MIT | School of Distance Education

Chapter 8

Descriptive, Predictive, Prescriptive and Diagnostic Analytics

Chapter Index

Learning Objectives

- Explain the importance of descriptive analytics
- Discuss the fundamentals of descriptive statistics
- Describe the need of predictive analytics
- Discuss about predictive modelling
- Elucidate the overview of prescriptive analytics
- Explain the need of diagnostic analysis

Descriptive Analytics

Using the data, descriptive analytics helps to answer "What has occurred in the corporation" and "What is going on now?".

Three crucial approaches to abridge and describe the raw data are:

Dashboards and MIS reporting

This technique gives condensed data giving information on "What has happened", "What's been going on?" and "How can it stand with the plan?".

Impromptu detailing

This technique supplements the past strategy in helping the administration to extract the information as required.

Drill-down reporting

This is the most complex piece of descriptive analysis and gives the capacity to delve further into any report to comprehend the information better.

Descriptive Statistics

Statistics involves collecting, organising, analysing, interpreting and presenting data.

Microsoft Excel supports statistical analysis in two ways.

- With statistical functions that are entered in worksheet cells directly or embedded in formulas.
- With the Excel Analysis Toolpak add-in to perform more complex statistical computations.

A population consists of all items of interest for a particular decision/investigation.

A sample is a subset of a population.

Understanding Statistical Notation

In general, x_i represents the i^{th} observation.

Population measures: Represented by Greek letters, such as σ (sigma), μ (mu), and π (pi).

Sample Statistics: Represented by italic letters such as by (x-bar), s, and p.

N: To represent the number of items in a population.
n: To represent the number of observations in a sample.

Summation (Greek capital sigma): means that the terms that follow it are added together.

Central Tendency

Central tendency is the measurement of a single value that attempts to describe a set of data by identifying the central position within that set of data.

Mean

- The mathematical average is called the mean which is the sum of the observations divided by the total number of observations.
- The population mean is represented by "μ" and the sample mean by $\bar{x}$.
- If the population contains N observations $x_1, x_2, \ldots, x_N$ then, $\mu = \frac{\sum_{i=1}^{n} x_i}{N}$
- The sample mean of n sample observations $x_1, x_2, \ldots, x_N$ is $= \frac{\sum_{i=1}^{n} x_i}{N}$
- One property of the mean is that the sum of the deviations of each observation from the mean is zero: $\sum_i (x_i - \bar{x}) = 0$. This means that the sum of the deviations above the mean is the same as the sum of the deviations below the mean.
- Excel Function: AVERAGE(Data range)

Median

- The measure of location that specifies the middle value when the data are arranged from least to greatest is the median.
- If the number of observations is odd (say 9), the median is the exact middle of the sorted numbers (5^{th} observation).
- If the number of observations is even (say 8), the median is the mean of the two middle numbers (that is 4^{th} and 5^{th} observation).
- Excel function: MEDIAN(data range)

Mode

- It is the observation/number/series that occurs the maximum number of times.
- You can easily identify the mode from a frequency distribution by identifying the value having the largest frequency or from a histogram by identifying the highest bar.
- Excel function: MODE.SNGL(Data range)

Midrange

This is simply the average of the greatest and least values in the data set.

Variability

A commonly used measure of dispersion is the variance.
The bigger the variance is, the more is the spread of the observations from the mean which indicates more variability in the data.

- The formula for the variance of a population is, $\sigma^2 = \frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}$; where x_i is the value of the i^{th} item, N is the number of items in the population, and μ is the population mean.
- Excel function to calculate population variance: VAR.P(Data range)
- The formula for the variance of a sample is, $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$; where n is the number of items in the sample and $\bar{x}$ is the sample mean.
- Excel function to calculate population variance: VAR.S(Data range)

Standard Deviation

The square root of the variance is the standard deviation.

The population standard deviation is computed as	$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \mu)^2}{N}}$
The sample standard deviation is computed as	$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{N}}$

Excel Function:

Population Standard Deviation: STDEV.P(Data Range)

Sample Standard Deviation: STDEV.S(Data Range)

The standard deviation has similar measure units that of same as the data units. Thus, it can be more easily related to the mean or other statistics measured in the same units.

Standardised Values

A z-score, or standardised value, provides a measure of the distance of the observation away from the mean, irrespective of the measurement units.

In a data set, z-score for the i^{th} observation is calculated as

- We subtract the sample mean from the i^{th} observation, x_i, and divide the result by the sample standard deviation.
- That is, $Z_i = \frac{x_i - \bar{x}}{s}$

- A z-score of 1.0 means that the observation is one standard deviation to the right of the mean.
- A z-score of –1.5 means that the observation is 1.5 standard deviations to the left of the mean.

- Excel function: STANDARDIZE (x, mean, standard_dev)

Coefficient of Variation

The coefficient of variation (CV) provides a relative measure of the dispersion in data relative to the mean and is defined as:

$$CV = \frac{Standard\ Deviation}{Mean}$$

Often, the coefficient of variation is multiplied by 100 to be expressed as a percentage.

The CV is useful when comparing the variability of two or more data sets when their scales differ.

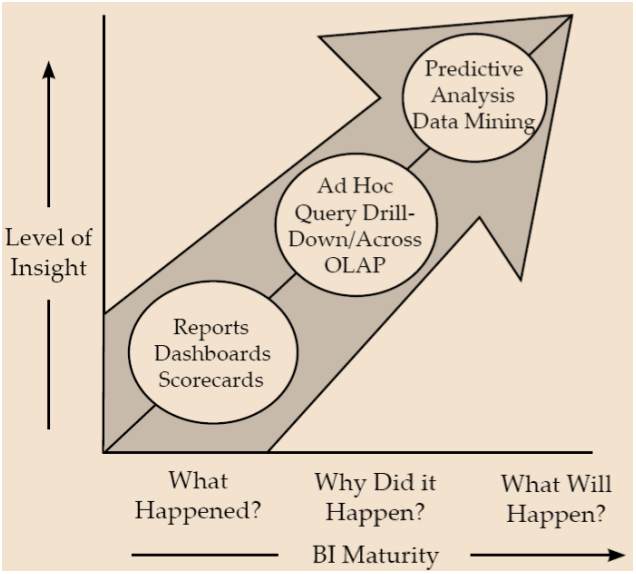
The smaller the coefficient of variation, the smaller the relative risk is for the return provided.

Return to Risk: The reciprocal of the coefficient of variation. If the objective is to maximise return, a higher return-to-risk ratio is often considered better.

Predictive Analytics

Predictive analytics is about understanding and predicting the future and answers the question 'What could happen?' by using statistical models and different forecast techniques.

The steps involved in predictive analytics



Source: http://www.witinc.com/predictive-analytics.id.355.htm

Predictive Modelling

Predictive modelling is the method of making, testing and authenticating a model to best predict the likelihood of a conclusion.

Predictive models in business context, are used to analyse historical facts and current data to better comprehend customer habits, partners and products and to classify possible risks and prospects for a company.

It practices many procedures, including statistical modelling, data mining and machine learning to aid analysts make better future business predictions.

Predictive Model

- They are representations of the relationship between how a member of a sample performs and some of the known characteristics of the sample.
- This model is used a lot in marketing which helps identify implied patterns which indicate customers' preferences.
- This model can perform calculations at the exact time that a customer performs a transaction.

The various business processes on predictive modelling are:

Creating the model

A software based solution allows you to make a model to multiple algorithms on the dataset.

Testing the model

Test the predictive model on the dataset. In some situations, the testing is done on previous data to the effectiveness of a model's prediction.

Authenticating the model

Authenticate the model results by means of business data understanding and visualisation tools.

Assessing the model

Assessing the best suited model from the used models and selecting the appropriate model tailored for the data.

Logic-Driven Models

Logic driven models are created on the basis of inferences and postulations which the sample space and existing conditions provide.

Example:

- Consider a customer who visits a restaurant around six times in a year and spends around ₹5000 per visit. The restaurant gets around 40% margin on per visit billing amount.
- The annual gross profit on that customer turns out to be **5000 × 6 × 0.40 = ₹12000**.
- 30% of the customers do not return each year, while 70% do return to provide more business to the restaurant.
- Assuming the average lifetime of a customer is 1/.3 = 3.33 years, the average gross profit for a typical customer turns out to be **12000 × 3.33 = ₹39,960.**

Using the above example we can logically arrive at a conclusion that

Economic Value of each Customer (V) = $\frac{R \times F \times M}{D}$

where, R = Revenue generated per customer; F = Frequency of visits per year; M = Profit margin; D = Defection rate

Data Driven Models

It finds the links between the state system variables (input and output) without clear knowledge of the physical attributes and behaviour of the system.

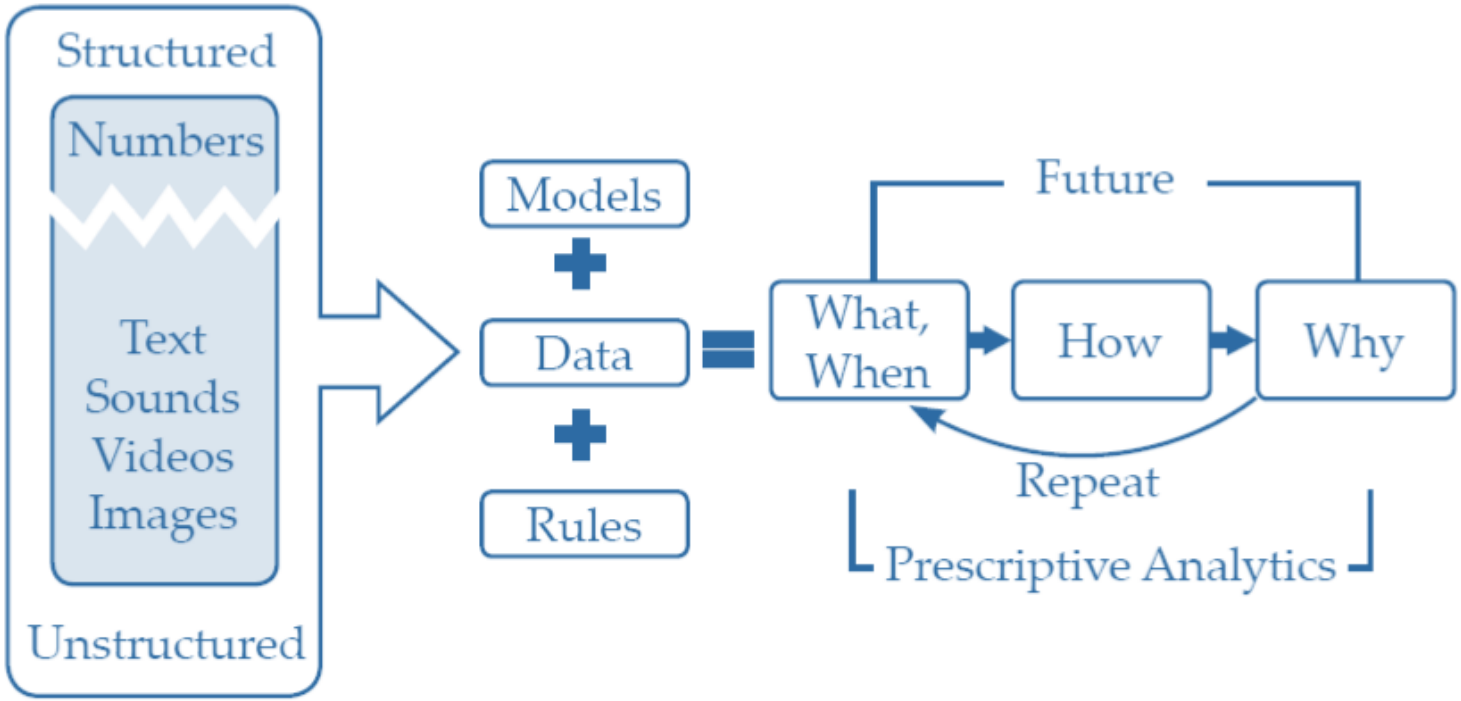
The data driven predictive modelling derives the modelling method based on the set of existing data and entails a predictive methodology to forecast the future outcomes.

Example: A company expecting losses in the current quarter due to the poor market performance and sentiments. In this case you are simply predicting the outcomes based on the data though the data inference is known.

Overview of Prescriptive Analytics

Using the optimisation technique, prescriptive analytics determines the finest substitute to minimise or maximise some equitable finance, marketing and many other areas.

Diagrammatic representation of the stages involved in the prescriptive analytics.



Prescriptive analytics

- Prescriptive analytics go beyond predictions, workforce optimisations and decision options.
- It is usually used to analyse complex data to analyse huge complex data to forecast outcomes, offer decision options and show alternative business impact.
- This analytics help enterprises to take decisions on how to take advantage of a future scenario or reduce a risk in the future course of time and represent the implication of each decision option.
- In real life, prescriptive analytics can automatically and continuously process new data to improve forecast accuracy and offer better decision options.

How Prescriptive Analytics Functions

Decision-making process of Prescriptive Analytics

- Identifying and breaking down every single potential choice
- Defining potential connections and associations between each of these choices with each other
- Identifying variables that could affect each of these choices (positively or negatively)

Prescriptive analytics utilises procedures like optimisation, game theory, simulation, and decision-analysis techniques.

Commercial Operations and Variability

Prescriptive analytics giving directors a chance to foresee what structures, messages and targets will yield ideal outcomes given the organisation's remarkable parameters, and after that choose which way will give the biggest returns.

Other business applications of prescriptive analytics are:	Optimising spend and rate of profitability (ROI) through exact customer profiling
	Providing important data for brand planning and go-to-market procedures
	Maximising campaign productivity, sales force arrangement and promotional activities
	Predicting and proactively overseeing market events
	Providing significant data for territory examination, customer deals and medical data

Research and Innovation

Research and Development exercises in a competitive industry includes:

- Demonstrating, anticipating and enhancing results from item utility
- Understanding sickness (or different zones of intrigue) patterns/movement
- Establishing ideal trial conditions through focused patient cohorts
- Increasing customer adherence to the item and diminishing compliance
- Understanding necessities for customised drug and different advancements
- Determining and setting up focused items and interventions
- Determining and setting up an ideal trial conditions through focused patient cohorts

Business Development

Key zones for prescriptive analytics are:

- Identifying and settling on choices about circumstances/rising ranges of unmet need
- Predicting the potential advantage
- Proactively following industry trends and actualising techniques to get an advantage
- Exploiting data analytics to distinguish particular buyer populations and regions that ought to be focused on
- Leveraging data analytics to distinguish key advancements for item improvement that will produce the biggest return for the investment
- Identifying likely purchasers to cut business improvement costs altogether, and imagine a scenario where situations for items, markets and purchasers could be an unmistakable differentiator for developing organisations

Customer Excellence

Prescriptive analytics can be utilised to improve purchaser excellence which includes:

- Predicting what purchasers will need and settling on key choices that address those necessities
- Segmenting purchasers and recognising and focusing on custom fitted messages to them
- Staying on top of competition and deciding (e.g., marketing, branding) about items that will prompt more desirable items and higher sales

Corporate Accounts

Corporate account functions can use prescriptive analytics to improve their capacity to settle on choices that help drive internal excellence & external strategy:

Internal excellence

- Viability and direction for non-item related activities; what choices ought to be made and what is the effect
- Viability and direction for item related activities; what choices ought to be made and what is the effect

External-facing key direction

- Utilising important data to demonstrate item esteem and build up a market valuing
- Utilising examination to build up a targeted on coupon strategy
- Recognising ideal price point alternatives and the effect of those choices on the income model for the item
- Better understanding the whole price cycle from rundown cost to repayment to inform the ideal pricing system
- Utilising an important competitor data to build up estimating and get market access

Supply Chain

Prescriptive analytics can help supply chain capacities to predict and make decisions in a few basic areas that includes:

- Forecasting future demand and pricing (e.g., supplies, material, fuel and different components affecting cost to guarantee proper supply)
- Utilising prescriptive analytics to illuminate stock levels, schedule plants, route trucks and different components in the supply chain cycle
- Modifying supplier threat by mining unstructured information regarding value-based information
- Better understanding historical demand examples and product course through supply chain channels, anticipating future examples and settling on choices on future state procedures

Governance, Risk and Compliance

Prescriptive analytics can help associations accomplish consistence through the capacity to anticipate up-coming dangers and settle on the proper mitigation choices.

Utilisation of prescriptive analytics in the region of governance, hazard and compliance incorporates:

- Improving internal review effectiveness
- Notifying third-party arrangement and management
- Classifying patterns related with outlandish spend (e.g., total spend working on this issue of pharma)
- Applying very much learned compliance controls

Diagnostic Analytics

Diagnostic analytics is used to find the root cause of a given situation. It can also be used to find the casual relationships between two or more data sets if the root cause is not detectable.

Functions which broadly cover functions of diagnostic analytics are:

1. **Identifying the problem and events worth investigating**

- Using results of descriptive analysis, the analyst must identify the areas which require further analysis and investigation since they are the ones which raise questions whose answers cannot be found by just looking at the data provided.
- Every one of these causes then can be analysed further using diagnostic analytics to find the root problems or causes.

2. Drilling into the analytics

- Once anomalies are distinct and recognised, the analyst must identify the data source which might be able to direct for the root cause of the anomalies.
- During this process, the analyst may have to look outside the selected data sets to find patters and directions.

3. Identifying casual relationships

- To explain the cause of identified anomalies, unseen relationships are identified by closely observing the events.
- Techniques and concepts such as probability theory, time-series data analytics filtering and regression analysis can be implemented and prove useful l for unravelling the true nature of data.
- Modern methods for diagnostic analytics use machine learning.
- Documentation of the diagnostic analysis must be done in a meaningful way which must state which issue was identified, what data sources were used to analyse and eliminate the issue and which casual relationships between data sets were identified during the analysis.

Thank You