**What is Predictive Analysis?**

- Analytics that help you forecast future performance and results. Historical Data to predict an Outcome or an Event.

Typically, Historical Data is used to build a Mathematical Model like Classifier, Predictive Model or a Regressor Which Capture the Important trends. And then current data is used on that Model to predict what will happen next. How to suggest action to take optimal Outcomes.

It Comprises Variety of Statistical Techniques like-

* Data Mining
* Predictive Modelling
* Machine Learning
* Define Problem Statement
* Data Collection
* Data Cleaning
* Data Analysis
* Build Predictive Model
* Validate Model
* Deployment

**Applications Where we Use Predictive Model**

Predictive Analysis can help in future projection that help in numerous Business ,etc.

We can use predictive analysis for lot of things.

* **Campaign Management :** figure out what kind of Target Audience based on previous data and previous campaign .
* **Customer Acquisition:**
* **Budgeting and Forecasting:**
* **Fraud Detection:**
* **Promotion:**
* **Pricing:**
* **Demand Planning:**

**Steps Involved in Predictive Analysis**

* Data Exploration (Data Assessment): Understanding of Data, unique values, columns, errors .
* Data Cleaning (missing , outliers, misspelled): reducing redundancy .
* Modelling : Build Model,
* Performance Analysis ( Evaluation): Evaluate best scoring model.

**Two Camps of Data Analysis**

Data analysis can be divided into two camps, according to the book [*R for Data Science*](https://r4ds.had.co.nz/):

1. **Hypothesis Generation** — This involves looking deeply at the data and combining your domain knowledge to generate [hypotheses](https://builtin.com/data-science/data-science-terms) about why the data behaves the way it does.
2. **Hypothesis Confirmation** — This involves using a precise mathematical model to generate falsifiable predictions with statistical sophistication to confirm your prior hypotheses.

**Types of Data Analysis**

Data analysis can be separated and organized into types, arranged in an increasing order of complexity.

1. Descriptive analysis
2. Diagnostic analysis
3. Exploratory analysis
4. Inferential analysis
5. Predictive analysis
6. Causal analysis
7. Mechanistic analysis
8. Prescriptive analysis

**1. DESCRIPTIVE ANALYSIS**

The goal of [descriptive analysis](https://builtin.com/data-science/descriptive-statistics) isto describe or summarize a set of data. Here’s what you need to know:

* Descriptive analysis is the very first analysis performed in the data analysis process.
* It generates simple summaries about samples and measurements.
* It involves common, descriptive statistics like measures of central tendency, variability, frequency and position.

**Descriptive Analysis Example**

Take the [Covid-19 statistics page](https://www.google.com/search?q=covid+19+statistics+google&rlz=1C5CHFA_enUS871US871&oq=covid+19+statistics+google&aqs=chrome..69i57j0i433i512j0i131i433i512j0i131i433i457i512j0i402l2j0i512j69i61.2993j0j4&sourceid=chrome&ie=UTF-8) on Google, for example. The line graph is a pure summary of the cases/deaths, a presentation and description of the population of a particular country infected by the virus.

Descriptive analysis is the first step in analysis where you summarize and describe the data you have using descriptive statistics, and the result is a simple presentation of your data.

MORE ON DATA ANALYSIS:[Data Analyst vs. Data Scientist: Similarities and Differences Explained](https://builtin.com/data-science/data-analyst-vs-data-scientist)

**2. DIAGNOSTIC ANALYSIS**

Diagnostic analysis seeks to answer the question “Why did this happen?” by taking a more in-depth look at data to uncover subtle patterns. Here’s what you need to know:

* Diagnostic analysis typically comes after descriptive analysis, taking initial findings and investigating why certain patterns in data happen.
* Diagnostic analysis may involve analyzing other related data sources, including past data, to reveal more insights into current data trends.
* Diagnostic analysis is ideal for further exploring patterns in data to explain anomalies.

**Diagnostic Analysis Example**

A footwear store wants to review its [website traffic](https://builtin.com/design-ux/track-essential-website-metrics) levels over the previous 12 months. Upon compiling and assessing the data, the company’s marketing team finds that June experienced above-average levels of traffic while July and August witnessed slightly lower levels of traffic.

To find out why this difference occurred, the marketing team takes a deeper look. Team members break down the data to focus on specific categories of footwear. In the month of June, they discovered that pages featuring sandals and other beach-related footwear received a high number of views while these numbers dropped in July and August.

Marketers may also review other factors like seasonal changes and company sales events to see if other variables could have contributed to this trend.

**3. EXPLORATORY ANALYSIS (EDA)**

Exploratory analysis involves examining or [exploring data](https://builtin.com/data-science/EDA-python) and finding relationships between variables that were previously unknown. Here’s what you need to know:

* EDA helps you discover relationships between measures in your data, which are not evidence for the existence of the correlation, as denoted by the phrase, “[Correlation doesn’t imply causation](https://builtin.com/data-science/correlation-is-not-causation).”
* It’s useful for discovering new connections and forming hypotheses. It drives design planning and data collection.

**Exploratory Analysis Example**

[Climate change](https://builtin.com/greentech) is an increasingly important topic as the global temperature has gradually risen over the years. One example of an exploratory data analysis on climate change involves taking the rise in temperature over the years from 1950 to 2020 and the increase of human activities and industrialization to find relationships from the data. For example, you may increase the number of factories, cars on the road and airplane flights to see how that correlates with the rise in temperature.

Exploratory analysis explores data to find relationships between measures without identifying the cause. It’s most useful when formulating hypotheses.

**4. INFERENTIAL ANALYSIS**

[Inferential analysis](https://builtin.com/learn/inferential-statistics) involves using a small sample of data to infer information about a larger population of data.

The goal of [statistical modeling](https://builtin.com/data-science/skewed-data) itself is all about using a small amount of information to extrapolate and generalize information to a larger group. Here’s what you need to know:

* Inferential analysis involves using estimated data that is representative of a population and gives a measure of uncertainty or [standard deviation](https://builtin.com/data-science/difference-between-standard-deviation-standard-error) to your estimation.
* The [accuracy](https://builtin.com/learn/tech-dictionary/data-integrity) of inference depends heavily on your sampling scheme. If the sample isn’t representative of the population, the generalization will be inaccurate. This is known as the [central limit theorem](https://builtin.com/data-science/understanding-central-limit-theorem).

**Inferential Analysis Example**

The idea of drawing an inference about the population at large with a smaller sample size is intuitive. Many statistics you see on the media and the internet are inferential; a prediction of an event based on a small sample. For example, a psychological study on the benefits of sleep might have a total of 500 people involved. When they followed up with the candidates, the candidates reported to have better overall attention spans and well-being with seven-to-nine hours of sleep, while those with less sleep and more sleep than the given range suffered from reduced attention spans and energy. This study drawn from 500 people was just a tiny portion of the 7 billion people in the world, and is thus an inference of the larger population.

Inferential analysis extrapolates and generalizes the information of the larger group with a smaller sample to generate analysis and predictions.

**5. PREDICTIVE ANALYSIS**

Predictive analysis involvesusing historical or current data to find patterns and make predictions about the future. Here’s what you need to know:

* The accuracy of the predictions depends on the input variables.
* Accuracy also depends on the types of models. A linear model might work well in some cases, and in other cases it might not.
* Using a variable to predict another one doesn’t denote a causal relationship.

**Predictive Analysis Example**

The 2020 US election is a popular topic and many [prediction models](https://builtin.com/machine-learning/predictive-behavior-modeling) are built to predict the winning candidate. FiveThirtyEight did this to forecast the 2016 and 2020 elections. Prediction analysis for an election would require input variables such as historical polling data, trends and current polling data in order to return a good prediction. Something as large as an election wouldn’t just be using a linear model, but a complex model with certain tunings to best serve its purpose.

Predictive analysis takes data from the past and present to make predictions about the future.

MORE ON DATA:[Explaining the Empirical for Normal Distribution](https://builtin.com/data-science/empirical-rule)

**6. CAUSAL ANALYSIS**

[Causal analysis](https://builtin.com/data-science/causality-vs-correlation-experiments)looks at the cause and effect of relationships between variables and is focused on finding the cause of a correlation. Here’s what you need to know:

* To find the cause, you have to question whether the observed correlations driving your conclusion are valid. Just looking at the surface data won’t help you discover the hidden mechanisms underlying the correlations.
* Causal analysis is applied in randomized studies focused on identifying causation.
* Causal analysis is the gold standard in data analysis and scientific studies where the cause of phenomenon is to be extracted and singled out, like separating wheat from chaff.
* Good data is hard to find and requires expensive research and studies. These studies are analyzed in aggregate (multiple groups), and the observed relationships are just average effects (mean) of the whole population. This means the results might not apply to everyone.

**Causal Analysis Example**

Say you want to test out whether a new drug improves human strength and focus. To do that, you perform randomized control trials for the drug to test its effect. You compare the sample of candidates for your new drug against the candidates receiving a mock control drug through a few tests focused on strength and overall focus and attention. This will allow you to observe how the drug affects the outcome.

Causal analysis is about finding out the causal relationship between variables, and examining how a change in one variable affects another.

**7. MECHANISTIC ANALYSIS**

Mechanistic analysis is used tounderstand exact changes in variables that lead to other changes in other variables. Here’s what you need to know:

* It’s applied in physical or engineering sciences, situations that require high [precision](https://builtin.com/product/accuracy-precision-product-management) and little room for error, only noise in data is measurement error.
* It’s designed to understand a biological or behavioral process, the pathophysiology of a disease or the mechanism of action of an intervention.

**Mechanistic AnalysisExample**

Many graduate-level research and complex topics are suitable examples, but to put it in simple terms, let’s say an experiment is done to simulate safe and effective nuclear fusion to power the world. A mechanistic analysis of the study would entail a precise balance of controlling and manipulating variables with highly accurate measures of both variables and the desired outcomes. It’s this intricate and meticulous modus operandi toward these big topics that allows for scientific breakthroughs and advancement of society.

Mechanistic analysis is in some ways a predictive analysis, but modified to tackle studies that require high precision and meticulous methodologies for physical or engineering science*.*

**8. PRESCRIPTIVE ANALYSIS**

[Prescriptive analysis](https://builtin.com/data-science/prescriptive-analytics) compiles insights from other previous data analyses and determines actions that teams or companies can take to prepare for predicted trends. Here’s what you need to know:

* Prescriptive analysis may come right after predictive analysis, but it may involve combining many different data analyses.
* Companies need advanced technology and plenty of resources to conduct prescriptive analysis. [AI](https://builtin.com/artificial-intelligence) systems that process data and adjust automated tasks are an example of the technology required to perform prescriptive analysis.

**Prescriptive Analysis Example**

Prescriptive analysis is pervasive in everyday life, driving the curated content users consume on social media. On platforms like TikTok and Instagram, [algorithms](https://builtin.com/software-engineering-perspectives/algorithm) can apply prescriptive analysis to review past content a user has engaged with and the kinds of behaviors they exhibited with specific posts. Based on these factors, an [algorithm seeks out similar content](https://builtin.com/software-engineering-perspectives/algorithmic-curation) that is likely to elicit the same response and recommends it on a user’s personal feed.

A tutorial on the different types of data analysis. | Video: Shiram Vasudevan

**When to Use the Different Types of Data Analysis**

* **Descriptive analysis** summarizes the data at hand and presents your data in a comprehensible way.
* **Diagnostic analysis** takes a more detailed look at data to reveal why certain patterns occur, making it a good method for explaining anomalies.
* **Exploratory data analysis** helps you discover correlations and relationships between variables in your data.
* **Inferential analysis** is for generalizing the larger population with a smaller sample size of data.
* **Predictive analysis** helps you make predictions about the future with data.
* **Causal analysis** emphasizes finding the cause of a correlation between variables.
* **Mechanistic analysis** is for measuring the exact changes in variables that lead to other changes in other variables.
* **Prescriptive analysis** combines insights from different data analyses to develop a course of action teams and companies can take to capitalize on predicted outcomes.

A few important tips to remember about data analysis include:

* Correlation doesn’t imply causation.
* EDA helps discover new connections and form hypotheses.
* Accuracy of inference depends on the sampling scheme.
* A good prediction depends on the right input variables.
* A simple linear model with enough data usually does the trick.
* Using a variable to predict another doesn’t denote causal relationships.
* Good data is hard to find, and to produce it requires expensive research.
* Results from studies are done in aggregate and are average effects and might not apply to everyone.​