CMA Inter

BDA

Business Data Analytics

Summary Notes

7596941141
WHATSAPP | CALL

GOURAV KABRA

CHAPTER 11
INTRODUCTION TO DATA SCIENCE FOR BUSINESS DECISION-MAKING

Meaning, Nature, Properties, Scope of Data

Data v/s Information v/s Knowledge	1. **Data**: *Raw alphanumeric values* collected through various methods. 2. **Information**: *Processed and organized data* that provides meaningful context. 3. **Knowledge**: *Information used to solve problems*, gained through experience and context.
Nature of Data	1. **Numerical Data**: Data expressed as a *number*. 2. **Descriptive Data**: *Qualitative* data, often in textual form. 3. **Graphic Data**: Data presented in the form of a *picture or graphic*.

Types of Data in Finance and Costing

Quantitative financial data v/s Qualitative financial data	1. **Quantitative Financial Data**: Data expressed in *numbers*, such as *stock price data* and *financial statements*. 2. **Qualitative Financial Data**: Data in qualitative formats like *text, videos, or audio*, such as the *management discussion and analysis* in a company's annual report.
Four Scales of Data [NOIR]	1. **Nominal Scale:** *Classifies observations* based on certain characteristics. Example: classifying equities into *small-cap, mid-cap,* and *large-cap* categories. 2. **Ordinal Scale:** Used for *classifying and ordering* observations (increasing or decreasing). Example: Ranking stocks based on *P/E ratio*. 3. **Interval Scale**: Categorizes and *ranks using equal intervals* between values. 4. **Ratio Scale**: Combines characteristics of *nominal*, *ordinal*, and *interval* scales. It has *equal intervals and a true zero*, allowing for magnitude description. Example: *Temperature scale*.

Digitization of Data and Information

Advantages of Digitization	Improved classification and indexing; Simultaneous access; Data reuse; Work processing; Business system integration; Backup and disaster recovery; Remote access; Increased productivity; Less storage space.

Steps to digitize any record [JAPDiCE]	1. **Justification**: Cost-benefit analysis 2. **Assessment**: Records, hardware/software, risk 3. **Planning**: Approach, documentation, resources, specifications, risk 4. **Digitization**: Capture, quality control, editing, storage, metadata 5. **Care of Records**: Permission, intellectual control, classification, maintenance 6. **Evaluation**: Systematic project merit assessment

Transformation of Data to Information

Process of converting data into information [COPIRU]	1. **Collection**: Standardized systems and trained staff 2. **Organizing**: Grouping and arranging data for users 3. **Processing**: Cleaning data by removing unnecessary elements 4. **Integration**: Creating data networks and master servers 5. **Reporting**: Translating data into accessible formats 6. **Utilization**: Using data to drive decisions and enhance business productivity

Professional Scepticism Regarding Data

Professional Scepticism in Data Analytics	1. Data analytics is valuable for decision-making, but ***managers should critically assess*** the insights. 2. A ***balance*** is needed between *costly scepticism and underusing analytics* to control costs.

Ethical Use of Data and Information

Basic Principles of Data Ethics [PITOO]	1. **Ownership**: Personal data belongs to the individual, and consent must be obtained before collecting or using it. 2. **Transparency**: Users should be informed about the purpose of data collection, such as tracking with cookies, through clear policies. 3. **Privacy**: Personally identifiable information (PII) must be protected with security measures like encryption and authentication. 4. **Intention**: Data should be collected and analysed ethically, not for exploiting weaknesses or causing harm. 5. **Outcomes**: Even with good intentions, data analysis should avoid unintended negative impacts on individuals.

CHAPTER 12
DATA PROCESSING, ORGANISATION, CLEANING & VALIDATION

Data Processing & its Phases

Data Processing	Data processing (DP) is the process of **organising**, **categorising**, and **manipulating data** to extract information.
Phases of DP	1. **Manual DP**: Data processed *without machine assistance*. 2. **Mechanical DP**: Uses *mechanical tools* (e.g., punch cards) for faster data processing than manual methods. 3. **Electronic DP**: Data processing *using computers and electronics*.

Functions of Data Processing [VSAARC]

1. **Data Validation**: Verifying if data values come from an *acceptable set* to ensure data quality.
2. **Data Sorting**: *Organizing data* in a meaningful order for easier analysis and visualization, applicable to both raw and aggregated data.
3. **Data Aggregation**: *Collecting and summarizing* data to enable efficient analysis of large datasets.
4. **Data Analysis**: *Cleaning, converting, and modelling* data to extract actionable insights for decision-making.
5. **Data Reporting**: *Structuring and presenting* raw data into a consumable format for evaluating performance, with a focus on accuracy and detail.
6. **Data Classification**: *Categorizing data for better usability* and security, making it searchable and trackable.

Risk Categories of Data & Systems

1. **Low Risk**: *Public data* with easy recovery.
2. **Moderate Risk**: Internal data, *not highly sensitive*.
3. **High Risk**: *Sensitive*, critical, or hard-to-retrieve data.

Data Organisation & Distribution

Data Organisation	Arranging unstructured data meaningfully using techniques like *classification*, *frequency distribution tables*, *image representations,* and *graphical representations.*
Structured Data v/s Unstructured Data	1. **Structured Data**: *Organized in tables*, easily integrated into databases and analytics software. 2. **Unstructured Data**: Raw, *unformatted information*, like plain text, with scattered content.

Data Distribution & Types

Data Distribution	A function that identifies all possible values for a variable and quantifies their relative frequency (probability of occurrence).
Types of Data Distribution	**Discrete Distributions**: 1. Discrete distributions result from countable data with a ***finite*** *number of values*. 2. **Examples:** Binomial Distribution, Poisson Distribution, Hypergeometric Distribution, Geometric Distribution. **Continuous Distributions**: 1. Continuous distributions involve an ***infinite*** *number of possible data points*, represented on a continuous scale. 2. **Examples:** Normal Distribution, Lognormal Distribution, F Distribution, Chi-Square Distribution, Exponential Distribution, T-Student Distribution.

Data Cleaning: Steps & Benefits

Data Cleaning	Data cleaning is the process of correcting or deleting inaccurate, corrupted, improperly formatted, duplicate, or incomplete data from a dataset. **Steps for Data Cleaning:** Remove duplicates > fix structural errors > filter outliers > handle missing data > validate data. **Benefits of Data Cleaning:** 1. Corrects errors 2. Reduces mistakes

Types of Data Validation

Types of Data Validation	1. **Data type check**: Age field allows only numeric values (e.g., 25, 30). 2. **Code check**: Postal code verification (e.g., 90210). 3. **Range check**: Temperature field must be between -50°C and 50°C. 4. **Format check**: Date field in "DD-MM-YYYY" format (e.g., 12-08-2024). 5. **Consistency check**: Delivery date cannot be earlier than the order date. 6. **Uniqueness check**: Email address should be unique in the database.

CHAPTER 13
DATA PRESENTATION: VISUALISATION & GRAPHICAL PRESENTATION

Data Visualization

Data Visualisation	Converting data into visual formats like maps or graphs.
The Right Way to Visualize Data! *[objective, audience, technology, & training]*	1. **Know the objective**: Decide if the data is conceptual or data-driven and if the goal is to explore or explain. 2. **Always keep the audience in mind:** Adjust detail based on who will view the visualizations. 3. **Invest in the best technology**: Set up a central data system (ERP) and choose tools that make it easy to create and explore visuals. 4. **Improve team's ability to visualise data**: Ensure staff know how to use data visualization tools effectively.
Objectives of Data Visualization *(Advantages)*	1. Make better decision analysis. 2. Faster decision analysis. 3. Analysing complicated data.

Data Presentation Architecture

A set of skills focused on identifying, finding, modifying, formatting, and presenting data to effectively convey meaning and provide insight.

Data Visualization Types

Dashboard	An interactive tool to manage metrics, visualize data, and generate reports for customers summarizing results.
Histogram	A chart showing the distribution of numerical data, with bars representing the frequency of each data interval.
Bar Chart	A chart that displays categorical data with rectangular bars proportional to the values they represent. 1. **Grouped Bar Chart**: Shows numerical values for two categorical variables, grouped together. 2. **Stacked Bar Chart**: Displays data comparisons with segments representing parts of a whole. 3. **Pareto Chart**: A bar chart ordered by frequency in descending order, with a line showing cumulative relative frequency.
Frequency Polygon	A graph that uses lines to connect midpoints of intervals in a histogram.
Line Chart	A chart showing data changes over time, connecting data points with a continuous line (e.g., asset price changes).
Bubble Line Chart	A variation of the line chart where varying-sized bubbles represent an additional dimension of data.

Scatter Plot	Visualizes joint variation between two numerical variables.
Pie Chart	A circular chart showing numerical proportions with arc lengths proportional to each value.
Maps & Density Maps	1. A map displays geographic data, helping visualize patterns in each location or region. 2. A density map shows patterns and concentrations of data points in a geographic area, useful for large datasets (e.g., cyclone-prone areas in India).
Gantt Chart	Displays project timelines, task dependencies, and resource allocation over time, applicable to any time-related data.
Tables	Used for displaying individual values in rows and columns.

Steps to Use Visualization in Report Design

The steps are aimed at creating a report that not only informs but also tells a **story** with data, ensuring it is engaging, clear, and impactful.

1. **Find a story**: Organize and analyse data to identify trends and outliers.
2. **Create a narrative**: Develop an engaging and logical story with context, highlighting key insights.
3. **Choose the right visualization**: Present data in the most effective chart or graph.
4. **Follow visual language**: Maintain consistency with the company's style and data visualization principles.
5. **Publicize the report**: Share the report with relevant external audiences.

Example: Annual Reports of Listed Companies

1. **Finding a story**: Displaying *stock performance* or *sales trends*.
2. **Creating a narrative**: Explaining how *new product launches* contributed to success.
3. **Choosing visuals**: Showing *graphs* for financial data.
4. **Following visual language**: Consistent *branding* and *design*.
5. **Publicizing the report**: Sharing on *websites* or via *press releases*.

Tools and Techniques of Visualisation and Graphical Presentation

1. **Tableau**: For creating interactive visualizations like *graphs & maps* from various data sources.
2. **Microsoft Power BI**: A *business intelligence tool* for centralized data access & reports.
3. **Microsoft Excel**: A versatile tool for visualizing data with charts like *bar, pie, & scatter plots*.
4. **QlikView**: A data discovery platform with a *drag-and-drop interface* for quick insights and easy integration from various data sources.

CHAPTER 14
DATA ANALYSIS AND MODELLING

Data Analytics: Steps & Benefits

Data Analytics	Data analytics involves *analysing raw data to identify patterns and extract useful insights*, often using machine learning, simulations, and automated systems.
Steps in Data Analytics	**Step 1:** Criteria for grouping data **Step 2:** Collecting the data **Step 3:** Organizing the data **Step 4:** Cleaning the data **Step 5:** Adopt the right type of data analytics process
Benefits	Better decisions, boosts efficiency & enhanced stakeholder service.

Data Mining and Implementation

Data Mining	Data mining is the process of analysing large data sets to find patterns and relationships, helping businesses predict trends and make better decisions.
Process of Data Mining	1. **Setting the business objective**: Identify the business challenge to guide data queries and parameters for the project. 2. **Preparation of data**: Collect and cleanse relevant data, removing noise like duplicates and outliers. 3. **Model building and pattern mining**: Use algorithms to classify or cluster data based on patterns (supervised or unsupervised learning). 4. **Result evaluation and implementation**: Analyse the findings and implement strategies to achieve business goals.
Techniques of Data Mining	1. **Association rules**: Discover relationships between variables, often used in market basket analysis (e.g., predicting product recommendations). 2. **Neural Networks**: Deep learning method that mimics the brain's structure to process data using layers of interconnected nodes. 3. **Decision tree**: A tree-like model used to classify or predict outcomes based on decisions. 4. **K-nearest neighbour (KNN)**: Classifies data points based on their proximity to similar data points.
Application of Data Mining in Finance	1. **Detecting financial crimes**: Data mining helps identify suspicious banking activities. 2. **Loan prediction**: Analyses data to predict loan repayment and creditworthiness. 3. **Target marketing**: Identifies optimal products for different customer segments. 4. **Data warehouse design**: Organizes large datasets for efficient analysis.

Descriptive Analytics

Descriptive Analytics	1. Descriptive analytics uses historical and current data to identify trends and relationships. 2. It relies on basic statistical techniques like averages and percentages, making it easier to understand for a wide audience.
Steps in Descriptive Analytics	**Step 1:** Decide the business metrics (sales growth). **Step 2:** Identification of data requirement (monthly sales data). **Step 3:** Preparation and collection of data (collect sales data). **Step 4:** Analysis of data (calculate sales growth %). **Step 5:** Presentation of data (use line chart).
Applications	Descriptive analytics is used to summarize past events, track social media engagement, report general trends, and compile survey data.

Diagnostic Analytics

Diagnostic Analytics	Diagnostic Analytics helps identify the causes behind trends and relationships between variables.
Advantages	1. Identifies root causes of issues. 2. Improves decision-making by understanding data patterns. 3. Helps optimize business operations by uncovering inefficiencies. 4. Enhances forecasting accuracy through trend analysis.

Predictive Analytics

Predictive Analytics	Predictive analytics focuses on forecasting future events by analysing past data patterns and trends, helping to predict what might happen in the future. The core of predictive analytics is probability, using techniques like data mining, statistical modelling, and machine learning to predict outcomes and their probabilities.
Application	1. **Inventory forecasting**: Predicts product demand. 2. **Customer service**: Personalizes offerings. 3. **Fraud detection**: Identifies trends and anomalies. 4. **Risk mitigation**: Improves applicant screening.
Predictive Analytics vs. Machine Learning	**Predictive Analytics**: Uses historical data and statistical methods to forecast future events. **Machine Learning**: A subset of AI that allows systems to learn from large data sets and improve over time without being explicitly programmed.

Prescriptive Analytics

Prescriptive Analytics	Prescriptive analytics recommends the best actions by predicting future outcomes and assessing potential consequences, requiring advanced expertise, and not often used in daily operations.

Applications	1. **GPS**: Recommends optimal routes based on travel time and road conditions. 2. **Oil & Manufacturing**: Monitors price fluctuations and improves equipment management. 3. **Healthcare**: Enhances patient care and analyses treatment cost-effectiveness. 4. **Insurance**: Evaluates customer risk for pricing and premiums.

Standards for Data Tagging and Reporting (XML, XBRL)

Extensible Markup Language (XML)	XML (Extensible Markup Language) is a format for storing and transferring data, enabling different systems to share information using a standardized language.
Extensible Business Reporting Language (XBRL)	XBRL (eXtensible Business Reporting Language) is a data format based on XML for exchanging standardized financial data, enabling automated, reliable extraction and report generation. **Benefits of XBRL:** 1. Ensures consistency and reduces errors through automation. 2. Cuts costs by simplifying report preparation. 3. Speeds up decision-making for financial entities. 4. Eases publishing of analyst and investor reports. 5. Enhances access, comparison, and analysis of financial data.

Cloud Computing

Cloud Computing	Cloud computing is the delivery of services like storage, processing, and data access over the internet, using remote servers instead of local storage or private data centres.
Types of Cloud Computing	1. **Private Cloud**: Exclusive cloud for one organization (e.g., internal data centres). 2. **Public Cloud**: Shared cloud with pay-as-you-go services (e.g., AWS). 3. **Hybrid Cloud**: Mix of private and public clouds for flexibility (e.g., sensitive data in private, processing in public).

Business Intelligence

Business Intelligence	Business Intelligence (BI) involves systems and tools that collect, store, and analyse company data to help make informed, data-driven decisions. It includes analytics, data mining, visualization, and best practices.
Procedures & Activities in BI	1. **Data Mining:** Extracting patterns from large datasets using analytics and machine learning. 2. **Reporting:** Sharing data analysis results for informed decision-making. 3. **Performance Metrics and Benchmarking:** Comparing current performance with past data to track progress.

	4. **Descriptive Analytics:** Analysing data to understand past events. 5. **Querying:** Extracting answers from data through specific queries. 6. **Statistical Analysis:** Using statistics to explore data patterns and trends. 7. **Data Visualization:** Converting data into charts and graphs for easier interpretation. 8. **Visual Analysis:** Real-time data exploration using visual tools to share insights. 9. **Data Preparation:** Organizing data from multiple sources for analysis.

Artificial Intelligence (AI)

Artificial Intelligence (AI)	AI simulates human intelligence in machines, using data and algorithms for problem-solving, including machine learning and deep learning for predictions.
Application of AI in Financial Services	1. **Investment Services:** Algorithmic trading, robo advisory, insurance pricing. 2. **Lending:** Lending operations, credit scores, default detection. 3. **Audit & Compliance:** Fraud detection, compliance, expense management. 4. **Customer Service:** Sales opportunities, KYC, churn prediction.
Weak AI & Strong AI	1. **Weak AI** (Narrow AI) is designed for specific tasks, like Siri and self-driving cars. 2. **Strong AI** includes Artificial General Intelligence (**AGI**), which mimics human intelligence, and Artificial Superintelligence (**ASI**), which surpasses it, both still theoretical.
Deep Learning vs. Machine Learning	1. **Deep Learning** automates feature extraction, uses large datasets, and requires minimal human input, handling both structured and unstructured data. 2. **Machine Learning** needs more human involvement for feature design and typically works with structured data.
Robotic Process Automation (RPA)	RPA uses software bots to automate rule-based business processes by mimicking human actions, operating 24/7 with high accuracy, and without changing existing systems.
Data Driven vs. Model Driven Approach to AI	In AI, the **data-driven** approach improves data quality and governance, while the **model-driven** approach focuses on developing new models and algorithms. Model-driven techniques have advanced more, leaving room for improvement in data governance.

Machine Learning & its Approaches

Machine Learning	Machine Learning is an AI application that allows systems to learn and improve from experience without explicit programming, enabling programs to access data and learn independently.

Approaches towards Machine Learning	1. **Supervised Learning**: Algorithms build a model using labelled data (inputs and expected outcomes) to make predictions, e.g., classification and regression. 2. **Unsupervised Learning**: Uses unlabelled data to find patterns like clustering or grouping based on similarities, e.g., cluster analysis. 3. **Semi-Supervised Learning**: Combines small labelled data with large unlabelled data to improve learning accuracy. 4. **Reinforcement Learning**: Software agents learn to maximize rewards by interacting with their environment, e.g., autonomous cars or game playing. 5. **Dimensionality Reduction**: Reduces the number of features in a dataset while retaining important variables, e.g., Principal Component Analysis (PCA).