Introduction to Data Science

What is Data Science?

Data Science is the field that uses **statistics, computer science, and domain knowledge** to extract insights and knowledge from data.

In simple terms, **Data Science is about turning raw data into useful information**.

What is Data?

Data is just **information**.
It can be **numbers, words, images, or facts** that describe something.

Examples:

- Your name, age, and phone number → that's data
- The number of goals in a football match → data
- The temperature today → data
- A list of what people buy in a supermarket → data
- Customer information in a bank.

Show example **Dataset** http://coding.co.ke/files/bank.csv

Science is a way of **studying and understanding how things work**.

Data Science = Data (information) + Science (Studying and Solving).

So, Data Science is all about:

Using information to study real-world problems and find smart solutions.

Think of it like this:

- **Data** gives us the facts (like numbers, names, or trends - Information).
- **Science** helps us explore those facts, understand them, and solve problems(Solutions)

Real-Life Example: Data Science in a Supermarket

Problem:

A supermarket like **Naivas** or **Quickmart** wants to:

- Avoid running out of popular products (like bread, milk, biscuits, yoghurts)
- Stop wasting money on items that don't sell
- Understand what customers like to buy.

Definition Reminder:

Data Science = Data (Information) + Science (Studying and Solving)

Data (Information):

The supermarket collects data from:

- Sales data (what people buy and when)
- Loyalty cards (who is buying and how often)
- Inventory (what's in stock and what's not)
- Seasonal trends (e.g. more juice bought in hot months, more snacks bought during holidays etc)

Science (Studying and Solving):

Data scientists analyze this data to:

- Find which products are most popular and at what times
- Predict how much milk or bread will be needed tomorrow
- Recommend what to stock more of — or less of
- Suggest/Predict offers (e.g. bread + margarine, juice + cakes)
- Stock more snacks during holidays when children have closed school.

Solution:

- The supermarket stocks exactly what people need, reducing waste
- Customers get personalized offers based on their shopping habits
- Profits go up because customers are happier and stock moves faster

More efficient supermarket, less waste, and better customer experience — **thanks to Data Science!**

Real-Life Example: Data Science in a Hospital

Data Science = Data (Information) + Science (Studying and Solving)

Problem:

Hospitals face challenges like:

- Long patient wait times
- Shortage of medicine
- Misdiagnosis or late diagnosis
- High treatment costs
- Overcrowding in emergency rooms

Data (Information) Collected in a Hospital:

- Patient records (age, symptoms, test results, diagnosis)
- Appointment schedules
- Medicine usage and inventory
- Equipment usage (like X-rays, MRI machines)
- Recovery time per disease
- Health trends (e.g., flu cases rising in July)

Science (Studying and Solving):

Data scientists use this data to:

- Predict disease outbreaks
- Help doctors diagnose faster using AI and machine learning

- Track which medicines are used most, and when to restock
- Reduce patient wait times by optimizing doctor schedules
- Identify which treatments work best for which patients

Aga Khan University Hospital and KNH (Kenyatta National Hospital) have digital health systems where patient data is stored electronically. This helps doctors quickly access a patient's history, track treatment progress, and make data-based medical decisions.

Examples of How Data Science is Used in the EPL:

1. **Player Performance Analysis:**
    - Teams like **Manchester City** and **Liverpool** use GPS and camera data to track every player's movement.
    - They know how far a player runs, how fast, and where they perform best.
2. **Injury Prevention:**
    - Clubs collect fitness data during training.
    - If a player shows signs of muscle strain (e.g. muscle stress), they get rest before an injury happens.
3. **Tactical Planning:**
    - Coaches use data from past matches to predict how the opponent will play.
    - Example: “Chelsea tends to play from the wings” → Plan to block crosses.

Applications of Data Science

1. Banking and Finance

Used to detect fraud, assess credit scores, and predict loan defaults. It also helps in recommending financial products to customers.

---

2. Retail and E-commerce

Stores like **Amazon** or **Jumia** use data science to recommend products, manage inventory, and personalize shopping experiences.

---

3. Healthcare

Helps doctors predict diseases early, choose better treatments, and manage hospital resources effectively.

---

4. Transport and Logistics

Used to plan delivery routes (like for G4S or DHL), track vehicle performance, and reduce fuel costs.

---

5. Telecom

Companies like **Safaricom** and **Airtel** use it to reduce dropped calls, manage network traffic, and send personalized offers to users.

---

6. Education

Helps track student performance, improve teaching methods, and create customized learning paths.

---

7. Entertainment and Streaming

Platforms like **Netflix** or **YouTube** use it to recommend shows, understand viewer habits, and improve content delivery.

---

8. Agriculture

Used to predict weather, monitor soil health, and recommend the best time to plant or harvest.

---

9. Government and Public Services

Used in traffic management, predicting crime areas, managing census data, and improving public services.

---

10. Sports

Teams analyze player performance, plan tactics, and prevent injuries using data — like in the **EPL** or **NBA**.

---

11. Human Resources (HR)

Companies use data to hire the right people, track employee performance, and predict when someone might quit.

Data Science Pipeline

Here's the typical Data Science pipeline (from start to finish): (How its done)

1. Problem Understanding/Understanding Data

What are we trying to solve or improve?
Example: "Why are sales dropping at our supermarket?"

---

2. Data Collection/Data Acquisition

Gather the information (data) needed to understand the problem.
Example: Get sales records, customer feedback, and inventory data.

---

3. Data Cleaning

Fix errors, remove duplicates, and make sure the data is correct and complete.
Example: Remove empty rows, fix wrong dates, fill missing values.

---

4. Insights & Reporting

Look for patterns, trends, or surprises in the data, Create Graphs and Plots
Example: Discover that milk sales drop every Wednesday.

---

5. Decision Making / Action /Monitoring

Use the findings to make smart choices or solve the problem, Monitor progress.
Example: Supermarket decides to offer midweek discounts based on the data.

Check Pandas_Seaborn PDF.