

ALABI Francis Toluwani

alabfrank@gmail.com

Data Analysis 2 Assignment

1. Describe the other types of databases with examples.

Beyond relational databases, several other database types exist:

a. NoSQL Databases: These are non-relational databases designed for flexibility and scalability. They handle unstructured, semi-structured, or rapidly changing data.

Examples: MongoDB, Cassandra, CouchDB, Firebase, DynamoDB.

b. Graph Databases: They store data in nodes and edges, making them ideal for relationship-heavy data.

Examples: Neo4j, Amazon Neptune, ArangoDB.

c. Time-Series Databases: Optimized for time-stamped or time-series data such as IoT metrics, logs, and financial data.

Examples: InfluxDB, TimescaleDB, Prometheus.

d. Columnar Databases: Store data by columns instead of rows, making them excellent for analytical queries.

Examples: Apache Cassandra, Amazon Redshift, Google BigQuery, ClickHouse.

e. Object-Oriented Databases: Store data as objects (like in object-oriented programming).

Examples: db4o, ObjectDB.

2. Describe the different methods of collecting data and tools used.

a. Surveys and Questionnaires: Used to collect first-party data from individuals.

Tools: Google Forms, SurveyMonkey, Typeform.

b. Interviews: One-on-one or group discussions used to gather qualitative insights.

Tools: Zoom, Microsoft Teams (recordings for analysis).

c. Observations: Watching subjects in their natural setting to collect behavioral data.

Tools: Cameras, note-taking apps, mobile devices.

d. Web Scraping: Extracting data from websites.

Tools: Python (BeautifulSoup, Scrapy), Octoparse.

e. Sensors & IoT Devices: Automatically collect environmental or machine data.

Tools: Smart sensors, IoT platforms.

f. Transactional Systems: Collect data from day-to-day business operations.

Tools: CRM systems, ERP systems, POS systems.

g. Databases and Open Data Repositories: Using publicly available or organizational datasets.

Tools: Kaggle, Google Public Datasets, Government portals.

3. Difference between Data Analysis, Data Science, and Data Engineering

Data Analysis

Focus: Interpreting existing data to uncover trends and insights.

Tasks: Cleaning, analyzing, visualizing, reporting.

Tools: Excel, Power BI, Tableau, SQL.

Output: Dashboards, reports, insights.

Data Science

Focus: Advanced analysis, modeling, and predictions using algorithms.

Tasks: Machine learning, statistical modeling, experimentation.

Tools: Python, R, TensorFlow, scikit-learn.

Output: Predictive models, forecasting systems.

Data Engineering

Focus: Building and maintaining data systems and pipelines.

Tasks: Data architecture, ETL pipelines, database optimization.

Tools: SQL, Python, Hadoop, Spark, Airflow.

Output: Reliable data infrastructure for analytics and science.

Summary:

Data Analysts analyze and interpret data.

Data Scientists build machine learning models and use algorithms.

Data Engineers build the infrastructure that makes data accessible.

4. Differences and similarities between Data Analysis and Business Intelligence Analysis

Similarities

Both use data to support business decision-making.

Both involve data cleaning, visualization, and reporting.

Both rely on tools like SQL, Excel, and BI dashboards.

Differences between Data Analysis(DA) and Business Intelligence (BI) Analysis

1. DA focuses on exploring and interpreting data while BI focuses on reporting and monitoring business performance.
2. DA answers deeper analytical questions (why, how) while BI answers high-level performance questions (what happened).
3. DA looks at both structured and sometimes unstructured data while BI only deals with structured enterprise data.
4. DA often involves more detailed and exploratory analysis while BI involves creating dashboards, KPIs, scorecards.
5. DA has broader role across industries and problem types while BI has narrower scope; specific to business operations.

5. Outline some applications of Data Analysis in different industries.

- a. Healthcare: Predicting disease outbreaks, Improving patient diagnosis and treatment, Hospital resource management.
- b. Finance: Fraud detection, Credit risk scoring, Stock market forecasting.
- c. Retail/E-commerce: Customer segmentation, Demand forecasting, Price optimization.
- d. Manufacturing: Predictive maintenance, Supply chain optimization, Quality control.

e. Marketing: Campaign performance measurement, Customer behavior analysis, Personalization and targeting.

f. Agriculture: Crop yield prediction, Weather pattern analysis, Soil monitoring.

g. Transportation & Logistics: Route optimization, Fleet management, Traffic analysis.

h. Government/Public Sector: Population analysis, Urban planning, Policy evaluation.