Unit I

Business analytics: Business analytics (BA) refers to the skills, technologies, and practices for continuous iterative exploration and investigation of past business performance to gain insight and drive business planning. Focuses on developing new insights and understanding of business performance based on data and statistical methods. Business Intelligence - Traditionally focuses on using a consistent set of metrics to both measure past performance and guide business planning or focuses on description, while business analytics focuses on prediction and prescription.

Table 1.3 Characteristics of Analytics, Business Analytics, and Business Intelligence

Characteristics	Analytics	Business Analytics (BA)	Business Intelligence (BI)
Business performance planning role	What is happening, and what will be happening?	What is happening now, what will be happening, and what is the best strategy to deal with it?	What is happening now, and what have we done in the past to deal with it?
Use of descriptive analytics as a major component of analysis	Yes	Yes	Yes
Use of predictive analytics as a major component of analysis	Yes	Yes	No (only historically)
Use of prescriptive analytics as a major component of analysis	Yes	Yes	No (only historically)
Use of all three in combination	No	Yes	No
Business focus	Maybe	Yes	Yes
Focus of storing and maintaining data	No	No	Yes
Required focus of improving business value and performance	No	Yes	No

~~~~~~

Importance of Business Analytics

- Business analytics is a methodology or tool to make a sound commercial decision. Hence it impacts functioning of the whole organization. Therefore, business analytics can help improve profitability of the business, increase market share and revenue and provide better return to a shareholder.
- Facilitates better understanding of available primary and secondary data, which again affect operational efficiency of several departments.
- Provides a competitive advantage to companies. In this digital age flow of information is almost equal to all the players. It is how this information is utilized makes the company competitive. Business analytics combines available data with various well thought models to improve business decisions.
- Converts available data into valuable information. This information can be presented in any required format, comfortable to the decision maker.

~~~~~~

Scope of Business Analytics: Business analytics has a wide range of application and usages. It can be used for descriptive analysis in which data is utilized to understand past and present situation. This kind of descriptive analysis is used to asses' current market position of the company and effectiveness of previous business decision. It is used for predictive analysis, which is typical used to asses' previous business performance. Business analytics is also used for prescriptive analysis, which is utilized to formulate optimization techniques for stronger business performance. For example, business analytics is used to determine pricing of various products in a departmental store based past and present set of information. Application - Business analytics has a wide range of application from customer relationship management, financial management, and marketing, supply-chain management, human-resource management, pricing and even in sports through team game strategies.

TYPES OF ANALYTICS – DESCRIPTIVE, PREDICTIVE, PRESCRIPTIVE

Table 1.2 Analytic Purposes and Tools

Type of Analytics	Purpose	Examples of Methodologies
Descriptive	To identify possible trends in large data sets or databases. The purpose is to get a rough picture of what generally the data looks like and what criteria might have potential for identifying trends or future business behavior.	Descriptive statistics, including measures of central tendency (mean, median, mode), measures of dispersion (standard deviation), charts, graphs, sorting methods, frequency distributions, probability distributions, and sampling methods.
Predictive	To build predictive models designed to identify and predict future trends.	Statistical methods like multiple regression and ANOVA. Information system methods like data mining and sorting. Operations research methods like forecasting models.
Prescriptive	To allocate resources optimally to take advantage of predicted trends or future opportunities.	Operations research methodologies like linear programming and decision theory.

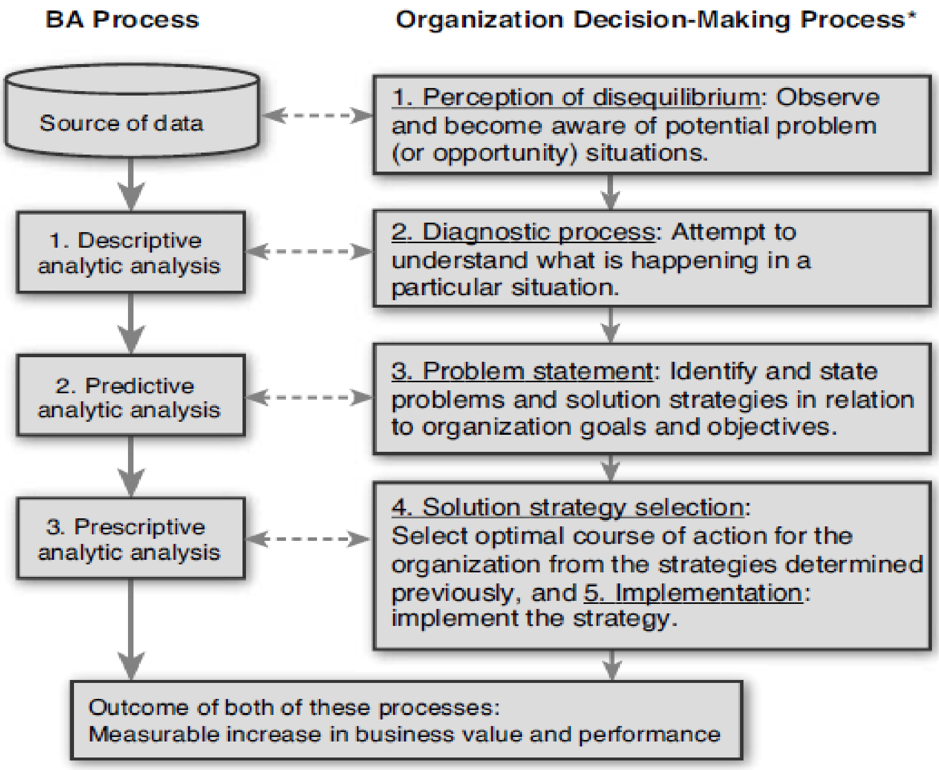

Figure 1.2 Comparison of business analytics and organization decision-making processes

Business Analytics Process: It involves asking questions, looking at data, and manipulating it to find the required answers. Since the approach to business is different for different organizations, their solutions and their ways to reach the solutions are also different. The image given below demonstrates the steps in Business Analytics process of a firm:

Relationship of Business Analytics Process and organisation,

6 Steps in the Business Analytics Process

1. **Step 1: Identifying the Problem:** The first step of the process is identifying the business problem. This is a crucial stage in Business Analytics as it is important to clearly understand what the expected outcome should be. When the desired outcome is determined, it is further broken down into smaller goals. Then, business stakeholders decide the relevant data required to solve the problem. Some important questions must be answered in this stage, such as: What kind of data is available? Is there sufficient data? And so on.
2. **Step 2: Exploring Data:** Once the problem statement is defined, the next step is to gather data (if required) and, more importantly, cleanse the data—most organizations would have plenty of data, but not all data points would be accurate or useful. Organizations collect huge amounts of data through different methods, but at times, junk data or empty data points would be present in the dataset. These faulty pieces of data can hamper the analysis. Hence, it is very important to clean the data that has to be analyzed.

   TO do this,

   - Do computations for the missing data,
   - Remove outliers,
   - And find new variables as a combination of other variables.
   - You may also need to plot time series graphs as they generally indicate patterns and outliers.
   - It is very important to remove outliers as they can have a heavy impact on the accuracy of the model that you create. Moreover, cleaning the data helps you get a better sense of the dataset.
3. **Step 3: Analysis:** Once the data is ready, the next thing to do is analyze it. Now to execute the same, there are various kinds of statistical methods (such as hypothesis testing, correlation, etc.) involved to find out the insights that you are looking for. You can use all of the methods for which you have the data. The prime way of analyzing is pivoting around the target variable, so you need to take into account whatever factors that affect the target variable. In addition to that, a lot of assumptions are also considered to find out what the outcomes can be. Generally, at this step, the data is sliced, and the comparisons are made. Through these methods, you are looking to get actionable insights.
4. **Step 4: Prediction and Optimization**: Business Analytics is all about being proactive. In this step, we use prediction techniques, such as neural networks or decision trees, to model the data. These prediction techniques will help to find out hidden insights and relationships between variables, which will further help to uncover patterns on the most important metrics. By principle, a lot of models are used simultaneously, and the models with the most accuracy are chosen. In this stage, a lot of conditions are also checked as parameters, and answers to a lot of 'what if…?' questions are provided.
5. **Step 5: Making a Decision and Evaluating the Outcome**: From the **insights** that you receive from your model built on target variables, **a viable plan of action will be established** in this step to meet the organization's goals and expectations. The said plan of action is then put to work, and the waiting period begins. We have to wait to see the actual outcomes of your predictions and find out how successful you were in your endeavours. Once you get the outcomes, you will have to measure and evaluate them.
6. **Step 6: Optimizing and Updating:** Post the implementations of the solution, the outcomes are measured as mentioned above. If we find some methods through which the plan of action can be optimized, then those can be implemented. If that is not the case, then you can move on with registering the outcomes of the entire process. This step is crucial for any analytics in the future because you will have an ever-improving database. Through this database, you can get closer and closer to maximum optimization. In this step, it is also important to evaluate the ROI (return on investment). Take a look at the diagram below of the life cycle of business analytics.

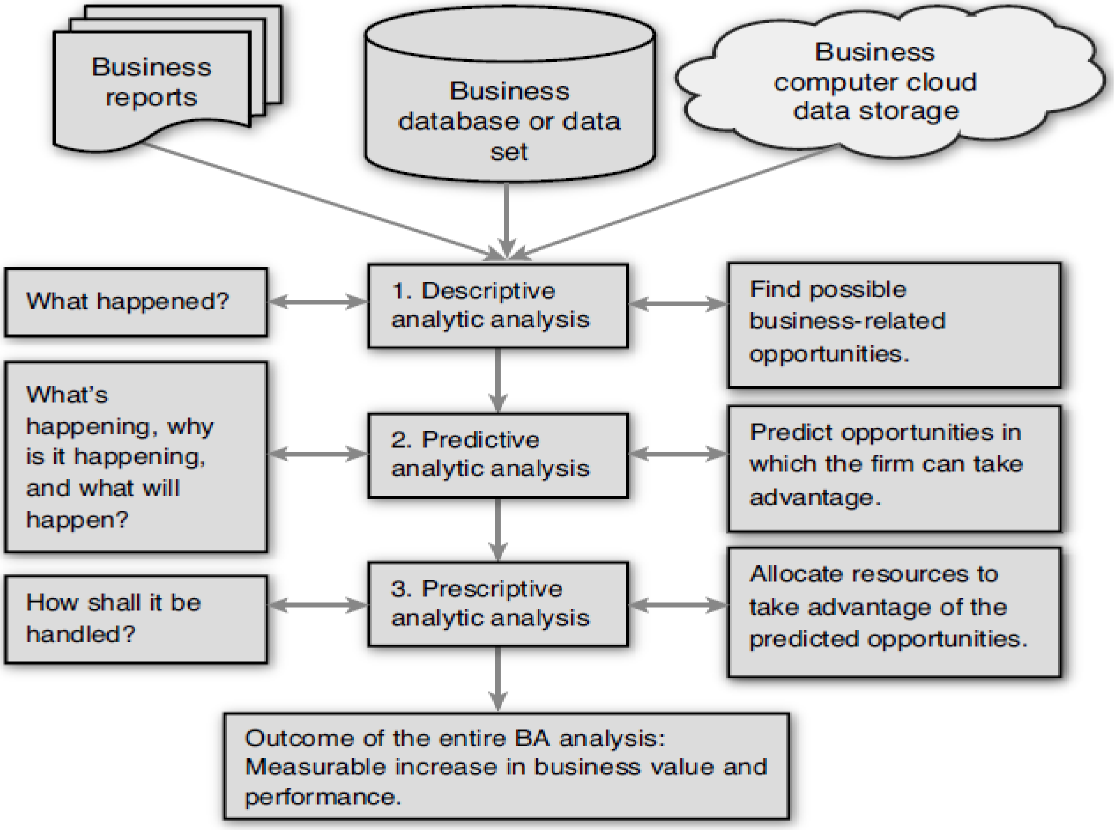

Figure 1.1 Business analytics process

~~~~~~

TYPES OF DATA
~~~~~~

Table 1.4 Types of Data Measurement Classification Scales

Type of Data Measurement Scale	Description
Categorical Data	Data that is grouped by one or more characteristics. Categorical data usually involves cardinal numbers counted or expressed as percentages. Example 1: Product markets that can be characterized by categories of "high-end" products or "low-income" products, based on dollar sales. It is common to use this term to apply to data sets that contain items identified by categories as well as observations summarized in cross-tabulations or contingency tables.
Ordinal Data	Data that is ranked or ordered to show relational preference. Example 1: Football team rankings not based on points scored but on wins. Example 2: Ranking of business firms based on product quality.
Interval Data	Data that is arranged along a scale, in which each value is equally distant from others. It is ordinal data. Example 1: A temperature gauge. Example 2: A survey instrument using a Likert scale (that is, 1, 2, 3, 4, 5, 6, 7), where 1 to 2 is perceived as equidistant to the interval from 2 to 3, and so on. Note: In ordinal data, the ranking of firms might vary greatly from first place to second, but in interval data, they would have to be relationally proportional.
Ratio Data	Data expressed as a ratio on a continuous scale. Example 1: The ratio of firms with green manufacturing programs is twice that of firms without such a program.

~~~~~~

Competitive advantages of Business Analytics

- Business analytics drives competitive advantage by generating economies of scale, economies of scope, and quality improvement.
- Taking advantage of the economies of scale is the first way organizations achieve comparative cost efficiencies and drive competitive advantage against their peers.
- Taking advantage of the economies of scope is the second way organizations achieve comparative cost efficiencies and drive competitive advantage against their peers.
- Measuring the impact of insights and activities on quality and using them to drive continuous improvements is the final way organizations leverage business analytics to drive competitive advantage.

~~~~~~

Statistical Notation - This web page describes how symbols are used on the Stat Trek web site to represent numbers, variables, parameters, statistics, etc.

~~~~~~

Descriptive Statistical methods - Descriptive statistics are used to describe the basic features of the data in a study. They provide simple summaries about the sample and the measures. Together with simple graphics analysis, they form the basis of virtually every quantitative analysis of data.

~~~~~~

Sampling and Estimation methods overview.

- In general, a sample size of 30 or larger can be considered large.
- An estimator is a formula for estimating a parameter. An estimate is a particular value that we calculate from a sample by using an estimator because an estimator or statistic is a random variable; it is described by some probability distribution.
- We refer to the distribution of an estimator as its sampling distribution. The standard deviation of the sampling distribution of the sample mean is called the standard error of the sample mean.

~~~~~~

The desirable properties of an estimator are

1. **Unbiasedness** - An estimator is said to be unbiased if its expected value is identical with the population parameter being estimated. That is if θ is an unbiased estimate of θ, then we must have E (θ) = θ. Many estimators are "Asymptotically unbiased" in the sense that the biases reduce to practically insignificant value (Zero) when n becomes
~~~~~~

sufficiently large. The estimator S2 is an example. It should be noted that bias is estimation is not necessarily undesirable. It may turn out to be an asset in some situations.

2. **Consistency** - If an estimator, say θ, approaches the parameter θ closer and closer as the sample size n increases, θ is said to be a consistent estimator of θ. Stating somewhat more rigorously, the estimator θ is said is be a consistent estimator of θ if, as n approaches infinity, the probability approaches 1 that θ will differ from the parameter θ by no more than an arbitrary constant. The sample mean is an unbiased estimator of μ no matter what form the population distribution assumes, while the sample median is an unbiased estimate of μ only if the population distribution is symmetrical. The sample mean is better than the sample median as an estimate of μ in terms of both unbiasedness and consistency.
3. **Efficiency** - The concept of efficiency refers to the sampling variability of an estimator. If two competing estimators are both unbiased, the one with the smaller variance (for a given sample size) is said to be relatively more efficient. Stated in a somewhat different language, an estimator θ is said to be more efficient than another estimator θ2 for θ if the variance of the first is less than the variance of the second. The smaller the variance of the estimator, the more concentrated is the distribution of the estimator around the parameter being estimated and, therefore, the better this estimator is.
4. **Sufficiency** - An estimator is said to be sufficient if it conveys much information as is possible about the parameter which is contained in the sample. The significance of sufficiency lies in the fact that if a sufficient estimator exists, it is absolutely unnecessary to considered any other estimator; a sufficient estimator ensures that all information a sample a sample can furnished with respect to the estimation of a parameter is being utilized. Many methods have been devised for estimating parameters that may provide estimators satisfying these properties. The two important methods are the least square method and the method of maximum likelihood.

~~~~~~

The two types of estimates of a parameter are

- **Point estimates** - A point estimate is a single number that we use to estimate a parameter.
- **Interval estimates** - An interval estimate is a range of values that brackets the population parameter with some probability.

~~~~~~

A confidence interval is an interval for which we can assert with a given probability 1 − α, called the degree of confidence, that it will contain the parameter it is intended to estimate. This measure is often referred to as the 100(1 − α)% confidence interval for the parameter. A 100(1 − α) % confidence interval for a parameter has the following structure: Point estimate ± Reliability factor × Standard error, where the reliability factor is a number based on the assumed distribution of the point estimate and the degree of confidence (1 − α) for the confidence interval and where standard error is the standard error of the sample statistic providing the point estimate.

~~~~~~
~~~~~~