Topic 1 Revision Notes

Business Analytics:

Definition:
-"Process of transforming data into actions through analysis and insights in the context of organisational decision making and problem solving"
-It is the use of data, information technology, statistical analysis, quantitative methods, and mathematical or computer-based models to help managers gain an improved insight about their business operations and make better, fact-based decisions
-Supported by various tools such as Microsoft excel, and other software packages

Importance of Analytics:
-Data, facts and analysis aid decision making, and that the decisions made on them are better than those made through gut instinct
-Decision making today is even more complicated, due to overwhelming data and information
-There is a strong relationship of use of analytics and profitability and revenye

Evolution of Business Analytics:
-Modern evolution of analytics began with the introduction of computers, as they provided the ability to store and analyze data easily.

Three major components of business analytics:

1. Descriptive Analysis **(WANT TO KNOW ABOUT PAST)**	-Most commonly used and most well understood type of analytics -Use data to understand past and present performance to make important decisions -Summarizes data into meaningful charts and reports
2. Predictive Analysis **(WANT TO KNOW ABOUT FUTURE)**	-Analyzes past performance in an effort to predict the future by examining historical data, detecting patterns or relationships in these data -Techniques include: regression and forecasting
3. Prescriptive Analysis **(MAKING DECISIONS-OPTIMIZATION)**	-Uses optimization to identify the best alternative to minimize or maximize some objective -Addresses questions such as: •How much should we produce to maximize profit? •What is the best way of shipping goods from our factory to minimize costs?

What is Statistics?

Statistics definition:
-"Statistics relates to the collection, analysis, interpretation, and presentation of data"
-Statistical methods are used to:
•Summarize a collection of data
•Draw inferences about an entire population
•Make predictions or forecasts
-Statistics is also the study of <mark>variation</mark> in data

<u>-Descriptive VS. Inferential statistics:</u>

1. Descriptive statistics:	-Are tabular, graphical, and numerical measures used to summarize data
2. Inferential statistics:	-The process of using data obtained from a sample to make estimates and test claims about the characteristics of a population

Variables:
-Characteristics of items or individuals
-EG. Gender, field of study, money in wallet, time spent in shower each day
-It is essential that all variables have an <u>operational definition</u>: which is defines how a variable is to be measured, otherwise confusion can occur.

Data:
-Observed characteristics of items of individuals.

Populations:
-A collection of all members of a group being investigated
-Two factors need to be specified when defining a population:
•1. The entity (EG. People or motor vehicles)
•2. The boundary

Sample:
-The portion of the population selected for analysis
-EG. Ten full time students selected for a focus group

Parameter:
-A numerical measure of some population characteristic
-EG. The average amount spent by all customers at the local shopping centre last weekend

Statistic:
-A numerical measure that describes a characteristic of a sample
-EG. The average amount spent by the 30 customers completing the market research survey

Data sources:

Four important sources of data:
-Data distributed by an organisation or an individual
-A designed experience
-A survey
-An observational study (such as a focus group)

Primary and Secondary sources:

Primary sources:	-When the data collector is the one using the data for analysis
	-EG. Internal company records, business transactions, customer market surveys
Secondary sources:	-When another organisation or individual has collected the data that is used for analysis by an organisation or individual
	-EG. Government and commercial sources, online research

Types of Data:

BIG DATA (Data deluge):
-Many companies have massive amounts of data at their disposal
-This data deluge is a result of:
•Automatic data collection
•Electronic instrumentation
•Online transactional processing
-There is growing recognition of the untapped value in these data bases
-Data is produced in great volumes, in a variety of forms, and is produced very quickly=BIG DATA

1. Categorical data (Qualitative data):
-Labels or names used to identify attributes of each entity
-Can be recorded in either numeric or nonnumeric formats
-EG. 'Yes or no', 'male or female' answers
-Usually counted or expressed as a portion or a percentage
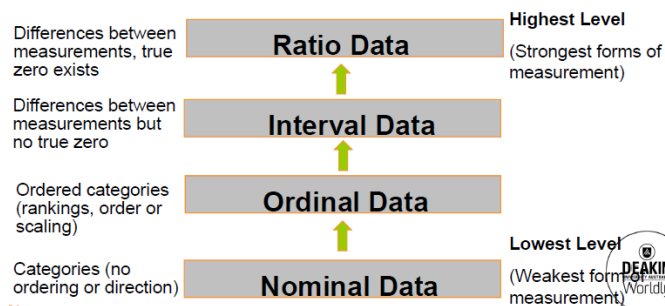
2. Numerical data (Quantitative data):
-Take numbers as their observed responses
-Numerical data can be converted to categorical data. EG Salary can be converted into low/medium/high. However you cannot convert categorical data back to numerical data
-There are two types of numerical data:

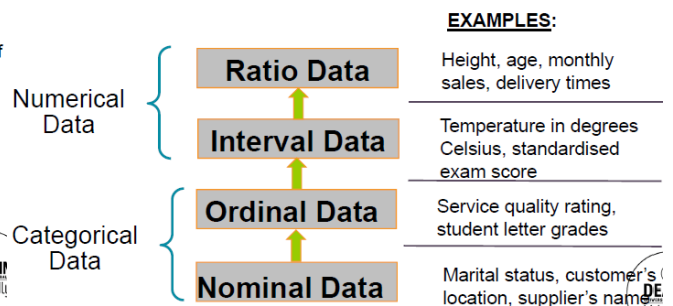| Discrete:	-If measuring how many (Whole numbers)
Continuous:	-If measuring how much (Decimal places)

Scales of Measurement:

Categorical Measurements	
Nominal:	-A classification of categorical data that implies no ranking -EG. Favorite soft drink, gender
Ordinal:	-Scale of measurement where values are assigned by ranking -EG. Rating customers service as 'very good, good, average, or poor'
Numerical Measurements	
Interval:	-A ranking of numerical data where differences are meaningful but there is no true zero point -EG. Shoe sizes 9, 9.5, 10
Ratio:	-A ranking of numerical data where differences between measurements involve a true zero point -EG. Length, weight, age, salary measurements

SCALES OF MEASUREMENT - SUMMARY

Differences between measurements, true zero exists — **Ratio Data** — Highest Level (Strongest forms of measurement)

Differences between measurements but no true zero — **Interval Data**

Ordered categories (rankings, order or scaling) — **Ordinal Data**

Categories (no ordering or direction) — **Nominal Data** — Lowest Level (Weakest form of measurement)

SCALES OF MEASUREMENT - SUMMARY

EXAMPLES:

Numerical Data:
- **Ratio Data** — Height, age, monthly sales, delivery times
- **Interval Data** — Temperature in degrees Celsius, standardised exam score

Categorical Data:
- **Ordinal Data** — Service quality rating, student letter grades
- **Nominal Data** — Marital status, customer's location, supplier's name

Two Broad Types of Data:

Cross-sectional data:	"Relates to a group of items or individuals at a given point of time"
Time ordered (time series) data:	"Relates to a particular entity or situation at different points of time"