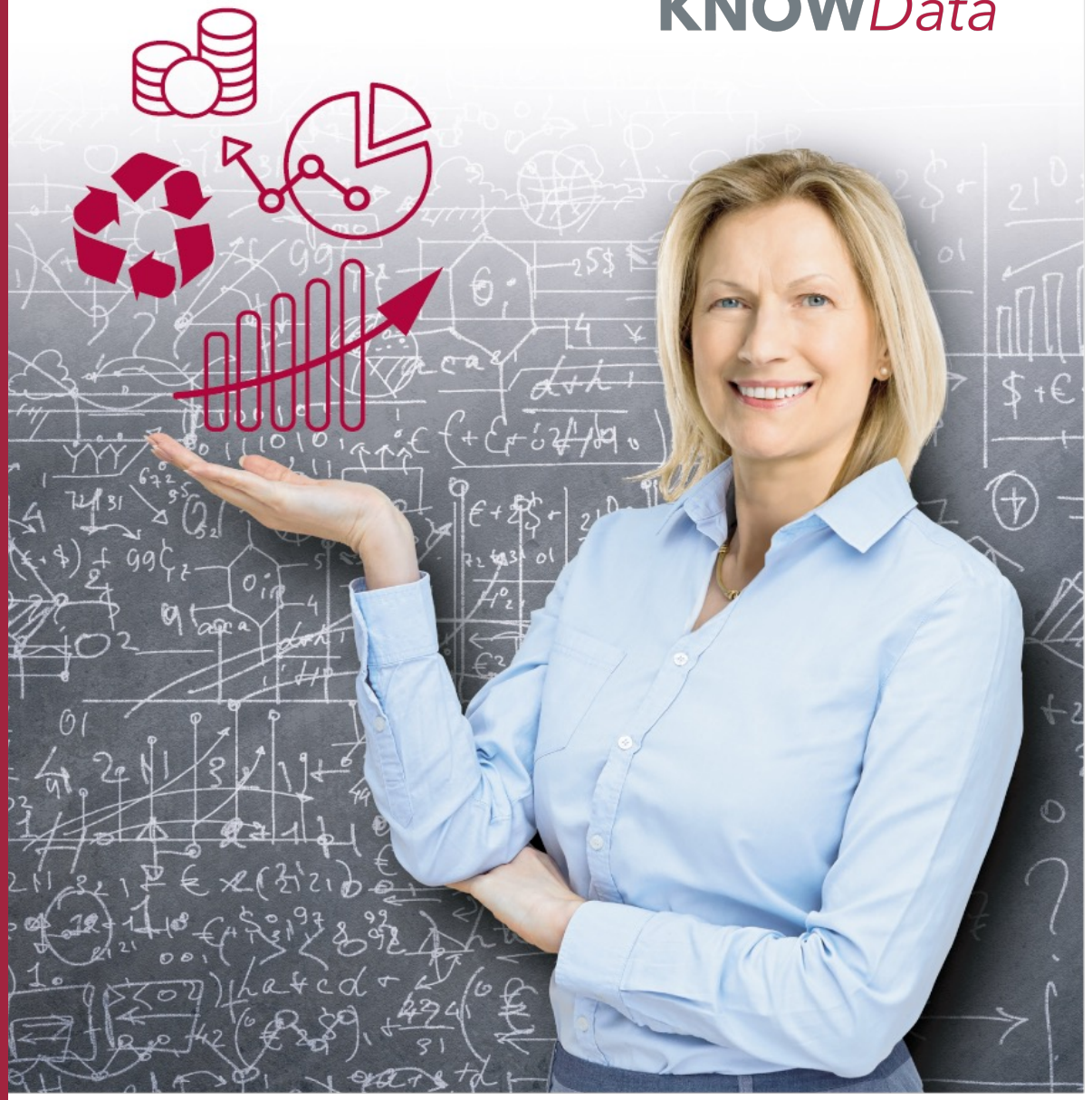
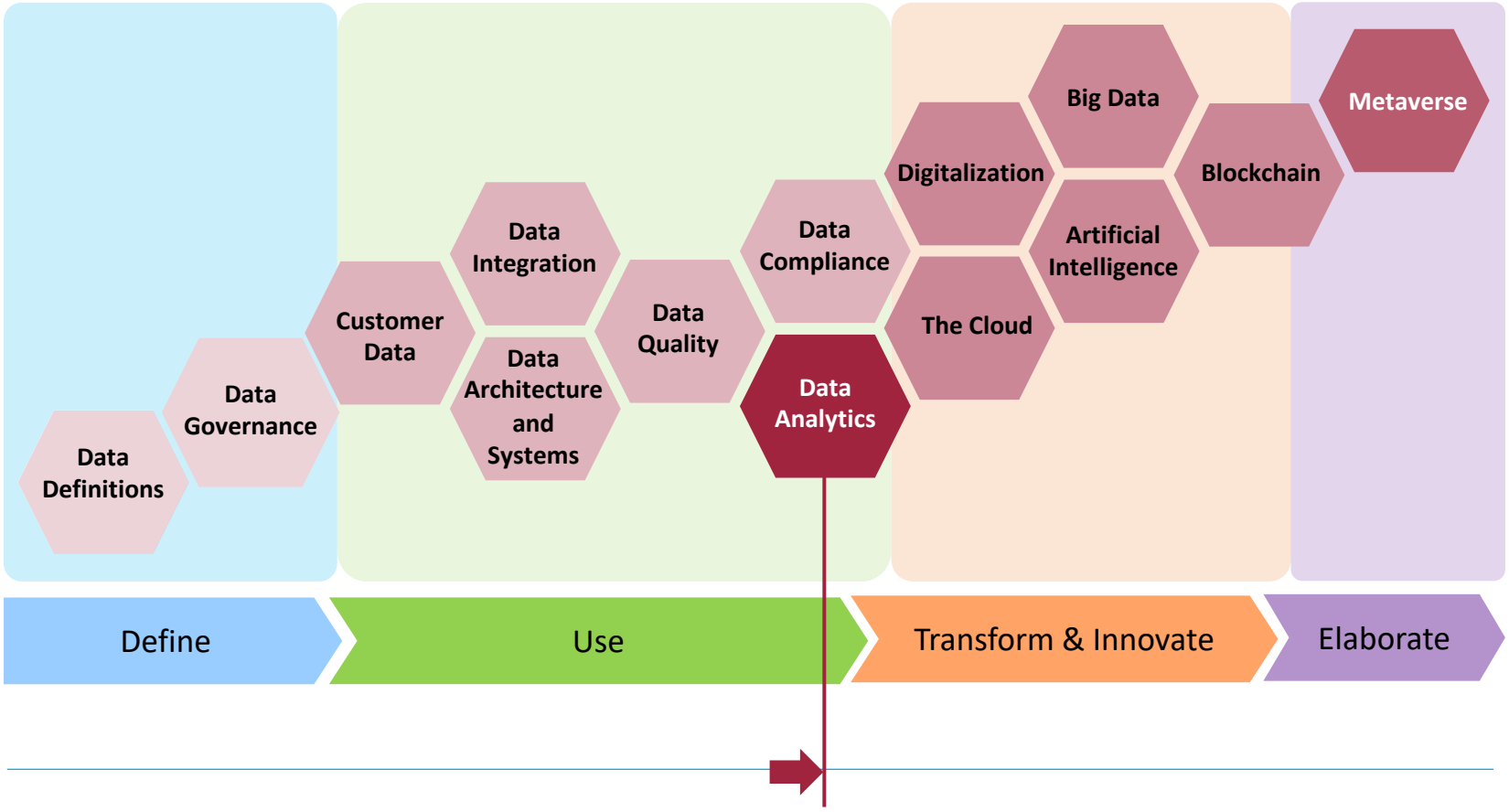
Data Analytics

KNOWData

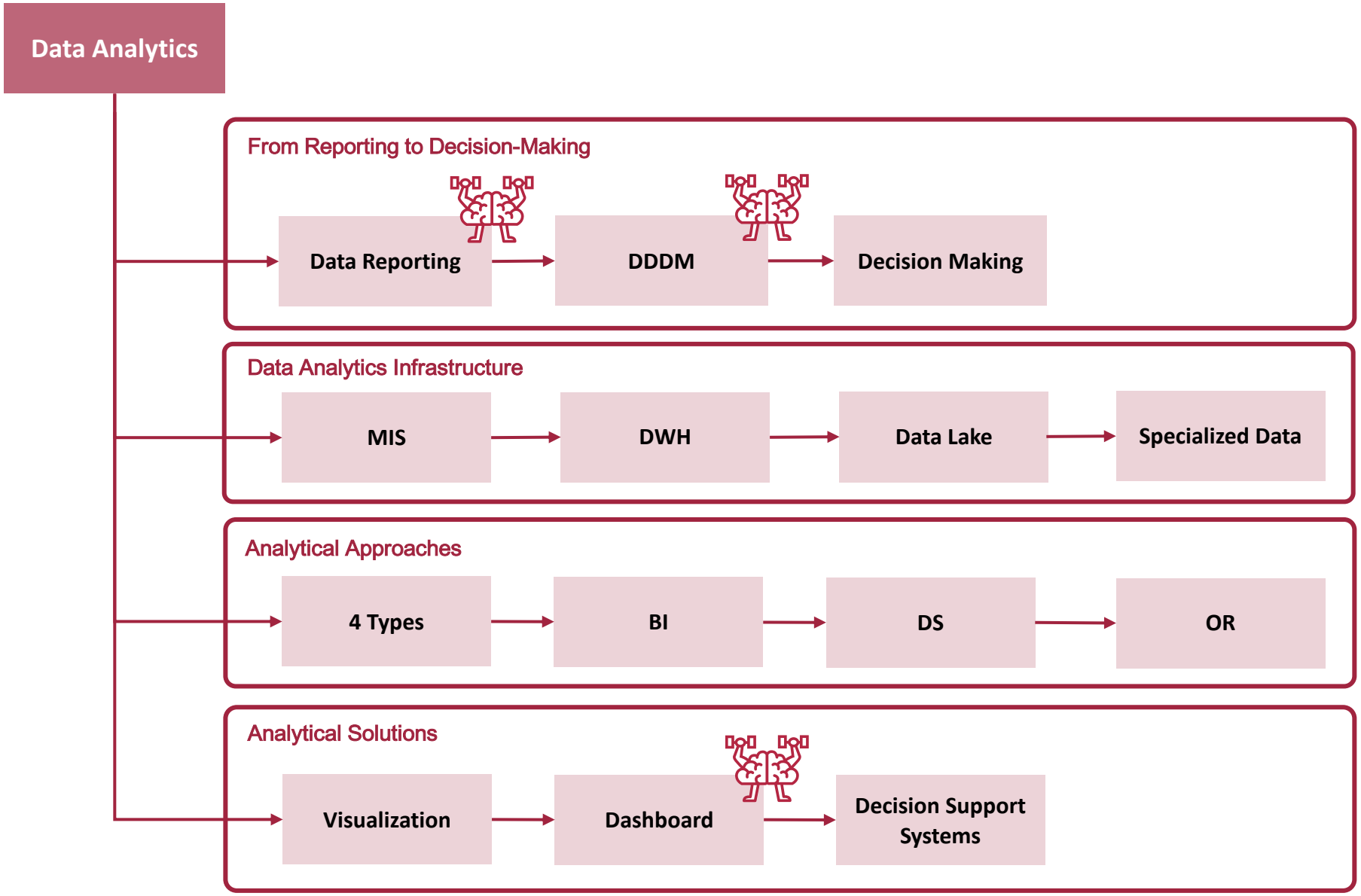


July 4, 2026

Lionel Pilorget


Data Definitions
Data Governance
Customer Data
Data Integration
Data Architecture and Systems
Data Quality
Data Compliance
Data Analytics
Digitalization
The Cloud
Big Data
Artificial Intelligence
Blockchain
Metaverse
Define
Use
Transform & Innovate
Elaborate

Structure of lesson 7



FINANCIAL SUMMARY
(Unaudited)

($ in millions, except percentages and per share data)	Q4-2024	Q1-2025	Q2-2025	Q3-2025	Q4-2025	YoY
Total automotive revenues	19,798	13,967	16,661	21,205	17,693	-11%
Energy generation and storage revenue	3,061	2,730	2,789	3,415	3,837	25%
Services and other revenue	2,848	2,638	3,046	3,475	3,371	18%
Total revenues	25,707	19,335	22,496	28,095	24,901	-3%
Total gross profit	4,179	3,153	3,878	5,054	5,009	20%
Total GAAP gross margin	16.3%	16.3%	17.2%	18.0%	20.1%	386 bp
Operating expenses	2,596	2,754	2,955	3,430	3,600	39%
Income from operations	1,583	399	923	1,624	1,409	-11%
Operating margin	6.2%	2.1%	4.1%	5.8%	5.7%	-50 bp
Adjusted EBITDA [1] [2]	4,333	2,814	3,401	4,227	4,154	-4%
Adjusted EBITDA margin [1] [2]	16.9%	14.6%	15.1%	15.0%	16.7%	-17 bp
Net income attributable to common stockholders (GAAP) [1]	2,128	409	1,172	1,373	840	-61%
Net income attributable to common stockholders (non-GAAP) [1] [3]	2,107	934	1,393	1,770	1,761	-16%
EPS attributable to common stockholders, diluted (GAAP) [1]	0.60	0.12	0.33	0.39	0.24	-60%
EPS attributable to common stockholders, diluted (non-GAAP) [1] [3]	0.60	0.27	0.40	0.50	0.50	-17%
Net cash provided by operating activities	4,814	2,156	2,540	6,238	3,813	-21%
Capital expenditures [4]	(2,780)	(1,492)	(2,394)	(2,248)	(2,393)	-14%
Free cash flow [4]	2,034	664	146	3,990	1,420	-30%
Cash, cash equivalents and investments	36,563	36,996	36,782	41,647	44,059	21%

[1] As a result of the adoption of the new crypto assets standard, the previously reported quarterly periods in 2024 have been recast.
[2] Beginning in Q1'25, Adjusted EBITDA (non-GAAP) is presented net of digital assets gains and losses and all prior periods have been adjusted.
[3] Beginning in Q1'25, Net income attributable to common stockholders (non-GAAP) is presented net of digital assets gains and losses and all prior periods have been adjusted.
[4] Beginning in Q1'25, Capital expenditures is presented inclusive of purchases of energy generation and storage systems and all prior periods have been adjusted.

Mandatory Data Reporting in Organizations

Financial Reports

- Annual Report: Includes financial statements, management discussion & analysis and an overview of the company's performance
- Audited Financial Statements: Prepared by independent auditors (required for public companies and some private firms)
- Core financial documents: Balance Sheet, Income Statement, and Cash Flow Statement

Tax and Legal Disclosures

- Tax Reports (e.g., IRS Form 1120 in the U.S.)
- Beneficial Ownership Reports (e.g., BOI in the U.S.): Discloses company ownership under anti-money laundering laws
- Country-by-Country Reporting (CbCR): For multinationals to disclose profits and taxes per jurisdiction

Compliance Reports

- Corporate Governance Report: Details board structure, policies, and compliance with governance codes
- Sustainability Report: Covers environmental, social, and governance performance (increasingly mandatory)
- Anti-Corruption & Compliance Reports: Required in regulated industries (e.g., banking, healthcare)

Industry-Specific Reports

- Banking & Finance: Basel III reports, Pillar 3 disclosures (risk exposure)
- Healthcare: HIPAA compliance reports (U.S.), clinical trial disclosures
- Energy & Utilities: Emissions reports, safety audits
- Public Companies: Insider trading disclosures (Form 4 in the U.S.)

Operational and Management Reports in a Company

Sales & Marketing Reports

- Sales Performance Reports: Tracks revenue by rep, region, product
- Customer Acquisition Cost (CAC) & Lifetime Value (LTV): Measures ROI on marketing
- Conversion Funnel Reports: Analyzes drop-offs in sales/marketing pipelines
- Campaign Performance Reports: Evaluates ROI of ads, email, social media

Operational Reports

- Production Efficiency Reports: Measures output vs. capacity (manufacturing)
- Inventory Management Reports: Tracks stock levels, turnover, and shortages
- Supply Chain & Logistics Reports: Monitors supplier performance, delays, costs
- Quality Control Reports: Tracks defects, compliance, and recalls

Risk & Compliance Reports

- Internal Audit Reports: Identifies control weaknesses and fraud risks
- Cybersecurity Threat Reports: Logs breaches, vulnerabilities, and IT risks
- Health & Safety Incidents: Tracks workplace accidents (OSHA compliance)

Human Resources Reports

- Employee Productivity Reports: Tracks performance metrics (e.g., sales per rep)
- Turnover & Retention Reports: Analyzes attrition rates and reasons
- Training & Development Reports: Measures skill gaps and training effectiveness
- Payroll Analysis: Tracks labor costs, overtime, bonuses

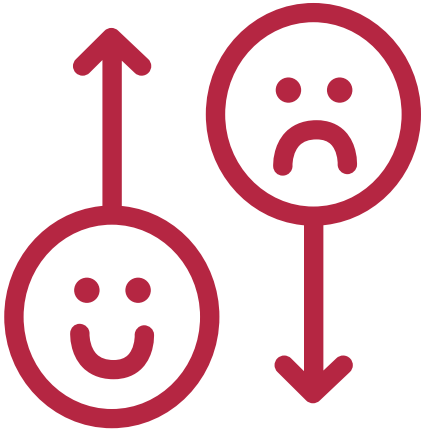
How Companies Measure Performance

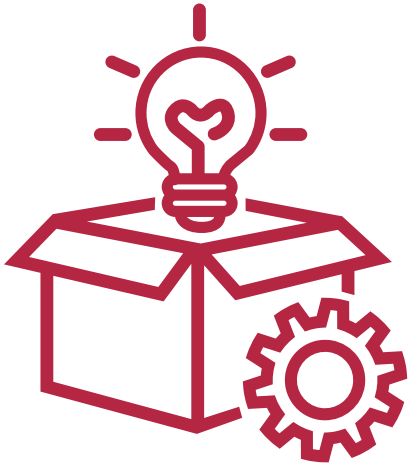
Some important ratios

Ratio	Formula	Description	Relevance
Current Ratio	Current Assets / Current Liabilities	Measures short-term liquidity	Indicates if a company can pay its short-term obligations *A ratio < 1 suggests liquidity risk*
Quick Ratio (Acid-Test)	(Current Assets - Inventory) / Current Liabilities	More stringent liquidity test (excludes inventory)	Assesses immediate solvency without selling inventory *(Ideal: ~1)*
Debt-to-Equity (D/E)	Total Debt / Total Equity	Evaluates financial leverage	High ratio = high risk (depends on industry) *Creditors prefer lower D/E*
Gross Profit Margin	(Revenue - Cost of Goods Sold) / Revenue	Shows % of revenue left after COGS	Measures core profitability before overhead *Higher = better pricing/cost control*
Net Profit Margin	Net Income / Revenue	% of profit after all expenses	Key for overall profitability *Low margins may indicate inefficiencies*
Return on Equity (ROE)	Net Income / Shareholders' Equity	Measures profitability relative to equity	Indicates how well management uses investors' capital *High ROE => efficient*
Return on Assets (ROA)	Net Income / Total Assets	Evaluates efficiency in using assets	Shows how well assets generate profit Useful for capital-intensive industries
Inventory Turnover	Cost of Goods Sold / Average Inventory	How quickly inventory sells	High turnover = efficient sales Low turnover = overstocking/obsolescence risk
Accounts Receivable Turnover	Net Credit Sales / Avg. Accounts Receivable	How fast customers pay	High ratio = efficient collections Low ratio = credit policy issues
Interest Coverage Ratio	EBIT / Interest Expense	Ability to pay interest on debt	Below 1.5 = risky (may struggle with debt payments)

Why is it important to report and analyze data?

Discuss with your neighbor. Write down 2 reasons.





Why Data Analysis Matters in Organizations

Area	Example
Retail	Collect customer information from cash registers, mailing lists, websites, and feedback cards for segmentation and targeting purposes
Healthcare	Record patient information from different units, including personal information, financial transactions with the hospital, and insurance data to Improve patient outcome
Construction	Purchase made during the construction timeline to identify cost-saving opportunities and to optimize schedules
Finance	Identify suspicious patterns or anomalies in financial transactions, enabling early detection and prevention of fraud
Public Administration	Understand the needs and characteristics of different population groups and develop targeted policies or programs based on demographic data
IT	Analyze large volumes of data to identify abnormal patterns or behaviors that could indicate potential cyber threats

Statistical Formulas Behind the Measures

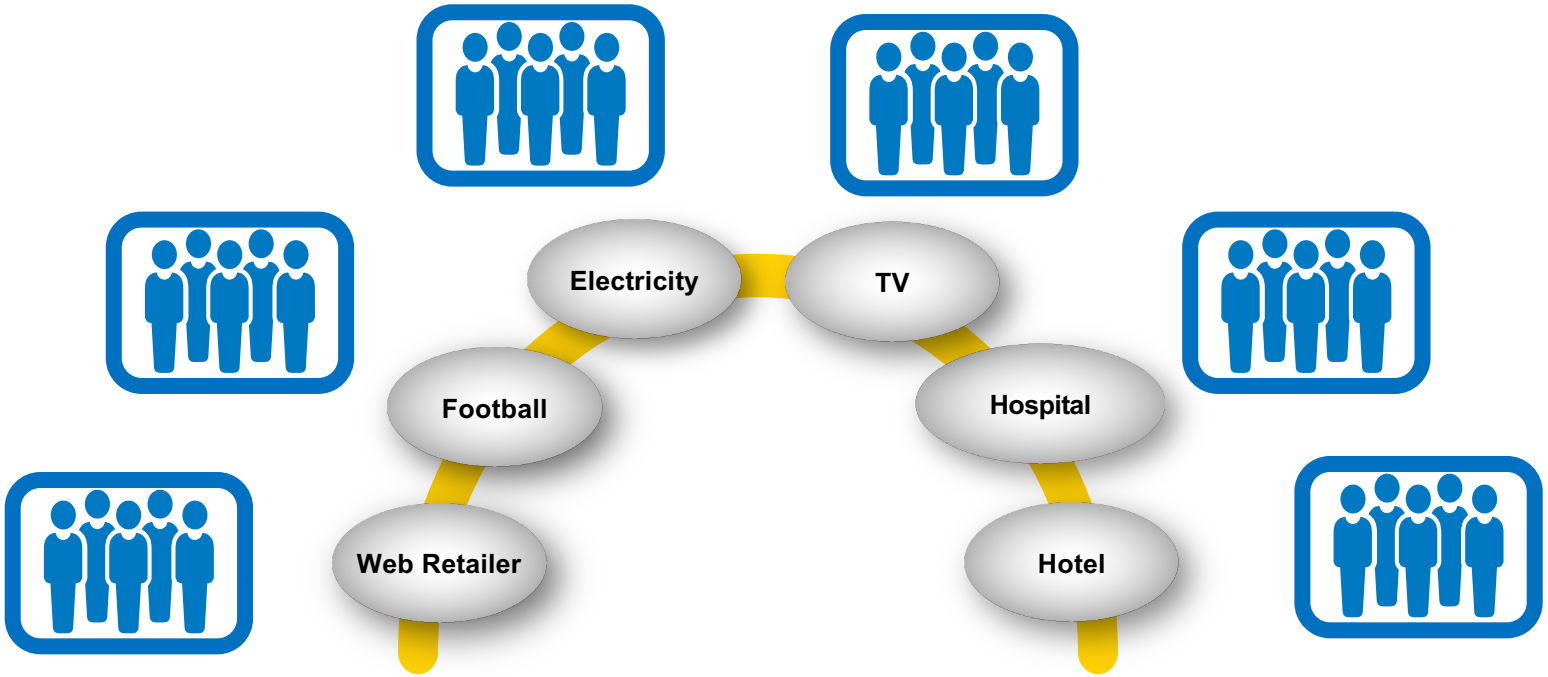
The mathematical foundation of performance measurement

Formula	Description	Possible Use case
Mean (Average) $\mu = (\Sigma x) / N$	Central value of a dataset	Financial forecasting, performance benchmarking
Median Middle value in ordered data	Robust measure of central tendency (less affected by outliers)	Salary analysis, skewed data (e.g., housing prices)
Standard Deviation (σ) $\sigma = \sqrt{(\Sigma(x - \mu)^2 / N)}$	Measures data dispersion	Risk assessment (e.g., stock volatility)
Expected Value (EV) $EV = \Sigma (x \times P(x))$	Mean outcome of probabilistic events	Decision-making under uncertainty (e.g., project Return on Investment)
Value at Risk (VaR) $VaR = \mu - (Z \times \sigma)$	Maximum potential loss at a confidence level	Financial risk management
Correlation Coefficient (r) $r = \Sigma[(x - \mu_x)(y - \mu_y)] / (\sigma_x \times \sigma_y)$	Measures linear relationship (-1 to 1)	Market research, sales vs. ad spend
Simple Linear Regression $Y = a + bX + \varepsilon$	Predicts dependent variable (Y) from independent (X)	Demand forecasting, pricing models
Defect Rate Defect Rate = (Defective Units / Total Units) × 100	% of faulty products	Six Sigma, manufacturing quality control
Capacity Utilization Utilization = (Actual Output / Max Output) × 100	Measures operational efficiency	Production planning

Define a Structure for Data Reporting

Which data to report, for which purpose, and for which decisions? (Stakeholders, Business needs, Data)

- A Web Retailer
- Champion Leagues
- Electricity Suppliers
- TV channel
- Hospital
- Hotel Chain

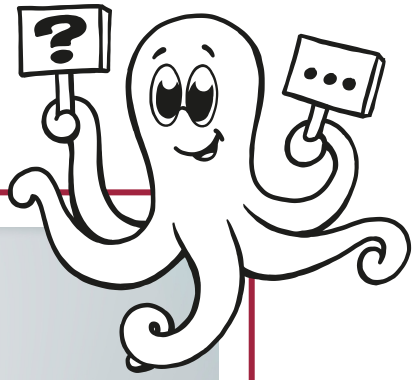


From Reporting to Better Business Decisions

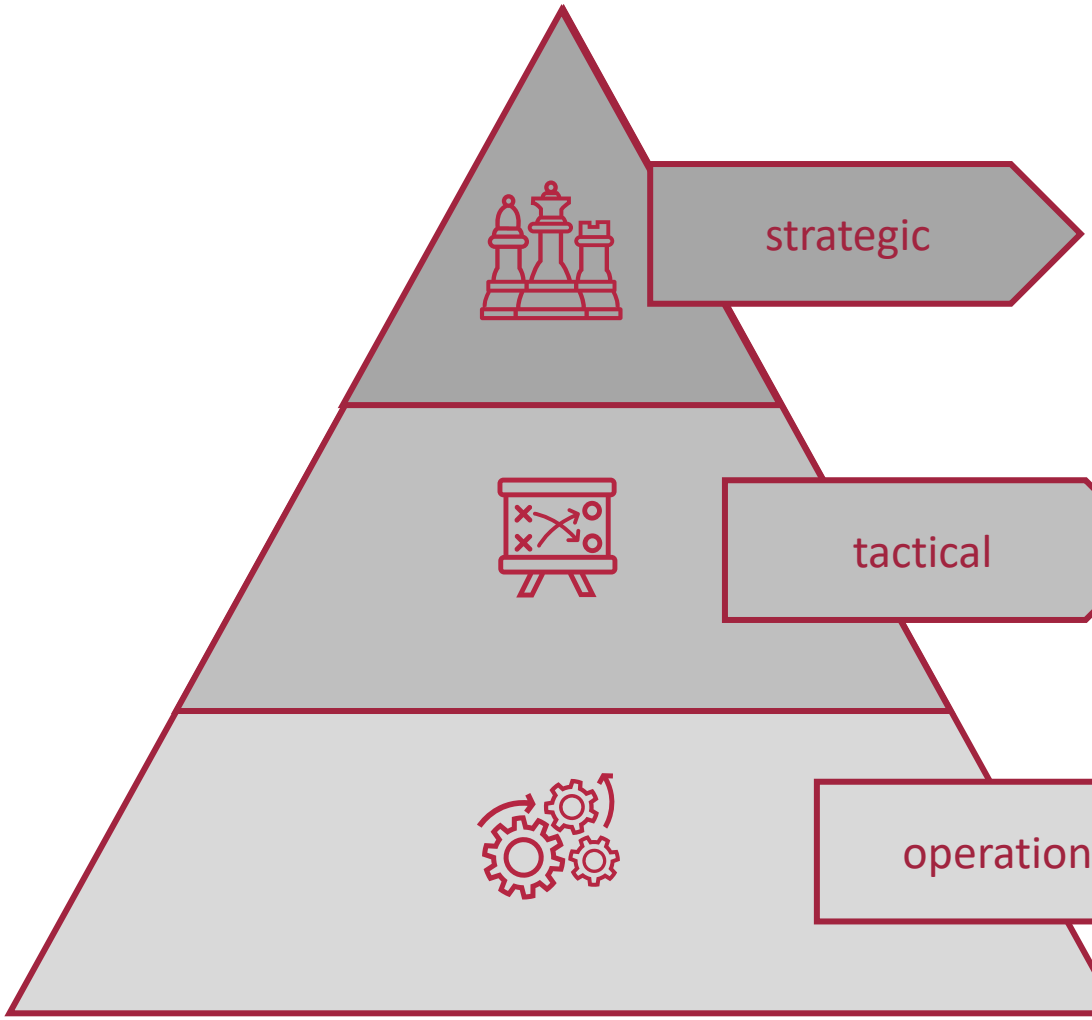
From intuition to evidence

What Kinds of Decisions Do Organizations Make?

Daily, medium-term, long-term



Decision-making in a Firm



- Long-term goals
- Entrance or exit from markets
- Acquisition of a competitor
- Investing or divesting in which business units

- Annual marketing budget
- Long-term contracts with suppliers
- Workforce scheduling across branches
- Production capacity based on demand forecasts

- Escalation of customer complaints
- Reordering products and restocking inventories
- Optimizing delivery route

Data Driven Decision Making (DDDM)

Translating a Question into a Data Problem

Classify the Question

Identify the **type** of question (descriptive, diagnostic, predictive, prescriptive) to determine the analytical approach

Break Down the Question

Deconstruct the question into **measurable components**

Define Data Requirements

Identify **data sources** and **metrics** needed to answer the question

Choose the Analytical Method

Map the question to a **data science or statistical technique**

Operationalize the Answer

Turn insights into **actionable outputs**

Fundamental Types of Questions by Purpose

Type	Definition	Primary Usage	Business Questions
Descriptive	Summarize *'what happened'* quantitatively	Reporting, performance tracking	• How many units did we sell last quarter? • What's our current inventory turnover rate?
Diagnostic	Identify causes or correlations behind events	Root-cause analysis, troubleshooting	• Why did customer churn spike in Q2? • Which marketing channel underperformed?
Predictive	Forecast future outcomes probabilistically	Risk assessment, demand planning	• What will next year's revenue be if trends continue? • Which customers are likely to default?
Prescriptive	Recommend actions to optimize outcomes	Strategic decision-making	• Should we open a new warehouse in Texas? • What's the optimal pricing strategy for Product X?

Storytelling

- **Narrative Structure**: create a narrative around the data

  → Organize information logically to guide the viewer through the insights, highlighting trends, patterns and outliers

- **Audience Engagement**: tailor the dashboard to the specific audience

  → This ensures that the story resonates with, enhances engagement and facilitates better decision-making

- **Actionable Insights**: present data and insights to drive action

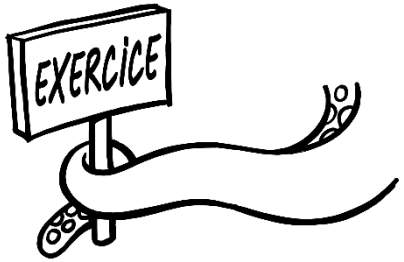
  → Dashboards can influence decision-making and prompt stakeholders to take appropriate actions based on the findings

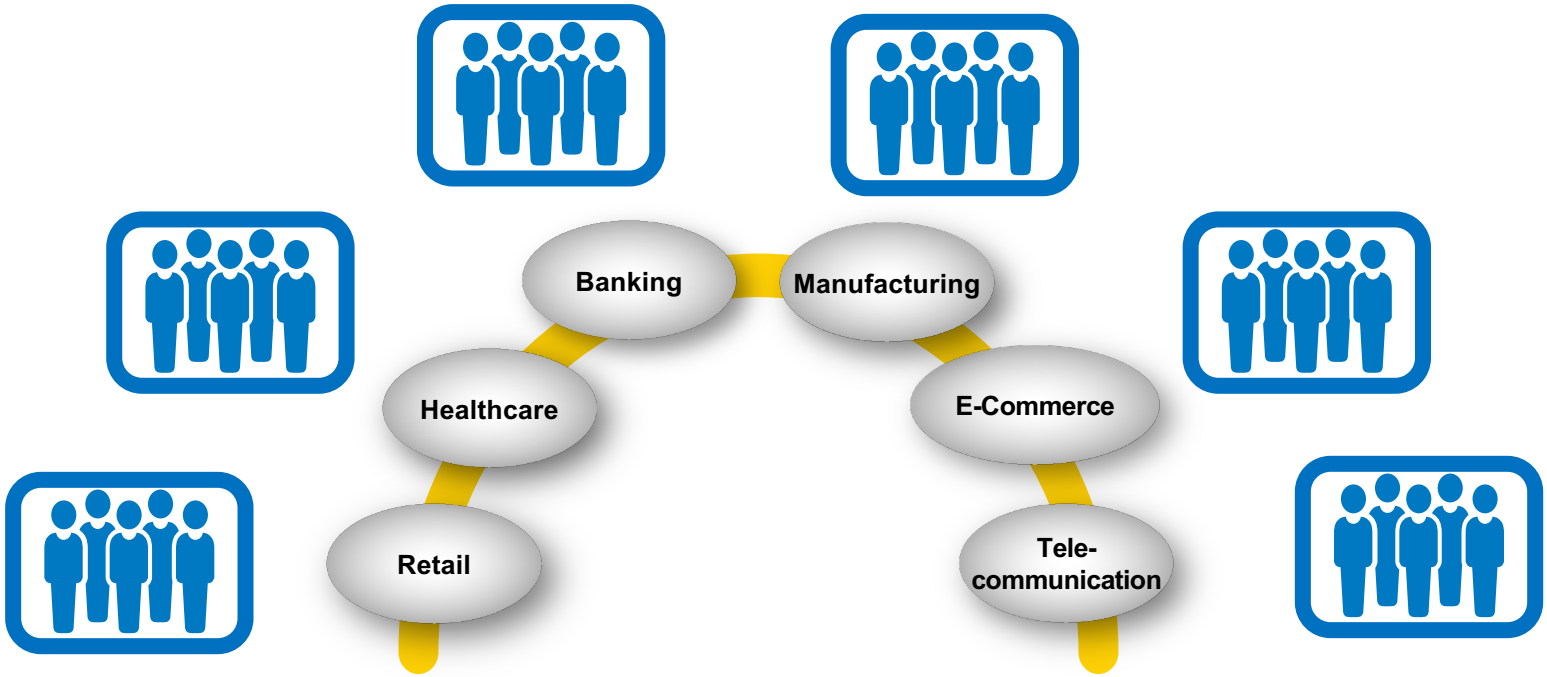
Drill-down into the data until you can answer the business question

Data Problem Decoder: Turning Questions into Analytical Actions

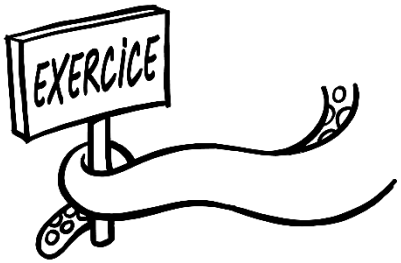
For each case (see next slide), identify:

- Question type (descriptive, diagnostic, predictive, prescriptive)
- Key variables (what data is needed?)
- Analytical approach (methods/tools)
- Expected output (dashboard, model, report)





Case Description



Type	Question
Retail	*Why did online sales decrease by 15% in December compared to last year?*
Healthcare	*Which patients are at high risk of readmission within 30 days?*
Banking	*What's the optimal loan approval threshold to minimize defaults while maximizing profit?*
Manufacturing	*How can we reduce equipment downtime in Factory A?*
E-Commerce	*What product recommendations increase average order value?*
Telecommunication	*Are there customer segments more likely to switch to competitors?*

Examples of Data-Driven Decision-Making

1. Leadership Development at Google

Google maintains a heavy focus on what it refers to as "*people analytics*". As part of one of its well-known people analytics initiative, called **Project Oxygen**, Google mined data from more than 10'000 performance reviews and compared data with employee retention rates. Google used the information to identify common behaviors of high-performing managers and created training programs to develop these competencies.

These efforts boosted median favorability scores for managers from 83% to 88%.

2. Real Estate Decisions at Starbucks

After hundreds of Starbwere closed in 2008, then-CEO Howard Schultz promised that the company would take a more analytical approach to identify future locations.

Starbucks now partners with a location-analytics company to pinpoint ideal store locations using data like demographics and traffic patterns. The organization also considers input from its regional teams before making decisions.

Starbucks uses this data to determine the likelihood of success for a particular location before taking a new investment.

3. Driving sales at Amazon

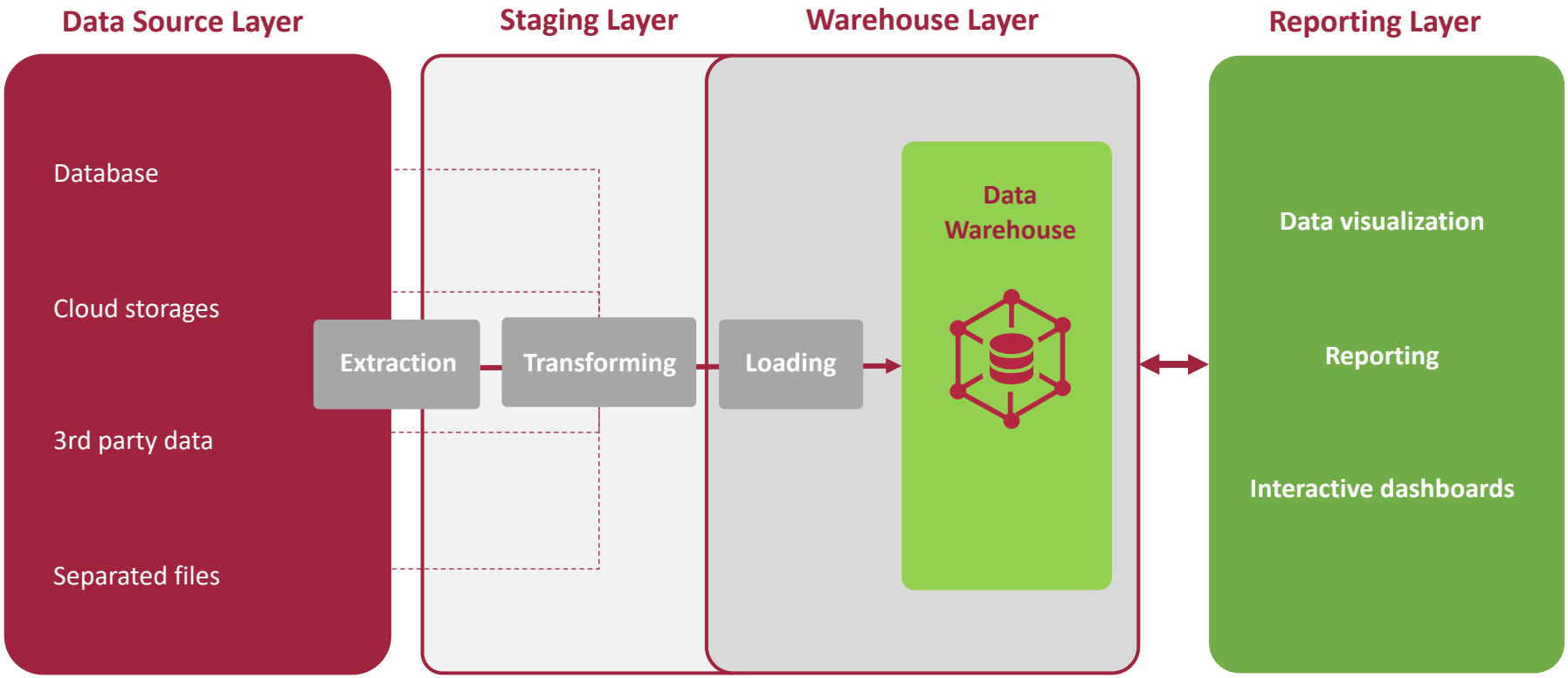
Amazon uses data to decide which products to recommend to customers based on prior purchases and patterns in search behavior.

Rather then blindly suggesting a product, Amazon uses data analytics and machine learning to drive its recommendation engine.

McKinsey estimated that, in 2017, 35% of Amazon's consumer purchases could be tied back to the company's recommendation system.

Source: online.hbs.edu/blog/post/data-driven-decision-making

Generic Architecture of Reporting Systems



What is a Management Information System (MIS)?

A **Management Information System (MIS)** is a computerized system that collects, processes, stores, and distributes structured data to support **operational and tactical decision-making** within an organization. It typically includes:

- **Integration with enterprise systems** (ERP, CRM, SCM),
- **Databases** (for storing structured data),
- **Reporting tools** (e.g., dashboards in Power BI, Tableau).

MIS generates **standardized, periodic reports** (e.g., daily, weekly, monthly) to monitor performance, track key metrics, and facilitate structured decision-making.

Possible Use Cases are:

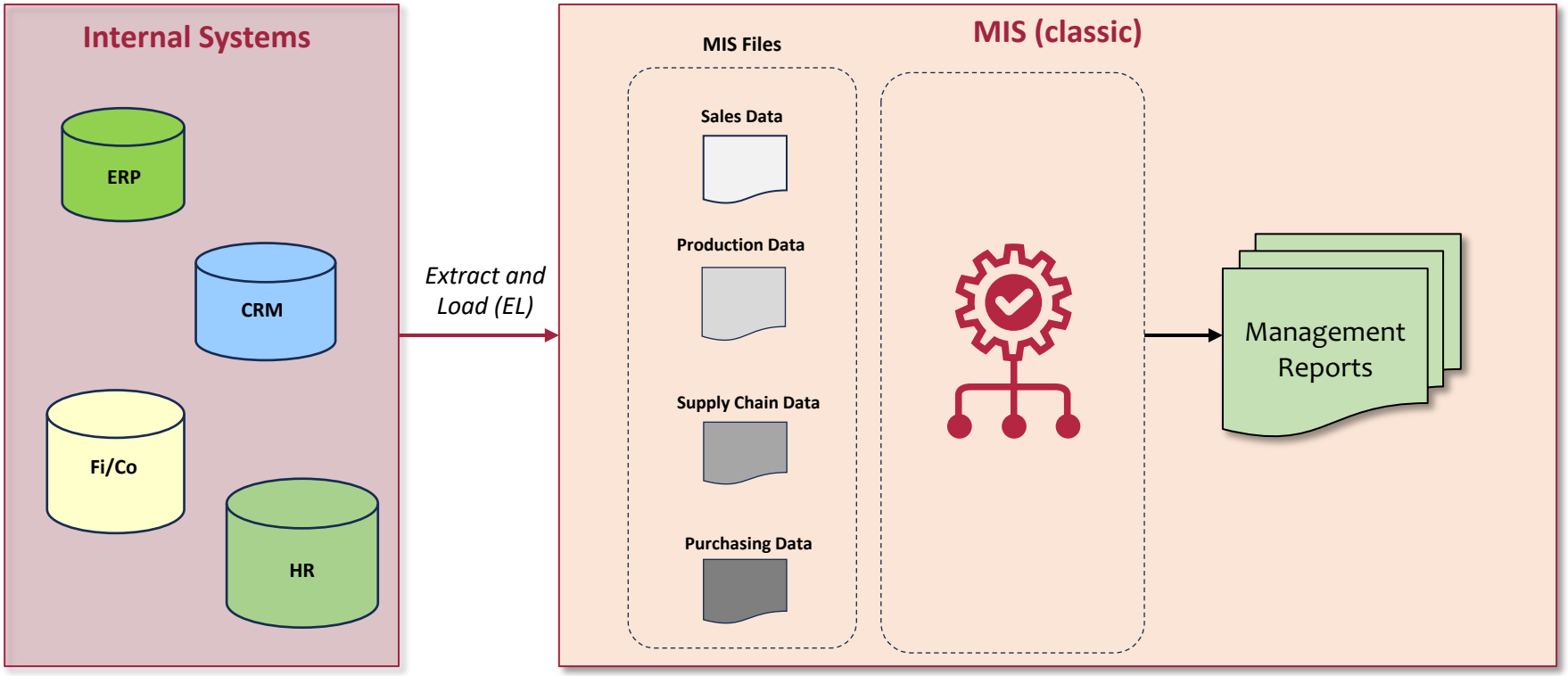
- Monthly sales performance reports
- Inventory tracking & stock-level analysis
- Budget vs. actual expenditure reports
- Employee productivity dashboards

Key Limitations are:

1. **Static & Historical Data** – MIS primarily relies on past data, limiting real-time insights
2. **Structured Data Dependency** – Requires well-defined data inputs and struggles with unstructured data (e.g., social media, text)
3. **Limited Predictive Capabilities** – Focuses on descriptive analytics rather than forecasting or prescriptive recommendations
4. **Rigid Reporting** – Reports are predefined, making ad-hoc analysis difficult without customization

MIS Data Processing

MIS = operational reporting



What is a Data Warehouse (DWH)?

A **Data Warehouse** is a centralized repository that consolidates large volumes of **historical, structured, and semi-structured data** from multiple sources (e.g., databases, ERP, CRM, IoT) to support **Analytical Reporting, Business Intelligence (BI), and Data-Driven Decision-Making (DDDM)**.

Possible Use Cases

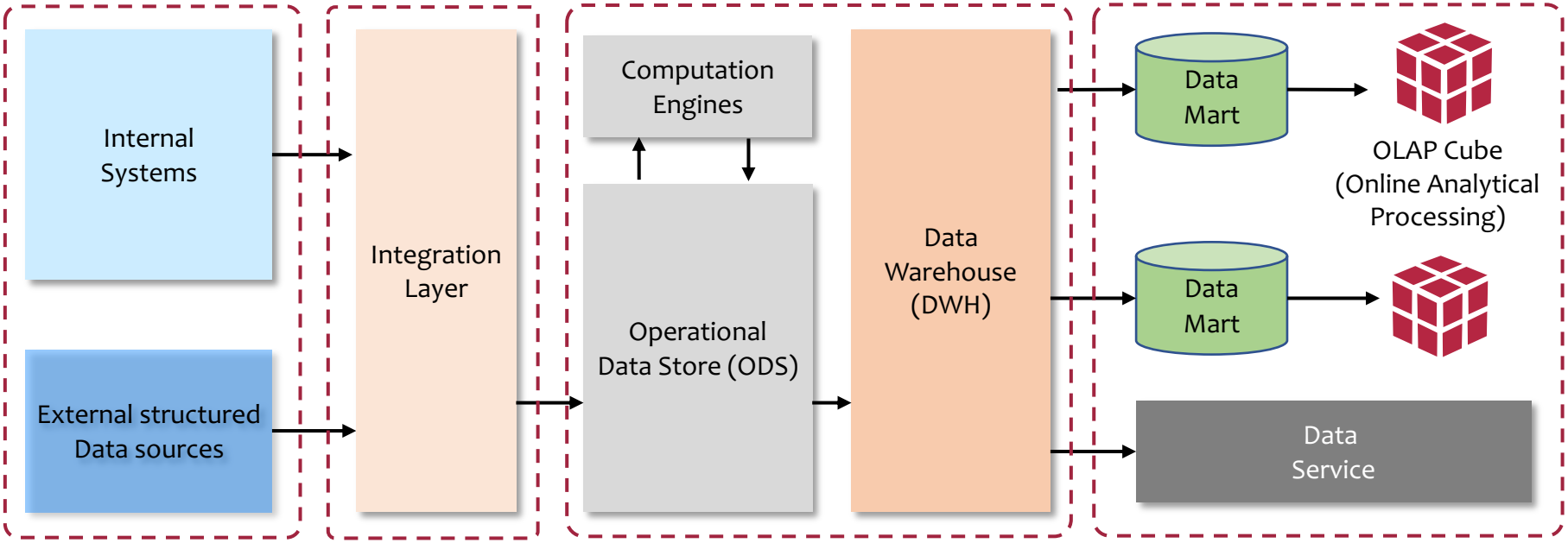
- Enterprise-wide trend analysis (e.g., yearly sales performance across regions)
- Customer behaviour analytics (e.g., segmentation, lifetime value)
- Regulatory/compliance reporting (e.g., financial audits)
- AI/ML training datasets (predictive analytics)

Key Limitations

1. Complex Setup: Requires careful schema design and governance
2. High Maintenance: Costs scale with data volume and transformation logic
3. Structured Focus: Less suited for raw/unstructured data (e.g., social media logs)
4. Not Real-Time: Data is batched and loaded periodically (latency vs. data lakes/streaming)

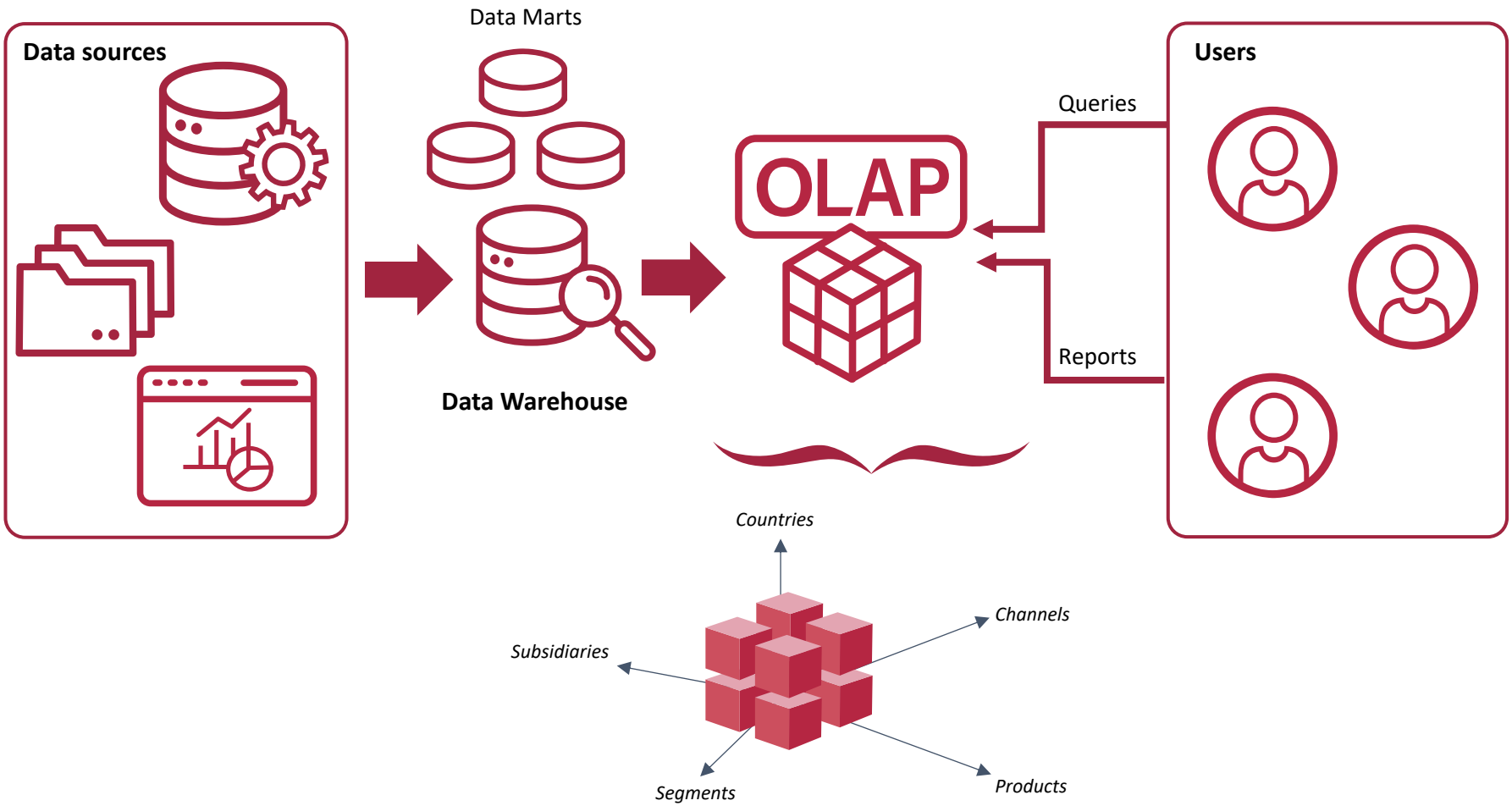
DWH Data Processing

DWH = structured analytics



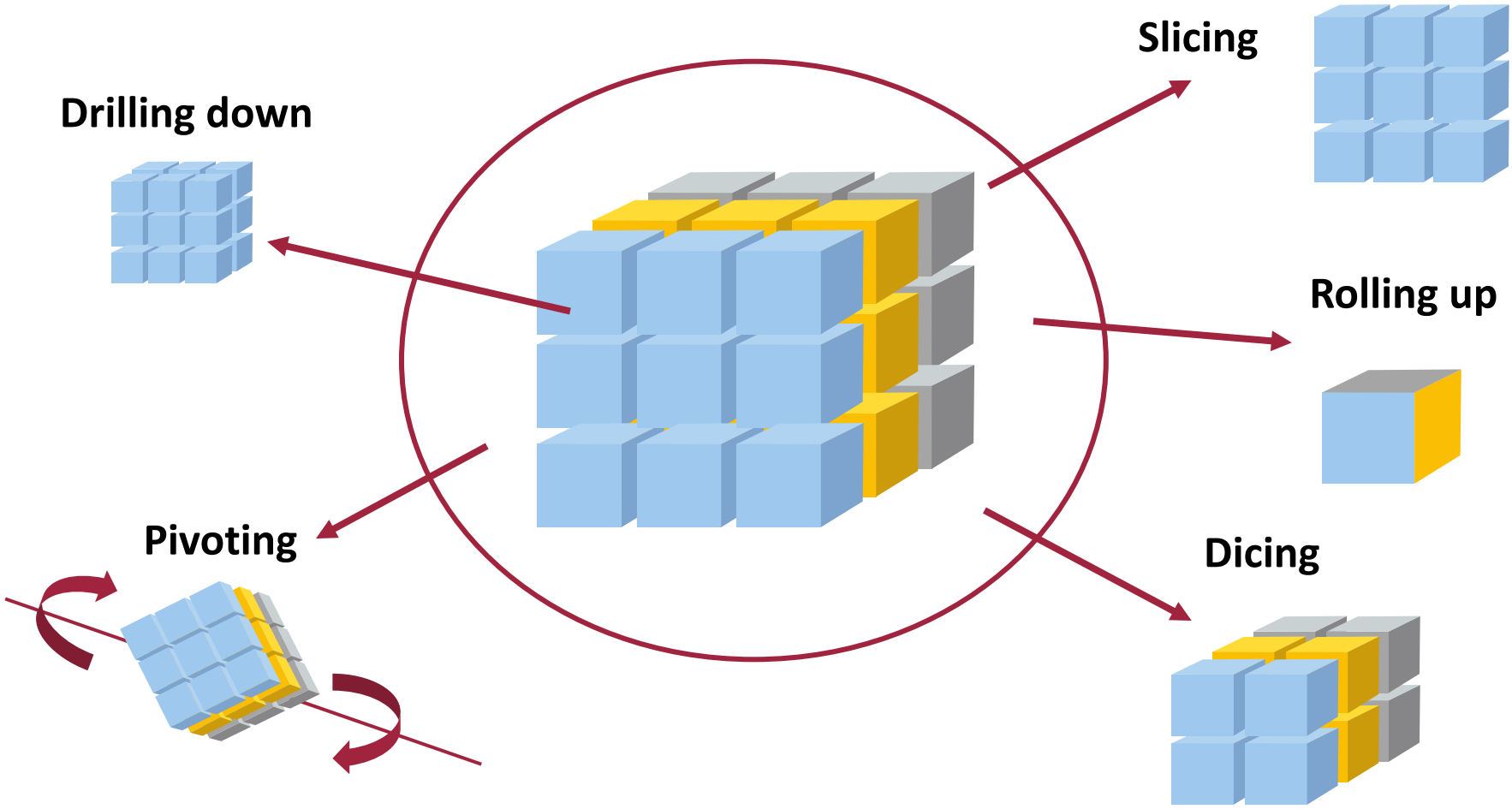
Online Analytical Processing (OLAP)

Technology for analysing business data from multiple perspectives and performing complex multidimensional analysis



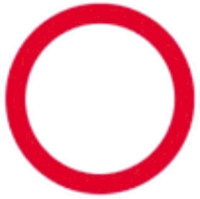
Drilling down, Pivoting, Slicing, Rolling up, Dicing

OLAP = multidimensional analysis



DWH Challenges

High Maintenance Costs



Security & Compliance Risks

Data Quality Issues

User Adoption & Training

Performance & Scalability Problems

Over-Engineering

What is a Data Lake?

A **Data Lake** is a **scalable, centralized storage system** that retains **raw data in its native format** (structured, semi-structured, or unstructured) for future processing and analysis.

Unlike traditional warehouses, it accommodates:

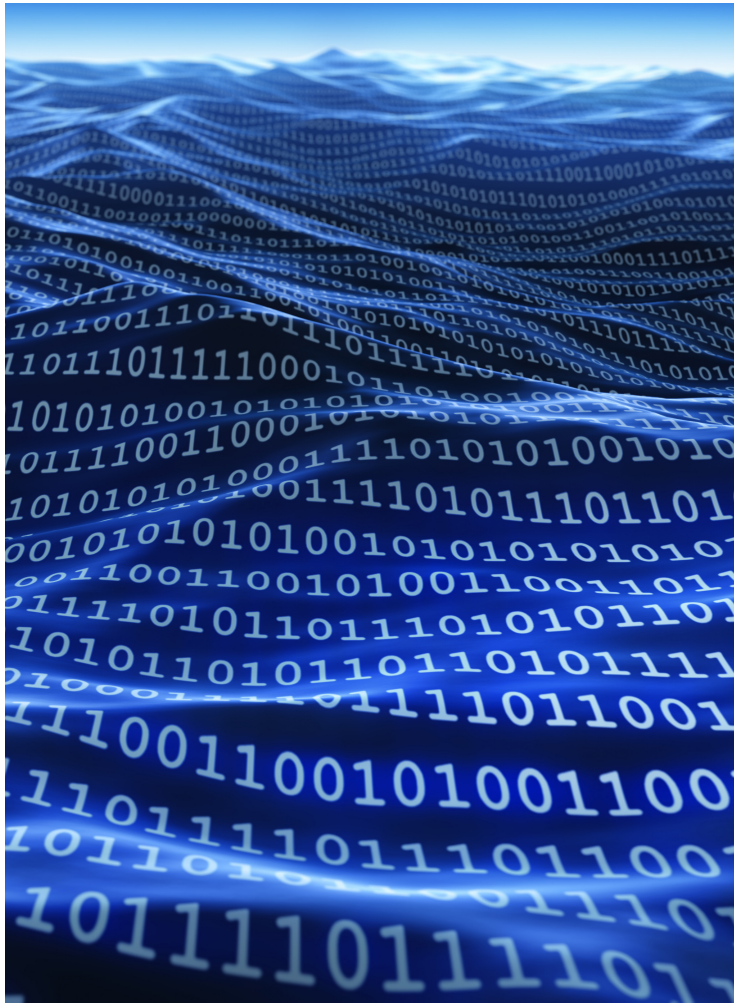
- **Massive volumes** (Big Data from IoT, logs, social media, etc.)
- **Schema-on-read** (no predefined structure until analysis)
- **Support for batch, real-time, and streaming data**

Possible Use Cases are:

- Advanced analytics (e.g., sentiment analysis on customer reviews)
- Machine Learning (training models on raw, unprocessed datasets)
- IoT & telemetry data (storing sensor/device data at scale)
- Data exploration (ad-hoc analysis without upfront modelling)

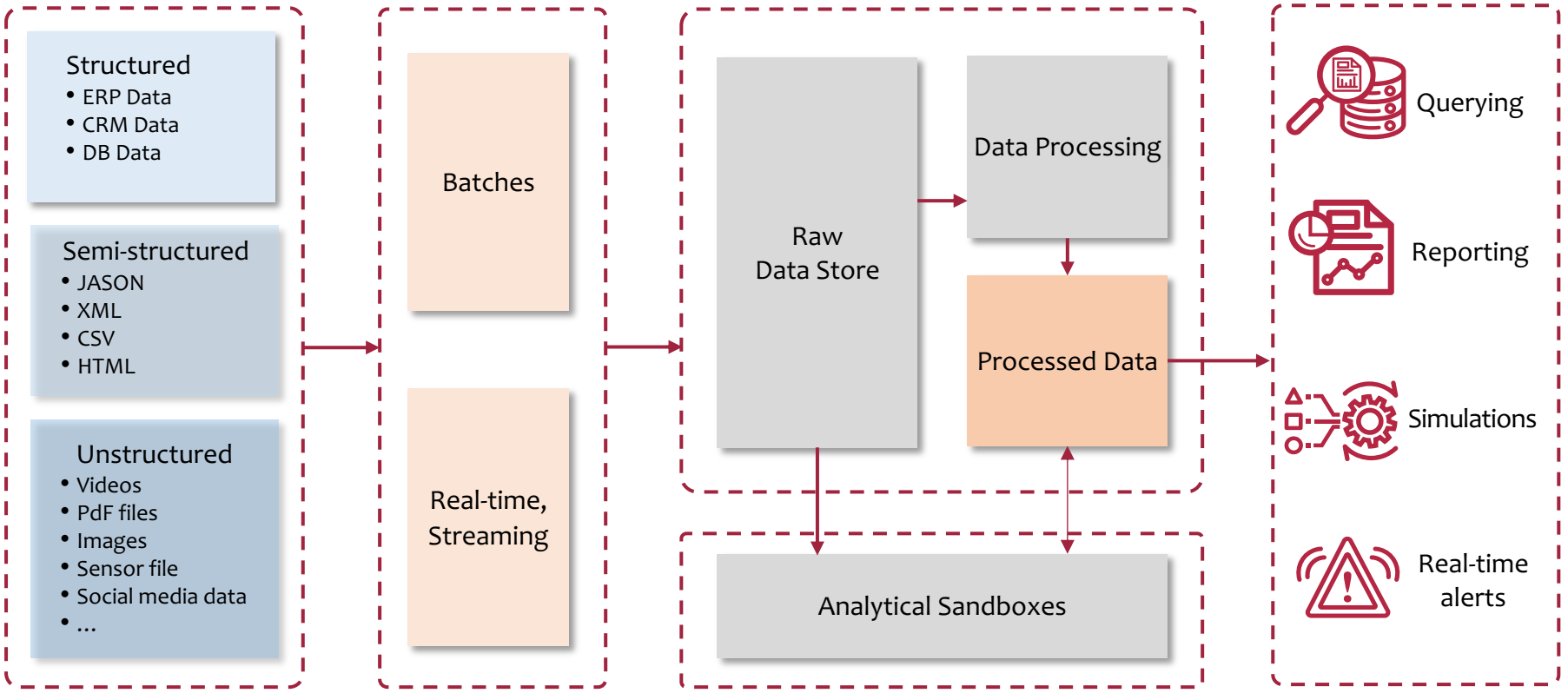
Key Components are:

- Distributed Storage (e.g., AWS S3, Azure Data Lake, Hadoop HDFS)
- Metadata Management (tagging, cataloguing for discoverability)
- Processing Frameworks (e.g., Apache Spark, Databricks for transformations)
- Integration with AI/ML tools (e.g., TensorFlow, PyTorch for model training)

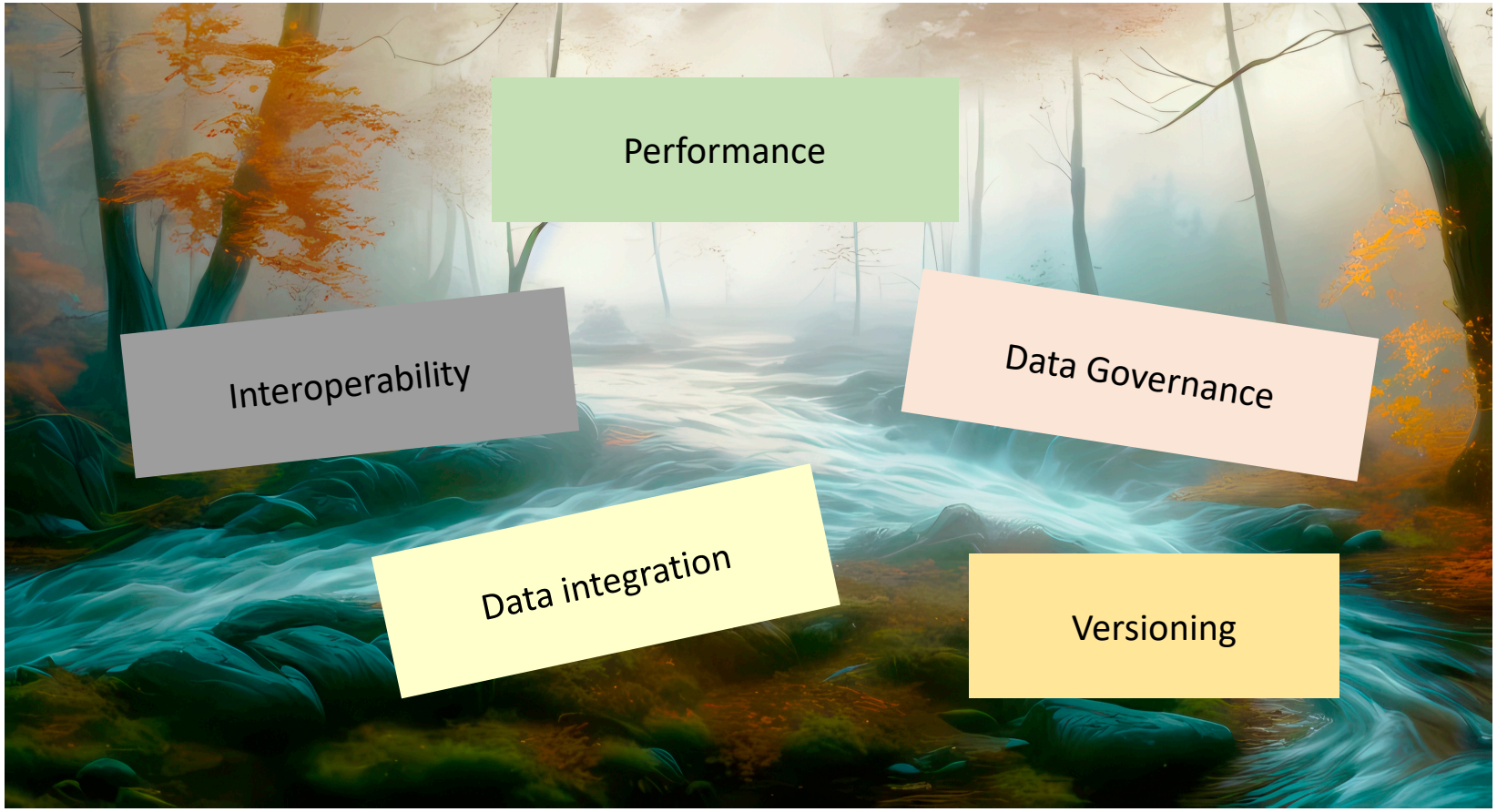


Example of a Data Lake Architecture

Data Lake = flexible and advanced analytics



Data Lake Challenges



Features of MIS, DWH and Data Lake

Feature	MIS (classic)	Data Warehouse (DWH)	Data Lake
Primary Purpose	Structured operational reports	Enterprise analytics & BI	Raw data storage and advanced analytics
Data Type	Structured, pre-processed	Structured, cleaned, historical	Raw (structured, semi-structured, unstructured)
Schema	Fixed schema	Schema-on-write (predefined model)	Schema-on-read (flexible)
Users	Managers, operational teams	Analysts, BI teams, executives	Data scientists, engineers
Flexibility	Low (static reports)	Moderate (ad-hoc queries)	High (exploratory, AI/ML-ready)
Latency	Periodic (daily/monthly)	Batch (hourly/daily updates)	Batch/real-time/streaming
Best For Organizations	Small to mid-sized businesses needing standardized reports (e.g., retail stores, local banks)	Medium to large enterprises requiring cross-departmental analytics (e.g., e-commerce, healthcare, finance)	Data-driven or tech-heavy companies (e.g., AI startups, IoT firms, social media platforms)
Use Cases	- Monthly sales reports - Inventory tracking	- Customer 360° analytics - Financial forecasting	- AI model training - Log analysis from servers

Macro-level Analytics: Beyond the Firm

Benchmarking, industry trends, and external data

Sustainability Analytics

Example of Sustainability Analytics

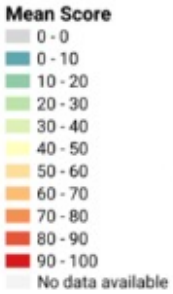
Regional Physical Hazard Metrics at CH Postal Code Level

Country	ZIP CODE	LOCATION COUNT	TIMESTEP Year	EXTREME HEAT	EXTREME COLD	EXTREME WIND	EXTREME PRECIPITATION	EXTREME SNOWFALL	WATER SCARCITY	TROPICAL CYCLONES	COASTAL FLOODING	FLUVIAL FLOODING	PLUVIAL FLOODING	RIVER LOW FLOW	WILD FIRE
Switzerland	xxx	10	2023	20	66	0	34	86	21	0	0	0	59	29	12
Switzerland	xxx	4	2023	16	68	0	47	92	22	0	0	0	0	28	4
Switzerland	xxx	3	2023	16	68	0	47	92	22	0	0	0	0	28	12
Switzerland	xxx	11	2023	15	68	0	40	90	22	0	0	0	11	26	7
Switzerland	xxx	11	2023	16	68	0	47	92	21	0	0	0	25	28	12
Switzerland	xxx	10	2023	16	68	0	47	92	21	0	0	0	49	28	11
Switzerland	xxx	3	2023	16	68	0	47	92	21	0	0	0	0	28	6
Switzerland	xxx	9	2023	16	68	0	47	92	21	0	0	0	10	28	4
Switzerland	xxx	15	2023	16	68	0	47	92	21	0	0	0	11	28	11
Switzerland	xxx	5	2023	14	75	0	34	74	21	0	0	0	21	26	4
Switzerland	xxx	20	2023	14	75	0	34	74	21	0	0	0	16	28	10
Switzerland	xxx	41	2023	14	75	0	34	74	21	0	0	0	37	28	3
Switzerland	xxx	17	2023	14	75	0	34	74	21	0	0	0	6	28	12
Switzerland	xxx	12	2023	14	66	0	34	74	21	0	0	0	12	28	12
Switzerland	xxx	12	2023	14	66	0	34	74	21	0	0	0	33	28	7
Switzerland	xxx	7	2023	14	75	0	34	74	21	0	0	0	16	28	12
Switzerland	xxx	13	2023	14	66	0	34	74	21	0	0	0	5	28	7
Switzerland	xxx	4	2023	20	66	0	34	86	21	0	0	0	0		13
Switzerland	xxx	5	2023	20	66	0	34	86	21	0	0	0	0	29	7
Switzerland	xxx	2	2023	16	68	0	47	90	21	0	0	0	59	28	11
Switzerland	xxx	12	2023	20	66	0	34	86	21	0	0	0	15	29	5
Switzerland	xxx	10	2023	16	68	0	47	90	21	0	0	0	44	28	11
Switzerland	xxx	2	2023	16	68	0	47	90	21	0	0	0	0	28	4
Switzerland	xxx	9	2023	16	68	0	47	90	21	0	0	0	0	28	11
Switzerland	xxx	12	2023	16	68	0	47	90	21	0	0	0	52	28	11

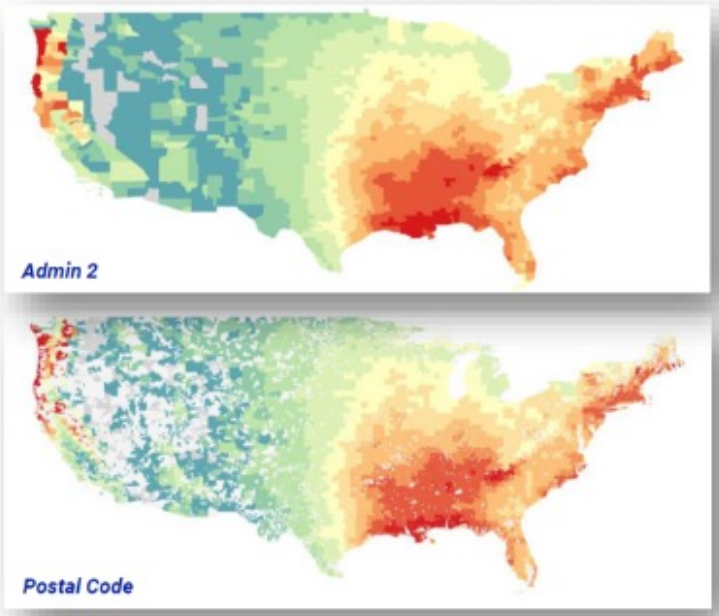
Score between 0 and 100, computed based on number of days per year over 100 years

Source: MSCI

Regional Physical Hazard Metrics - Spatial Resolution

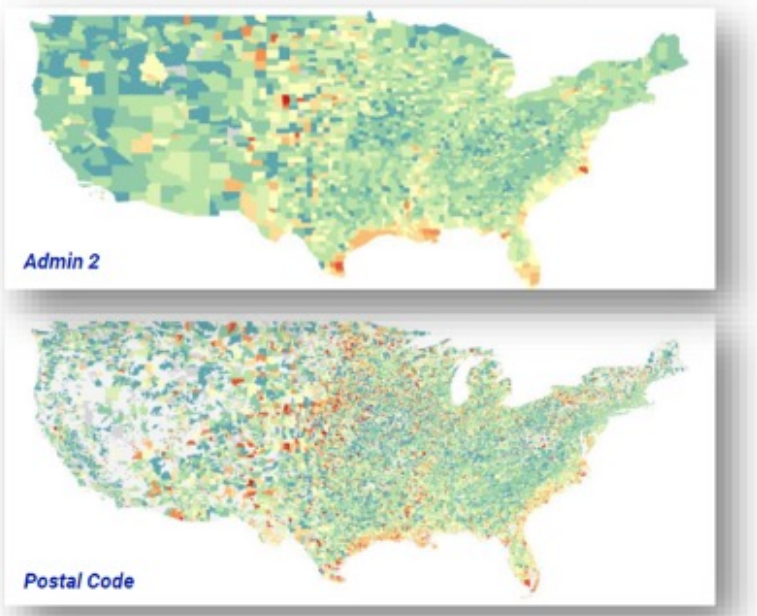


Heavy Precipitation Scores for Base Year (2023)



Exposure to **chronic** hazards is more consistent over broader scales, so admin 1 or 2 data levels may be sufficient

Pluvial Flooding Scores for Base Year (2023)



Acute hazards are event-based and because exposure can vary significantly over small scales, more granular data levels may be preferable

Source: MSCI

Geospatial Analytics

11,000 to 15,000 satellites orbiting Earth

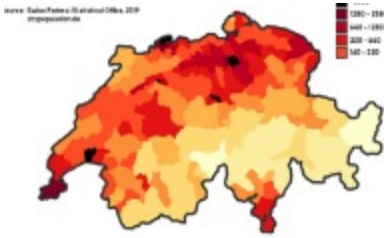
Geospatial Analytics: Some Use Cases

Smart Cities

Deforestation Tracking

Precision Farming

Customer demographics



Fleet Tracking

Border Surveillance

Health Care Data Analytics

Key Applications of Healthcare Data Analytics

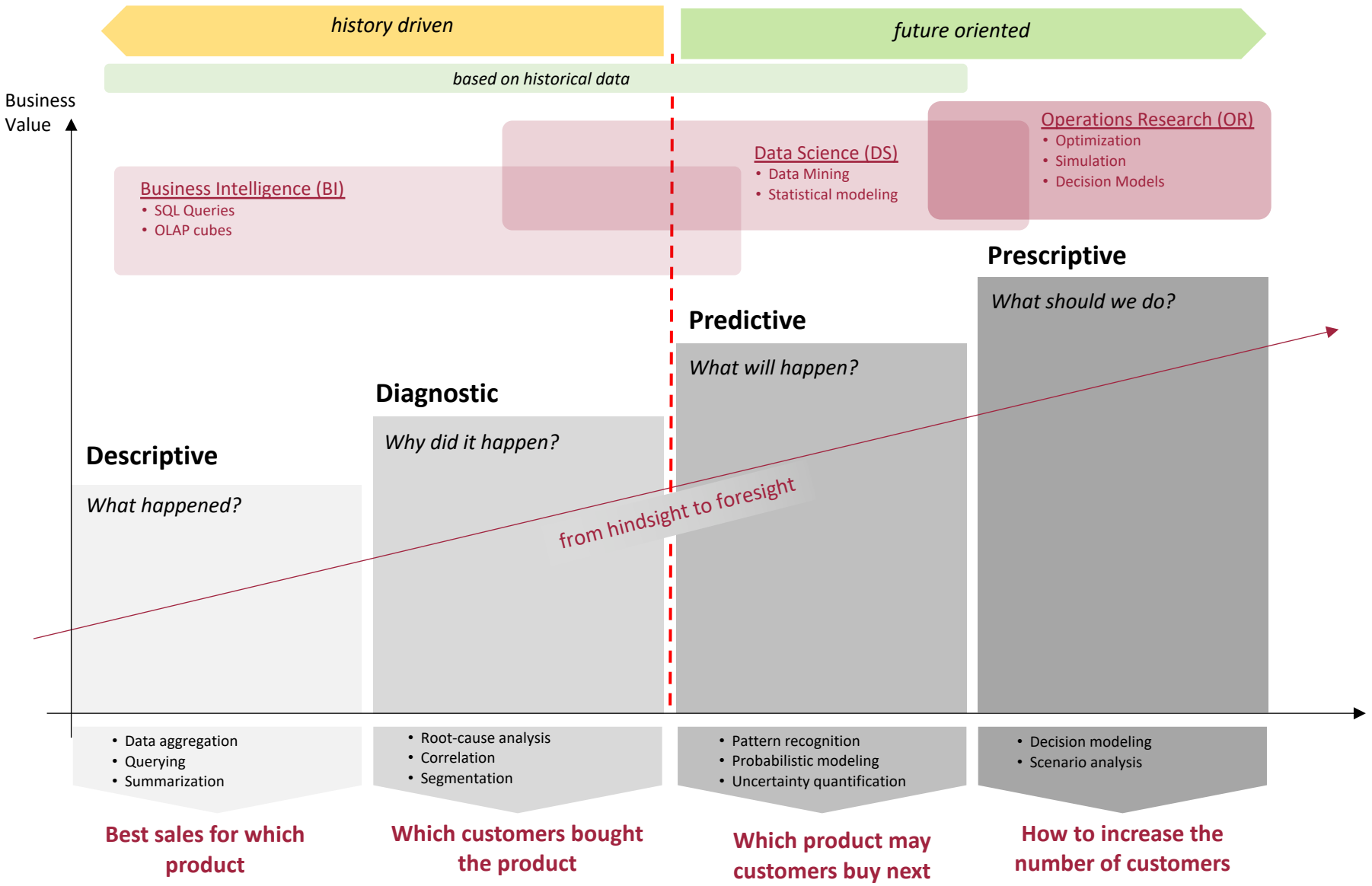
Application	Use Case	Example
Predictive Analytics	Forecasting disease outbreaks or readmissions	Mayo Clinic's AI predicts admissions at the Intensive Care Unit
Diagnostic AI	Detecting tumours (MRI) or diabetic retinopathy	Google Health's AI matches radiologists in breast cancer detection
Population Health	Tracking chronic diseases (diabetes, hypertension)	IBM Watson analyses Electronic Health Records for at-risk patients
Telemedicine Analytics	Optimizing virtual care workflows	Teladoc uses Natural Language Processing to triage patient calls
Genomics and Precision Med	Matching treatments to DNA (e.g., CAR-T therapy)	Tempus Labs uses AI for oncology analytics
Operational Analytics	Reducing hospital-acquired infections (HAIs)	GE Healthcare's Command Centre cuts HAIs by 20%

MRI: Magnetic Resonance Imaging
HAI: Hospital-Acquired Infections

Financial and Risk Data Analytics

Examples of Financial and Risk Data Analytics

Analytical Approaches



Capability Mapping

Analytics types = what question is asked
Disciplines (BI, DS, OR) = how the question is solved

„how“ →

„what“ ↓

Analytics Type	BI	DS	OR
Descriptive	✓✓	✓	✗
Diagnostic	✓✓	✓✓	✓
Predictive	✓	✓✓✓	✓
Prescriptive	✗	✓	✓✓✓

Definition of Business Intelligence (BI)

BI is the set of technologies and practices that convert historical and current data into insights about past and present performance — answering the questions 'What happened?' and 'Why did it happen?' to support operational and tactical decisions

History of Business Intelligence

Pre-BI Era
(Pre-1970s)

Manual Reporting & Early Computing
- 1958: IBM researcher Hans Peter Luhn first coined the term "Business Intelligence" in a paper describing an automated system for disseminating information

BI 1.0
(1979s – 1990s)

The Era of Data Warehousing and Static Reports
- IT-controlled, batch-processed reports.
- Limited to historical data analysis
- 1990s: First BI tools: Cognos, BusinessObjects

BI 2.0
(2000s–2010s)

Self-Service BI and Dashboards
- Interactive dashboards
- Faster, ad-hoc analysis
- Still mostly descriptive analytics (what happened?)

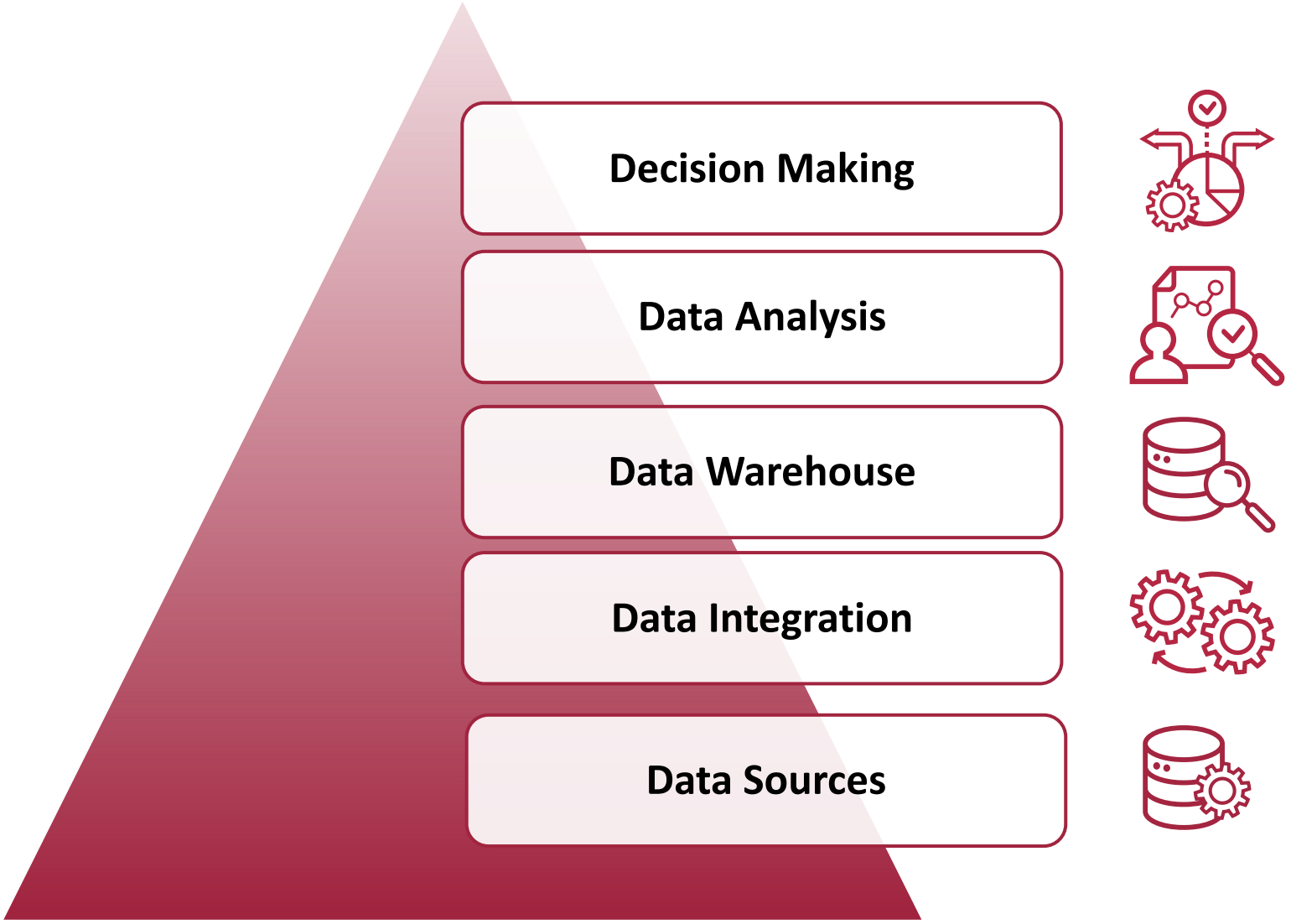
BI 3.0
(2010s–2020)

AI, Big Data and Predictive Analytics
- Predictive analytics ("what will happen?“)
- Real-time data processing
- Embedded BI, inside apps like Salesforce, Slack

BI 4.0
(2020s–Present)

Augmented and Autonomous BI
- Autonomous BI (prescriptive analytics)
- Conversational analytics (chat-based queries)
- Hyper-personalization (BI tailored to individual users)

Building Business Intelligence



Example of BI Use: Starbucks store location

Starbucks carefully analyses several data points to identify what might make a successful store location.
This includes:

- Consumer demographics
- Population density
- Average income levels
- Traffic patterns
- Public transport hubs
- Types of businesses in the location under consideration

BI also has its hands in Starbucks' " Starbucks Rewards" program, where it's leveraged to track consumer behaviour and generate tailored customer offers during the times that buyers are predicted to be most likely to purchase a beverage. This not only drives sales but also improves brand loyalty as Starbucks is able to give their customers exactly what they want, right when they want it.

What is Data Science (DS)?

BI tells what happened.
DS tells what will happen and what to do about it.

Dimension	BI	DS
Question	What happened? Why?	What will happen? What should we do?
Focus	Past and present	Future and action
Data	Structured, historical	Structured and unstructured
Techniques	Reporting, OLAP, dashboards	Predictive models, recommendations, classifications
User	Managers, analysts	Data scientists

At the heart of Data Science lies **Data Mining**, the process of discovering non-trivial patterns and relationships in large datasets.

What is Data Mining?

Process of analyzing large datasets to uncover non-trivial patterns and relationships using statistical and database techniques, generating insights that are not immediately apparent.

Data Mining is one of the core methodological foundations of Data Science.

Data Mining Process

According to the CRoss Industry Standard Process for Data Mining (CRISP-DM)

Source: Wikipedia

Data Mining Techniques

what you want to achieve

Technique	Description	Usage	Use Cases
Classification	Assigns predefined labels to data based on features	Supervised learning (= labelled data)	Spam detection, credit scoring, medical diagnosis
Regression	Predicts continuous numerical values based on input variables	Supervised learning	Sales forecasting, house price prediction, risk assessment
Clustering	Groups similar data points without predefined labels	Unsupervised learning (no labels)	Customer segmentation, anomaly detection, image recognition
Association Rules	Discovers relationships between variables (e.g., "If X, then Y")	Pattern mining in transactional data	Market basket analysis (e.g., "Customers who buy chips also buy soda")
Anomaly Detection	Identifies rare or unusual data points deviating from the norm	Supervised and Unsupervised	Fraud detection, network intrusion, manufacturing defects
Sequential Pattern	Finds patterns where events lead to subsequent events over time	Time-series and sequence analysis	Clickstream analysis, patient treatment pathways, stock market trends
Text Mining	Extracts insights from unstructured text (NLP)	Natural Language Processing (NLP)	Sentiment analysis, chatbots, document classification
Dimensionality Reduction	Reduces number of features while preserving key information	Preprocessing for efficiency	Image compression, feature selection for ML models

Data Mining Algorithms

how you want to achieve

Algorithm	Description	Technique	Use Cases
Decision Trees	Splits data into branches based on feature values to make predictions	Classification, Regression	Customer churn prediction, medical diagnosis
Random Forest	Ensemble of decision trees for improved accuracy and robustness	Classification, Regression	Credit scoring, stock market analysis
K-Means	Partitions data into 'k' clusters based on centroid proximity	Clustering	Market segmentation, image compression
DBSCAN	Groups data based on density, identifying outliers automatically	Clustering	Anomaly detection, geographic data analysis
Linear Regression	Models linear relationships between dependent and independent variables	Regression	Sales forecasting, real estate pricing
K-Nearest Neighbours (KNN)	Classifies data based on majority vote of its 'k' nearest neighbours	Classification	Handwriting recognition, recommendation systems
Principal Component Analysis (PCA)	Reduces data dimensionality while preserving variance	Dimensionality Reduction	Facial recognition, noise reduction in datasets
Apriori	Finds frequent item sets and derives association rules (e.g., "If X, then Y")	Association Rules	Retail product recommendations, cross-selling strategies

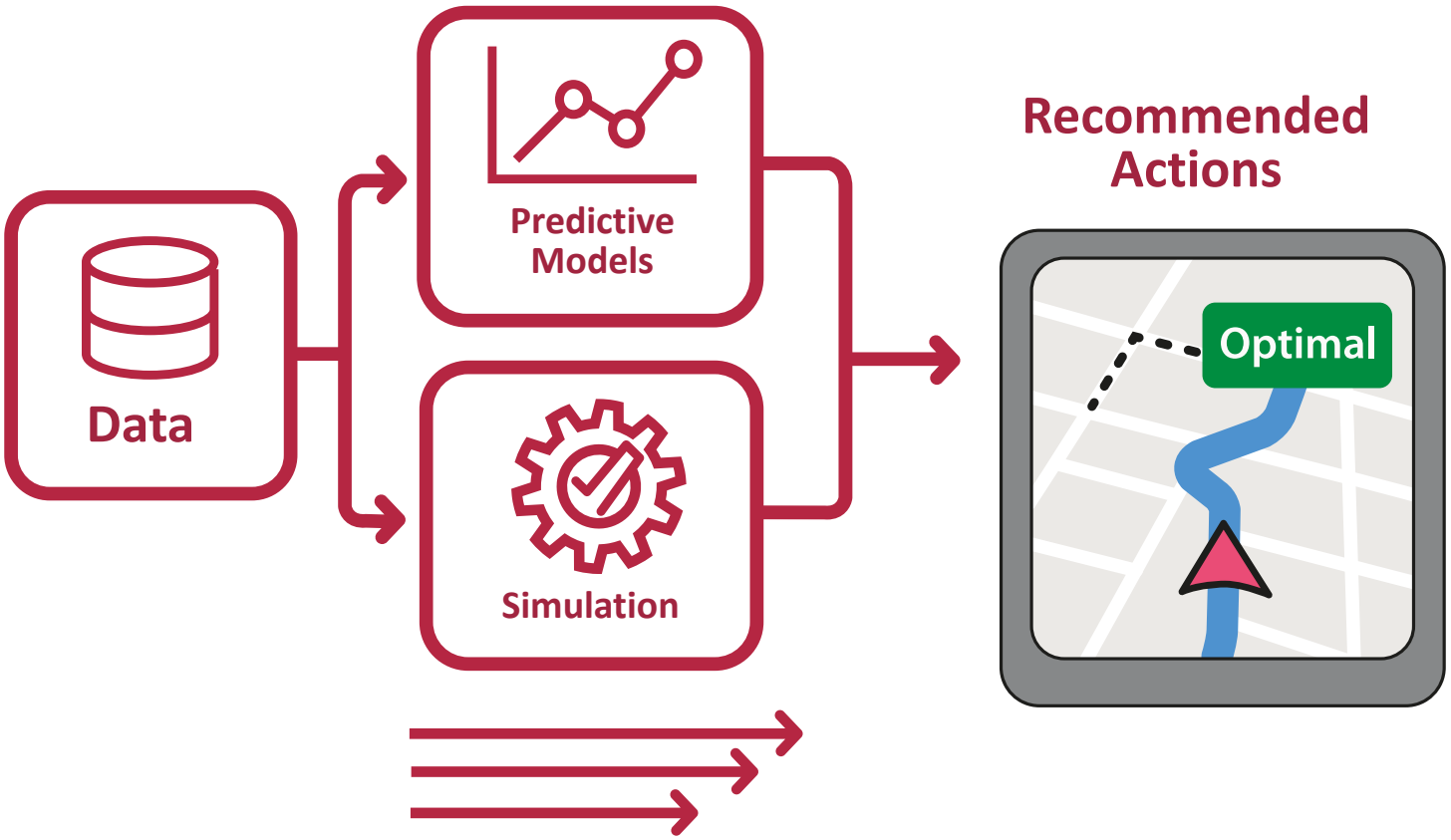
Data Mining as a Hurdle Race

- **High Dimensionality**
- **Interpretability**
- **Dynamic Data**
- **Algorithm Selection**

Data Mining helps us predict the future. But prediction alone doesn't make decisions. To choose the best action among many options, we need **Operations Research**.

What is Operations Research (OR)?

OR optimizes decisions under constraints



Simulations: Risk-Free Experimentation

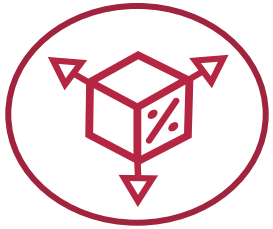
Unlike static reports, simulations show a range of possible futures

Rules and Assumptions

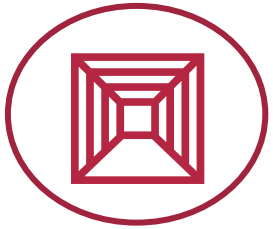
- Deductive reasoning
- Empirical knowledge
- Assumptions

Data

Computer Model of System



Simulated Scenarios



Tests "what-if" scenarios by modeling possible outcomes under different conditions

Predictions

Most Used Simulation Techniques

Technique	Description	Key Applications	Tools
Monte Carlo Simulation	Uses random sampling to model probabilities of outcomes in uncertain systems	Financial risk analysis, project scheduling	@RISK, Python
Discrete Event Simulation (DES)	Models system operations as sequences of events over time	Logistics optimization, hospital workflows	AnyLogic, Simio, Arena
Agent-Based Modeling (ABM)	Simulates autonomous agents' interactions to study emergent system behaviours	Traffic flow, epidemic spread, market dynamics	NetLogo, Mesa (Python)
System Dynamics	Captures feedback loops and time delays in complex systems	Climate change models, supply chain resilience	Venzins, Stella
Finite Element Analysis (FEA)	Breaks structures into small elements to simulate physical stresses	Aerospace engineering, structural design	ANSYS, COMSOL
Molecular Dynamics	Simulates atomic/molecular motion using physics laws	Drug discovery, material science	GROMACS, LAMMPS
Computational Fluid Dynamics (CFD)	Simulates fluid flow and heat transfer	Aerodynamics, HVAC system design	OpenFOAM, ANSYS Fluent

Example of a Monte Carlo Simulation

Estimating Profit for a New Product Launch
Profit=(Unit Price−Unit Cost) × Sales Volume

Identify the **uncertain variables** that impact profit:

- **Sales volume** (units sold): between 5,000 and 10,000 units
- **Unit price**: between $90 and $110
- **Unit cost**: between $60 and $80

Using probability distributions:

- Sales volume: randomly sample from a uniform distribution between 5,000 and 10,000
- Unit price: sample from a normal distribution with mean $100
- Unit cost: sample from a normal distribution with mean $70,

Do this for **10,000 iterations**.

For each iteration:

- Plug the randomly sampled values into the profit formula
- Store the resulting profit

Example iteration:

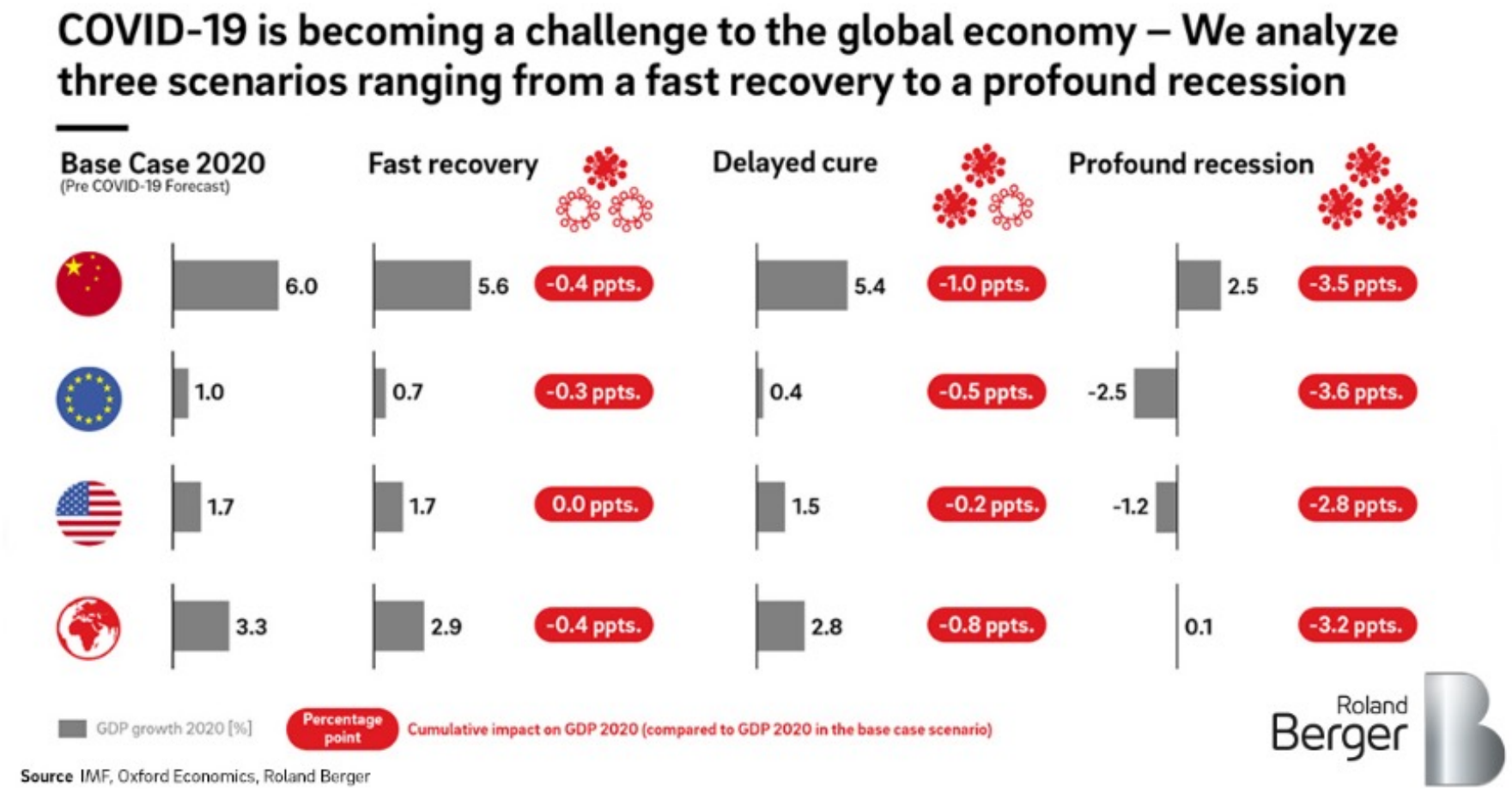
- Unit Price = $104.20 / Unit Cost = $72.50 / Sales Volume = 7,800
- Profit = ($104.20 - $72.50) × 7,800 ≈ **$246,360**

Calculate the **mean**, **median**, **standard deviation**, and **percentiles** (e.g., 5th and 95th) of all simulated profits

Visualize results with a histogram to see the probability distribution of profit

Scenarios for Pandemic Crises

During COVID-19, prescriptive models helped policymakers balance health and economic outcomes - in real time and under extreme uncertainty



Advantages and Limitations of Prescriptive Analytics

Advantages

Actionable Recommendations
Goes beyond "what might happen" to suggest **what you should do**

Optimized Decision-Making
Helps identify the **best possible outcomes** under constraints, saving time, cost, or resources

Scenario Testing & What-if Analysis
Allows businesses to simulate different actions and **compare potential outcomes**

Increased Efficiency & Automation
Enables **automated decision systems**, reducing human error and speeding up processes

Strategic Advantage
When implemented well, it offers a **competitive edge** by aligning decisions with goals and predictive insights

Personalization at Scale
In marketing or customer service, prescriptive analytics can **tailor actions** to individuals (e.g., product recommendations, pricing)

Limitations

Complexity of Models
Requires advanced tools, skills, and often **cross-disciplinary expertise** (e.g., data science + operations research)

Data Dependency
Garbage in, garbage out — inaccurate or incomplete data can lead to poor recommendations

High Implementation Costs
Involves significant investment in **technology, infrastructure, and talent**

Model Interpretability
Optimization or AI models can be **black boxes**, making it hard for stakeholders to trust or understand the logic

Over-reliance on Automation
Blindly following model recommendations without domain context can be risky

Ethical & Bias Risks
Algorithms may unintentionally embed bias, especially in **sensitive decision-making** areas (e.g., hiring, loans, criminal justice)

Examples of Prescriptive Analytics in Action

Venture Capital: Investment Decisions

Investment decisions can be strengthened by algorithms that weigh risks and recommend whether to invest.

One example in the venture capital space is an experiment, as explained in the Harvard Business Review, that tested the effectiveness of an algorithm's decisions about which startups to invest in as compared to angel investors' decisions.

The findings were nuanced. The algorithm outperformed angel investors who were less experienced at investing and less skilled at controlling their cognitive biases; however, angel investors outperformed the algorithm when they were experienced in investing and able to control their cognitive biases.

This experiment sheds light on the complementary role prescriptive analytics must play in making decisions and its potential to aid decision-making when experience isn't present and cognitive biases need flagging. An algorithm is only as unbiased as the data it is trained with, so human judgment is required whether using an algorithm or not.

Sales: Lead Scoring

Prescriptive analytics plays a prominent role in sales through lead scoring, also called lead ranking. Lead scoring is the process of assigning a point value to various actions along the sales funnel, enabling you, or an algorithm, to rank leads based on how likely they are to convert into customers.

Actions you can assign value include:

- Page views
- Email interactions
- Site searches
- Content engagement, such as attending webinars, downloading e-books, or watching videos

When assigning each action a point value, assign the highest number of points to those that imply purchase intent (for instance, visiting a product page) and negative points to those that reveal non-purchase intent (for instance, viewing job postings on your site).

This can help prioritize outreach to leads most likely to convert into customers, potentially saving your organization time and money.

Content Curation: Algorithmic Recommendations

If you've ever scrolled through a social media platform or dating app, you've likely experienced prescriptive analytics firsthand through algorithmic content recommendations.

Businesses' algorithms gather data based on your engagement history on their platforms (and potentially others, too). The combinations of your previous behaviors can act as triggers for an algorithm to release a specific recommendation. For instance, if you regularly watch shoe review videos on YouTube, the platform's algorithm will likely analyze that data and recommend you watch more of the same type of video or similar content you may find interesting.

On social media, TikTok's "For You" feed is one example of prescriptive analytics in action. The company's website explains that a user's interactions on the app, much like lead scoring in sales, are weighted based on indication of interest. "For example," TikTok's website says, "if you finish a video, that's a strong indicator that you're interested. Videos are then ranked to determine how likely you'll be interested in each video and delivered to each unique 'For You' feed."

This prescriptive analytics use case can make for higher customer engagement rates, increased customer satisfaction, and the potential to retarget customers with ads based on their behavioral history.

Source: online.hbs.edu/blog/post/prescriptive-analytics

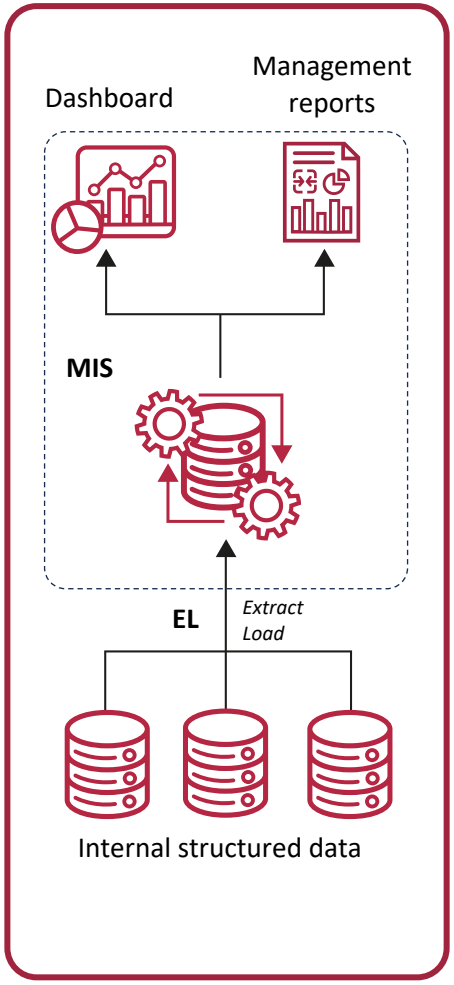
Prescriptive Analytics as Weather Forecast?

Prescriptive analytics does not guarantee perfect decisions.
It provides probabilistic guidance under uncertainty.

From Descriptive to Prescriptive

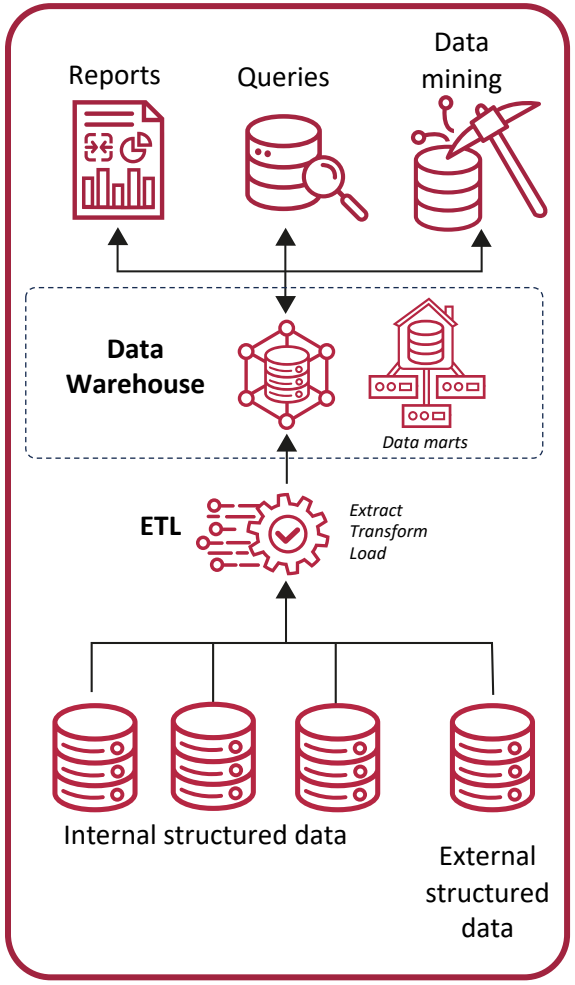
Descriptive analytics

What happened?



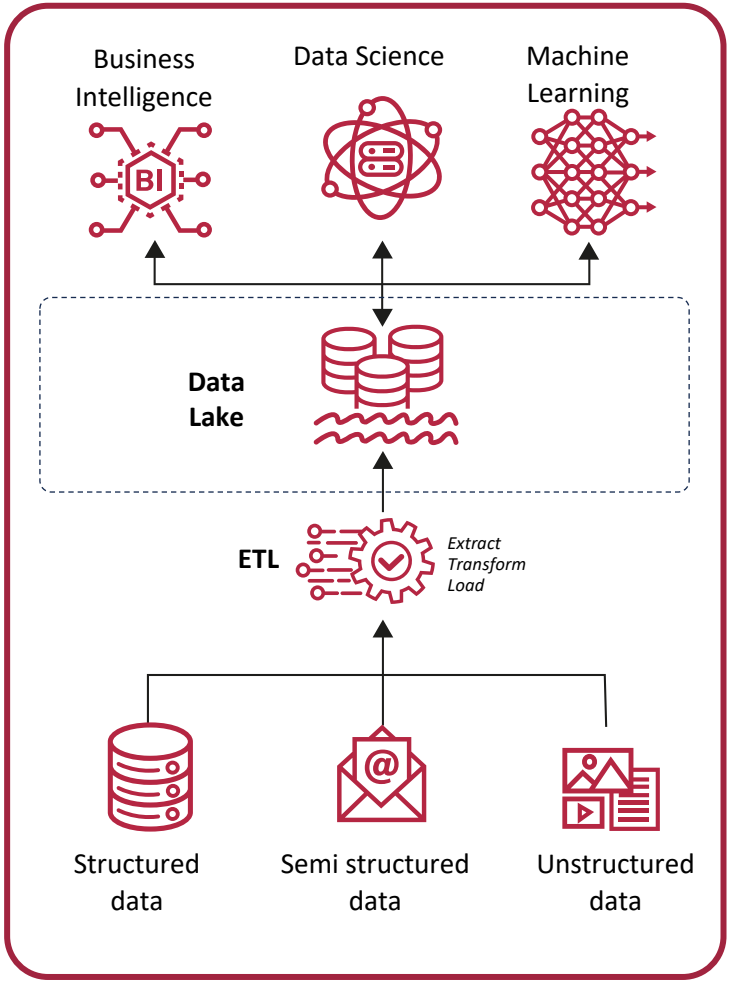
Diagnostic analytics

Why did it happen?



Predictive + Prescriptive analytics

What will happen and what should we do?



Data Visualization

We have analyzed data. Now let us present to make decisions happen.

The Psychology of Colors: Green - Yellow - Red

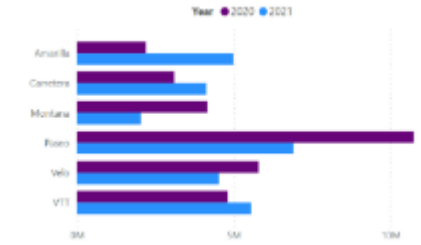
Color choice is cultural and context-dependent

Flag of Mali ☺

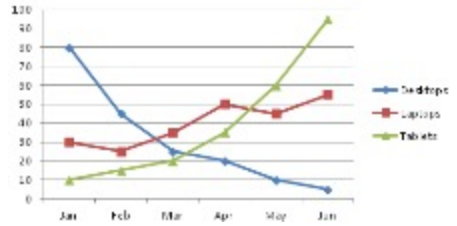
Types of Data Visualization

Bar chart

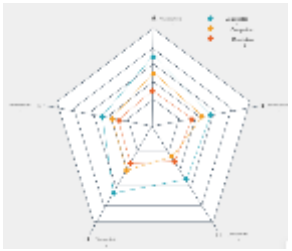
Column Chart



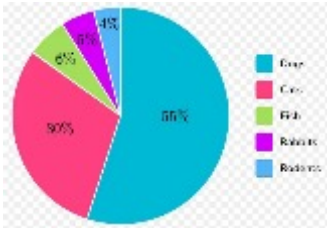
Line Chart



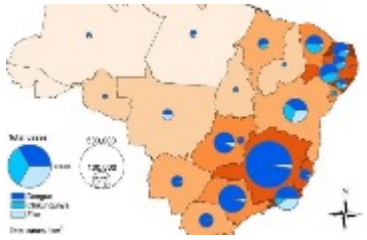
Radar Chart



Pie Chart



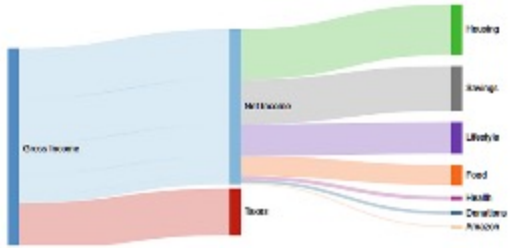
Bubble Map



Area Chart

Candlestick Chart

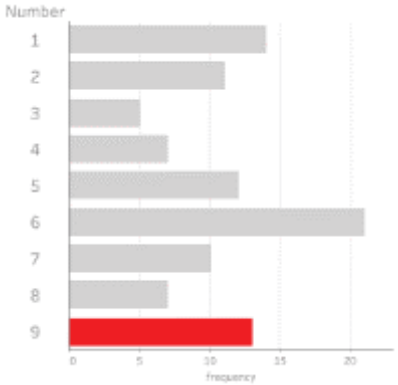
Sankey Diagram

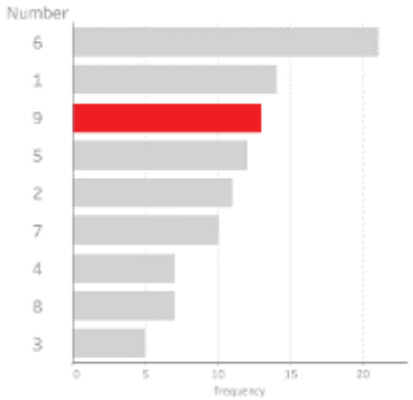


How to Visualize Data Effectively?

How many “9”?

2 2 5 6 7 1 1 6 9 1
9 1 7 5 5 5 6 2 5 9
4 5 2 9 6 9 7 6 4 6
8 1 5 7 8 5 6 6 6 7
7 2 3 6 8 9 1 7 9 1
3 8 6 8 4 5 6 9 4 5
4 9 9 2 3 7 1 9 1 2
3 7 8 1 6 1 5 6 1 6
5 6 6 8 6 6 9 1 2 6
3 2 4 2 6 9 4 2 7 1



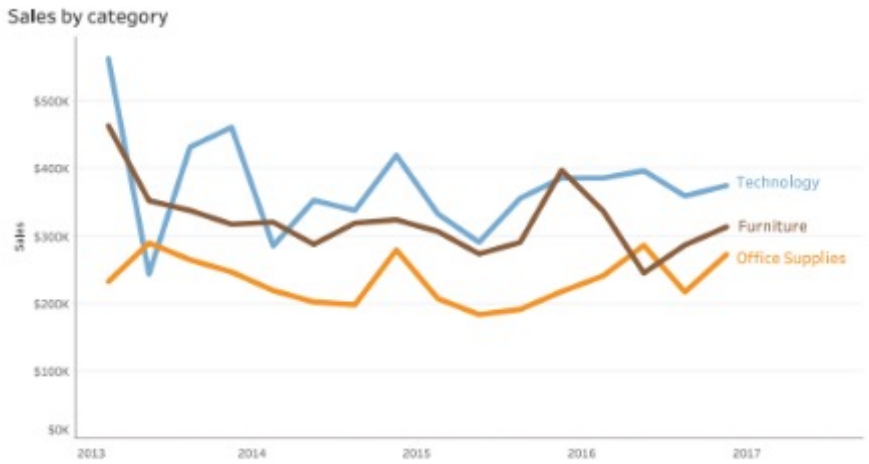


Source: S. Wexler, J. Shaffer, and A. Cotgreave, The Big Book of Dashboards

How to Interpret Figures?

Sales Statistics

Category	2013 Q1	2013 Q2	2013 Q3	2013 Q4	2014 Q1	2014 Q2	2014 Q3	2014 Q4
Furniture	$463,988	$352,779	$338,169	$317,735	$320,875	$287,934	$319,537	$324,319
Office Supplies	$232,558	$290,055	$265,083	$246,946	$219,514	$202,412	$198,268	$279,679
Technology	$563,866	$244,045	$432,299	$461,616	$285,527	$353,237	$338,360	$420,018
Category	**2015 Q1**	**2015 Q2**	**2015 Q3**	**2015 Q4**	**2016 Q1**	**2016 Q2**	**2016 Q3**	**2016 Q4**
Furniture	$307,028	$273,836	$290,886	$397,912	$337,299	$245,445	$286,972	$313,878
Office Supplies	$207,363	$183,631	$191,405	$217,950	$241,281	$286,548	$217,198	$272,870
Technology	$333,002	$291,116	$356,243	$386,445	$386,387	$397,201	$359,656	$375,229



Source: S. Wexler, J. Shaffer, and A. Cotgreave, The Big Book of Dashboards

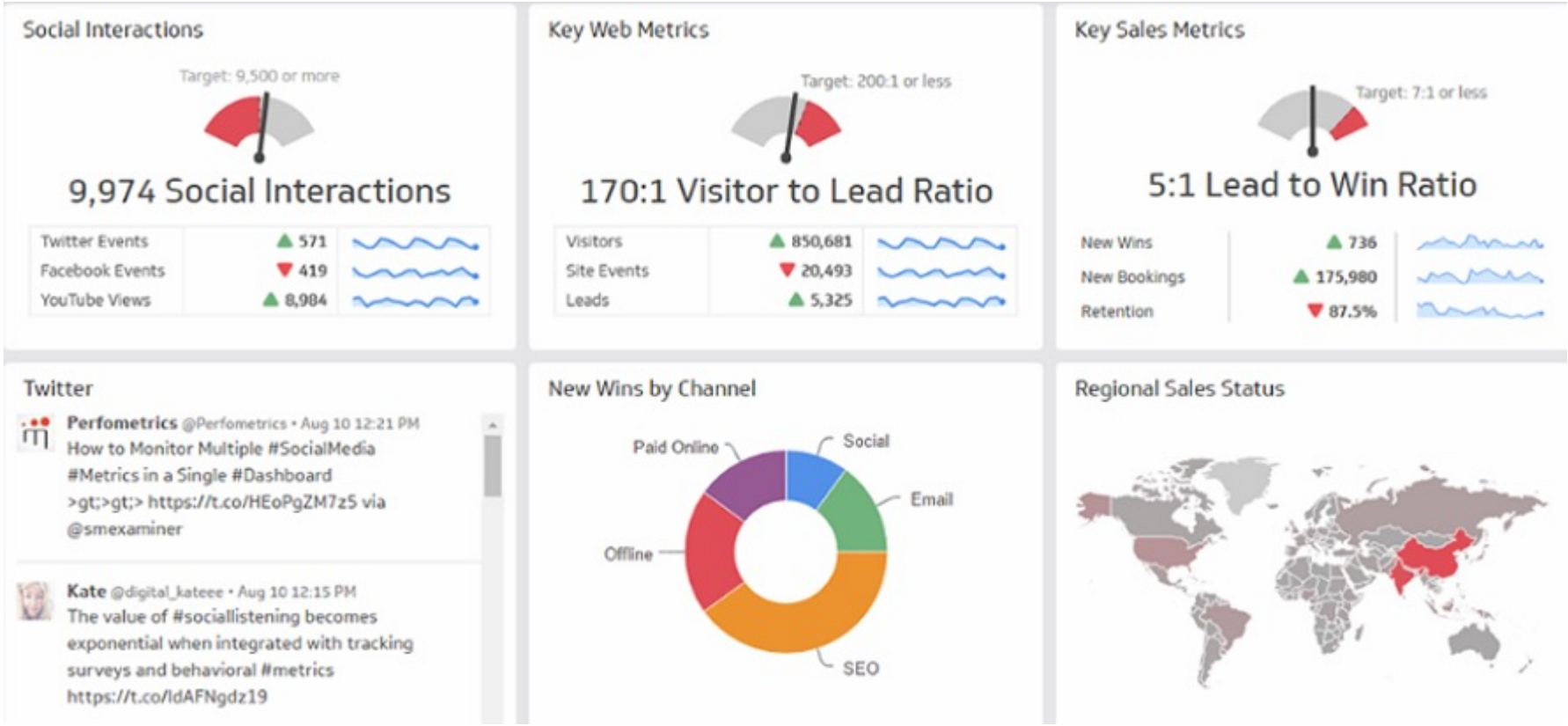
References for Best Data Visualization

- History of Bruce Springsteen
- Apollo
- Keuzestress
- Marvel Cinematic Universe
- The Many Moons of Jupiter
- Newsmap
- The Big Mac Index
- CF Weather Charts
- Galaxy of Covers
- Red Bull Party Visualization
- Figures in the Sky
- The Women of Data Viz

Source: www.xenonstack.com/insights/data-visualization

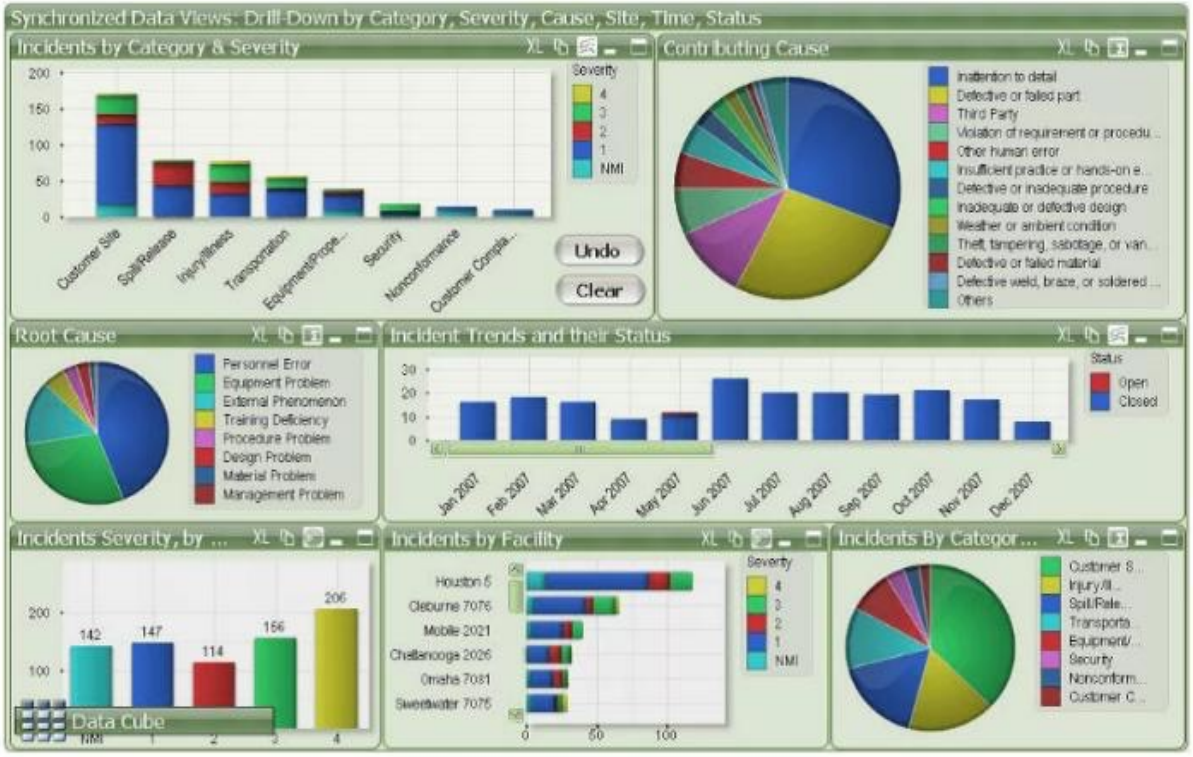
Dashboard

A dashboard is a visual display of data used to monitor conditions and/or facilitate understanding



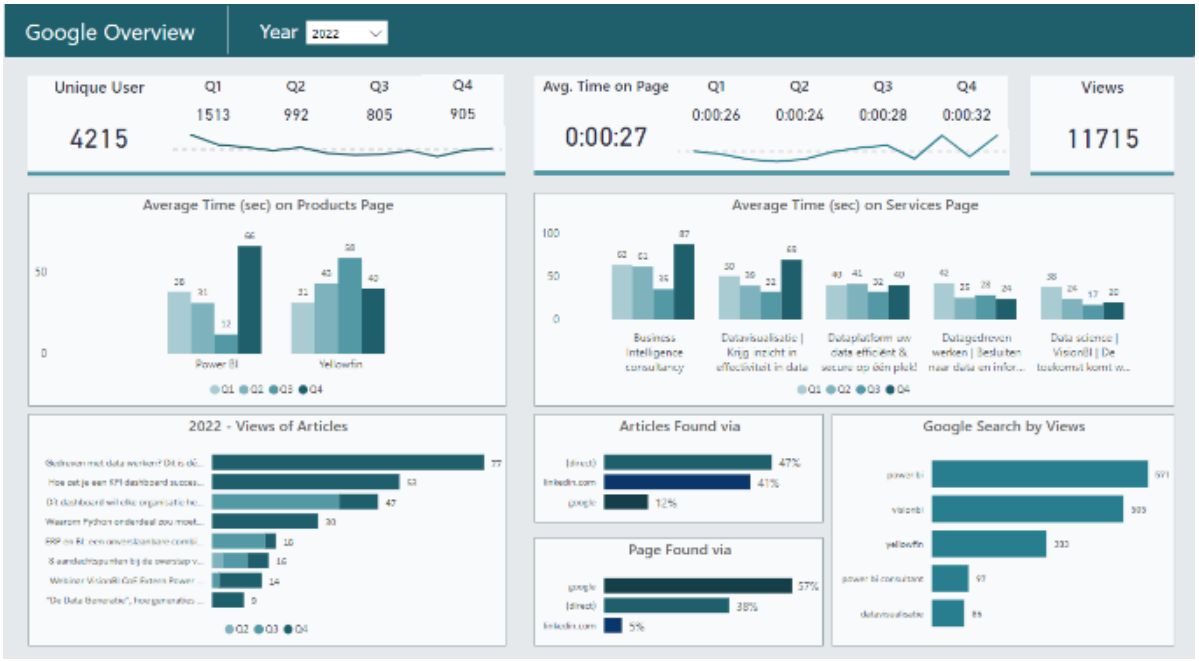
Source: S. Wexler, J. Shaffer, and A. Cotgreave, The Big Book of Dashboards

Example of a Bad Dashboard



- Overloaded
- Too many colours
- Difficult to read, as the reader gets overwhelmed
- Wrong chart type, e.g. pie charts have too many slices

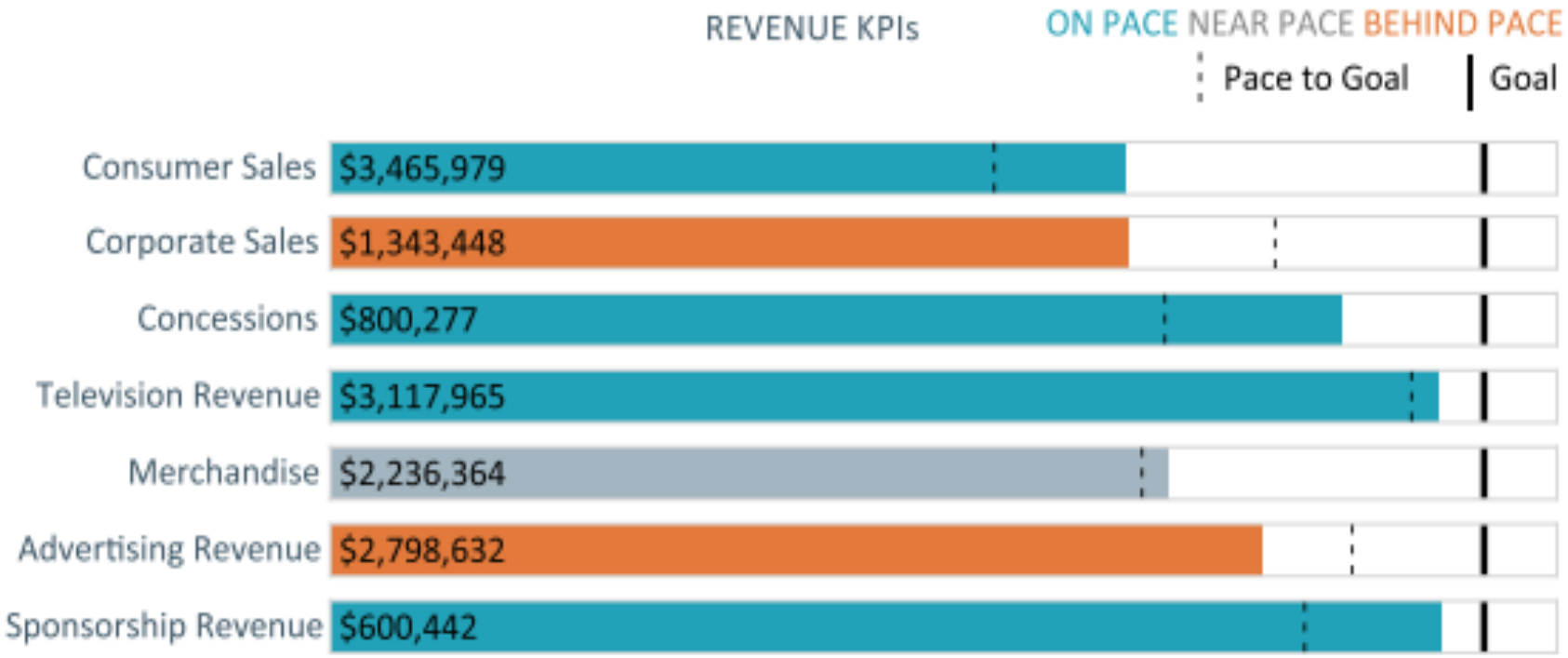
Example of a Good Dashboard



- Well structured
- Corporate colours
- Good amount of KPI's and graphs
- Reader gets a good starting point for the analysis

Executive Summary at a Glance

Pace chart



Source: S. Wexler, J. Shaffer, and A. Cotgreave, The Big Book of Dashboards

Run Your Company in Your Pocket

Source: www.designmind.com/blog/business-intelligence/power-bi-to-enhance-data-exploration

Notifications and Alerts

Notification

Informational messages about system events, updates, or status changes that may not require immediate action

Rule: If a new high-value deal (>$50k) is closed

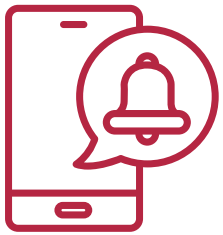
```
WHEN opportunity_status changes to "Closed Won"
AND opportunity_amount > 50000
THEN trigger "Major Deal Closed" notification
```

Alert

Urgent messages about critical conditions that require immediate attention

Rule: If monthly expenses exceed budget by more than 15%

```
WHEN SUM(expenses) > (budget_amount * 1.15)
AND time_period = current_month
THEN trigger "Budget Overrun" alert
```



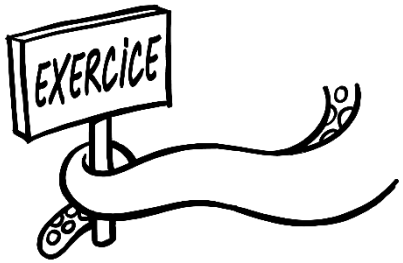
New major deal closed: Acme Corp - 72,000
Salesrep: Sarah Johnson
Q2 total sales = 420,000.

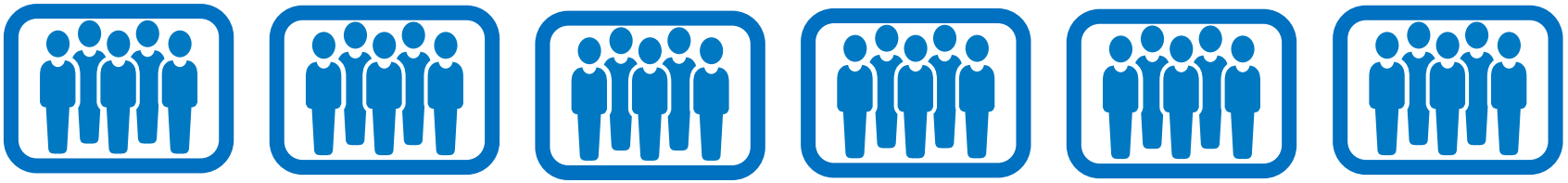
CRITICAL: Department expenses have exceeded budget by 18% this month
Current spend: 115,000 vs Budget: 97,500

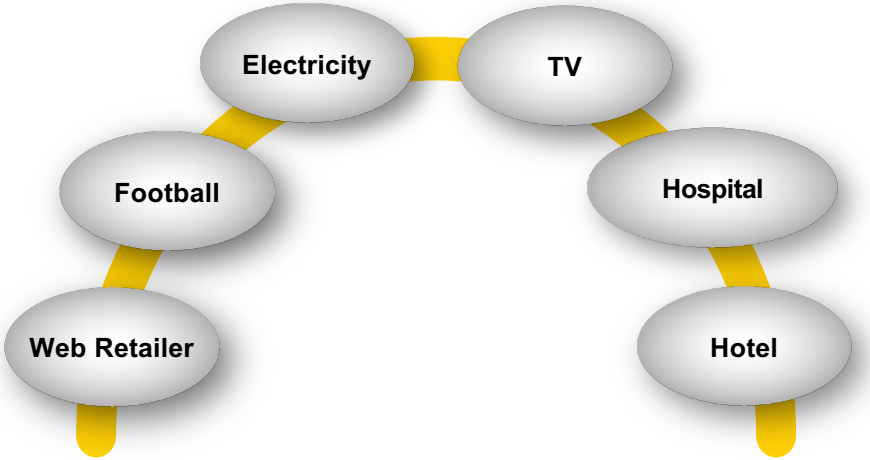
Design a Dashboard

Give a representation of 5 indicators in a dashboard

- A Web Retailer
- Champion Leagues
- Electricity Suppliers
- TV channel
- Hospital
- Hotel Chain







Types of Decision-Support Systems (DSS)

Visualization and dashboards are key components of decision-support systems

Type	Characteristics	Application Areas	Examples
Data-Driven DSS	• Focuses on querying and analysing large datasets • Uses OLAP, data mining	Retail, marketing, operations	• Analysing sales trends by region • Customer segmentation
Model-Driven DSS	• Relies on mathematical models (simulation, optimization) • "What-if" analysis	Finance, logistics, engineering	• Portfolio risk analysis • Supply chain optimization
Knowledge-Driven DSS	• Uses AI, rule-based systems, or expert systems • Provides recommendations	Healthcare, diagnostics, legal	• Medical diagnosis support • Legal case evaluation
Document-Driven DSS	• Manages unstructured data (text, reports, emails) • NLP-powered search	Research, compliance, knowledge management	• Searching legal precedents • Compliance document analysis
Collaborative DSS	• Supports group decision-making • Includes voting, shared workspaces	Corporate strategy, project management	• Boardroom decision-making • Team-based project planning
Web-Based DSS	• Cloud-hosted, accessible via browsers • Real-time data integration	E-commerce, remote teams, global enterprises	• Dynamic pricing for online retailers • Remote team budgeting tools

NLP: Natural Language Processing

Executive Information System (EIS) versus Executive Support System (ESS)

Aspect	EIS	ESS
Definition	Simplified system for monitoring KPIs and trends via dashboards	Interactive system for strategic decision-making, often with simulations
Primary Focus	Historical/Current: Tracking performance metrics	Forward-looking: Scenario modelling, "what-if" analysis
Data Sources	Primarily internal (ERP, CRM, data warehouses)	Internal + external (market data, economic indicators)
User Interaction	Low (pre-defined reports, static dashboards)	High (drill-down, ad-hoc queries, predictive tools)
Key Features	• Summarized reports • Visualization (charts, heatmaps) • Alerts	• Predictive analytics • External data integration • AI-driven insights
Applications	• Financial performance tracking • Operational oversight	• Mergers & Acquisitions • Long-term strategy • Risk management

Gartner Magic Quadrant for Analytics and Business Intelligence Platforms

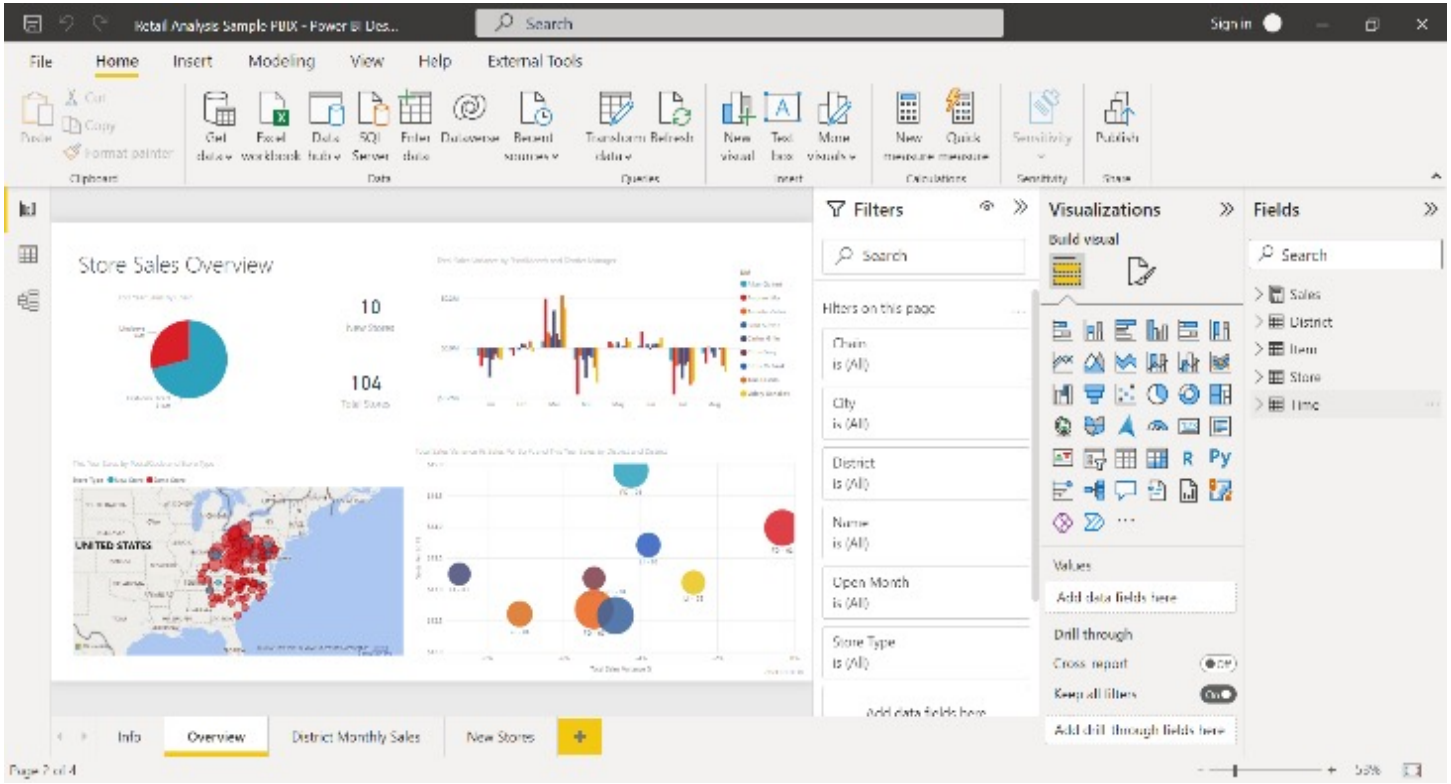
BI tools in practice

Source: Gartner

Example: Power BI Desktop (Microsoft)

The most common uses for Power BI Desktop are :

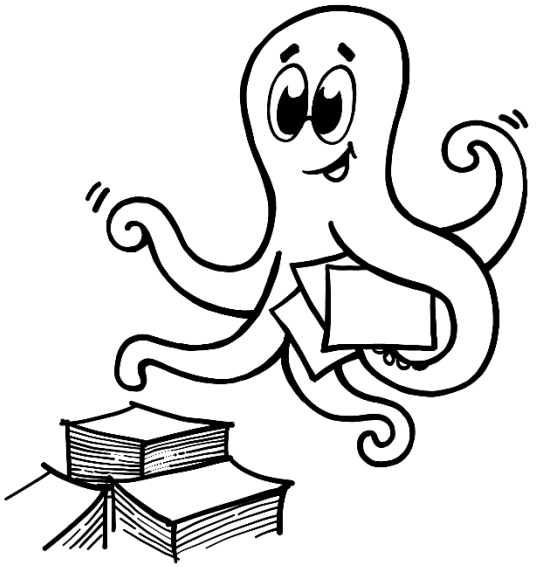
- Transform and clean data to create a data model
- Create visuals, such as charts or graphs that provide visual representations of the data
- Create reports that are collections of visuals on one or more report pages
- Share reports with others



Source: learn.microsoft.com/en-us/power-bi/fundamentals/desktop-what-is-desktop

Take away and must know!

- To distinguish between legal, operational, and managerial reporting
- To translate business questions into data problems
- To differentiate between MIS, Data Warehouses and Data Lakes
- To explain the differences between BI, DS and OR
- To critically evaluate data visualizations
- To design effective dashboards for decision-making



Literature

- K. C. Laudon, J.P. Laudon (2012) Management Information Systems MANAGING THE DIGITAL FIRM, Prentice Hall
- R. D. Peng & E. Matsui (2015) The Art of Data Science - A Guide for Anyone Who Works with Data, Skybrude Consulting, LLC
- S. Wexler, J. Shaffer, and A. Cotgreave (2017) The Big Book of Dashboards, John Wiley & Sons, Inc., Hoboken, New Jersey

KNOWLEDGE

www. know-ledge.ch