

Introduction to Data Analytics

PROFESSIONAL CERTIFICATE IBM
Data Analyst

Agenda

What is data analysis

Data analysis process

Data sources & structure

Modern data ecosystem

Data analysis primary types

Data professionals

A Day in the life of a data analyst

Data analyst required skills

What is Data Analysis

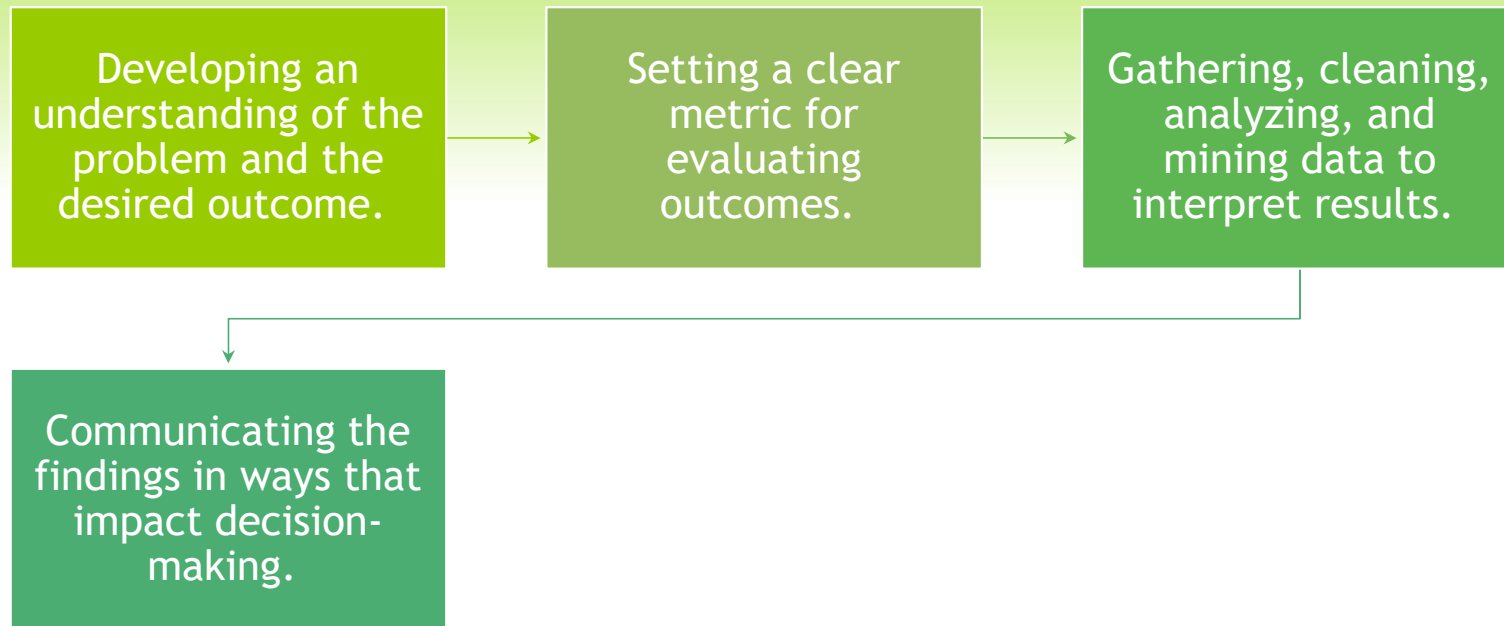
Data analysis helps businesses to

- Understand their past performance and informs their decision-making for future actions

- looking for patterns in financial transactions to detect fraud, using recommendation

- Engines to drive conversion, mining, social media posts for customer voice or brands personalizing their offers based on customer behavior analysis, business leaders realized that

Data analysis process



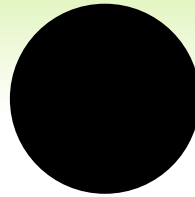
Data Sources & Structure



HEALTH



BIOMETRICS



HOUSEHOLD
DATA IN THE
CASE OF IOT
DEVICES



STRUCTURED



UNSTRUCTURED

Modern Data Ecosystem



Data



Enterprise Data
Environment



End-users

Data Analyst Ecosystem

Gather
Data

Clean
Data

Mine
Data

Visualize
Data



Primary Types of Data Analysis



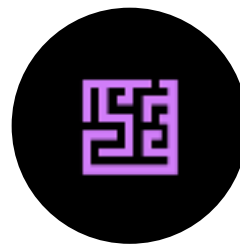
Descriptive Analytics
helps answer questions
about what happened
over a given period



Diagnostic analytics
helps answer the
question. Why did it
happen?



Predictive analytics
helps answer the
question, What will
happen next?



Prescriptive Analytics
helps answer the
question, What should
be done about it?

Types of Data Analysis Activity

- A company is interested in being able to forecast the weather using meteorological data and predicting the increase in water demand in a particular city or region. **Predictive Analysis**

An insurance company is interested in gathering more information on its customers such as their age, genders, and nationalities.

Types of Data Analysis Activity

- Netflix wants to continue to recommend new shows/movies to their users (actionable predictions) using data gathered on their choices, purchases, or website navigation based on their clicks.

An unusually high number of people have been admitted to the ER over a short span of time at a local hospital.

Data professionals

- Data engineers
- Data analysts

Data scientists

Business analysts

Business intelligence or BI analysts

A Day in the Life of a Data Analyst

Acquiring

Acquiring data that best serves the use case.

Preparing and analyzing data to understand what it represents.

Interpreting

Interpreting and effectively communicating the message to stakeholders

Ensuring

Ensuring that the process is documented for future reference and repeatability.

Data Analyst Required Skills



Technical skills: spreadsheets, statistical tools, visualization tools, programming and querying languages, and the ability to work with different types of data repositories and big data platforms.



Soft Skills: include the ability to work collaboratively, communicate effectively, tell a compelling story with data, and garner support and buy-in from stakeholders. Curiosity to explore different pathways

A Day in the Life of a Data Analyst Activity

What do you think a day in the life of a Data Analyst would look like in the following roles?

1. Data analyst

Data Scientist

Data Engineer

Business Analyst

Fraud Analyst

Data Types



Structured



Semi-structured



Unstructured

Structured

- Data that follows a rigid format and can be organized neatly into rows and columns is structured data.

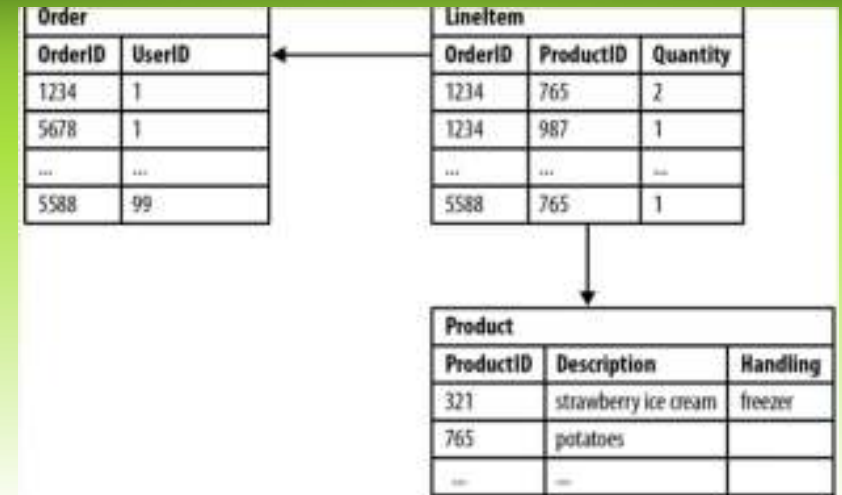
This is the data that you see typically in databases and spreadsheets, for example.

Dates

Phone numbers

ZIP codes

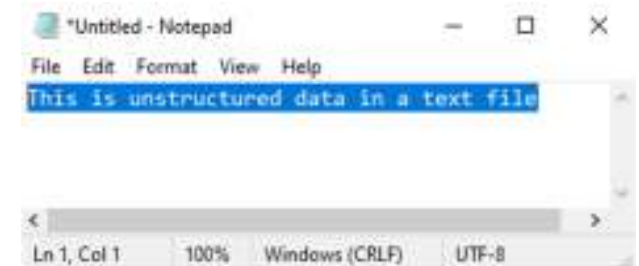
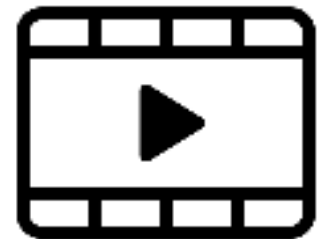
Customer names



A	B	C	D
an_id	name	city	commis
5001	James Hoog	New York	(
5002	Nail Knite	Paris	(
5005	Pit Alex	London	(
5006	Mc Lyon	Paris	(
5007	Paul Adam	Rome	(
5003	Lauson Hen	San Jose	(

Unstructured Data

- Complex, and mostly qualitative information that is impossible to reduce to rows and columns.



Semi-structured Data

- Semi-structured data is a mix of data that has consistent characteristics and data that doesn't conform to a rigid structure.

For example, emails.

An email has a mix of structured data, such as the name of the sender and recipient, but also has the contents of the email, which is unstructured data.



Sources of Data

Data sources have never been as dynamic and diverse as they are today

- Relational Database
- Non-Relational Database
- APIs
- Web Services
- Data Streams
- Social Platforms
- Sensor Devices

Data Repositories: Breakout Room Activity 10 min

A data repository is a general term used to refer to data that has been collected, organized, and isolated so that it can be used for business operations or mined for reporting and data analysis.

Databases

Data warehouses

Data marts

Data lakes

Big data stores.


Big Data Processing Tools

- The Big Data processing technologies provide ways to work with large sets of structured, semi-structured, and unstructured data so that value can be derived from big data.

Apache Hadoop

Apache Hive

Apache Spark

The background of the slide features a green gradient. On the left side, there are several thin, curved, light-colored lines that sweep across the upper portion of the image, creating a sense of motion or light rays. The rest of the background is a smooth, horizontal gradient of green shades.

Any questions