



Data Science

Introduction

Making sense of Data

Takeaways from the course

@amity1415



Introduction to
Data Science



What is
Data?



What is Data
Science?



Industry
applications



Problems solved
by Data Science



Project Life
Cycle



Data Types



Conclusion

Agenda

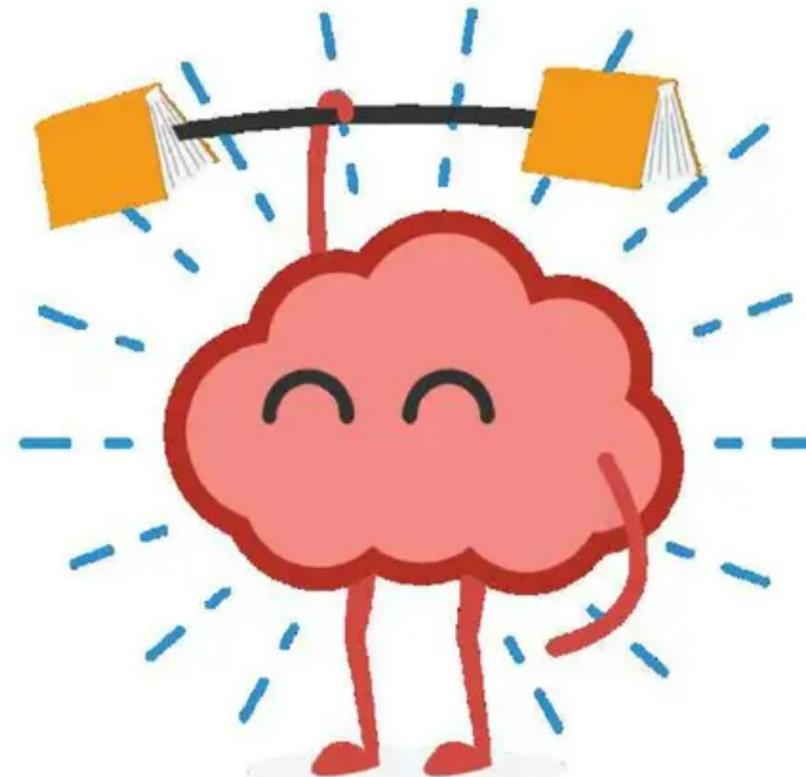
@amity1415



- Example to reflect on
- What is Data?
- What is Data science?
- Who are Data Scientists?
- Deeper discussions
- Problems that DS Solves
- Applications
- Project life cycles
- More on Data

Let's jog our minds with an exercise!

@amity1415



Example to reflect on

@amity1415

Guess the
Customer Profile!

PURCHASE BEAVIOUR			PROFILE
iPhone	Beer	Cornflakes	
iPhone	Diapers	Beer	
iPhone	Diapers	Biscuits	



Example to reflect on

@amity1415

Guess the
Customer Profile!

PURCHASE BEAVIOUR			PROFILE
iPhone	Beer	Cornflakes	Single M/F
iPhone	Diapers	Beer	Parent M/F
iPhone	Diapers	Biscuits	Parent M/F



Example to reflect on

@amity1415

Guess the
Customer Profile!

PURCHASE BEAVIOUR			ALSO BUY	PROFILE
iPhone	Beer	Cornflakes	Biscuits	Single M/F
iPhone	Diapers	Beer	Aftershave	Parent M/F
iPhone	Diapers	Biscuits	Cosmetics	Parent M/F



Example to reflect on

@amity1415

Guess the
Customer Profile!

PURCHASE BEAVIOUR			%	ALSO BUY	PROFILE
iPhone	Beer	Cornflakes	97%	Biscuits	Single M/F
iPhone	Diapers	Beer	98.4%	Aftershave	Parent M/F
iPhone	Diapers	Biscuits	91%	Cosmetics	Parent M/F



Example to reflect on

@amity1415

What do you
recommend?

iPhone

Beer

Cornflakes

?

?

?

Hint !!

PURCHASE BEAVIOUR			%	ALSO BUY	PROFILE
iPhone	Beer	Cornflakes	97%	Biscuits	Single M/F
iPhone	Diapers	Beer	98.4%	Aftershave	Parent M/F
iPhone	Diapers	Biscuits	91%	Cosmetics	Parent M/F

Example to reflect on

@amity1415



amazon.com

Recommended for You

Customers who viewed this item also viewed these products

Dualit Food XL1500 Processor \$560 Add to cart	Kenwood kMix Manual Espresso Machine ★★★★★ \$250 Select options	Weber One Touch Gold Premium Charcoal Grill-57cm \$225 Add to cart	NoMU Salt Pepper and Spice Grinders \$3 View options

*Source: Venturebeat

What is Data: Examples

@amity1415



Sequence of bits

8	4	8
1	7	5
6	2	3

Table of Numbers



Characters on a page



Recording of sounds



Moon rock specimen

What is Data Science?

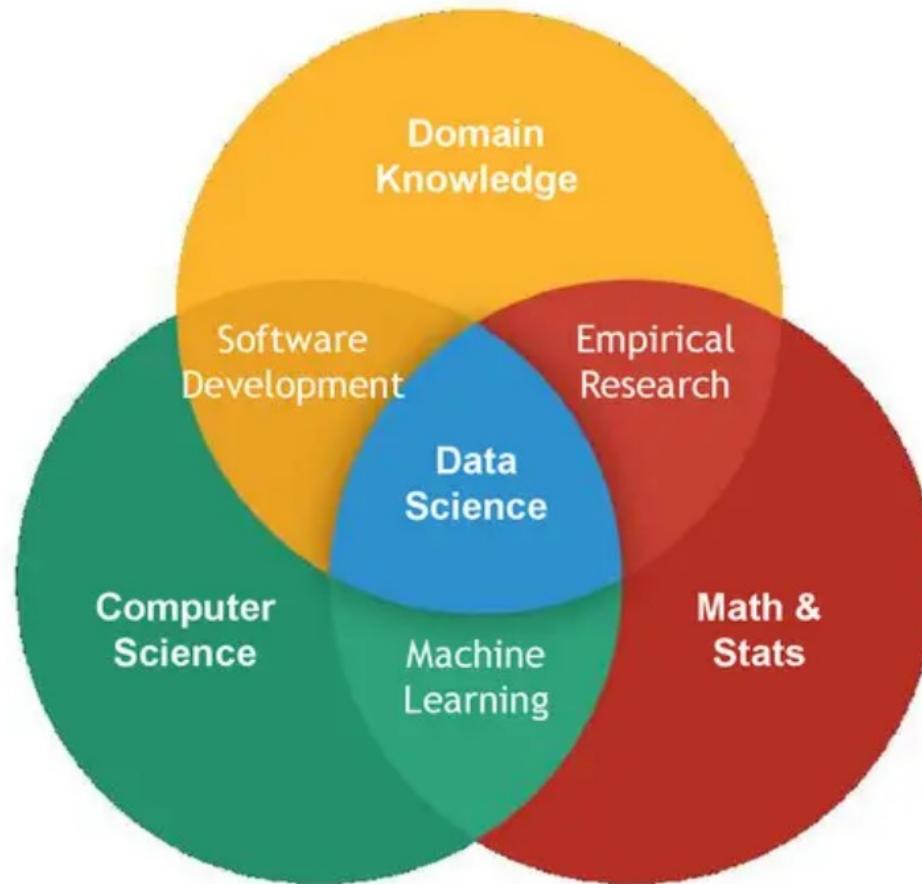
Data science is the study
of:

Where information
comes from

What it represents,
and

How it can be turned into a
valuable resource

for business strategies.



What is the point of it?

@amity1415

Analytics/BI



Shows **what** has happened

Data Science(EDA)



Shows **why** it happened

Machine Learning



Shows **how** we can predict & automate decision making systems

Problems that Data Scientists solve

@amity1415

**Is this
weird?**

Anomaly



Is this mail in your inbox
spam?



Is someone riding bike on
a walkway?

Problems that Data Scientists solve

@amity1415

How much
or
How many?

Regression



How much will this used
car sell for?



Predict how much product
sales will be achieved in this
month?

Problems that Data Scientists solve

@amity1415

**How is this
organized?**

Clustering



Which viewers like the
same kind of movies?



Which pet owner groups
have the same purchase
behaviour for pet foods?

Problems that Data Scientists solve

@amity1415

**What
should I do
now?**
Reinforcement
learning



How far should the
Robot Jump?



Should I break,
accelerate, or continue
at that yellow light?

Data Science: Applications in Industry

@amity1415

Telecom



- Customer Acquisition Strategies
- Churn Analysis and Control
- Up-sell / Cross-sell
- Product Bundling

Data Science: Applications in Industry

@amity1415

Healthcare



- Clinical trials of new drugs
- A / B Testing
- Genetics Analysis
- Epidemic Forecasting and Control

Data Science: Applications in Industry

@amity1415

Banking & Finance



- Fraud detection and prevention
- Customer Segmentation
- Risk management
- Portfolio Optimization

Energy & Utilities



- Predictive maintenance for machinery
- Predict theft and loss prevention
- Churn analysis and prevention
- Auto balance resources and dynamic routing

Data Science: Applications in Industry

@amity1415

Retail & e-commerce



- Shelf-space Optimization
- Market Basket Analysis
- Product Bundling
- Promotions
- Up-sell / Cross-sell

How do I start a Data Science project

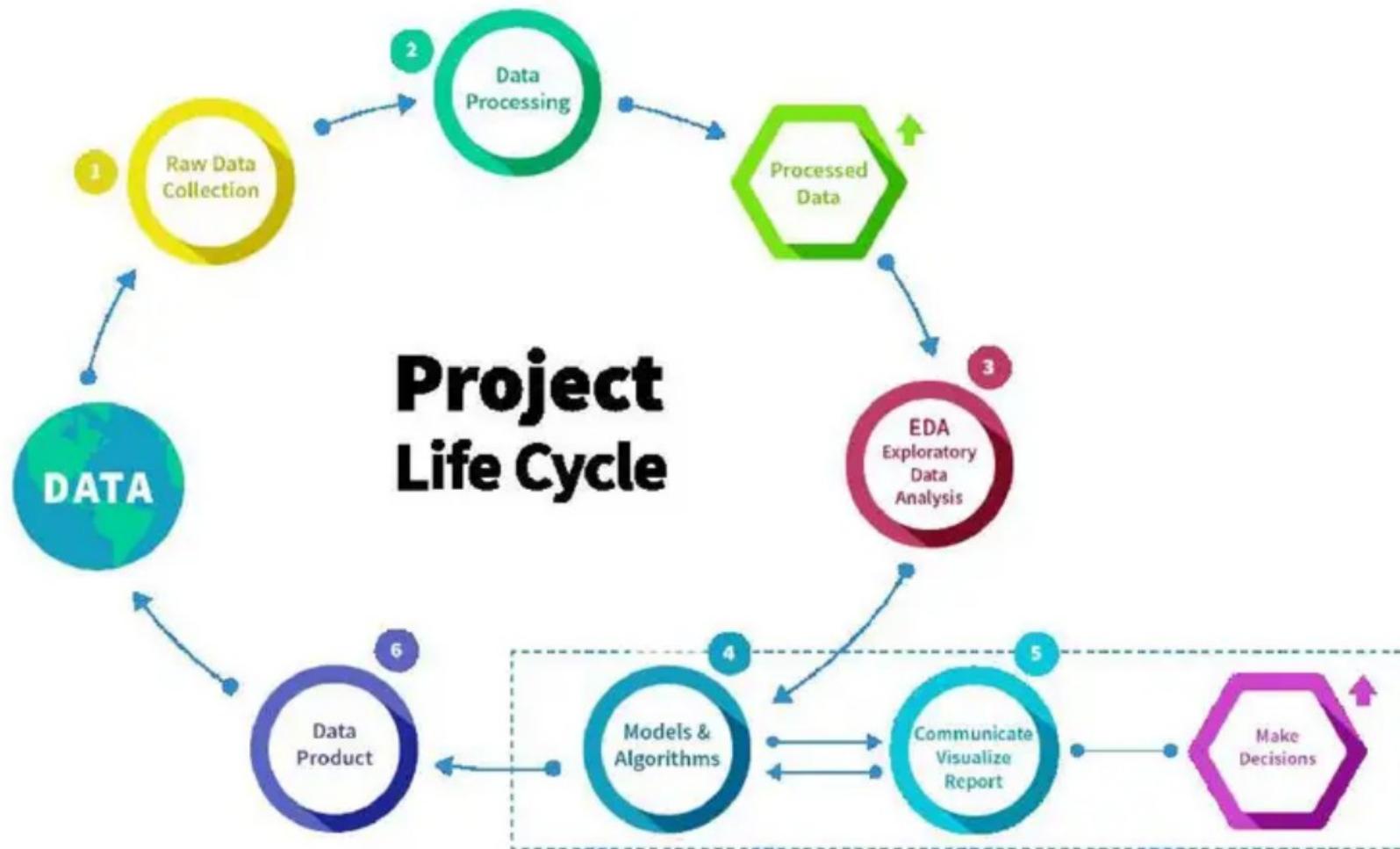
@amity1415



**Just follow
the upcoming
simple steps...**

Data Science Process

@amity1415

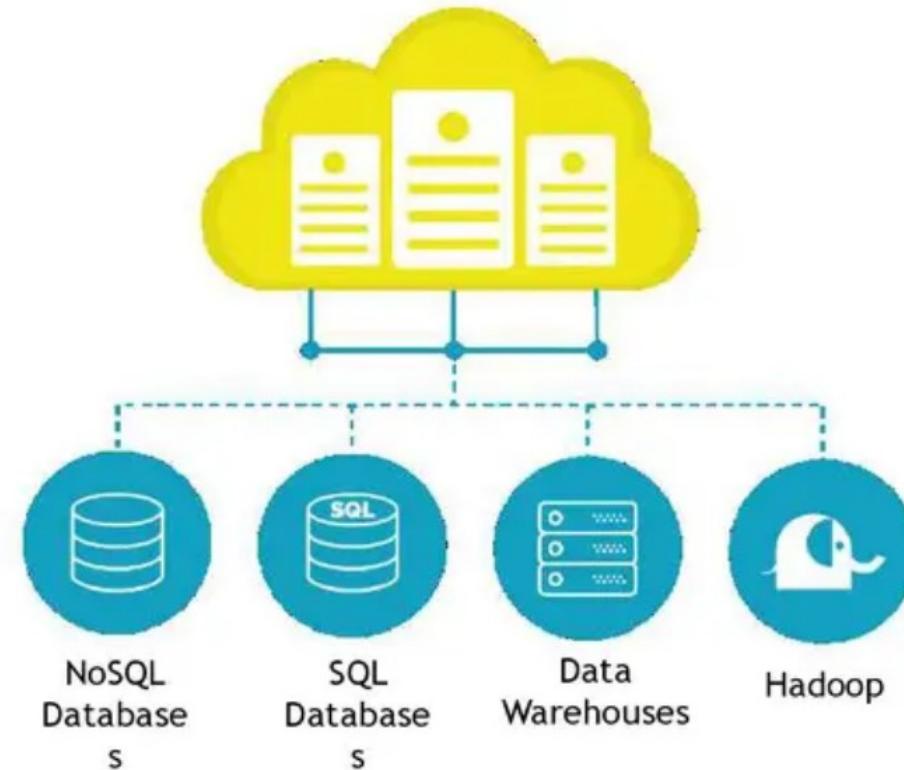


Data Science Process

@amity1415

01

Raw Data collection
&



Data Science Process

@amity1415

02

Data Cleaning & Processing

Pre-processing / Cleaning



Removing Outliers



Missing Data



Malicious Data



Erroneous Data



Irrelevant Data



Inconsistent Data



Formatting

Data Science Process

@amity1415

02

Data Cleaning & Processing

Data Cleaning Example

#	First Name	Last Name	Age	Gender	Mobile	Telephone	Address
1	David	Wong	26	Male		02 -2110063	3-12 soi S Rochtevi
2	Bili	Chrlnasom	43	Male	08-9999-8765		354 SOi 4 Nono SukumV'lt
3	Noncv	Nomochue	29	Female	08-1111-3456	02 -3 10-8633	Sukumv1I Klonutoev Banukok
4	Mie	Nakanishi	26	FeMale	0-1230- 9965	02 -260-5544	
5	David	1sh1ko	26	Mole	08 7511-1234		

Un-standardized
Missing/Misspelled
Duplications

#	First Name	Last Name	Age	Gender	Mobile	Telephone	Address
1	David	Ishikawa	26	Male	QS.7511-1234	02-211-0063	3-12 soi 5 Rochtevi
2	Bill	Chrinasom	43	Male	08-9999-8765	02-416-0011	354 soi4 Nono Sukumvit
3	Noncv	Nomdluc	29	Female	08-1111-3456	02-310-8633	S-IO A Mansion 104 SOI 4 Sukumvil
4	Mie	Nakan shi	26	Female	0-1234-9965	02-260-5544	71 soi Rotnrao

Standardized
Populated
Duplicate removed

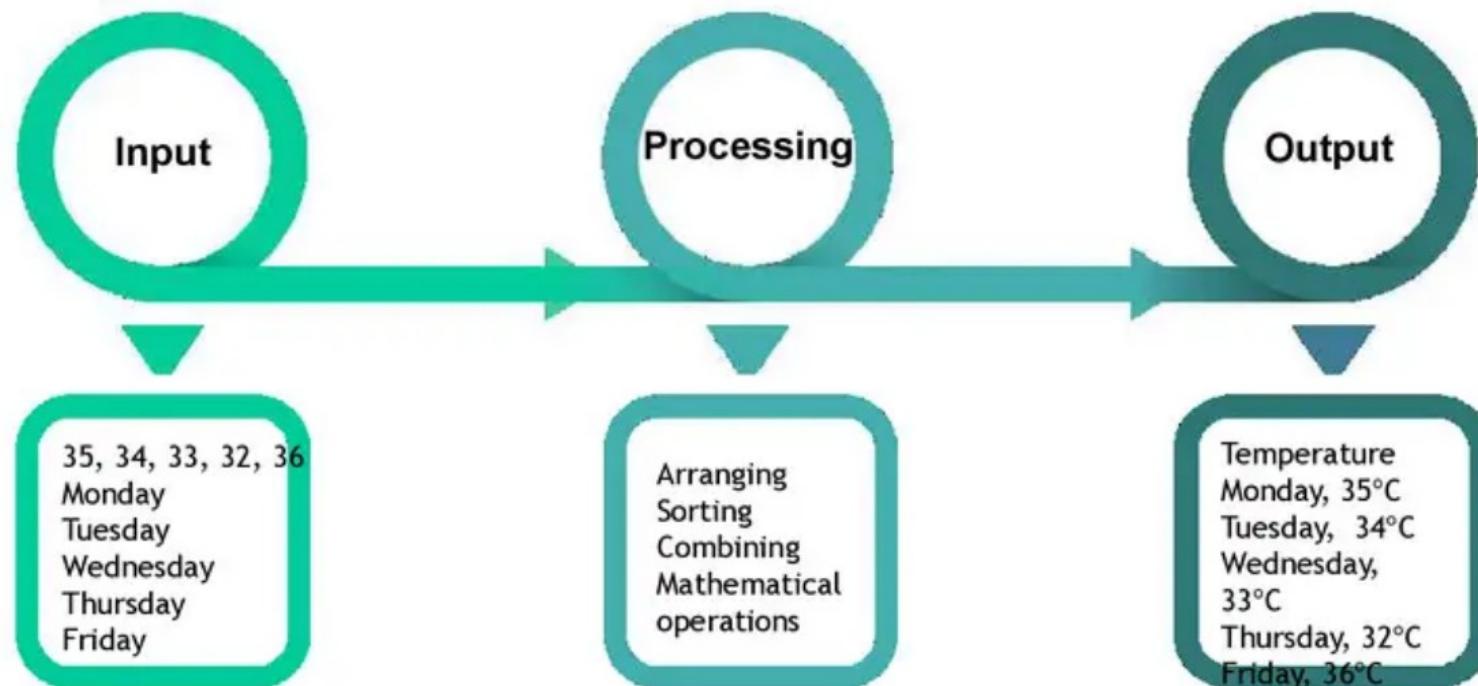
Data Science Process

@amity1415

02

Data Cleaning & Processing

Data Processing Example

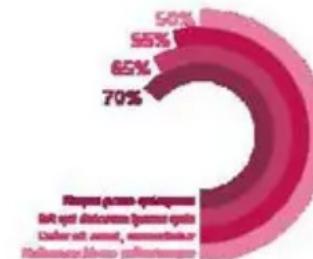
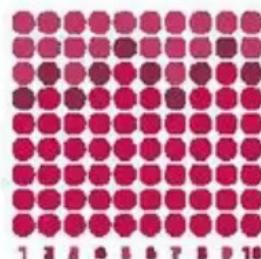
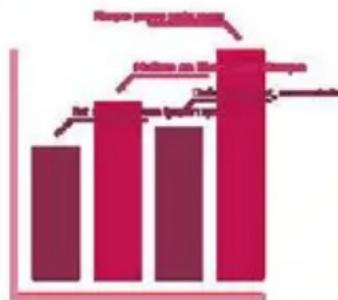


Data Science Process

@amity1415

03

Exploratory Data Analysis (EDA)



Data Science Process

@amity1415

03

Exploratory Data Analysis (EDA)

Objectives



Discover
Patterns



Spot
Anomalies



Frame
Hypothesis



Check
Assumptions

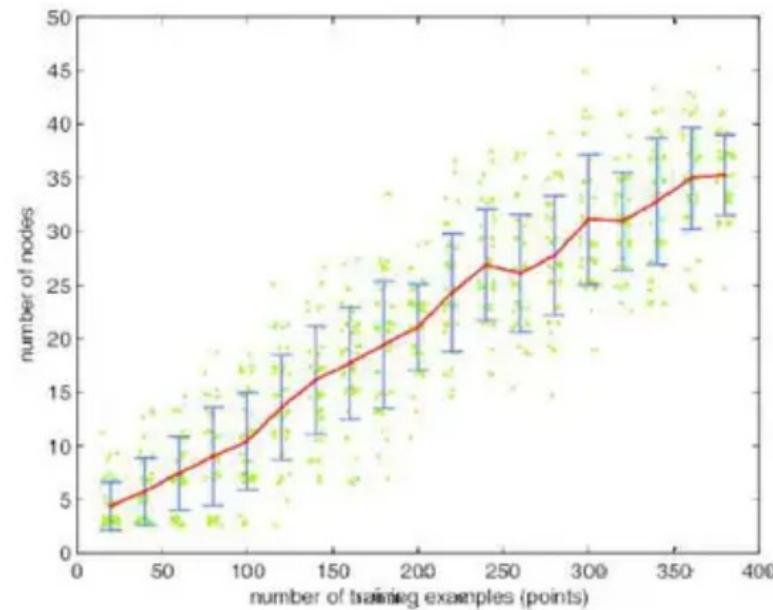
Data Science Process

@amity1415

04

Models & Algorithms

Create multiple models to solve the business problem, and find the most appropriate one (Speed vs Accuracy)



Data Science Process

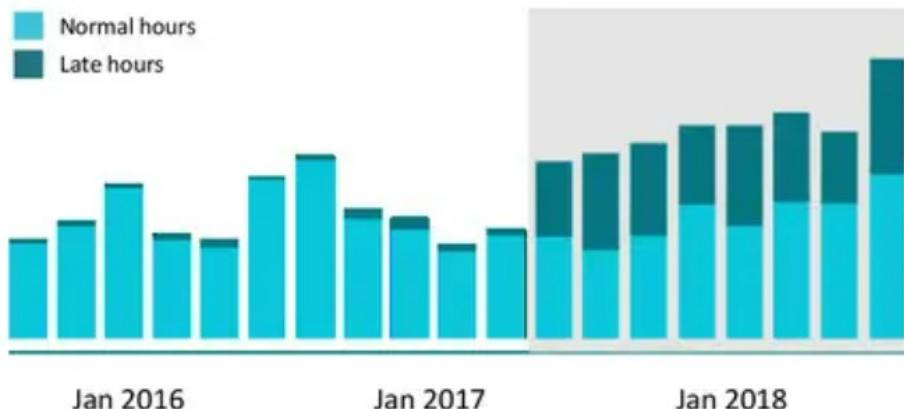
@amity1415



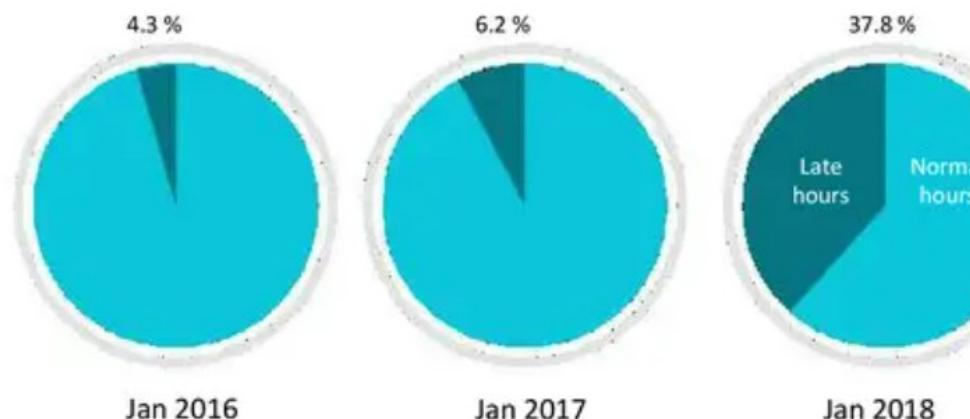
Communicate, Visualize & Report

Take action and deploy findings in the real world

Significant increase in late night technician repairs since Jan 2016



Late night technician repairs increase to almost 40% in Jan 2018



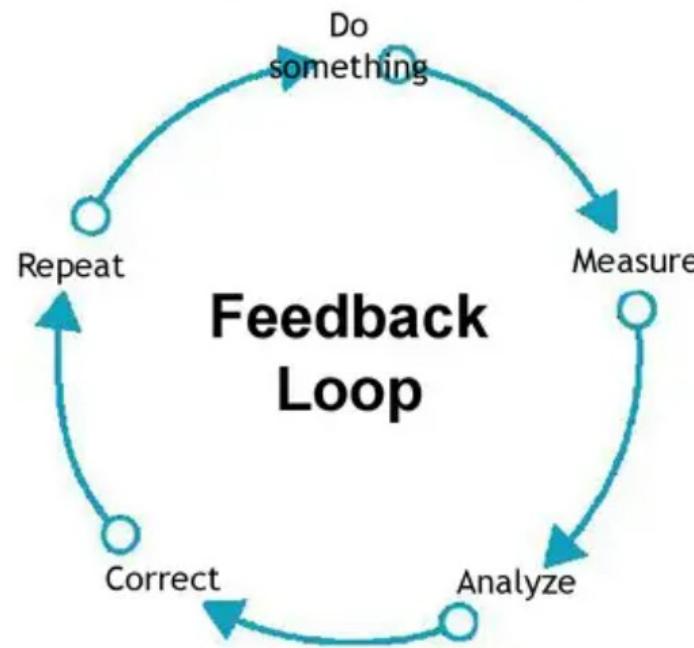
Data Science Process

@amity1415

05

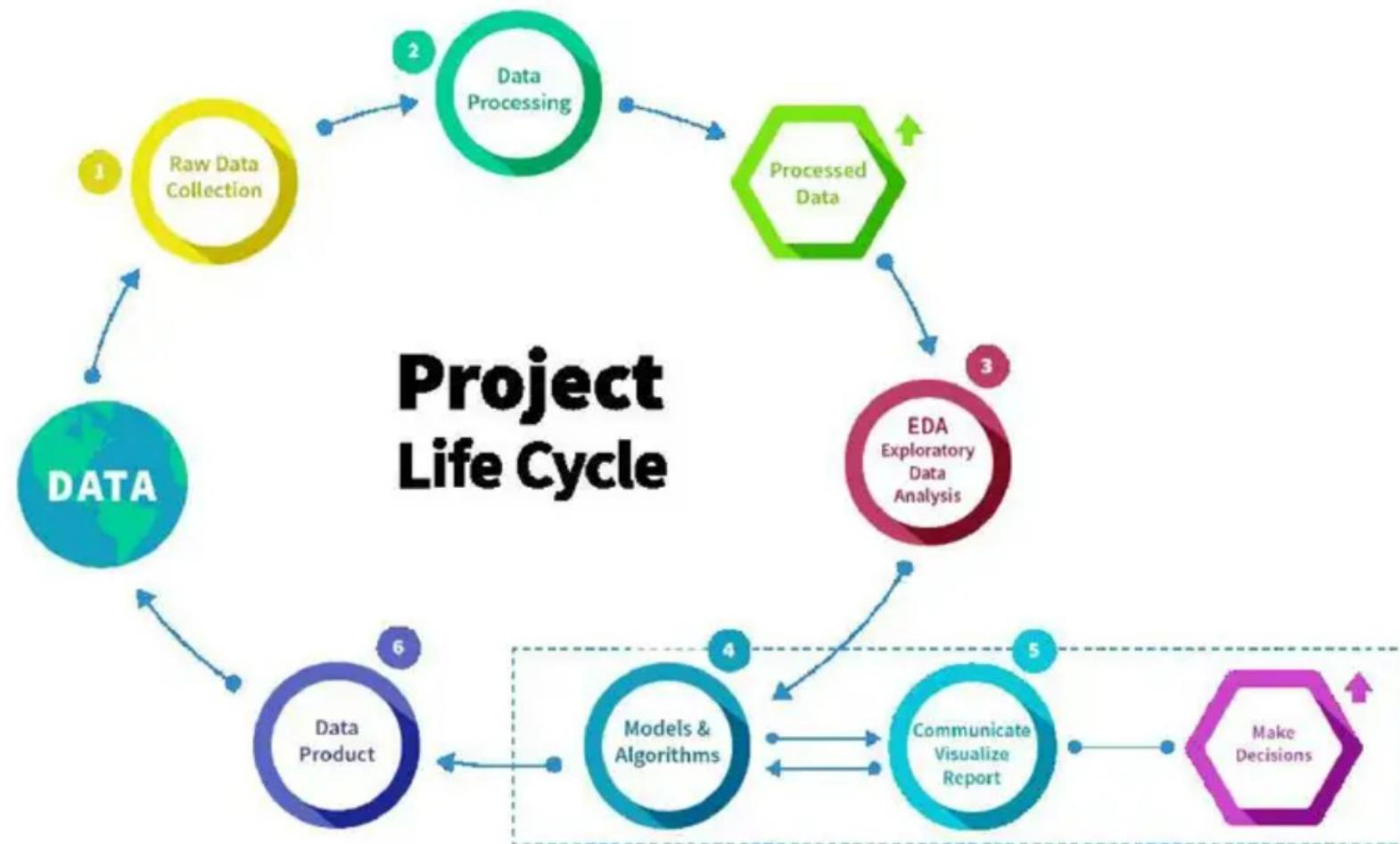
Communicate, Visualize & Report

Present results to target stakeholders



Data Science Process

@amity1415



What is at the heart of Data Science?

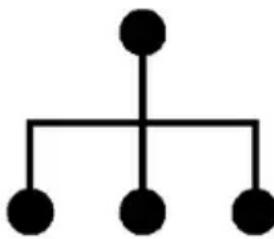
@amity1415



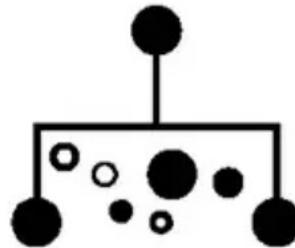
DATA

Types of Data

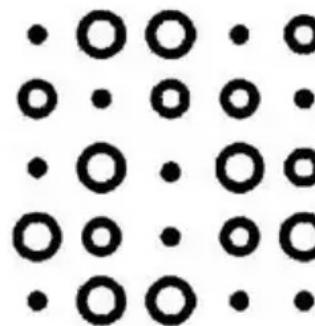
@amity1415



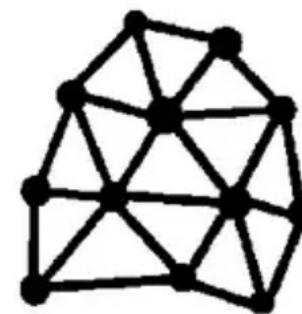
Structured
data



Semi-structured
data



Unstructured
data



Graph
data



Streaming
data

Types of Data

@amity1415

Structured data - Transactions

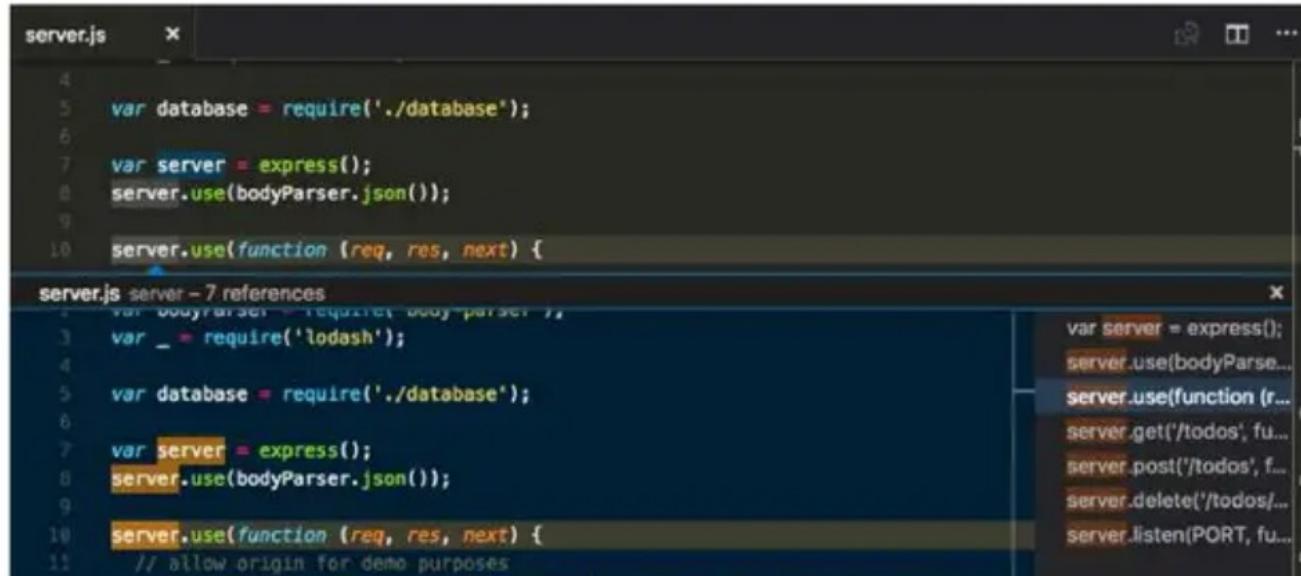
Transactions Before November 21 2016

All	Payment Type	Event	Amount	Payment Type	Account	Payment Customer	Gateway	Gateway Transaction ID	Info
KBBJ-071116-1	✓ PAID		\$89.60	ACCOUNT VOUCHER1	👤 Dolly	Dolly	None	5820cc52a177b	<button>查看详情</button>
GFTM-170616-1	✓ PAID		\$42.00	AMEX ****0005 John Smith	🌐 website	John Smith	None	1466209337	<button>查看详情</button>
CSZV-170616-1	✓ PAID		\$61.60	AMEX ****0005 Brendan McClure	🌐 website	Brendan McClure	None	1466213357	<button>查看详情</button>
FCDQ-170616-1	✓ PAID		\$123.20	AMEX ****0005 Brendan McClure	🌐 website	Brendan McClure	None	1466213761	<button>查看详情</button>
CHZG-290915-1	✓ PAID		\$316.25	Cash Norris Cole	🌐 website	Norris Cole	Vend	7ec56036-6258-4	<button>查看详情</button>

Types of Data

@amity1415

Semi structured data - JavaScript code



```
server.js  x

4  var database = require('./database');
5
6  var server = express();
7  server.use(bodyParser.json());
8
9  server.use(function (req, res, next) {
10
11    // allow origin for demo purposes
12  })

server.js server - 7 references
1  var bodyParser = require('body-parser');
2  var _ = require('lodash');
3
4  var database = require('./database');
5
6  var server = express();
7  server.use(bodyParser.json());
8
9  server.use(function (req, res, next) {
10
11    // allow origin for demo purposes
12  })

var server = express();
server.use(bodyParser.json());
server.use(function (req, res, next) {
server.get('/todos', function (req, res) {
server.post('/todos', function (req, res) {
server.delete('/todos/:id', function (req, res) {
server.listen(PORT, function () {
```

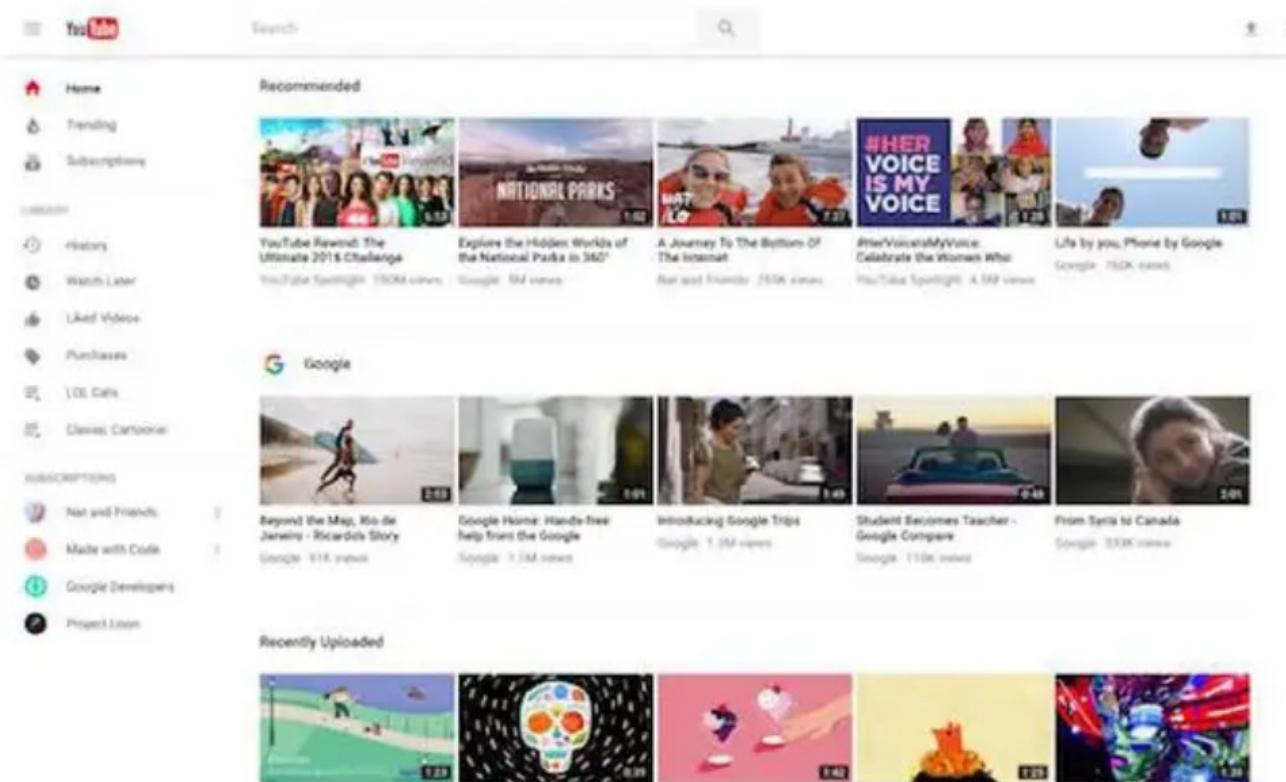
{.js}
JavaScript



Types of Data

@amity1415

Unstructured data - YouTube Feed



Types of Data

@amity1415

Graph data - Facebook friend network



Types of Data

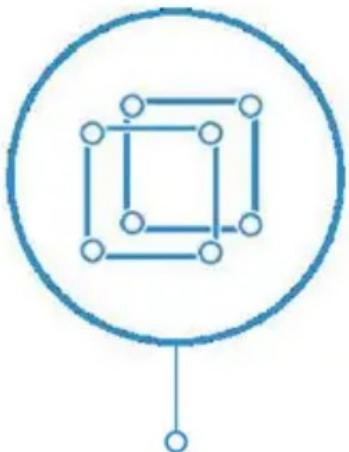
@amity1415

Streaming data - Stock prices



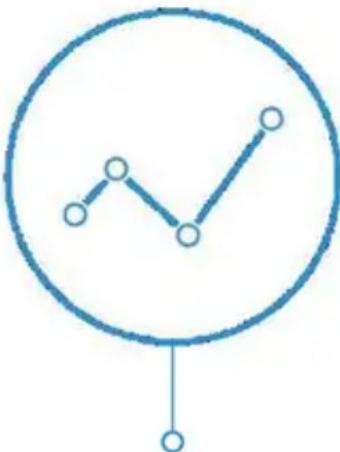
Data Quality Issues

@amity1415



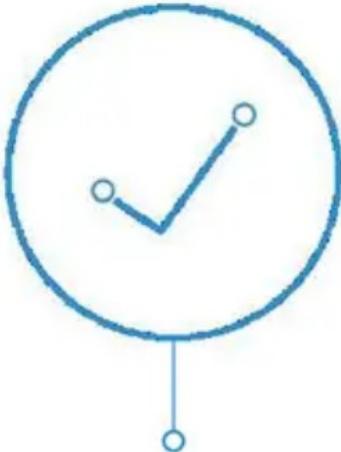
Duplicate

y
Redundancy
leading to
resource
wastage



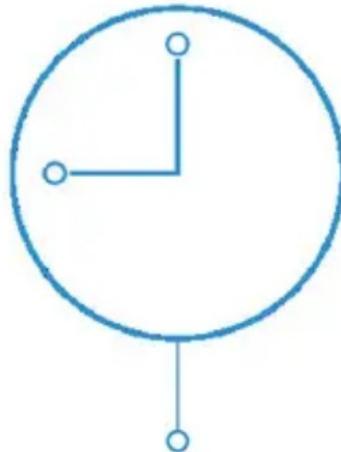
Inconsistency

y
Withdrawal of INR
10/- not
reflecting in Net
Banking



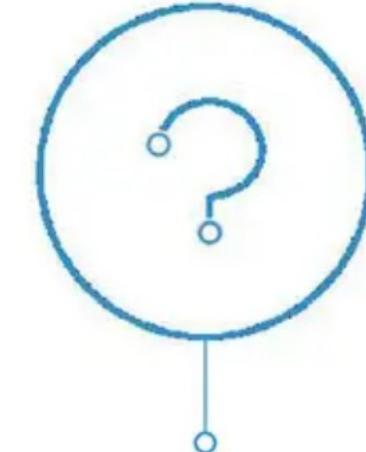
Correctness

s
Age/Income as
a negative
number



Timelines

s
Feedback
forms given
to students
from
instructor



Missing

values
Stock prices
risen, but
displaying
low on
front-end