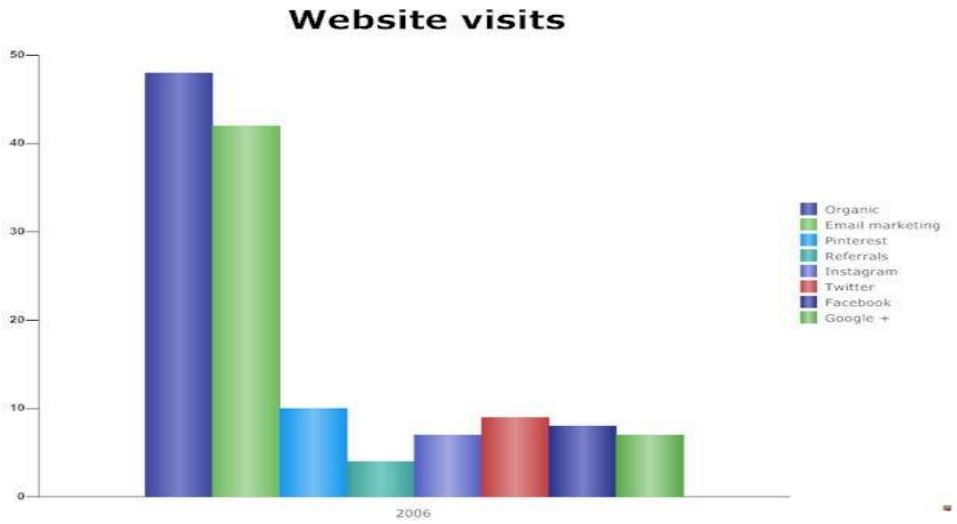
A diagram is a circular diagram, split into sections like wedges, which shows the amount. The complete valuation of each coin is 100% and is a proportional portion of the whole.

5. Bar Charts

A bar graph is a visual instrument which utilizes bars to match information between cities. bars are also called a bar chart or bar diagram. A bar chart can be executed horizontally or vertically. What we need to understand is that the longer the bar is, the more valuable it is. Two axes are the bar graphs. The horizontal axis (or x-axis) is shown on a graph of the vertical bar, as shown above. They are years in this instance. The vertical axis is the magnitude.

Bar charts have three main attributes:

- The graph shows classes on one axis and on the other a separate value. The objective is to demonstrate the connection between the two axes.
- Bar diagrams can also display over moment large information modifications.



Visualization Tools for Big Data

1. Power BI

Power BI is a company analysis option that enables you to view and share your information or integrate them into your app or blog. Connect to hundreds of information sources and live dashboards and accounts to take your information to life.

Microsoft Power BI is used to discover perspectives into the information of an organization. Power BI can communicate, convert and wash information into the data model and generate chart or diagram to display information graphics.

2. Tableau

Tableau has been utilized in the business intelligence industry as a strong and rapidly increasing information vision instrument. It makes it readily understandable to simplify raw information.

Data analysis with Tableau is very quick and the visualizations are in the shape of dashboards and tablets. The information produced using Tableau can be comprehended at every stage in an organisation by the specialist. It even enables a non-technical user a personalized dashboard to be created.

The best feature Tableau are

Data Blending

Real-time analysis

Collaboration of data

3. Grafana

Grafana is a metrics & visualizing package of open source analysis. It is used most frequently for moment serial data visualization for infrastructure and implementation analysis, but many use it in other areas including agricultural equipment, domestic automation, climate, and process control.

Grafana is a temporary information sequence display instrument. A graphical description can be obtained from a lot of gathered information of the position of a business or organization

Interaction Techniques

Data interaction techniques refer to various methods and functionalities that allow users to engage with and manipulate data effectively.

They enable users to explore, analyze, and understand data through visualization, filtering, sorting, searching, drill-down, grouping, interactive dashboards, tooltips, linked data, collaboration features, interactive documentation, natural language querying, and predictive analytics.

Interaction techniques essentially involve data entry and manipulation, and thus place greater emphasis on input than output.
Output is merely used to convey affordances and provide user feedback. The use of the term *input technique* further reinforces the central role of input. Conversely, techniques that mainly involve data exploration and thus place greater emphasis on output are called visualization techniques.

What are the benefits of Interactive Data Visualization?

The benefits of Interactive Data Visualization are listed below:

Identifying Causes and Trends Quickly

Today's 93% of human communication is visual, and it tells that human eyes are processing images 60,000 times more than the text-based data.

Relationships Between Tasks and Business Operations

By interacting with data to put the focus on specific metrics, decision-makers are able to compare specific throughout definable timeframes.

Telling Story Through Data

By allowing users to interact with data present in a clear visual manner, a data-intensive story becomes visible.

- Basic Idea: Dynamic generation of the visualizations or interaction with the visualization for a more effective exploration of the data.
- Overview
  - Data-to-Visualization Mapping
  - Projections
  - Filtering (Selection, Querying)
  - Linking & Brushing
  - Zooming
  - Detail on Demand

- Dynamic projection
  - dynamically change the projections to explore multi-dimensional data sets
  - projection pursuit, which finds well-separated clusters in scatterplot
- Interactive Filtering
  - browsing, can be difficult for big data sets
  - querying, need to specify a subset

- Zooming
- Distortion
  - e.g., fisheye view
- Brushing and linking
  - requires well-integrated system for visualization
  - selection from one visualization is fed into another one, selected instances highlighted in some way

Systems and Applications:

applications of data visualization

1.Healthcare Industries

A dashboard that visualises a patient's history might aid a current or new doctor in comprehending a patient's health. It might give faster care facilities based on illness in the event of an emergency.

Instead than sifting through hundreds of pages of information, data visualization may assist in finding trends.

Health care is a time-consuming procedure, and the majority of it is spent evaluating prior reports. By boosting response time, data visualisation provides a superior selling point. It gives matrices that make analysis easier, resulting in a faster reaction time.

2.Business intelligence

When compared to local options, cloud connection can provide the cost-effective "heavy lifting" of processor-intensive analytics, allowing users to see bigger volumes of data from numerous sources to help speed up decision-making.

Because such systems can be diverse, comprised of multiple components, and may use their own data storage and interfaces for access to stored data, additional integrated tools, such as those geared toward business intelligence (BI), help provide a cohesive view of an organization's entire data system (e.g., web services, databases, historians, etc.).

Multiple datasets can be correlated using analytics/BI tools, which allow for searches using a common set of filters and/or parameters. The acquired data may then be displayed in a standardised manner using these technologies, giving logical "shaping" and better comparison grounds for end users.

3.Military:

It's a matter of life and death for the military; having clarity of actionable data is critical, and taking the appropriate action requires having clarity of data to pull out actionable insights.

The adversary is present in the field today, as well as posing a danger through digital warfare and cybersecurity. It is critical to collect data from a variety of sources, both organised and unstructured. The volume of data is enormous, and data visualisation technologies are essential for rapid delivery of accurate information in the most condensed form feasible. A greater grasp of past data allows for more accurate forecasting.

Dynamic Data Visualization aids in a better knowledge of geography and climate, resulting in a more effective approach. The cost of military equipment and tools is extremely significant; with bar and pie charts, analysing current inventories and making purchases as needed is simple.

4.Finance Industries

For exploring/explaining data of linked customers, understanding consumer behaviour, having a clear flow of information, the efficiency of decision making, and so on, data visualisation tools are becoming a requirement for financial sectors.

For associated organisations and businesses, data visualisation aids in the creation of patterns, which aids in better investment strategy. For improved business prospects, data visualisation emphasises the most recent trends.

5.Data science

Data scientists generally create visualisations for their personal use or to communicate information to a small group of people. Visualization libraries for the specified programming languages and tools are used to create the visual representations.

Open source programming languages, such as Python, and proprietary tools built for complicated data analysis are commonly used by data scientists and academics. These data scientists and researchers use data visualisation to better comprehend data sets and spot patterns and trends that might otherwise go undiscovered.

6.Marketing

In marketing analytics, data visualisation is a boon. We may use visuals and reports to analyse various patterns and trends analysis, such as sales analysis, market research analysis, customer analysis, defect analysis, cost analysis, and forecasting. These studies serve as a foundation for marketing and sales.

Visual aids can assist your audience grasp your main message by visually engaging them and visually engaging them. The major advantage of visualising data is that it can communicate a point faster than a boring spreadsheet.

In b2b firms, data-driven yearly reports and presentations don't fulfil the needs of people who are seeing the information. They are unable to grasp the art of engaging with their audience in a meaningful or memorable manner. Your audience will be more interested in your facts if you present them as visual statistics, and you will be more inclined to act on your discoveries.

7.Food delivery apps

When you place an order for food on your phone, it is given to the nearest delivery person. There is a lot of math involved here, such as the distance between the delivery executive's present position and the restaurant, as well as the time it takes to get to the customer's location.

Customer orders, delivery location, GPS service, tweets, social media messages, verbal comments, pictures, videos, reviews, comparative analyses, blogs, and updates have all become common ways of data transmission.

Users may obtain data on average wait times, delivery experiences, other records, customer service, meal taste, menu options, loyalty and reward point programmes, and product stock and inventory data with the help of the data.

8.Real estate business

Brokers and agents seldom have the time to undertake in-depth research and analysis on their own. Showing a buyer or seller comparable home prices in their neighbourhood on a map, illustrating average time on the market, creating a sense of urgency among prospective buyers and managing sellers' expectations, and attracting viewers to your social media sites are all examples of common data visualisation applications.

If a chart is difficult to understand, it is likely to be misinterpreted or disregarded. It is also seen to be critical to offer data that is as current as feasible. The market may not alter overnight, but if the data is too old, seasonal swings and other trends may be overlooked.

Clients will be pulled to the graphics and to you as a broker or agent if they perceive that you know the market. If you display data in a compelling and straightforward fashion, they will be drawn to the graphics and to you as a broker or agent.

9.Education

Users may visually engage with data, answer questions quickly, make more accurate, data-informed decisions, and share their results with others using intuitive, interactive dashboards.

The ability to monitor students' progress throughout the semester, allowing advisers to act quickly with outreach to failing students. When they provide end users access to interactive, self-service analytic visualisations as well as ad hoc visual data discovery and exploration, they make quick insights accessible to everyone – even those with little prior experience with analytics.

10.E-commerce

In e-commerce, any chance to improve the customer experience should be taken. The key to running a successful internet business is getting rapid insights. This is feasible with data visualisation because crossing data shows features that would otherwise be hidden.

Your marketing team may use data visualisation to produce excellent content for your audience that is rich in unique information. Data may be utilised to produce attractive narrative through the use of infographics, which can easily and quickly communicate findings.

Patterns may be seen all throughout the data. You can immediately and readily detect them if you make them visible. These behaviours indicate a variety of consumer trends, providing you with knowledge to help you attract new clients and close sales.