

## Name of the Skill

“**Data Analysis**” is considered in continuous improvement and allows you to make sound decisions about the process, product or service. Generally collecting and reviewing data, and making inferences from the data is called Data Analysis [1]. There are several types of Data Analysis techniques that exist based on technology. However, the major types of data analysis are: Text Analysis, Statistical Analysis, Diagnostic Analysis, Predictive Analysis, and Prescriptive Analysis [3]. Each of these types of data analysis are connected and depend to some degree on each other. Each serves a different purpose, and provides different insights. Moving from descriptive analysis to predictive and prescriptive analysis requires much more technical skills, but also provides your organization with more insight.



## Classification of Skill

Data analysis can be classified as a “**Hard Technical Skill**” as it requires training and experience to master. Hard skills are gained and improved by practice, training, and education. Hard skills are learned abilities acquired and enhanced through practice, repetition, and education which are important because they increase productivity and efficiency [5].

## Prerequisites for Skill

There are several background areas that can help a person to be expertise in this skill: Statistical Programming experience with R or Python ,Structured thinking approach and Critical Thinking ,Passion for solving problems Willingness to learn statistical concepts , Love for numbers and quantitative stuff , Grit to keep on learning , Data Visualization , Presentation Skills .

## Related Software Engineering Area(s)

Data analysis can be used in different software engineering areas. It is considered as one of the fundamental skills during "software Process Improvement" in **Software Engineering Process** area. Also in **Software Engineering Management** and **Software Engineering Foundations**.

## Rationale for Skill

Data analysis as an internal arrangement function done by data analysts through presenting numbers and figures to management, involves a more detailed approach in recording, analyzing, disseminating and presenting data findings in a way that is easy to interpret and make decisions for the business. Data analysis is one of the fundamental skills for software process improvement and helps analysts to make decisions on process improvement and behavior prediction, increasing profitability and drive effective decision-making.

## Roles for Skill

**Process Improvement Analyst:** Develop procedures to improve operational efficiency through process improvement. Analyze current process standards and metrics to deliver enhancement solutions. To develop appropriate process control changes, perform data gathering, root cause analysis, and performance trends.

**Research Assistant, Business Analyst, Data Analyst, Software Engineer, Data Scientist** are other roles that can be considered for the data analysis skill.

## Work Related to Skill

Gathering and analyzing data from processes will begin to uncover some unexpected insights. Such experiences can then be used to build process changes and strengthen the software models and methods and increase performance. Based on the gathered information and data, you can make decisions for process improvement. The process of Data Analysis consists of the following phases: [3]

**Data Requirement Gathering:** You need to ponder why you want to do this data analysis? All you need to find out about the purpose or purpose of doing the Analysis. You have to determine what sort of data analysis you want to carry out! You have to decide in this phase what to analyze and how to measure it, you have to understand why you are investigating, and what measures to use to do this analysis.

**Data Collection:** After gathering the requirements, you'll get a clear idea of what you need to measure and what your findings should be. Now is the time to collect the data on the basis of the needs. When collecting your data, remember to process or organize the collected data for

**Analysis.** As you have collected data from different sources, you must keep a log with the date and source of the data being collected.

**Data Cleaning:** Now whatever data is collected may not be useful or irrelevant to your analysis objective, therefore it should be cleaned up. The gathered data may contain duplicate records, white spaces or errors. The data should be cleaned, without error. This phase must be done before Analysis because your Analysis output will be closer to your expected outcome, based on data cleaning.

**Data Analysis:** Once the data is collected, cleaned and processed, Analysis is ready. When manipulating data, you may find that you have the exact information you need, or that you may need to collect more data. During this process, you will use tools and software for data analysis that will help you understand, interpret and draw conclusions based on the requirements.

**Data Interpretation:** It's finally time to interpret your results, after analyzing your data. You may select the way the data analysis can be presented or conveyed either clearly in phrases, or maybe a table or map. Then use your data analysis findings to determine the best course of action.

**And Data Visualization:** Visualization of data is very common in your day-to-day life; it often appears as charts and graphs. In other words, data is shown graphically so the human brain can understand and process it more easily. Visualization of data often used to spot unknown facts and trends. You can find a way to find meaningful information by observing the relationships and comparing data sets.

## Real-World Example



As a real world example, **Process Mining** is a unique data analysis technique which can strengthen Agile Project Management during Software Development. Through Process Mining, companies can not only achieve a long-term change in the software development process but will also gain a more solid overall understanding of their software [4].

Process mining can be used to discover agile processes followed by agile teams to set the baselines and assess fidelity, or to gain feedback to enhance agility.

## Role of Academia or Industry in Cultivating the Skill

As demand for data analysis and data science continues to rise, academia is trying to find out how best to contribute to the training of a workforce. Strong demand for data analytics graduates makes data science courses a prerequisite for all students in this area. On the other hand, the growth of data analysis in academia is primarily a function of its vast private sector expansion,

and many of the major developments in data science over the past decade have been made by large private companies such as Google, Facebook and LinkedIn [3].

## Tools Supporting the Skill

The list of top 10 of data analytics tools, both open source and paid version, based on their popularity, learning and performance.



**R programming** is the leading analytics tool in the industry and widely used for statistics and data modeling. It can easily manipulate your data and present in different ways.



**Tableau Public** is a free software that connects any data source be it corporate Data Warehouse, Microsoft Excel or web-based data, and creates data visualizations, maps, dashboards.



**Python** is an object-oriented scripting language which is easy to read, write, maintain and is a free open source tool.



**SAS** is a programming environment and language for data manipulation and a leader in analytics



**Apache Spark** is a fast large-scale data processing engine and executes applications in Hadoop clusters 100 times faster in memory and 10 times faster on disk.



**Excel** is a basic, popular and widely used analytical tool almost in all industries. Whether you are an expert in Sas, R or Tableau, you will still need to use Excel.



**RapidMiner** is a powerful integrated data science platform that performs predictive analysis and other advanced analytics like data mining, text analytics, machine learning and visual analytics without any programming.



**KNIME** is a leading open source, reporting, and integrated analytics tools that allow you to analyze and model the data through visual programming



**QlikView** has many unique features like patented technology and has in-memory data processing, which executes the result very fast to the end users and stores the data in the report itself.



**Splunk** is a tool that analyzes and search the machine-generated data. Splunk pulls all text-based log data and provides a simple way to search through it.

# Skill Self-Assessment

7 Out of 10

Although I have a good knowledge and experience about the pre-requisites and tools of this skill, being expertise in this area require minimum 1-2 years of quantitative analysis work experience and more than 2 years experience handling, manipulating and analyzing data and creating analytical reports.

## References

- [1] <https://www.quality-assurance-solutions.com/what-is-data-analysis.html>
- [2] <https://businesspartnermagazine.com/5-reasons-why-data-analysis-is-important-for-every-business/#:~:text=Knowing%20your%20Target%20Customers,on%20and%20for%20which%20customer%20S.>
- [3] <https://www.guru99.com/what-is-data-analysis.html>
- [4] <https://www.linkedin.com/pulse/process-mining-strengthen-agile-project-management-during-salzmann/>
- [5] <https://www.investopedia.com/terms/h/hard-skills.asp>
- [6] <https://www.proschoolonline.com/blog/top-10-data-analytics-tools>
- [7] <https://ieeexplore.ieee.org/document/7975917>