STATISTICS FOR DATA SCIENCE

Introduction to Data Science

D. Uma

Department of Computer Science and Engineering

umaprabha@pes.edu

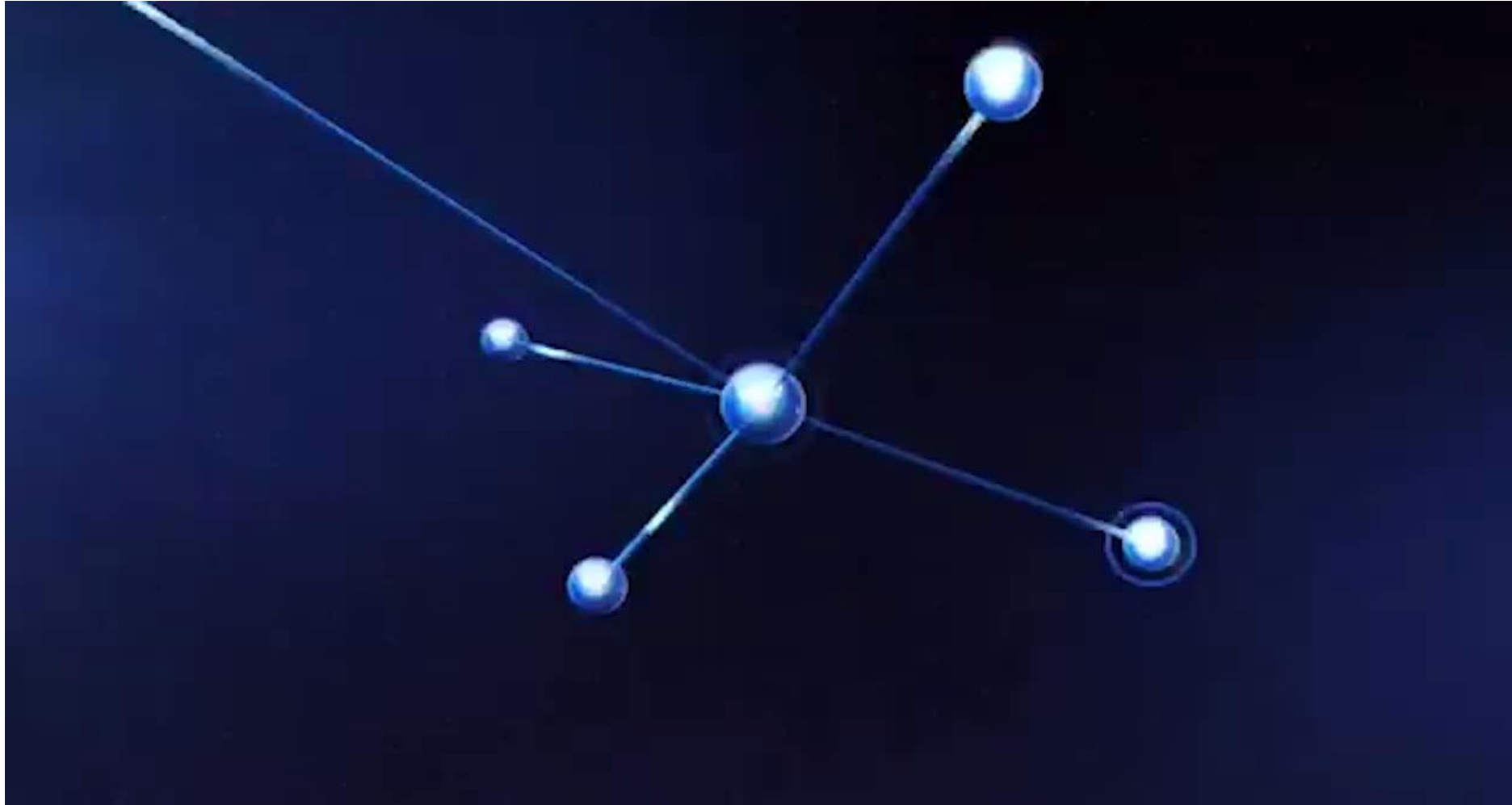# STATISTICS FOR DATA SCIENCE

Introduction to Data Science

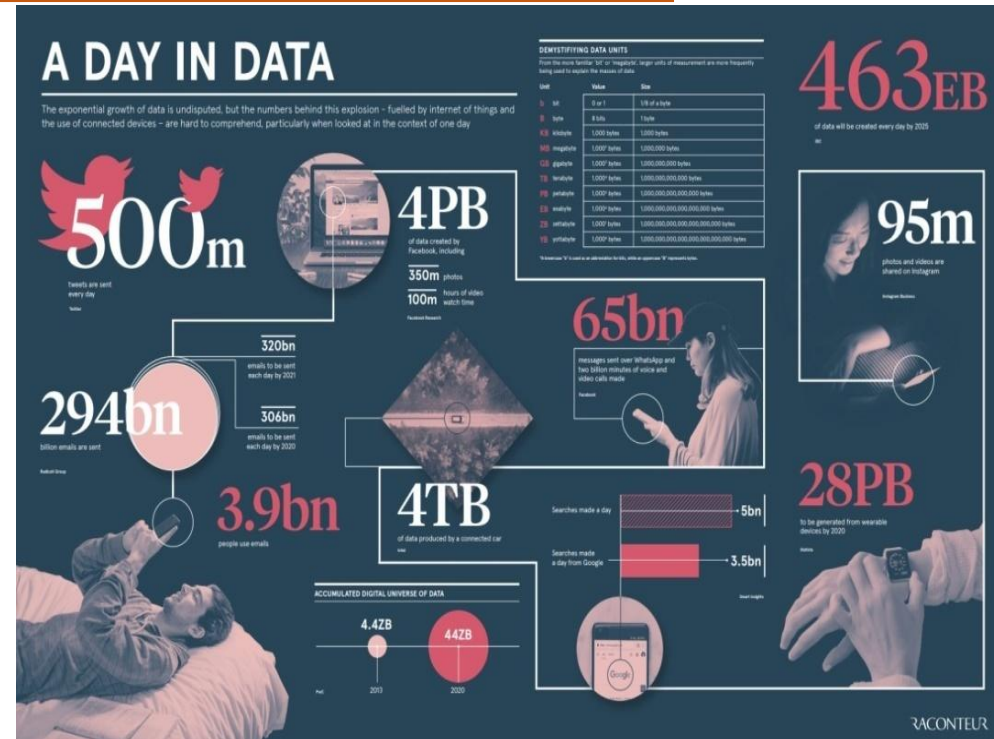**Prof. Uma D**

Department of Computer Science and Engineering

How Data Gets Generated?
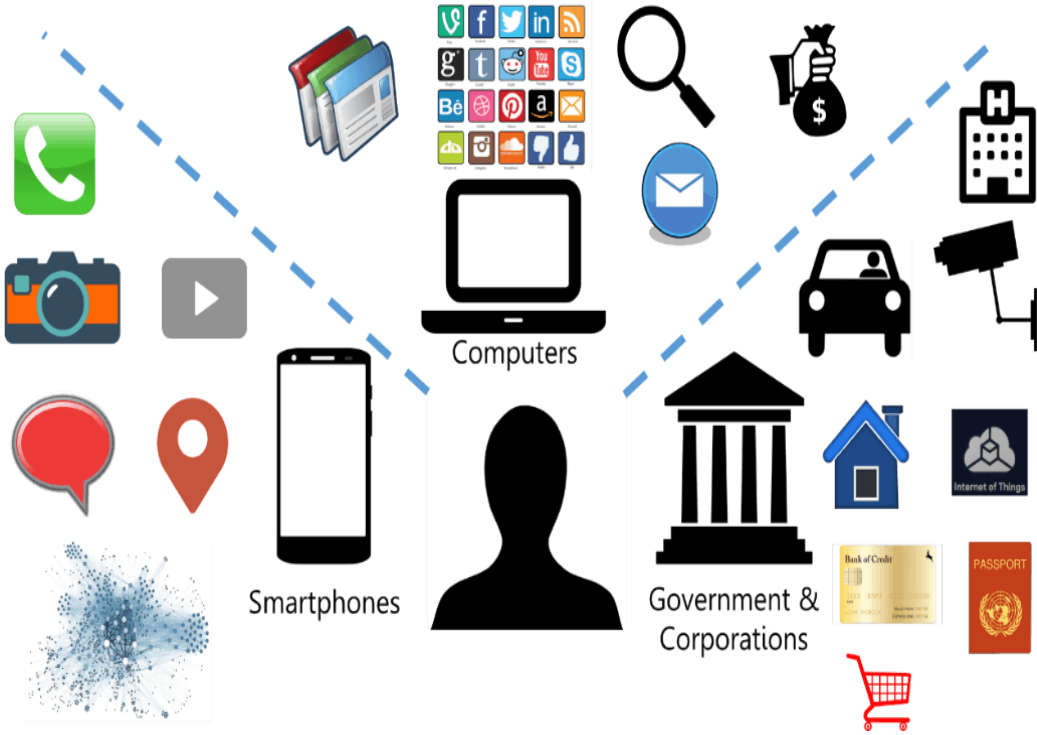
Data Generation: A Day In Data

500 million tweets are sent.

294 billion emails are sent.

4 petabytes of data are created on Facebook.

65 billion messages are sent on WhatsApp.

5 billion searches are made.

By 2025,its estimated that 463 exabytes of data will be created **each day** globally.
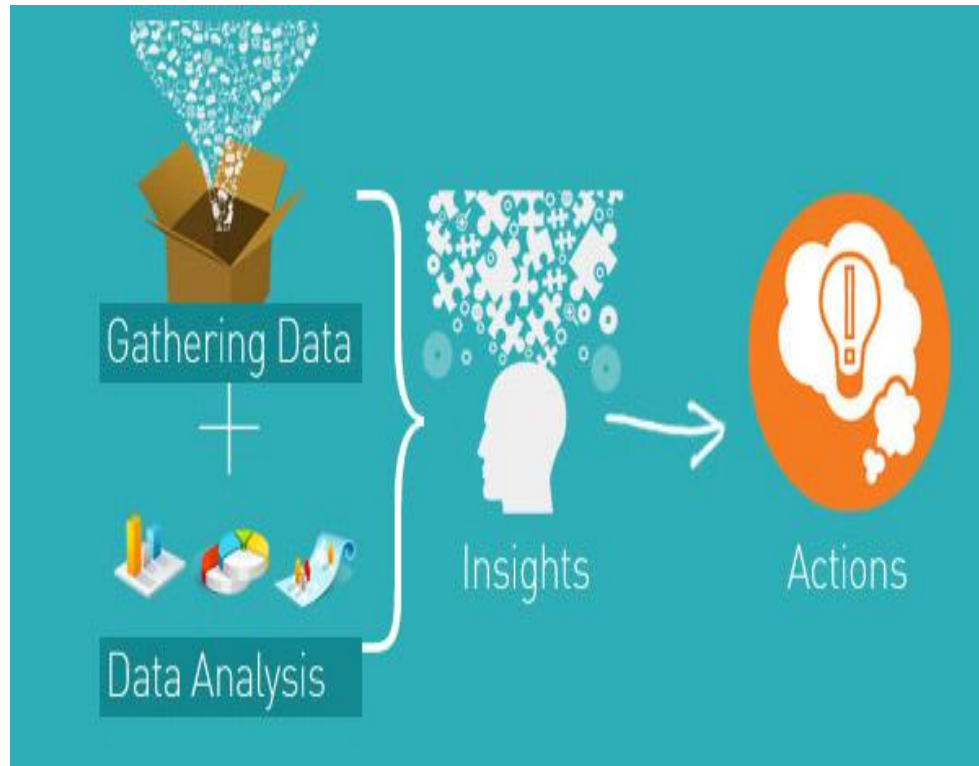
Sources and Forms of Data

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

Data Science

"Where there is data smoke, there is business fire."
— Thomas Redman

Data Science in Self-driving Cars

Source: Netflix
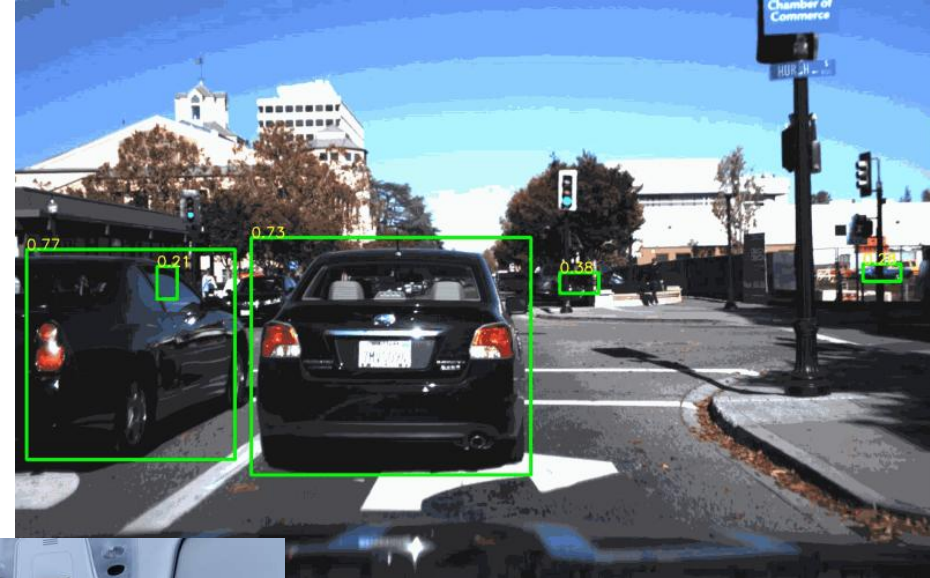
Data Science in Airlines Industry

Dear Flyer, We regret to inform you that your flight has been cancelled due to delay from Airbus on account of engine delivery

1. Due to lack of data available, flights are often delayed or cancelled at the last minute

2. Due to improper route planning customers don't get the flight for desired time and duration

3. Incorrect decisions in selection of right equipment leads to unplanned delays and cancellations

Using Data Science, we can achieve the following:

Route Planning: Whether to schedule direct or connecting flights

Predictive analytics model can be built to foresee flight delays

Promotional offers depending on customer booking patterns

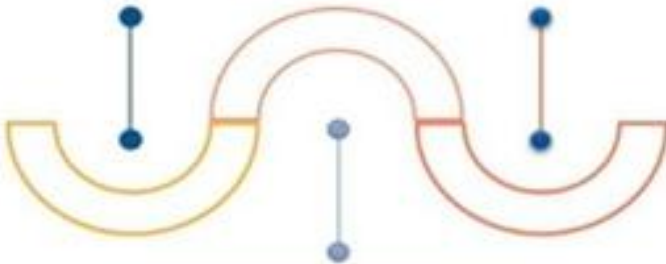Deciding which class of planes to purchase for better performance

ISO-GRAPHICS

01
02
03
04

Logistics companies like FedEx are using Data Science models for operational efficiency

Discover the best routes to ship

The best suited time to deliver
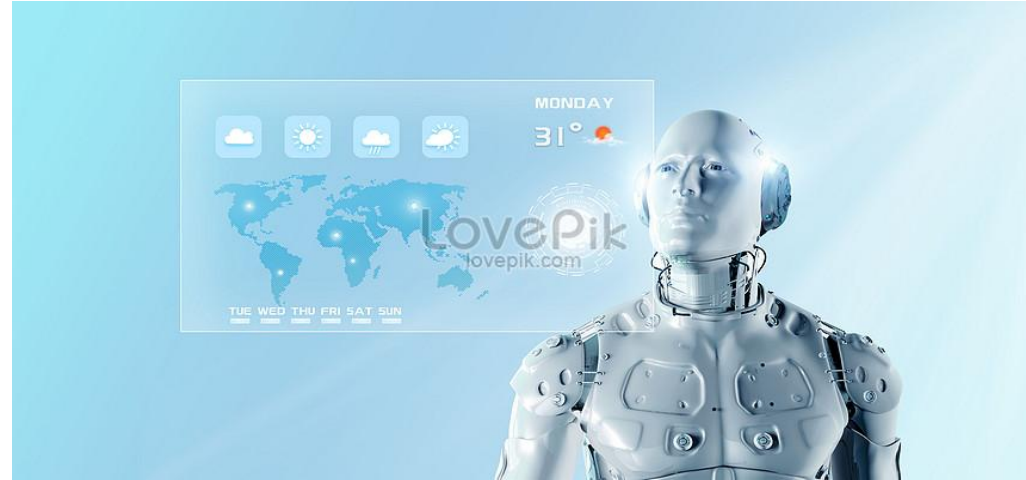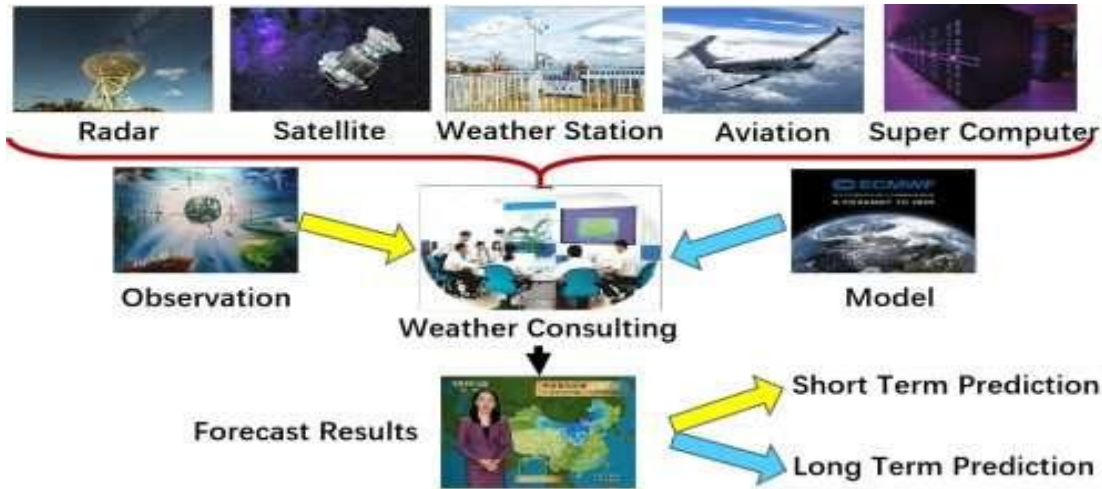
The best mode of transport

Data from ships, aircrafts, radars, satellites can be collected and analyzed to build models.

These models will not only forecast the weather but also help in predicting the occurrence of any natural calamities.

It will help you to take appropriate measures beforehand and save many precious lives.

Source: phys.org/news

Data Science can answer a lot of other questions as well!

Which viewers like the same kind of TV shows?

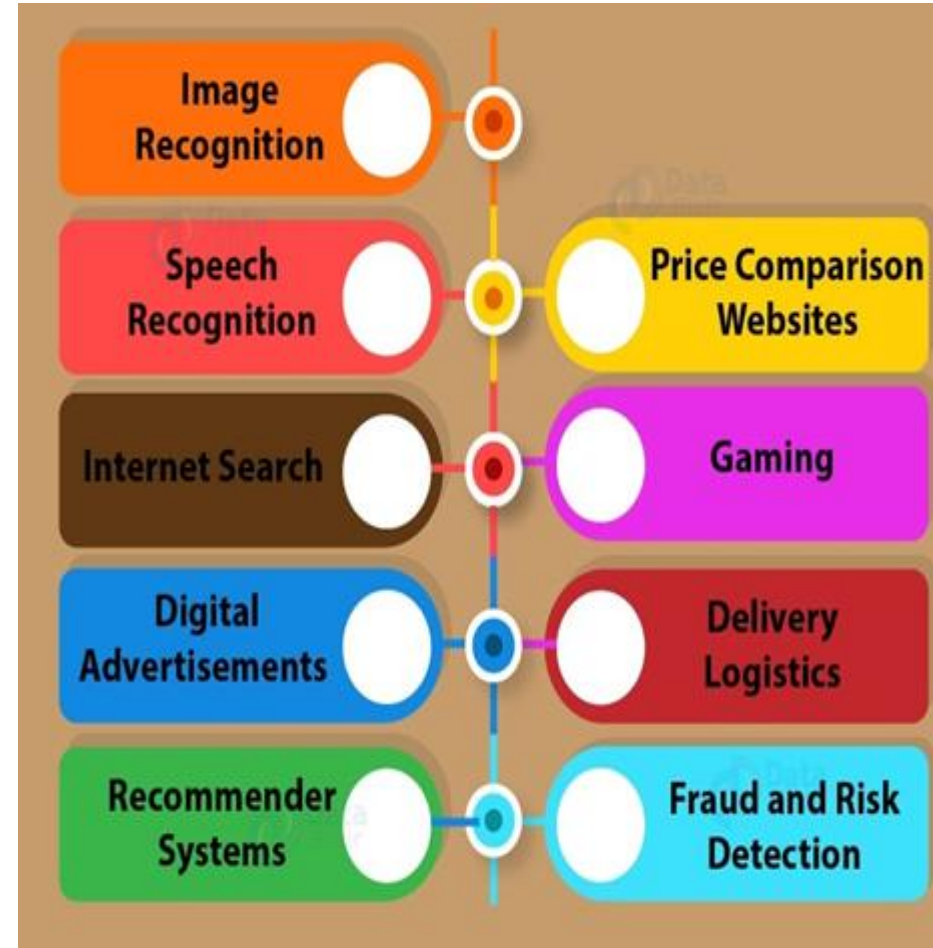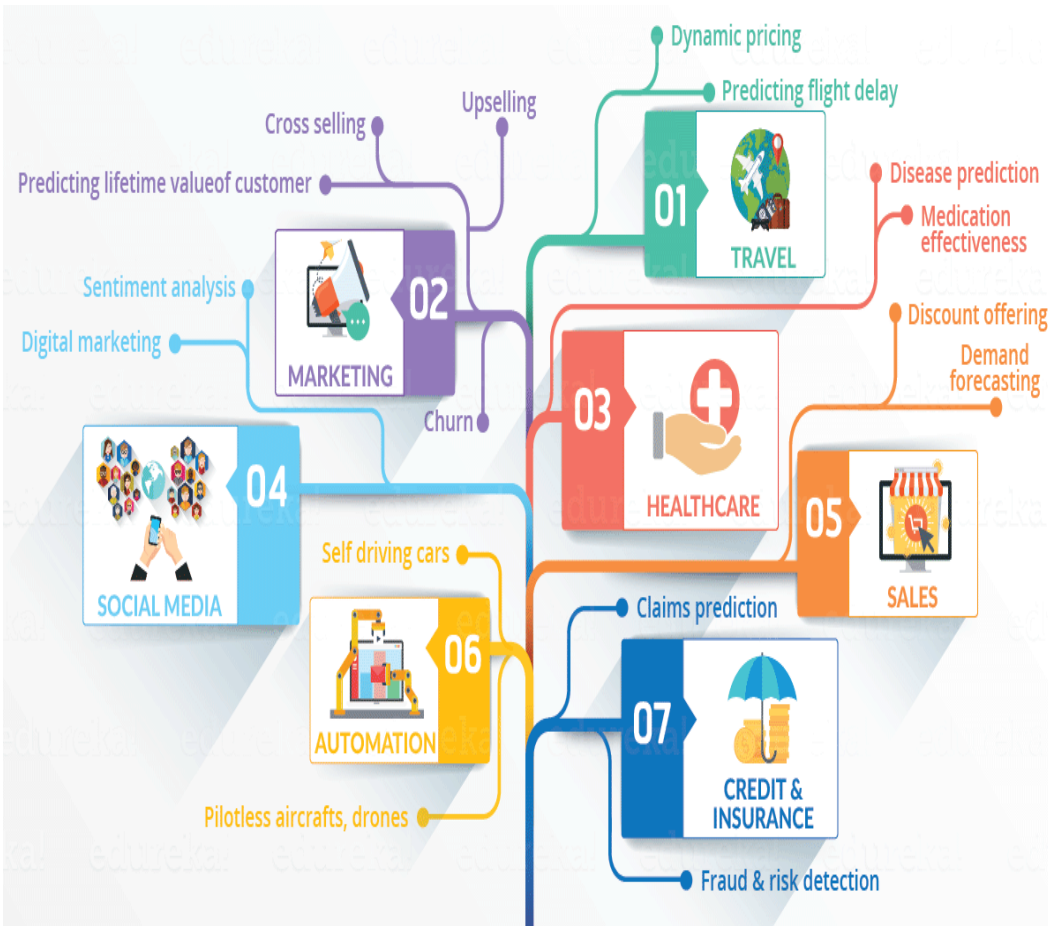Will this refrigerator fail in the next 3 years: Yes or No?

Which route should my cab take so that I reach faster?

Who will win the elections?

Source: Simplilearn

STATISTICS FOR DATA SCIENCE

Applications of Data Science
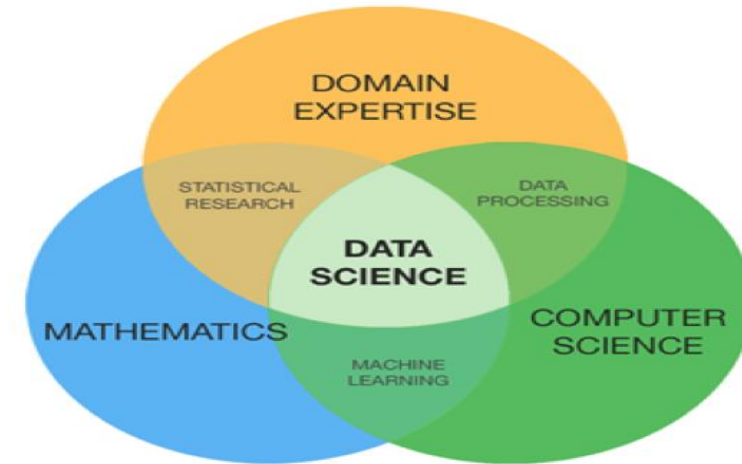
What is Data Science???

Data Science is an inter-disciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from many structured and unstructured data.

What is Data Science???

Data Science is a blend of various tools, algorithms, and machine learning principles with the goal to discover hidden patterns from the raw data.

Data Science is a concept to unify statistics, analysis, machine learning, domain knowledge and related methods in order to understand and analyze actual phenomena with data.
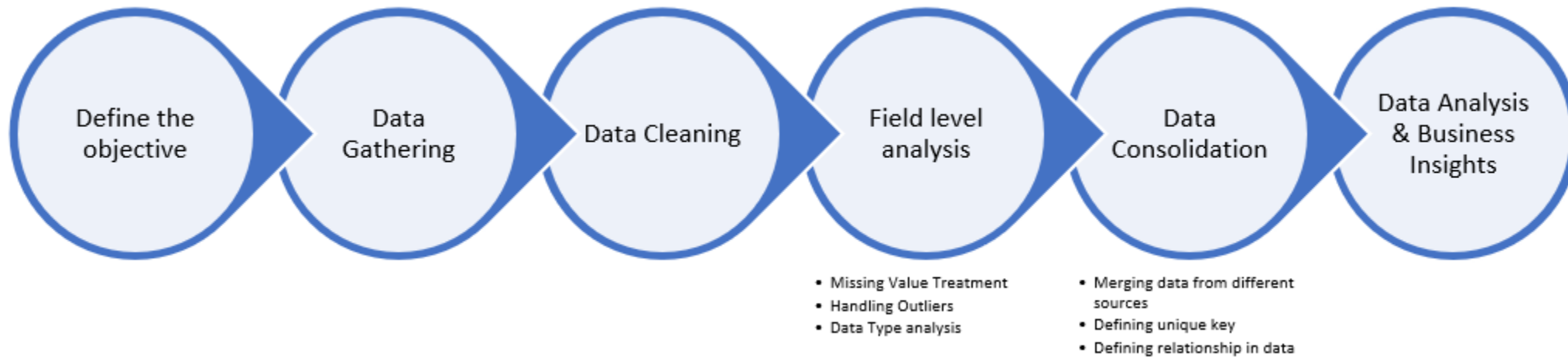
The term "data scientist" was coined as recently as 2008 .

Data Scientist are the people who make sense out of all the data that are available and figure out just what can be done with it.

What Does A Data Scientist Do?

Define the objective	Data Gathering	Data Cleaning	Field level analysis	Data Consolidation	Data Analysis & Business Insights

Field level analysis:
- Missing Value Treatment
- Handling Outliers
- Data Type analysis

Data Consolidation:
- Merging data from different sources
- Defining unique key
- Defining relationship in data

STATISTICS FOR DATA SCIENCE

Data Scientist

Asking the right questions and exploring the data

Modeling the data using various algorithms

Finally communicating and visualizing the results
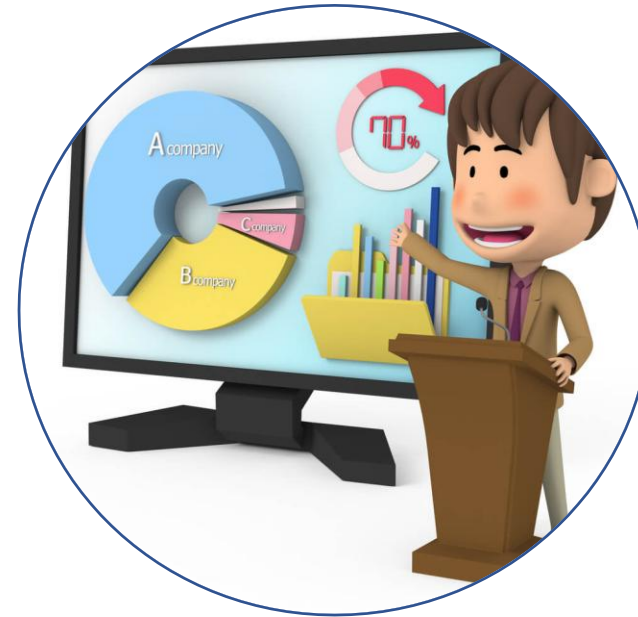
Prerequisites For Data Scientist

CURIOSITY

COMMON SENSE

COMMUNICATION SKILLS
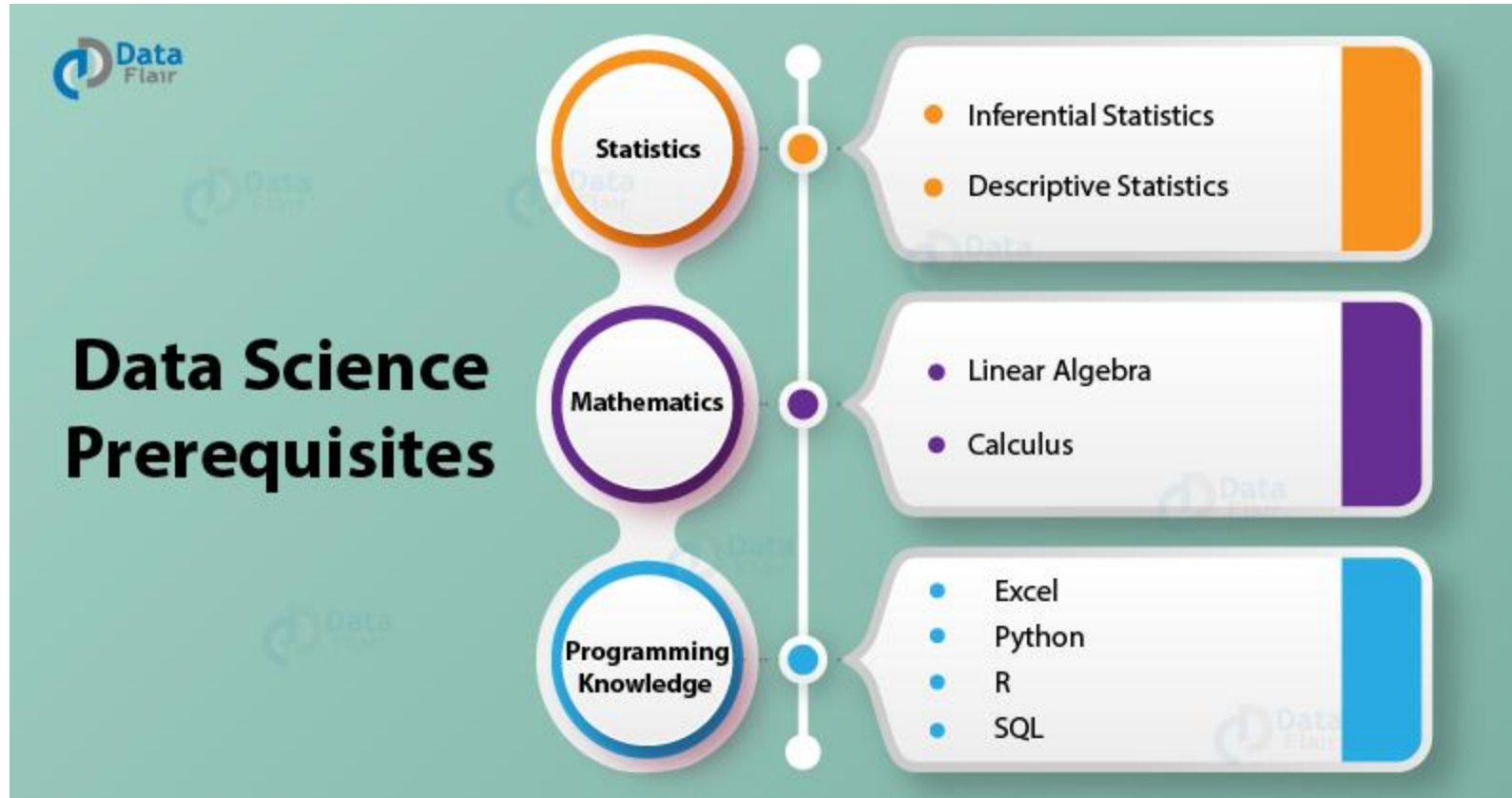
Prerequisites For Data Scientist

28%	4,524	$120,931	#1
Demand Increase by 2020	Number of Job Openings	Average Base Salary	Best Job in America 2016, 2017, 2018

Sources : Glassdoor, Forbes

Simple Linear Regression
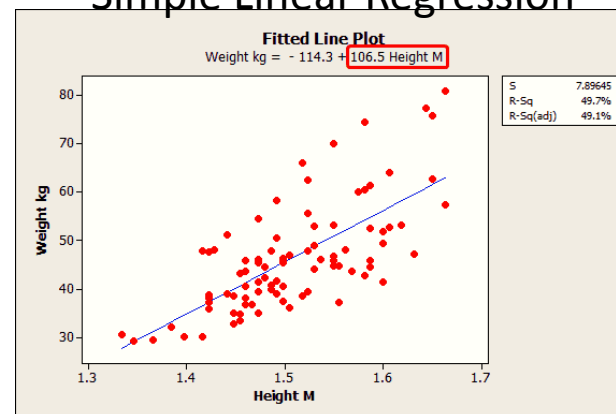
THANK YOU

D. Uma

Department of Computer Science and Engineering

umaprabha@pes.edu

+91 99 7251 5335