Introduction to Data Science

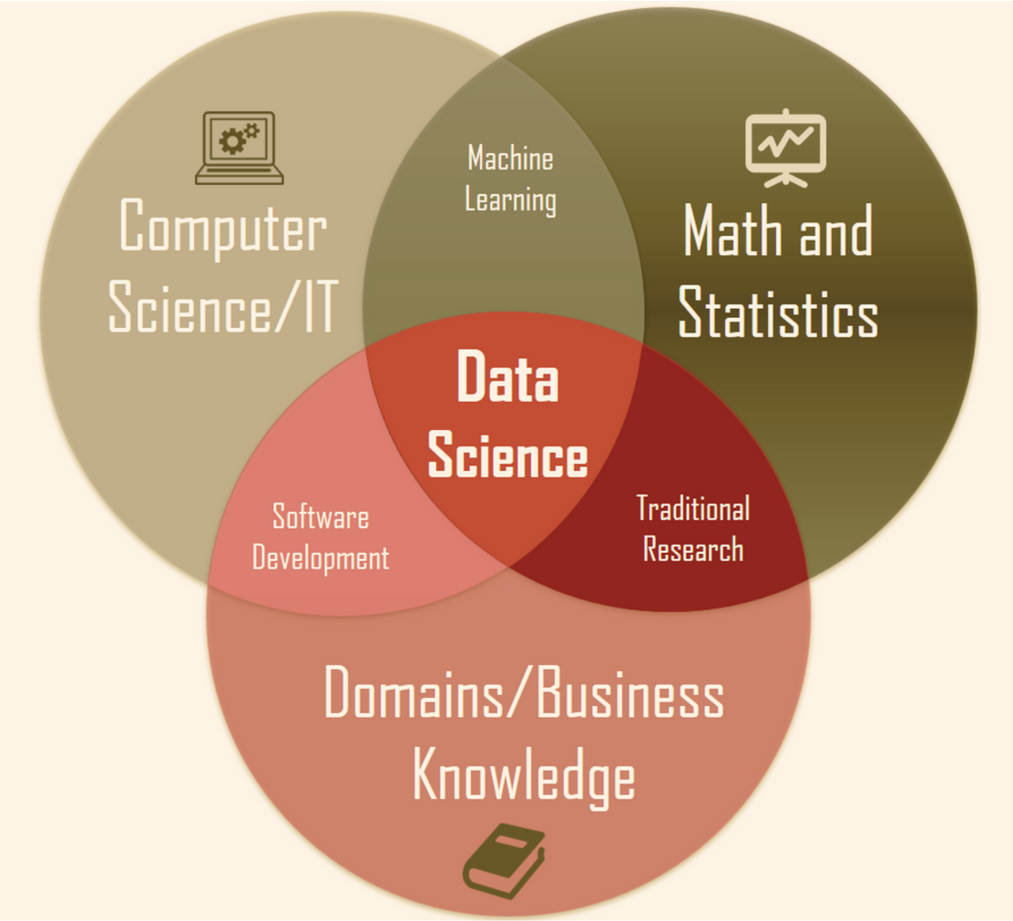
AI Saturdays Lagos

Presented By:
Emefa Duah (Mrs)

Outline

1. What is data science
2. What is not data science
3. Data Science vs Machine Learning
4. Data science examples
5. The skillset of data scientists
6. Data science pipeline

What is Data Science

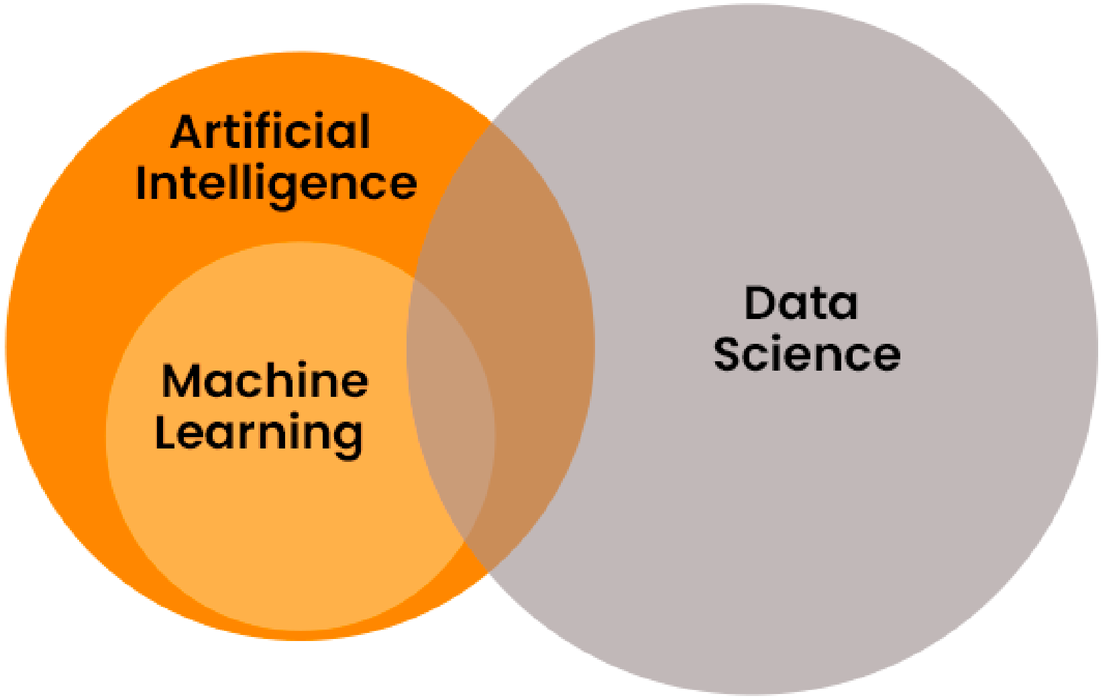


It is a **multidisciplinary** approach that combines principles and practices from the fields of mathematics, statistics, artificial intelligence, and computer engineering to analyze large amounts of data.

What is Not Data Science

- Data science is not Statistics.
- Data science is not machine learning.
- Data science is not big data

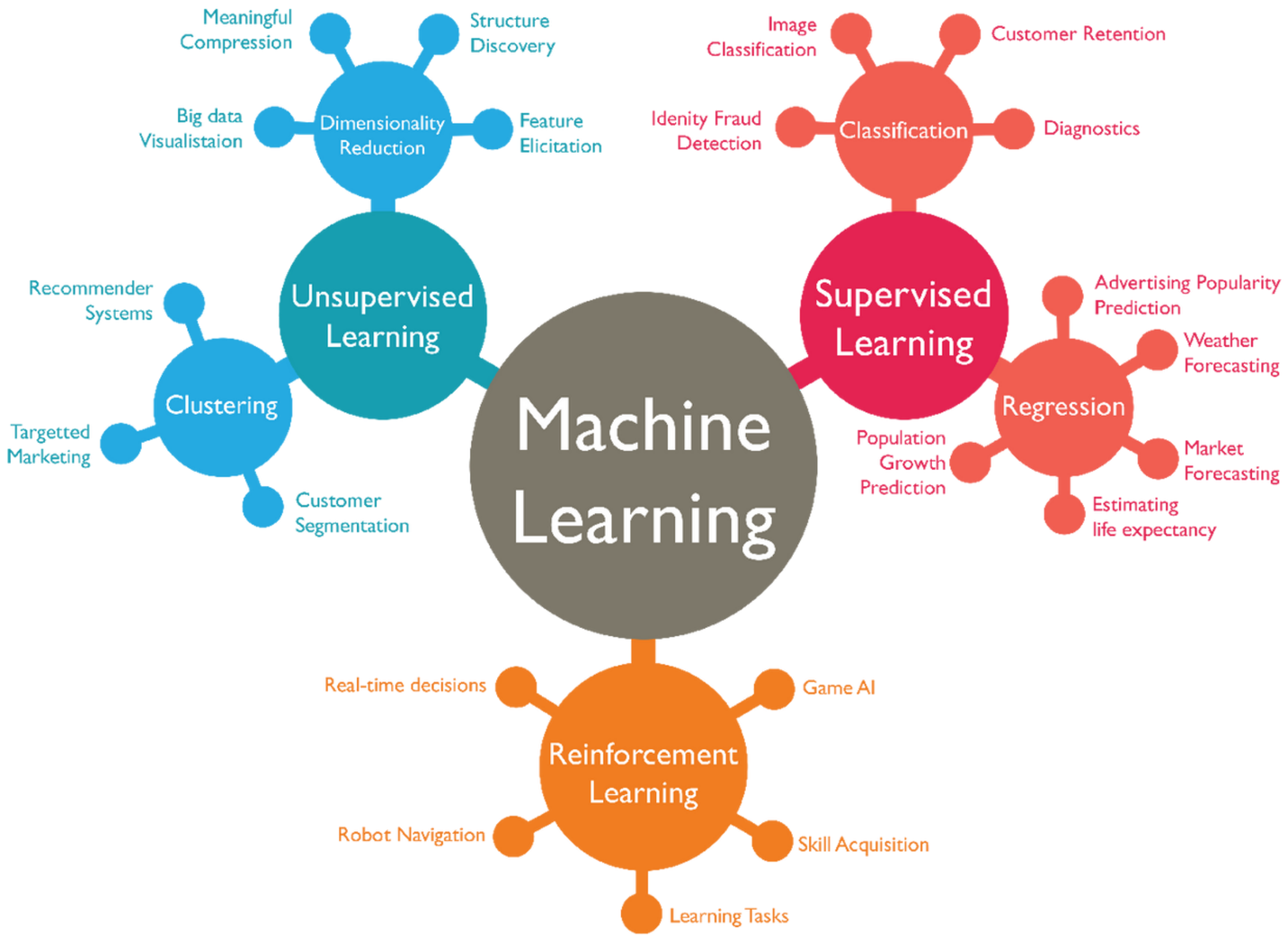
Data Science vs Machine Learning



Machine Learning is the science of getting machines to interpret, process and analyze data in order to solve real-world problems.

The three types of ML includes Supervised, unsupervised and Reinforcement learning.

Data Science vs Machine Learning(Cont'd)



Data Science vs Machine Learning (Cont'd)

Machine learning	Data science
Machine Learning focuses on enabling machines to self-learn and execute any task	Data science focuses on using data and complex machine learning algorithms to help businesses analyze and understand trends.
Machine learning is a branch of Artificial Intelligence	Machine learning is a part of Data science. Its one of the skills you need as a Data scientist

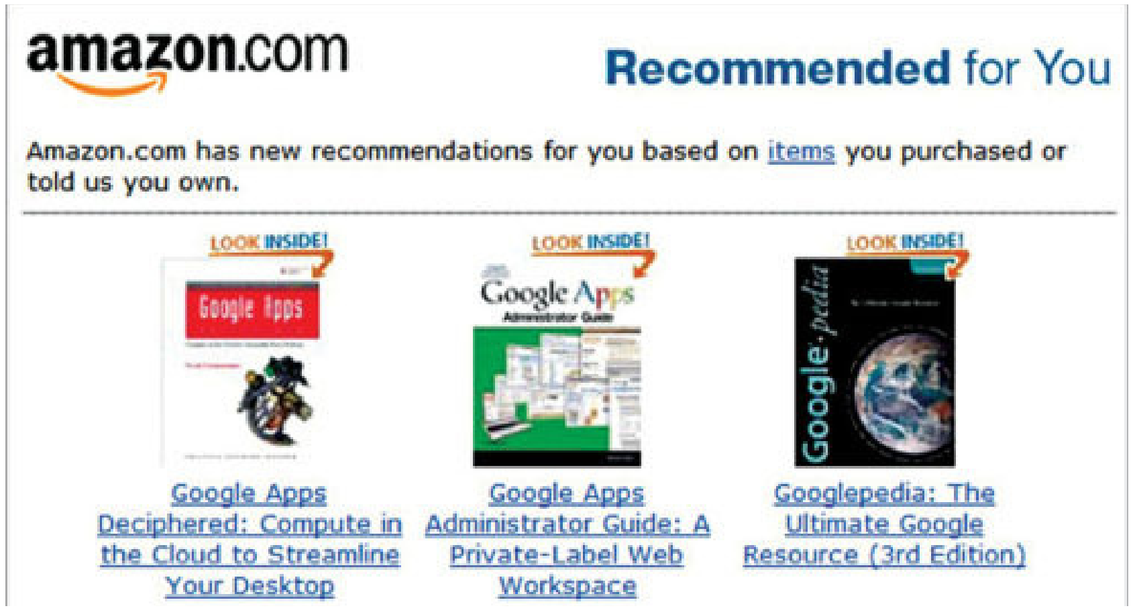

QUIZ
TIME!

Data science examples (I)

Website Recommendations

- This method is used by internet companies such as Amazon, Twitter, Google Play, Netflix, LinkedIn etc.

- Recommendations are based on subscriber's prior search results.



Data science examples (II)

Virtual Assistants and Health Robots

- AI-powered smartphone apps, which are mostly chatbots

Data science examples (III)

Text and Advanced Image Recognization

- Image recognition may be found on Facebook, Instagram, and Twitter, among other social media platforms.

- Virtual assistants like Siri, google assistant etc.

Any more examples ? Lets discuss

The skillset of data scientists

To be a data scientist, you'll need to be able to gather, clean and analyze data, then present your findings. You'll need both technical and soft skills

Technical skills

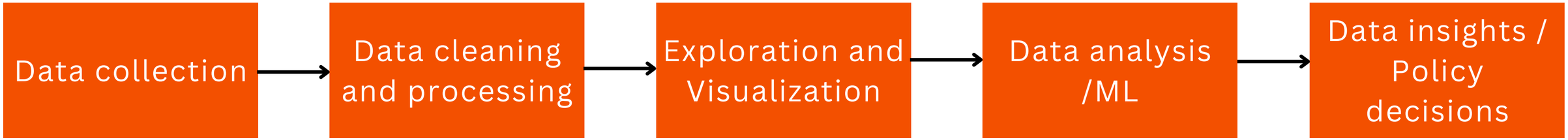
- Programming skills
- SQL
- Statistics, Probability and Data Analytics
- Data visualization
- Business strategy
- ML and AI

Soft skills

- Good communication
- Analytical thinking
- Story telling
- Continuous learning
- Attention to detail
- Teamwork

Data science pipeline

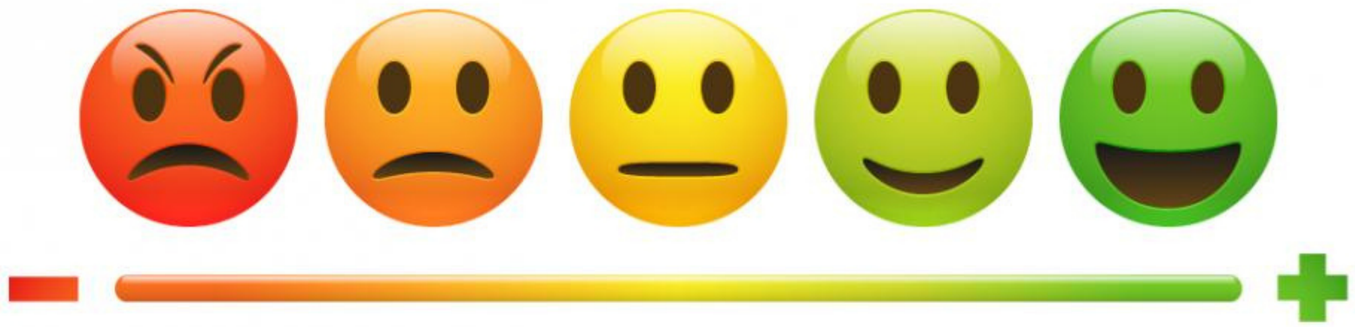
The data science pipeline refers to the process and tools used to gather raw data from multiple sources, analyze it, and present the results in an understandable format.



In reality, this is virtually never a linear process

Feedback Form: Learning Experience

https://docs.google.com/forms/d/e/1FAIpQLSfoYXRTHJfjQo1Xj2PrgoNLKWOAUeiVl4MLrRrOACtYFzM8cg/viewform



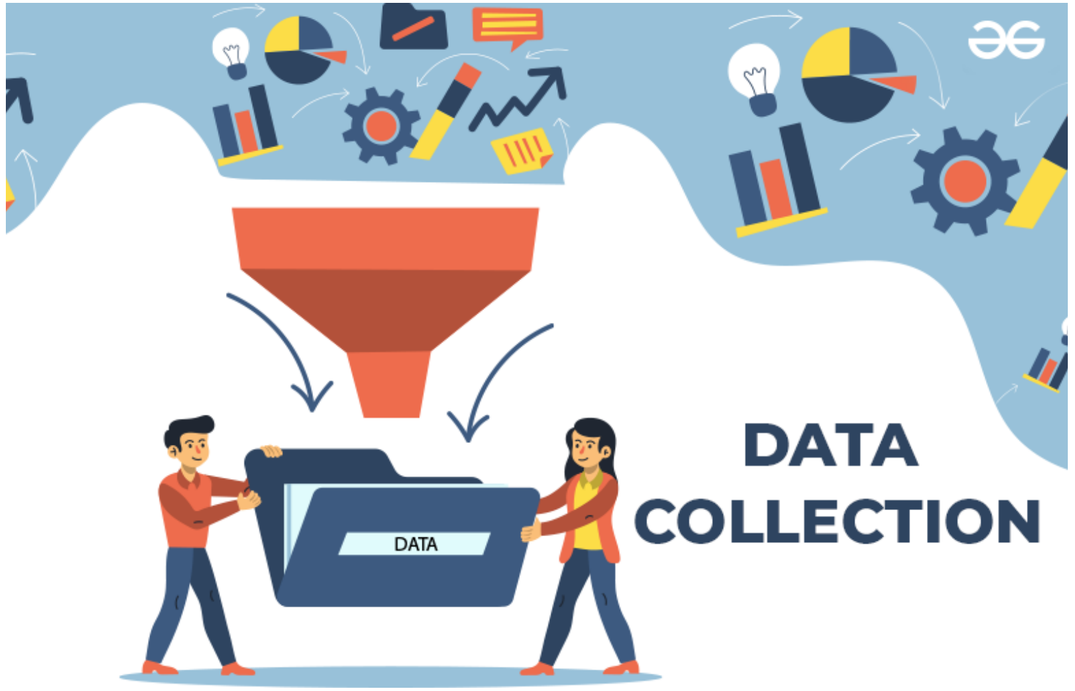
Data Collection

Data collection is the process of gathering data for use in business decision-making, strategic planning, research and other purposes.

There are two major types of data collections methods.

Primary - surveys, interviews, focus groups, statistical methods, polls etc.

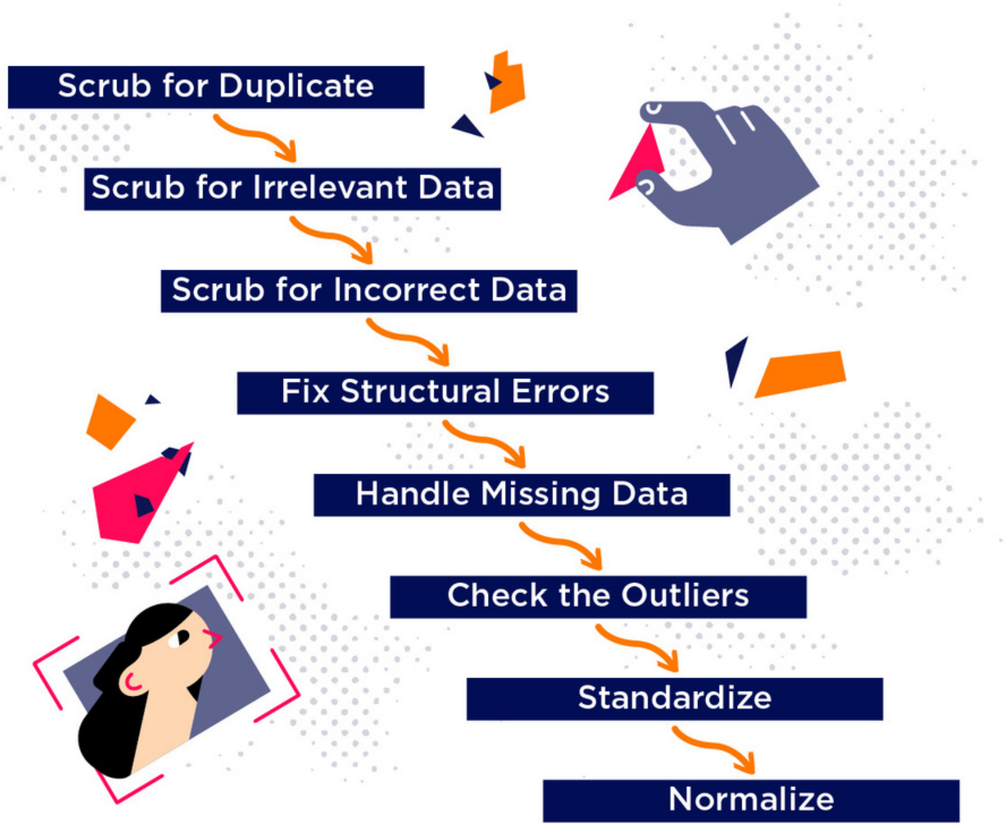
Secondary - Sales report, financial reports, internet (Data Scraping)



Data cleaning and processing

Data cleaning is the process of fixing or removing incorrect, corrupted, incorrectly formatted, duplicate, or incomplete data within a dataset.

Data processing occurs when data is collected and translated into usable information.



Exploration and Visualization

Data exploration techniques enable users to easily identify variables that are likely to have interesting observations.

Data visualization is the representation of data through use of common graphics, such as charts, plots etc.

Data Analysis

Data analysis is the process of systematically applying statistical and/or logical techniques to describe and illustrate and evaluate data.

The four types of data analysis are:

- Descriptive Analysis
- Diagnostic Analysis
- Predictive Analysis (ML)
- Prescriptive Analysis

Data insights / Policy decisions

Data insights are found through the deep analysis of patterns and statistics within data.

Through the interpretation of these patterns, organizations transform insights into forecasting customer needs, therefore promoting more effective decision-making.

Thank you. :)

Any Questions?