

# **DATA SCIENCE WITH PYTHON**

## *INTRODUCTION*

**APPAPPAN MURUGESAN**

**PHONE: 7010774329**

**WHATSAPP: 8870483557**

*DATA*

*SCIENCE*



*DATA*

*SCIENCE*

- Distinct pieces of information
- Data is any sequence of one or more symbols given meaning by specific act of interpretation. Data requires interpretation to become information.
- Data is a set of values of qualitative or quantitative variables



# TYPE OF DATA

## Unstructured Data



### HIGHLIGHTS

- In a moment for the ages, Dhoni clobbered Kulasekara high over the long-on fence to spark widespread celebrations in India
- The architects of India's triumph on that night in Mumbai were Zaheer, Gambhir and Dhoni
- Tendulkar called it the proudest moment of his life

*Indian cricketers pose with the World Cup trophy after beating Sri Lanka in the final (Getty Images)*

# TYPE OF DATA

## *Structured Data*

Subcode	Sub Name	Internal	External	Credits Total	Pass/Fail
T0221	POWER SYSTEMS - I	15	50	4	Pass
T0222	ELECTRICAL MACHINES - II	16	42	4	Pass
T0223	IC & PULSE AND DIGITAL CIRCUITS (LAB)	24	47	2	Pass
T0224	ELECTRICAL MACHINES - I (LAB)	22	44	2	Pass
V0122	ENVIRONMENTAL STUDIES	17	36	4	Pass
V0221	MANAGERIAL ECO. & FINANCIAL ANALYSIS	16	50	4	Pass
V0222	LINEAR & DIGITAL IC APPLICATIONS	15	44	4	Pass
V0223	CONTROL SYSTEMS	17	60	4	Pass

# TYPE OF DATA

## *Semi Structured Data*

**{JSON}**  
JavaScript Object Notation



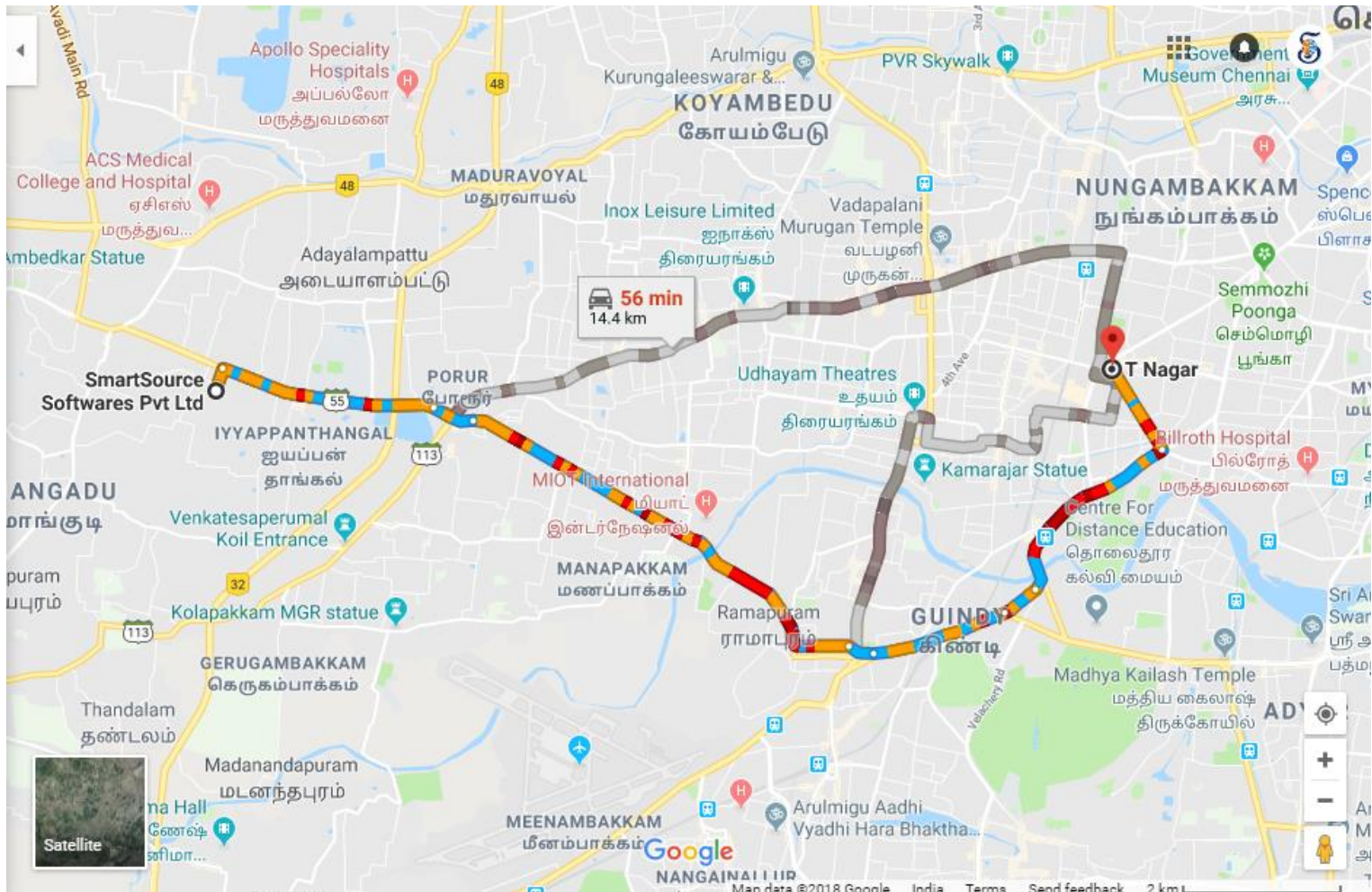
# WHY DATA NEED?



???



# WHY DATA NEED?





# WHY DATA NEED?

## MARK YOUR CALENDAR

**January 15:** Makara Sankranti

**January 26:** Republic Day

**February 13:** Maha Shivarathri

**March 29:** Mahaveer Jayanti

**March 30:** Good Friday

**April 18:** Basava Jayanti

**May 1:** May Day

**June 16:** Kutub-E-Ramzan

**August 15:** Independence Day

**August 22:** Bakrid

**September 13:** Vinayaka Vrata

**September 21:** Last day of  
Moharram

**October 2:** Gandhi Jayanti

**October 8:** Mahalaya Amavasye

**October 18:** Maha Navami, Ayudh  
Puja

**October 19:** Vijayadashami

**October 24:** Valmiki Jayanti

**November 1:** Kannada Rajyotsava

**November 6:** Naraka Chaturdashi

**November 8:** Bali Padyami

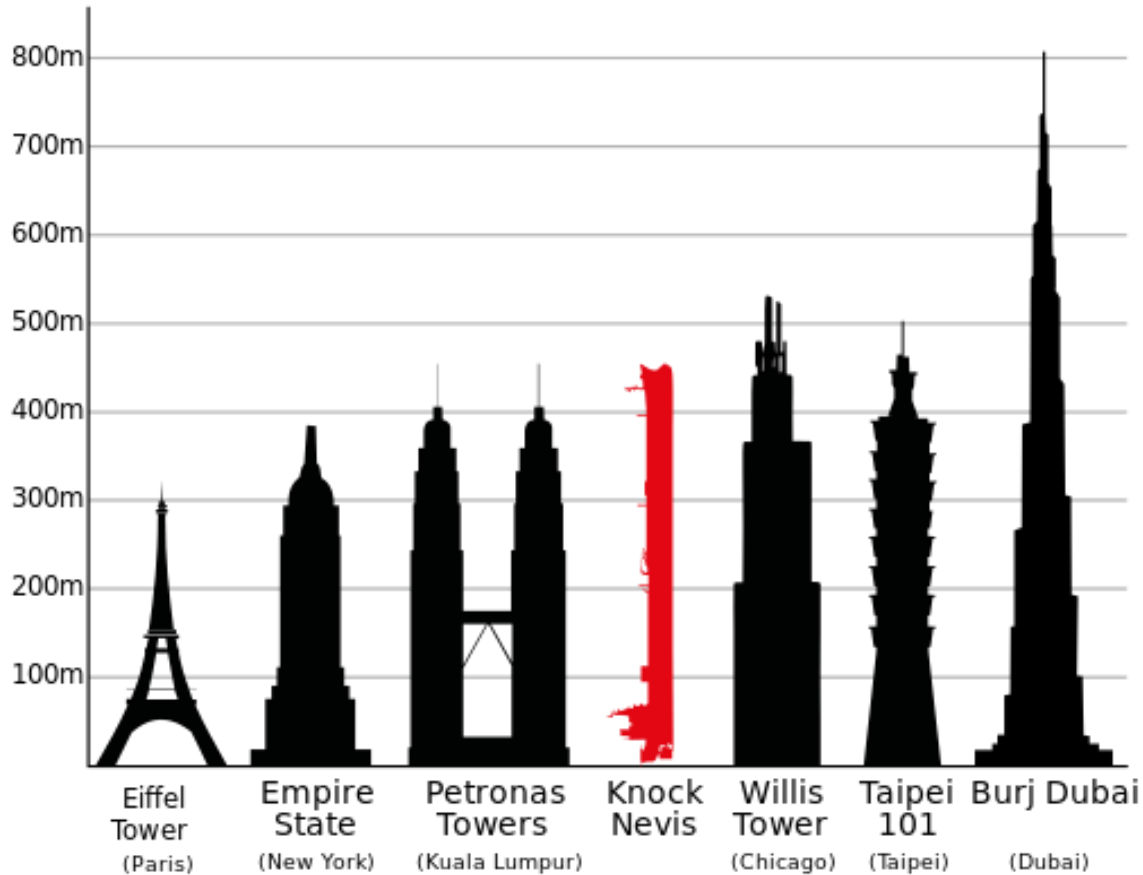
**November 21:** Eid-Milad

**November 26:** Kanakadasa  
Jayanti

**December 25:** Christmas Day

*Plan*

# WHY DATA NEED?



*Compare*

# WHY DATA NEED?

Oil prices in Rupees	1-Apr	2-Apr
Mumbai	81.59	81.69
Delhi	73.73	73.83
Kolkata	76.44	76.54
Chennai	76.48	76.59
Diesel Price	1-Apr	2-Apr
Mumbai	68.77	68.89
Delhi	64.58	64.69
Kolkata	67.27	67.38
Chennai	68.12	68.24

*Under  
stand*

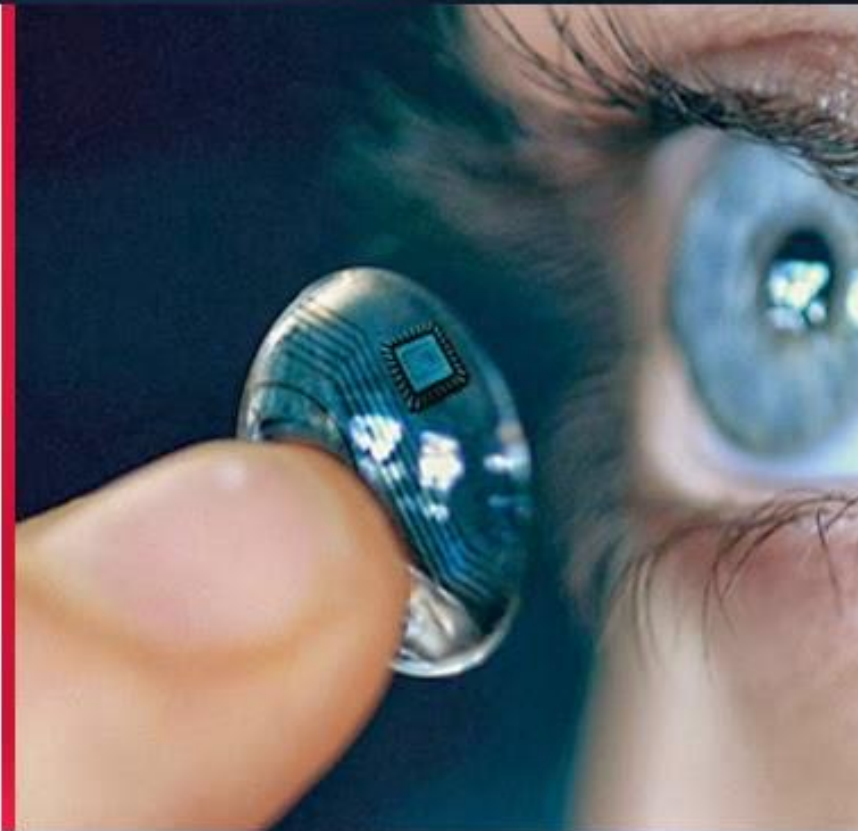
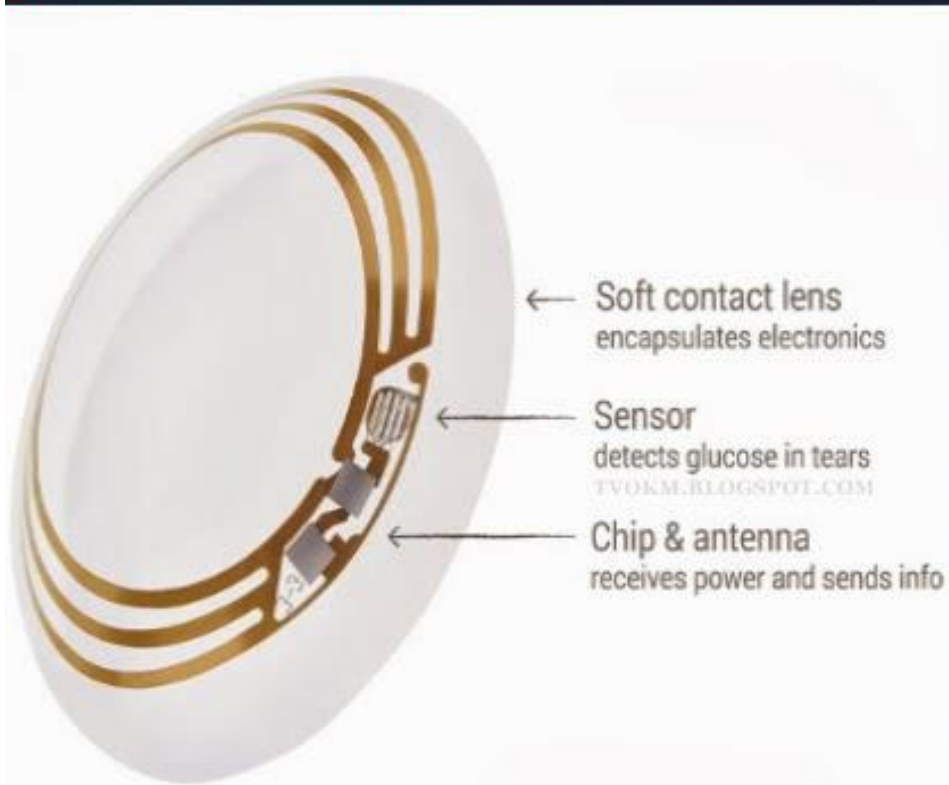
# WHY DATA NEED?



# BEST SOURCE

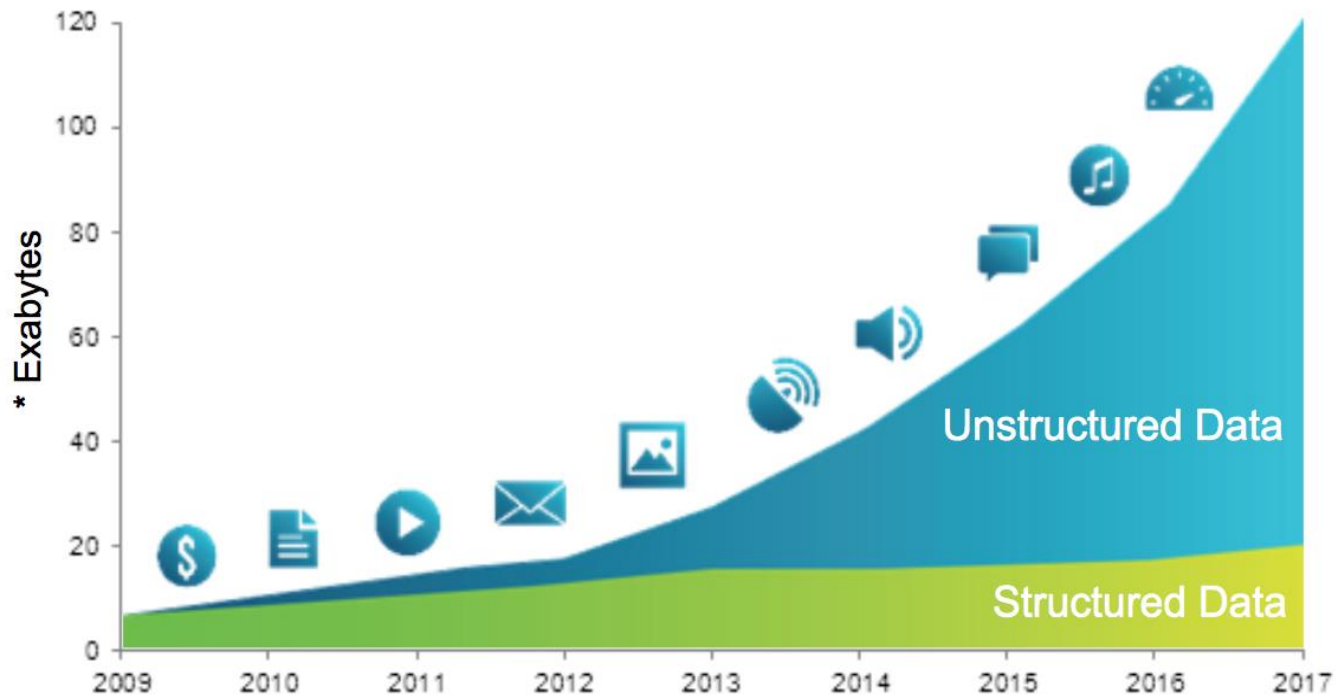


## Google Smart Content Lenses



# Data Growth

*Problem - Traditional and Legacy Storage Designed for Transactional, Not Unstructured Data*



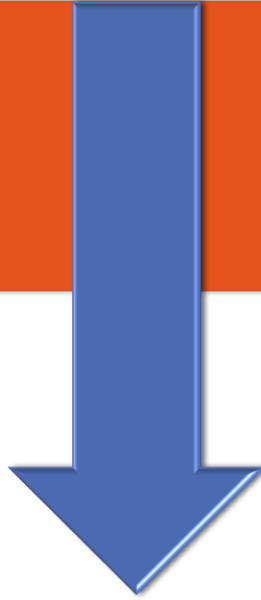
\*1 exabyte = 1,000 petabytes = 1 million terabytes = 1 billion gigabytes **Source: IDC**

- Unstructured data growth of
- 60–80% per year
- creates Web-scale storage needs



*DATA*

*SCIENCE*



*DATA*

*SCIENCE*

# SCIENCE

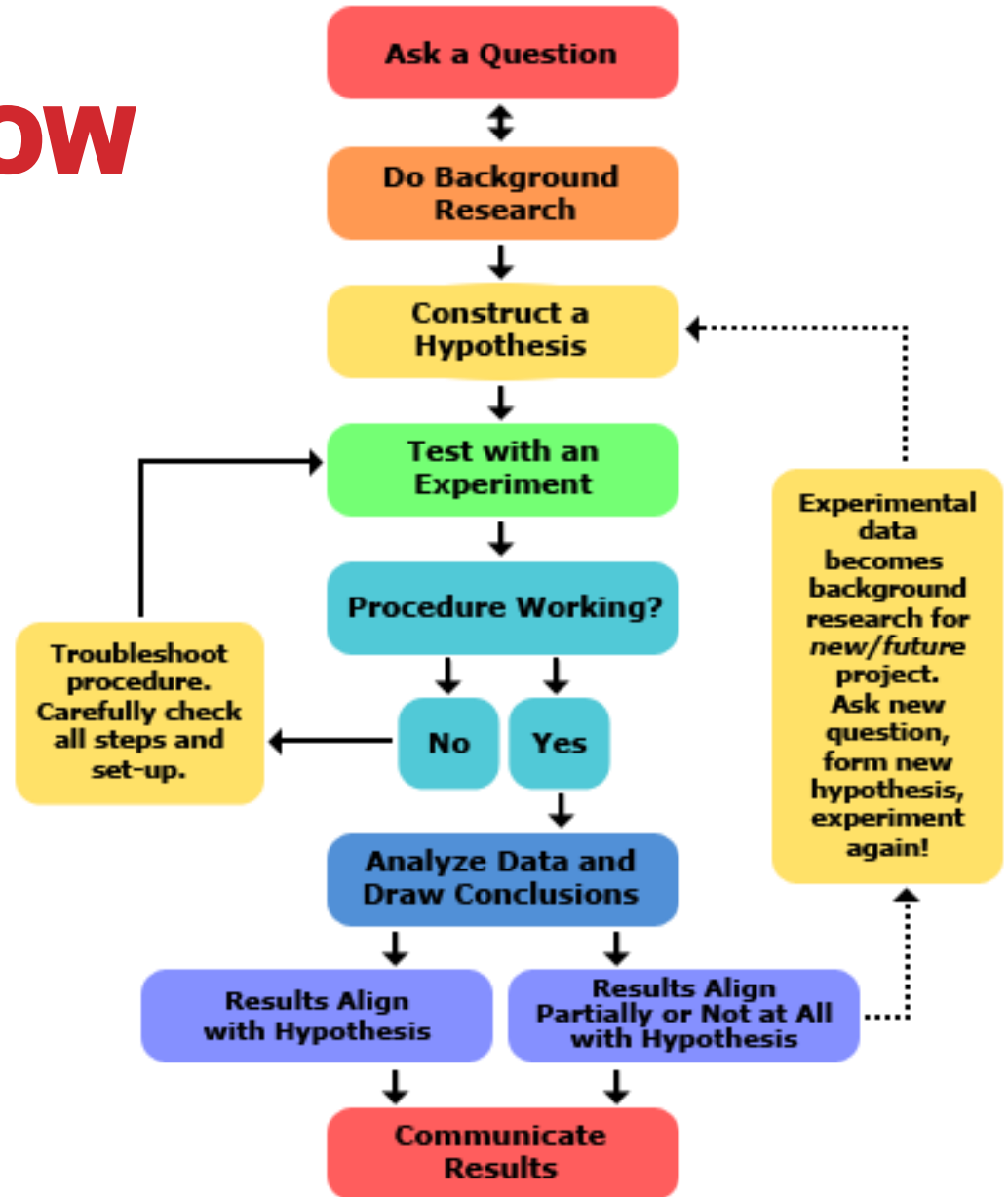
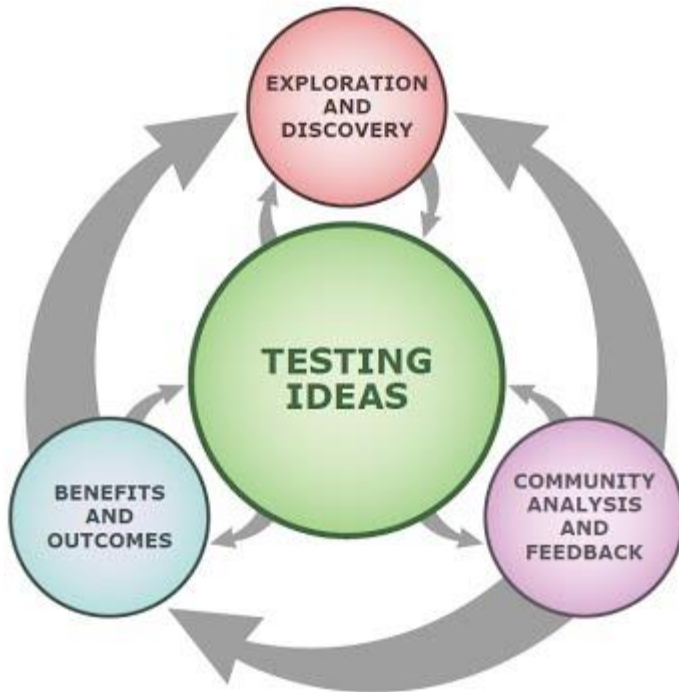
- *The intellectual and practical activity encompassing the systematic study of the structure and behavior of the physical and natural world through observation and experiment.*
- *knowledge of any kind*
- *Science is both a body of knowledge and a process*
- *Science is the concerted human effort to understand, or to understand better, the history of the natural world and how the natural world works, with observable physical evidence as the basis of that understanding*
- *It is the systematic and organized inquiry into the natural world and its phenomena*

# USAGE OF SCIENCE

- *finding new knowledge*
- *Practically Understanding*
- *predictions*
- *improves human life at every level*
- *learn about enormously important and useful topics, such as our health, the environment, and natural hazards*

# PROCESS FLOW

How science works



*DATA*

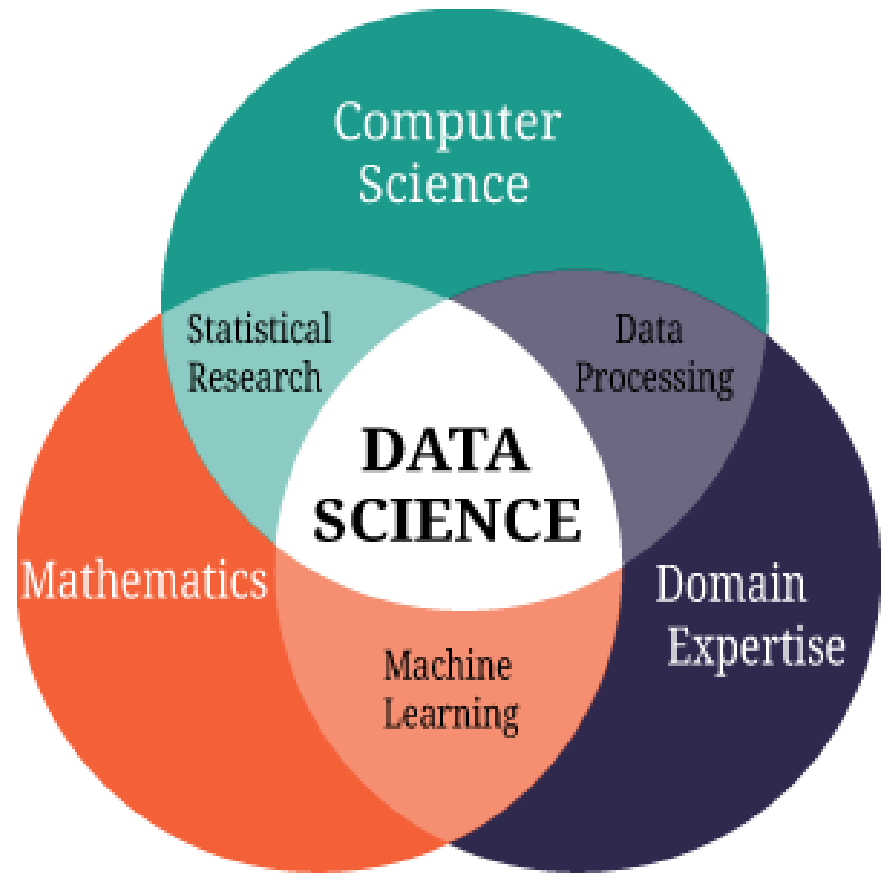
*SCIENCE*



*DATA SCIENCE*

# WHAT IS DATA SCIENCE?

- *uses scientific methods, processes, algorithms and systems*
- *extract knowledge*
- *insights from data in various forms, both structured and unstructured*





# WHY IT'S SO IMPORTANT

1. *Adding value to all the business models by using statistics and deep learning to make better decisions*
2. *Predict possible situations and risk*
3. *Human brain has certain cognitive limitations in terms of information fetching and processing, memory, recognition, prediction and so on*

# EXAMPLES

## Customer Prediction

System can be trained based on customer behavior pattern to predict the likelihood of a customer buying a product



## Service Planning

Restaurants can predict how many customers will visit on a weekend and plan their food inventory to handle the demand



# WHY PYTHON?

1. Primary language of Tech Giants ( Google, FaceBook ...)
2. Easy to Learn
3. Scalability
4. Variety of Data Science libraries  
(NumPy, SciPy, Pandas, StatsModels, Scikit-Learn ...)
5. Python Community
6. Graphics & Visualization
7. Python is faster than R in executing scripts
8. It is much easier to integrate with other applications like Hadoop-HDFS, Spark , Apache Kafka (Apache Kafka), etc. to create better data pipelines.

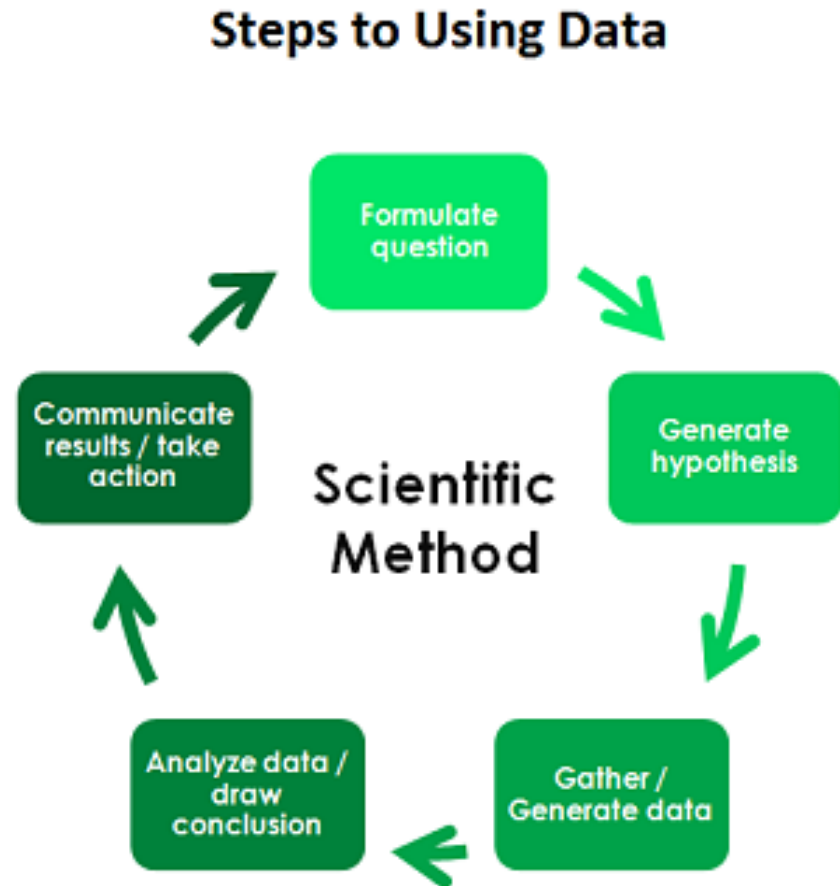
# WHY PYTHON?





# SCIENTIFIC METHOD

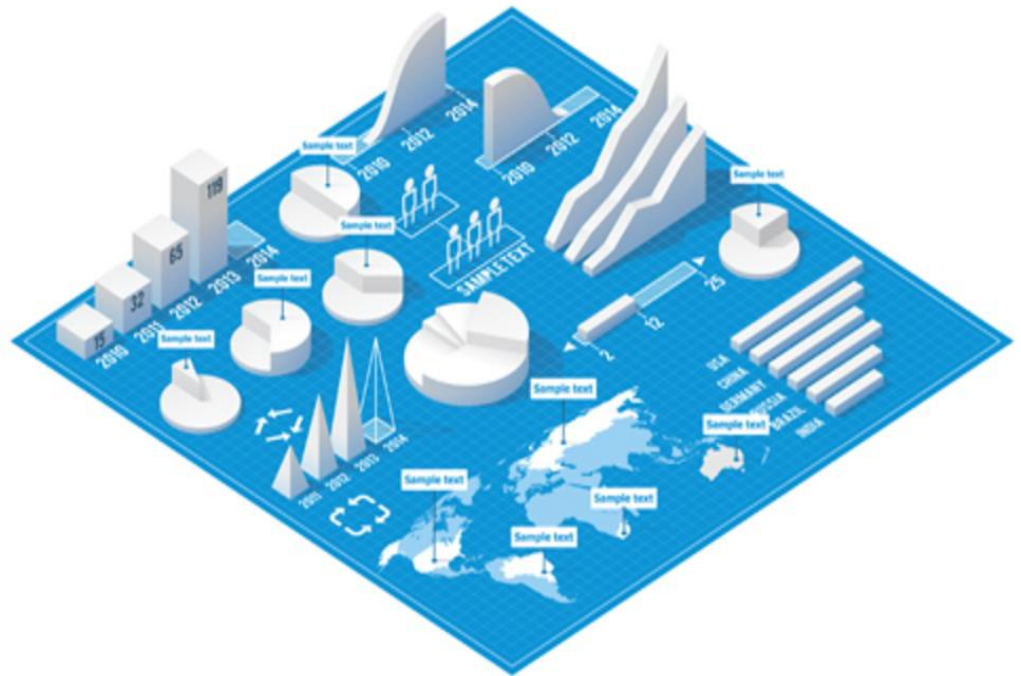
The scientific method is body of techniques for objectively investigating phenomena, *acquiring new knowledge, or correcting and integrating previous knowledge.* This method includes the collection of empirical evidence, subject to specific principles of reasoning



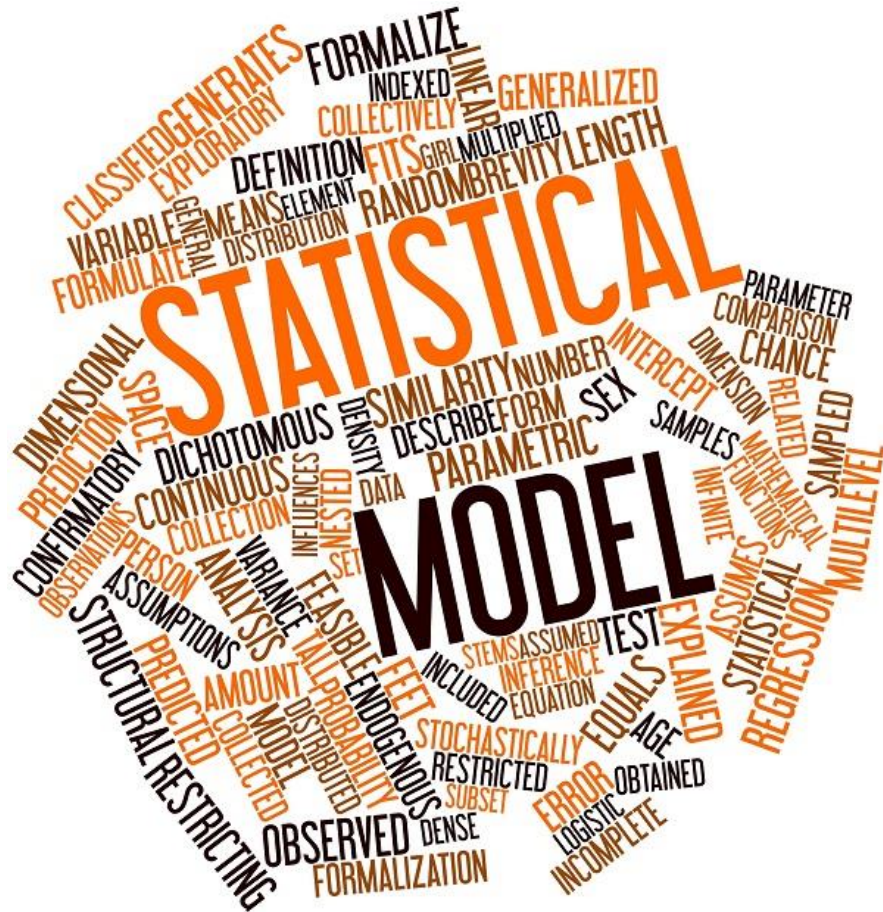


# DATA VISUALIZATION

Data visualization refers to the techniques used to communicate **data or information by encoding it as visual objects** (e.g., points, lines or bars) contained in **graphics**. The goal is to communicate **information clearly and efficiently** to users.



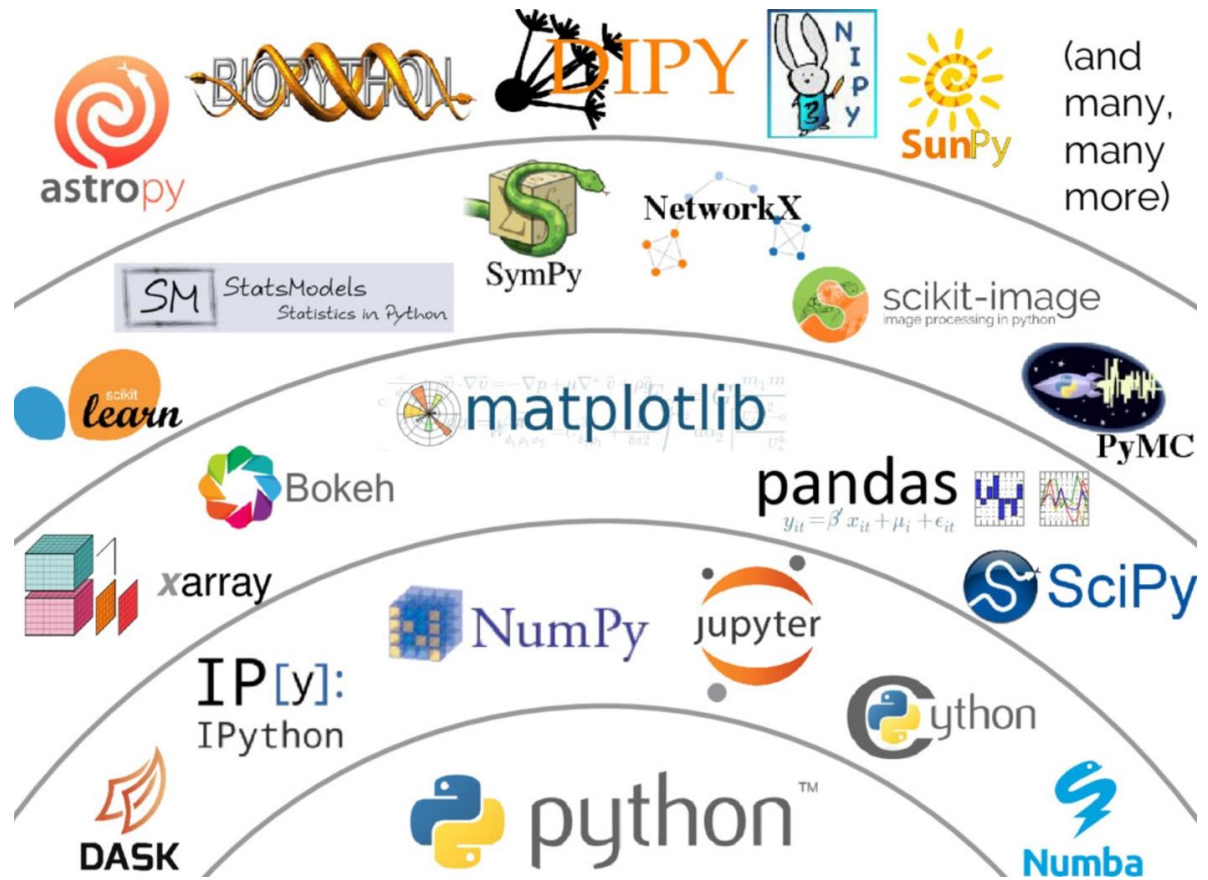
# STATISTICAL MODELING



A statistical model is a mathematical model that embodies a set of statistical **assumptions** concerning the generation of some sample data and similar **data** from a larger population. A statistical model represents, often in considerably idealized form, the data-generating process

# STATISTICAL COMPUTING

it's about  
processing  
large data  
sets using  
statistical  
techniques





# DATA TECHNOLOGY

A technology that **delivers information from various data sources**, including big data sources such as Hadoop and distributed data stores in real-time and near-real time

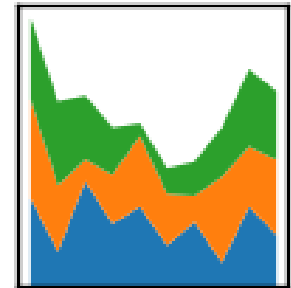
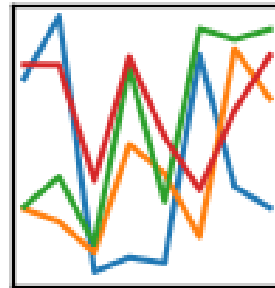
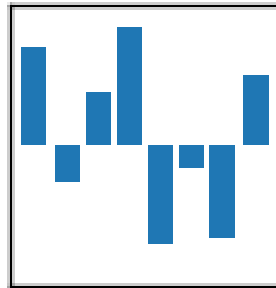


# DATA ANALYSIS

A process *of inspecting, cleansing, transforming, and modeling data* with the goal of *discovering useful information*, informing conclusions, and supporting decision-making

# pandas

$$y_{it} = \beta' x_{it} + \mu_i + \epsilon_{it}$$



**COND ...**