Big Data and Its Value in Decision-Making

For both end user and organization

What is Big Data?

Big data refers to datasets that are so huge, fast moving and different that traditional data management tools cannot effectively capture, process ,or analyze them. These data sets often combine structured data, semi structured data and unstructured data. (Google Cloud, n.d.)

Big Data is defined by the 5V's:

- Volume – refers to the massive scale of data generated every day (Robinson & Gillis, 2023).
- Velocity – describes the rapid speed at which data is produced and processed (Zitter, 2024).
- Variety – highlights the wide range of data types and formats, both structured and unstructured (GeeksforGeeks, 2025).
- Veracity – relates to the reliability and accuracy of the data being used (Robinson & Gillis, 2023).
- Value – represents the meaningful insights and benefits organizations gain through data analysis (SAS, 2023.; Salesforce, 2024).

The value of data for decision making

For Organizations

- Strategic Advantage: Big Data analytics show patterns, correlations, and trends that enable evidence-based decision-making (Salesforce, 2024).
- Operational Efficiency: Organizations use data insights to optimize supply chains, manage risks, and reduce costs (DashDevs, 2024).
- Innovation & Growth: Data fosters new business models, product development, and customer engagement strategies (Salesforce, 2024

For End Users

- Personalized Services: Retail personals and online platforms use data to recommend products and tailor customer experiences (Salesforce, 2024).
- Improved Quality of Life: In healthcare, Big Data supports predictive analytics for early diagnosis and better treatment plans (SAS, 2023).
- Trust & Transparency: When data is used responsibly, it builds user confidence in services and decision-making processes (Deloitte, 2024).

Why it Matters?

The Big Data enables an organization to transform its intuitive decisions and the strategies to become data-driven thus offering flexibility in evolving markets. To people, it is a sign of more customized, efficient and reliable services. The growing capacity to process data nearly in real-time is also changing industries, where instant insights have become a competitive requirement now (Deloitte, 2024).

Real World Examples

- Healthcare: Hospitals leverage Big Data for predictive modeling to improve patient outcomes.
- Finance: Banks use transaction analysis to detect fraud in real time.
- Retail: Companies like Amazon use customer data to personalize shopping experiences

Data-Driven Decision-Making:

Evaluating the Advantages, Challenges, and Impact of Data-Driven Decision-Making" Corelating to selected data set

Advantages

1.Objective insights: The dataset allows legislators and businesses to base decisions on evidence instead of assumptions. For example, it clearly shows that higher education levels increase the likelihood of earning more than $50K. This can guide investment in training and education policies.
(Provost and Fawcett, 2013)

2.Identifying patterns: It highlights inconsistencies(e.g., gender or occupation-based income differences), which can inform diversity and inclusion initiatives in the workplace. (Marr, 2016)

3.Better resource allocation: If the data shows that certain occupations correlate with higher or lower incomes, job training programs can be designed accordingly

4.Better resource allocation: If the data shows that certain occupations correlate with higher or lower incomes, job training programs can be designed accordingly
(Brynjolfsson and McElheran, 2016)

Challenges

1.Data bias: If the dataset over-represents certain groups (e.g., men vs women), results may not generalise fairly. Biased data can lead to biased decisions.
(O'Neil, 2016)

2.Correlation vs causation: Just because higher education is linked to higher income doesn't mean education alone causes it. Other factors (social background, geography) may be missing from the dataset.

3.Data quality & completeness: Missing values, outdated records, or self-reported data can reduce reliability.
(Kitchin, 2014)

4.Ethical issues: Using demographic data (like gender, race, marital status) for decision-making raises privacy and fairness concerns.
(Zwitter, 2014)

Impact on users

Data-driven decision-making can improve fairness and transparency by identifying links between education, income, and opportunities (Provost and Fawcett, 2013). This helps shape policies that support equality and access to resources. However, misuse of sensitive data can threaten privacy and reinforce inequalities if bias is present (O'Neil, 2016; Zwitter, 2014).

Zuhry Jaldeen

Impact on Organisations

For organizations, data supports more efficient strategies, improved workforce planning, and stronger competitiveness by reducing uncertainty (Brynjolfsson and McElheran, 2016). At the same time, reliance on incomplete or misinterpreted data can lead to flawed decisions and damage trust, highlighting the need for ethical safeguards (Kitchin, 2014).

Zuhry Jaldeen

Data-Driven Decision-Making:

Statistical and Graphical Techniques

Statistical Techniques

- Regression Analysis – Widely used to identify relationships between variables, such as predicting customer spending from income levels (James et al., 2013).
- Clustering techniques (such as k-means and hierarchical clustering) are used to categorize data points with shared characteristics. They are widely applied in areas like customer segmentation and detecting fraudulent activities (Tan et al., 2019).
- Classification (e.g., decision trees, logistic regression, random forests) – Predicts categorical outcomes, such as whether a customer will churn or stay (Hastie, Tibshirani and Friedman, 2009).
- Time-Series Analysis – Detects trends and patterns over time, heavily applied in finance, supply chain, and demand forecasting (Chatfield, 2016).

Graphical & Visualization Techniques

- Dashboards & Interactive Visuals (Power BI, Tableau) – Present real-time insights for business decision-making (Marr, 2016).
- Heatmaps – Highlight intensity of values, often used in website analytics or customer behaviour tracking.
- Scatterplots & Boxplots – Show relationships and distributions, helping spot outliers and correlations.
- Geospatial Maps – Combine location with data, useful in logistics, marketing, and public health.
- Network Graphs – Visualise complex connections, such as social media interactions or supply chain networks (Kitchin, 2014).

Industry uses a mix of statistical methods to uncover patterns and predictive models, and visualization techniques to make insights accessible, interactive, and actionable.

Data-Driven Decision-Making:

Industry-Leading Tools for Big Data Analysis and Visualization: Review and Real-World Applications

Review of Tools & Software

Power BI (Microsoft)

- Business intelligence tool for interactive dashboards and real-time reporting, integrated with Microsoft Office and Azure.

Example: Coca-Cola uses Power BI to track global supply chain performance (Marr,

Zuhry Jaldeen

Tableau

- Popular for advanced, interactive data visualisation and storytelling.

Example: Charles Schwab uses Tableau to provide advisors with customised client dashboards (Tableau, 2023).

Zuhry Jaldeen

Python (Pandas, Matplotlib, Plotly, Seaborn, scikit-learn)

- Open-source programming language widely applied in machine learning, automation, and analytics.

Example: Netflix leverages Python for recommendation systems and user behaviour analysis (Marr, 2016).

Zuhry Jaldeen

R

- Specialised for statistics and predictive modelling, often used in healthcare and academia (James et al., 2013).

Zuhry Jaldeen

Google BigQuery

- Cloud-based big data warehouse optimised for fast SQL queries across massive datasets.

Example: Spotify uses BigQuery to analyse streaming patterns and improve user experience (Marr, 2016).

Zuhry Jaldeen

SAS

- Enterprise-grade analytics platform with strong compliance features.

Example: HSBC employs SAS for fraud detection and risk modelling (Marr, 2016; SAS, 2023).

Zuhry Jaldeen

Amazon Web Services (AWS Analytics)

Includes services like Redshift and QuickSight for cloud-based analytics and visualisation.

Example: Airbnb uses AWS Redshift to process large volumes of booking and user data (Marr, 2016).

Zuhry Jaldeen

Comparison & Real-Life Impact

- **BI Tools (Power BI, Tableau, Qlik):** Dashboards for executives, fast KPI monitoring.
- **Programming (Python, R):** Flexible analytics & ML for researchers, healthcare, tech.
- **Enterprise (SAS):** Compliance-focused, vital in finance & government.
- **Big Data (Hadoop, Spark, BigQuery, AWS):** Large-scale processing for e-commerce & cloud.

Overall: Organisations use a hybrid ecosystem BI for decision-making, programming for deep analysis, and big data platforms for scale. This integration boosts accuracy, speed, and innovation.