

A Survey on the Evolution of Data Analytics and the Future in Data Science

Deepa Mary Thomas, Senior Associate, IT Audit, Deloitte

Abstract:- This paper cornerstone on the concepts of Data Analytics, different types of Data Analytics and the evolution of Data Analytics into the latest Prescriptive Data Analytic model.

I hope this paper will be very helpful for everyone who wants to get a clear idea about the concepts of Data Analysis and the Data Analytics

➤ Paper Overview

The respective paper is classified into 5 segments. Section I describes about Data Analytics in an overall way. Section II briefly discuss on each of the phase in the Data Analytics process. Section III states on the 4 different types of Data Analytics. Section IV describes on the evolution of Data Analytics and finally the paper conclude with the section V as the future vision leading to Data Science. The Reference section is included for the easy access of the data contents being discussed.

Keywords:- AI (Artificial Intelligence), DA (Data Analytics), EDA (Exploratory Data Analysis)

I. INTRODUCTION

The Paper mainly discusses on the Importance of Data Analysis and the Data Analytics in the real world of computations. After reading this paper you will definitely get some ideas about the types of Data Analytics and the scoped in underlying areas with the Data Analytics and how to shape your future with the Data Science.

II. DATA ANALYSIS AND DATA ANALYTICS

Data Analysis and Data Analytics both are needed for performing with varying sorts of analytical works in the audit firms.

Data Analysis, as the word says it performs analysis of data. Analysis is really a heuristic activity, where scanning through all the data from top to bottom fashion, the analyst gains some insight. As we perform data analysis more and more, the analyst will be able to get some ideas about each characteristics of data set and thereby we can say that whether the data derived is accurate or not.

As we all know the process in data analysis, that is in this type of analysis the process starts with inspection of data, then follows with a cleaning process (that is detecting and correcting of inaccurate records), followed by the transformation and modeling of data. The goal of the process is to identify useful information with suitable informing of conclusions and with a strong support of decision-making techniques.

Data Analytics on the other hand is different from data analysis.

As research says, Data analytics applications includes many different phases other than just analyzing of the data sets being obtained. Particularly on advanced analytics projects (in Banks, Telecomms, Oil companies and many more firms), much of the required operations takes place upfront, like as the process of gathering of data, integrating of different data sets obtained and preparing and combining of the data sets into one master table and later on developing, debugging and testing and then revising various types of analytical models to ensure that they produce only efficient and accurate results.

Data analytics (DA) is the process of examining various and huge data sets, and after gathering of the data sets then cleanse the data according to the business needs and finally give a conclusion about the data they contain. The data analytics uses many different systems and softwares at each of the phases to give a solid conclusion or results in each of the phases.

Data analytics technologies and techniques are widely used in commercial industries to enable organizations to make more-informed business decisions

Data analytic works can help with business team to increase with their operational efficiency, followed by a higher revenue growth. The outcome of data analytic work can be based on the emerging market trends and also because of the way work can be customized, it will result with more customer satisfaction and also this type of work will give a competitive edge to work with more accuracy. The ultimate goal of taking part with the data analytical work is to give improvement with the business performance.

In short we can say that the Data Analytics is the superset, in which Data Analysis is a part or Data Analysis is the subset of Data Analytics.

➤ The detailed Data Analytic process is shown as below

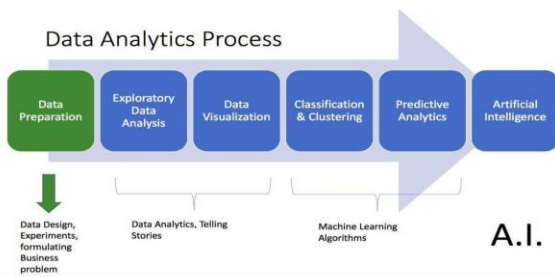


Fig 1

The Data Analytics process starts with the Data Preparation phase. In this phase the data will get collected after having discussions with the respective business teams. Before the data being collected from the respective business or clients, there will be a data gathering phase in which we will have a detailed discussion regarding what data inputs we are looking from the clients. Based on this detailed discussions and the ongoing understanding of data requirements needed, we would like to get the data inputs from the respective clients or entities.

In the data preparation phase, initial processing (of checking of data fields) will happen before the data is being cleaned

Data preparation is the primary process in the analytical work and is more on relative with the understanding of business requirements. In this phase we acquire data sets followed by a formal cleansing of the data being obtained and finally consolidating the large data sets obtained into one file or table and this cleansed data table will be the output of the data preparation phase. This output data table will be meaning and can be used further for our analysis. This data table will be finalized based on the business requirements and the understanding of business needs.

The second phase leads into the Exploratory Data Analysis phase. In Data Analytics, people may wonder how to get started after receiving a large dataset. This is where Exploratory Data Analysis (EDA) comes into action. The main task in this phase is to gain an understanding of data and find clues from the data and to formulate Assumptions and Hypothesis for our modelling. Then in this phase, to check the Quality of data for further processing and cleaning if necessary.

The main task of Exploratory Data Analysis is as follows,

- EDA is the initial step for all Data Analytics projects
- EDA “is an approach to analyze datasets to summarize their main characteristics, often with visual methods.
- EDA - knowing about the data sets, gaining a certain amount of familiarity with the data, before one starts to extract insights from it

After Exploratory Data Analysis, the next phase is Data Visualization. Data Visualization is the graphical representation of information and data. By using visual elements like charts, graphs, and maps, data visualization tools provide an accessible way to see and understand trends, outliers, and patterns in data. This phase involves producing images that communicate relationships among the represented data to viewers of the images

One of the data visualization tool, Power BI is a business analytics service by Microsoft. It aims to provide interactive visualizations and business intelligence capabilities with an interface simple enough for end users to create their own reports and dashboards

Then the Classification and Clustering of data sets will come into action. In the entities, the business people are the owners of the data and they are actually classifying the data based on the criticality and the relevance of the data sets. Although the data classification and clustering techniques have certain similarities, the difference lies in the fact that classification uses predefined classes in which objects are assigned, while clustering identifies similarities between objects, and then groups it according to those characteristics which are in common. We can say that data classification in Data Analytics means, it is a technique where we categorize data into a given number of classes. The main goal of a classification problem is to identify the category/class to which a new data will fall under. Classification can be performed on structured or unstructured data.

Clustering aspects briefly concludes that it is a process of classifying the population (or data sets) or data points into a number of different groups such that data points in the same groups are more similar to other data points in the same group than those in other groups. The aim of clustering is to segregate groups with similar characteristics and assign them into different clusters.

Classification and Clustering is part of Data Mining techniques in short.

Some of the Data Mining tools which are available in market are as follows



Fig 2

The final and more detailed phase with the Data Analytics is the Artificial Intelligence (AI). It is the broader concept of machines being able to carry out tasks in a way that we would consider “smart”. Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. The overall research goal of artificial intelligence is to create technology that allows computers and machines to function in an intelligent manner

We all know the example - 'Alexa'. She can be queried about the weather, stream news and music on demand and serves as a robotic assistant that responds to voice commands to control home lighting. Other platforms being used are as follows, Microsoft Azure AI Platform, Google Cloud AI Platform, IBM Watson, Infosys Nia, Dialogflow and BigML

III. TYPES OF DATA ANALYTICS

Depending on different stage of the workflow and the requirement of Data Analysis, there are four main kinds of Data Analytics – **Descriptive, Diagnostic, Predictive and Prescriptive**

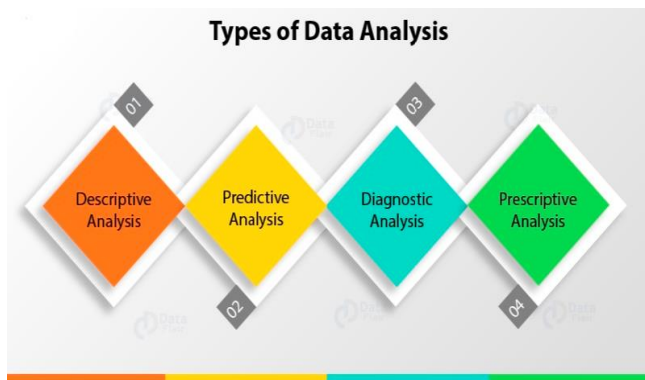


Fig 3

➤ *Descriptive Data Analysis*

It is the simple way to describe our data, that means, “**What happened to the Data**” or try to “**Figure it out what is going on**”

Descriptive statistics are important because of the fact that if we simply present our huge raw data sets, it would be very hard to make us visualize and conclude what the data was showing, especially if the data sets are huge in size and is impossible to conclude with human eyes.

Here comes the role of Descriptive statistics. This type of statistics help us to present the data in a very more meaningful way, which allows simpler interpretation of the data (with the help of graphic analysis – like graphs, charts, pie, stacked bar charts etc)

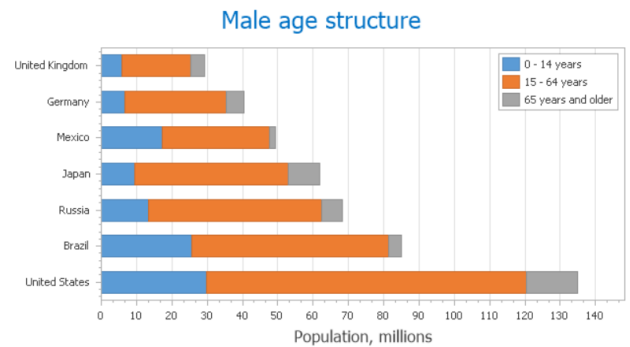


Fig 4

➤ *Diagnostic Data Analysis or Deeper Data Analysis*

It is a form of advanced analytics that examines data or content to answer the question, “**Why did it happen?**” It is characterized by techniques such as Drill-Down (to root cause), Data Discovery, Data Mining and Correlations

Drill Down - *Drill down* is a capability that takes the user from a more general view of the *data* to a more specific one. For example, a report that shows sales revenue by state can allow the user to select a state, click on it and see sales revenue by county or city within that state.

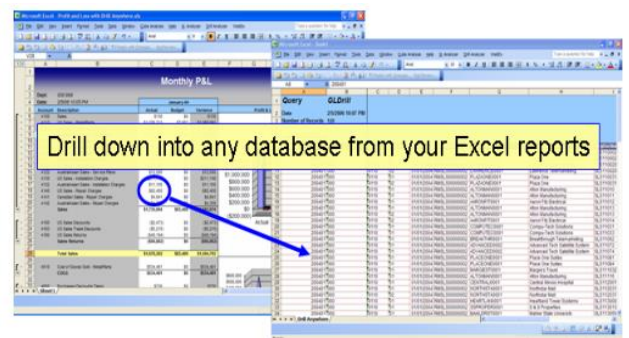


Fig 5

Data Discovery - Data discovery is the collection and analysis of data from various sources to gain insight from hidden patterns and trends. Through the data discovery process, data is gathered, combined, and analyzed in a sequence of steps. The goal is to make messy and scattered data clean, understandable, and user-friendly

Data Mining - Data Mining is defined as a process used to extract usable data from a larger set of any raw data. It implies analyzing data patterns in large batches of data using one or more software. Data mining is also known as Knowledge Discovery in Data (KDD)

Correlations - Correlation is often used as a preliminary technique to discover relationships between variables

➤ *Predictive Data Analysis*

It's a type of Data Analysis and is more complex in nature.

In the Predictive analytics it forecasts “**what will happen**” in the near future.

Predictive analytics encompasses a variety of statistical techniques from data mining, predictive modelling, machine learning and artificial intelligence that analyze current and historical facts to make predictions about future or otherwise unknown events



Fig 6

➤ *Prescriptive Data Analysis*

This is the advanced form of Data Analytics. Prescriptive Analysis provide intelligent recommendations for the optimal next steps for almost any application or business process to drive desired outcomes or accelerate results.

Prescriptive Analysis makes use of machine learning to help businesses decide a course of action based on a computer program’s predictions. Prescriptive analytics works with predictive analytics, which uses data to determine near-term outcomes

➤ *Potential value to business against the complexity of analysis*

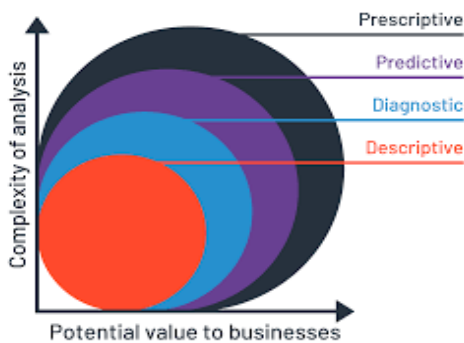


Fig 7

➤ *Pictorial representation for “types” of Data Analysis*

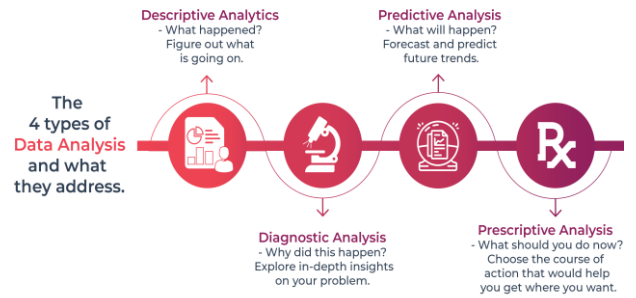


Fig 8

Based on the difficulty of the analysis and the wide aspects of the value of each analysis we can plot the types of data analytics as follows

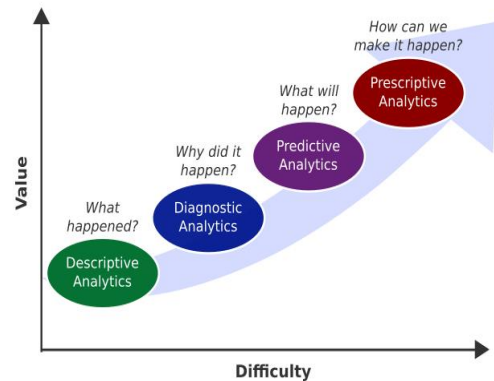


Fig 9

IV. EVOLUTION OF DATA ANALYTICS

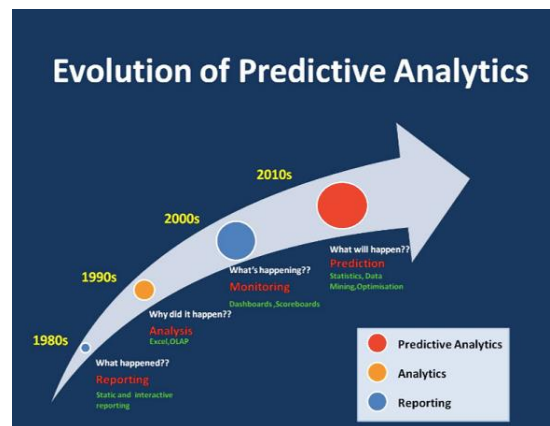


Fig 10

The pictorial representation of the evolution of Data Analytics shows that the concept of Data Analytics started in the early 1980s. In 1980’s the Data Analytics is used in such a way that only reporting is used to happen. That means what is happening with the data being obtained. In this period only the static and interactive reporting is being carried out. After this type of Data Analytic modeling, the Data Analytic is being moved into the second phase that is with early 1990’s more of Analysis (Analytics) came into

existence. In this period, it focuses on “why did it happen” to the data. Then in 2000 onwards, the Monitoring of data happens. The dashboards and the scoreboards are being used for the same. With this type of analysis, a clear idea of what’s happening to the data is being understood. Then after 2010 onwards, the Prediction with the data and the data inputs being implemented with. That means, what will happen with the data is the main question being asked in the period after 2010. The different methods of statistics, data mining and the optimization is being used in this period. Now we are in the era with the more detailed data analytics and that is of nature Prescriptive. In this period we are training our machines to be smarter and focusing on the computations to happen with less time and less efforts. So we can conclude that we are in the period with more of AI.

➤ Advantage of Data Analytics

Data Analytics in firms involves mostly with the analysis of complete sets of data being obtained to find out the inconsistencies and trends for further exploration, as well as to provide the evidence.

If we say the entity as an audit firm, then one of the advantage of data analytics is the higher-quality audit evidence, since the auditor can now examine far more data than had previously been possible with the audit sampling. Data analytics in audit firms can include many benefits including giving a better coverage and assurance of the data being obtained, also can enable a much better and risk focused audits. With the support of data analytics we will have a clear understanding of the data being collected and also can deep dive and can conclude on the outputs.

V. FUTURE SCOPE – LEADING TO DATA SCIENCE

As we all know, Data science draws knowledge from several fields like data mining, data clustering, machine learning and big data and uses many different scientific principles and methods, various processes and many variant algorithms and systems to extract knowledge and insights from many structural, messy and unstructured data.

With the concept of Data Science we can select with varying and of unique carrier opportunities. The below picture shows the intersection of the fields coming under the Data Science

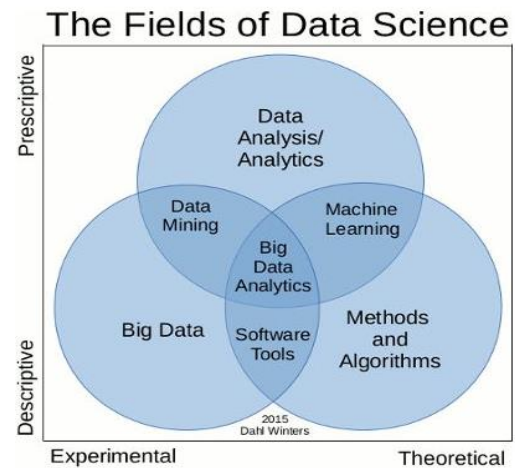


Fig 11

So one with a focused mind and if he/she is more into with the Data Science, then they can become a Data Scientist. We can conclude that the Data scientists work closely with the business stakeholders and their business needs and the respective requirements to solve the problems and will finalize on the goals and conclude on with the collected data sets how to achieve on the predefined goals. The data scientists will design on various data models and their processes, write respective algorithms and make predictive models, to extract the data the business needs, then help analyze the data and share insights with peers.

REFERENCES

- [1]. <http://www.journalofaccountancy.com/issues/2007/jul/ariskbasedapproachtojournalentrytesting.html>
- [2]. <http://www.ifac.org/system/files/downloads/a012-2010-iaasb-handbook-isa-240.pdf>
- [3]. **Exploratory Data Analysis: A Practical Guide and Template for Structured Data** - <https://towardsdatascience.com/exploratory-data-analysis-eda-a-practical-guide-and-template-for-structured-data-abfbf3ee3bd9?gi=f2c75f1c52c>
- [4]. Concept of Generalized Audit Softwares(ACL and IDEA) in Journal Entry Testing, by Deepa Mary Thomas: http://www.ijeam.com/55_1.php