In our data-rich age, understanding how to analyze and extract true meaning from the digital insights available to our business is one of the primary drivers of success.

Despite the colossal volume of data we create every day, a mere 0.5% is actually analyzed and used for [data discovery](https://www.datapine.com/blog/what-are-data-discovery-tools/), improvement, and intelligence. While that may not seem like much, considering the amount of digital information we have at our fingertips, half a percent still accounts for a huge amount of data.

With so much data and so little time, knowing how to collect, curate, organize, and make sense of all of this potentially business-boosting information can be a minefield – but [online data analysis](https://www.datapine.com/data-analysis-tools) is the solution.

To help you understand the potential of analysis, the meaning, and how you can use it to enhance your business practices, we will answer a host of important analytical questions. Not only will we explore data analysis methods and techniques, but we’ll also look at different types of data analysis while demonstrating how to perform analysis in the real world with a 15-step blueprint for success.

**What Is Data Analysis?**

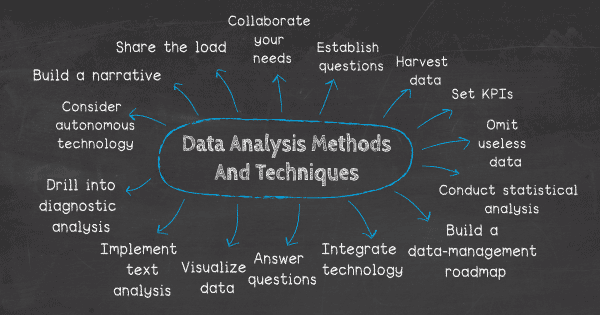
Data analysis is a process that relies on methods and techniques to taking raw data, mining for insights that are relevant to the business’s primary goals, and drilling down into this information to transform metrics, facts, and figures into initiatives for improvement.

There are various methods for data analysis, largely based on two core areas: *quantitative data analysis methods*and*data analysis methods in qualitative research.*

To explain the key differences between qualitative and quantitative data, here’s a video for your viewing pleasure:

Gaining a better understanding of different techniques for data analysis, and methods in quantitative research as well as qualitative insights, will give your information analyzing efforts a more clearly defined direction, so it’s worth taking the time to allow this particular knowledge to sink in. Additionally, you will be able to create a comprehensive [analytical report](https://www.datapine.com/blog/analytical-report-example-and-template/) that will skyrocket your analysis processes.

**How To Analyze Data: Take Advantage Of The Top 15 Data Analysis Methods & Techniques**



Now that we’ve answered the question, ‘what is data analysis?’, considered the different types of analysis methods, it’s time to dig deeper into how to do data analysis by working through these 15 essential elements.

**1. Collaborate your needs**

Before you begin to analyze your data or drill down into any analysis techniques, it’s crucial to sit down collaboratively with all key stakeholders within your organization, decide on your primary campaign or strategic goals, and gain a fundamental understanding of the types of insights that will best benefit your progress or provide you with the level of vision you need to evolve your organization.

**2. Establish your questions**

Once you’ve outlined your core objectives, you should consider which questions will need answering to help you achieve your mission. This is one of the most important data analytics techniques as it will shape the very foundations of your success.

To help you ask the right things and ensure your data works for you, you have to ask the right [data analysis questions](https://www.datapine.com/blog/data-analysis-questions/).

**3. Harvest your data**

After giving your data analytics methodology real direction and knowing which questions need answering to extract optimum value from the information available to your organization, you should decide on your most valuable data sources and start collecting your insights – the most fundamental of all data analysis techniques.

**4. Set your KPIs**

Once you’ve set your data sources, started to gather the raw data you consider to offer potential value, and established clear-cut questions you want your insights to answer, you need to set a host of key performance indicators (KPIs) that will help you track, measure, and [shape your progress](https://www.mindtools.com/pages/article/improving-business-processes.htm) in a number of key areas.

KPIs are critical to both analysis methods in qualitative and quantitative research. This is one of the primary methods of analyzing data you certainly shouldn’t overlook.

To help you set the best possible KPIs for your initiatives and activities, explore our collection of [key performance indicator examples](https://www.datapine.com/kpi-examples-and-templates/).

**5. Omit useless data**

Having bestowed your data analysis techniques and methods with true purpose and defined your mission, you should explore the raw data you’ve collected from all sources and use your KPIs as a reference for chopping out any information you deem to be useless.

Trimming the informational fat is one of the most crucial methods of data analysis as it will allow you to focus your analytical efforts and squeeze every drop of value from the remaining ‘lean’ information.

Any stats, facts, figures, or metrics that don’t align with your business goals or fit with your [KPI management](https://www.datapine.com/blog/kpi-management-and-best-practices/) strategies should be eliminated from the equation.

**6. Conduct statistical analysis**

One of the most pivotal types of analysis is statistics.

This kind of analysis method focuses on aspects including cluster, cohort, regression, factor, and neural networks and will ultimately give your data analysis methodology a more logical direction.

Here is a quick glossary of these vital statistical analysis terms for your reference:

* **Cluster:**The action of grouping a set of elements in a way that said elements are more similar (in a particular sense) to each other than to those in other groups – hence the term ‘cluster’. This approach is used to provide additional context to a trend or dataset.
* **Cohort:**A subset of behavioral analytics that takes insights from a given data set (e.g. a web application or CMS) and instead of looking at everything as one wider unit, each element is broken down into related groups. By using this analysis methodology, it’s possible to gain a wealth of insight into consumer needs or a firm understanding of a wider target group.
* **Regression:**A definitive set of statistical processes centered on estimating the relationships among particular variables to gain a deeper understanding of particular trends or patterns. This is an analysis methodology that is incredibly powerful when focused on predictive analytics.
* **Factor:**A statistical practice utilized to describe variability among observed, correlated variables in terms of a potentially lower number of unobserved variables called ‘factors’. The aim here is to uncover independent latent variables. An ideal method of analysis for streamlining specific segments of data.
* **Neural networks:**A neural network is a form of machine learning (ML) focused on predicting the outcome of a specific variable. This concept is far too comprehensive to summarize – but this [explanation](https://www.techradar.com/uk/news/what-is-a-neural-network) will help paint you a fairly comprehensive picture. Later in the article, we will also focus more on this term.
* **Data modeling:**A method of analysis that is the umbrella term for engineering metrics and insights for additional value, direction, and context by using a mix of strategies and techniques. When considering how to analyze data, adopting a data modeling mindset is essential to success - as such, it’s an area that is worth exploring in greater detail.
* **Predictive analytics:**Again, while this is a broad term (and something we will revisit in due course), when you’re looking at how to do analysis the right way, you should ensure this term is firmly ingrained in your mind. Predictive analytical techniques are essential as they will empower you to spot invaluable trends as they unfold and [develop initiatives](https://hbr.org/2016/05/where-predictive-analytics-is-having-the-biggest-impact) that will not only enhance your various operational processes but also help you gain an all-important edge on the competition.

**7. Build a data management roadmap**

While, at this point, this particular step is optional (you will have already gained a wealth of insight and formed a fairly sound strategy by now), creating a data governance roadmap will help your data analysis methods and techniques become successful on a more sustainable basis. These roadmaps, if developed properly, are also built so they can be tweaked and scaled over time.

Invest ample time in [developing a roadmap](https://inform.tmforum.org/features-and-analysis/2016/08/6-step-roadmap-good-data-governance/) that will help you store, manage, and handle your data internally, and you will make your analysis techniques all the more fluid and functional – one of the most powerful types of data analysis methods available today.

**8. Integrate technology**

There are many ways to analyze data, but one of the most vital aspects of analytical success in a business context is integrating the right [decision support software](https://www.datapine.com/decision-support-system) and technology.

Robust analysis platforms will not only allow you to pull critical data from your most valuable sources while working with dynamic KPIs that will offer you actionable insights; it will also present the information in a digestible, visual, interactive format from one central, [live dashboard](https://www.datapine.com/live-dashboards). A data analytics methodology you can count on.

By integrating the right technology for your statistical method data analysis and core data analytics methodology, you’ll avoid fragmenting your insights, saving you time and effort while allowing you to enjoy the maximum value from your business’s most valuable insights.

For a look at the power of software for the purpose of analysis and to enhance your methods of analyzing data, glance over our selection of [dashboard examples](https://www.datapine.com/dashboard-examples-and-templates/).

**9. Answer your questions**

By considering each of the above efforts, working with the right technology, and fostering a [cohesive internal culture](https://www.datamaticsinc.com/5-ways-improve-workplace-culture/) where everyone buys into the different ways to analyze data as well as the power of digital intelligence, you will swiftly start to answer your most burning business questions. Arguably, the best way to make your data concepts accessible across the organization is through data visualization.

**10. Visualize your data**

[Online data visualization](https://www.datapine.com/data-visualization-tools) is a powerful tool as it lets you tell a story with your metrics, allowing users across the business to extract meaningful insights that aid business evolution – and it covers all the different ways to analyze data.

The purpose of data analysis is to make your entire organization more informed and intelligent, and with the right platform or dashboard, this is simpler than you think, as demonstrated by our [marketing dashboard](https://www.datapine.com/dashboard-examples-and-templates/marketing).

**Primary KPIs:**

* Bounce Rate
* Average Sessions Duration
* Goal Conversion Rates
* Landing Page Conversion Rates

Delving deeper than the user data served up by Google Analytics (GA) alone, this visual, dynamic, and interactive [online dashboard](https://www.datapine.com/online-dashboard) displays the behavior of your users and site visitors, presenting a wealth of metrics based on KPIs that explore session duration, page bounce rates, landing page conversion rates, and goal conversion rates, making a comprehensive [marketing report](https://www.datapine.com/blog/daily-weekly-monthly-marketing-report-examples/) that a user can additionally interact with and adjust.

This centralized mix of information provides a real insight into how people interact with your website, content, and offerings, helping you to identify weaknesses, capitalize on strengths, and make[data-driven decisions](https://www.datapine.com/blog/data-driven-decision-making-in-businesses/) that can benefit the business exponentially.

**11. Implement text analysis**

A vast quantity of data that businesses collect is unstructured. While having access to a breadth of data-driven insight is essential to enhancing your business intelligence (BI) capabilities, without implementing techniques of data analysis to give your metrics structure, you will only ever be scraping the surface.

Text analysis, also known in the industry as text mining, is the process of taking large sets of textual data and arranging it in a way that makes it easier to manage. By working through this cleansing process in stringent detail, you will be able to extract the data that is truly relevant to your business and use it to develop actionable insights that will propel you forward.

Modern analysis tools and techniques accelerate the process of text analytics, helping to collect and curate insights in a way that is efficient and results-driven.

Collectively, we create a colossal 2.5 quintillion bytes of digital data every single day, and a large segment is text-based. By investing in [data analyst tools](https://www.datapine.com/articles/data-analyst-tools-software) and techniques that will help you extract insight from various word-based data sources, including product reviews, articles, social media communications, and survey responses, you will gain invaluable insights into your audience, as well as their needs, preferences, and pain points.

By gaining this level of knowledge, you will be able to create campaigns, services, and communications that meet the needs of your prospects on a personal level, growing your audience while boosting customer retention. One of the most effective data analysis tools and techniques you will ever invest in.

**12. Drill into diagnostic analysis**

When it comes to lessons on how to do analysis, drilling down into diagnostic analysis is essential. Designed to provide direct and actionable answers to specific questions, this is one of the world’s most important methods in research, among its other key organizational functions such as [retail analytics](https://www.datapine.com/retail-analytics), e.g.

A powerful form of data discovery, diagnostic data analytics empowers analysts and business executives by helping them gain a firm contextual understanding of *why something happened.*If you know why something happened as well as *how it happened,*you will be able to pinpoint the exact ways of tackling the issue or challenge. To gain a practical understanding, it’s vital that you gain a foundational knowledge of the following two areas:

**a) Predictive analytics**

If you understand why a trend, pattern, or event happened through data, you will be able to develop an informed projection of how things may unfold in particular areas of the business. By doing so, you will be able to formulate initiatives or launch campaigns ahead of the curve, beating your competitors to the punch. Also, if you can use the predictive aspect of diagnostic analytics to your advantage, you will be able to prevent potential problems or inefficiencies from spiraling out of control, nipping potential problems in the bud. There are [BI reporting tools](https://www.datapine.com/bi-reporting) that have predictive analytics options already implemented within them, but also made user-friendly so that you don't need to calculate anything manually or perform the robust and advanced analysis yourself.

**b) Prescriptive analytics**

Another of the most effective data analysis methods in research, prescriptive data techniques cross over from predictive analysis in the way that it revolves around using patterns or trends to develop responsive, practical business strategies.

By drilling down into prescriptive analysis, you will play an active role in the data consumption process by taking well-arranged sets of visual data and using it as a powerful fix to emerging issues in a number of key business areas, including marketing, sales, customer experience, HR, fulfillment, finance, [logistics analytics](https://www.datapine.com/logistics-analytics), and others.

**13. Consider autonomous technology**

Autonomous technologies, such as artificial intelligence (AI) and machine learning (ML), play a significant role in the advancement of understanding how to analyze data more effectively.

Gartner predicts that by 2021, [80%](https://www.gartner.com/en/documents/3875982/100-data-and-analytics-predictions-through-2022) of emerging technologies will be developed with AI foundations. This is a testament to the ever-growing power and value of autonomous technologies. At present, neural networks and intelligence alarms are driving the autonomous revolution in the world of data-driven analytics.

One of the techniques of data analysis of the modern age, intelligent alarms provide automated signals based on particular commands or occurrences within a dataset. For example, if you’re monitoring [supply chain KPIs](https://www.datapine.com/blog/supply-chain-metrics-and-kpis/), you could set an intelligent alarm to trigger when invalid or low-quality data appears. By doing so, you will be able to drill down deep into the issue and fix it swiftly and effectively.

A neural network is a branch of machine learning - a form of data-driven analytics that attempts, with minimal intervention, to understand how the human brain would process insights and predict values. Neural networks learn from each and every data transaction, meaning that they evolve and advance over time.

**14. Build a narrative**

Now that we’ve discussed and explored the technical applications of data-driven analysis, we’re going to look at how you can bring all of these elements together in a way that will benefit your business - starting with a little something called data storytelling.

The human brain responds incredibly well to strong stories or narratives. Once you’ve cleansed, shaped, and visualized your most invaluable data using various [BI dashboard tools](https://www.datapine.com/bi-dashboard-tools), you should strive to tell a story - one with a clear-cut beginning, middle, and end.

By doing so, you will make your analytical efforts more accessible, digestible, and universal, empowering more people within your organization to use your discoveries to their actionable advantage.

**15. Share the load**

Last but certainly not least in our advice on how to make data analysis work for your business, we discuss *sharing the load*.

Expanding on our previous point, by using technical methods to give your data more shape and meaning, you will be able to provide a platform for wider access to data-driven insights.

If you work with the right tools and dashboards, you will be able to present your metrics in a digestible, value-driven format, allowing almost everyone in the organization to connect with and use relevant data to their advantage.

Modern data dashboards consolidate data from various sources, providing access to a wealth of insights in one centralized location, no matter if you need to monitor [recruitment metrics](https://www.datapine.com/blog/recruitment-metrics-and-dashboards/) or generate reports that need to be sent across numerous departments. Moreover, these cutting-edge tools offer access to dashboards from a multitude of devices, meaning that everyone within the business can connect with practical insights remotely - and share the load.

Once everyone is able to work with a data-driven mindset, you will catalyze the success of your business in ways you never thought possible. And when it comes to knowing how to make data analysis, this kind of collaborative approach is essential.

**Data Analysis In The Big Data Environment**

Big data is invaluable to today’s businesses, and by using different methods for data analysis, it’s possible to view your data in a way that can help you turn insight into positive action.

To inspire your efforts and put the importance of big data into context, here are some insights that you should know – facts that will help shape your big data analysis techniques.

* A [10% boost](http://www.baselinemag.com/analytics-big-data/slideshows/surprising-statistics-about-big-data.html) in data accessibility will result in more than $65 million extra net income for your average Fortune 1000 company.
* [90%](https://www.mediapost.com/publications/article/291358/90-of-todays-data-created-in-two-years.html) of the world’s big data was created in the past five years.
* According to Accenture, [79%](https://www.accenture.com/gb-en/_acnmedia/Accenture/Conversion-Assets/DotCom/Documents/Global/PDF/Industries_14/Accenture-Big-Data-POV.pdf) of notable business executives agree that companies that fail to embrace big data will lose their competitive position and could face extinction. Moreover, 83% of business execs have implemented big data projects to gain a competitive edge.

Data analysis concepts may come in many forms, but fundamentally, any solid methodology will help to make your business more streamlined, cohesive, insightful and successful than ever before.

To summarize, here are the top 15 steps for data analysis techniques and methods:

1. Collaborate your needs
2. Establish your questions
3. Harvest your data
4. Set your KPIs
5. Omit useless data
6. Conduct statistical analysis
7. Build a data management roadmap
8. Integrate technology
9. Answer your questions
10. Visualize your data
11. Implement text analysis
12. Drill into diagnostic analysis
13. Consider autonomous technology
14. Build a narrative
15. Share the load

## Measuring Quantitative Data

Quantitative analysis methods rely on the ability to accurately count and interpret data based on hard facts. Our first three methods for upping your analysis game will focus on quantitative data:

**1. Regression analysis**

Regression studies are excellent tools when you need to make predictions and forecast future trends. Regressions measure the relationship between a dependent variable (what you want to measure) and an independent variable (the data you use to predict the dependent variable). While you can only have one dependent variable, you can have a nearly limitless number of independent ones. Regressions also help you uncover areas in your operations that can be optimized by highlighting trends and relationships between factors.

**2. Hypothesis testing**

Also known as “T Testing,” this analysis method lets you compare the data you have against hypotheses and assumptions you’ve made about your operations. It also helps you forecast how decisions you could make will affect your organization. T Testing lets you compare two variables to find a correlation and base decisions on the findings. For instance, you may assume that more hours of work are equivalent to higher productivity. Before implementing longer work hours, it’s important to ensure there’s a real connection to avoid an unpopular policy.

**3. Monte Carlo simulation**

As one of the most popular ways to calculate the effect of unpredictable variables on a specific factor, Monte Carlo simulations use probability modeling to help predict risk and uncertainty. To test a hypothesis or scenario, a Monte Carlo simulation will use random numbers and data to stage a variety of possible outcomes to any situation based on any results. This is an incredibly useful tool across a variety of fields including project management, [finance](https://www.sisense.com/dashboard-examples/finance/), engineering, logistics, and more. By testing a variety of possibilities, you can understand how random variables could affect your plans and projects.

Measuring Qualitative Data

Unlike quantitative data, qualitative information requires moving away from pure statistics and toward more subjective approaches. Yet, you can still extract useful data by employing different data analysis techniques depending on your demands. Our final two techniques focus on qualitative data:

**4. Content analysis**

This method helps to understand the overall themes that emerge in qualitative data. Using techniques like color coding specific themes and ideas helps parse textual data to find the most common threads. Content analyses can work well when dealing with data such as user feedback, interview data, open-ended surveys, and more. This can help identify the most important areas to focus on for improvement.

**5. Narrative analysis**

This kind of analysis focuses on the way stories and ideas are communicated throughout a company and can help you better understand the organizational culture. This might include interpreting how employees feel about their jobs, how customers perceive an organization, and how operational processes are viewed. It can be useful when contemplating changes to corporate culture or planning new marketing strategies.

There is no gold standard for statistical analysis or right way to do it. The method you choose should always reflect the data you’ve collected, and the type of insights you want to extract. Matching the right data and analysis helps uncover better insights to optimize your organization.

What is the first thing that comes to mind when we see data? The first instinct is to find patterns, connections, and relationships. We look at the data to find meaning in it.

Similarly, in research, once data is collected, the next step is to get insights from it. For example, if a clothing brand is trying to identify the latest trends among young women, the brand will first reach out to young women and ask them questions relevant to the research objective. After collecting this information, the brand will analyze that data to identify patterns — for example, it may discover that most young women would like to see more variety of jeans.

Data analysis is how researchers go from a mass of data to meaningful insights. There are many different data analysis methods, depending on the type of research. Here are a few methods you can use to analyze quantitative and qualitative data.

**Analyzing Quantitative Data**

**Data Preparation**

The first stage of analyzing data is data preparation, where the aim is to convert raw data into something meaningful and readable. It includes four steps:

**Step 1: Data Validation**

The purpose of data validation is to find out, as far as possible, whether the data collection was done as per the pre-set standards and without any bias. It is a four-step process, which includes…

* **Fraud**, to infer whether each respondent was actually interviewed or not.
* **Screening**, to make sure that respondents were chosen as per the research criteria.
* **Procedure**, to check whether the data collection procedure was duly followed.
* **Completeness**, to ensure that the interviewer asked the respondent all the questions, rather than just a few required ones.

To do this, researchers would need to pick a random sample of completed surveys and validate the collected data. (Note that this can be time-consuming for surveys with lots of responses.) For example, imagine a survey with 200 respondents split into 2 cities. The researcher can pick a sample of 20 random respondents from each city. After this, the researcher can reach out to them through email or phone and check their responses to a certain set of questions.

**Step 2: Data Editing**

Typically, large data sets include errors. For example, respondents may fill fields incorrectly or skip them accidentally. To make sure that there are no such errors, the researcher should conduct [basic data checks](https://blog.socialcops.com/academy/resources/4-data-checks-clean-data/), [check for outliers](https://blog.socialcops.com/academy/resources/find-deal-outliers-data-set/), and edit the raw research data to identify and clear out any data points that may hamper the accuracy of the results.

For example, an error could be fields that were left empty by respondents. While editing the data, it is important to make sure to remove or fill all the empty fields. ([Here are 4 methods to deal with missing data.](https://blog.socialcops.com/academy/resources/4-methods-missing-data/))

**Step 3: Data Coding**

This is one of the most important steps in data preparation. It refers to grouping and assigning values to responses from the survey.

For example, if a researcher has interviewed 1,000 people and now wants to find the average age of the respondents, the researcher will create age buckets and categorize the age of each of the respondent as per these codes. (For example, respondents between 13-15 years old would have their age coded as 0, 16-18 as 1, 18-20 as 2, etc.)

Then during analysis, the researcher can deal with simplified age brackets, rather than a massive range of individual ages.

**Quantitative Data Analysis Methods**

After these steps, the data is ready for analysis. The two most commonly used quantitative data analysis methods are descriptive statistics and inferential statistics.

**Descriptive Statistics**

Typically descriptive statistics (also known as descriptive analysis) is the first level of analysis. It helps researchers summarize the data and find patterns. A few commonly used descriptive statistics are:

* **Mean**: numerical average of a set of values.
* **Median**: midpoint of a set of numerical values.
* **Mode**: most common value among a set of values.
* **Percentage**: used to express how a value or group of respondents within the data relates to a larger group of respondents.
* **Frequency**: the number of times a value is found.
* **Range**: the highest and lowest value in a set of values.

Descriptive statistics provide absolute numbers. However, they do not explain the rationale or reasoning behind those numbers. Before applying descriptive statistics, it’s important to think about which one is best suited for your research question and what you want to show. For example, a percentage is a good way to show the gender distribution of respondents.

Descriptive statistics are most helpful when the research is limited to the sample and does not need to be generalized to a larger population. For example, if you are comparing the percentage of children vaccinated in two different villages, then descriptive statistics is enough.

Since descriptive analysis is mostly used for analyzing single variable, it is often called univariate analysis.

**Analyzing Qualitative Data**

Qualitative data analysis works a little differently from quantitative data, primarily because qualitative data is made up of words, observations, images, and even symbols. Deriving absolute meaning from such data is nearly impossible; hence, it is mostly used for exploratory research. While in quantitative research there is a clear distinction between the data preparation and data analysis stage, analysis for qualitative research often begins as soon as the data is available.

**Data Preparation and Basic Data Analysis**

Analysis and preparation happen in parallel and include the following steps:

1. **Getting familiar with the data**: Since most qualitative data is just words, the researcher should start by reading the data several times to get familiar with it and start looking for basic observations or patterns. This also includes transcribing the data.
2. **Revisiting research objectives**: Here, the researcher revisits the research objective and identifies the questions that can be answered through the collected data.
3. **Developing a framework**: Also known as coding or indexing, here the researcher identifies broad ideas, concepts, behaviors, or phrases and assigns codes to them. For example, coding age, gender, socio-economic status, and even concepts such as the positive or negative response to a question. Coding is helpful in structuring and labeling the data.
4. **Identifying patterns and connections**: Once the data is coded, the research can start identifying themes, looking for the most common responses to questions, identifying data or patterns that can answer research questions, and finding areas that can be explored further.

**Qualitative Data Analysis Methods**

Several methods are available to analyze qualitative data. The most commonly used data analysis methods are:

* **Content analysis**: This is one of the most common methods to analyze qualitative data. It is used to analyze documented information in the form of texts, media, or even physical items. When to use this method depends on the research questions. Content analysis is usually used to analyze responses from interviewees.
* **Narrative analysis**: This method is used to analyze content from various sources, such as interviews of respondents, observations from the field, or surveys. It focuses on using the stories and experiences shared by people to answer the research questions.
* **Discourse analysis**: Like narrative analysis, discourse analysis is used to analyze interactions with people. However, it focuses on analyzing the social context in which the communication between the researcher and the respondent occurred. Discourse analysis also looks at the respondent’s day-to-day environment and uses that information during analysis.
* **Grounded theory**: This refers to using qualitative data to explain why a certain phenomenon happened. It does this by studying a variety of similar cases in different settings and using the data to derive causal explanations. Researchers may alter the explanations or create new ones as they study more cases until they arrive at an explanation that fits all cases.

These methods are the ones used most commonly. However, other data analysis methods, such as conversational analysis, are also available.

**What is Data Analysis? Definition & Example**

The systematic application of statistical and logical techniques to describe the data scope, modularize the data structure, condense the data representation, illustrate via images, tables, and graphs, and evaluate statistical inclinations, probability data, to derive meaningful conclusions, is known as Data Analysis. These analytical procedures enable us to induce the underlying inference from data by eliminating the unnecessary chaos created by the rest of it. The generation of data is a continual process; this makes data analysis a continuous, iterative process where the collection and performing data analysis simultaneously. Ensuring data integrity is one of the essential components of data analysis.

There are various examples where data analysis is used ranging from transportation, risk and fraud detection, customer interaction, city planning healthcare, web search, digital advertisement, and more.

Considering the example of healthcare as we have noticed recently that with the outbreak of the pandemic Coronavirus hospitals are facing the challenge of coping up with the pressure in treating as many patients as possible, considering data analysis allows to monitor machine and data usage in such scenarios to achieve efficiency gain.

Before diving any more in-depth, make the following pre-requisites for proper Data Analysis:

* Ensure availability of the necessary analytical skills
* Ensure appropriate implementation of data collection methods and analysis.
* Determine the statistical significance
* Check for inappropriate analysis
* Ensure the presence of legitimate and unbiased inference
* Ensure the reliability and validity of data, data sources, data analysis methods, and inferences derived.
* Account for the extent of analysis

**Data Analysis Methods**

There are two main methods of Data Analysis:

**1. Qualitative Analysis**

This approach mainly answers questions such as ‘why,’ ‘what’ or ‘how.’ Each of these questions is addressed via quantitative techniques such as questionnaires, attitude scaling, standard outcomes, and more. Such kind of analysis is usually in the form of texts and narratives, which might also include audio and video representations.

**2. Quantitative Analysis**

Generally, this analysis is measured in terms of numbers. The data here present themselves in terms of measurement scales and extend themselves for more statistical manipulation.

The other techniques include:

**3. Text analysis**

Text analysis is a technique to analyze texts to extract machine-readable facts. It aims to create structured data out of free and unstructured content. The process consists of slicing and dicing heaps of unstructured, heterogeneous files into easy-to-read, manage and interpret data pieces. It is also known as text mining, text analytics, and information extraction.

The ambiguity of human languages is the biggest challenge of text analysis. For example, the humans know that “Red Sox Tames Bull” refers to a baseball match, but if this text is fed to a computer without background knowledge, then it would generate several linguistically valid interpretations, and sometimes people not interested in baseball might have trouble understanding it too.

**4. Statistical analysis**

Statistics involves data collection, interpretation, and validation. Statistical analysis is the technique of performing several statistical operations to quantify the data and apply statistical analysis. Quantitative data involves descriptive data like surveys and observational data. It is also called a descriptive analysis. It includes various tools to perform statistical data analysis such as SAS (Statistical Analysis System), SPSS (Statistical Package for the Social Sciences), Stat soft, and more.

**5. Diagnostic analysis**

The diagnostic analysis is a step further to statistical analysis to provide more in-depth analysis to answer the questions. It is also referred to as root cause analysis as it includes processes like data discovery, mining and drill down and drill through.

The diagnostic analysis is a step further to statistical analysis to provide more in-depth analysis to answer the questions. It is also referred to as root cause analysis as it includes processes like data discovery, mining and drill down and drill through.

The functions of diagnostic analytics fall into three categories:

* **Identify anomalies:**After performing statistical analysis, analysts are required to identify areas requiring further study as such data raise questions that cannot be answered by looking at the data.
* **Drill into the Analytics (discovery):**Identification of the data sources helps analysts explain the anomalies. This step often requires analysts to look for patterns outside the existing data sets and requires pulling in data from external sources, thus identifying correlations and determining if any of them are causal in nature.
* **Determine Causal Relationships:**Hidden relationships are uncovered by looking at events that might have resulted in the identified anomalies. Probability theory, regression analysis, filtering, and time-series data analytics can all be useful for uncovering hidden stories in the data.

**6. Predictive analysis**

Predictive analysis uses historical data and feds it into the machine learning model to find critical patterns and trends. The model is applied to the current data to predict what would happen next. Many organizations prefer it because of its various advantages like volume and type of data, faster and cheaper computers, easy-to-use software, tighter economic conditions, and a need for competitive differentiation.

The following are the common uses of predictive analysis:

* **Fraud Detection:**Multiple analytics methods improves pattern detection and prevents criminal behavior.
* **Optimizing Marketing Campaigns:**Predictive models help businesses attract, retain, and grow their most profitable customers. It also helps in determining customer responses or purchases, promoting cross-sell opportunities.
* **Improving Operations:**The use of predictive models also involves forecasting inventory and managing resources. For example, airlines use predictive models to set ticket prices.
* **Reducing Risk:** Credit score that is used to assess a buyer’s likelihood of default for purchases is generated by a predictive model that incorporates all data relevant to a person’s creditworthiness. Other risk-related uses include insurance claims and collections.

**7. Prescriptive Analysis**

Prescriptive analytics suggests various courses of action and outlines what the potential implications could be reached after predictive analysis. Prescriptive analysis generating automated decisions or recommendations requires specific and unique algorithmic and clear direction from those utilizing the analytical techniques.

**Data Analysis Process**

Once you set out to collect data for analysis, you are overwhelmed by the amount of information that you find to make a clear, concise decision. With so much data to handle, you need to identify relevant data for your analysis to derive an accurate conclusion and make informed decisions. The following simple steps help you identify and sort out your data for analysis.

1. Data Requirement Specification - define your scope:

* + Define short and straightforward questions, the answers to which you finally need to make a decision.
  + Define measurement parameters
  + Define which parameter you take into account and which one you are willing to negotiate.
  + Define your unit of measurement. Ex – Time, Currency, Salary, and more.

2. Data Collection

* + Gather your data based on your measurement parameters.
  + Collect data from databases, websites, and many other sources. This data may not be structured or uniform, which takes us to the next step.

3. Data Processing

* + Organize your data and make sure to add side notes, if any.
  + Cross-check data with reliable sources.
  + Convert the data as per the scale of measurement you have defined earlier.
  + Exclude irrelevant data.

4. Data Analysis

* + Once you have collected your data, perform sorting, plotting, and identifying correlations.
  + As you manipulate and organize your data, you may need to traverse your steps again from the beginning, where you may need to modify your question, redefine parameters, and reorganize your data.
  + Make use of the different tools available for data analysis.

5. Infer and Interpret Results

* + Review if the result answers your initial questions
  + Review if you have considered all parameters for making the decision
  + Review if there is any hindering factor for implementing the decision.
  + Choose data visualization techniques to communicate the message better. These visualization techniques may be charts, graphs, color coding, and more.

Once you have an inference, always remember it is only a hypothesis. Real-life scenarios may always interfere with your results. In the process of Data Analysis, there are a few related terminologies that identity with different phases of the process.

**1. Data Mining**

This process involves methods in finding patterns in the data sample.

**2. Data Modelling**

This refers to how an organization organizes and manages its data.

**Data Analysis Techniques**

There are different techniques for Data Analysis depending upon the question at hand, the type of data, and the amount of data gathered. Each focuses on strategies of taking onto the new data, mining insights, and drilling down into the information to transform facts and figures into decision making parameters. Accordingly, the different techniques of data analysis can be categorized as follows:

1. Techniques based on Mathematics and Statistics

* **Descriptive Analysis**: Descriptive Analysis takes into account the historical data, Key Performance Indicators, and describes the performance based on a chosen benchmark. It takes into account past trends and how they might influence future performance.
* **Dispersion Analysis**: Dispersion in the area onto which a data set is spread. This technique allows data analysts to determine the variability of the factors under study.
* **Regression Analysis**: This technique works by modeling the relationship between a dependent variable and one or more independent variables. A regression model can be linear, multiple, logistic, ridge, non-linear, life data, and more.
* **Factor Analysis**: This technique helps to determine if there exists any relationship between a set of variables. In this process, it reveals other factors or variables that describe the patterns in the relationship among the original variables. Factor Analysis leaps forward into useful clustering and classification procedures.
* **Discriminant Analysis**: It is a classification technique in data mining. It identifies the different points on different groups based on variable measurements. In simple terms, it identifies what makes two groups different from one another; this helps to identify new items.
* **Time Series Analysis**: In this kind of analysis, measurements are spanned across time, which gives us a collection of organized data known as time-series.

2. Techniques based on Artificial Intelligence and Machine Learning

* **Artificial Neural Networks:** a Neural network is a biologically-inspired [programming paradigm](https://hackr.io/blog/programming-paradigms) that presents a brain metaphor for processing information. An Artificial Neural Network is a system that changes its structure based on information that flows through the network. [ANN](https://en.wikipedia.org/wiki/Artificial_neural_network) can accept noisy data and are highly accurate. They can be considered highly dependable in business classification and forecasting applications.
* **Decision Trees**: As the name stands, it is a tree-shaped model that represents a classification or regression models. It divides a data set in smaller subsets simultaneously developing into a related decision tree.
* **Evolutionary Programming**: This technique combines the different types of data analysis using evolutionary algorithms. It is a domain-independent technique, which can explore ample search space and manages attribute interaction very efficiently.
* **Fuzzy Logic**: It is a data analysis technique based on probability which helps in handling the uncertainties in data mining techniques.

3. Techniques based on Visualization and Graphs

* **Column Chart, Bar Chart**: Both these charts are used to present numerical differences between categories. The column chart takes to the height of the columns to reflect the differences. Axes interchange in the case of the bar chart.
* **Line Chart**: This chart is used to represent the change of data over a continuous interval of time.
* **Area Chart**: This concept is based on the line chart. It additionally fills the area between the polyline and the axis with color, thus representing better trend information.
* **Pie Chart**: It is used to represent the proportion of different classifications. It is only suitable for only one series of data. However, it can be made multi-layered to represent the proportion of data in different categories.
* **Funnel Chart**: This chart represents the proportion of each stage and reflects the size of each module. It helps in comparing rankings.
* **Word Cloud Chart:** It is a visual representation of text data. It requires a large amount of data, and the degree of discrimination needs to be high for users to perceive the most prominent one. It is not a very accurate analytical technique.
* **Gantt Chart**: It shows the actual timing and the progress of activity in comparison to the requirements.
* **Radar Chart**: It is used to compare multiple quantized charts. It represents which variables in the data have higher values and which have lower values. A radar chart is used for comparing classification and series along with proportional representation.
* **Scatter Plot**: It shows the distribution of variables in the form of points over a rectangular coordinate system. The distribution in the data points can reveal the correlation between the variables.
* **Bubble Chart**: It is a variation of the scatter plot. Here, in addition to the x and y coordinates, the area of the bubble represents the 3rd value.
* **Gauge:** It is a kind of materialized chart. Here the scale represents the metric, and the pointer represents the dimension. It is a suitable technique to represent interval comparisons.
* **Frame Diagram**: It is a visual representation of a hierarchy in the form of an inverted tree structure.
* **Rectangular Tree Diagram**: This technique is used to represent hierarchical relationships but at the same level. It makes efficient use of space and represents the proportion represented by each rectangular area.
* **Map**
  + **Regional Map:** It uses color to represent value distribution over a map partition.
  + **Point Map:** It represents the geographical distribution of data in the form of points on a geographical background. When the points are the same in size, it becomes meaningless for single data, but if the points are as a bubble, then it additionally represents the size of the data in each region.
  + **Flow Map:** It represents the relationship between an inflow area and an outflow area. It represents a line connecting the geometric centers of gravity of the spatial elements. The use of dynamic flow lines helps reduce visual clutter.
  + **Heat Map**: This represents the weight of each point in a geographic area. The color here represents the density.

**Data Analysis Tools**

There are several data analysis tools available in the market, each with its own set of functions. The selection of tools should always be based on the type of analysis performed, and the type of data worked. Here is a list of a few compelling tools for Data Analysis.

**1. Excel**

It has a variety of compelling features, and with additional plugins installed, it can handle a massive amount of data. So, if you have data that does not come near the significant data margin, then Excel can be a very versatile tool for data analysis.

**2. Tableau**

It falls under the BI Tool category, made for the sole purpose of data analysis. The essence of Tableau is the Pivot Table and Pivot Chart and works towards representing data in the most user-friendly way. It additionally has a data cleaning feature along with brilliant analytical functions.

**3. Power BI**

It initially started as a plugin for Excel, but later on, detached from it to develop in one of the most data analytics tools. It comes in three versions: Free, Pro, and Premium. Its PowerPivot and DAX language can implement sophisticated advanced analytics similar to writing Excel formulas.

**4. Fine Report**

Fine Report comes with a straightforward drag and drops operation, which helps to design various styles of reports and build a data decision analysis system. It can directly connect to all kinds of databases, and its format is similar to that of Excel. Additionally, it also provides a variety of dashboard templates and several self-developed visual plug-in libraries.

**5. R & Python**

These are programming languages which are very powerful and flexible. R is best at statistical analysis, such as normal distribution, cluster classification algorithms, and regression analysis. It also performs individual predictive analysis like customer behavior, his spend, items preferred by him based on his browsing history, and more. It also involves concepts of machine learning and artificial intelligence.

**6. SAS**

It is a programming language for data analytics and data manipulation, which can easily access data from any source. SAS has introduced a broad set of customer profiling products for web, social media, and marketing analytics. It can predict their behaviors, manage, and optimize communications.

**Conclusion**

This is a complete beginner guide about What is Data Analysis? Data Analysis is the key to any business, whether it be starting up a new venture, making marketing decisions, continuing with a particular course of action, or going for a complete shut-down. The inferences and the statistical probabilities calculated from data analysis help to base the most critical decisions by ruling out all human bias. Different analytical tools have overlapping functions and different limitations, but they are also complementary tools. Before choosing a data analytical tool, it is essential to take into account the scope of work, infrastructure limitations, economic feasibility, and the final report to be prepared.