MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Unit 1: Introduction

Mamatha.H.R

Department of Computer Science and Engineering

PES UNIVERSITY
CELEBRATING 50 YEARS

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Unit 1:Introduction

Mamatha H R
Department of Computer Science and Engineering

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Course content

Unit 1: Descriptive Statistics and Sampling Distributions **14 Hours**

Sampling and Descriptive Statistics: Introduction, Motivating Examples and Scope. Statistics: Introduction, Types of Data, Types of Experiments—Controlled and Observational Study, Types of Statistics, Descriptive and Inferential Statistics, Sampling: Sampling Methods, Sampling Errors.
Sampling Distributions: Chebyshev's inequality, Linear function of random variables, Normal Probability Plots, Introduction to Generation of Random Variates and mention the types, Acceptance-Rejection method, Sampling Distribution, The Central Limit Theorem and Applications.
Self Learning: Mean, Median, Mode

Unit 2: Point Estimation, MLE, Confidence Intervals **14 Hours**

Point Estimation, MLE: Principles of Point Estimation – Mean Squared Error for Bernoulli, Binomial, Poisson, Normal, Maximum Likelihood Estimate for Bernoulli, Binomial, Poisson, Normal. Introduction to Multivariate Normal Distribution, MAP distribution.
Confidence Intervals: Interval Estimates for Mean of Large and Small Samples, Student's t Distribution, Interval Estimates for Proportion of Large and Small Samples, Confidence Intervals for the Difference between Two Means, Confidence interval for difference between Two Proportions, and Interval Estimates for Paired Data. Factors affecting Margin of Error.

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Course content

Unit 3: Hypothesis Testing and Simple Linear Regression **14 Hours**

Hypothesis Testing: Introduction, Large-Sample tests for a Population Mean, Drawing conclusions from the results of Hypothesis tests, the relationship between confidence interval and hypothesis tests, Tests for population proportion, Small-Sample Tests for a Population Mean. Distribution-Free Tests, Chi-squared Test, Fixed Level Testing, Type I and Type II Errors, Power of a Test, Factors Affecting Power of a Test.

Simple Linear Regression: Introduction, Correlation, Fitting Least Squares Line, Checking Assumption

Unit 4: Engineering Optimization **14 Hours**

Introduction to Optimization, Types of Optimization, Mathematical concepts of Objective Function, Constraints and Decisions, Constrained Linear Optimization- LPP – Formulation, Graphical Method, The Simplex Method, Discrete Variable Optimization – Integer Programming Problem, Branch and Bound Method, Constrained Non-linear Optimization – Lagrangian Multipliers method, Kuhn Tucker Conditions, Unconstrained non-linear Optimization – Gradient Descent Method, Evolutionary Computation - Genetic Algorithms.

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Course content: Applications

Unit 1: Applications:

1. Poisson distribution: Calculation of number of calls received in a specified time duration in call centers, Modeling the Arrival of Patients in an Emergency Room
2. Variance, Standard Deviation: Monitoring Transaction Variability in E-Commerce, Identifying the customer satisfaction in online shopping
3. Central Limit Theorem: Estimating Average Waiting Time in Public Transport, Load Balancing in Distributed Systems and Internet Traffic Prediction

Unit 2: Applications:

1. Application of Point Estimation (MLE) in Email Spam Detection
2. Application of Mean Squared Error (MSE) in Sensor Calibration
3. Application of Confidence Intervals in Clinical Trials

Unit 3: Applications:

1. Hypothesis Testing in E-commerce Conversion Rates.
2. Chi-Square Test for analyzing the association between vaccination and recovery of the patients considering COVID data, Analyzing the relationship between gender and preference for a product purchase.
3. Identifying Type 1 and Type 2 Errors in Spam Mail Classification.
4. Application of Simple Linear Regression in Real Estate Price Prediction, Stock Price Prediction

Unit 4: Applications:

1. Minimize a Loss functions in Neural Networks using Batch gradient descent (Unconstrained Optimization)
2. Lagrange Multipliers to find local maxima and minima of a function subject to equations constrains (Constrained Optimization)
3. Case study on Bayesian Optimization with Discrete Variables (Discrete Variable optimization)
4. Use Genetic Algorithms to optimize Production Scheduling in a manufacturing environment, focusing on minimizing total production costs while meeting job deadlines and machine constraints. Evaluate the GA's effectiveness against traditional scheduling methods.

Tools and Textbooks

Tools / Languages/Libraries: Jupyter Notebook, Python, Pandas, Matplotlib, Scipy, Seaborn, BeautifulSoup, Numpy, Scikit learn.

Text Book(s):

1. "Statistics for Engineers and Scientists", William Navidi, McGraw Hill Education, India, 6th Edition, 2024.
2. "Engineering Optimization Theory and Practice", Singiresu S. Rao, John Wiley & Sons, Inc, 5th Edition, 2020

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Evaluation Policy

	Conduction	Reduced to
ISA 1	40	20
ISA 2	40	20
Assignment	20	10
ESA	100	50

Assignment Components

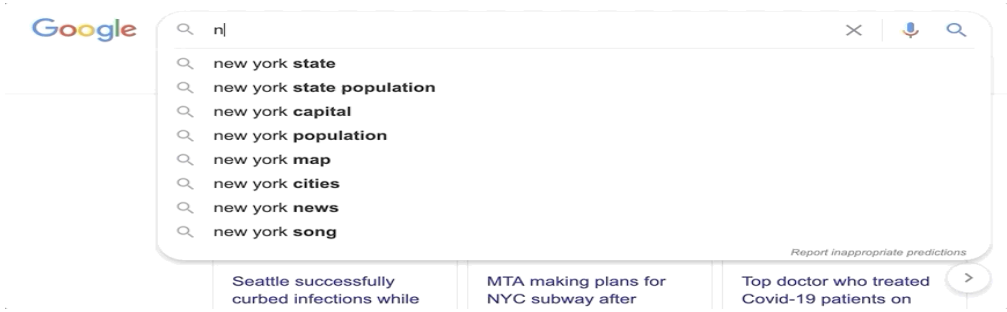
Banana Level Problem	2 Marks (Conducted for 4 marks)
Orange Level Problem	3 Marks (Conducted for 6 marks)
Jackfruit level problem (Datathon)	5 Marks (Conducted for 10 marks)

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Data Science

- Have you ever wondered how YouTube recommends videos of your liking?
- How Google's autocomplete works?
- How Gmail filters your emails into spam and non-spam categories?

Data Science enables us to solve real-world problems by leveraging the power of data. Many everyday applications—such as recommendation systems, fraud detection, and predictive analytics—would not be possible without access to relevant data. In essence, **Data Science** is the discipline of extracting meaningful insights and solutions from data.

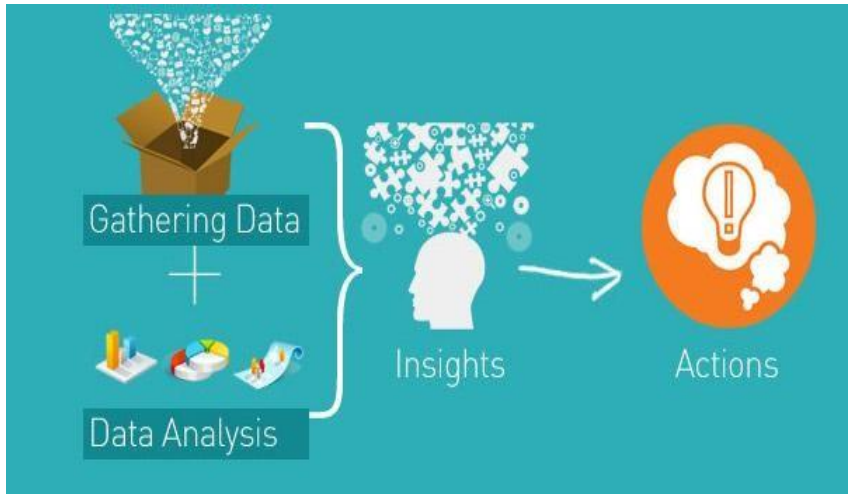


Source: https://coralogix.com/blog/elasticsearch-autocomplete-with-search-as-you-type/

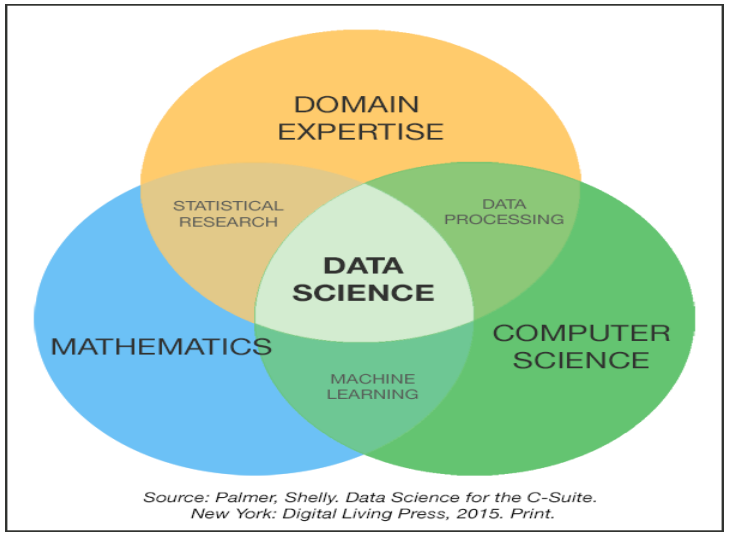
Data Science

Data Science is an interdisciplinary field dedicated to extracting knowledge and actionable insights from data.

- By combining techniques from Statistics, Computer Science, Mathematics, and Domain-specific knowledge (e.g., Business or Healthcare), data science enables informed decision-making and problem-solving across a wide range of sectors.
- The insights derived from data are applied to address complex challenges, optimize processes, and drive innovation in fields such as finance, healthcare, marketing, education, and more.



Source: theblog.adobe..com

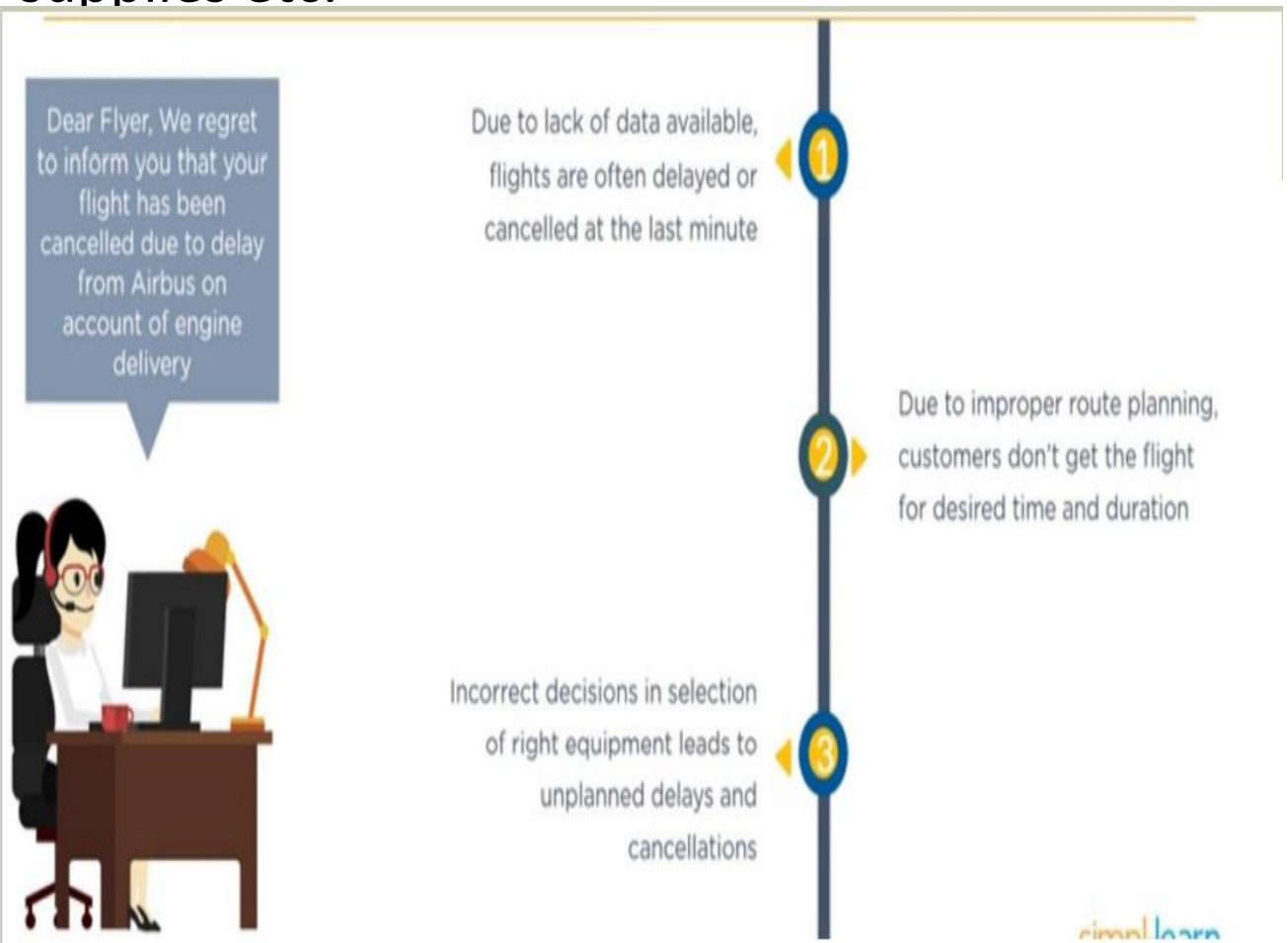


Applications of Data Science

Source: edureka.com

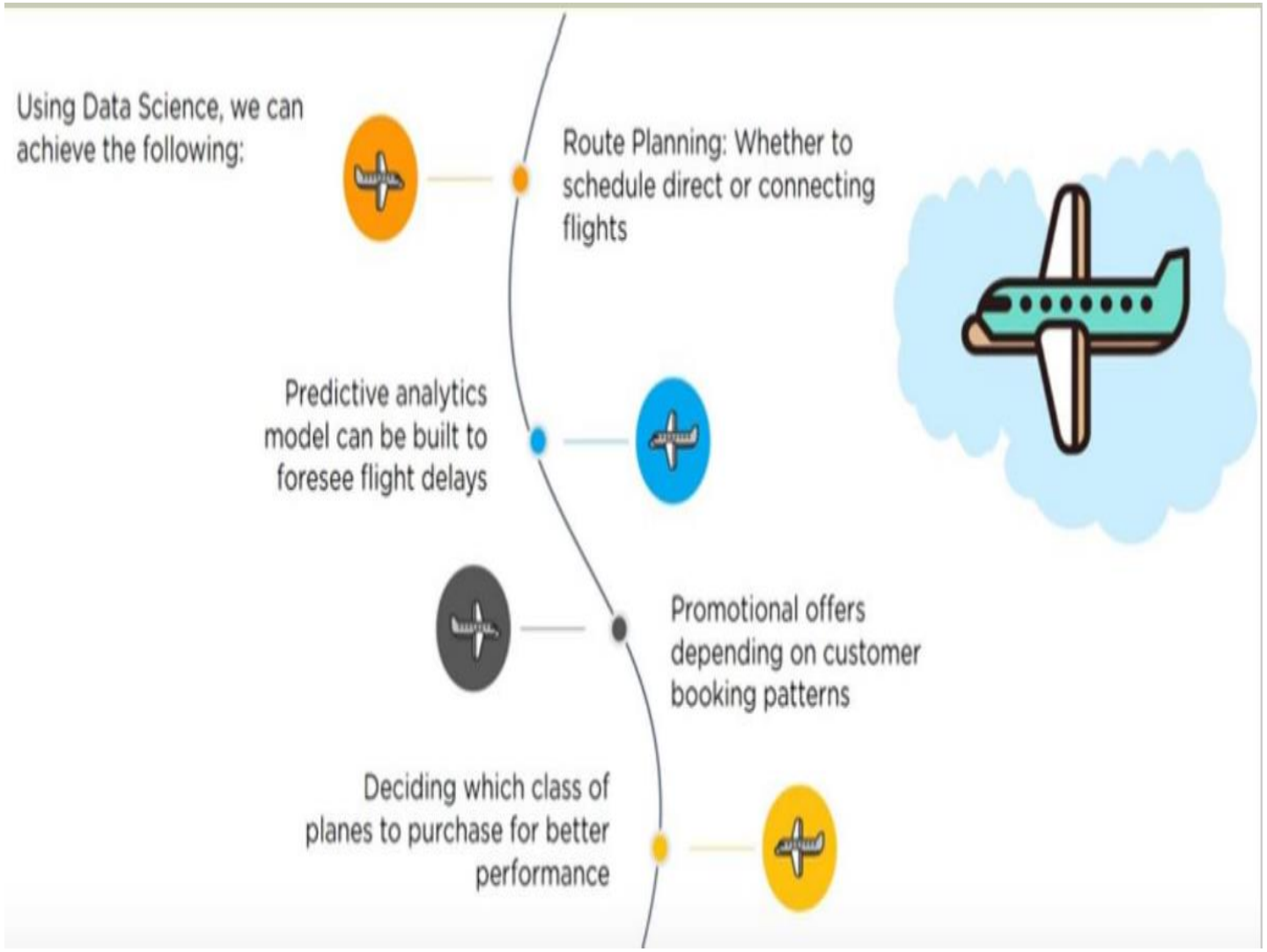
Applications of Data Science : Data Science in Airlines Industry

Data Science is used for various purposes like: route planning, revenue management, prediction on in-flight sales and food supplies etc.



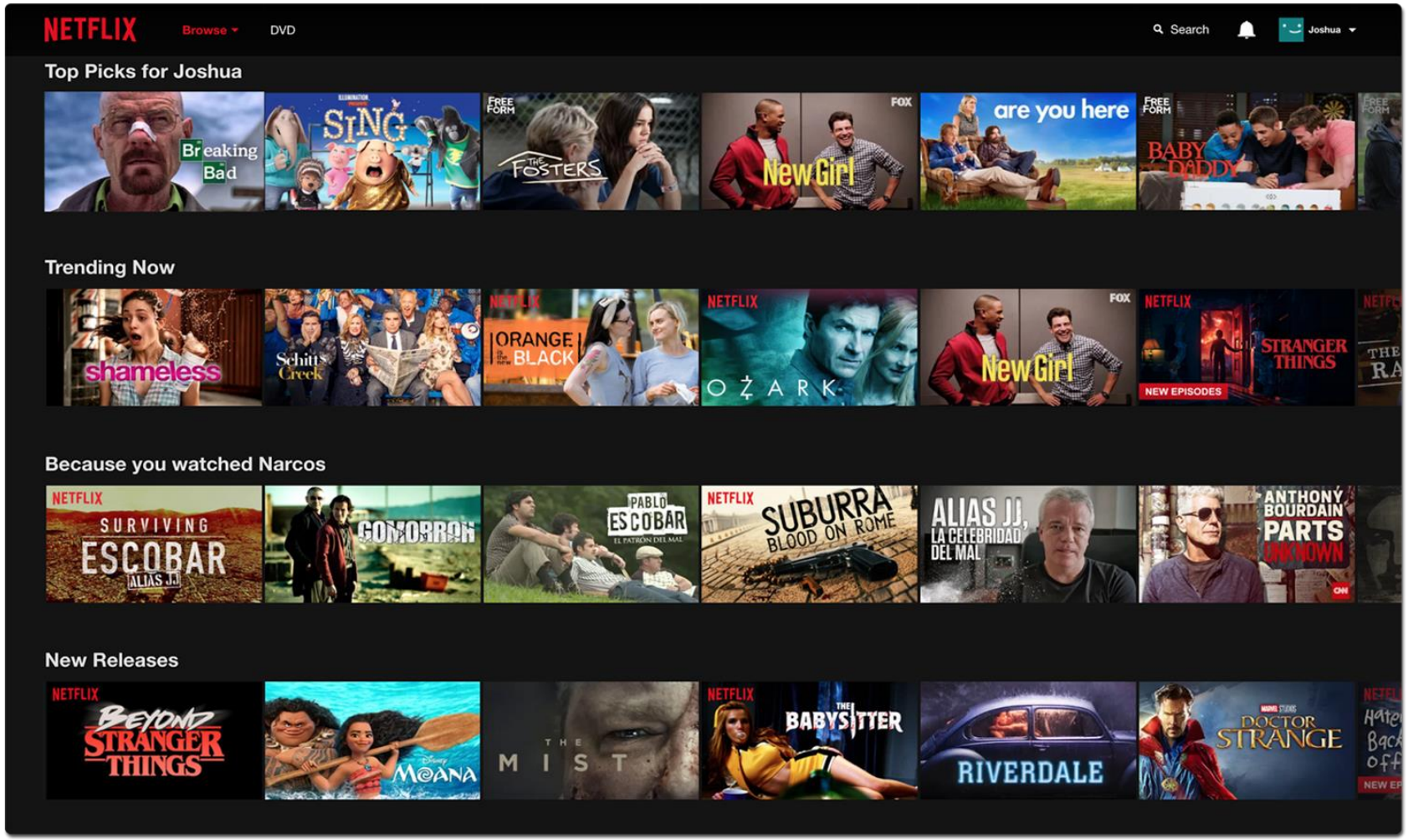
Sources: Simplilearn, datasciencecentral.com

Applications of Data Science : Data Science in Airlines Industry



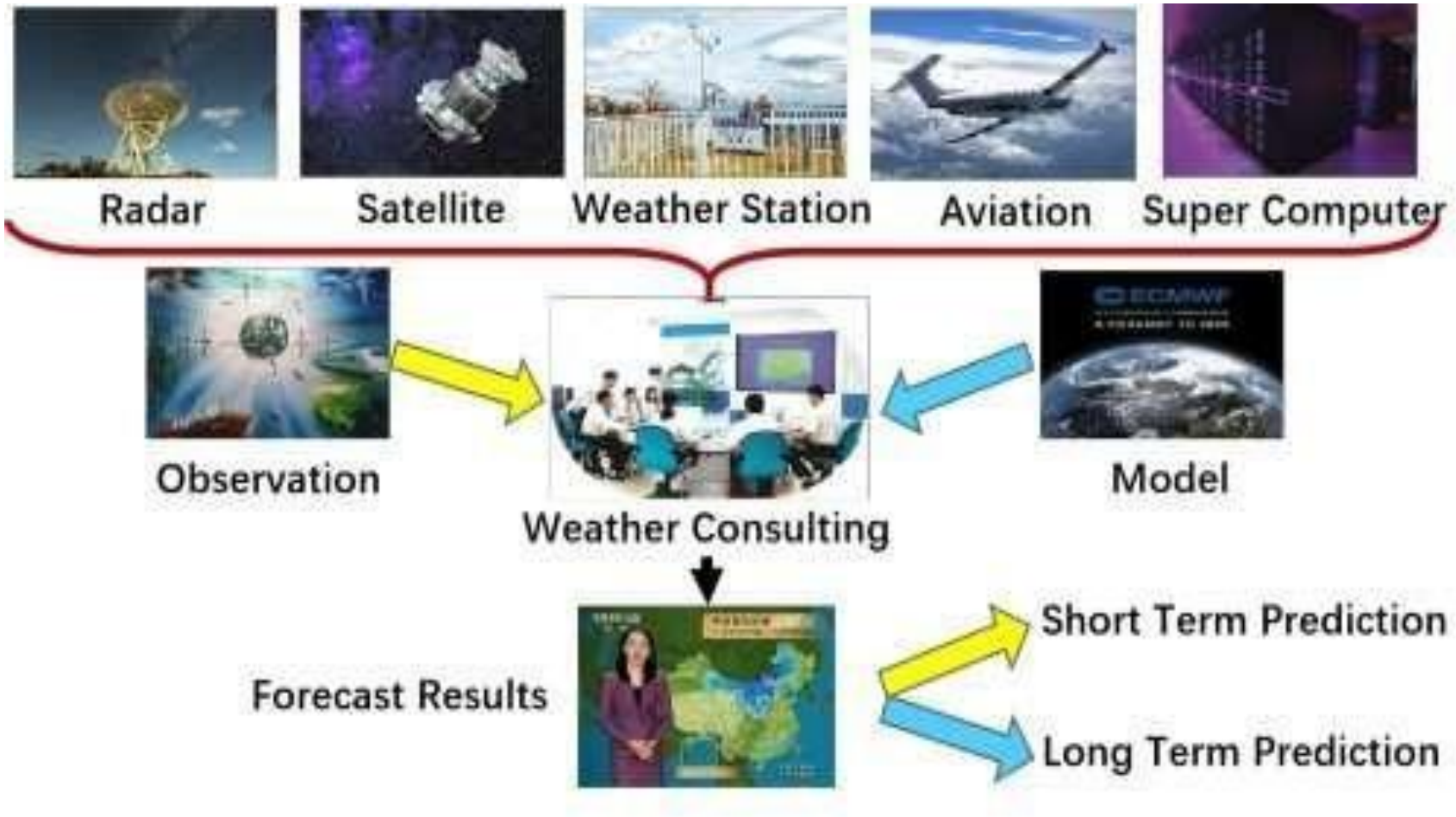
Sources: Simplilearn, datasciencecentral.com

Applications of Data Science : Data Science in Recommender systems



Source: https://medium.com/swlh/recommendations-in-time-context-93b32f73d98d

Applications of Data Science : Data Science in Weather Forecasting



Source: phys.org/news

Applications of Data Science : Data Science in Sports

Source: https://arstechnica.com/information-technology/2015/10/big-data-an-it-buzzword-that-is-actually-producing-results/

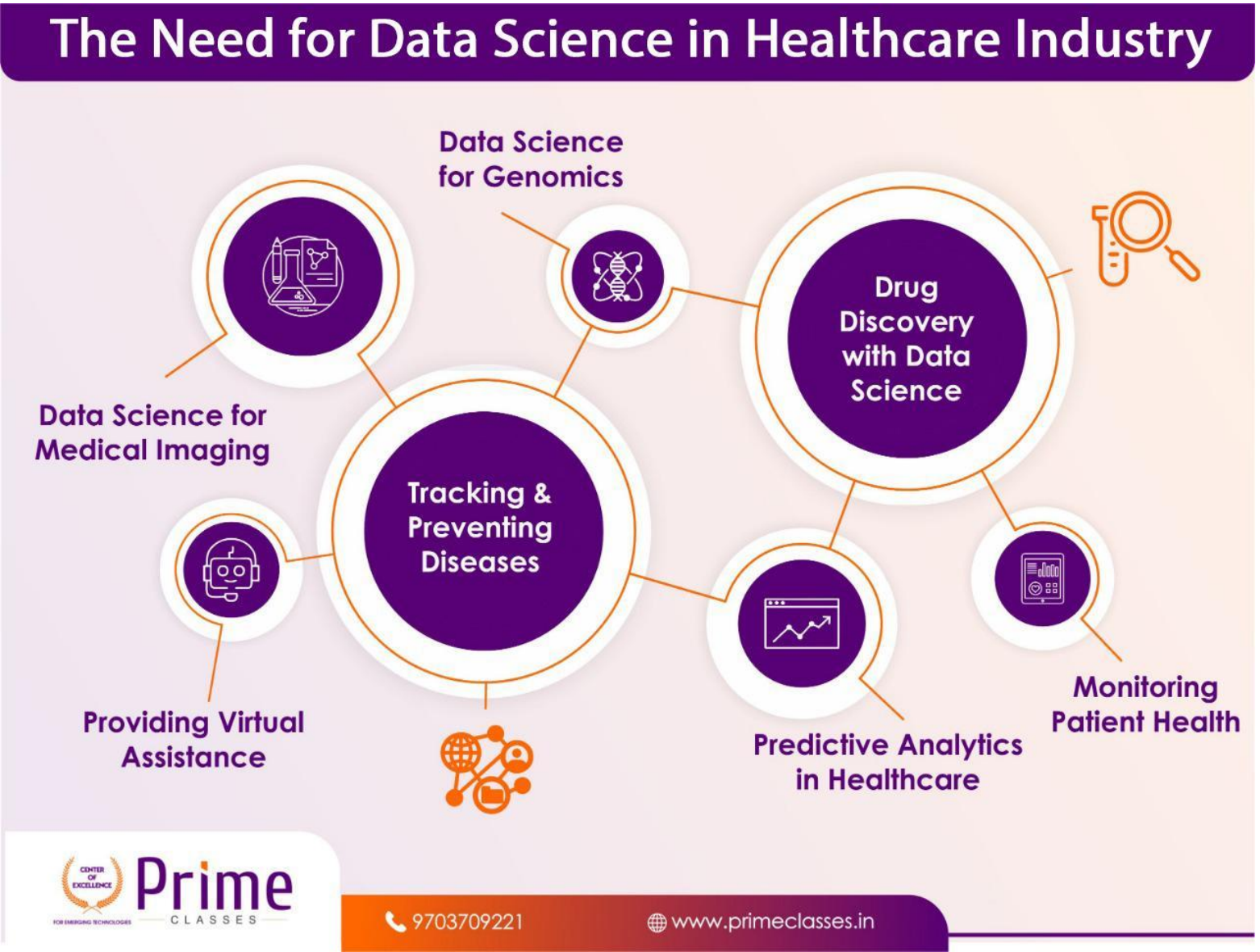
Applications of Data Science : Data Science in Politics

Political parties and their strategists have realized the **importance of mining real-time demographic and polling data**.

The various data points may include voter sentiment, mass emotions, citizen concerns in different constituencies, popular outlooks in various states, etc. Political parties can use these insights to,

- pull voter donations
- convert undecided voters
- enroll young volunteers
- organize resources
- social media campaigns
- improve effectiveness of electioneering activities etc.

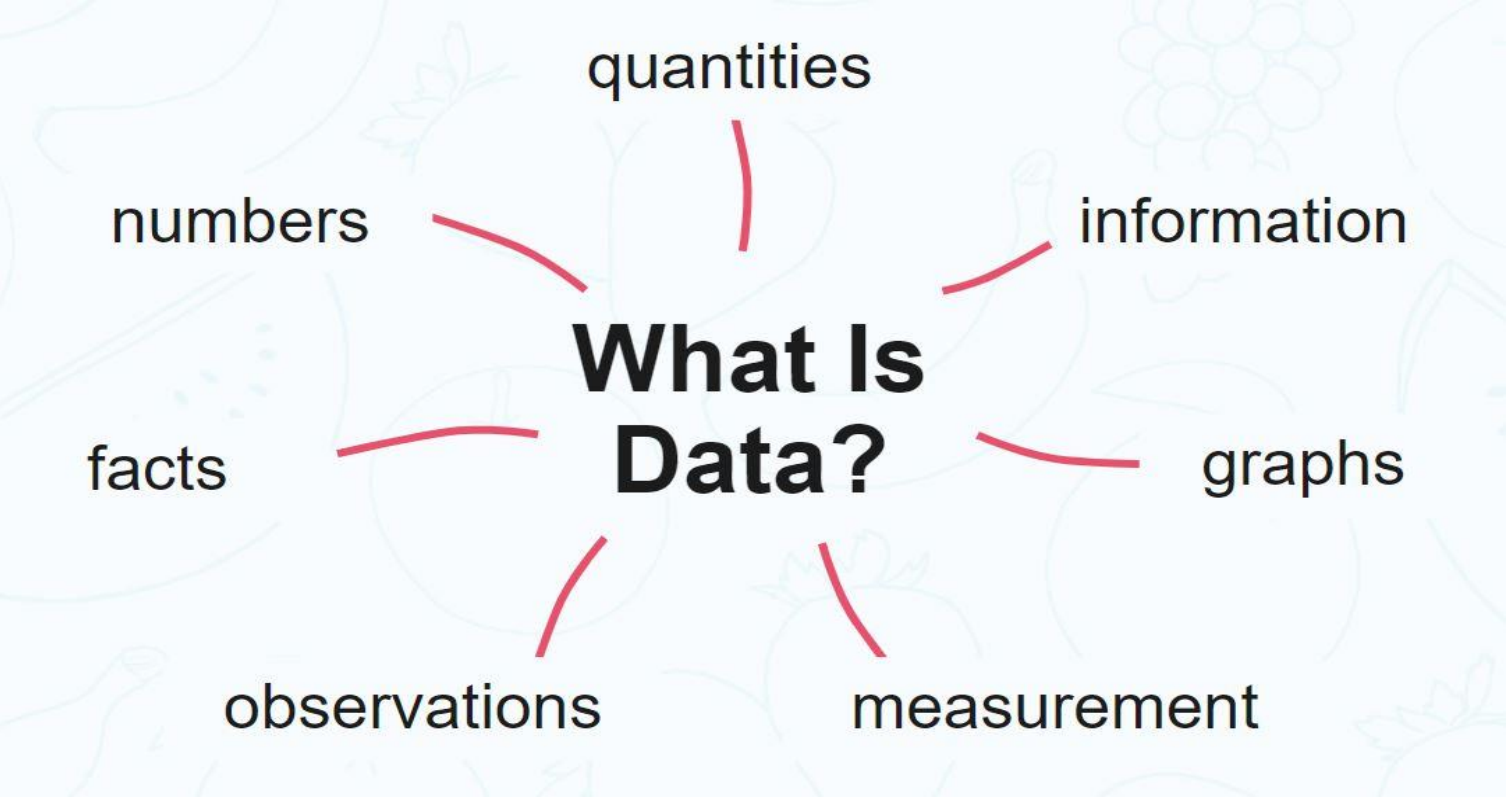
Applications of Data Science : Data Science in Healthcare & Medicine



Source:
http://www.primeclasses.in/blog/2019/08/26/the-need-for-data-science-in-healthcare-industry/

Data

Technically, data refers to **individual facts, statistics, or items of information**, often numeric, that are collected through observation.



Source: https://www.twinkl.de/teaching-wiki/data

Data vs Information

➔ **Data**

- Raw facts, usually formatted in a special way.
- Based on records, observations etc.
- Unorganized.

➔ **Information**

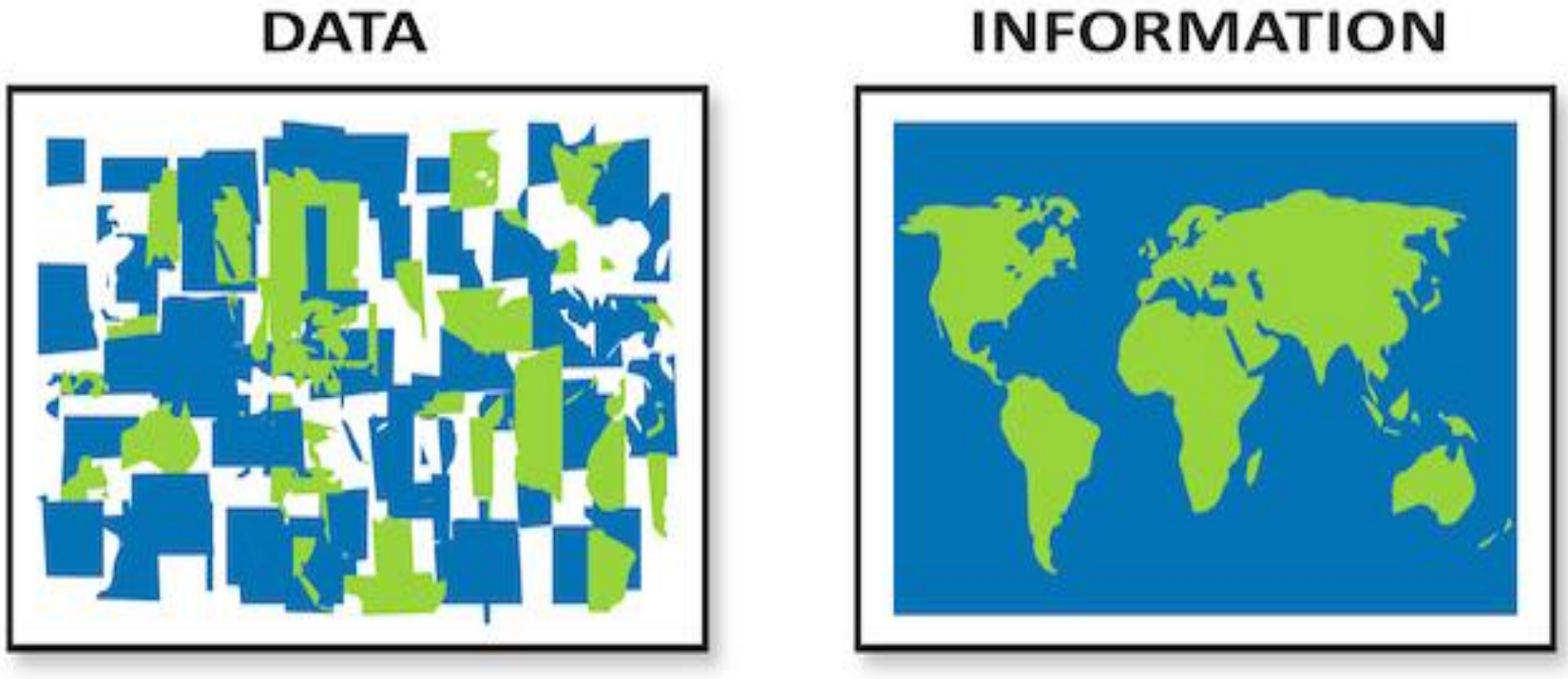
- A collection of facts organized in such a way that they have additional value beyond the value of the facts themselves.
- Based on analysis of data.
- Organized and always depends on data.

Ex : Data – thermometer readings of temperature taken every hour: (16.0, 17.0, 16.0, 18.5, 17.0,15.5….)

[on transformation]

Information – today's high: 18.5, today's low: 15.5

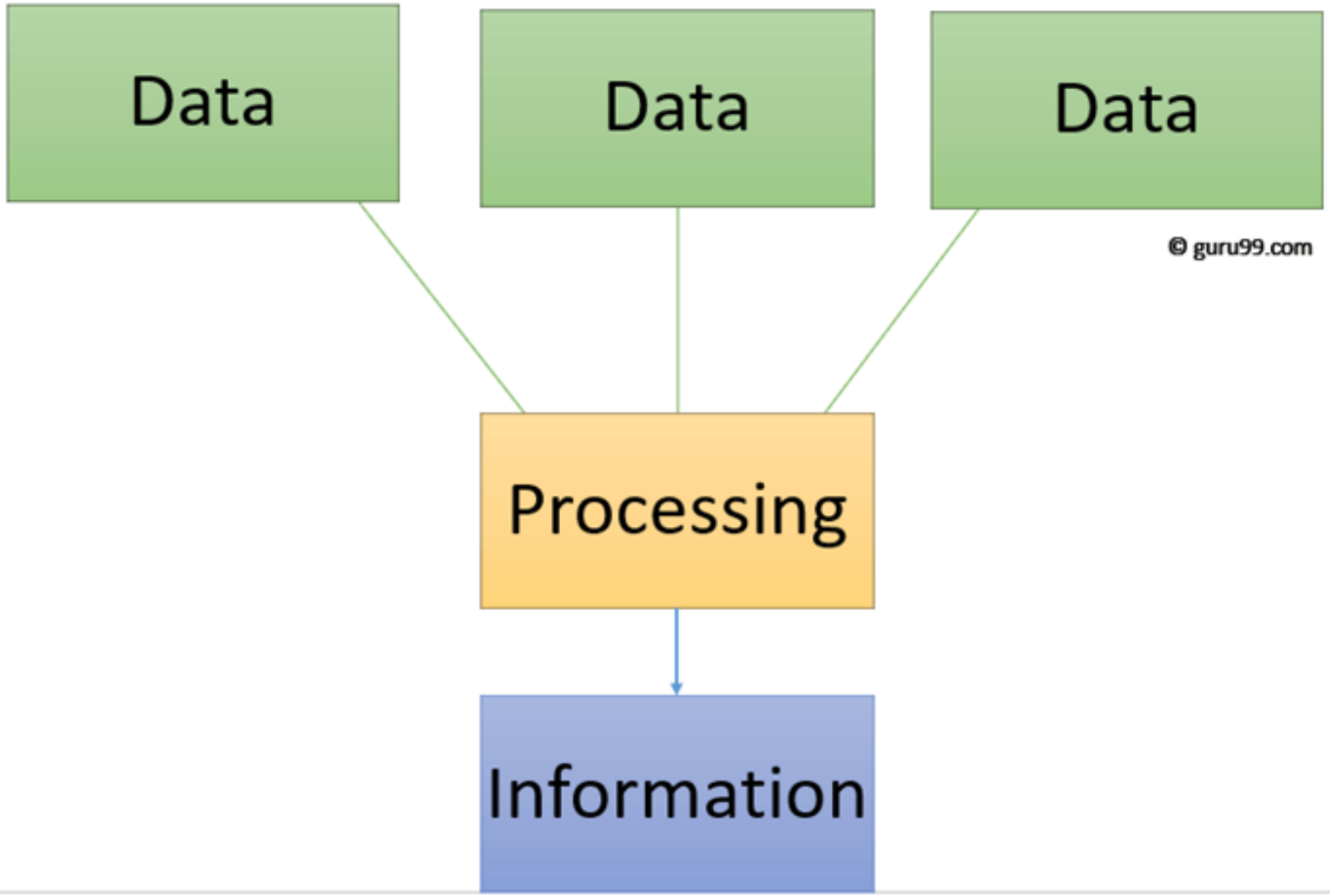
Data vs Information



Source: https://effectualsystems.com/data-need-information/

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Data ⟶ Information

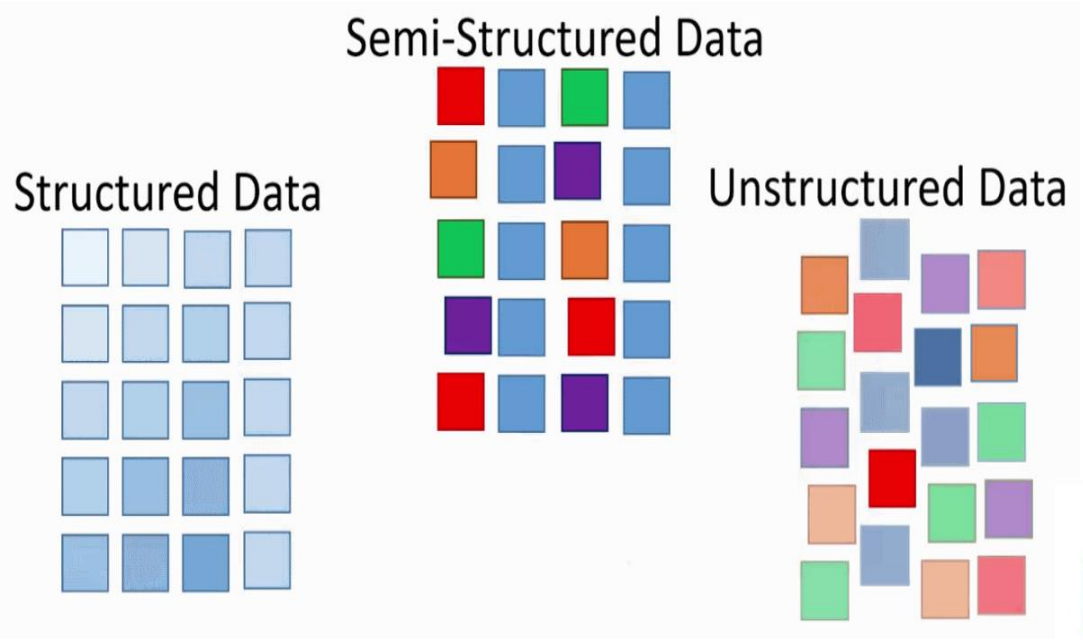


Source: guru99.com

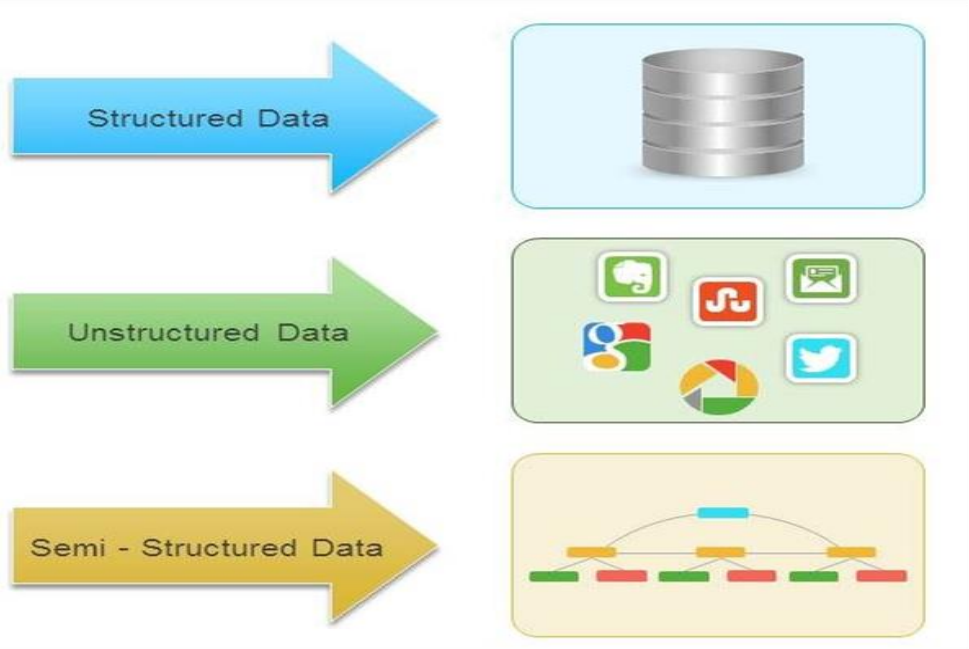
Types of Data

Data	Represented by
Alphanumeric data	Numbers, letters, and other characters
Image data	Graphic images or pictures
Audio data	Sound, noise, tones
Video data	Moving images or pictures

Structured, Unstructured & Semi-structured Data



Source: https://towardsdatascience.com/data-extraction-from-a-pdf-table-with-semi-structured-layout-ef694f3f8ff1



Source: slidegeeks.com

Structured, Unstructured & Semi-structured Data

Structured Data:

Structured data is the data whose elements are addressable for effective analysis. The data is organized into a formatted repository that is typically a database. Ex: Relational data.

Semi-Structured Data:

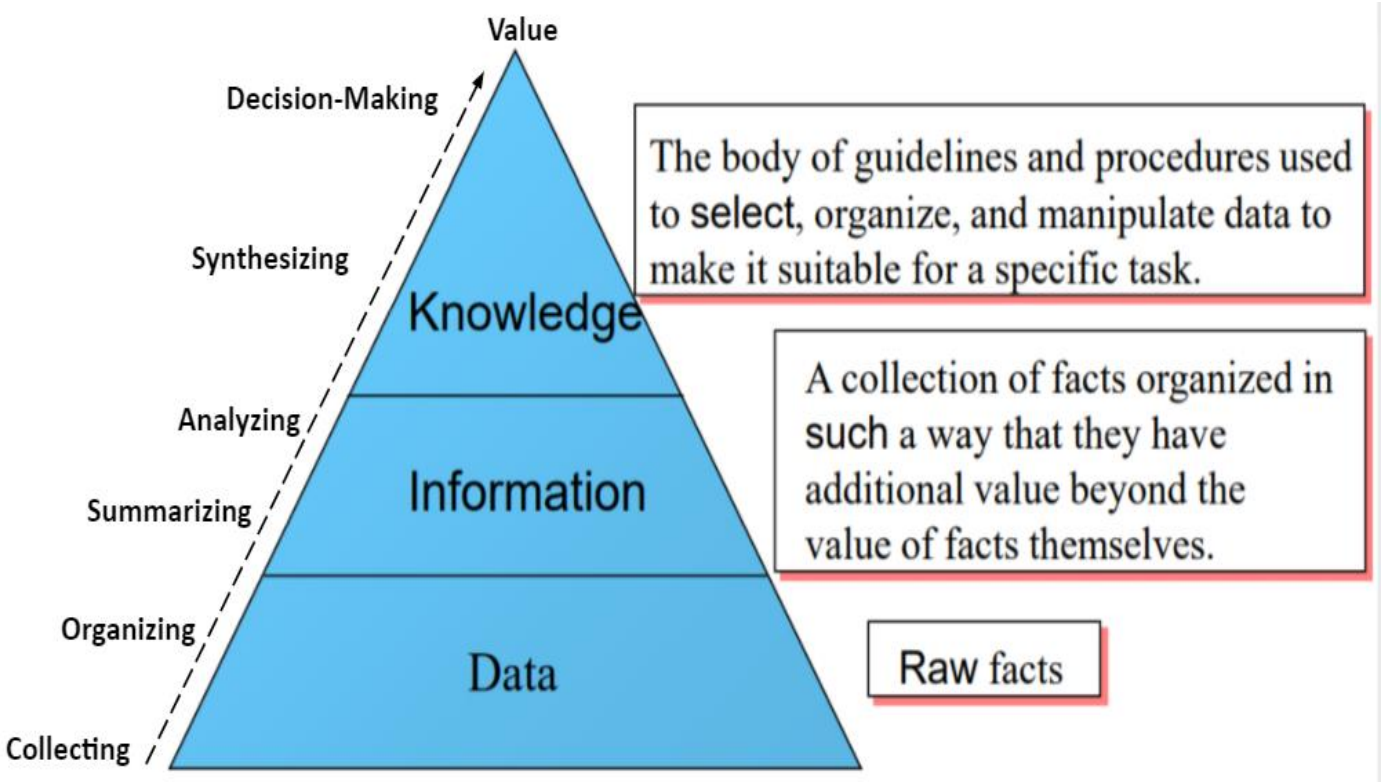
It is the data that doesn't reside in relational database but has some organizational properties that make it easier to analyse. Ex: XML data.

Unstructured Data:

It is the data which is not organized in a predefined manner or doesn't have a predefined data model, thus not a good fit for a mainstream relational database.

Ex: Word, pdf, text etc.

Information Concepts

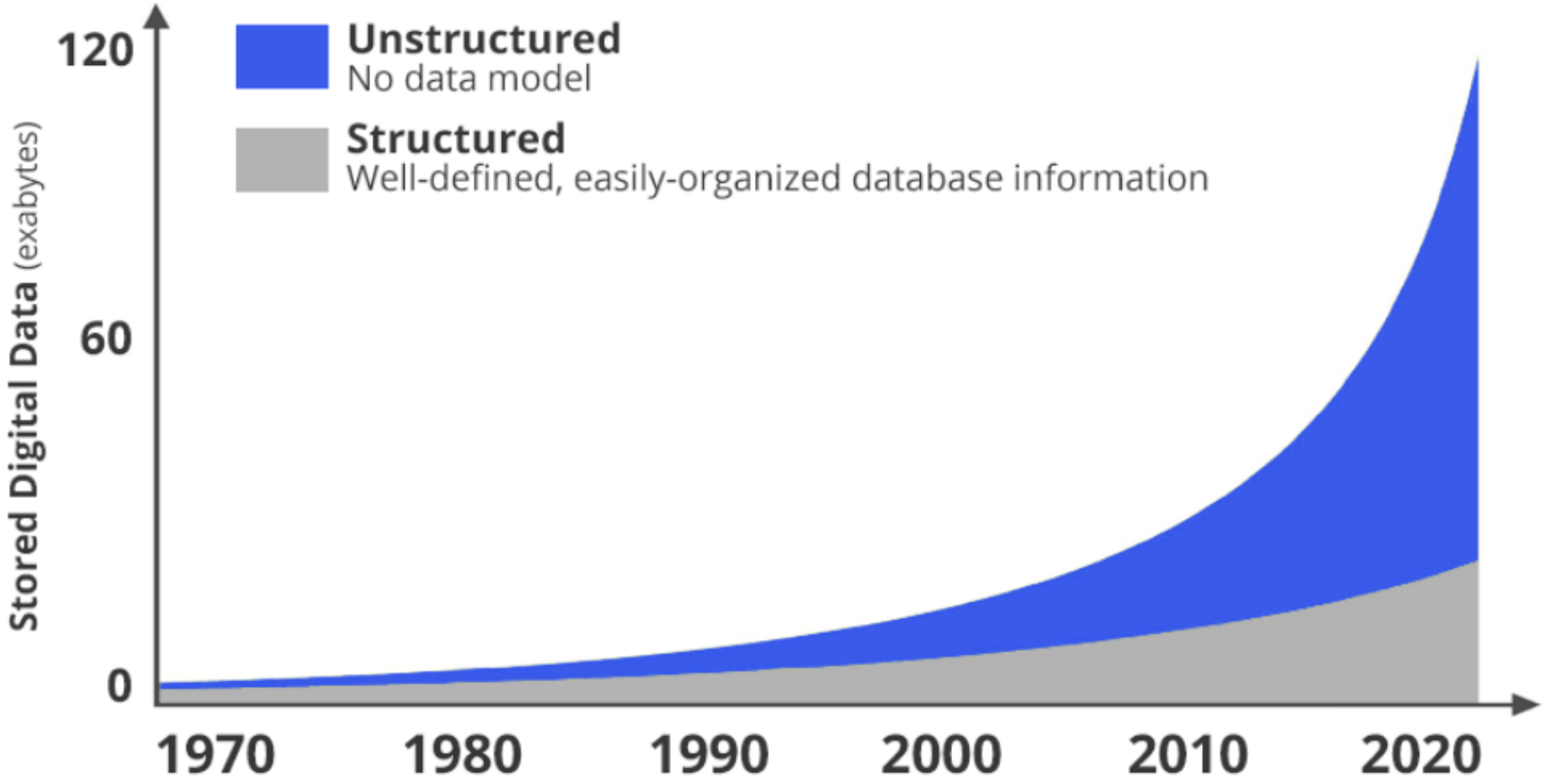


As we move up the pyramid:Data becomes more refined and useful.It gains value and relevance for decisions and actions.

Source: https://learningforsustainability.net

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Why do we need Data Science?



Source: https://static.seekingalpha.com/uploads/2020/1/14/50485001-15789998083991578_origin.png

Why do we need Data Science?

The main reason why we need data science is the **ability to process and interpret data**. This enables users and industries to **make informed decisions** as well as helps in their **growth, optimization, and performance**.

We know that, unstructured data is generated everywhere, every second. **Unstructured data** isn't well organized or easy to access. But its **growth is enormous** and importance of analyzing and drawing inferences from this type of data is crucial.

Data Science provides a number of methods and techniques to deal with such data.
This certainly helps many businesses and industries significantly to improve their productivity.

How is Data generated?

There is tons of data getting generated each day.
Some of the major sources from which data is generated are:
web, databases, media, IoT, cloud etc.
Insight into data generation in a day over the internet:

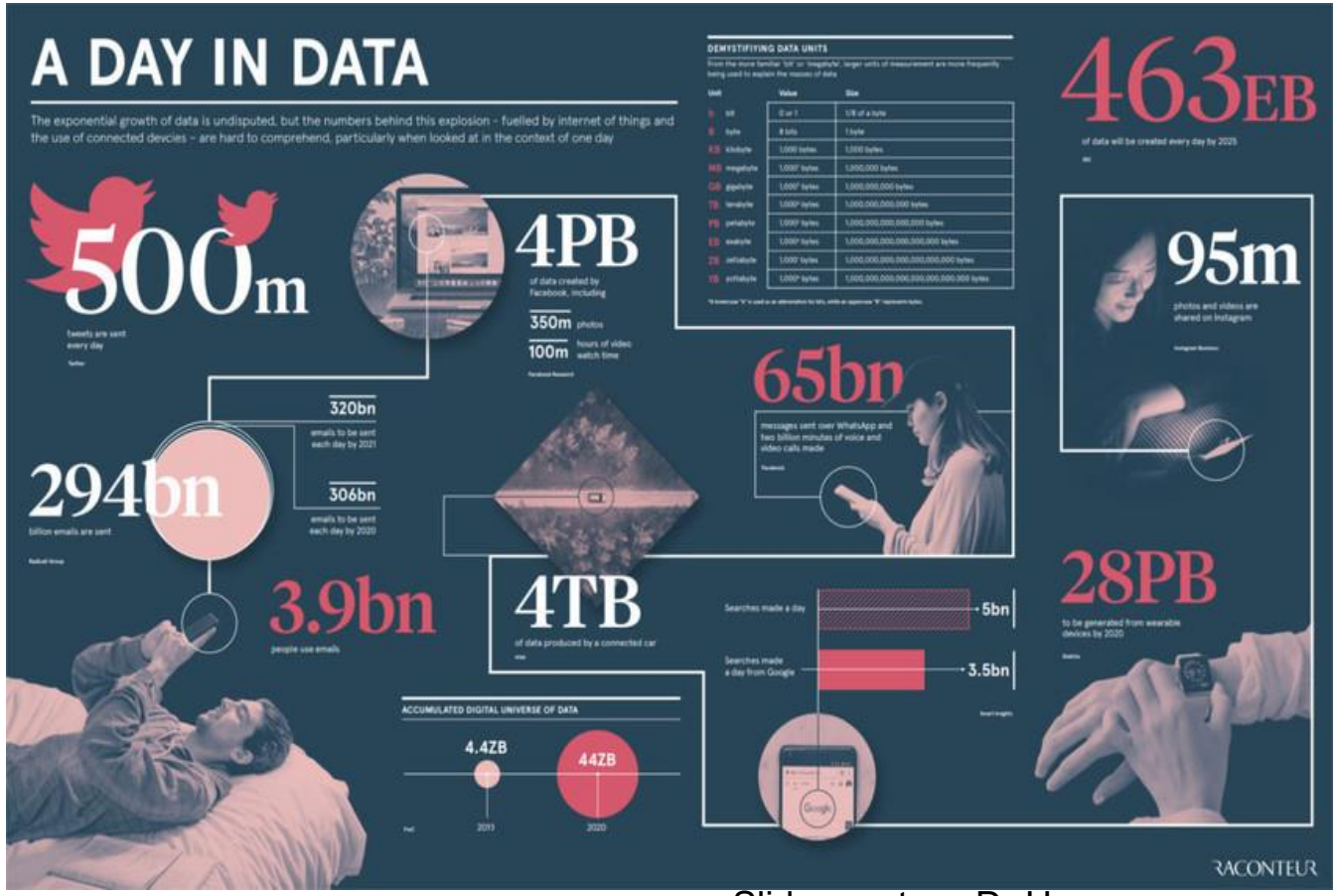
- 500 million tweets are sent
- 294 billion emails are sent
- 4 petabytes of data are created on Facebook
- 4 terabytes of data are created from each connected car
- 65 billion messages are sent on WhatsApp
- 5 billion searches are made

Slide courtesy:Dr.Uma

How is Data generated?

By 2025, it's estimated that 463 exabytes of data will be created each day globally

– that's the equivalent of 212,765,957 DVDs per day!



Source: theblog.adobe..com

Slide courtesy:Dr.Uma

Data generation

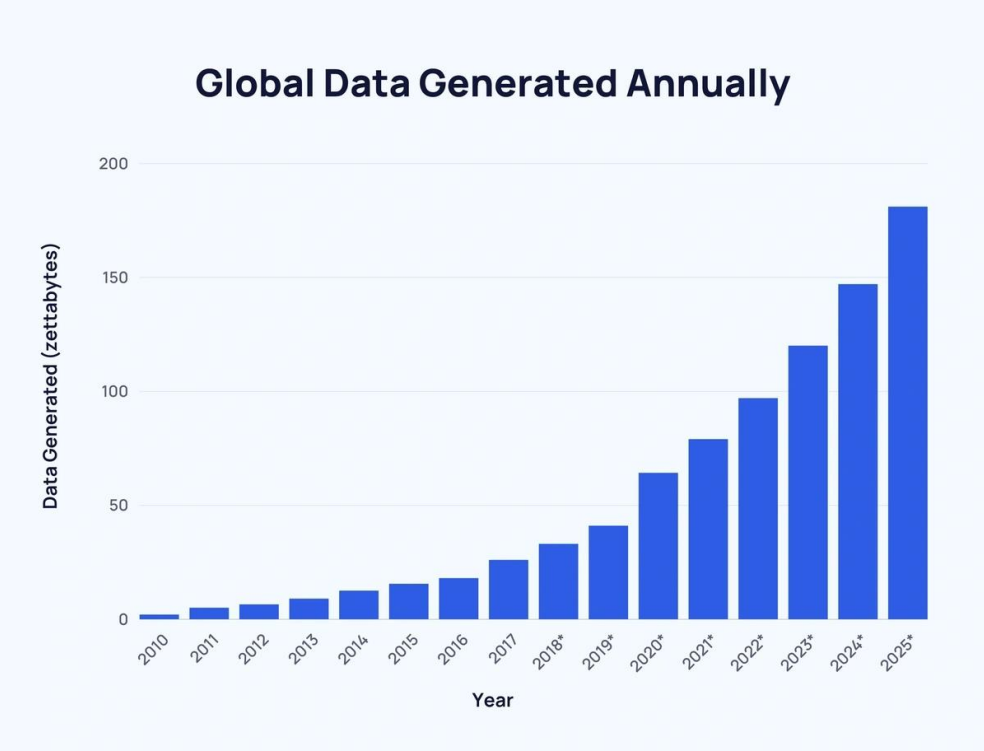
- In 2014, Oscars-host Ellen DeGeneres' "celeb selfie" tweet that was viewed **26 million times** across the Web during a 12-hour period.
- **More than one billion hours** of TV shows and movies are streamed from Netflix per month.
- Walmart, handles **more than 1 million customer transactions** every hour, feeding databases estimated at **more than 2.5 petabytes**. (the equivalent of 167 times the books in America's Library of Congress)
- Facebook, is home to **40 billion photos**.

Growth in Data generation

The total **amount of data created, captured, copied and consumed globally** has been **exponentially increasing**.

In 2020, the amount of data created & replicated was higher than expected caused by the increased demand due to the pandemic.

Up to **2025**, global data creation is projected to grow to **more than 180 zettabytes**.



https://explodingtopics.com/blog/data-generated-per-day

Data generation: Conclusion

- The world is entering the **zettabyte era**, where data is not just massive in volume but also increasingly **complex, unstructured**, and **real-time**.
- This explosion of data underlines the **critical role of Data Science, AI, and Cloud Computing** in managing, analyzing, and extracting value from such vast information sources.

How much of data is put into use?

Though there is a huge amount of data getting generated each day, it shall serve no purpose if it is left unused.

This can further lead to **information overload** where there is an overabundance of information but it is not put into work due to lack of time, resources, understanding of the information, irrelevance of the information or other reasons.

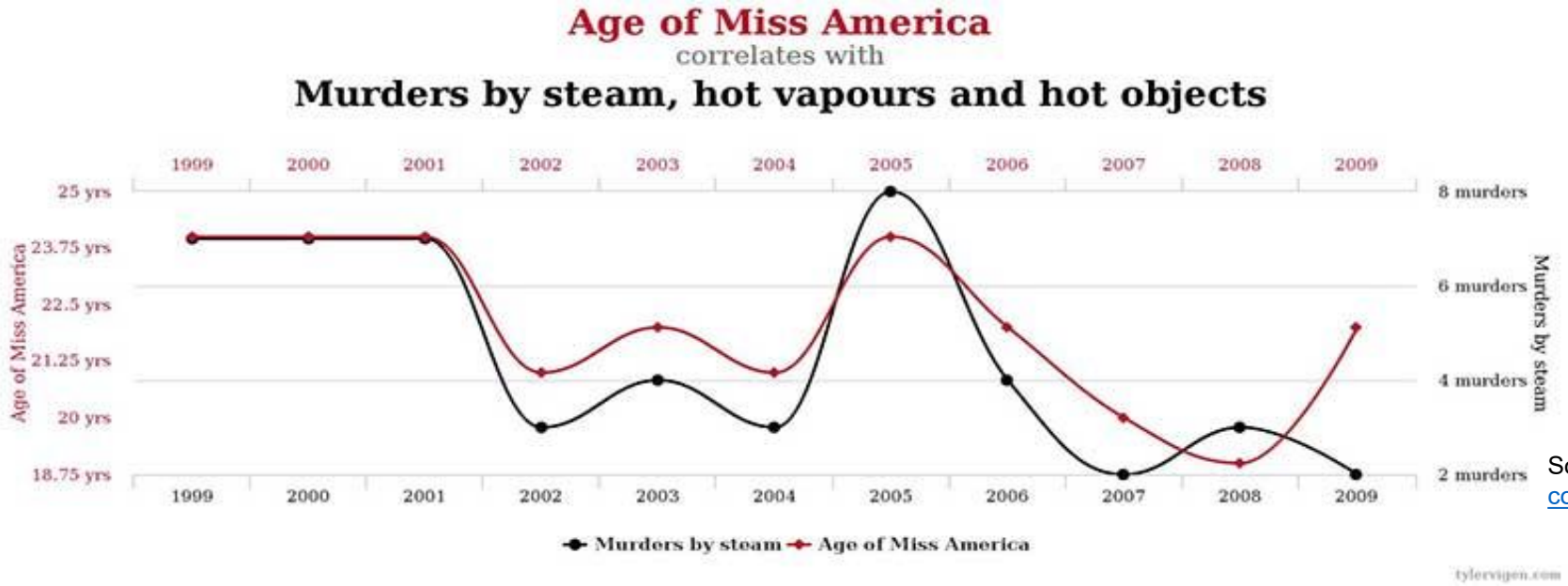
Thus, it is important to **understand the data and know how to utilize it** in the right manner.

But is data all we need?

The graph below shows a **cause & effect relationship** between 'Age of Miss America' and 'Murders by steam, hot vapour and hot objects' which **practically doesn't seem correct**.

Thus, we see that the presence of **interesting patterns need not imply their correctness**.

Blindly applying various processes and techniques on data can result in incorrect inferences.



Source: https://i2.wp.com/boingboing.net/wp-content/uploads/2016/02/chart.jpg?fit=800%2C315&ssl=1

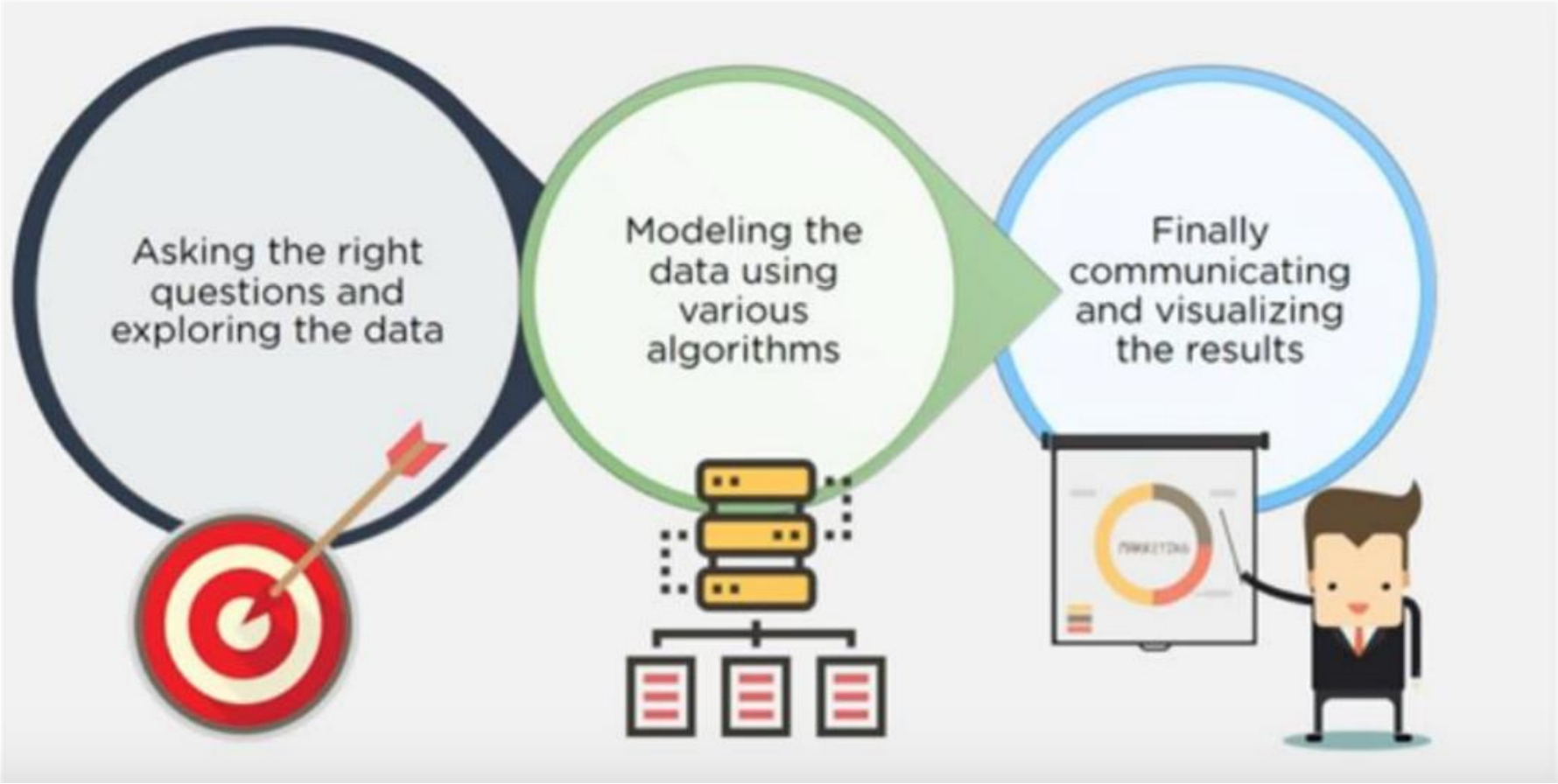
Learn how to use data

The above examples help us understand that we need to **learn how to utilize and handle** the available data in the right manner to be able to **arrive at correct results** and **draw meaningful inferences**.

- ➔ **Explore**: identify patterns
- ➔ **Predict**: make informed guesses
- ➔ **Infer**: quantify what you know

MATHEMATICS FOR COMPUTER SCIENCE ENGINEERS

Learn how to use data



Source:slidesharecdn.com

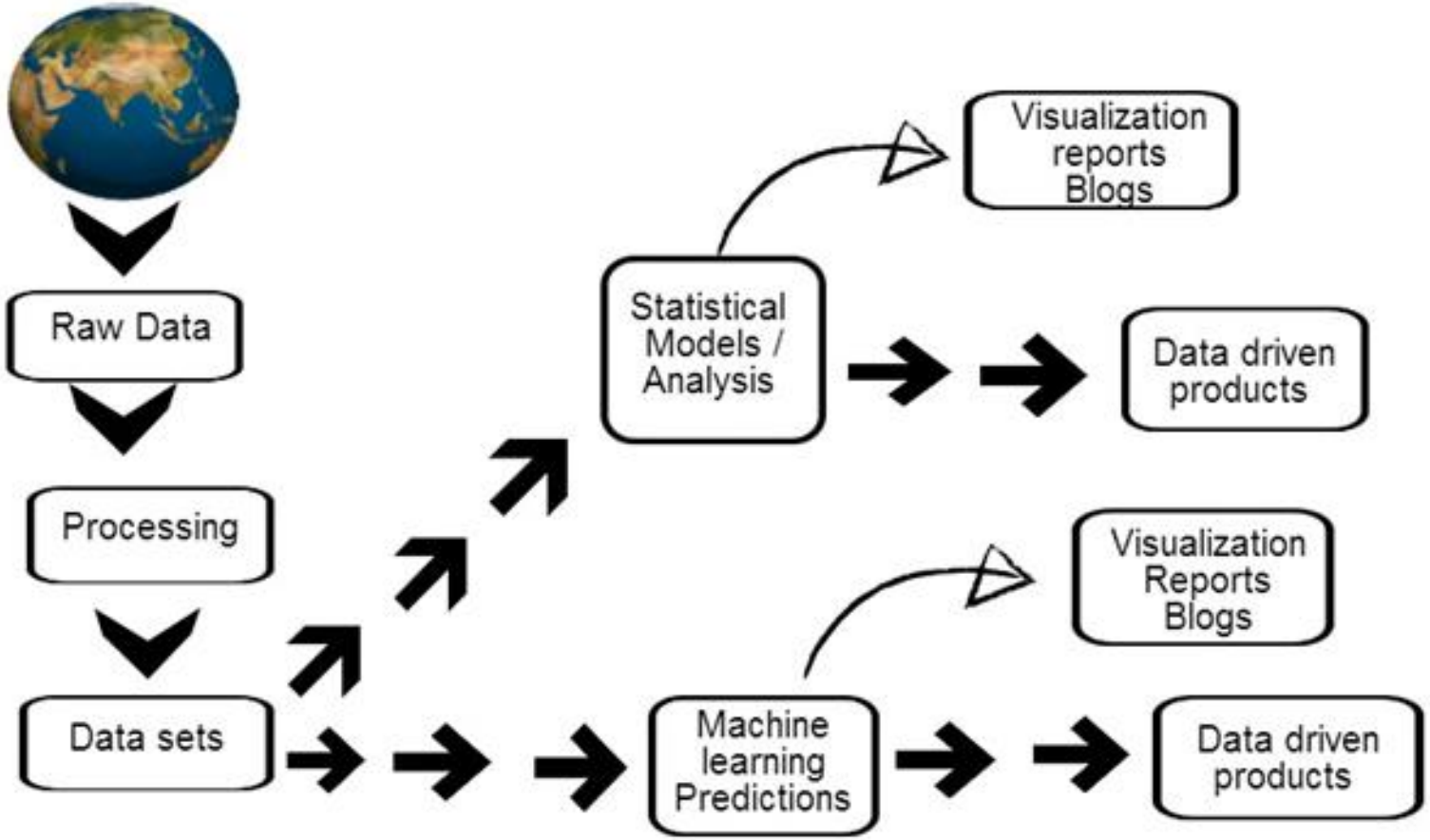
Data Scientist

Data Scientists in simple words are those who make sense out of all the data that are available and figure out the things that can be done with it.

Source: proschoolonline.com

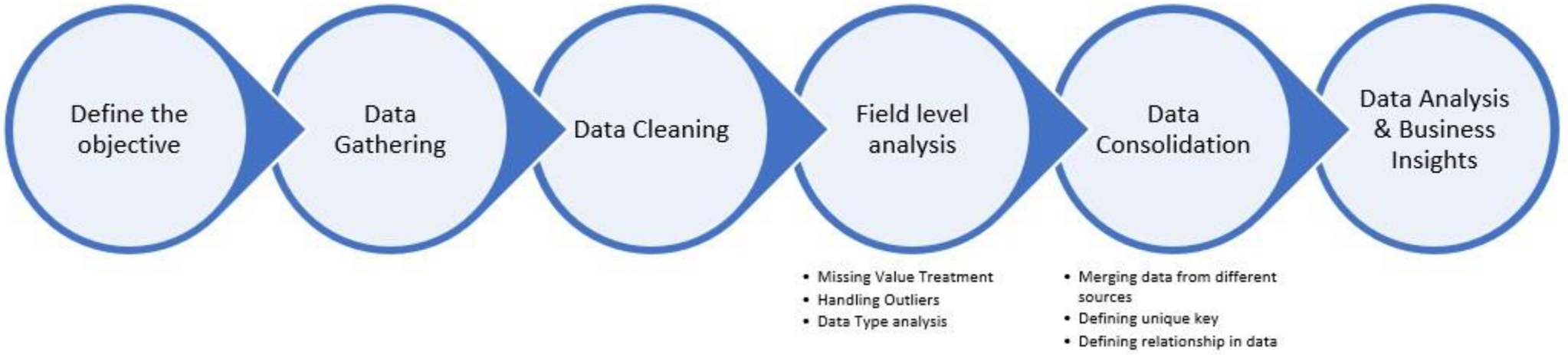
What does a Data Scientist do?

They are responsible for **collecting, analyzing, modelling and interpreting large amounts of data**. Their role combines Computer Science, Mathematics, Statistics etc.



Source: https://edvancer.in/wp-content/uploads/2015/11/76c99311fc4be19bf4353cfc3c2e94b2.png

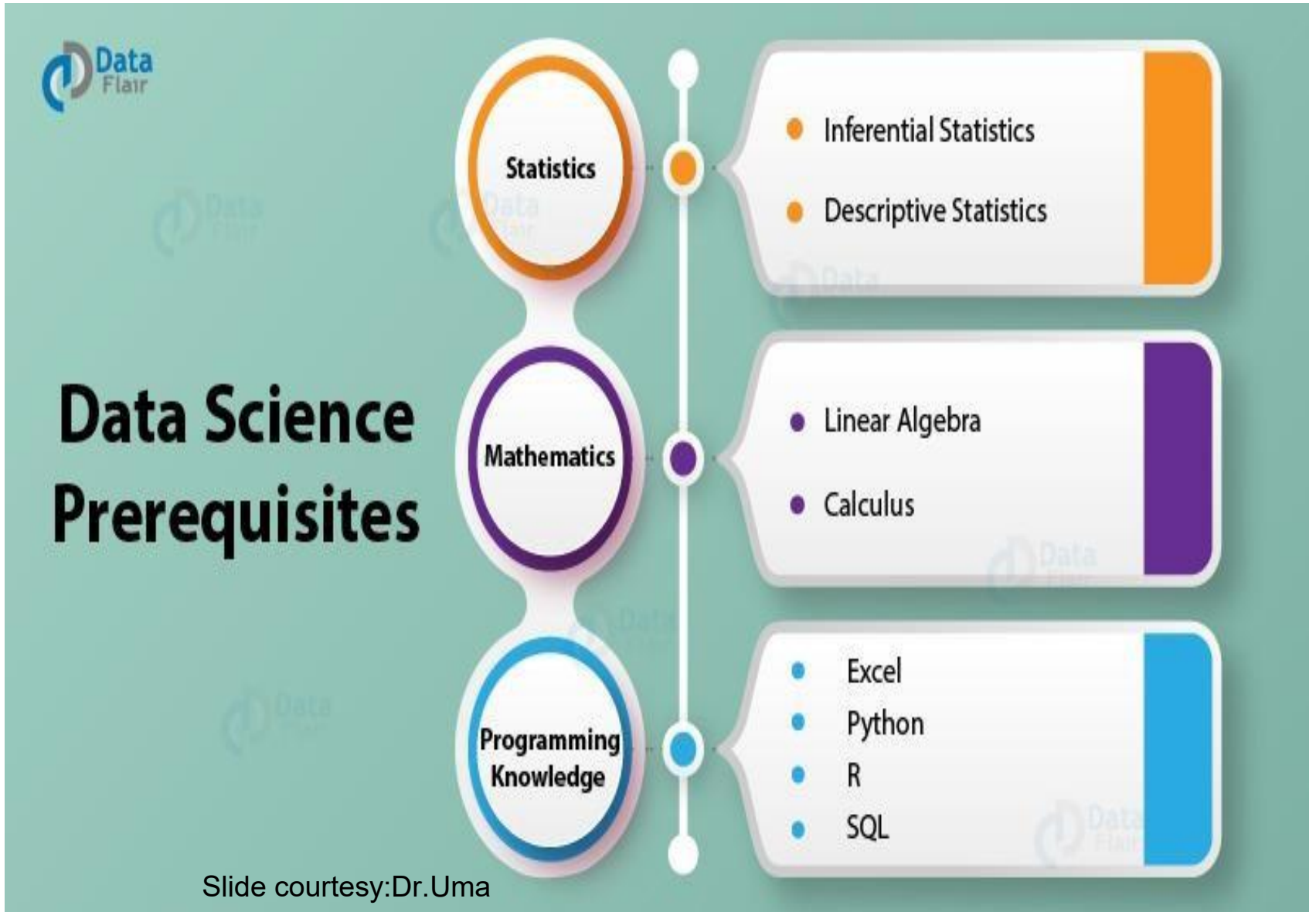
What does a Data Scientist do?



Source: medium.com

Slide courtesy:Dr.Uma

Prerequisites for a Data Scientist



Slide courtesy:Dr.Uma

Source: data-flair.training

Demand for Data Scientist

The demand for **Data Scientists in 2025** is expected to be **exceptionally strong**, both **globally** and **in India**, due to the ever-increasing volume of data and the rise of AI-powered decision-making across sectors.

Global Demand (2025)

- According to **The U.S. Bureau of Labor Statistics (BLS)**:

Data scientist roles will grow by 35% between 2022 and 2032, adding more than 59,000 new jobs in the U.S. alone.

- **World Economic Forum (Future of Jobs Report)** predicts:

Data science and AI roles will be among **the top 10 fastest-growing job categories** through 2025.

- **Estimated global openings by 2025**:

11.5 million+ data-related roles, including Data Scientists, Data Engineers, ML Engineers, and AI specialists.

Demand for Data Scientist

In India-Specific Demand (2025)

- NASSCOM and AIM research (2024-25) reports:

India is projected to have **over 1.5–2 lakh** unfilled roles in Data Science & Analytics in 2025 due to a **skills gap**.

- **Top hiring sectors in India**:
  - IT and Software Services
  - Banking, Financial Services, and Insurance (BFSI)
  - E-commerce and Retail
  - Healthcare and Pharma
  - Manufacturing and Logistics
  - EdTech and Startups
- **Top cities for hiring**: Bengaluru, Hyderabad, Pune, Mumbai, New Delhi

How is it different from what Statisticians have been doing?

Both Statisticians and Data Scientists work closely with data.

- Statisticians **use mathematical equations and statistical models** to analyze data and arrive at conclusions.
- Data Scientists however f**ocus on delivering actionable results** and sometimes need to deploy the model to the production system.

Data Science vs Data Analysis

Data Science:

•**Scope:**

Data science is a broader, interdisciplinary field that encompasses data analysis, data engineering, and machine learning.

•**Focus:**

It focuses on extracting knowledge, building predictive models, and developing new algorithms to solve complex problems.

•**Techniques:**

Data scientists utilize advanced statistical modeling, machine learning, and programming skills (like Python or R).

•**Goal:**

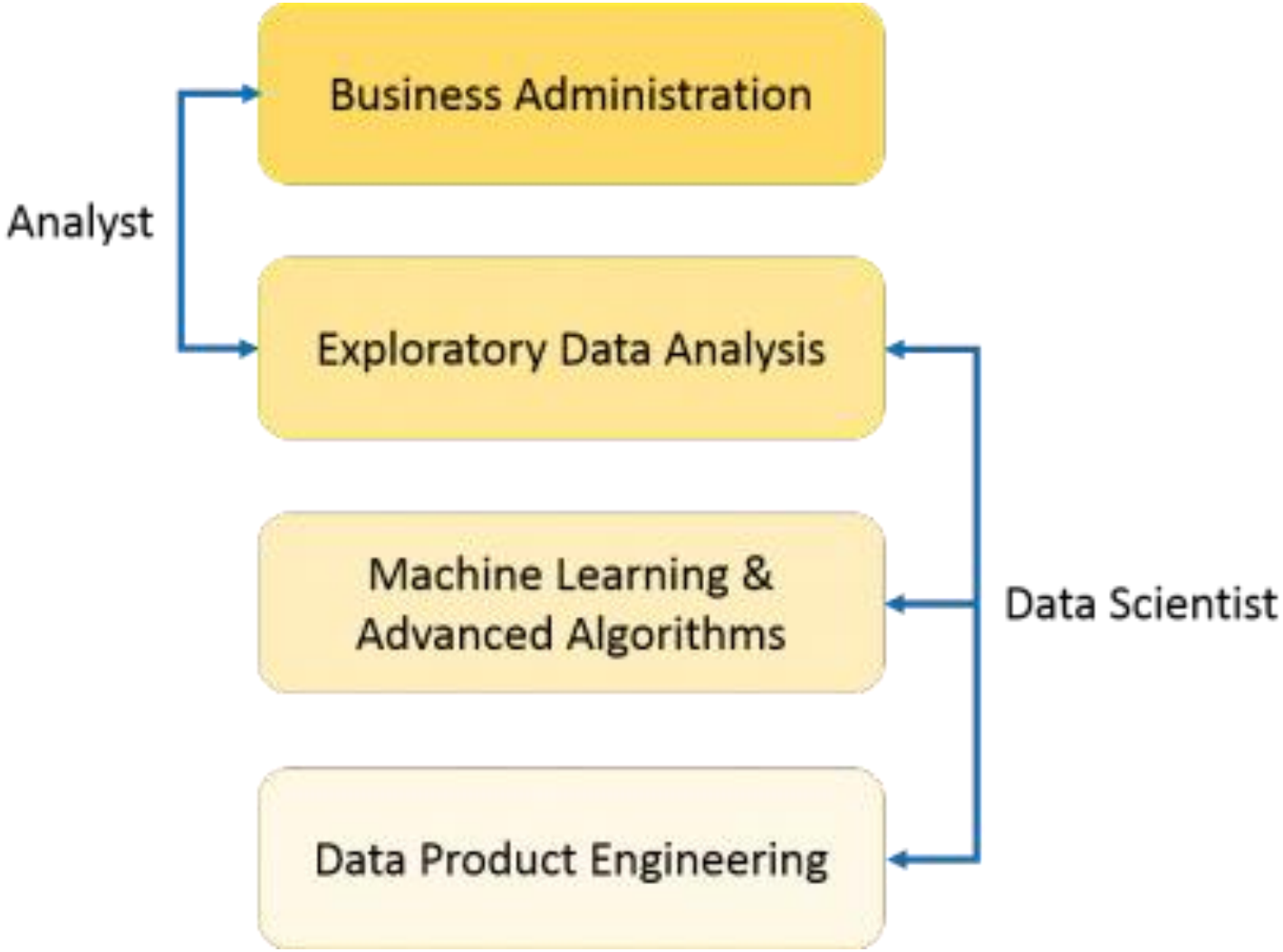
To uncover hidden patterns, predict future trends, and create new data-driven solutions.

•**Example:**

Building a recommendation engine for a streaming service or predicting customer churn using machine learning.

- Data Analysis includes **descriptive analytics and prediction to a certain extent**.

Data Science vs Data Analysis



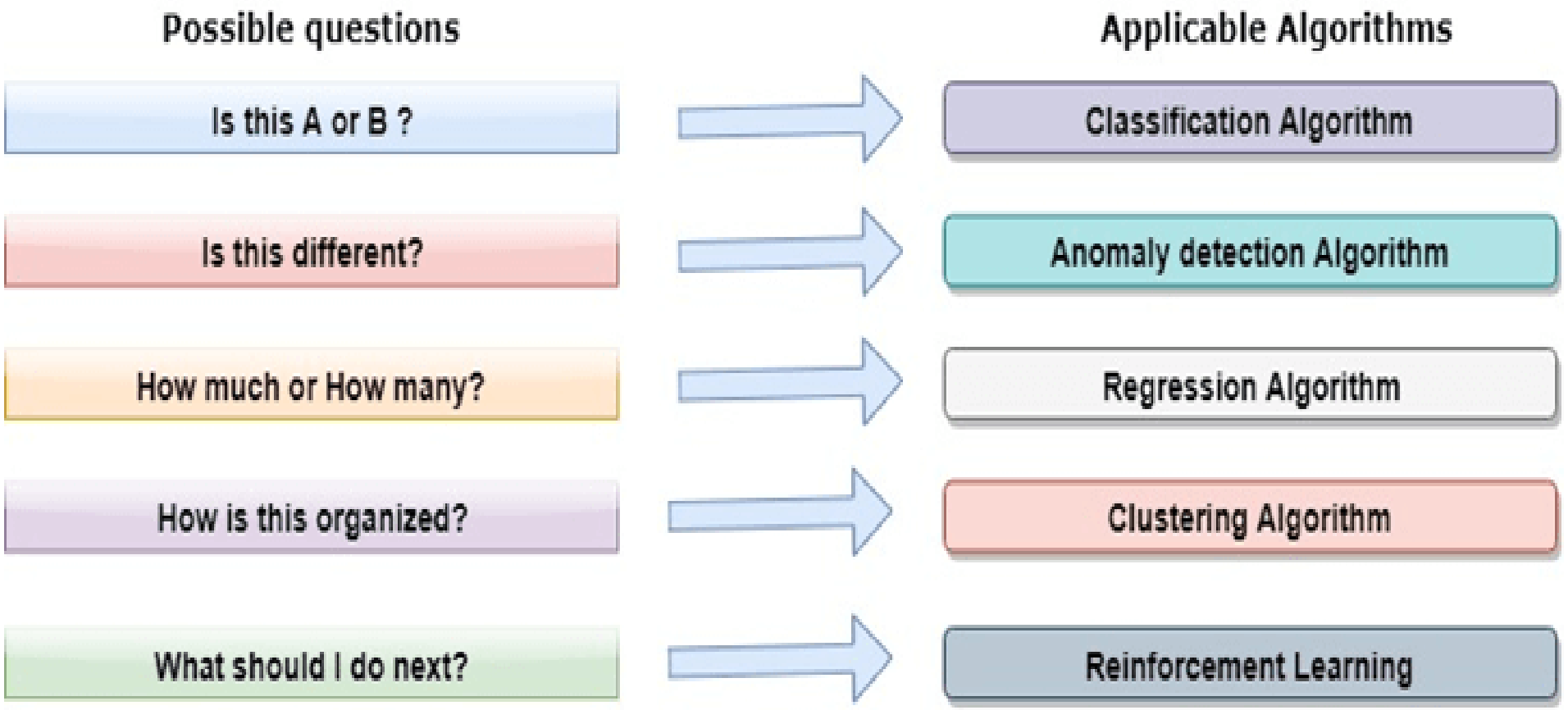
Source:
https://d1jnx9ba8s6j9r.cloudfront.net/blog/wp-content/uploads/2017/01/Data-Analyst-vs-Data-Science-1-422x300.png

Common tasks in Data Science

So Data Science is mainly needed for:

Source: Simplilearn

Common tasks in Data Science



Source: https://static.javatpoint.com/tutorial/data-science/images/how-to-solve-a-problem-in-data-science.png

References

Text Book:

Statistics for Engineers and Scientists, William Navidi,6th Edition , McGraw Hill Education, India

THANK YOU

Dr.Mamatha H R
Professor, Department of Computer Science
mamathahr@pes.edu