ABHYASMITRA.IN

STUDY NOTES

BUSINESS ANALYTICS

Unit II – Analytical Decision-Making

2.1 Analytical Decision-Making Process

Analytical decision-making is the process of using data, evidence, and logical analysis – rather than gut feeling or guesswork – to make business decisions. It involves a structured, step-by-step approach to gather information, analyse it, and choose the best course of action.

> **Definition**
> *"Analytical decision-making is a structured process that uses quantitative data, statistical tools, and logical reasoning to evaluate options and select the course of action most likely to achieve business objectives."*

The Analytical Decision-Making Process – Steps

- **Step 1 – Identify the Business Problem:** Clearly define the problem or decision that needs to be made. Without a well-defined problem, analysis will be unfocused and unhelpful.
- **Step 2 – Decompose into Analytical Questions:** Break the big problem into smaller, specific questions that can be answered with data.
- **Step 3 – Gather Relevant Data:** Identify and collect the data needed to answer the questions. Assess data quality and availability.
- **Step 4 – Analyse the Data:** Apply appropriate analytical techniques (descriptive statistics, visualisation, predictive modelling) to find patterns and answers.
- **Step 5 – Interpret Results and Generate Insights:** Translate analytical findings into actionable business insights. Avoid over-interpreting or misinterpreting the data.
- **Step 6 – Make the Decision:** Select the best course of action based on the insights. Consider business context, constraints, and stakeholder input.
- **Step 7 – Implement and Monitor:** Execute the decision and measure the results against expected outcomes. Use findings to refine the model for future decisions.

2.2 Characteristics of the Analytical Decision-Making Process

Not all decision-making qualifies as "analytical". The following characteristics distinguish a rigorous analytical decision-making process from an intuitive or ad-hoc one.

- **Data-Driven:** Decisions are based on objective data evidence, not personal opinions or gut feelings.
- **Structured and Systematic:** Follows a defined, repeatable process rather than being random or unorganised.
- **Objective and Unbiased:** Uses statistical methods to minimise human cognitive bias in the decision process.
- **Measurable:** Outcomes are quantified and success is defined in measurable terms (KPIs, metrics).

- **Iterative:** The process is continuously improved based on feedback and new data.
- **Transparent:** The methods, data sources, and assumptions used in the analysis are documented and open to review.
- **Action-Oriented:** The end goal is always a specific, actionable decision or recommendation.
- **Contextual:** Data is always interpreted within the business context to ensure relevance.

2.3 Breaking Down a Business Problem into Key Questions

One of the most important skills in business analytics is the ability to translate a vague or complex business problem into specific, answerable questions. This is the bridge between business strategy and data analysis.

Why Decompose Problems into Questions?

- A complex business problem is overwhelming to analyse all at once.
- Specific questions focus the analysis on what truly matters.
- Answerable questions can be directly mapped to data and analytical methods.
- It ensures that the analysis is relevant and actionable.

How to Break Down a Business Problem

Example: Business Problem – "Our customer retention rate has been declining for 6 months." Key questions to answer:

- **Q1:** Which customer segments are churning at the highest rate?
- **Q2:** When in the customer lifecycle is churn most likely to occur?
- **Q3:** What are the most common reasons customers give for leaving?
- **Q4:** Are churning customers also the least engaged with our product features?
- **Q5:** Which channels or campaigns were used to acquire churning customers?
- **Q6:** What is the financial impact of current churn levels?

Each question can then be mapped to specific data sources, analytical methods, and tools, making the overall problem manageable.

MECE Principle (Mutually Exclusive, Collectively Exhaustive)

When decomposing a problem into questions, the MECE principle ensures:

- **Mutually Exclusive:** Each question covers a separate, non-overlapping aspect of the problem.
- **Collectively Exhaustive:** Together, all questions cover the entire problem with no important aspect left out.

2.4 Characteristics of Good Analytical Questions

Not every question is a good analytical question. Good analytical questions have specific properties that make them answerable with data and useful for business decision-making.

Characteristic	Description	Good Example	Poor Example
Specific	Clearly defined scope	Which product category had the lowest margin in Q3?	How is our business doing?
Measurable	Can be answered with numbers or data	What is the monthly customer churn rate?	Are customers happy?
Answerable	Can actually be answered with available data	Which marketing channel has the best ROI?	What will happen in 10 years?
Relevant	Directly linked to a business decision	Which regions should we expand into?	What colour should our logo be?
Time-bound	Has a clear time frame	What was the conversion rate in January?	What is our conversion rate ever?

Additional Characteristics

- **Non-leading:** The question should not assume a specific answer. Avoid "Why is our marketing ineffective?" Instead: "How effective is our marketing?"
- **Falsifiable:** The question should have an answer that could potentially disprove a hypothesis.
- **Actionable:** The answer should lead to a decision or action that can be taken by the business.

2.5 Skills of a Good Business Analyst

A business analyst is the professional who bridges the gap between business strategy and data analysis. They translate business problems into analytical questions, conduct analysis, and communicate insights to decision-makers.

Technical Skills

- **Data Analysis:** Proficiency in data analysis tools such as Microsoft Excel, SQL, Python, or R.
- **Statistical Knowledge:** Understanding of statistics including mean, median, standard deviation, hypothesis testing, and regression.
- **Data Visualisation:** Ability to create clear, informative charts and dashboards using tools like Power BI, Tableau, or matplotlib.
- **Database Management:** Ability to query databases using SQL to extract and manipulate data.
- **Business Intelligence Tools:** Experience with BI tools like Power BI, Tableau, Qlik, or MicroStrategy.

Business Skills

- **Domain Knowledge:** Understanding of the specific industry or business function (finance, marketing, operations, etc.).
- **Business Acumen:** Ability to understand business strategy, goals, and constraints.
- **Problem Decomposition:** Skill in breaking complex business problems into specific analytical questions.
- **Requirements Gathering:** Ability to understand and document stakeholder needs accurately.

Soft Skills

- **Communication:** Ability to explain complex analytical findings in simple, clear language to non-technical stakeholders.
- **Storytelling with Data:** Creating a compelling narrative around data insights to drive action.
- **Critical Thinking:** Questioning assumptions, challenging data, and thinking logically.
- **Attention to Detail:** Spotting errors, inconsistencies, and anomalies in data.
- **Collaboration:** Working effectively with data engineers, data scientists, and business stakeholders.

2.6 Basic Tools of Business Analytics

Business analysts use a variety of tools at different stages of the analytics process. These tools range from simple spreadsheets to advanced visualisation platforms and statistical software.

Microsoft Excel

Excel is the most widely used data analysis tool in business. Despite being relatively simple, it is extremely powerful for small to medium-sized datasets.

- **Data Storage:** Organise and store data in tables.
- **Formulas and Functions:** Perform calculations, lookups, and text manipulations (SUM, AVERAGE, VLOOKUP, IF, COUNTIF).
- **PivotTables:** Quickly summarise and cross-tabulate large datasets.
- **Charts:** Create bar charts, line graphs, pie charts, scatter plots.
- **Data Analysis ToolPak:** Built-in statistical functions for regression, descriptive statistics, and histograms.
- **Power Query:** Import, clean, and transform data from multiple sources.

Excel is best for: Quick analysis, small datasets (up to ~1 million rows), and reports for non-technical audiences.

Tableau

Tableau is one of the world's most popular data visualisation and business intelligence tools. It enables users to create interactive, visually stunning dashboards without writing code.

- **Drag-and-Drop Interface:** Connect to data and create visualisations by simply dragging fields onto the canvas.
- **Wide Data Connectivity:** Connects to Excel, SQL databases, Google Sheets, Salesforce, cloud data warehouses, and more.
- **Interactive Dashboards:** Create dashboards where users can filter, drill down, and explore data interactively.
- **Calculated Fields:** Create custom calculations and metrics directly in Tableau.
- **Tableau Public:** Free platform for sharing visualisations publicly.

Tableau is best for: Data visualisation, executive dashboards, and interactive reporting for business users.

Power BI

Microsoft Power BI is a cloud-based business intelligence and data visualisation tool that is deeply integrated with the Microsoft ecosystem (Excel, Azure, SharePoint, Teams).

- **Power Query (M Language):** Clean, transform, and load data from multiple sources.
- **DAX (Data Analysis Expressions):** A powerful formula language for creating calculated columns, measures, and tables.
- **Interactive Reports and Dashboards:** Highly interactive visualisations with drill-through and cross-filtering.
- **Natural Language Q&A:** Ask questions in plain English and get instant visual answers.
- **Power BI Service:** Share and collaborate on reports in the cloud.

Power BI is best for: Microsoft-ecosystem users, real-time dashboards, and enterprise business intelligence.

2.7 Statistical Analysis and Hypothesis Testing

Statistical analysis is the process of collecting, organising, analysing, and interpreting numerical data to discover patterns, relationships, and insights. It is the mathematical backbone of business analytics.

Descriptive Statistics

Descriptive statistics summarise and describe the main features of a dataset.

- **Mean (Average):** The sum of all values divided by the count. Represents the central tendency.
- **Median:** The middle value when data is sorted. Less affected by extreme outliers than the mean.
- **Mode:** The most frequently occurring value in the dataset.

- **Standard Deviation:** Measures how spread out the values are from the mean. A high SD means data is widely spread.
- **Variance:** The square of the standard deviation. Another measure of spread.
- **Range:** The difference between the maximum and minimum values.
- **Percentiles and Quartiles:** Divide data into portions (25th, 50th, 75th percentile) to understand its distribution.

Hypothesis Testing – Concept

Hypothesis testing is a statistical method used to make decisions based on data. It tests whether there is enough evidence to support or reject a claim about a population based on a sample.

> **Key Terms in Hypothesis Testing**
> *Null Hypothesis (H_0): The default assumption that there is no effect or no difference.*
> *Alternative Hypothesis (H_1): The claim we are trying to prove.*
> *p-value: The probability of observing the results if the null hypothesis were true. If $p < 0.05$ (significance level), we reject H_0.*
> *Type I Error (α): Rejecting H_0 when it is actually true (false positive).*
> *Type II Error (β): Failing to reject H_0 when it is actually false (false negative).*

Example: A marketing team claims that a new email campaign increases click rates. H_0: The campaign has no effect on click rates. H_1: The campaign increases click rates. After running the campaign and collecting data, if the p-value is less than 0.05, we reject H_0 and conclude that the campaign did have a statistically significant effect.

Note: Numerical calculations for hypothesis tests (t-test, chi-square, ANOVA) are NOT expected in the syllabus – only the concept is required.

2.8 Data Visualisation – Concept

> **Definition**
> *"Data visualisation is the graphical representation of information and data using visual elements like charts, graphs, and maps to make complex data easier to understand, analyse, and communicate."*

The human brain processes visual information 60,000 times faster than text. Data visualisation leverages this by turning rows and columns of numbers into intuitive visual patterns that reveal trends, outliers, and relationships instantly.

Why Data Visualisation is Important

- **Simplifies Complexity:** Hundreds of rows of data can be summarised in a single chart.

- **Reveals Patterns:** Trends, seasonality, and outliers that are invisible in raw data become obvious in charts.
- **Facilitates Communication:** Non-technical stakeholders can easily understand visual insights.
- **Speeds Decision-Making:** Decision-makers can quickly grasp the situation without reading lengthy reports.
- **Enables Exploration:** Interactive visualisations allow users to explore data themselves.

Common Chart Types and When to Use Them

Chart Type	Best Used For	Example
Bar Chart	Comparing categories	Sales by product category
Line Chart	Showing trends over time	Monthly revenue over 2 years
Pie Chart	Showing proportions (use sparingly)	Market share by company
Scatter Plot	Showing correlation between two variables	Advertising spend vs Sales
Histogram	Showing data distribution	Distribution of customer ages
Heat Map	Showing intensity or pattern in a matrix	Website clicks by day and hour
Box Plot	Showing spread and outliers in data	Salary distribution by department
Treemap	Showing hierarchical proportions	Revenue breakdown by region and product
Waterfall Chart	Showing sequential addition/subtraction	Monthly profit/loss breakdown

2.9 Popular Data Visualisation Tools

Many tools are available for data visualisation, ranging from simple spreadsheet tools to sophisticated enterprise BI platforms.

Tableau

Tableau is widely considered the gold standard for business intelligence visualisation. It is drag-and-drop, highly interactive, and connects to almost any data source. Tableau Public (free) and Tableau Desktop (paid) are the main versions.

Microsoft Power BI

Power BI is Microsoft's BI and visualisation tool. It integrates seamlessly with the Microsoft Office suite. It has a strong free tier and is widely used in enterprise environments.

Google Looker Studio (formerly Data Studio)

Google's free, cloud-based visualisation tool. Connects easily to Google Analytics, Google Sheets, BigQuery, and other Google services. Ideal for marketing analytics.

QlikView / Qlik Sense

Qlik offers powerful in-memory data processing for fast interactive dashboards. Known for its associative data model that allows highly flexible data exploration.

Python Libraries (matplotlib, seaborn, plotly)

For data scientists and analysts who use Python, these libraries enable code-based creation of custom charts and interactive visualisations. Plotly and Dash are popular for creating interactive web dashboards.

R (ggplot2)

R's ggplot2 library is widely used in academic and research settings for creating publication-quality statistical charts.

Microsoft Excel / Google Sheets

Both are capable of basic charts (bar, line, pie, scatter) and remain the most accessible tools for quick analysis in business settings.

2.10 Exploratory Data Analysis (EDA)

> **Definition**
> *"Exploratory Data Analysis (EDA) is an approach of analysing datasets using statistical summaries and graphical representations in order to understand the main characteristics of the data before applying any formal modelling."*

EDA was introduced by statistician John Tukey in 1977. It is the "first look" at data – before forming any hypothesis or applying any model. EDA helps analysts understand the structure, patterns, anomalies, and relationships in data.

Goals of EDA

- Understand the size and shape of the data (how many rows, columns, what types of variables).
- Find missing values and understand their extent.
- Identify outliers and unusual observations.
- Understand the distribution of each variable (normal, skewed, bimodal).
- Discover correlations and relationships between variables.
- Generate hypotheses that can be tested with formal statistical methods.

EDA Techniques

- **Summary Statistics:** Calculate mean, median, mode, standard deviation, minimum, maximum for all numerical variables.

- **Frequency Tables:** Count occurrences of each category for categorical variables.
- **Histograms:** Visualise the distribution of a numerical variable.
- **Box Plots:** Show the spread of data and identify outliers.
- **Scatter Plots:** Explore the relationship between two numerical variables.
- **Correlation Matrix:** A table showing the correlation coefficient between every pair of numerical variables.
- **Pair Plots:** A grid of scatter plots for all pairs of variables, useful for multi-variable exploration.
- **Bar Charts:** Show the distribution of categorical variables.

2.11 Data Cleaning

Data cleaning (also called data cleansing or data scrubbing) is the process of detecting and correcting errors, inconsistencies, and inaccuracies in a dataset. It is one of the most time-consuming tasks in data analytics – typically taking 60–80% of an analyst's total time.

> **Definition**
> *"Data cleaning is the process of detecting and fixing corrupt, inaccurate, or irrelevant records from a dataset to improve data quality and ensure accurate analysis."*

Common Data Quality Problems

- **Missing Values:** Data fields that are blank or null. Example: Customer age is empty for 15% of records.
- **Duplicate Records:** The same data is recorded more than once. Example: The same customer appears twice with different IDs.
- **Inconsistent Formatting:** The same data is stored in different formats. Example: Dates stored as "01/01/2024", "1 Jan 2024", and "2024-01-01".
- **Incorrect Data Types:** Numbers stored as text, or dates stored as plain numbers.
- **Outliers and Errors:** Extreme or impossible values. Example: A customer age of 250 or a negative sale amount.
- **Structural Errors:** Typos or inconsistent capitalisation. Example: "Mumbai", "mumbai", and "MUMBAI" stored as three different cities.
- **Irrelevant Data:** Variables that are not useful for the current analysis.

Data Cleaning Steps

- **Step 1:** Remove or merge duplicate records.
- **Step 2:** Handle missing values – fill with mean/median/mode, use forward fill, or drop rows/columns if too many values are missing.

- **Step 3:** Standardise formats – ensure dates, currencies, and categorical values are in a consistent format.
- **Step 4:** Fix data types – convert text-stored numbers to numeric format.
- **Step 5:** Handle outliers – investigate extreme values and decide whether to keep, correct, or remove them.
- **Step 6:** Remove irrelevant columns – drop variables that are not useful for analysis.
- **Step 7:** Validate the cleaned data – check summary statistics and distributions again to confirm the data is clean.

2.12 Data Inspection

Data inspection is the first step in any data analysis project. It involves a systematic examination of the dataset to understand its structure, content, and quality before any cleaning or analysis begins.

What to Check During Data Inspection

- **Shape of the Data:** How many rows (records) and columns (variables) does the dataset have?
- **Column Names:** Are the column names meaningful and consistent?
- **Data Types:** What type is each column (integer, decimal, text, date, boolean)? Are the types correct?
- **Missing Values:** Which columns have missing data? How many missing values are there? What percentage of the column is missing?
- **Unique Values:** For categorical columns, how many unique values are there? Are there unexpected values or typos?
- **Summary Statistics:** What is the minimum, maximum, mean, and standard deviation of each numeric column? Does anything look wrong?
- **Sample Records:** Look at the first and last few rows of the dataset to spot obvious issues.
- **Correlations:** Are there highly correlated variables that might cause problems for modelling?

Tools for Data Inspection

- **Microsoft Excel:** Filter, sort, and use formulas (COUNTBLANK, COUNTIF) to inspect data manually.
- **Python (pandas):** Commands like .head(), .info(), .describe(), .isnull().sum() give a quick overview.
- **R:** Functions like str(), summary(), is.na() are used for data inspection.
- **SQL:** Queries using COUNT, DISTINCT, GROUP BY help inspect data in databases.
- **Tableau/Power BI:** Loading data and checking for errors or anomalies visually.