Business Analytics Week 1 Lecture

Introduction to Business Analytics and Data Visualisation

Importance of analytics

- Data, facts and analysis are powerful tools to aid decision making
- Decision making is complicated by the overwhelming amount of data and information available

Scope of business analytics

- Business analytics begins with the collection, organisation, and manipulation of data and is supported by three major components:
  - ➢ Descriptive analytics
  - ➢ Predictive analytics
  - ➢ Prescriptive analytics
- Although the tools used in descriptive, predictive, and prescriptive analytics are different many applications involve all three

Types of Analytics

Descriptive analytics

- Uses data to understand past and present performance and make informed decisions. Most commonly used and well understood type of analytics
- Uses fundamental tools and methods of data analysis focusing on:
  - ➢ Descriptive statistical measures and Data visualisation
  - ➢ Probability distributions / Confidence Interval
  - ➢ Hypothesis test

Predictive analytics

- Analyses past performance in an effort to predict the future by examining historical data, detecting patterns or relationships in these data, and then extrapolating these relationships forward in time.
- Techniques include:
  - ➢ Regression and Forecasting
  - ➢ Data Mining and Machine Learning
  - ➢ Inferential statistics

Prescriptive analytics

- Uses optimization to identify the best alternative to minimize or maximize some objective.
- Techniques include the use of mathematical models:
  - ➢ Decision analysis
  - ➢ Optimization

➢ Simulation

Data for business analytics

- **Data:** numerical or textual facts and figures that are collected through some type of measurement process.
- **Information:** result of analysing data; that is, extracting meaning from data to support evaluation and decision making.

Data sources

- **Primary sources:** internal company records and business transactions, automated data-capturing equipment, customer market surveys
- **Secondary sources:** government and commercial data sources, custom research providers, online research

Big data

- Big data to refer to massive amounts of business data from a wide variety of sources, much of which is available in real time, and much of which is uncertain or unpredictable.
- Volume: amount of data we produce [2.5 Exabyte a day and doubling in every 40 months] • Variety: Big data takes the form of messages, updates and images posted to social networks, readings from sensors, GPS signals from cell phones and more.
- Velocity: means both how fast data is being produced and how fast the data must be processed to meet the demand.

Dataset, entities variable and records

- The data collected in a particular study are referred to as a dataset.
- The people, places or things for which we store and maintain information are called entities.
- A variable (or attribute) is a characteristic of interest for the entities.
- The set of measurements collected for a particular entity is called a record (or observation).

Types of data

- The statistical analysis that is appropriate depends on the type of data.
- For analysis purposes data can be classified as being categorical or numerical.
- In general, there are more alternatives for statistical analysis when the data are numerical

Categorical Data

- Labels or names used to identify an attribute of each entity.
- Often referred to as qualitative data.
- Can be recorded in either numeric or nonnumeric format.
- Appropriate statistical analyses are rather limited.
- Usually counted or expressed as a proportion or a percentage

Numerical Data

- Numerical data indicate how many or how much and often referred to as quantitative data. Ordinary arithmetic operations are meaningful for numerical data.
  Discrete: if measuring how many
- Continuous: if measuring how much Numerical data can be converted to categorical data;
- e.g., Salary can be converted into 'Low', 'Middle' and 'High'. But cannot convert 'High' Salary back into a specific salary figure.

Scale of measurement

Nominal data

- Data are labels or names used to identify an attribute of the entity. A non-numeric label or numeric code may be used.
- Customers can be classified by their geographical location (New South Wales, Victoria, Western Australia, South Australia, Tasmania, ACT and NT).

Ordinal data

- The data have the properties of nominal data and the categories have a meaningful rank. A nonnumeric label or numeric code may be used.
- Rating customer service as "Poor", "Average", "Good", "Very Good" or "Excellent". Ordinal data have no fixed unit of measurement.
- Thus, cannot make meaningful statements about the difference between categories.
- Cannot say that the difference between "Excellent" and "very Good" is the same as between "Good" and "Average"

Interval data

- The data have the properties of ordinal data, and the interval between observations is expressed in terms of a fixed unit of measure. Interval data are always numeric and have no true Zero.
- Both Fahrenheit and Celsius scales represent specific measure of distance – degrees of temperature but have no true zero.
- Thus, cannot make meaningful ratios
- We cannot say that 50 degrees is twice as hot as 25 degrees.

Ratio data

- The data have all the properties of interval data and the ratio of two values is meaningful.
- This scale must contain a zero value that indicates that nothing exists for the variable at the zero point.