

Why data science is important?

Ikang Fadhlí

Mathematician, Market Researcher, and
Computational Scientist



A little bit of myself



Who am I?

Bachelor in Math ITB

Find me at:
ikanx101.github.io

IG: mr.ikanx
Telegram: ikanx.bot

1st work as Market Researcher in 2008 at **Marketing Research Indonesia**, the biggest market research firm in banking industry.



After moving to several other market research firms, now I work at @nutrifood as **Market Research Specialist**.



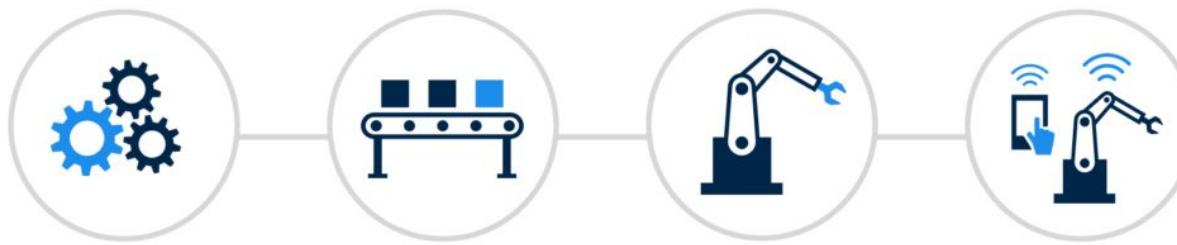
Let me start with this:

Have you ever heard of these terms?

- **Industry 4.0**
- **Data scientist**
- **Data engineer**
- **Data analyst**
- **Machine learning**
- **Artificial intelligence**
- **Big data**

So what is actually industry 4.0?

The Four Industrial Revolutions



Industry 1.0

Mechanization and the introduction of steam and water power

Industry 2.0

Mass production assembly lines using electrical power

Industry 3.0

Automated production, computers, IT-systems and robotics

Industry 4.0

The Smart Factory. Autonomous systems, IoT, machine learning

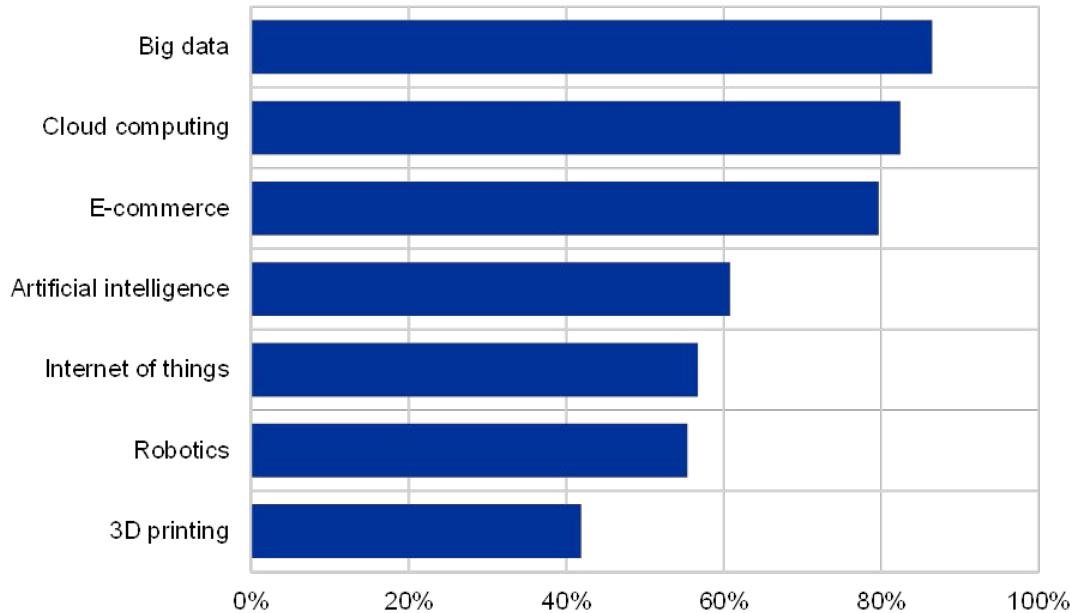
So what is actually industry 4.0?

The Four Industrial Revolutions





What is behind industry 4.0?



Focus in:





Competing on Analytics

By Professor Thomas Davenport

Every companies can sell same products, can provide same services.

So, what is the differences?

<https://hbr.org/2006/01/competing-on-analytics>

Employees hired for their expertise with numbers or trained to recognize their importance are armed with the best evidence and the best quantitative tools. As a result, they make the best decisions.

Although numerous organizations are embracing analytics, only a handful have achieved this level of proficiency. But analytics competitors are the leaders in their varied fields—consumer products, finance, retail, and travel and entertainment among them. Analytics has been instrumental to Capital One, which has exceeded 20% growth in earnings per share every year since it became a public company. It has allowed Amazon to dominate online retailing and turn a profit despite enormous investments in growth and infrastructure. In sports, the real secret weapon isn't steroids, but stats, as dramatic victories by the Boston Red Sox, the New England Patriots, and the Oakland A's attest.

We have mentioned it several times.

But what is data?



Data adalah representasi faktual dari suatu pengamatan.



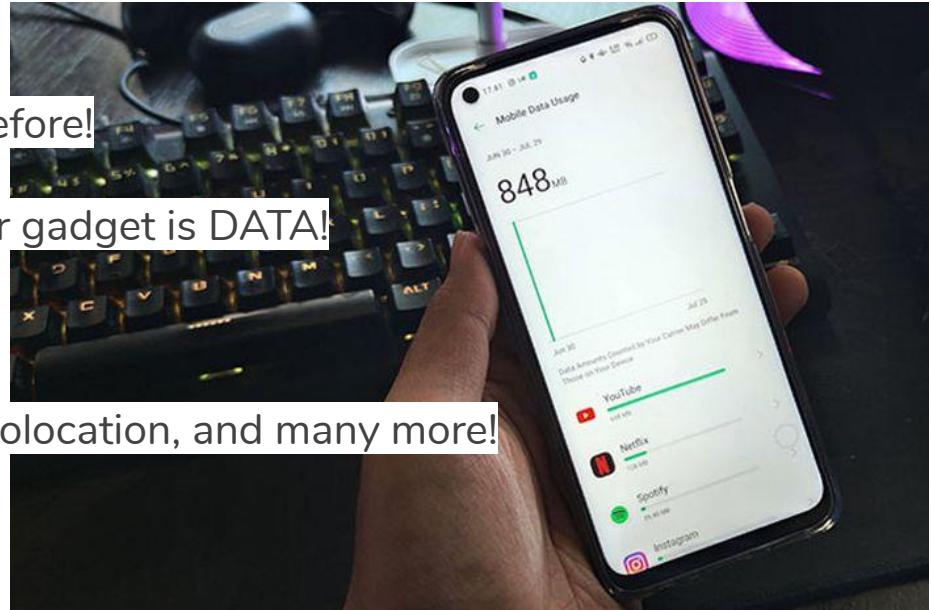
Data is not about Excel Table again nowadays!

It is more complicated even before!

Every bit we produce with our gadget is DATA!

Yes!

Your photos, videos, audio, geolocation, and many more!





How to handle those types of data?

For some of you, processing data with Excel is easy and fun.
Others may feel this is tiring work.

But for unstructured data such as photos, videos, audios,
geolocation it is impossible to be analyze using **traditional tools**
like Excel.



Did I say Excel is traditional? Yes....!



This leads us to a new job titles:

1 Data Engineer

A Data Engineer is a person who specializes in preparing data for analytical usage.

2 Data Analyst

A data analyst is a person who extract information from a given pool of data.

3 Data Scientist

A data scientist is a person who possess knowledge of statistical tools and programming skills.

Moreover, a data scientist possesses knowledge of **machine learning** algorithms.



Still confused? Let me explain in a simple way: Covid 19 Data.

Kasus

Lokasi	Dikonfirmasi ↓	Kasus baru (60 hari terakhir)	Kasus per 1 jt orang	Sembuh	Meninggal Dunia
🌐 Seluruh dunia	11.579.837		1.489	6.269.860	536.814
🇺🇸 Amerika Serikat	2.992.350		9.080	900.629	132.399
🇧🇷 Brasil	1.626.071		7.694	978.615	65.556
🇮🇳 India	719.665		529	439.947	20.160
🇷🇺 Rusia	687.862		4.687	454.329	10.296
🇵🇪 Peru	305.703		9.514	197.619	10.772
🇨🇱 Chili	298.557		15.625	264.371	6.384

Data Engineer:

A person who prepares the platform for data to be stored and analyzed.

Data Analyst:

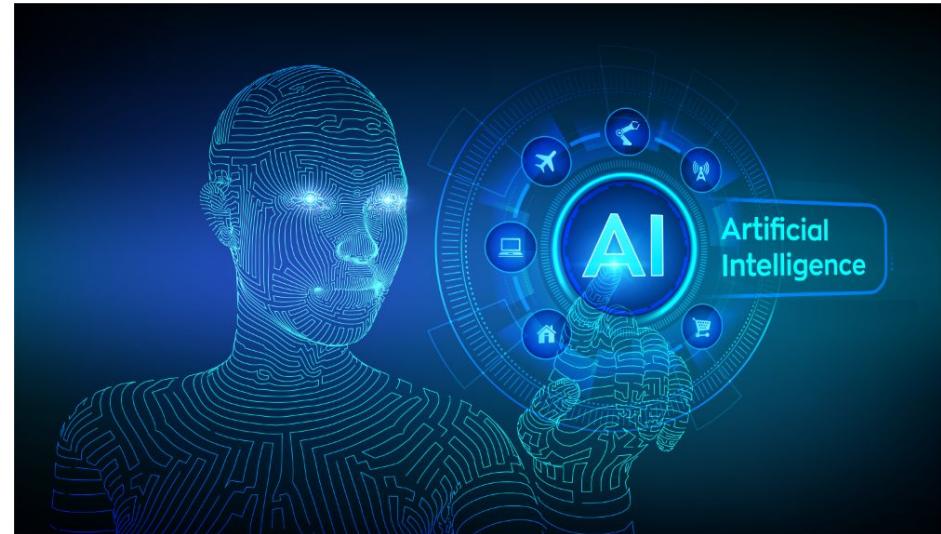
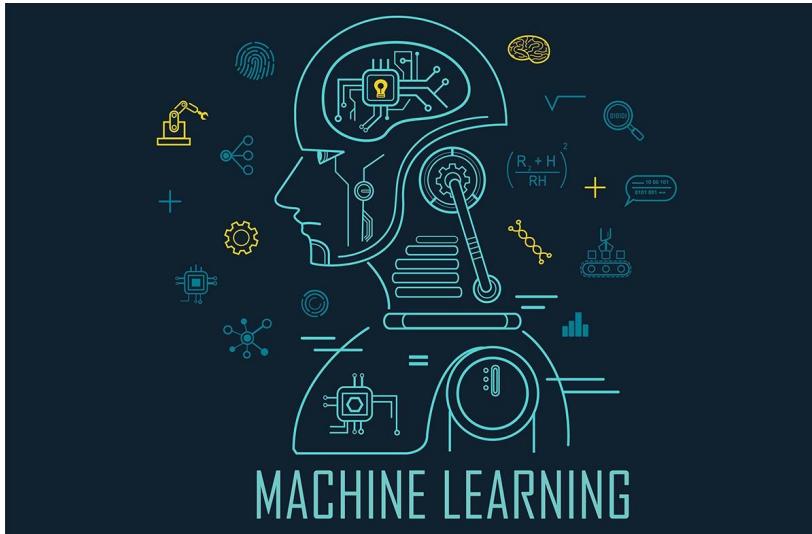
A person who makes a description or narration from the data.

Data Scientist:

A person who makes a math or statistical model to predict or deep dive analysis..



This leads us to a new terms:



It is like teenagers talking about sex.

Many people talk about it, but how many people doing it?



Is this AI?

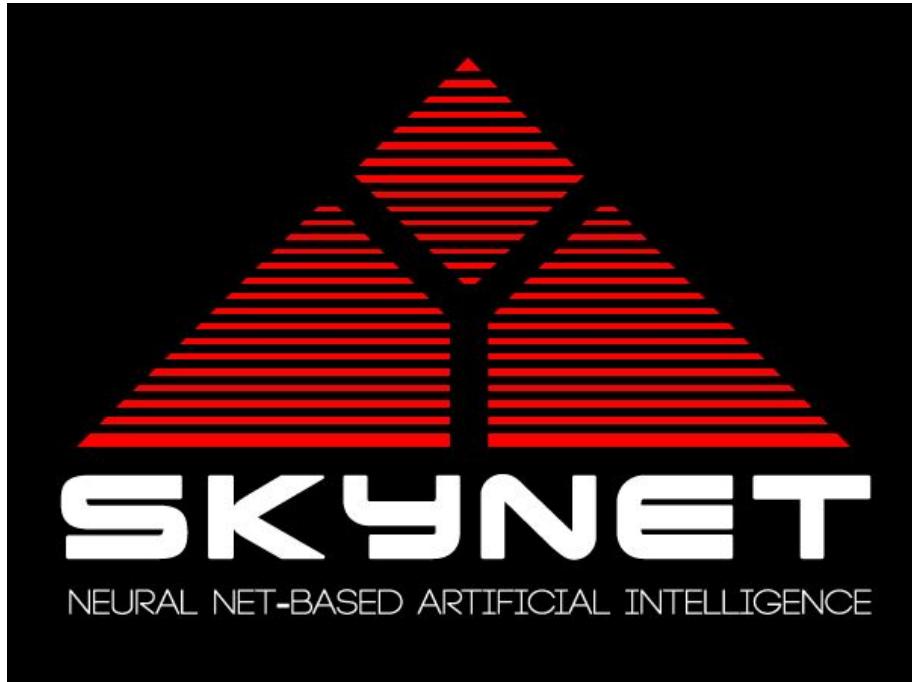


AI can cause world war III

<https://www.businessinsider.sg/elon-musk-thinks-ai-will-be-the-cause-of-world-war-iii-2017-9/?r=US&lR=T>



Is this AI?



Okay Google:

Are you skynet?



Is this AI?



AI will take our jobs in the future!

Economic transformation not annihilation

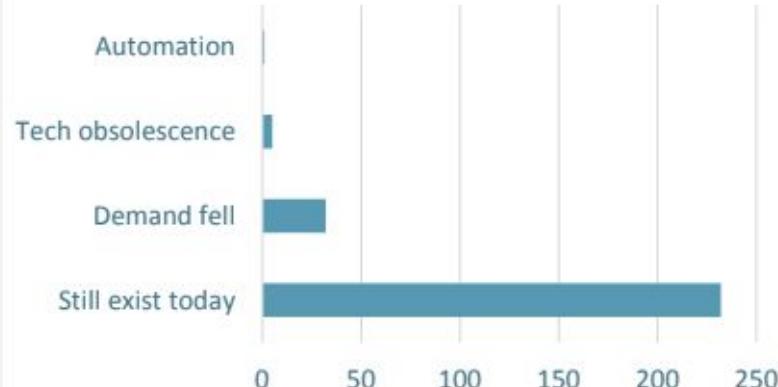
Previous studies have suggested:

- 45% of daily tasks currently done by humans could be automated (1)
- 47% of jobs in the US will be lost to automation (2)

The reality:

- Only 14% of jobs in OECD countries are "highly automatable." (3)

What happened to the 270 jobs listed in the 1950 Census?



1) Source: How Many of Your Daily Tasks Could be Automated? Chui, Mireyka, Miremadi, 2015

2) Source: The Future of Unemployment, Frey, Osborne, 2013

3) Source: Automation, Skills Use and Training, Nedelkoska and Quintini, 2018

Source: How Computer Automation Affects Occupations: Technology, Jobs, And Skills, Bessen, 2016



Cobotics and simbio-intelligence

Repetitive jobs that are being replaced are a *design* problem

Do you remember when people thought the ATM would lead to a drastic reduction in bank employees?



Actually, more jobs were created in the banking industry after the introduction of the ATM!

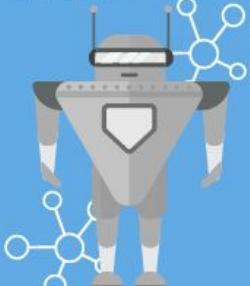
Source: Learning by Doing: The Real Connection Between Innovation, Wealth, and Welfare; Bessen, 2013



What is actually Artificial Intelligence and Machine Learning?

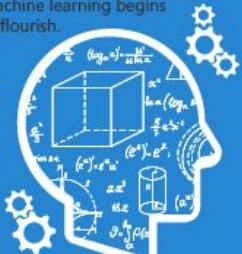
ARTIFICIAL INTELLIGENCE

Early artificial intelligence stirs excitement.



MACHINE LEARNING

Machine learning begins to flourish.



DEEP LEARNING

Deep learning breakthroughs drive AI boom.



1950's 1960's 1970's

1980's 1990's 2000's 2010's

Did you know:

Term AI has been used since the 1950s as umbrella terms for applied science such as Maths, Statistics, Physics, Engineering and Computer science.

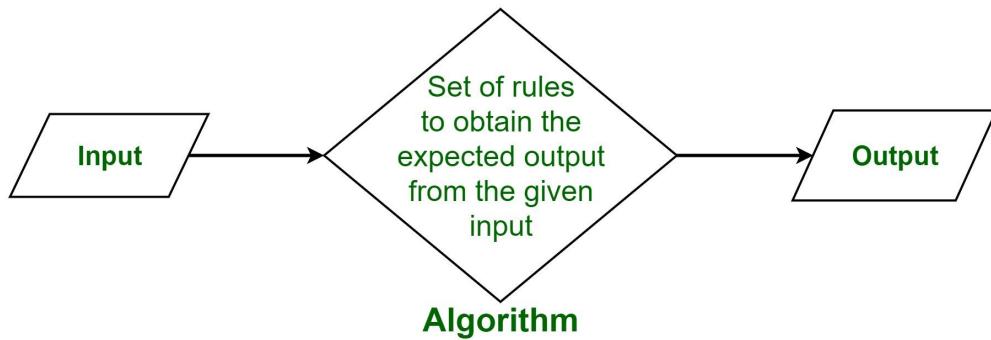
So, what is AI, ML and DL?

It is just a bunch lines of **algorithm** codes doing well-defined jobs!



This leads us to a new term again...

What is Algorithm?



Imagine a computer that you give commands to do everything you want.

You give commands to computer, not vice versa!



Now what's the benefit of data science?

Problem Statement!

Start from:



The benefits:

As an individual



As a company





As a company:

Main question:

What is the purpose of all companies?



As a company:

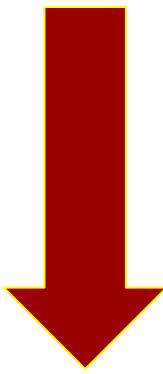
Main question:

What is the purpose of all companies?

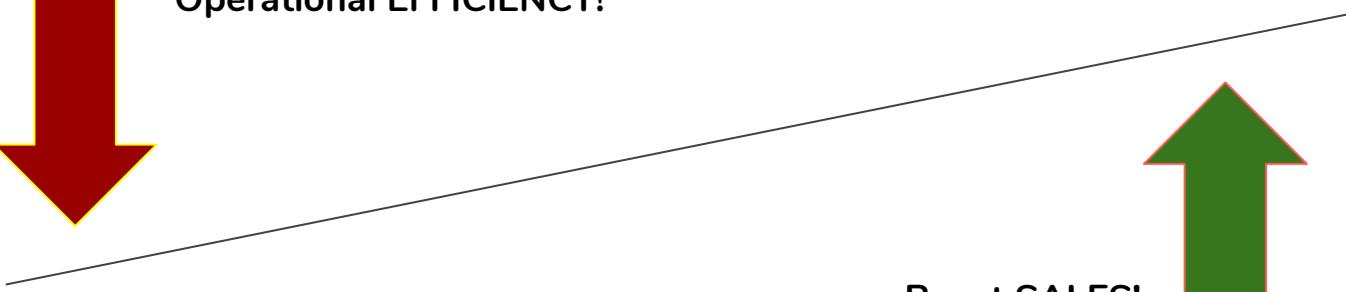




How to gain profit?



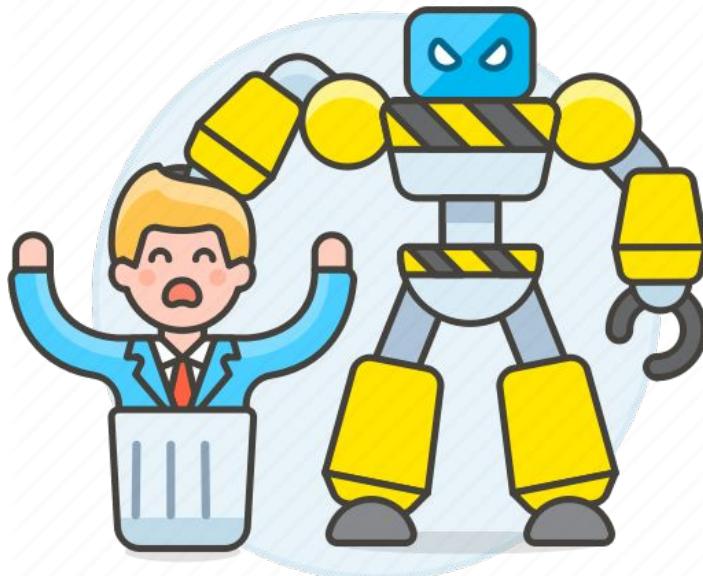
Operational EFFICIENCY!



Boost SALES!



Data science application in: operational EFFICIENCY



Replace repetitive work with algorithms
(robotic process automation).

Algorithm works faster, flawless, and
emotionless.

But, if our work is taken over by
algorithm, what will we do?



Example:

Automatic Data Entry using AI

Screenshot of RStudio showing a data entry interface for BreadTalk transaction data.

The interface includes:

- Header:** BreadTalk® RUKO SUMMARECON BEKASI
- Toolbar:** RStudio, File, Edit, Code, View, Plots, Session, Build, Debug, Profile, Tools, Help, Addins
- File Explorer:** Skrip.R, final, All Func.R
- Data View:** A table showing transaction details:

	nama.outlet	tanggal.transaksi	waktu.transaksi	nama.produk	qty	harga
1	BreadTalk RUKO SUMMARECON BEKASI	10 May 19	16:32:47	Bread Butter Pudding	1	11500
2	BreadTalk RUKO SUMMARECON BEKASI	10 May 19	16:32:47	Oream Bruille	1	14000
3	BreadTalk RUKO SUMMARECON BEKASI	10 May 19	16:32:47	Choco Croissant	1	10500
4	BreadTalk RUKO SUMMARECON BEKASI	10 May 19	16:32:47	Bank Of Chocolat	1	7500

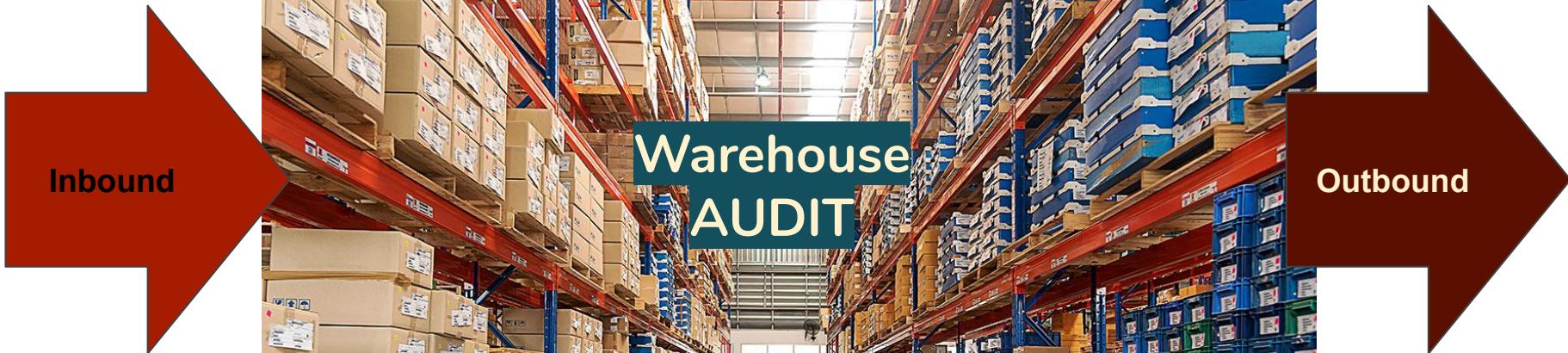
Text Overlay: Thank You
Please Come Again
Pembelian Anda gratis jika tidak
dibeli saku atau anehila Anda

Bottom Right: A small screenshot of a barcode or QR code area with numbers 68, 69, 70, 71, 72 visible.



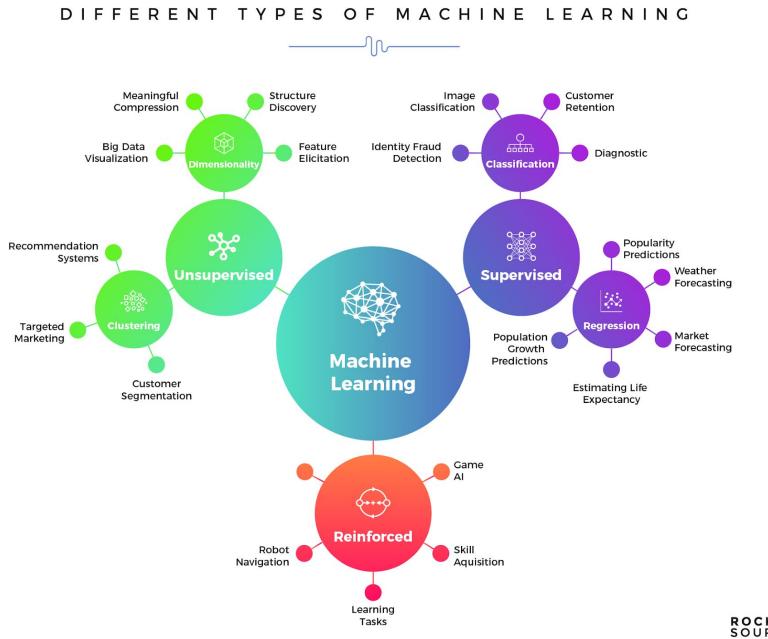
Example:

Automatic Inbound vs Outbound Matchmaking





Data science application in: boost SALES



Various prediction techniques and math models can be used to target products, consumers, distribution channels, how to advertise, prices, etc.



Example: Product bundling



If a consumer buys bread, what other products will he buy?

CROSS BASKET ANALYSIS: Determine which products can be sold in a bundle.



As an individual:

Main question:

What do we want from technology?

It helps us solve daily
problems!





When you get lost...

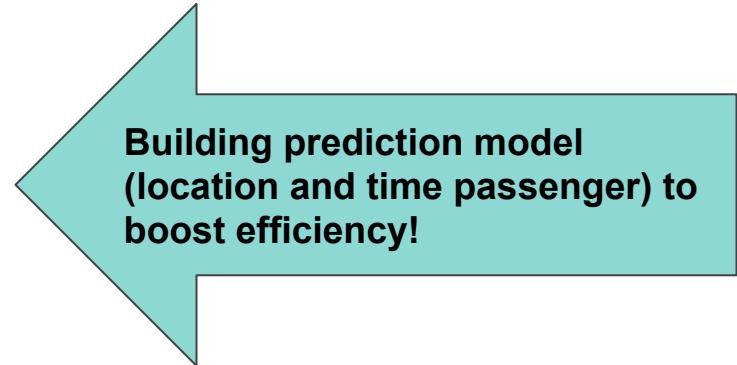


Google Maps





**These apps share same service, but
what is the difference?**



When you need something...

H. E-commerce Landscape in Indonesia





When you need to be entertained...



Spotify®



Without us knowing, we are assisted by an algorithm!



Google Maps



So, are we controlled by algorithms?

And all of your apps!



These apps are built on algorithms that have certain goals that can help its users!

```
import os
os.chdir('c:\\\\Users\\\\yourname\\\\desktop') # fails
def process_data(data): # doesn't scale
    data = complicated_function(data) # not documented
    data.to_pickle('data.pkl') # not good for big data
    data = pd.read_csv('file-i-dont-have.csv') # fails
process_data(data) # should be a DAG
```



Now it's our turn!

Simple data science projects!

Creating data science projects to do data-based journalism.



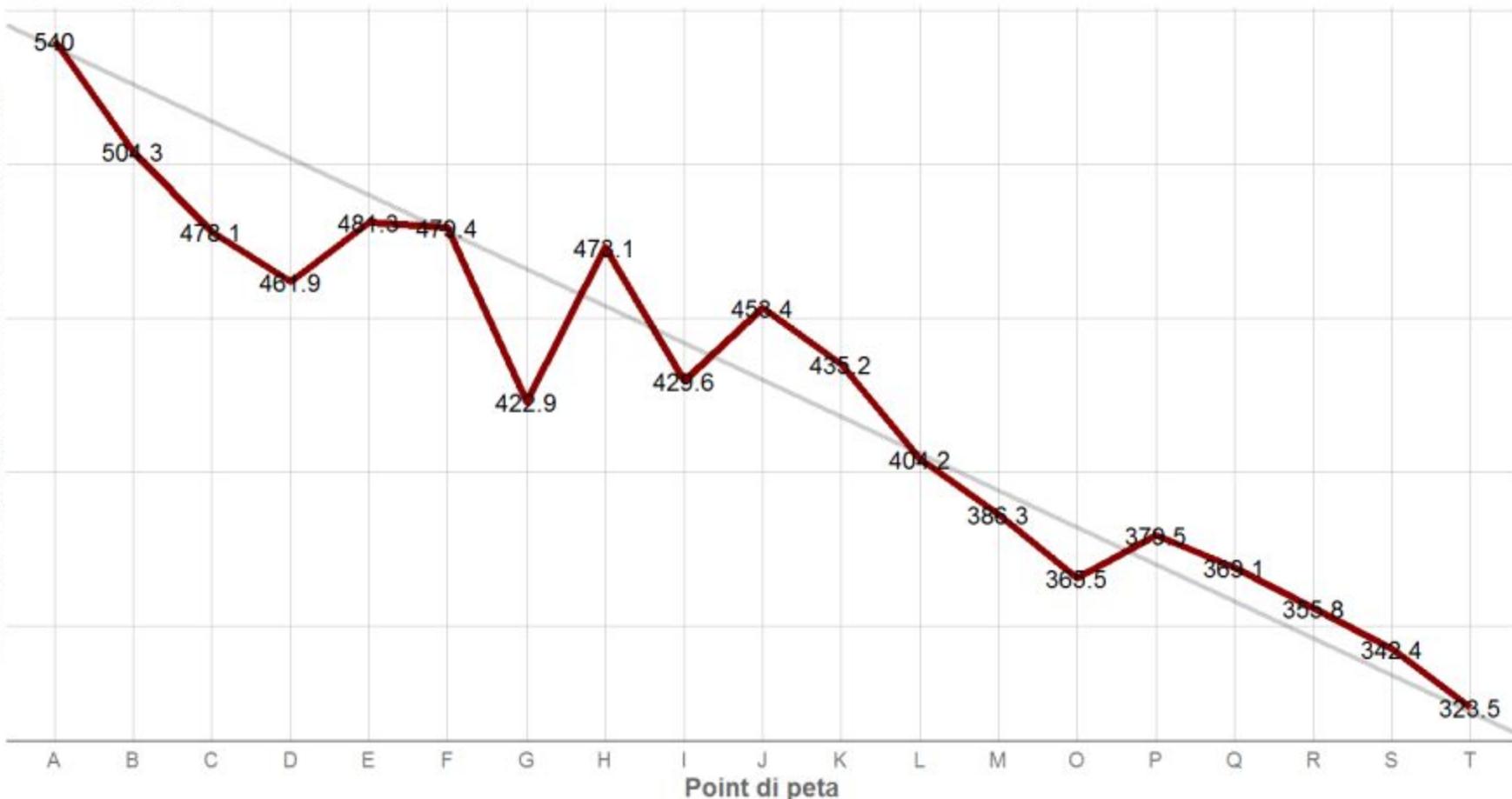
Cipularang Traffic Accidents always occurred in 90s km.

Is it really supernatural?

Hati - hati di Tol Purbaleunyi KM 97 - KM 92

Banyak orang yang tidak sadar ada "turunan" di sana...

Elevation: Ketinggian di Atas Permukaan Laut





Now it's our turn!

Simple data science projects!

Automatic data scraping from various source!



In Jan 2020, I want to sell my Daihatsu Terios TX Adventure.

What is the best price I can offer to buyers?



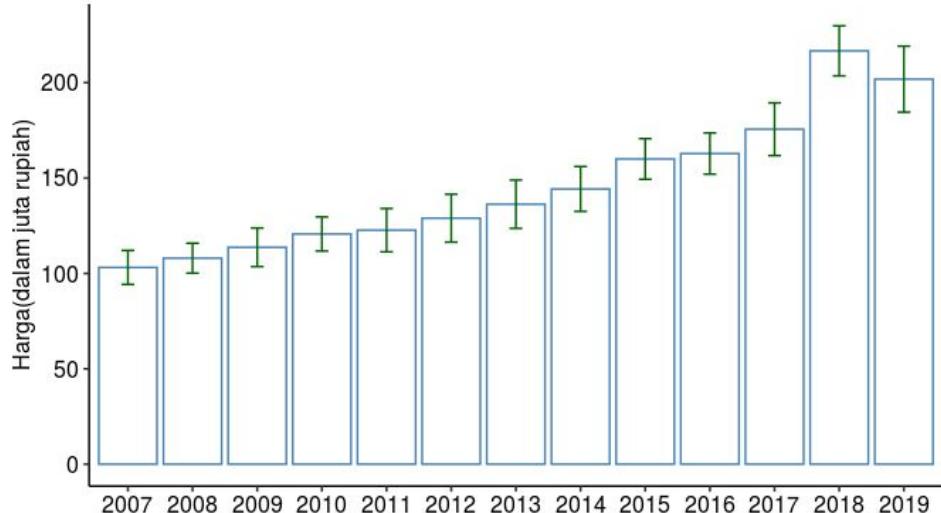
So I need to benchmark.

There's 548 Daihatsu Terios listed in the market!



Harga Daihatsu Terios Bekas 2007 - 2019

source: situs Carmudi Indonesia



Scraped 03-12-19 15:00
Visualised using R
i k A n x



Now it's our turn!

Simple data science projects!

Best travelling routes!



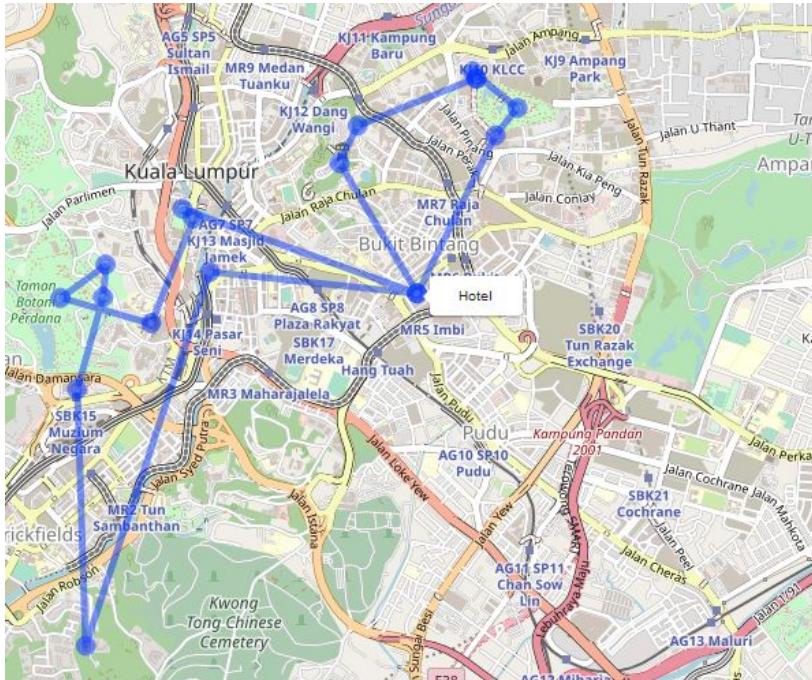
The couple are on their honeymoon to Kuala Lumpur.

Before going, they listed down several tourist destinations there.

If they only stay for 2 days, is there an optimal travel route?

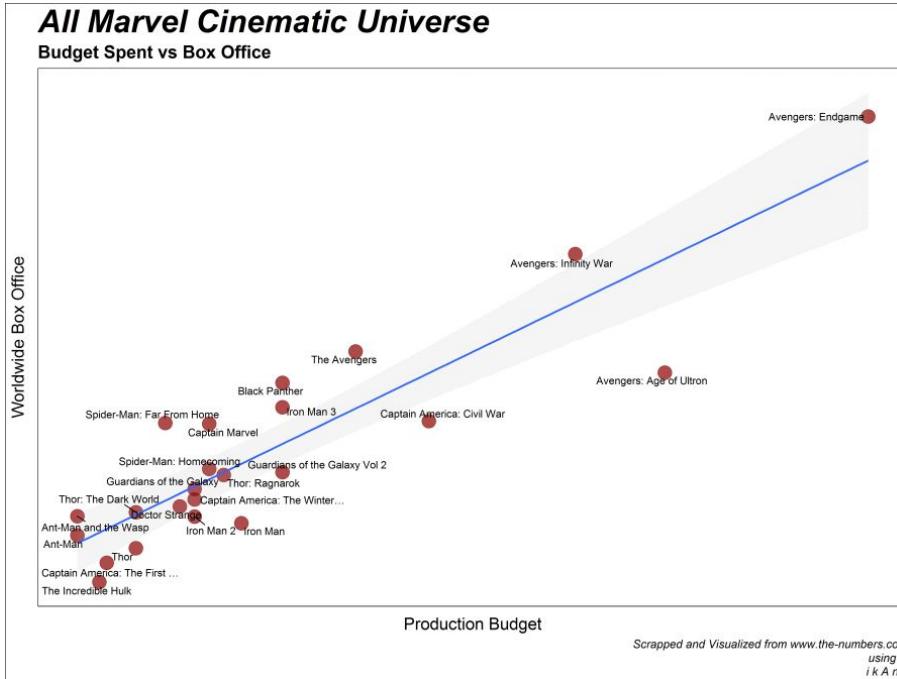


Travelling Salesperson Problem Algorithm



The algorithm works by grouping destination points with the shortest distance.

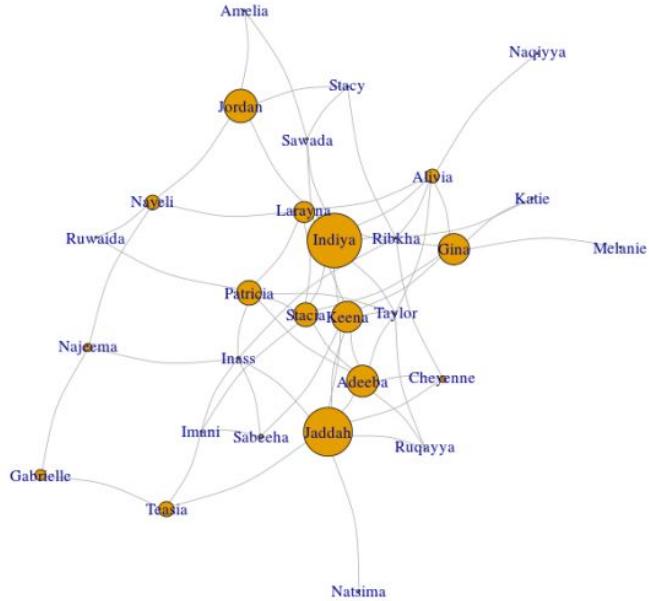
Other simple data science project we can do.



The bigger the budget, the bigger it gets!



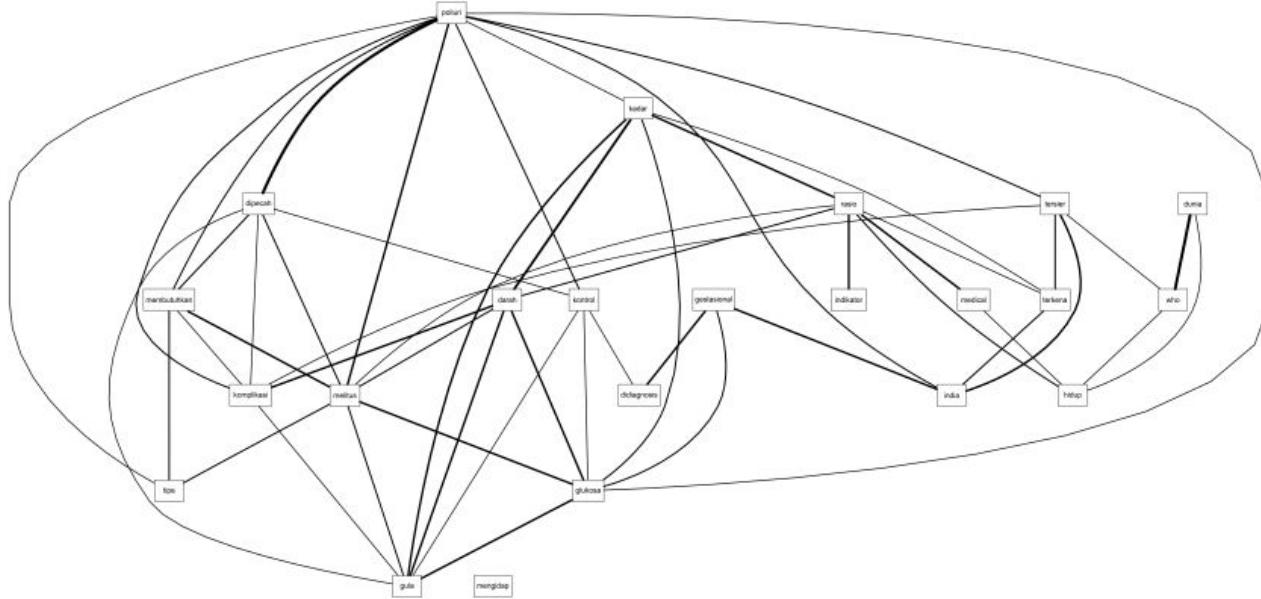
Other simple data science project we can do.



Building the best team!



Other simple data science project we can do.

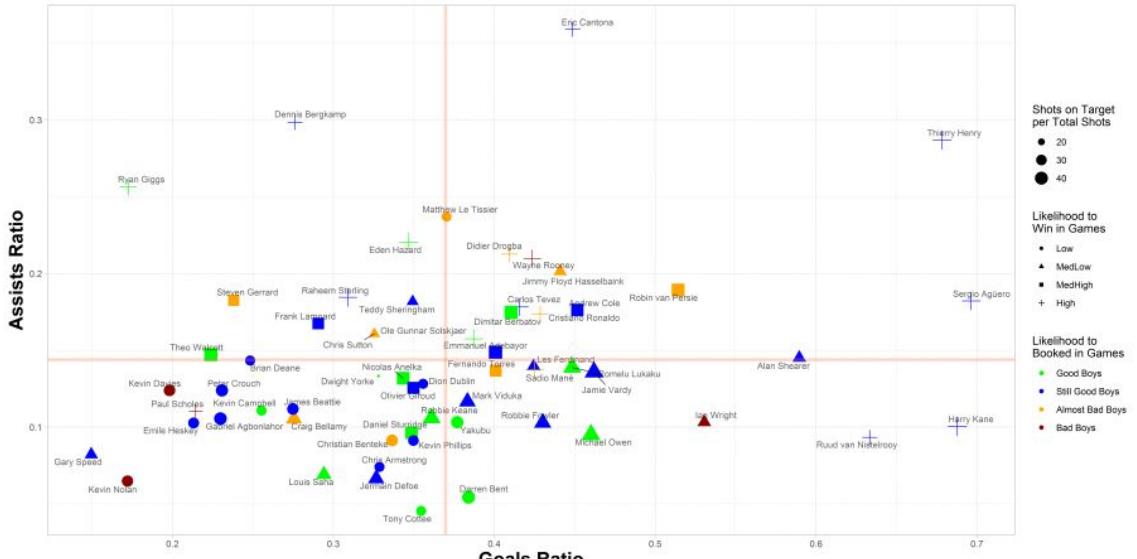


Read 341 articles
about diabetes and find
word association!

Other simple data science project we can do.

Top 60 Alltime Goal Scorer from Premier League

source: www.premierleague.com



What makes striker legendary?

Thank you

Let's discuss...