

# Data Analytics Pipeline

Introduction to Data Analysis With Excel and SQL



Data Science Learning Studio

09 January 2022





**Bagoes Suryo Nugroho**

Panggil aja **Mas Goes**



**1920-2020**

**PERINGATAN  
100 TAHUN  
PENDIDIKAN  
TINGGI TEKNIK  
DI INDONESIA**



**Felilo**



# Aktivitas Saat Ini

**Lion**  **parcel**

**REFERAN!**  
**DIANDI**



1

Data Engineer at Lion Parcel

2

Strategic Partnership at DSI

3

Agile Innovation Facilitator at CIAS

4

Individual Data Consultant



# Bab Materi

1

Introduction to SQL and Excel

2

Introduction to Excel for Analysis

3

Introduction to SQL for Analysis

4

Hands on !

Data Analysis Using SQL and Excel



# THE DATA SCIENCE HIERARCHY OF NEEDS

LEARN/OPTIMIZE

AGGREGATE/LABEL

EXPLORE/TRANSFORM

MOVE/STORE

COLLECT

AI,  
DEEP  
LEARNING

A/B TESTING,  
EXPERIMENTATION,  
SIMPLE ML ALGORITHMS

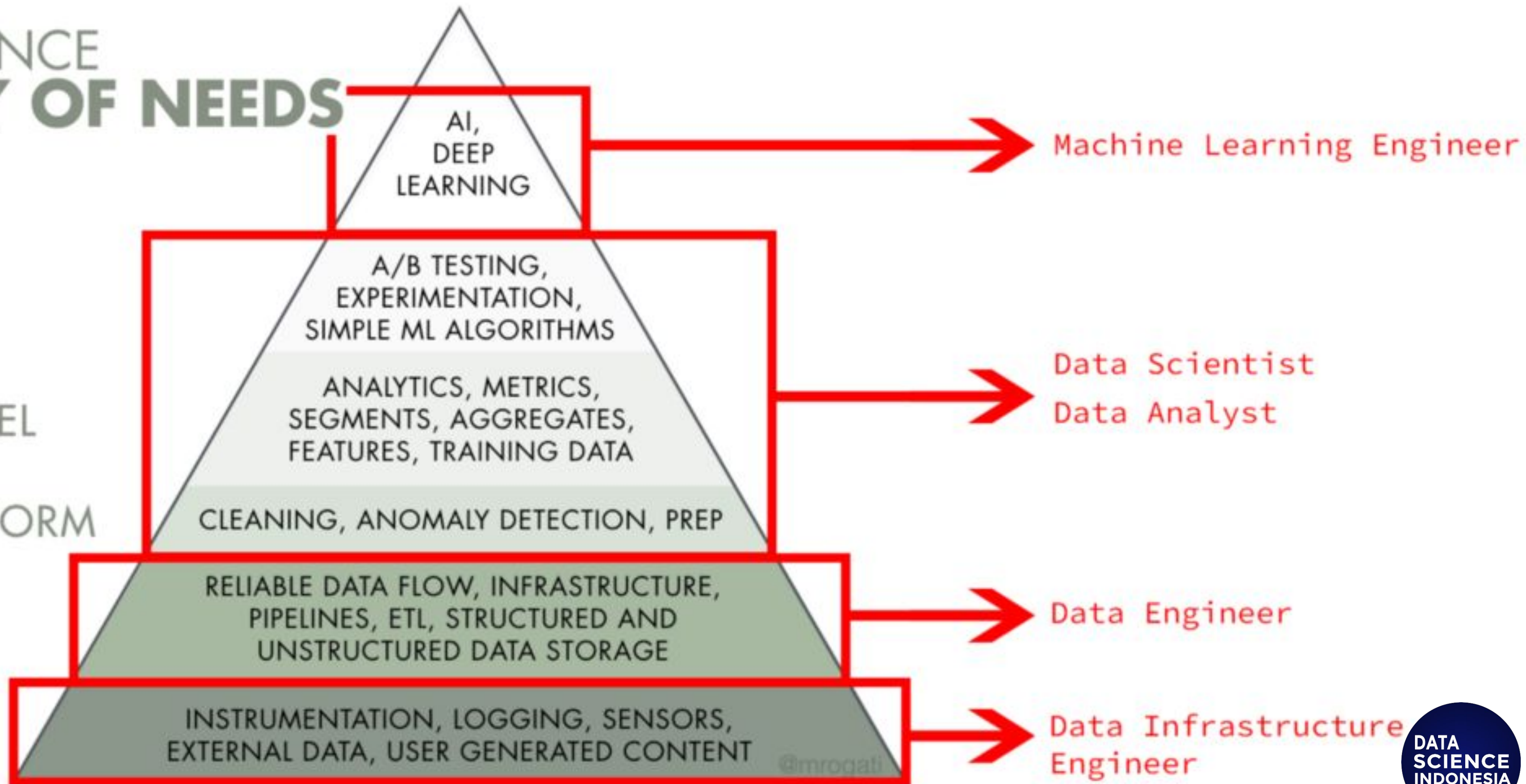
ANALYTICS, METRICS,  
SEGMENTS, AGGREGATES,  
FEATURES, TRAINING DATA

CLEANING, ANOMALY DETECTION, PREP

RELIABLE DATA FLOW, INFRASTRUCTURE,  
PIPELINES, ETL, STRUCTURED AND  
UNSTRUCTURED DATA STORAGE

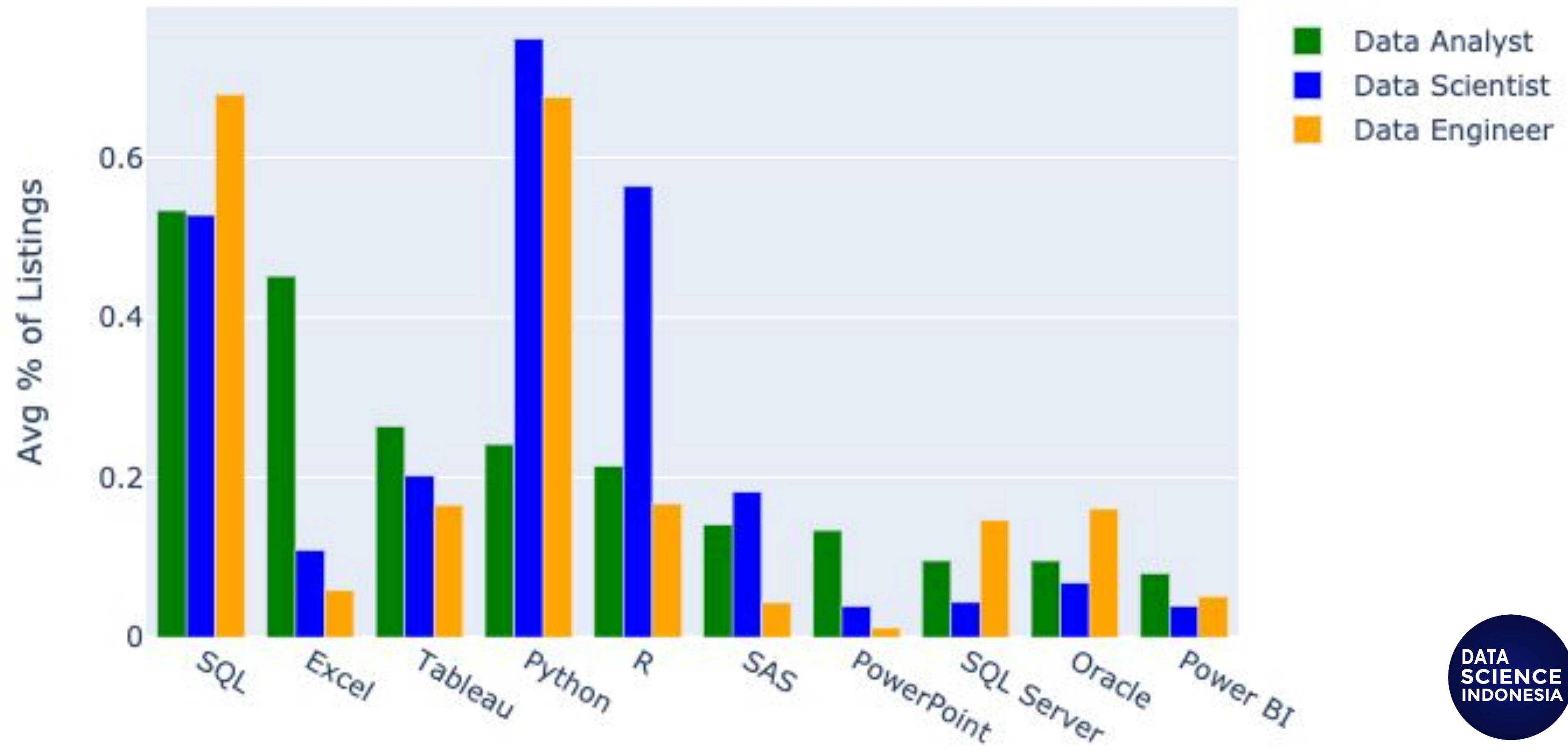
INSTRUMENTATION, LOGGING, SENSORS,  
EXTERNAL DATA, USER GENERATED CONTENT

# Pyramid of Needs





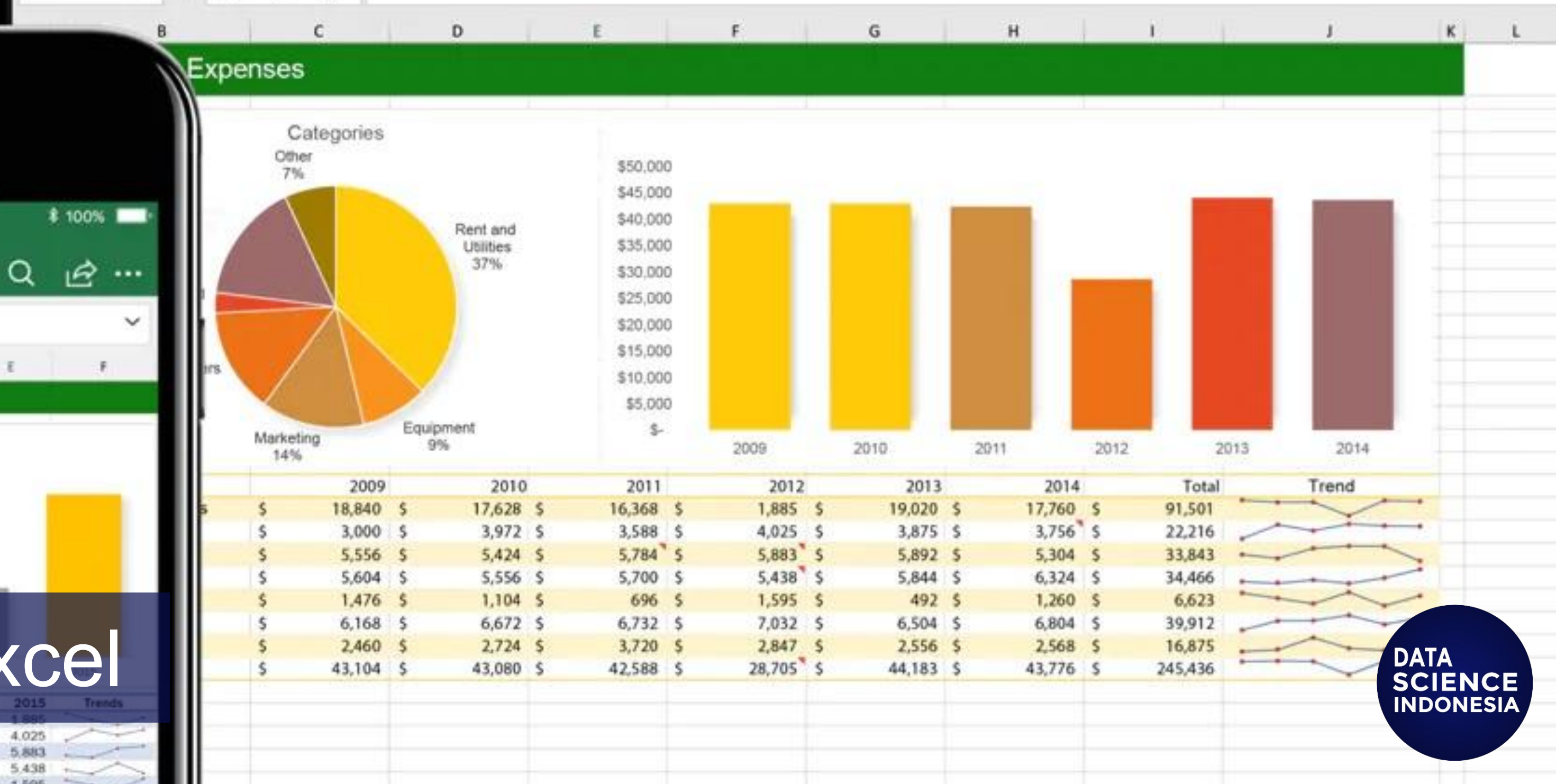
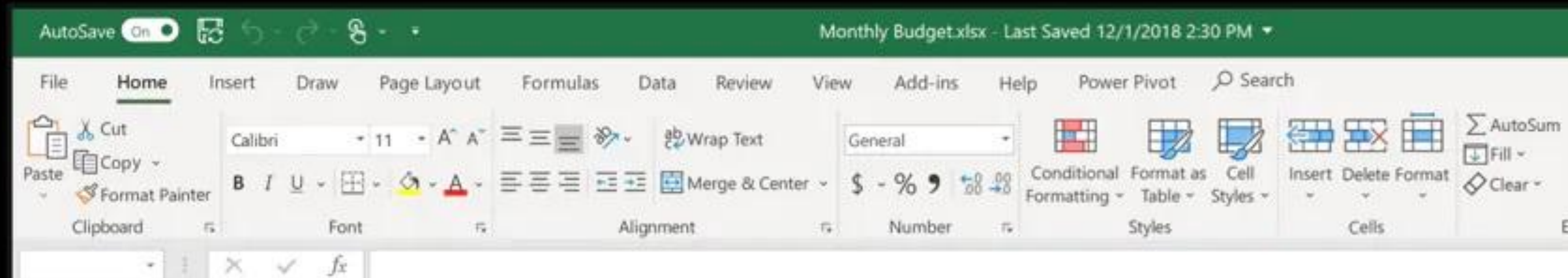
# Most Common Keywords for Data Analyst vs. Data Scientist, Data Engineer



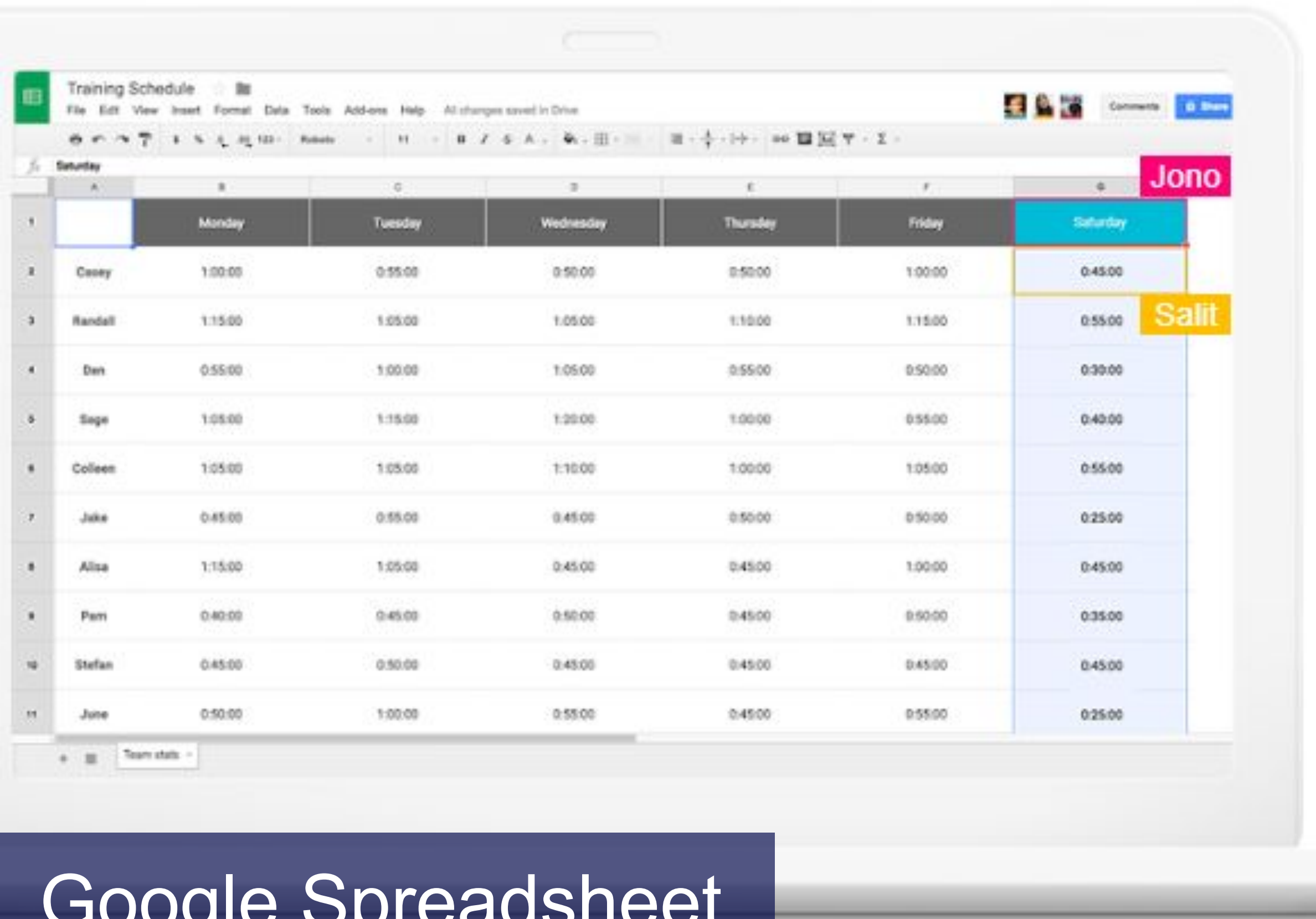


Excel





Microsoft Excel



Google Spreadsheet

Google Spreadsheet

DATA  
SCIENCE  
INDONESIA



# Microsoft Excel

- 1 Berbayar dan Offline\*  
File disimpan secara offline di local
- 2 Kurang baik untuk kerjaan Kolaboratif. Tidak bisa sharing sheets.
- 3 Kemampuan Analisis dan Visualisasinya yang sangat baik karena ada banyak fungsi yang ada

# Google Sheets

- 1 Gratis dan Online.  
File disimpan di Google Drive
- 2 Lebih baik untuk kerjaan kolaboratif. Dapat sharing sheets.
- 3 Kemampuan Analisis tidak selengkap Ms Excel dan beberapa Visualisasi dilakukan secara manual



Format CSV - XLSX



# CSV



1

Hanya menyimpan data tanpa ada formatting, operasi, macro,dll

2

Import Data CSV lebih cepat,mudah dan memakan memori yang sedikit

3

Dapat dibuka di Excel, Notepad, dan Text Editor lainnya

# XLSX



1

Dapat menyimpan operasi, macro, tabel, formatting,dll

2

Import Data XLSX lebih lambat,kompleks, dan memakan memori yang banyak

3

Hanya dapat dibuka di Excel



Invoice 29 Juli

.XLSX



Saved to Drive

File Edit View Insert Format Data Tools Help

Last edit was on September 4, 2020

100% \$ % .0 .00 123 Default (Ca... 11 B I S A

C15

fx

A

B

C

D

E

F

G

H

I

J

TOTAL

ANGKA

PENDAPATAN KOTOR

#NAME?

MODAL BERSIH

#NAME?

PROFIT PUSAT

#NAME?

PROFIT MITRA

#NAME?

ONGKIR

#NAME?

Profit mitra 11650

PENDAPATAN BERSIH

Rp

1,525,872

Profit tambahan dari tokped

Modal actual

1300000

Selisih

#NAME?

Profit

#NAME?

#NAME?

Ongkir

#NAME?

Kurir

90000

#NAME?

Bensin

0

#NAME?

0

Makan

30000

Bensin

15000

Parkir

10000

55000

#NAME?

Kelebihan  
bombay

0

plastik  
bensin  
makan  
parkir  
supirmodal + petty Rp  
modal actual Rp

Rp - selisih

Data di Excel?

DATA  
SCIENCE  
INDONESIA

TOTAL FINAL

Database

DATABASE HARGA

LIST BELANJA

AWAL

TEMPLATE PUSAT

Dita

Intan



- Users (7)
  - AHMED
  - ALI
  - FULL\_ACCES
  - OSMAN
  - PUBLIC
  - SUHAIB
  - SYSDBA
- Unicorn
- Employee
  - Query Window
  - Tables (12)
    - COUNTRY
    - JOB
    - DEPARTMENT
    - CUSTOMER
    - EMPLOYEE
    - PROJECT
    - SALES
    - EMPLOYEE\_PROJECT
    - PROJ\_DEPT\_BUDGET
    - SALARY\_HISTORY
    - SALES2
    - COUNTRY2
  - Generators (2)
    - EMP\_NO\_GEN
    - CUST\_NO\_GEN
  - Triggers (4)
  - Views (1)
  - Stored Procedures (11)

Employee: Select first 1000 from COUNTRY Employee: Select first 1000 from DEPARTMENT



```
1 select first 1000 * from DEPARTMENT;  
2  
3 select first 1000 * from COUNTRY;  
4
```

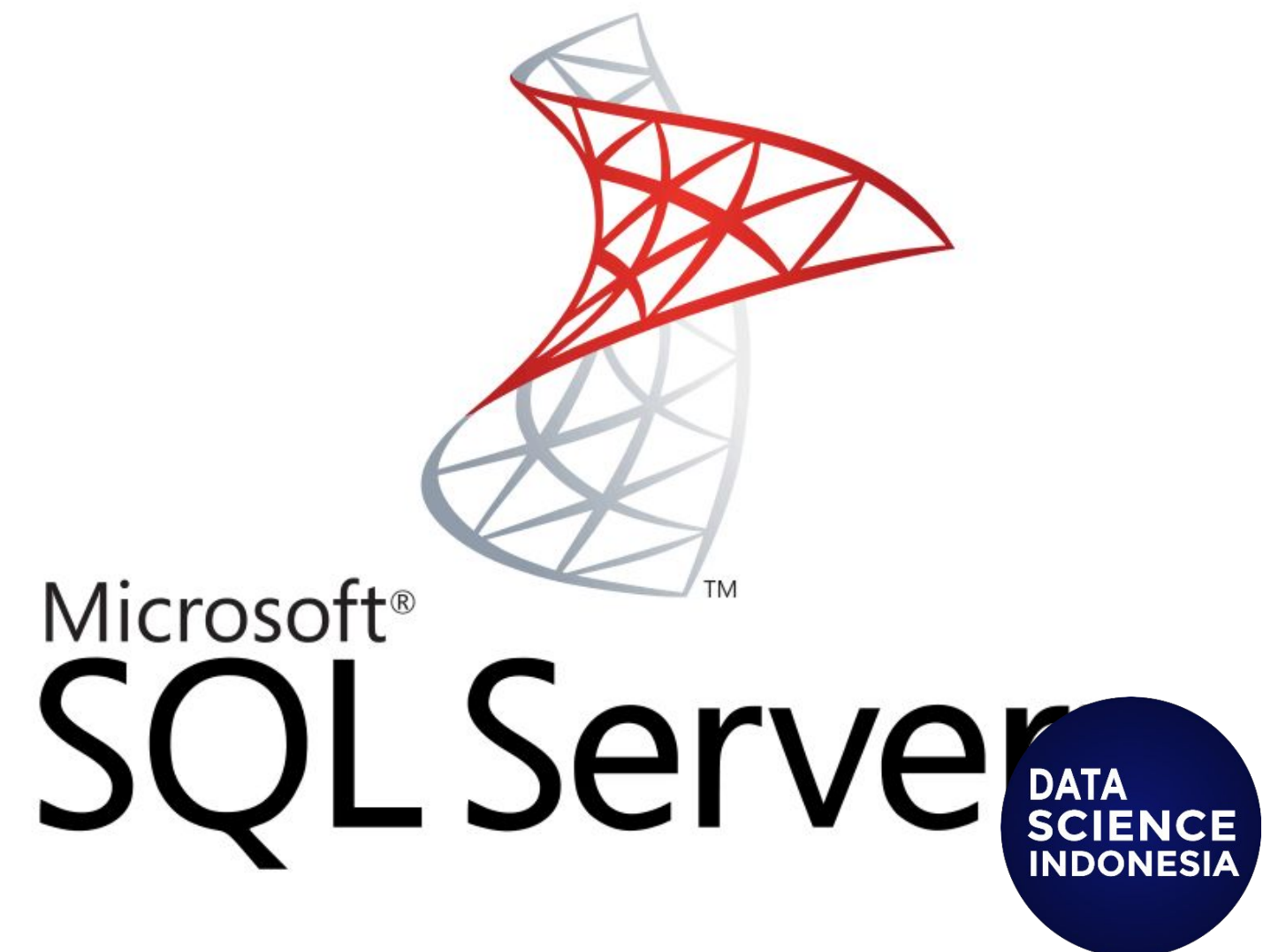
## Standar Query Language (SQL)

Result # 1 Result # 2

DEPT_NO	DEPARTMENT	HEAD_DEPT	MNGR_NO	BUDGET
000	Corporate Headquarters		105	0.00\$
100	Sales and Marketing	000	85	0.00\$
600	Engineering	000	2	0.00\$
900	Finance	000	46	0.00\$
180	Marketing	100		
620	Software Products Div.	600		
621	Software Development	620		




Jenis Database  
SQL






Query history
Saved queries
Job history
Transfers
Scheduled queries
Reservations
BI Engine

Resources  + ADD DATA 

- world\_bank\_health\_population
- world\_bank\_intl\_debt
- world\_bank\_intl\_education
- world\_bank\_wdi
  - country\_series\_definitions
  - country\_summary
  - footnotes
  - indicators\_data
  - series\_summary
  - series\_time
- worldpop
  - population\_grid\_1km

Query editor + COMPOSE NEW QUERY  HIDE EDITOR  FULL SCREEN

```
1 SELECT * FROM `bigquery-public-data.worldpop.population_grid_1km` WHERE country_name = 'Spain' and population > 5000
```

Run
Save query
Save view
Schedule query
More
This query will process 858.6 GB when run.

Query results [SAVE RESULTS](#) [EXPLORE DATA](#) ▼

Query complete (30.5 sec elapsed, 425.1 GB processed)

[Job information](#)   [Results](#)   [JSON](#)   [Execution details](#)

Row	country_name	geo_id	population	longitude_centroid	latitude_centroid	alpha_3_code	geog
1	Spain	wpsp3yz2kzkk	5180.38623046875	2.786250003841822	41.970416684653465	ESP	POLYGON((2.7820833372 41.96625003351, 2.7904166705 41.96625003351, 2.8070833370 41.96625003351, 2.8154166703 41.96625003351, 2.8237500037 41.96625003351, 2.8320833370 41.96625003351, 2.8399166703 41.96625003351, 2.8477500037 41.96625003351, 2.8554166703 41.96625003351, 2.8631666703 41.96625003351, 2.8708333370 41.96625003351, 2.8785833370 41.96625003351, 2.8862500037 41.96625003351, 2.8939166703 41.96625003351, 2.9016666703 41.96625003351, 2.9093833370 41.96625003351, 2.9170833370 41.96625003351, 2.9247916669 41.96625003351, 2.9325000037 41.96625003351, 2.9402083370 41.96625003351, 2.9479166669 41.96625003351, 2.9556250037 41.96625003351, 2.9633333370 41.96625003351, 2.9710416669 41.96625003351, 2.9787500037 41.96625003351, 2.9864583370 41.96625003351, 2.9941666669 41.96625003351, 3.0018750037 41.96625003351, 3.0095833370 41.96625003351, 3.0172916669 41.96625003351, 3.0250000037 41.96625003351, 3.0327083370 41.96625003351, 3.0404166669 41.96625003351, 3.0481250037 41.96625003351, 3.0558333370 41.96625003351, 3.0635416669 41.96625003351, 3.0712500037 41.96625003351, 3.0789583370 41.96625003351, 3.0866666669 41.96625003351, 3.0943750037 41.96625003351, 3.1020833370 41.96625003351, 3.1097916669 41.96625003351, 3.1175000037 41.96625003351, 3.1252083370 41.96625003351, 3.1329166669 41.96625003351, 3.1406250037 41.96625003351, 3.1483333370 41.96625003351, 3.1560416669 41.96625003351, 3.1637500037 41.96625003351, 3.1714583370 41.96625003351, 3.1791666669 41.96625003351, 3.1868750037 41.96625003351, 3.1945833370 41.96625003351, 3.2022916669 41.96625003351, 3.2100000037 41.96625003351, 3.2177083370 41.96625003351, 3.2254166669 41.96625003351, 3.2331250037 41.96625003351, 3.2408333370 41.96625003351, 3.2485416669 41.96625003351, 3.2562500037 41.96625003351, 3.2639583370 41.96625003351, 3.2716666669 41.96625003351, 3.2793750037 41.96625003351, 3.2870833370 41.96625003351, 3.2947916669 41.96625003351, 3.3025000037 41.96625003351, 3.3102083370 41.96625003351, 3.3179166669 41.96625003351, 3.3256250037 41.96625003351, 3.3333333370 41.96625003351, 3.3410416669 41.96625003351, 3.3487500037 41.96625003351, 3.3564583370 41.96625003351, 3.3641666669 41.96625003351, 3.3718750037 41.96625003351, 3.3795833370 41.96625003351, 3.3872916669 41.96625003351, 3.3950000037 41.96625003351, 3.4027083370 41.96625003351, 3.4104166669 41.96625003351, 3.4181250037 41.96625003351, 3.4258333370 41.96625003351, 3.4335416669 41.96625003351, 3.4412500037 41.96625003351, 3.4489583370 41.96625003351, 3.4566666669 41.96625003351, 3.4643750037 41.96625003351, 3.4720833370 41.96625003351, 3.4797916669 41.96625003351, 3.4875000037 41.96625003351, 3.4952083370 41.96625003351, 3.5029166669 41.96625003351, 3.5106250037 41.96625003351, 3.5183333370 41.96625003351, 3.5260416669 41.96625003351, 3.5337500037 41.96625003351, 3.5414583370 41.96625003351, 3.5491666669 41.96625003351, 3.5568750037 41.96625003351, 3.5645833370 41.96625003351, 3.5722916669 41.96625003351, 3.5800000037 41.96625003351, 3.5877083370 41.96625003351, 3.5954166669 41.96625003351, 3.6031250037 41.96625003351, 3.6108333370 41.96625003351, 3.6185416669 41.96625003351, 3.6262500037 41.96625003351, 3.6339583370 41.96625003351, 3.6416666669 41.96625003351, 3.6493750037 41.96625003351, 3.6570833370 41.96625003351, 3.6647916669 41.96625003351, 3.6725000037 41.96625003351, 3.6802083370 41.96625003351, 3.6879166669 41.96625003351, 3.6956250037 41.96625003351, 3.7033333370 41.96625003351, 3.7110416669 41.96625003351, 3.7187500037 41.96625003351, 3.7264583370 41.96625003351, 3.7341666669 41.96625003351, 3.7418750037 41.96625003351, 3.7495833370 41.96625003351, 3.7572916669 41.96625003351, 3.7650000037 41.96625003351, 3.7727083370 41.96625003351, 3.7804166669 41.96625003351, 3.7881250037 41.96625003351, 3.7958333370 41.96625003351, 3.8035416669 41.96625003351, 3.8112500037 41.96625003351, 3.8189583370 41.96625003351, 3.8266666669 41.96625003351, 3.8343750037



Redshift query editor v2

+ Create

Load data

Filter resources

Serverless

mybi-cluster

redshift-cluster-20

redshift-cluster-anz

redshift-cluster-ds2

redshift-debu-cluster-1

- dev
- sample\_data\_dev
- tpcds\_100g

redshift-debu-rbac

redshiftpersonademo-redsh...

redshift-producer

Cluster redshift-debu-cluster-1 (awsuser) Database dev

load-data-event-dfa1

load-data-users-1dfe

tickitsales

load-data-customer\_activity-b52d

load-data-customer\_activity-601d

tpch 1-10 query

tpch 11-22 query

Run

Limit 100

Explain

Save

Shortcuts

1 COPY dev.demo.sales FROM 's3://debuaws/tickitdb/category\_pipe.txt' IAM\_ROLE 'arn:aws:iam::265361228846:role/DebuRedShiftRole' FORMAT AS CSV DELIMITER '|' QUOTE '' REGION AS 'us-east-1'

Result 1

Chart

Summary

ERROR:Invalid digit, Value 'S', Pos 0, Type: Integer

Elapsed time: 0 seconds

Result set query:

COPY dev.demo.sales FROM 's3://debuaws/tickitdb/category\_pipe.txt' IAM\_ROLE 'arn:aws:iam::265361228846:role/DebuRedShiftRole' FORMAT AS CSV DELIMITER '|' QUOTE '' REGION AS 'us-east-1'

Amazon Redshift



»

Publish all

✓ Validate all

↻ Refresh

🗑 Discard all

🏠

🗄

💡

🔊

🔗

🛠

Data

Filter resources by name

▶ Storage accounts

◀ Databases

▶ ContosoDW

▶ Tables

▶ External tables

▶ staging.factGreenCab

▶ staging.factYellowCab

▶ staging.factholiday

▶ staging.factweather

▶ staging.factFHV

▶ dbo.dimHoliday

▶ dbo.GreenCab

▶ dbo.Fhv

▶ dbo.weather

▶ dbo.YellowCab

▶ dbo.dimNYCLocation

▶ dbo.PredictedValues

▶ Views

DWSQLQuery1

▶ Run

⬆ Publish

🔗 Query plan

Connect To ContosoDW

↻ Use database ContosoDW

1 SELECT [Datepickup]

2 ,fhvrides

3 , Yellowrides

4 , Greenrides

5 FROM [dbo].[vwMarketShareByDay]

6 order by datepickup asc

Results

Messages

View

Table

Chart

Export results




Chart type

Line

X axis column

Datepickup

Y axis columns

fhvrides, Yellowrides, Greenrides

Legend position:

center - bottom

Y axis label

Y axis minimum label

00:00:08 Query executed successfully.

Microsoft Azure Synapse Analytics

DATA SCIENCE INDONESIA



```

1 // MongoDB Playground
2 // To disable this template go to Settings | MongoDB | Use Default Template For Playground.
3
4 // Select the database to use.
5 use('test');
6
7 // Insert a few documents into the sales collection.
8 db.sales.insertMany([
9   { '_id' : 1, 'item' : 'abc', 'price' : 10, 'quantity' : 2, 'date' : new Date('2014-03-01T08:00:00Z') },
10  { '_id' : 2, 'item' : 'jkl', 'price' : 20, 'quantity' : 1, 'date' : new Date('2014-03-01T09:00:00Z') },
11  { '_id' : 3, 'item' : 'xyz', 'price' : 5, 'quantity' : 10, 'date' : new Date('2014-03-15T09:00:00Z') },
12  { '_id' : 4, 'item' : 'xyz', 'price' : 5, 'quantity' : 20, 'date' : new Date('2014-04-04T11:21:39.736Z') },
13  { '_id' : 5, 'item' : 'abc', 'price' : 10, 'quantity' : 10, 'date' : new Date('2014-04-04T21:23:13.331Z') },
14  { '_id' : 6, 'item' : 'def', 'price' : 7.5, 'quantity' : 5, 'date' : new Date('2015-06-04T05:08:13Z') },
15  { '_id' : 7, 'item' : 'def', 'price' : 7.5, 'quantity' : 10, 'date' : new Date('2015-09-10T08:43:00Z') },
16  { '_id' : 8, 'item' : 'abc', 'price' : 10, 'quantity' : 5, 'date' : new Date('2016-02-06T20:20:13Z') },
17 ]);
18
19 // Run an aggregation to view total sales for each product in 2014.
20 const aggregation = [
21   { $match: { d: { $gte: new Date('2014-01-01'), $lt: new Date('2015-01-01') } } },
22   { $group: { _id: 'date', '$quantity' : { $sum: '$quantity' } } } ],
23 ];
24 db.sales.aggregate(aggregation);
25

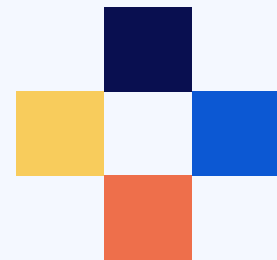
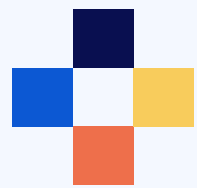
```

# NoSQL - MongoDB



```
1 SELECT
2     id, stockmarket, stockname, stockholder, stocktime, investoriban
3 FROM sqlshack_regex
4 WHERE investoriban ~ '^.*$'
```

	id [PK] integer	stockmarket text	stockname text	stockholder text	stocktime timestamp without time zone	investoriban text
1	667	NASDAQ	Popular, Inc.	Rochette Guilfoyle	2018-05-27 16:58:37	PL19 0214 4300 8029 3204 5272 3724
2	908	NYSE	Assured Guaran...	Sammy Luety	2020-11-11 03:10:30	PK66 LIKZ DIYK NYGZ FUGM M4JJ
3	1239	NYSE	GasLog Partner...	Latisha Fumagall	2017-12-07 08:47:56	FR17 9602 2435 91IZ JZ1T E7BY F69
4	4848	NYSE	State Street Cor...	Rock Godbert	2019-08-05 13:17:30	NL42 VRGQ 7633 4289 04
5	1872	NYSE	CBS Corporation	Demetri Tynemo...	2018-02-28 21:41:46	HU15 1293 4607 8941 9706 4939 5676
6	2492	NYSE	Care.com, Inc.	Janeen Semonin	2017-11-04 07:32:33	KZ03 527J RFOL ASMP 4VLD
7	2681	NYSE	Aegon NV	Ingaborg Pinchin	2018-05-27 19:28:21	FR65 7904 0244 56CF KM7D RSLA S32
8	2917	NYSE	Innovative Indu...	Merv Brass	2020-03-24 10:06:25	LI87 4802 4WE0 MQMK NKWT C
9	3534	NYSE	American Hom...	Fidelity Novic	2019-10-19 06:36:20	CZ34 6278 9268 8166 5597 5039
10	3534	NYSE	EPUBLI...	Gayelord Allin	2019-03-01 16:40:43	BA28 2853 1449 4039 5790
11	3534	NYSE	Bank	Bartram Marple	2020-02-18 23:26:21	KZ05 596R 810K QGZ9 CKOR
12	4942	NYSE	Brookfield DTL...	Portia MacAndie	2019-06-07 08:39:31	LV11 YREW P54N TSDQ 4DB1 Q



# Ringkasan Excel vs SQL


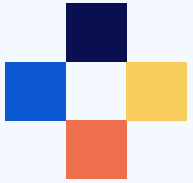
## SQL

- SQL adalah **bahasa**
- Data disimpan di dalam **database**
- Untuk real project dengan data besar, SQL **berbayar**
- Ukuran data yang bisa disimpan bisa *diupgrade* sesuai kebutuhan

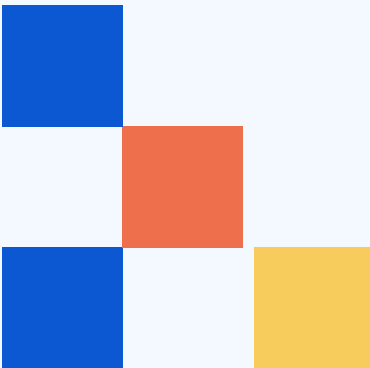
## Excel

- Excel adalah **program**
- Data disimpan di **lokal** atau **cloud** (jika menggunakan Google Sheet)
- Untuk real project dengan data besar, Excel dapat **gratis** (Google Spreadsheet)
- Batas maksimum baris data sampai 1,2 juta





# Kenapa memakai Excel vs SQL ?



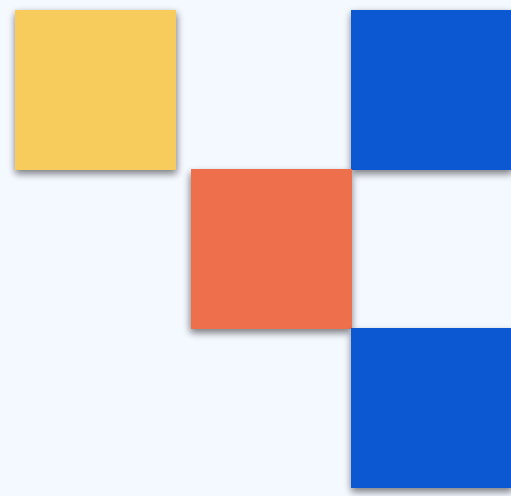
## SQL

- Jumlah data yang ada besar dan bergerak cepat
- SQL dibuat khusus untuk Data
- Lebih memungkinkan melaksanakan **Collaborative Work**
- Tingkat keamanan yang lebih baik

## Excel

- Jumlah data yang sedikit dan relatif lambat
- Hampir seluruh orang familiar dengan Excel
- Multifungsi (Penyimpanan Data, Analisis Data, dan melakukan hal lain diluar data)
- Mudah dipelajari

# Yuk Belajar Excel



## Formula

Operasi Excel yang menggunakan rumus bawaan excel

## Equation

Operasi Excel yang menggunakan cell-cell untuk menghitung suatu nilai

## Vlookup

**Vertical Lookup**, mencari suatu nilai pada tabel berdasarkan acuan nilai tertentu

## Pivot Table

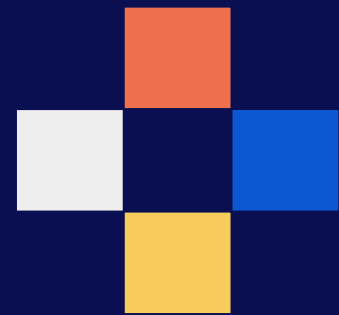
Resume atau ringkasan singkat dari data yang kita miliki

## Chart

Membuat visualisasi data yang telah kita miliki ke dalam grafik

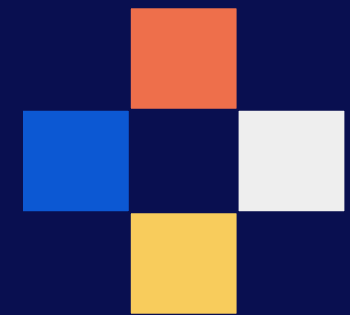


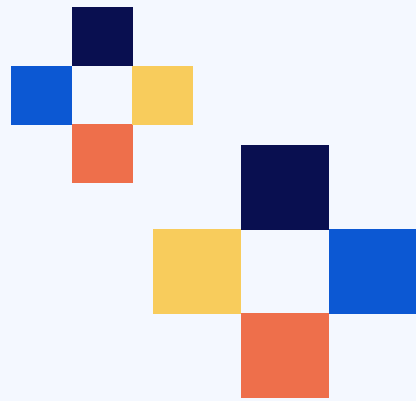
# Excel Hands On !



---

**Data Analytics using Excel**





# Yuk Belajar SQL

## Operasi SQL

- Filtering
- Ordering
- Alias
- Aggregate
- Groupby
- Subquery
- Join
- Conditional

## Database SQL

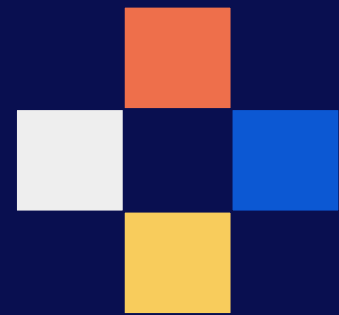
- Create
- Read
- Update
- Delete



# SQL Hands On !

---

Data Analytics using SQL





# THANK YOU



BSN

For more information,  
let's talk!

Contact person:

Bagoes  
[bagoessuryo19@gmail.com](mailto:bagoessuryo19@gmail.com)

[linkedin.com/in/bagoes-suryo-nugroho/](https://linkedin.com/in/bagoes-suryo-nugroho/)

