Lesson 01: Data Analytics Introduction

During this course, you are going to learn some important basics about data analytics that will ~~let~~ allow you to master it in a very short time. We will discuss important topics and demonstrate how to combine them. Also, you will find some real-world examples explained to give you a clear idea of how organizations can extract information from the process of analyzing data today. So, by the end of this lesson you ~~will~~ should be able to explain:

* What is data and data analytics?
* Difference between Data Analytics and Data Science.
* Evolution of Data Analytics.
* Importance of data analytics to study and use in our life.
* How Data Analytics Work?
* ~~Famous~~ Most popular Data Analytics tools.
* ~~Various real-world examples.~~

Before starting let’s have a look on Facebook and see how things ~~go~~ works. ~~Imagine~~ Your Facebook profile has many various data like your age, relationship, interests, watched movies, type of your favorite music and so on. All these data together will pass through a certain process that will help Facebook displaying some advertisement which you may be interested in. For example, let’s say a Facebook user “Andy” who lives in California and he is a big fan of Elvis Presley, Wanda Jackson, and Roy Orbison songs. Another Facebook user “Nancy” who lives in Kentucky and she is a big fan of Jennifer Lopez, Shakira, and Ricky Martin songs.

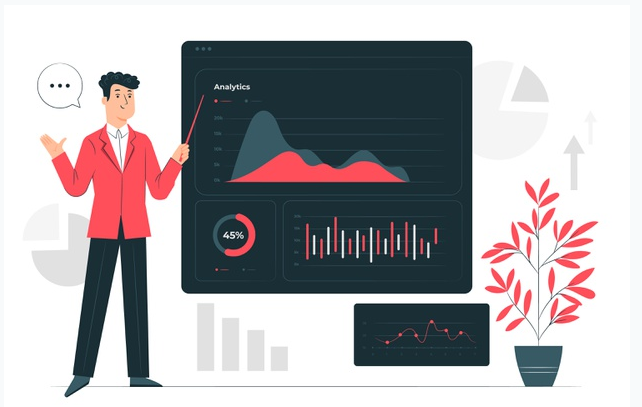
Now a Facebook ad[[1]](#footnote-1) may appear to Nancy for the Katy Perry concert that will be in Louisville, meanwhile, Andy may see a Facebook ad for a store located in Los Angeles selling a medley CD containing Chuck Berry's greatest hits. As you can notice the two ads are different in content, however, they are based on each Facebook user's interests, type of music ~~he or she~~ they listen to and some other data. So, what ~~is~~ does data mean?

What is Data?

Nowadays, data is the fuel of everything that exists in our life where every decision you made, every option you choose and every step you take is based on some sort of data that may be unrelated. However, these data together and after being processed will ~~produce~~ ~~you~~ generate the needed information ~~that will~~ to help you taking the optimum choice and decision. So basically, **data** is defined as a set or a group of facts like words, numbers, various interests, type of music you like, etc. Usually, data in its first form is called raw data[[2]](#footnote-2). Now, these data need to be processed in such a way it is transformed into more useful information which will be in an understandable format.

Back again to the previous Facebook example. Typically, the process that was applied to the Facebook users’ data will help the Facebook business grows by attracting more advertisers. Now, the most important question arises, what is the process that is used by Facebook and how it was used? The answer is Data Analytics.

What is Data Analytics?

**Data analytics** is considered one of the most trending and uprising terminologies in the information technology field these days. It is the process of converting or translating the data into useful information by drawing results and conclusions. Typically, without using this process, data itself is just useless and there will be no useful or meaningful information coming out of it.

Typically, data analytics is considered the method of turning hundreds of raw data into meaningful thoughts that can be easily applied to some business processes. Nowadays, data analytics has played an important role in most corporates’ growth strategies and business developments. All of this happened because data analytics allows detecting patterns in data that can be used to highlight certain areas that will help to improve better and effective strategies. Also, it will help in exploring new and possible future trends.

Additionally, data analytics can be used in many various areas and industries to increase productivity and business gain. Typically, the extracted data is cleaned and arranged into categories to analyze various patterns. Various techniques and tools are used which may vary depending on the organization's needs. Corporations like Twitter, Google and Facebook can gather various data from online activities. Then these data can be analyzed for better service optimization or tailored advertisements and many other services. Moreover, data only by itself has no clear meaning and does not reveal much particularly when viewed and examined without any convenient method. To understand these data meaning there should be a way or method to refine it, clean it, arrange it and explain it better. All of the previous stages are included in the process of analyzing data.

In general, we are surrounded by a huge amount of raw data that has incredibly increased in size. These raw data and in its current format have no meaning. However, to get some useful information from these raw data, it needs to be polished, cleaned, transformed and modeled to get various insights from these raw data which will help in better decision making and so many other benefits. However, and before discussing the data analytics evolution and how it becomes a popular trend nowadays, let’s differentiate first between two important terminologies which are Data Science and Data Analytics.

Data Science Vs Data Analytics

Typically, **Data Science** is a multidisciplinary field to study where the information comes from, what does this information represent and even more how this information can be translated into a more valuable and useful resource. In other words, we can say that data science is about revealing findings data. Also, data science can deal with a large amount of data that ~~may~~ could be structured or unstructured.

To sum up, you can say that the term data science is extremely similar to the term science which is a large term that has a great number of specialties. Similarly, data science is a widely used term that includes various methods and models to help you get information. Moreover, a data scientist's main concern is to find the right questions to ask with less concern for searching for specific answers.

On the other hand, if data science is the place or home that carries all the methods and tools, then **Data Analytics** is considered a certain room in that home.Consequently, data analytics is data more concentrated and specific than data science. Typically, data analytics is more focused on a specific goal in mind that you need to sort the data and search for methods to support this goal, instead of just searching for a connection between this data. Additionally, data analytics is simply solving and fixing problems for questions you truly know that you do not ~~know~~ have the answers to. Also, data analytics is based on getting results that will lead to instant enhancements and improvements.

Now you ~~must~~ should have a clear idea of the difference between Data Science and Data Analytics. Now let’s move on to the next section to check the history of data analytics.

Evolution of Data Analytics

Statistics is the root or basis for the foundation of data analytics and it is supposed that statistics were firstly used by Ancient Egyptians while they were building the pyramids. Historically, statistics played a very important role across the world for all governments. For example, the censuses creation that is used for many different planning activities related to the government as taxation. The next step with all the collected data is to analyze it, then you will be able to discover some beneficial information; like determining the population growth in a certain area will help governments decide whether or not to increase the number of hospitals in this given area.

Some say that data analytics began to appear in the 19th century. However, in the late 1960s, ~~data analytics~~ it started to slightly arise as the support and decision-making systems started to receive more attention.

The real turning point was in the 1980s with the appearance of the relational database systems. Additionally, using SQL to select and retrieve data from the database, users were able to analyze their data easily. Also, the process of getting data to become easier than before and this helped to spread the use of database systems.

In the late 1980s, the amount of gathered data started to increase remarkably. That’s when the Data Warehouse[[3]](#footnote-3) and Business Intelligence (BI)[[4]](#footnote-4) terms were proposed. Using BI helps you better in decision making with searching, gathering and analyzing the stored data.

Around the 1990s, data mining[[5]](#footnote-5) appeared which introduced new ways of analyzing more data other than usual methods. Hence, a new business trend started to show up where organizations began to predict any potential customer needs, for example, based on the customer’s purchasing style.

The second turning point was in the 2000s, where the service of the Google search engine boomed out and started to process big data then analyze it. It was just like magic; you google something and the response comes out with just a few seconds with the result that you most likely needed. Moreover, with the Big Data[[6]](#footnote-6) development alongside the Cloud[[7]](#footnote-7), Data Warehouses and much other software and hardware enhancements, Data Analytics emerged on the surface significantly.

In the 2010s, two important services which are Amazon Redshift[[8]](#footnote-8) and Google BigQuery[[9]](#footnote-9) came out at a very affordable price. For the time being, almost all the companies even startups can easily obtain and build an infrastructure for big data analysis at a very low cost.

To put it differently, computer technology and data analysis are connected. They are developing, improving and affecting each other. They are in a proportional relationship, as the gathered data gets larger then new data analysis techniques will be introduced.

Why Study Data Analytics?

Every day more organizations are gathering more data than ever before. These organizations might be governmental or business or healthcare or even charity or any other organization. These gathered data include many insights into the organization’s operations and more. Hence, data analytics is needed to extract these insights. Also, this will require some statistical, analytical and computational skills.

Advantage of Data Analytics

Simply data analytics is here because there is great information ~~is~~ buried in the data and it’s waiting to be uncovered and discovered. Again, data itself is just having no useful meaning, it needs some cleaning, arranging and polishing to see what is hidden beneath ~~this data~~ it. Once you discover the hidden information you will be able to:

* Gather all the ~~hidden~~ unknown insights then analyzed to clear business requirements.
* Take decisions in the most critical situations easily and rapidly.
* Deeply understand the client requirements correspondingly you will be able to build successful business relationships with your clients.
* Implement and apply preventative measures to increase risk management.
* Rapidly react to any change in the market or business hence flexibility improvement.
* Have a deep and clear look into financial performance.
* Reduce unnecessary costs and hence increase the organization's profit.

How data analytics can help business?

Think about the Pepsi corporation, Pepsi company may gather data from retailers, points of sale, various client surveys, and so on. All this gathered information will provide the company management various insights that can be used to enhance the financial, marketing, development strategies, and so many other company-related aspects. For example, based on the gathered information the company may suggest tailored advertisements and promotions, better marketing plan, value chain enhancements, and so many other decisions to improve the customer experience and increase company profit.

Typically, data that is used in the process of analytics needs to be saved, analyzed properly, remove not use data and arranging it into logical groups. Then connect key data points to find similar patterns and helpful information that might be needed in the future. As well, using data analytics companies can better understand themselves, design new and better strategies based on the data analytics results and findings which can be translated and mapped into actionable insights.

There are so many various key areas where using a technique of analytics can be a great win for the business. Looking inside the company operation is one of the first areas that several companies look at. Some companies regularly set objectives, milestones, and goals to reach, however, tracking them is not that easy. On the other hand, focusing on the business and data teams it will be easy to set key performance indexes and have clear methods to measure them.

Indeed, data analytics can remarkably enhance customer support and relations. Moreover, customer engagement and service are not just based on the fact of understanding what the customer needs at a certain time, but also, what the customer may need in the future. By using predictive data analytics, a company will be able to see and locate the areas where a customer is satisfied, the area where improvement is needed and above all, where these options may lead in the future. Additionally, company marketing can be greatly improved using data analytics. It will be so easy to track the marketing plan's success by measuring something like social media interactions or users clicks on some advertisements and some other measures which will deeply help in planning for future marketing campaigns.

How Things Work Using Data Analytics

Generally, the data analytics process starts with an idea you need to investigate or a decision to make or even a question that needs an answer. Then first things first, you will have to gather data. Once you have data in your hand, it will be time to explore it and look deep inside it. Consequently, you will be able to check for errors in the data like blanks or useless words like ‘sfasafs’ which has no meaning or missing data like if you have Paris and there is no corresponding country then obviously you can put France as a country and so on.

Moreover, and as you move on, you will need to visualize the data as it is considered one of the best ways to explore things. Once you have a good visual structure for the data you will next be going to test it by for example applying some filters and sorting to check if the data has a meaning. As soon as you find a meaning for the data now you can study the results to make a decision or get an answer for a certain question. Sometimes this step may lead to more questions, in other words, you answer a question and another one pops up. Hence, the data analytic cycle goes on as if it is a never-ending cycle which is a big challenge and you need to fully understand that it is an ongoing cycle or process.

Generally, in real life, the data analytic process does not follow these perfect predefined steps and even more, you can jump between these steps easily and flexibly. Instead, what happens is that you start collecting data, exploring it and maybe at some point you realize that you need to get more data, so you move back to the first stage and collect more data again. But again, and for the most part, when you think about the data analytics process below are the most stages that you need to follow:

1. Identify the problem and ask questions.
2. Gathering as much data as you can.
3. Explore and clean collected data to remove errors.
4. Visualize data for better and rapid understandings.
5. Test the results and check if they hold up and have a meaning or not.
6. Study the results and get answers to your questions.

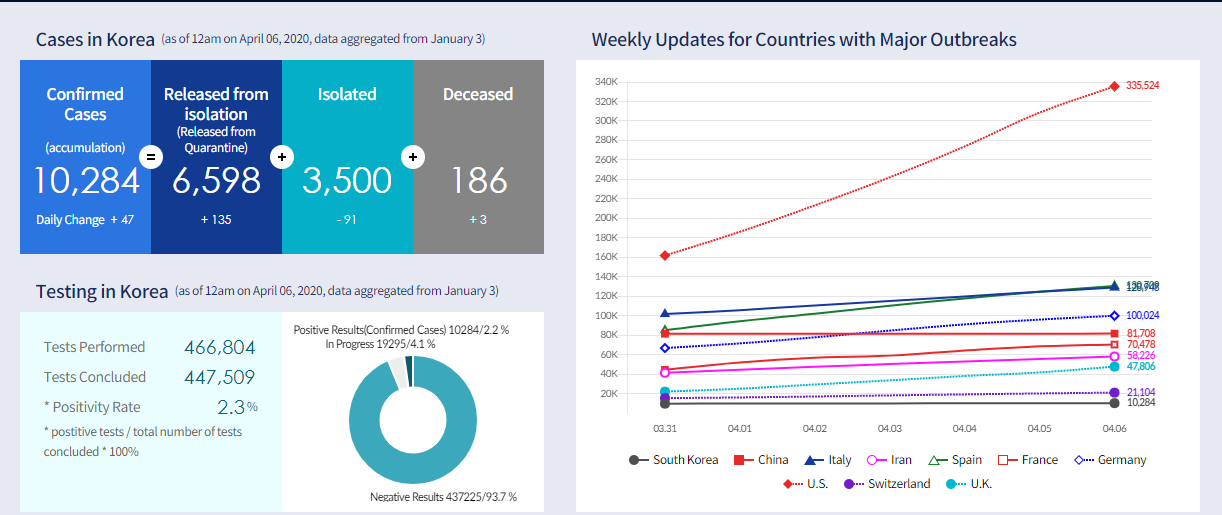
Data Analytics Tools

Due to the increase in demand for using data analytics, various tools have emerged with many and different functionalities. Below are the top tools that can be used in data analytics:

* [Microsoft Excel](https://products.office.com/en/excel) is one of the most famous and widely used tools in the data analytics market. Typically, Microsoft Excel is chosen and used by many clients to analyze their internal data.
* [Python](https://www.python.org/) is an object-oriented programming language which is an open-source tool. Python is easy to use, read, write and preserve. Moreover, Python offers many visualization and machine learning libraries that can be used to analyze data. Also, it can be used alongside SQL Server, JSON, MongoDB, etc.
* [R Programming](https://www.r-project.org/) is the leading data analytics tool that is used for data modeling and statistics. The R language can run on many different platforms like Windows, Mac OS, and Linux. Moreover, it offers tools that can automatically and based on the user requirements to install all packages.
* [SAS](https://www.sas.com/en_us/home.html) is a programming language used for data analytics and manipulation. The SAS is easily accessible and can perform data analyzing from various sources.
* [KNIME](https://www.knime.com/) is an open-source platform which is a shortcut for Konstanz Information Miner. KNIME allows you to perform data modeling and analyzing alongside visual programming and a reporting and integration platform.

Real-World Examples of Data Analytics

South Korea and Coronavirus Pandemic



One great example of using data analytics to save human lives in the recent South Korea experience in fighting the Coronavirus or “COVID-19” that widely spread in the whole world in late 2019. South Korea has a huge Big Data platform run by the government. This platform stores all the residents and foreign citizens' data. These data come from every organization found in the country like airports, tourism companies, mobile phone operators, financial services, hospitals, various governmental organizations, and many other organizations. Additionally, these various data that are available from many various sources are integrated and analyzed in such a way the result is used in fighting the Coronavirus spread.

For example, if a person was subjected to a Coronavirus medical test and his or her result was positive[[10]](#footnote-10), then all this person's neighbors will be notified by all his or her movements and activities through the previous two weeks. Consequently, neighbors will be very cautious and isolate themselves to avoid spreading the virus or even seek medical help. People get notified through a mobile application or SMS. Meanwhile, all the health providers in South Korea will receive all the related patient information. Moreover, it will be so easy to track the patient’s contacts and everyone who has dealt with this patient in the past 14 days. Hence, provide medical observation, test and help if needed.

Even more, the South Korean government was able to determine and detect regions with the highest Coronavirus infection rate and this was based on the analyzed data results. Hence, the government was able to increase awareness campaigns in these regions especially.

Google and Data Analytics

Consider the great Google corporation, Google is one of the largest and most famous tech companies in the globe and it is a multi-billion-dollar company worth. However, Google does not charge for the great majority of its offered services. Have a look at Gmail, Google Search, Google Analytics and so many other free services that are extremely expensive to deliver. Google would be charging very huge sums of money if they were offering their products for purchase. But instead, Google invested in data and hence they considered data as their product to be purchased. So, while you are searching google and navigating the web easily, Google tracks all the data provided by you and then understand what you and other various people are searching for. This helps Google to target ads in such a way when you search for something, Google will show you the most relevant information and advertisement. For the advertisers, this will lead to a high increase in their sales and better use of their money.

Facebook and Data Collection

Facebook the well-known company that earns tons of money for its free application. ~~Hate it or love it~~ whether you like it or not, Facebook is free and has many services that are for free. However, what matters is the data. Facebook collects many various data and information that enable them to better sever the Facebook user, display the best-related ad and so many other services. Hence, Facebook value your data and it ~~became~~ was able to turn your ~~data~~ it into ~~cash~~ profit.

To sum up, data analytics is ~~that it is~~ the study of analysis. It is considered an important tool for obtaining useful business insights. Also, it offers tailored solutions and responses to various types of customers. Nowadays, the use of data analytics has gradually improved and expanded offering a lot of benefits to everyone. Additionally, learning data analytics does not need a high degree in statistics. All that is needed is to find and get the right tool, secure and fast platform to analyze data, then you will be able to explain the final data produced. Hence, you will start to make better decisions that will give you a great chance at your competition.

\* Consider adding some sort of activity/excercise for students. I want them to search, look for something, write results or give opinion,… anything that will make they will spend some time while or after reading the lesson.

1. Facebook ad is a shortcut for the sponsored advertisement shown by Facebook. [↑](#footnote-ref-1)
2. Raw data is a type of data that is not processed yet to be used and sometimes it is called source data. [↑](#footnote-ref-2)
3. Data warehouse is a database system that is used for reporting and some analytical purposes. Typically, it stores historical data from various sources. [↑](#footnote-ref-3)
4. Business Intelligence is a group of processes and steps that transfers raw data into useful and meaningful information. Basically, business intelligence is a complete software that is able to convert data into some actionable knowledge. [↑](#footnote-ref-4)
5. Data mining is the process of arranging large data groups to discover patterns and setup relationships to solve and fix problems. [↑](#footnote-ref-5)
6. Big data is a field that is used to analyze and extract information from data sets that are too complex and large to be dealt with by an ordinary data-processing software. [↑](#footnote-ref-6)
7. Cloud computing simply means storing data and applications and accessing them over the Internet instead of storing and accessing them from your computer's hard drive. Typically, the cloud idiom is just another metaphor for the Internet. [↑](#footnote-ref-7)
8. Amazon Redshift is one of Amazon products. It is a data warehouse product that is cloud based. [↑](#footnote-ref-8)
9. Google BigQuery is a Google product that is which is a data warehouse. Basically, BigQuery enables you to execute the SQL queries using the great processing power of the Google's infrastructure. Simply all you need to do is to move your data into the BigQuery platform and let Google handle the hard work. [↑](#footnote-ref-9)
10. If a patient has a testing positive for the Covid-19 then he or she is infected with the Coronavirus. [↑](#footnote-ref-10)