Data Analytics - Overview

Presented by Shriranga Kulkarni [Shriranga.Kulkarni@gmail.com]

https://www.linkedin.com/in/shrirangakulkarni
What is Data Analytics?

Analytics is the use of:
- data,
- information technology,
- statistical analysis,
- quantitative methods, and
- mathematical or computer-based models

to help managers gain improved insight about their business operations and make better, fact-based decisions.

Business Analytics (BI) is a subset of Data Analytics
What is Business Analytics?

**Business Analytics Applications**
- Management of customer relationships
- Financial and marketing activities
- Supply chain management
- Human resource planning
- Pricing decisions
- Sport team game strategies
What is Business Analytics?

Importance of Business Analytics

- There is a strong relationship of BA with:
  - profitability of businesses
  - revenue of businesses
  - shareholder return
- BA enhances understanding of data
- BA is vital for businesses to remain competitive
- BA enables creation of informative reports
Scope of Business Analytics

- Descriptive analytics
  - uses data to understand past and present
- Predictive analytics
  - analyzes past performance
- Prescriptive analytics
  - uses optimization techniques
Retail Markdown Decisions

Most department stores clear seasonal inventory by reducing prices.
The question is:
When to reduce the price and by how much?

Descriptive analytics: examine historical data for similar products (prices, units sold, advertising, ...)

Predictive analytics: predict sales based on price

Prescriptive analytics: find the best sets of pricing and advertising to maximize sales revenue
Data for Business Analytics

- **DATA**
  - collected facts and figures

- **DATABASE**
  - collection of computer files containing data

- **INFORMATION**
  - comes from analyzing data
Data for Business Analytics

- **Metrics** are used to quantify performance.
- **Measures** are numerical values of metrics.
- **Discrete** metrics involve counting
  - on time or not on time
  - number or proportion of on time deliveries
- **Continuous** metrics are measured on a continuum
  - delivery time
  - package weight
  - purchase price
## Data for Business Analytics

### A Sales Transaction Database File

![Figure 1.1](attached_image)

**Entities**
- Cust ID
- Region
- Payment
- Transaction Code
- Source

**Fields or Attributes**
- Transaction Code
- Source
- Amount
- Product
- Time Of Day

**Records**

<table>
<thead>
<tr>
<th>Cust ID</th>
<th>Region</th>
<th>Payment</th>
<th>Transaction Code</th>
<th>Source</th>
<th>Amount</th>
<th>Product</th>
<th>Time Of Day</th>
</tr>
</thead>
<tbody>
<tr>
<td>10001</td>
<td>East</td>
<td>Paypal</td>
<td>93816645</td>
<td>Web</td>
<td>$20.19</td>
<td>DVD</td>
<td>22:19</td>
</tr>
<tr>
<td>10002</td>
<td>West</td>
<td>Credit</td>
<td>74083490</td>
<td>Web</td>
<td>$17.65</td>
<td>DVD</td>
<td>13:27</td>
</tr>
<tr>
<td>10003</td>
<td>North</td>
<td>Credit</td>
<td>64942368</td>
<td>Web</td>
<td>$23.98</td>
<td>DVD</td>
<td>14:27</td>
</tr>
<tr>
<td>10004</td>
<td>West</td>
<td>Paypal</td>
<td>70560957</td>
<td>Email</td>
<td>$23.51</td>
<td>Book</td>
<td>15:38</td>
</tr>
<tr>
<td>10005</td>
<td>South</td>
<td>Credit</td>
<td>36208817</td>
<td>Web</td>
<td>$16.33</td>
<td>Book</td>
<td>15:21</td>
</tr>
<tr>
<td>10006</td>
<td>West</td>
<td>Paypal</td>
<td>20978903</td>
<td>Email</td>
<td>$17.30</td>
<td>DVD</td>
<td>13:11</td>
</tr>
<tr>
<td>10007</td>
<td>East</td>
<td>Credit</td>
<td>80103311</td>
<td>Web</td>
<td>$177.72</td>
<td>Book</td>
<td>21:59</td>
</tr>
<tr>
<td>10008</td>
<td>West</td>
<td>Credit</td>
<td>14132683</td>
<td>Web</td>
<td>$21.76</td>
<td>Book</td>
<td>4:04</td>
</tr>
<tr>
<td>10009</td>
<td>West</td>
<td>Paypal</td>
<td>40128225</td>
<td>Web</td>
<td>$15.92</td>
<td>DVD</td>
<td>19:35</td>
</tr>
<tr>
<td>10010</td>
<td>South</td>
<td>Paypal</td>
<td>49073721</td>
<td>Web</td>
<td>$23.39</td>
<td>DVD</td>
<td>13:26</td>
</tr>
</tbody>
</table>
What is Big Data?

- Information from multiple internal and external sources:
  - Transactions
  - Social media
  - Enterprise content
  - Sensors
  - Mobile devices

- Companies leverage data to adapt products and services to:
  - Meet customer needs
  - Optimize operations
  - Optimize infrastructure
  - Find new sources of revenue
  - Can reveal more patterns and anomalies
Types of Data
• When collecting or gathering data we collect data from individuals cases on particular variables.

• A variable is a unit of data collection whose value can vary.

• Variables can be defined into types according to the level of mathematical scaling that can be carried out on the data.

• There are four types of data or levels of measurement:

<table>
<thead>
<tr>
<th>1. Categorical (Nominal)</th>
<th>2. Ordinal</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. Interval</td>
<td>4. Ratio</td>
</tr>
</tbody>
</table>
## Classifying Data Elements in a Purchasing Database

### Table 1.1: Purchasing Database Overview

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Order No</th>
<th>Item No.</th>
<th>Item Description</th>
<th>Item Cost</th>
<th>Quantity</th>
<th>Cost per order</th>
<th>A/P Terms</th>
<th>(Month) Order Date</th>
<th>Arrival Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spacetime Technologies</td>
<td>A0111</td>
<td>6489</td>
<td>C-Ring</td>
<td>$3.00</td>
<td>900</td>
<td>$2,700.00</td>
<td>25</td>
<td>10/10/11</td>
<td>10/18/11</td>
</tr>
<tr>
<td>Sleepin Inc.</td>
<td>A0115</td>
<td>5319</td>
<td>Shielded Cable/ft.</td>
<td>$1.10</td>
<td>17,500</td>
<td>$19,250.00</td>
<td>30</td>
<td>08/20/11</td>
<td>09/31/11</td>
</tr>
<tr>
<td>Sleepin Inc.</td>
<td>A0123</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,250</td>
<td>$15,957.50</td>
<td>30</td>
<td>08/25/11</td>
<td>09/01/11</td>
</tr>
<tr>
<td>Sleepin Inc.</td>
<td>A0204</td>
<td>5319</td>
<td>Shielded Cable/ft.</td>
<td>$1.10</td>
<td>16,500</td>
<td>$18,150.00</td>
<td>30</td>
<td>09/15/11</td>
<td>10/05/11</td>
</tr>
<tr>
<td>Sleepin Inc.</td>
<td>A0205</td>
<td>5077</td>
<td>Side Panel</td>
<td>$195.00</td>
<td>120</td>
<td>$23,400.00</td>
<td>30</td>
<td>11/02/11</td>
<td>11/13/11</td>
</tr>
<tr>
<td>Sleepin Inc.</td>
<td>A0207</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,200</td>
<td>$15,750.00</td>
<td>30</td>
<td>09/01/11</td>
<td>09/10/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0223</td>
<td>4224</td>
<td>Bolt-nut package</td>
<td>$3.95</td>
<td>4,500</td>
<td>$17,775.00</td>
<td>30</td>
<td>10/15/11</td>
<td>10/20/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0433</td>
<td>5417</td>
<td>Control Panel</td>
<td>$255.00</td>
<td>500</td>
<td>$127,500.00</td>
<td>30</td>
<td>10/20/11</td>
<td>10/27/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0443</td>
<td>1243</td>
<td>Airframe fasteners</td>
<td>$4.25</td>
<td>10,000</td>
<td>$42,500.00</td>
<td>30</td>
<td>08/08/11</td>
<td>08/14/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0446</td>
<td>5417</td>
<td>Control Panel</td>
<td>$255.00</td>
<td>400</td>
<td>$103,500.00</td>
<td>30</td>
<td>09/01/11</td>
<td>09/10/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A0533</td>
<td>9752</td>
<td>Gasket</td>
<td>$4.05</td>
<td>1,500</td>
<td>$6,075.00</td>
<td>25</td>
<td>09/20/11</td>
<td>09/25/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A0555</td>
<td>6489</td>
<td>C-Ring</td>
<td>$3.00</td>
<td>1,100</td>
<td>$3,300.00</td>
<td>25</td>
<td>10/05/11</td>
<td>10/10/11</td>
</tr>
</tbody>
</table>
Classifying Data Elements in a Purchasing Database

Figure 1.2

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
</tr>
</thead>
<tbody>
<tr>
<td>Purchase Orders</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Supplier</td>
<td>Order No</td>
<td>Item No.</td>
<td>Item Description</td>
<td>Item Cost</td>
<td>Quantity</td>
<td>Cost per order</td>
<td>A/P Terms</td>
<td>(Month)</td>
<td>Order Date</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A0111</td>
<td>8489</td>
<td>C-Ring</td>
<td>$3.00</td>
<td>900</td>
<td>$2,700.00</td>
<td>25</td>
<td>10/10/11</td>
<td>10/18/11</td>
</tr>
<tr>
<td>Steep Inc.</td>
<td>A0115</td>
<td>5319</td>
<td>Shielded Cable</td>
<td>$1.10</td>
<td>17,500</td>
<td>$19,250.00</td>
<td>30</td>
<td>06/20/11</td>
<td>06/31/11</td>
</tr>
<tr>
<td>Steep Inc.</td>
<td>A0123</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,250</td>
<td>$15,537.50</td>
<td>30</td>
<td>08/25/11</td>
<td>09/01/11</td>
</tr>
<tr>
<td>Steep Inc.</td>
<td>A0204</td>
<td>5319</td>
<td>Shielded Cable</td>
<td>$1.10</td>
<td>16,500</td>
<td>$18,150.00</td>
<td>30</td>
<td>09/15/11</td>
<td>10/05/11</td>
</tr>
<tr>
<td>Steep Inc.</td>
<td>A0205</td>
<td>5077</td>
<td>Side Panel</td>
<td>$15.00</td>
<td>120</td>
<td>$3,600.00</td>
<td>30</td>
<td>11/01/11</td>
<td>11/01/11</td>
</tr>
<tr>
<td>Steep Inc.</td>
<td>A0207</td>
<td>4312</td>
<td>Bolt-nut package</td>
<td>$3.75</td>
<td>4,200</td>
<td>$15,750.00</td>
<td>30</td>
<td>09/01/11</td>
<td>09/10/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0223</td>
<td>4224</td>
<td>Bolt-nut package</td>
<td>$3.95</td>
<td>4,500</td>
<td>$17,775.00</td>
<td>30</td>
<td>10/15/11</td>
<td>10/20/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0433</td>
<td>5417</td>
<td>Control Panel</td>
<td>$25.00</td>
<td>500</td>
<td>$12,500.00</td>
<td>30</td>
<td>10/20/11</td>
<td>10/27/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0443</td>
<td>1243</td>
<td>Airframe Fasteners</td>
<td>$4.25</td>
<td>10,000</td>
<td>$42,500.00</td>
<td>30</td>
<td>08/08/11</td>
<td>08/14/11</td>
</tr>
<tr>
<td>Alum Sheeting</td>
<td>A0446</td>
<td>5417</td>
<td>Control Panel</td>
<td>$25.00</td>
<td>400</td>
<td>$10,000.00</td>
<td>30</td>
<td>09/01/11</td>
<td>09/10/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A0553</td>
<td>9752</td>
<td>Gasket</td>
<td>$4.05</td>
<td>1,500</td>
<td>$6,075.00</td>
<td>25</td>
<td>09/20/11</td>
<td>09/25/11</td>
</tr>
<tr>
<td>Spacetime Technologies</td>
<td>A0555</td>
<td>6489</td>
<td>C-Ring</td>
<td>$3.00</td>
<td>1,100</td>
<td>$3,300.00</td>
<td>25</td>
<td>10/05/11</td>
<td>10/10/11</td>
</tr>
</tbody>
</table>
**Categorical (Nominal) data**

- **Nominal or categorical** data is data that comprises of categories that *cannot* be rank ordered – each category is just different.

- The categories available **cannot be placed in any order** and no judgement can be made about the relative size or distance from one category to another.

  - Categories bear no quantitative relationship to one another
  - **Examples:**
    - customer’s location (America, Europe, Asia)
    - employee classification (manager, supervisor, associate)

- What does this mean? **No mathematical operations can be performed on the data relative to each other.**

- Therefore, nominal data reflect **qualitative differences** rather than quantitative ones.
### Nominal data

Examples:

<table>
<thead>
<tr>
<th>What is your gender? <em>(please tick)</em></th>
<th>Did you enjoy the film? <em>(please tick)</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>Yes</td>
</tr>
<tr>
<td>Female</td>
<td>No</td>
</tr>
</tbody>
</table>
Nominal data

• Systems for measuring nominal data must ensure that each category is **mutually exclusive** and the system of measurement needs to be **exhaustive**.

• Variables that have only two responses i.e. Yes or No, are known as **dichotomies**.
Ordinal data

• Ordinal data is data that comprises of categories that can be rank ordered.
• Similarly with nominal data the distance between each category cannot be calculated but the categories can be ranked above or below each other.
   No fixed units of measurement
   Examples:
    - college football rankings
    - survey responses
      (poor, average, good, very good, excellent)
• What does this mean? Can make statistical judgements and perform limited maths.
Example:

How satisfied are you with the level of service you have received? (please tick)

Very satisfied
Somewhat satisfied
Neutral
Somewhat dissatisfied
Very dissatisfied
Interval and ratio data

- Both interval and ratio data are examples of **scale data**.
- Scale data:
  - data is in numeric format ($50, $100, $150)
  - data that can be **measured on a continuous scale**
  - the **distance** between each can be observed and as a result **measured**
  - the data can be **placed in rank order**.
Interval data

- Ordinal data but with constant differences between observations
- Ratios are not meaningful
- Examples:
  - **Time** - moves along a continuous measure or seconds, minutes and so on and is without a zero point of time.
  - **Temperature** - moves along a continuous measure of degrees and is without a true zero.
  - **SAT scores**
Ratio data

- Ratio data measured on a *continuous scale* and *does* have a natural zero point.
  - Ratios are meaningful
  - Examples:
    - monthly sales
    - delivery times
    - Weight
    - Height
    - Age
Types of Analytics
Decision Models

Model:
- An abstraction or representation of a real system, idea, or object
- Captures the most important features
- Can be a written or verbal description, a visual display, a mathematical formula, or a spreadsheet representation
Decision Models

Figure 1.3
A decision model is a model used to understand, analyze, or facilitate decision making.

Types of model input
- data
- uncontrollable variables
- decision variables (controllable)
Decision Models

**Descriptive Decision Models**
- Simply tell “what is” and describe relationships
- Do not tell managers what to do

**An Influence Diagram for Total Cost**
Descriptive Analytics

- Descriptive analytics, such as reporting/OLAP, dashboards, and data visualization, have been widely used for some time.
- They are the core of traditional BI.

What has occurred?

Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and predictive analytics.
Decision Models

A Break-even Decision Model

\[ TC(\text{manufacturing}) = 50,000 + 125*Q \]
\[ TC(\text{outsourcing}) = 175*Q \]

Breakeven Point:
Set \[ TC(\text{manufacturing}) = TC(\text{outsourcing}) \]
Solve for \[ Q = 1000 \text{ units} \]
Decision Models

- **Predictive Decision** Models often incorporate uncertainty to help managers analyze risk.
- Aim to predict what will happen in the future.
- **Uncertainty** is imperfect knowledge of what will happen in the future.
- **Risk** is associated with the consequences of what actually happens.
Predictive Analytics

- Algorithms for predictive analytics, such as regression analysis, machine learning, and neural networks, have also been around for some time.
- Prescriptive analytics are often referred to as advanced analytics.

**What will occur?**

- Marketing is the target for many **predictive analytics** applications.
- Descriptive analytics, such as data visualization, is important in helping users interpret the output from predictive and prescriptive analytics.
Decision Models

A Linear Demand Prediction Model
As price increases, demand falls.

Figure 1.8
Decision Models

A Nonlinear Demand Prediction Model
Assumes price elasticity (constant ratio of % change in demand to % change in price)
**Prescriptive Decision Models** help decision makers identify the best solution.

- **Optimization** - finding values of decision variables that minimize (or maximize) something such as cost (or profit).
- **Objective function** - the equation that minimizes (or maximizes) the quantity of interest.
- **Constraints** - limitations or restrictions.
- **Optimal solution** - values of the decision variables at the minimum (or maximum) point.
Prescriptive Analytics

- Prescriptive analytics are often referred to as advanced analytics.
- Regression analysis, machine learning, and neural networks
- Often for the allocation of scarce resources

What should occur?

- For example, the use of mathematical programming for revenue management is common for organizations that have “perishable” goods (e.g., rental cars, hotel rooms, airline seats).
- Harrah’s has been using revenue management for hotel room pricing for some time.
Organizational Transformation

• Brought about by opportunity or necessity
• The firm adopts a new business model enabled by analytics
• Analytics are a competitive requirement
2013 Academic Research

• A 2011 TDWI report on Big Data Analytics found that 85% of respondents indicated that their firms would be using advanced analytics within three years.

• A 2011 IBM/MIT Sloan Management Review research study found that top performing companies in their industry are much more likely to use analytics rather than intuition across the widest range of possible decisions.
Conditions that Lead to Analytics-based Organizations

- The nature of the industry
- Seizing an opportunity
- Responding to a problem
Complex Systems

- Tackle complex problems and provide individualized solutions
- Products and services are organized around the needs of individual customers
- Dollar value of interactions with each customer is high
- There is considerable interaction with each customer
- Examples: IBM, World Bank, Halliburton
Volume Operations

- Serves high-volume markets through standardized products and services
- Each customer interaction has a low dollar value
- Customer interactions are generally conducted through technology rather than person-to-person
- Are likely to be analytics-based
- Examples: Amazon.com, eBay, Hertz
BI Applications

• Analysis of clickstream data
• Customer profitability analysis
• Customer segmentation analysis
• Product recommendations
• Campaign management
• Pricing
• Forecasting
• Dashboards
The Nature of the Industry

- Online retailers like Amazon.com and Overstock.com are high volume operations who rely on analytics to compete.
- When you enter their sites a cookie is placed on your PC and all clicks are recorded.
- Based on your clicks and any search terms, recommendation engines decide what products to display.
- After you purchase an item, they have additional information that is used in marketing campaigns.
- Customer segmentation analysis is used in deciding what promotions to send you.
- How profitable you are influences how the customer care center treats you.
- A pricing team helps set prices and decides what prices are needed to clear out merchandise.
- Forecasting models are used to decide how many items to order for inventory.
- Dashboards monitor all aspects of organizational performance.