

# DATA - Another Mysterious World!

**Premanand S**  
Machine Learning Enthusiast

May 7, 2021

# Roadmap

- Data in minutes!
- DATA means?
- Data - Intro
- Data - History
- Data - Types
- Data - Characteristics
- How DATA is used for analytics - valuable insights?
- Steps to do with Data
- Data - Tools and Techniques
- Data - Engineer
- Data Engineers Vs Other roles
- Data - Some important jargon's

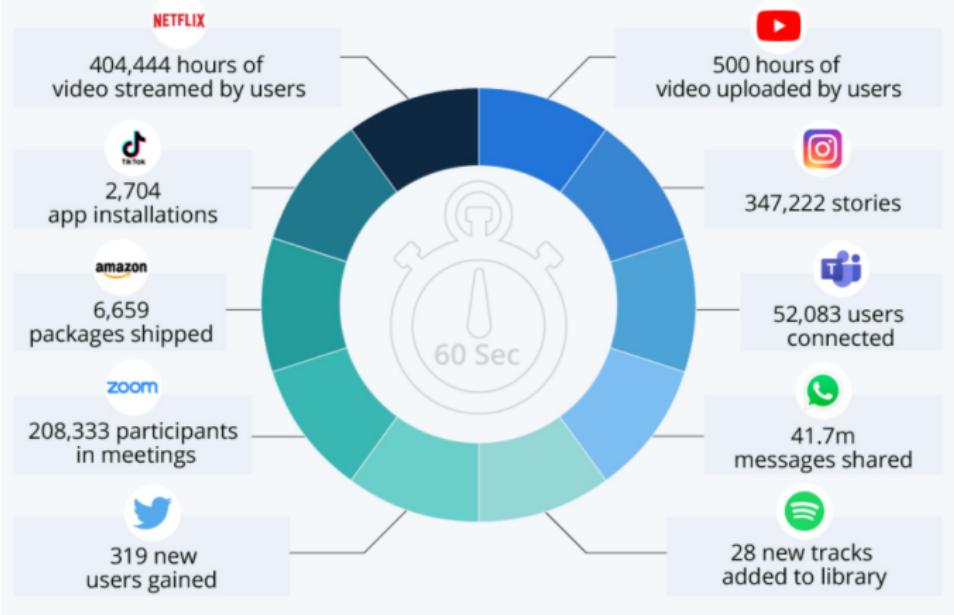
# Data here, Data there, Data everywhere!



# Data in minutes!

## A Minute on the Internet in 2020

Estimated amount of data created  
on the internet in one minute



Source: Visual Capitalist

# DATA means?!

# Data - Intro

- Its like a "CELL" for Human or like "FUEL" for transportation
- Data – Data Science (AI, ML, DL or Whatever)
- Our whole day in real time - So much of DATA
- DATA is everywhere like air!

# Importance of Data - MEME Explanation!



# What is Data?

- '**datum**' - Single piece of information
- **Data** - Any Signal - Wisdom / Information ?
- **Source** - Smart Phone, Laptop, Sensors, Smart devices & so on...
- **Types** - Image, Audio, Video, Pdf, Word, Excel & so on...
- **Application** - Communication, Social Media, E-Commerce, Sports, Groceries, News, Entertainment, Games & so on...

# Data from our App

## Signal Vs Telegram, WhatsApp Facebook Messenger: What Data Does Each App Collect from your phone?

Signal	Telegram	WhatsApp	Facebook
<p><b>None.</b> (The only personal data Signal stores is your phone number, and it makes no attempt to link that to your identity.)</p> 	<p>Contact Info Contacts User ID</p> 	<p>Device ID User ID Advertising Data Purchase History Coarse Location Phone Number Email Address Contacts Product Interaction Crash Data Performance Data Other Diagnostic Data Payment Info Customer Support Product Interaction Other User Content</p> 	<p>Purchase History Other Financial Info Precise Location Coarse Location Physical Address Email Address Name Phone Number Other User Contact Info Contacts Photos or Videos Gameplay Content Other User Content Search History Browsing History User ID Device ID Product Interaction Device ID Signal</p> <p>Advertising Data Other Usage Data Crash Data Performance Data Other Diagnostic Data Other Data Types Browsing History Health Fitness Payment Info Photos or Videos Audio Data Gameplay Content Customer Support Other User Content Search History Sensitive Info iMessage Email address Phone number Search history</p> 

# Data - History - Ancient World!



The Ishango bone, thought to be an early form of tallying



Cuneiform, one of the earliest known writing systems

## Data - History - Earlier Interpretations

## The table of casualties used by Graunt in his data analysis

*Natural and Political*  
**OBSERVATIONS**  
Mentioned in a following INDEX,  
and made upon the  
Bills of Mortality.

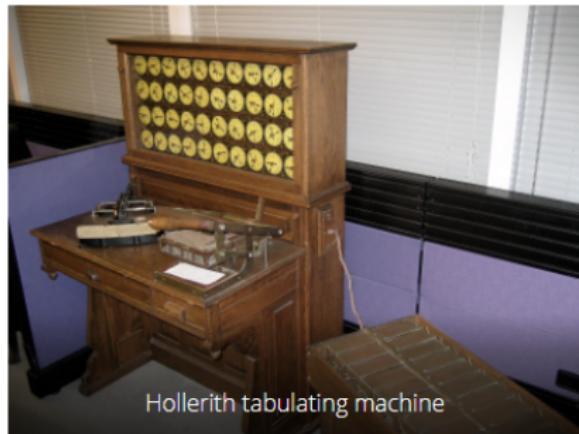
By JOHN GRAUNT,  
Citizen of  
LONDON.

With reference to the Government, Religion, Trade,  
Growth, Ayre, Distress, and the several Changes of the  
said COUNTRY.

— Nee, me ar mireur Tarba, labere.  
Commeau paxeur LeBarbus —

L O N D O N ,  
Printed by Thos. Raycroft, for John Martin, James Allardyce,  
The title page of *Observations on the Bills of Mortality*

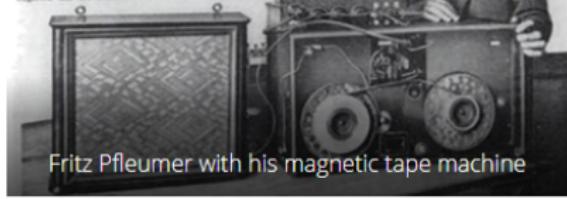
# Data - History - Data processing problems



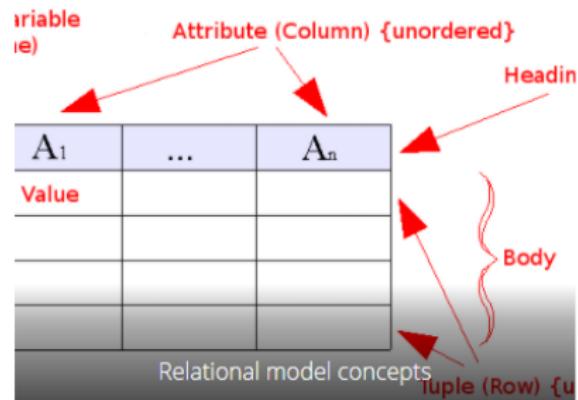
S	10	On	S	A	C	E	a	e	e	g	EB	SB	Ch	Sy	U	S
E	15	Off	IS	B	D	F	b	d	f	h	SY	X	Fp	Cn	R	
N	20			0	0	0	0	0	0	0	0	0	0	0	0	0
0	25	A	1	1	1	1	1	1	1	1	1	1	1	1	1	1
5	30	B	2	2	2	2	2	2	2	2	2	2	2	2	2	2
0	3	C	3	3	3	3	3	3	3	3	3	3	3	3	3	3
1	4	D	4	4	4	4	4	4	4	4	4	4	4	4	4	4
2	C	E	5	5	5	5	5	5	5	5	5	5	5	5	5	5
A	D	F	6	6	6	6	6	6	6	6	6	6	6	6	6	6
B	E	G	7	7	7	7	7	7	7	7	7	7	7	7	7	7
a	F	H	8	8	8	8	8	8	8	8	8	8	8	8	8	8
b	c	I	9	9	9	9	9	9	9	9	9	9	9	9	9	9

## Data - History - Question of storage

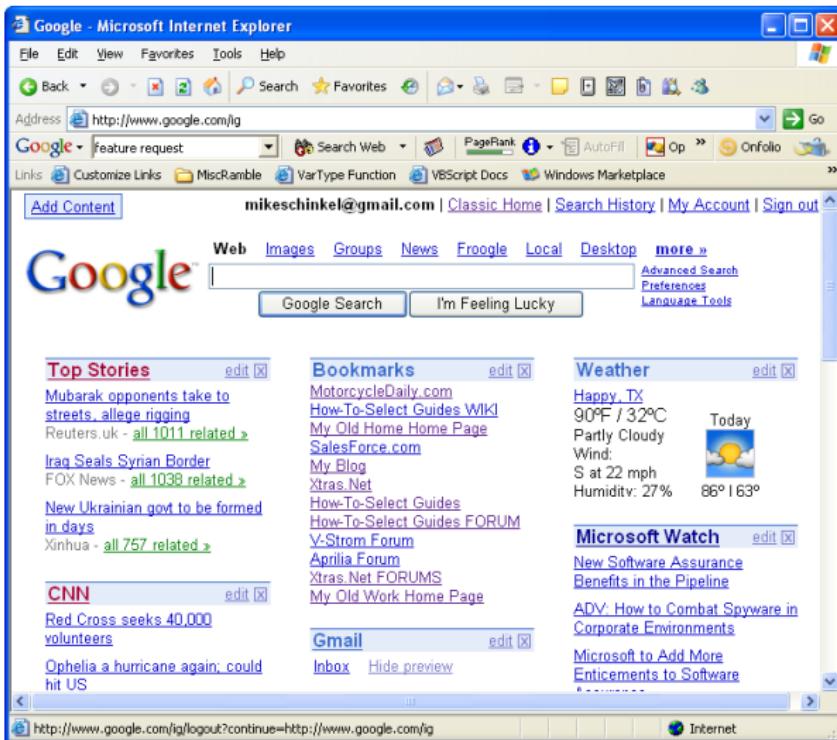
**Tierisches Papier.** Der Düsseldorfer Ingenieur Pfeiffer hat ein künstliches Tierisches Papier, Tinte und Papier zu fixieren. „Auf Zedernholz beschwieg sich ein Strohfilz Papier, ähnlich wie das Vorland der Schneckenhaube.“ Durch eine besondere Anordnung der Magnete kann man auf demselben Papier verschiedene Theile magnetisch transformieren. Diese magnetisierten Theile haften an dem Stahlkohlen. Bei der Wiedergabe wirken die magnetisierten Theile, die jetzt das Papier sind, auf den Magneten des Schreibstiftes, die jetzt das Lautholz sind, und so wird das geschriebene Papier wiederhergestellt. Die Magnete sind aus Eisen, die Stahlkohlen aus Stahl, Gestein, die Tinte aus einer wasserfesten Farbe. Eine 300 m lange Rolle des von Pfeiffer hergestellten Tierischen Papiers kostet für etwa 25.- bis 30.- Mark. Ein großer Vorteil dieses Papiers ist, dass es selbst bei sehr kalter Abstreuung, insbesondere Gras- und Grapfelgras, noch kommen klar bleibt. Das Papier fragt nach dem Preis 10.- Mark pro Meter. Durch Ueberstreichen mit einem Magneten kann das Tintbild gelöscht werden, und das Papier kann mit dem gleichen Magneten erneut verföhrt werden, so dass die Handlung wiederholt werden kann.“



## Fritz Pfeumer with his magnetic tape machine



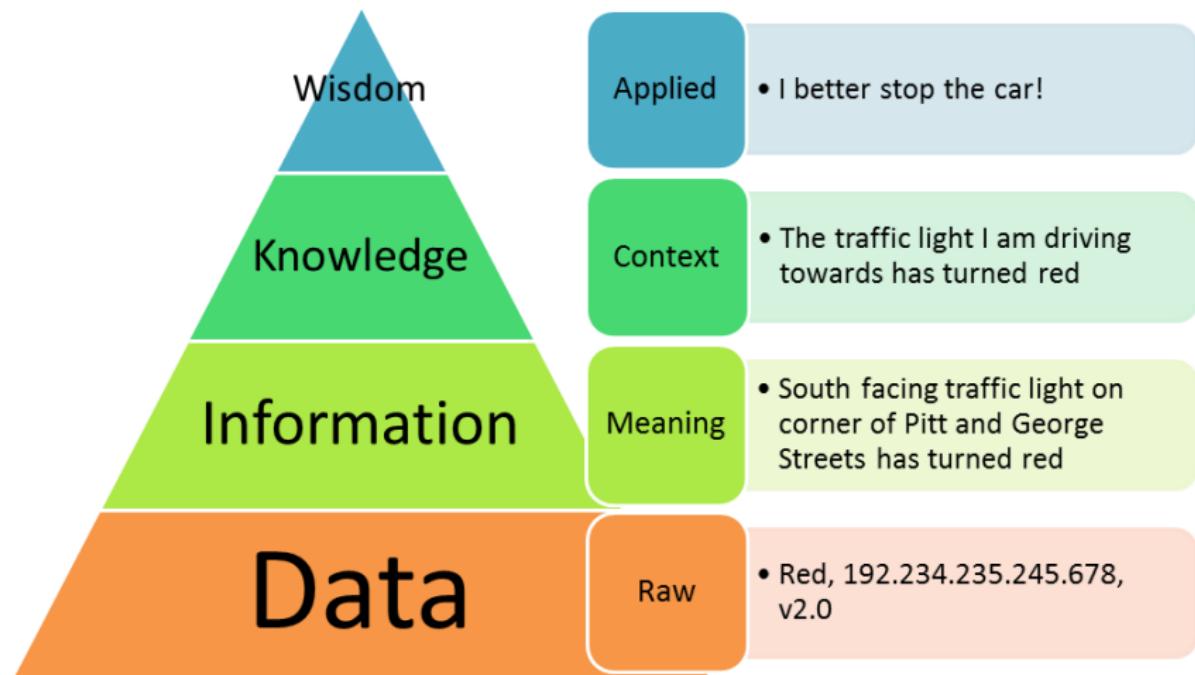
# Data - History - The Internet



# Data - History - Growth of AI, ethics



# Data in a better understanding!



# Data - Types

## Types of Data

### Quantitative

Data that can be measured with numbers, such as duration or speed

#### Discrete

Whole numbers that can't be broken down, such as a number of items

#### Continuous

Numbers that can be broken down, such as height or weight

#### Interval

Numbers with known differences between variables, such as time

#### Nominal

Data used for naming variables, such as hair colour

#### Ordinal

Data used to describe the order of values, such as 1 = happy, 2 = neutral, 3 = unhappy

#### Ratio

Numbers that have measurable intervals where difference can be determined, such as height or weight

### Qualitative

Non-numerical data that is categorical, such as yes/no responses or eye colour

# DATA - Characteristics

- Clean
- Complete
- Comprehensive
- Chosen
- Credible
- Calculable
- Currency

# How Data is used for analytics - valuable insights ?!

- Descriptive Analysis
- Diagnostic Analysis
- Predictive Analysis
- Prescriptive Analysis

# Descriptive Analysis



# Diagnostic Analysis

## Global Allergy Diagnostics Market 2020-2024



The market is **FRAGMENTED** with several players occupying the market share

### Health Care Industry:

The Health Care sector will see **POSITIVE IMPACT** due to the COVID-19 outbreak, and the industry is expected to register a high growth rate compared to the global GDP growth.

### Market impact:

This market will have **POSITIVE IMPACT** due to pandemic



All market estimates to be revisited and updated in Q4-2020, based on the revaluation of the impact, as the pandemic spread plateaus

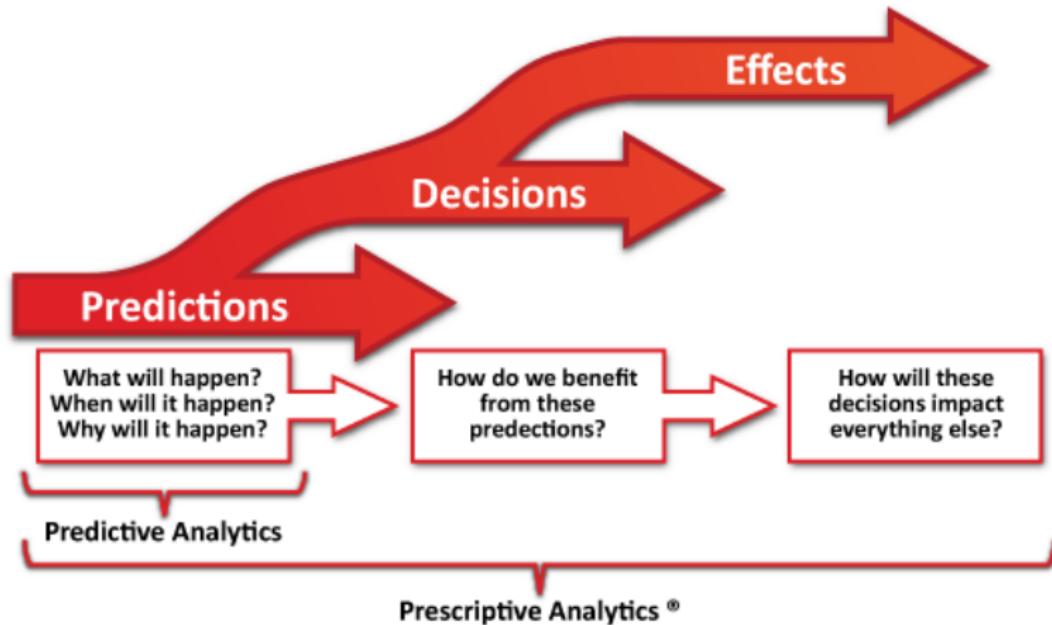
17000+ Reports covering niche topics. Read them at



# Predictive Analysis



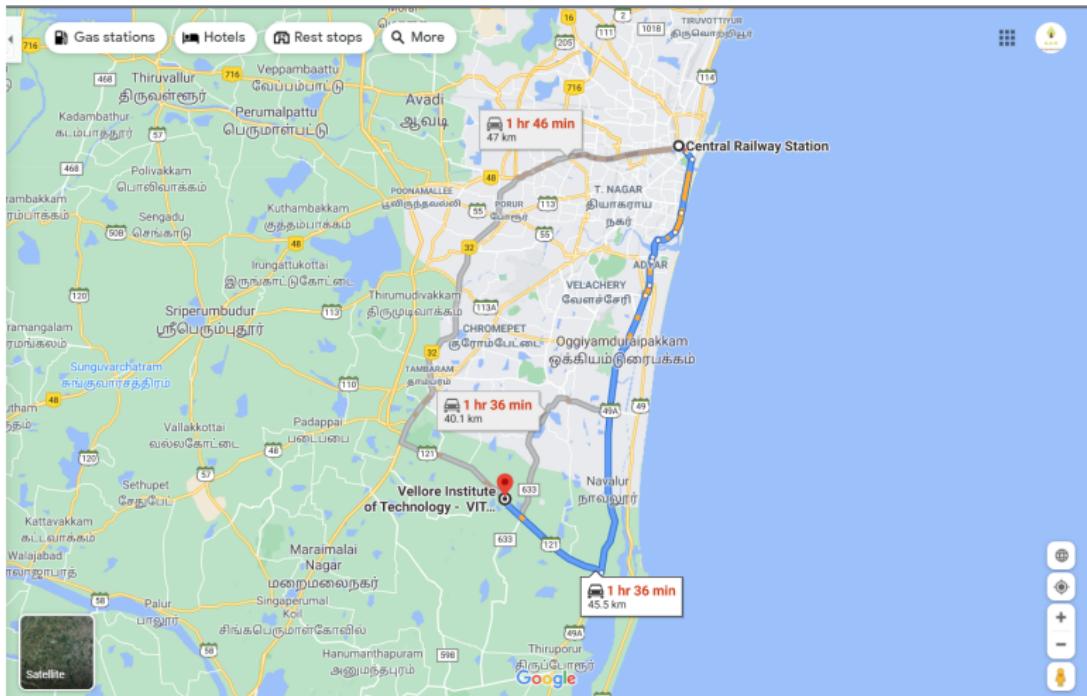
# Prescriptive Analysis



# Steps to do with Data

- Collect DATA – Different sources
- What to do with DATA – Visualization, Decisions, Hypothesis, Prediction, Insights so on... .

# Examples - Traffic Scenario



# Examples - Recommender system

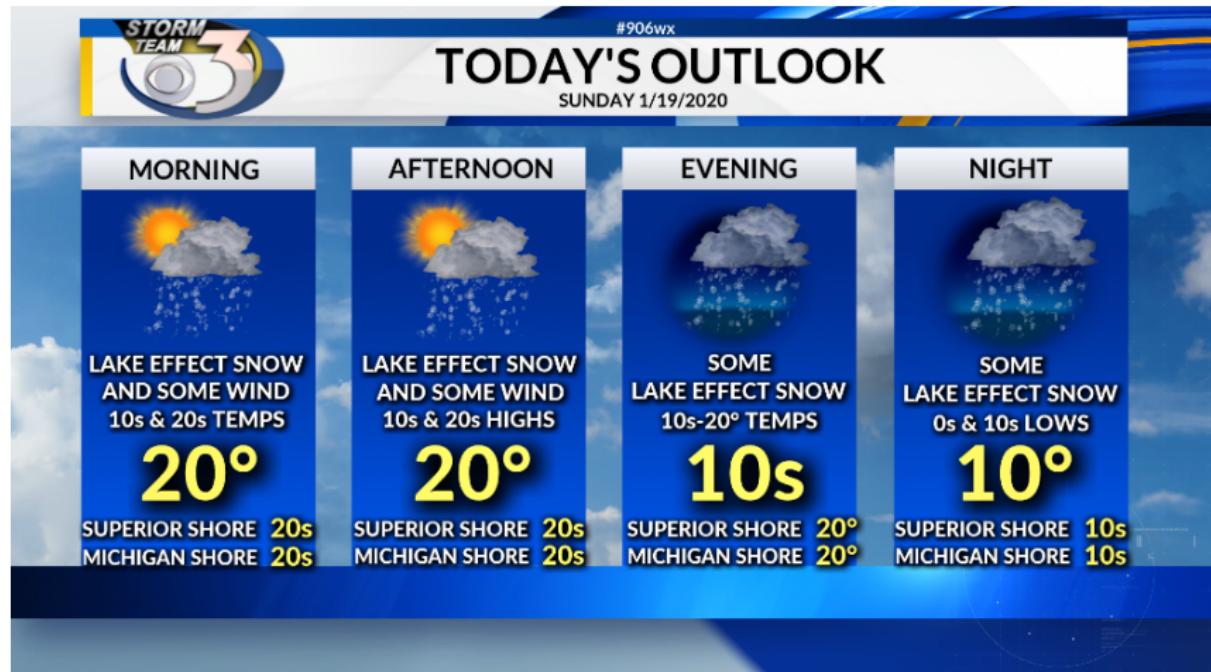
The screenshot shows the Amazon.in homepage with a 'Recommended for you, Premanand' section. The section displays four categories of products:

- Electronics**: 34 ITEMS. Includes images of a power supply unit, multiple green and black connectors, and electronic components like a breadboard and a microcontroller board.
- Home & Kitchen**: 100 ITEMS. Includes images of a blue pedestal fan, a small white device, and various kitchen tools like screwdrivers and wrenches.
- Office Products**: 6 ITEMS. Shows a large black stapler.
- Home Improvement**: 10 ITEMS. Includes images of a set of wrenches, a coiled cable, a pink glove, and a power strip.

Below this section, there are two more rows of product recommendations, partially visible:

- Electronics**: Includes a small image of a decorative plaque.
- Home & Kitchen**: Includes images of a digital multimeter displaying "6.94 V 0.00 A" and a digital timer displaying "89".
- Office Products**: Includes images of a bag of coffee beans and a small black device.
- Home Improvement**: Includes images of two large black control panels or meters.

# Examples - Weather Forecasting



## Examples - Medical Diagnosis



# Examples - Smart Cities



# Examples - Financial Market



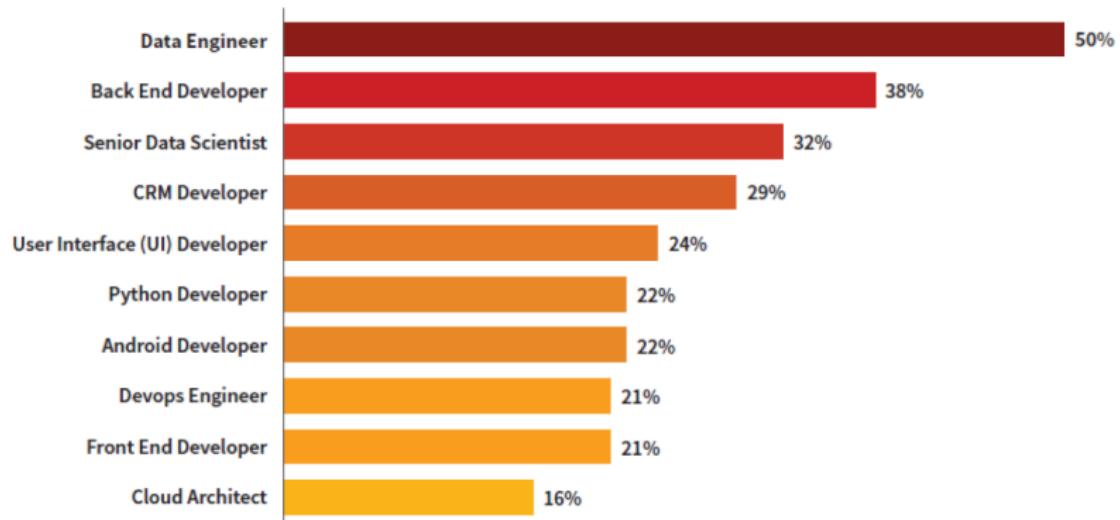
# Data - Tools and Techniques

- Basic Data Manipulation Analysis
- Big Data
- Data Mining
- Machine Learning / Deep Learning
- Data Visualization
- Data Collection and Preprocessing

# Data - Engineer

## FASTEST GROWING TECH OCCUPATIONS

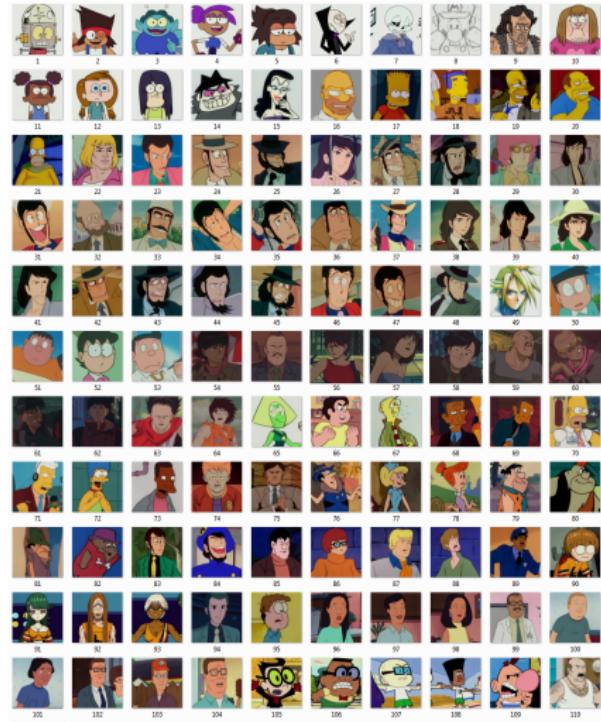
YEAR-OVER-YEAR GROWTH



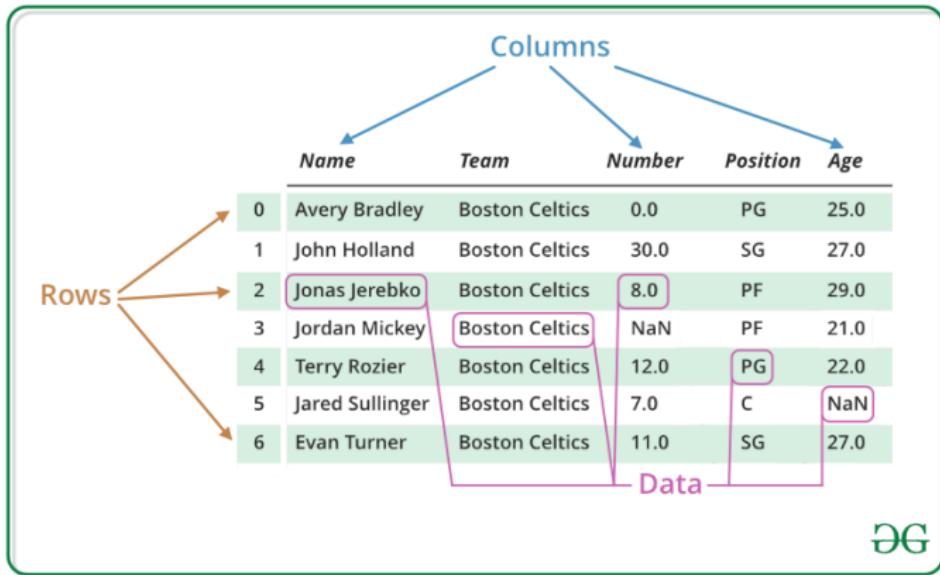
# Data Engineer Vs Other roles

- Data role: Data scientist, Data engineer, Data Analyst, Machine Learning Engineer and Machine Learning Scientist.
- Companies are targeting those who can engineer the data more than those who can only model the data.

## Datasets

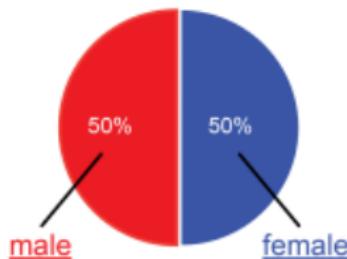


# Data Frames



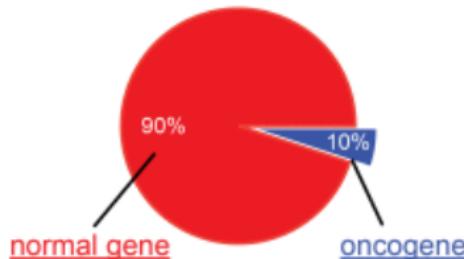
# Balanced and Imbalanced Datasets

Example of balanced and imbalanced data



Negatives ≈ Positives

Balanced



Negatives > Positives

Imbalanced

# Data Cleaning and Enrichment



# Key points to get into Data Science

- Language - Python or R
- Basic Maths and Statistics
- Data Visualization
- Machine Learning
- Deep Learning
- Other tools

mail me: er.anandprem@gmail.com

ring me: +91 73586 79961

follow me: Linkedin

Website: tango-learning

**Learning gives Creativity, Creativity leads to Thinking, Thinking provides Knowledge, and Knowledge makes you Great - Dr APJ Abdul Kalam**