Data Analytics

Module-1

Introduction to Data Analytics

- What is Business

Introduction to Data Analytics

- Business is the act of doing something productive to serve someone's needs and thus earn a living and make the world a better place.

- Business activities are recorded on paper or using electronic media, and then these records become data.

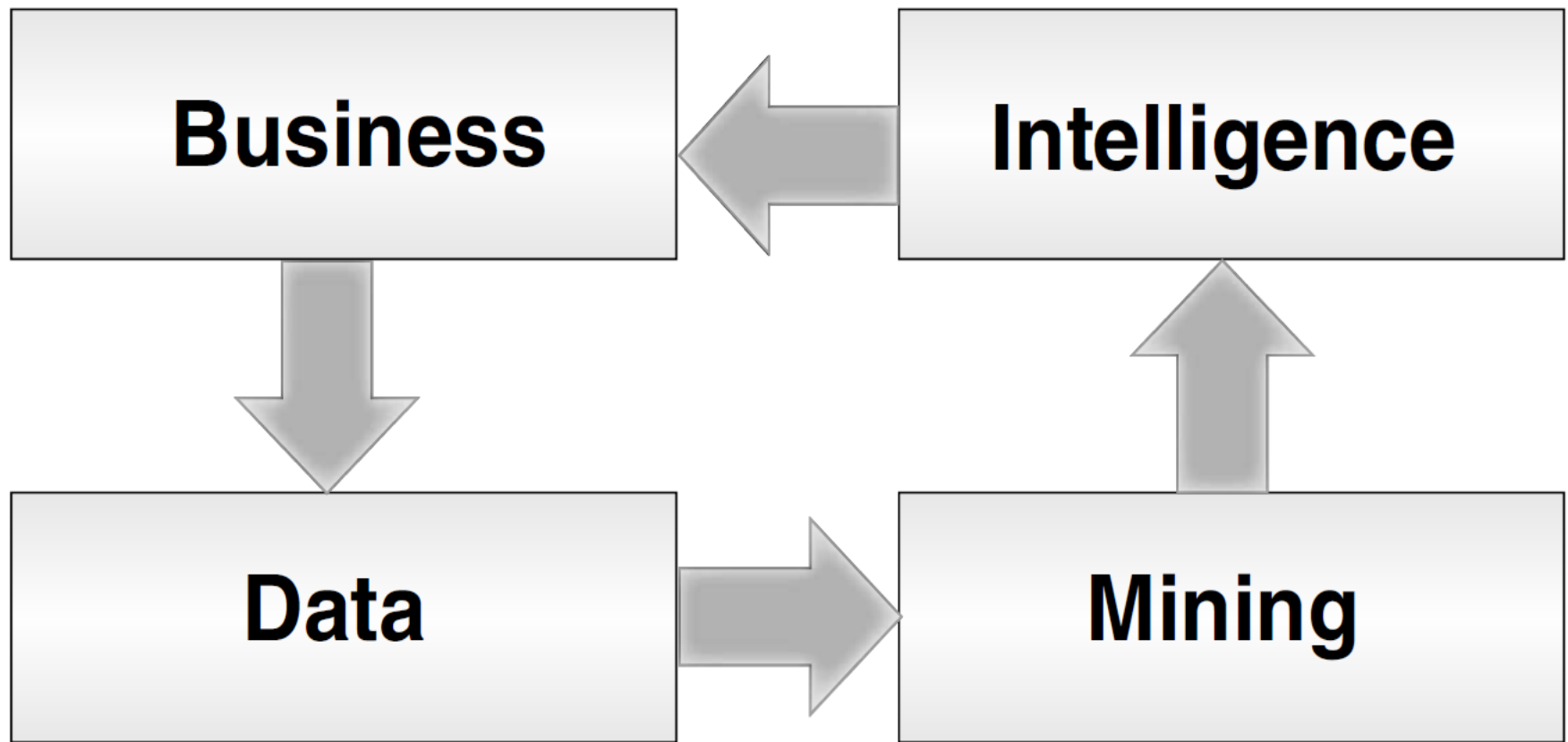- There is more data from customers' responses and on the industry as a whole.

FIGURE 1.1 Business Intelligence and Data Mining (BIDM) Cycle

Introduction to Data Analytics

- All this data can be analyzed and mined using special tools and techniques to generate patterns and intelligence, which reflect how the business is functioning.

- These ideas can then be fed back into the business so that it can evolve to become more effective and efficient in serving customer needs; and the cycle goes on

Introduction to Data Analytics

- These insights and intelligence can help design better
  - promotion plans
  - product   bundles
  - store layouts

- which in turn lead to a better-performing business

Analytics in a Retail store

The vice president of sales

- would want to track the sales to date against monthly targets, the performance of each store and product category the top store managers for that month

The vice president of finance

- would be interested in tracking daily revenue, expense, and cash flows by store; comparing them against plans; measuring cost of capital; and so on.

PATTERN RECOGNITION

- A pattern is a design or model that helps grasp something.

- Patterns help connect things that may not appear to be connected.

- Patterns help cut through complexity and reveal simpler understandable trends.

Pattern types

- Spatial pattern w.r.t space

- TEMPORAL pattern- w.r.t time

- A functional pattern may involve test-taking skills.

- Finding a Pattern
- Uses of Pattern

DATA PROCESSING CHAIN

- -Data can be modeled and stored in a database.

- -Relevant data can be extracted from the operational data stores according to certain reporting and analyzing purposes, and stored in a data warehouse.

- -The data from the warehouse can be combined with other sources of data, and mined using data mining techniques to generate new insights.

- -The insights need to be visualized and communicated to the right audience in real time for competitive advantage.
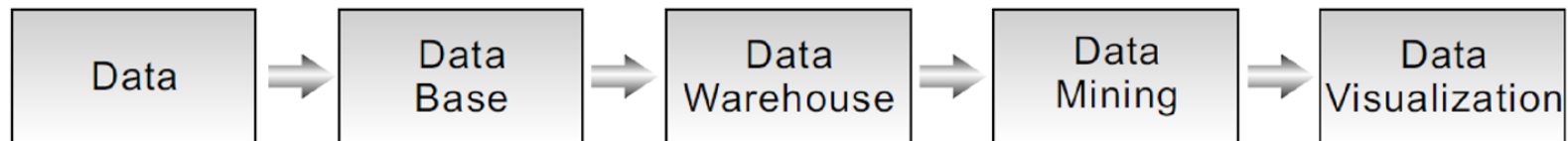

FIGURE 1.2 Data Processing Chain

Dataset 1.1

Movies Transactions Database				
Order No.	Date Sold	Product Name	Location	Amount
1	April 2015	Monty Python	US	$9
2	May 2015	Gone With the Wind	US	$15
3	June 2015	Monty Python	India	$9
4	June 2015	Monty Python	UK	$12
5	July 2015	Matrix	US	$12
6	July 2015	Monty Python	US	$12
7	July 2015	Gone With the Wind	US	$15
8	Aug 2015	Matrix	US	$12
9	Sept 2015	Matrix	India	$12
10	Sept 2015	Monty Python	US	$9
11	Sept 2015	Gone With the Wind	US	$15
12	Sept 2015	Monty Python	India	$9
13	Nov 2015	Gone With the Wind	US	$15

Movies Sales Data Warehouse			
Row No.	Quarter Sold	Product Name	Amount
1	Q2	Gone With the Wind	$15
2	Q2	Monty Python	$30
3	Q3	Gone With the Wind	$30
4	Q3	Matrix	$36
5	Q3	Monty Python	$30
6	Q4	Gone With the Wind	$15
7	Q4	Monty Python	$18

Movies Sales by Quarters – Cross-tabulation				
Quarter/Product	Gone With the Wind	Matrix	Monty Python	Total Sales Amount
Q2	$15	0	$30	$45
Q3	$30	$36	$30	$96
Q4	$15	0	$18	$33
Total Sales Amount	$60	$36	$78	$174

Based on the cross-tabulation above, one can readily answer some product sales questions, like:

1. Which is the best-selling movie by revenue? – *Monty Python*
2. Which is the best quarter by revenue this year? – *Q3*
3. Any other patterns? – *Matrix movie sells only in Q3 (seasonal item)*

Data Mining

- One should select the right data (and ignore the rest), organize it into a precise and imaginative framework that brings relevant data together, and then apply data mining techniques to deduce the right insight.

Data Mining

- Data can be analyzed at multiple levels of granularity and could lead to a large number of interesting combinations of data and interesting patterns.

- Some of the patterns may be more meaningful than the others. Such highly granular data is   often used, especially in finance and high-tech areas, so that one can gain even the slightest edge over the competition

Mining

- Decision Trees
- Regression
- Artificial Neural Networks
- Cluster Analysis
- Association Rule Mining

Data Visualization

FIGURE 1.3 Sample Executive Dashboard

Few considerations when presenting using data:

■ Present the conclusions and not just report the data.

■ Choose wisely from a palette of graphs to suit the data

.

■ Organize the results to make the central point stand out.

■ Ensure that the visuals accurately reflect the numbers. Inappropriate visuals can create misinterpretations and misunderstandings.

Few considerations when presenting using data:

.

■ Make the presentation unique, imaginative and memorable.

- Executive dashboards are designed to provide information on select few variables for every executive.
- They use graphs, dials, and lists to show the status of important parameters.
- These dashboards also have a drill-down capability to enable a root-cause analysis of exception situations
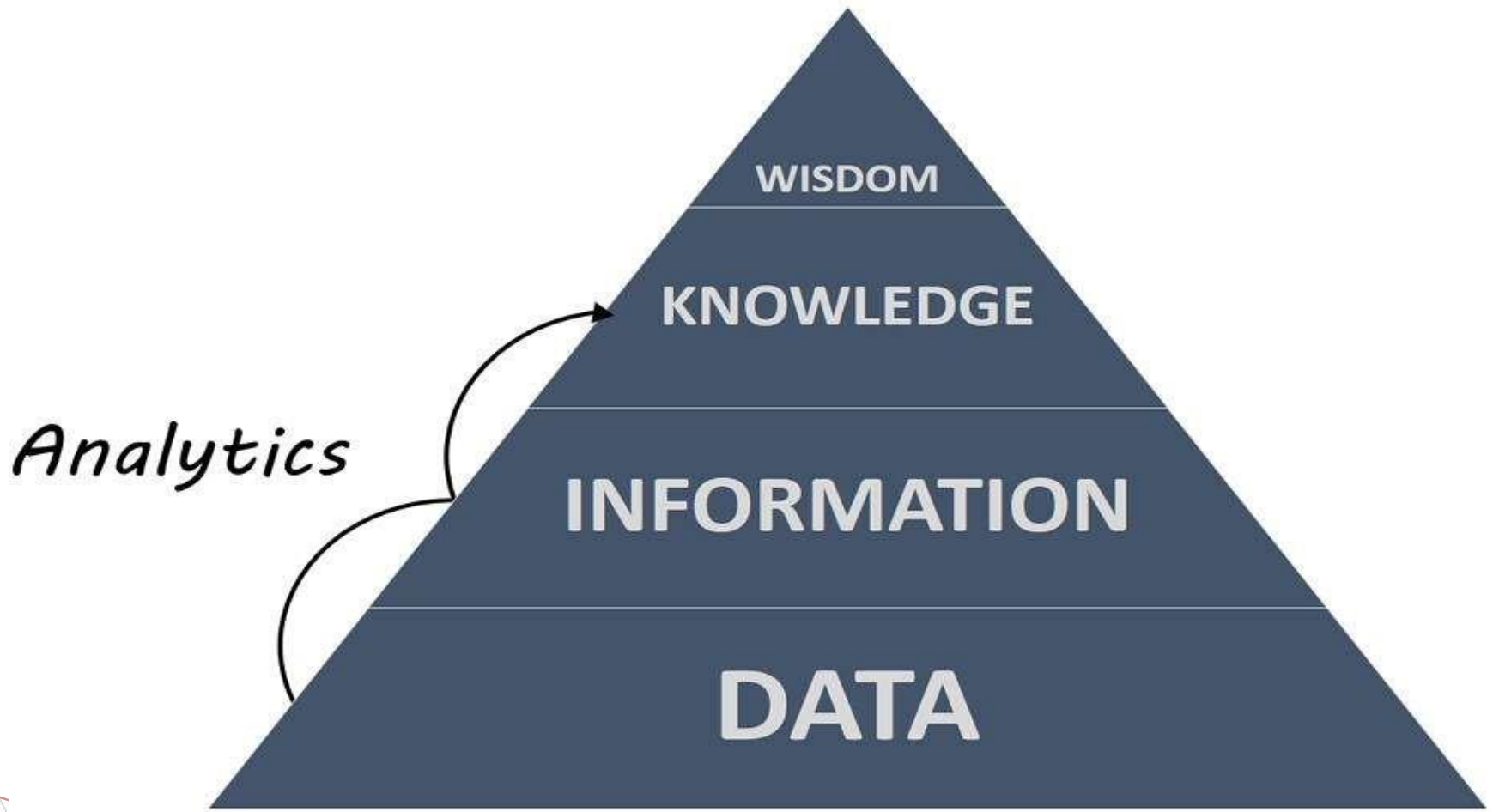
Data Analytics

- Data Analytics is a technology oriented term, involves the use of tools and techniques to find novel useful patterns from data.
- Involves organizing and modeling the data " to test a particular hypothesis" or "answer a particular question".
- Knowledge of machine learning and statistics can be useful in this role.

Data Analysis

- Data Analysis is actually a subset of Data Analytics which helps us to understand the data by questioning and to collect useful insights from the information already available.

Data Analytics Vs Data Analysis

Sr. No.	Data Analytics	Data Analysis
1.	It is described as a traditional form or generic form of analytics.	It is described as a particularized form of analytics.
2.	It includes several stages like the collection of data and then the inspection of business data is done.	To process data, firstly raw data is defined in a meaningful manner, then data cleaning and conversion are done to get meaningful information from raw data.
3.	It supports decision making by analyzing enterprise data.	It analyzes the data by focusing on insights into business data.
4.	It uses various tools to process data such as Tableau, Python, Excel, etc.	It uses different tools to analyze data such as Rapid Miner, Open Refine, Node XL, KNIME, etc.
5.	Descriptive analysis cannot be performed on this.	A Descriptive analysis can be performed on this.
6.	One can find anonymous relations with the help of this.	One cannot find anonymous relations with the help of this.
7.	It does not deal with inferential analysis.	It supports inferential analysis.

Impact of Data analytics in business applications

- 5 Big Benefits of Data and Analytics for Positive Business Outcomes

1. Proactivity & Anticipating Needs: of customer

companies need to capture and reconcile multiple customer identifiers such as cell phone, email and address, to one single customer ID.

Customers are increasingly using multiple channels in their interactions with companies, hence both traditional and digital data sources must be brought together to understand customers' behaviours.

companies deliver contextually relevant, real-time experiences.

2. Mitigating Risk & Fraud:

Security and fraud analytics helps to protect all physical, financial and intellectual assets from misuse by internal and external threats

3. Delivering Relevant Products:

- Effective data collation from 3rd party sources where individuals publicise their thoughts and opinions, combined with analytics will help companies stay competitive when demand changes or new technology is developed as well as facilitate anticipation of what the market demands to provide the product before it is requested.

4. Personalisation & Service:

Being able to react in real time and make the customer feel personally valued is only possible through advanced analytics. Big data offers the opportunity for interactions to be based on the personality of the customer, by understanding their attitudes and considering factors such as real-time location to help deliver personalisation in a multi-channel service environment.

5. Optimizing & Improving the Customer Experience

Applying analytics for designing, controlling the process and optimizing business operations in the production of goods or services ensures efficiency and effectiveness to fulfil customer expectations and achieve operational excellence.

- improve field operations productivity and efficiency as well as optimize an organisational workforce according to business needs and customer demand.

- Optimum utilisation of data and analytics will also ensure that continuous improvements are instigated on an on-going basis as a result of end-to-end view and measurement of key operational metrics

Type of Analytics

- descriptive analytics,
- predictive analytics
- prescriptive analytics,
- **Diagnostic analytics**,

descriptive analytics, which tracks key performance indicators (KPIs) to understand the present state of a business;

predictive analytics, which analyzes trend data to assess the likelihood of future outcomes; and

prescriptive analytics, which uses past performance to generate recommendations for handling similar situations in the future.

Diagnostic analytics, which is like descriptive analytics. It analyzes the state of a business and diagnoses why certain events or outcomes happened.

TERMINOLOGY AND CAREERS

- Many overlapping terms are used in the market place

- Business Intelligence is a business oriented term, and so is Decision Science.

- data analytics or data mining as a component.

- Big Data is special type of data – very large, fast and complex. Storing and managing all this data is a slightly more technical task and it uses more recent and innovative tools

- Cloud computing is an attractive solution to store and process Big Data.

- Data Science is a new discipline born in the early 2000s. The scope includes the entire data processing chain.

- A data scientist would ideally be familiar with all aspects of the discipline while specializing in a part of the area.

- This helps increase the supply of data scientists which has been called the hottest job of the decade.