Menguasai Ilmu Data Science Dasar untuk Millenial

William Yohanes Sirait, S.Si

SEKOLAHMU

Biography

Nama : William Yohanes Sirait, S.Si

TTL : Tangerang, 10 Agustus 1992

Pendidikan : Sarjana Fisika FMIPA

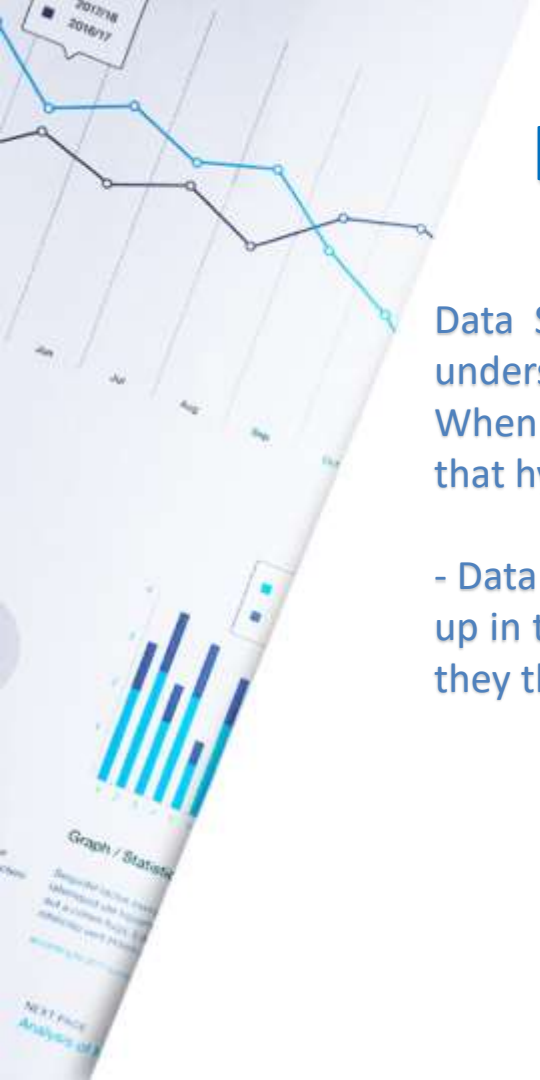
- Universitas Negeri Sebelas Maret[2016]
- Sertifikasi “Data Science Practioner”IBM Indonesia [2021]

Pekerjaaan : Guru Sains di SMAK Kanaan Tangerang [2017-Sekarang]

- Pengajar Ekstrakulikuler “Data Science”

Kontak : - Williamyohanes59@gmail.com (email)

- 087819774819 (Whatsapp)
- William yohanes (linkedIn)



History of Data Science

Data Science is a process, not an event. Its is the process of using data to understand different things, to understand the world.
When you have a model, or a hypothesis of a problem and you try to validate that hypothesis or model with your data. (IBM).

- Data Science involves data and some science. The definition, or the name, came up in the 80s and 90s when some professors were looking into the statistics, and they thoughts it would be better to call it data science.

Data Science: The Sexiest Job in the 21st Century

In the data-driven world, data scientists have emerged as a hot commodity. The chase is on to find the best talent in data science. Already, experts estimate that millions of jobs in data science might remain vacant for the lack of readily available talent. The global search for skilled data scientists is not merely a search for statisticians or computer scientists. In fact, the firms are searching for well-rounded individuals who possess the subject matter expertise, some experience in software programming and analytics, and exceptional communication skills.

Our digital footprint has expanded rapidly over the past 10 years. The size of the digital universe was roughly 130 billion gigabytes in 1995. By 2020, this number will swell to 40 trillion gigabytes.[6] Companies will compete for hundreds of thousands, if not millions, of new workers needed to navigate the digital world. No wonder the prestigious *Harvard Business Review* called data science "the sexiest job in the 21st century."[7]

A report by the McKinsey Global Institute warns of huge talent shortages for data and analytics. "By 2018, the United States alone could face a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts with the know-how to use the analysis of big data to make effective decisions."[8]

Because the digital revolution has touched every aspect of our lives, the opportunity to benefit from learning about our

There is still the need to convince the C-suite executives of the benefits of data and analytics. It appears that the senior management might be a step or two behind the middle management in being informed of the potential of analytics-driven planning. Professor Peter Fader, who manages at the Customer Analytics Initiative at Wharton, knows that executives reach the C-suite without having to interact with data. He believes that the real change will happen when executives are well-versed in data and analytics.[10]

SAP, a leader in data and analytics, reported from a survey that 92% of the responding firms in its sample experienced a significant increase in their data holdings. At the same time, three-quarters identified the need for new data science skills in their firms. Accenture believes that the demand for data scientists may outstrip supply by 250,000 in 2015 alone. A similar survey of 150 executives by KPMG in 2014 found that 85% of the respondents did not know how to analyze data. "Most organizations are unable to connect the dots because they do not fully understand how data and analytics can transform their business," Alwin Magimay, head of digital and analytics for KPMG UK, said in an interview in May 2015.[11]

Bernard Marr writing for *Forbes* also raises concerns about the insufficient analytics talent. "There just aren't enough people with the required skills to analyze and interpret this information—transforming it from raw numerical (or other) data into actionable insights—the ultimate aim of any Big Data-driven initiative," he wrote.[12] Bernard quotes a survey by Gartner of business leaders of whom more than 50% reported the lack of in-house expertise in data science.

Bernard reported on Walmart, which turned to crowdsourcing for its analytics need. Walmart approached Kaggle to host a competition for analyzing its proprietary data. The retailer provided sales data from a shortlist of stores and asked the competitors to develop better forecasts of sales based on promotion schemes.

Given the shortage of data scientists, the employers are willing to pay top dollars for the talent. Michael Chui, a principal at McKinsey, knows this too well. Data science "has become relevant to every company ... There's a war for this type of talent," he said in an interview.[13] Take Paul Minton, for example. He was making $20,000 serving tables at a restaurant. He had majored in math at college. Mr. Minton took a three-month programming course that changed everything. He made over $100,000 in 2014 as a data scientist for a web startup in San Francisco. "Six figures, right off the bat ... To me, it was astonishing," said Mr. Minton.[14]

Could Mr. Minton be exceptionally fortunate, or are such high salaries the norm? Luck had little to do with it; the *New York Times* reported $100,000 as the average base salary of a software engineer and $112,000 for data scientists.

Data Scientist Average Salary in Indonesia 2021

How much money does a Data Scientist make in Indonesia?

Average Monthly Salary

19,400,000 IDR

(232,000,000 IDR yearly)

LOW	AVERAGE	HIGH
10,500,000 IDR	19,400,000 IDR	29,200,000 IDR

Switch to yearly salaries

A person working as a **Data Scientist** in **Indonesia** typically earns around **19,400,000 IDR** per month. Salaries range from **10,500,000 IDR** (lowest) to **29,200,000 IDR** (highest).

This is the average monthly salary including housing, transport, and other benefits. Data

The Profesion of Data Science

Data Science

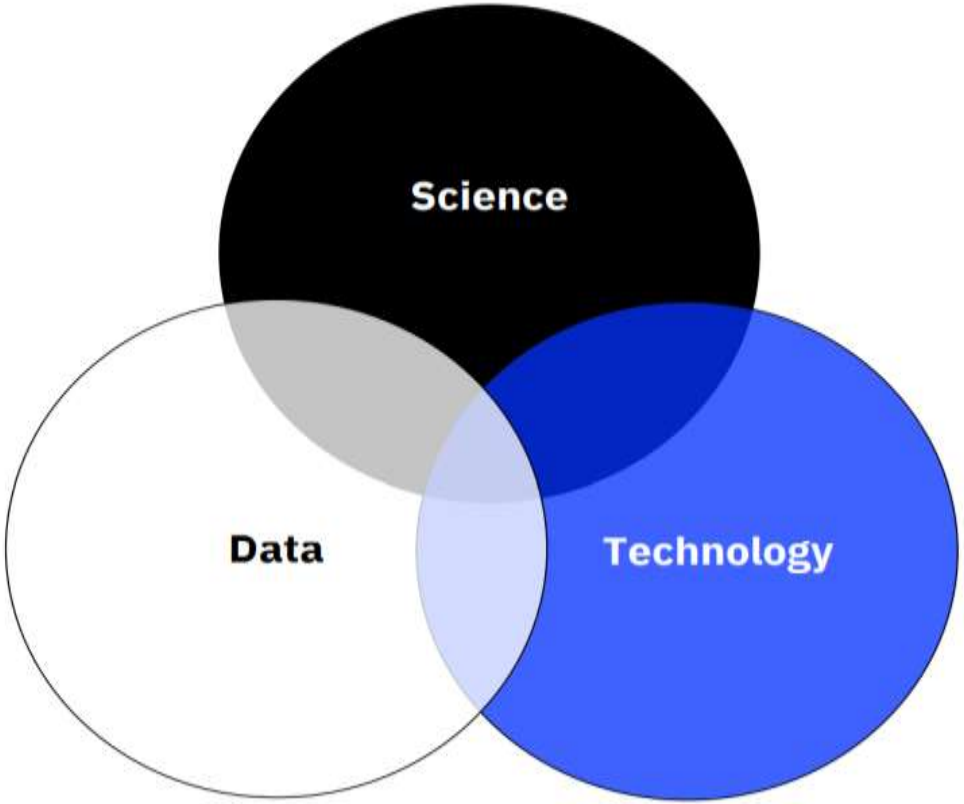
- Machine learning
- Statistical modeling
- Experiment design
- Statistics, research, mathematics

Data Journalism

- Domain expertise
- Strategic problem solving
- Business acumen
- Communication skills
- Visualization skills
- Decision making based on insights

Data Engineering

- Database and data storage
- Scripting language
- Artificial Intelligence
- Cloud Infrastructure
- Statistical computing



Team behind data science

Sally

Data Analyst

28 years old | 6 years of experience

Maria

Data Scientist

40 years old | 15 years of experience

Tom

Data Engineer

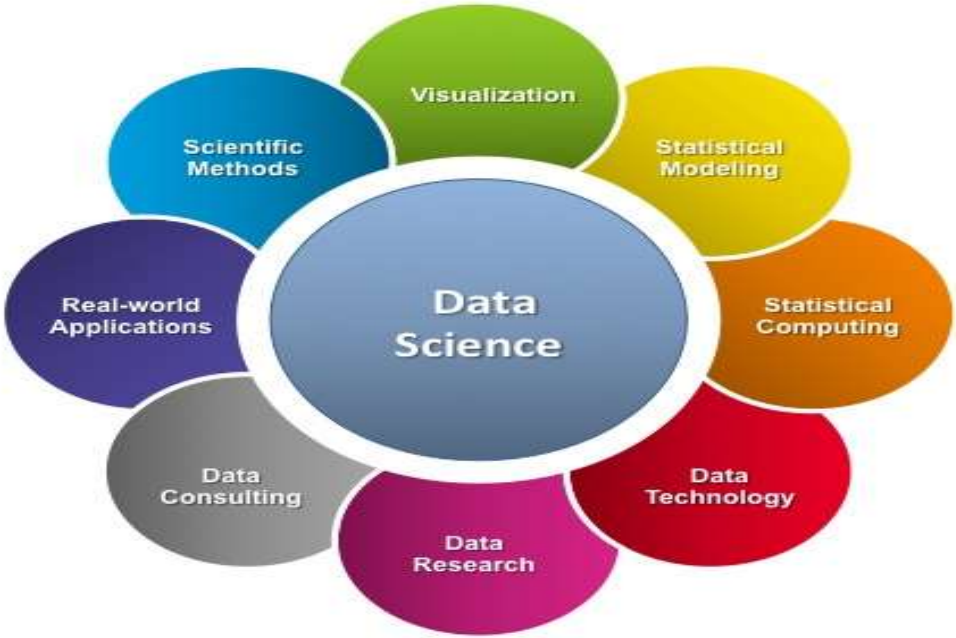
36 years old | 10 years of experience

Jokowi Klaim Google hingga Microsoft Mau Bangun Pusat Data di Indonesia

Andhika Prasetia - detikNews

Jumat, 28 Feb 2020 14:50 WIB

Data Science Framework

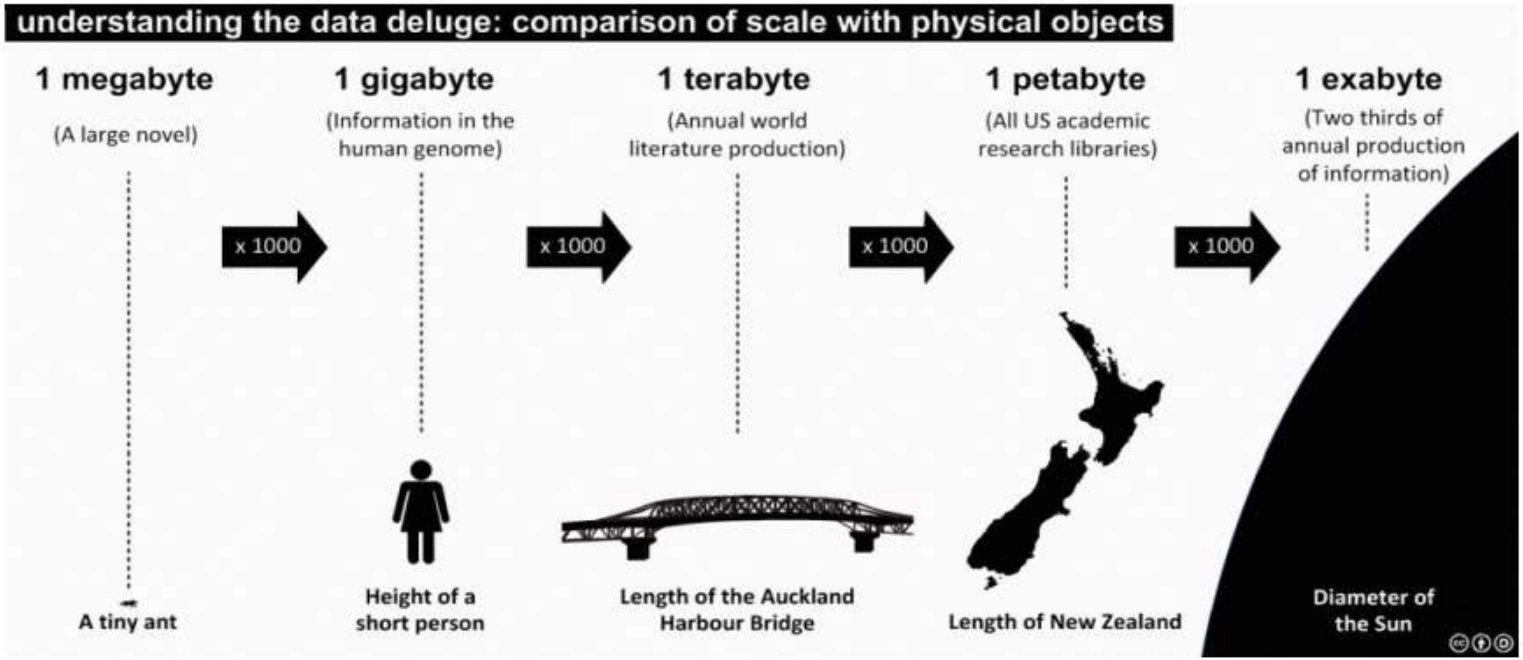


Big Data

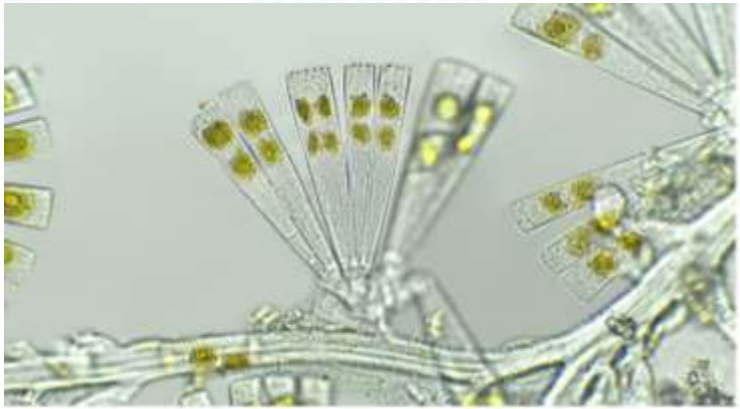
Big Data is high-volume, high-velocity, and/or high-variety information assests that demand cost effective, innovative forms of information processing that enable processing tenable enhanced insight, decision making and process automation. (The Research Film Gartner)

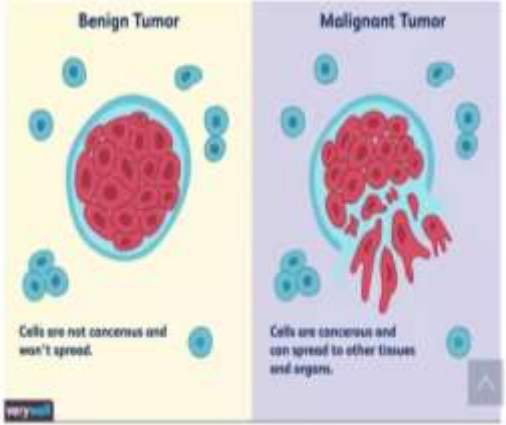
Karakteristik Big Data :

1. Velocity : is the speed of the data, or the speed at which data accumulates
2. Veracity : is the conformity to facts and accuracy
3. Volume : is the scale of the data or the increase in the amount of data stored
4. Value : our ability and need to turn data into Value. Values isn't just profit. It may be medical or social benefits, or customer, employee, or personal satisfaction
5. Variety is the diversity of the data. We have structured data that fits neatly into rows and columns, or relational databases and unstructured data that is not organized in a pre-define way, for example tweets, blogspots, pictures, numbers and even video data


understanding the data deluge: comparison of scale with physical objects
1 megabyte
(A large novel)
A tiny ant
x 1000
1 gigabyte
(Information in the human genome)
Height of a short person
x 1000
1 terabyte
(Annual world literature production)
Length of the Auckland Harbour Bridge
x 1000
1 petabyte
(All US academic research libraries)
Length of New Zealand
x 1000
1 exabyte
(Two thirds of annual production of information)
Diameter of the Sun

Application of Data Science in daily life

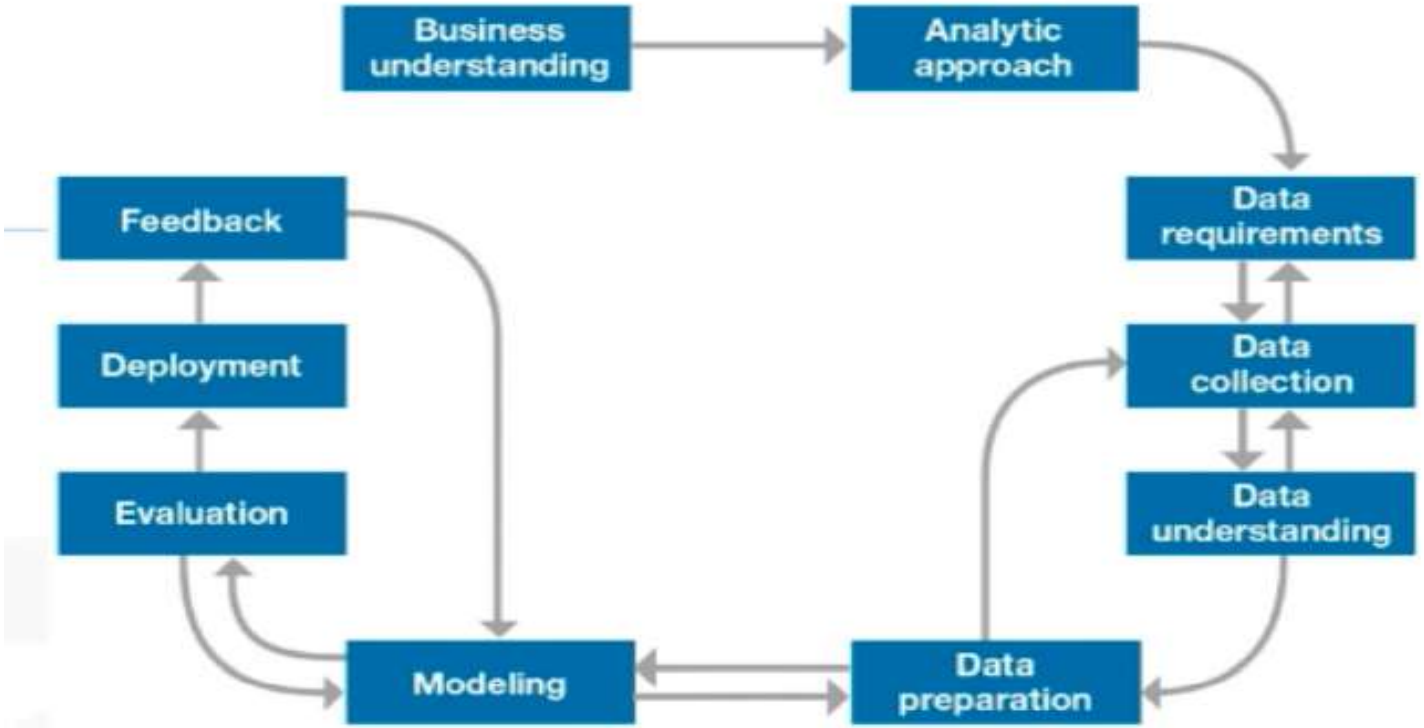




The Importance Knowledge of Data Science

1. Melatih kemampuan siswa untuk melek “riset” sejak dini
2. Melatih kemampuan untuk mampu mempresentasikan hasil riset dengan baik dan benar sesuai kaidah keilmiahan
3. Membantu mahasiswa untuk menyelesaikan tugas besar dengan baik dan memuaskan
4. Membantu mahasiswa untuk sukses dalam tugas besar/skripsi

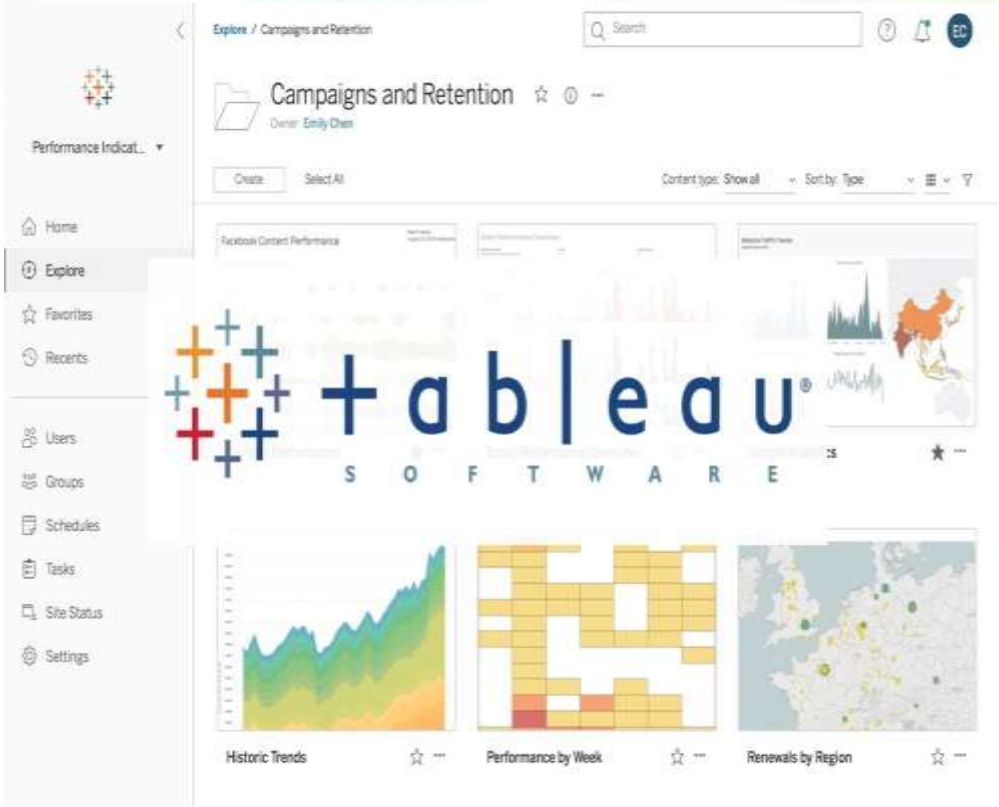
Data Science Methodology



Data Science Tools

Platform	Apache Spark, hadoop
Tools	tableau, jupyter, RStudio, QlikView
Frameworks	Spark MLlib, learn, TensorFlow
Language	python, R, Java, Scala
Format	Parquet, pandas
SaaS	Google Machine Learning, Amazon Machine Learning, Azure Machine Learning

Output Program



What were the global affects of Tuberculosis (TB) between 1990 and 2013?

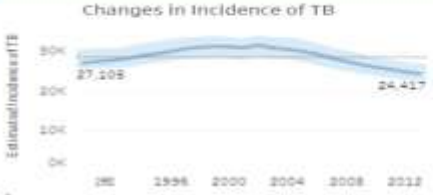
Year: 1990

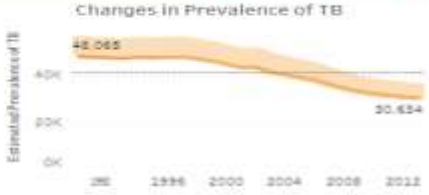
Region: (All)

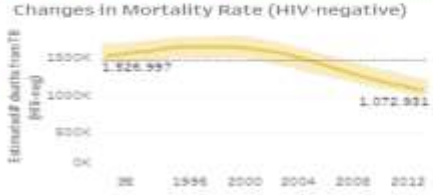
Worldwide death rates due to TB

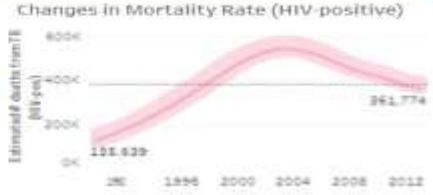
TB Case Detection Rate

General affects of TB over the years










File Data Worksheet Dashboard Story Analysis Map Format Server Window Help
Show Me
Dashboard Layout
Device Preview
Size
Legal Landscape (1150 x 7...
Sheets
YOY Sales by Reg...
Customer Acquis...
Customer Acquis...
New Customers
CYTD vs. PYTD
Adjustable CYTD ...
CMTD vs. PMTD
Search Parameter
Slope Graph
Slope Graph 2
Quadrants
Donut Graph
Standard Deviation
Waterfall Chart
Bump Chart
Market Basket A...
Conditional Form...
New Max
Top 10 and Botto...
Multi-Tab
Calendar
Yearly Min/Max v...
Objects
Horizontal
Image
Vertical
Web Page
Text
Blank
Tiled
Floating
Show dashboard title
interworks
Sales Performance
West
Central
South
East
Sales
Jan '11
Dec '14
$725,458
$501,240
$391,722
$678,781
Sales by State
© OpenStreetMap contributors
Sales by SubCategory
Phones
Chairs
Storage
Tables
Binders
Machines
Accessories
Copiers
Bookcases
Appliances
Furnishings
Paper
Supplies
Art
Envelopes
Labels
Fasteners
0K
50K
100K
150K
200K
250K
300K
Select Measure
Sales over Time for All
West
Dec '13
$32,617
Central
Sep '11
$34,409
South
Mar '11
$32,911
East
Nov '14
$40K
$20K
$0K
Jan '11
Jan '12
Jan '13
Jan '14
Jan '15

Program Activity

1. Instalasi aplikasi “Tableau” dan memahami tools dan fungsi pada “Tableau”
2. Mengenal jenis-jenis data statistik pada Data Science
3. Memahami kiat-kiat dalam memvisulisasikan Data
4. Mempresentasikan hasil olahan dan visulisasi data yang diperoleh dari sumber dataset di Internet (additional : Dasar Machine learning dan software ML yaitu Orange)

The Benefit from Program

1. Free Registrasi dan Instalasi aplikasi “Tableau”
2. Free 3 Ebook (IBM Skills Academy) tentang Data Science + Open Course Slide tiap pertemuan.
3. Konsultasi gratis via WA
4. Sertifikat

Data Visualization & Presentation

LECTURE 6

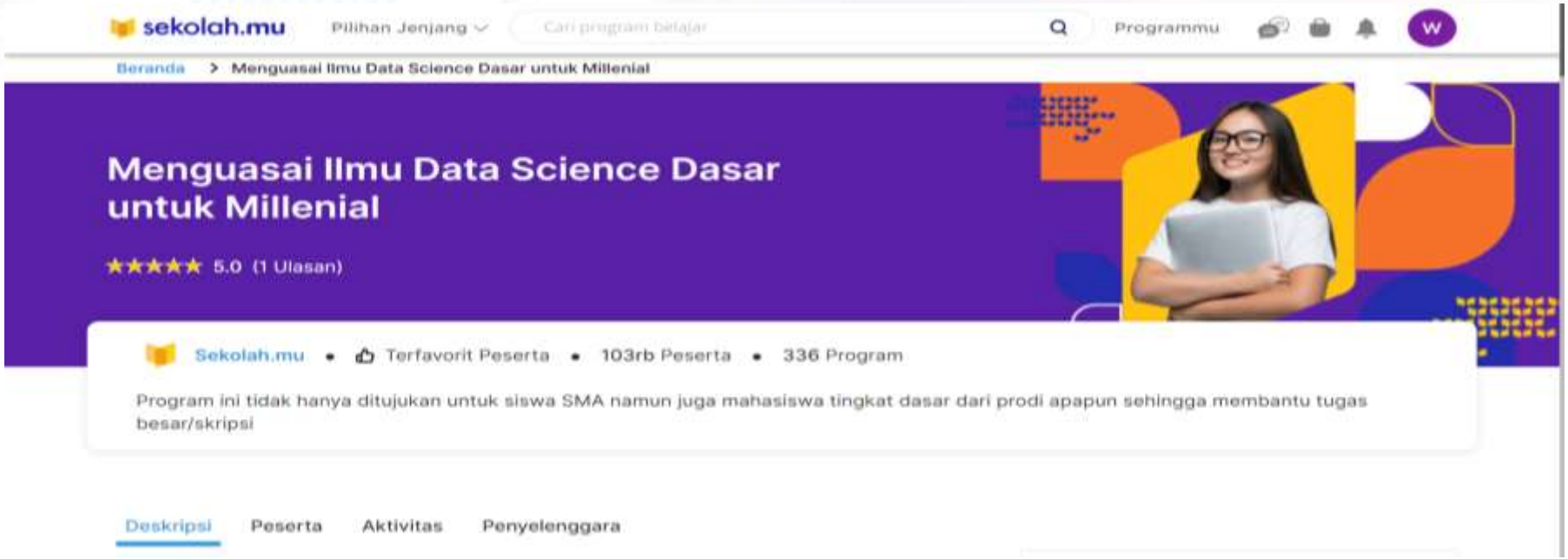
Version: 2020.04.08

IBM

IBM Design Thinking

PERTEMUAN XXVIII

The Homepage of Program



https://www.sekolah.mu/program/menguasai-ilmu-data-science-dasar-untuk-millenial

Reference

https://aceemerge.com/data-science/
https://news.detik.com/berita/d-4918523/jokowi-klaim-google-hingga-microsoft-mau-bangun-pusat-data-di-indonesia
https://interworks.com/blog/dyu/2016/10/05/scaling-tableau-dashboards-high-res-monitors/
https://www.thedataschool.co.uk/janasobini-vetharuban/my-experience-of-common-practices-that-led-to-ineffective-dashboard-designs-in-tableau
http://www.salaryexplorer.com/salary survey.php?loc=101&loctype=1&job=10202&jobtype=3

Haider, Murtaza, 2015. *Getting Started with Data Science*. IBM Press