

Week 1

Thursday, April 22, 2021 4:01 PM

Key Players in the Data Ecosystem

To summarize, in simple terms:

Data engineering - converts raw data into usable data.

Data analytics - uses this data to generate insights.

Data scientists - use data analytics and data engineering to predict the future using data from the past, business analysts and business intelligence analysts use these insights and predictions to drive decisions that benefit and grow their business.

Interestingly, it's not uncommon for data professionals to start their career in one of the data roles and transition to another role within the data ecosystem by supplementing their skills.

Defining Data Analytics

We will explore **four primary types** of data analysis, each with a different goal and place in the data analysis process.

Descriptive Analytics helps answer questions about what happened over a given period of time by summarizing past data and presenting the findings to stakeholders. It helps provide essential insights into past events. **For example**, tracking past performance based on the organization's key performance indicators or cash flow analysis.

Diagnostic analytics helps answer the question, Why did it happen? It takes the insights from descriptive analytics to dig deeper to find the cause of the outcome. **For example**, a sudden change in traffic to a website without an obvious cause or an increase in sales in a region where there has been no change in marketing.

Predictive analytics helps answer the question, What will happen next? Historical data and trends are used to predict future outcomes. Some of the areas in which businesses apply predictive analysis are risk assessment and sales forecasts.

It's important to note that the purpose of predictive analytics is not to say what will happen in the future, it's objective is to forecast what might happen in the future. All predictions are probabilistic in nature.

Prescriptive Analytics helps answer the question, What should be done about it? By analyzing past decisions and events, the likelihood of different outcomes. Is estimated on the basis of which a course of action is decided. Self-driving cars are a **good example** of Prescriptive Analytics. They analyze the environment to make decisions regarding speed, changing lanes, which route to take, etc. Or airlines automatically adjusting ticket prices based on customer demand.

Gas prices, the weather or traffic on connecting routes.

Key Steps in Data Analysis Process

Understanding the problem and desired result - Defining where you are and where you want to be

Setting a clear metric - Deciding what will be measured and how it will be measured

Gathering data - Identifying data you require, the sources from which you will access this data, and

the best tools for the job

Cleaning data - Fixing quality issues in the data and standardizing data coming in from multiple sources

Analyzing and Mining Data - Extracting, analyzing, and manipulating data from different perspectives to understand trends, identify correlations, and find patterns and variations.

Interpreting results - Interpreting results, evaluating defendability of analysis and circumstances under which analysis may not hold true.

Presenting your findings - Communicating and presenting your findings in clear, impactful and convincing ways.

Viewpoints: What is data analysis?

- Collecting information
- Analyzing Data
- Confirming Hypothesis
- Storytelling with Data
- Use of Information to Make Decisions
- Analyzing sets of data to understand what's going on
- Understanding where a business is coming from, it's present, and future direction
- Communicating Business Insights

Responsibilities of a Data Analyst

- Acquiring Data
- Creating queries to extract required data
- Filtering, cleaning, standardizing, and reorganizing data
- Using statistical tools and techniques
- Analyzing patterns
- Preparing reports and charts
- Creating appropriate documentation

Technical Skills

- Expertise in using spreadsheets (Microsoft Excel or Google Sheets)
- Proficiency in statistical analysis and visualization tools and software (SPSS, Power BI, SAS, Oracle Visual Analyzer)
- Proficiency in programming languages (R, Python, C++, Java, Matlab)
- Good knowledge of SQL and ability to work with data in relational and NoSQL databases
- The ability to access and extract data from data repositories (Data Marts, Data Lakes, Data Pipelines)
- Familiarity with Big Data processing tools (Hadoop, Hive, and Spark)

Functional Skills

- Proficiency in Statistics
 - Analyze data, validate the analysis, identify fallacies and logical errors

- Analytical skills
 - Research and interpret data, theorize, make forecasts
- Problem-solving skills
 - Come up with possible solutions for a given problem
- Probing skills
 - Identify and define the problem statement and desired outcome
- Data Visualization skills
 - Create clear and compelling visualizations to present the analysis
- Project Management skills
 - Manage the process, people, dependencies and timelines

Functional Skills

- Work collaboratively with business and cross-functional teams
- Communicate effectively to report and present your findings
- Tell a compelling and convincing story
- Gather support and buy-in for your work
- Allowing new questions to surface and challenging your own assumptions and hypotheses
- Having a sense of the future based on pattern recognition and past experiences