

**TUGAS RANCANG
BUSINESS INTELLIGENCE
COVID-19 WORLD VACCINATION PROGRESS MENGGUNAKAN
REGRESSION LINEAR**



Dosen : Ramos Somya, S.Kom., M.Cs.

Diusulkan oleh :

Andrew C. Handoko (672019250)

Arya Damar Pratama (672019227)

Deffa Ferdian Alif Utama (672019163)

Elifas Gavra Harnanda (672019223)

Zefanya Ardika (672019225)

**UNIVERSITAS KRISTEN SATYA WACANA
FAKULTAS TEKNOLOGI INFORMASI
TEKNIK INFORMATIKA**

2022

DAFTAR ISI

DAFTAR ISI	1
ABSTRAK	2
BAB I PENDAHULUAN	3
1.1 Latar belakang	3
1.2 Tujuan	3
1.3 Manfaat	3
BAB II PERANCANGAN.....	4
2.1 Pemodelan Data Warehouse	4
2.2 Aplikasi	5
BAB III PEMBAHASAN	6
3.1 Proses ETL	6
3.1.1 Extract	6
3.1.2 Transform	7
3.1.3 Loading	9
3.2 OLAP	15
3.2.1 Slicing & Dicing	15
3.2.2 Roll up & Drill down	17
3.2.3 Pivot	19
3.2.4 Filtering	20
3.3 Hasil Aplikasi	21
3.3.1 Perhitungan Aplikasi	25
3.3.2 Perhitungan Manual	26
3.3.3 Rules	27
BAB IV PENUTUP	28
4.1 Simpulan	28

ABSTRAK

Seiring berkembangnya virus Covid-19 yang terjadi di Wuhan dan berkembang di seluruh dunia, Pencegahan penyebaran virus Covid-19 salah satunya dilakukan dengan vaksinasi yang dilakukan hampir di seluruh negara. vaksin mempunyai macam-macam jenis yang berbeda oleh karena itu negara bebas dalam menentukan jenis vaksin yang ingin digunakan. Oleh sebab itu melalui proses vaksinasi yang dilakukan setiap negara membuat bertambah banyaknya data yang belum dikelompokkan yang sebenarnya data tersebut dapat dianalisis dan dijadikan sebuah informasi dalam menentukan sebuah kebijakan oleh karena itu pembuatan tugas ini bertujuan untuk menghasilkan sebuah sistem data *warehouse* yang dapat digunakan untuk pengambilan dan analisa data yang cepat agar dapat mengakses dan menanyakan data yang relevan untuk menginformasikan mengenai rekam jejak vaksinasi setiap negara. Dan bertujuan untuk mencari hubungan antar rasio vaksinasi dan angka kematian menggunakan regresi liner.

Kata Kunci : SARS-CoV-2, Analisis Data, Data Warehousing, Vaksinasi, Regresi linear

ABSTRACT

Along with the development of the Covid-19 virus that occurred in Wuhan and developed throughout the world, one of the ways to prevent the spread of the Covid-19 virus is through vaccinations that are carried out in almost all countries. There are various types of vaccines, therefore countries are free to determine the type of vaccine they want to use. Therefore, through the vaccination process carried out by each country, there is an increase in the number of data that has not been grouped, which actually can be analyzed and used as information in determining a policy. Therefore, making this task aims to produce a data warehouse system that can be used for retrieval and rapid data analysis in order to access and query relevant data to inform each country's vaccination track record. And aims to find the relationship between vaccination ratio and mortality rate using linear regression

Keywords : SARS-CoV-2, Data Analysis, Data Warehousing, Vaccination, Linear regression

BAB I

PENDAHULUAN

1.1 Latar Belakang

Virus Covid-19 yang saat ini sedang mewabah dengan perkembangan yang sangat cepat di seluruh dunia, menyebabkan terjadinya pandemi dimana-mana. Banyak cara pencegahan telah dilakukan, dimana salah satunya adalah dengan vaksinasi. Vaksin merupakan suatu cara khusus yang digunakan untuk meningkatkan kekebalan tubuh seseorang terhadap suatu penyakit, dengan cara memasukkan cairan vaksin khusus ke dalam tubuh. Vaksinasi Covid-19 sendiri bertujuan untuk membangun sebuah kekebalan tubuh (*herd immunity*) agar masyarakat dapat menjalani kesehariannya dengan normal dan produktif kembali.

Vaksin Covid-19 sendiri mempunyai beberapa jenis, dimana beberapa jenis vaksin tersebut seperti Sinovac, AstraZeneca, Sinopharm, Moderna dan lain sebagainya. Dengan banyaknya jenis vaksin dan tiap negara tentu menggunakan jenis vaksin dan jumlah yang berbeda. Serta jumlah masyarakat yang harus mendapatkan vaksin di tiap negara juga berbeda. Maka dari itu suatu tempat untuk menampung tiap data agar bisa dianalisis diperlukan. Sehingga data-data tersebut harus dibuatkan sebuah *warehouse* untuk dapat melihat informasi secara mendalam. Hal ini diperlukan agar pengelola maupun instansi yang bersangkutan dapat mengelola, menganalisis dan mendapat gambaran kondisi dari informasi mengenai vaksinasi.

Dalam pengerjaan ini, kelompok kami menggunakan data yang diambil dari website *Kaggle*. Dengan dataset yang berjudul “*Covid-19 World Vaccination Progress*”. Data yang kami peroleh ini masih perlu dilakukan normalisasi menjadi beberapa tabel yang kemudian dapat diolah.

1.2 Tujuan

Tujuan dari pembuatan tugas ini adalah untuk membangun data *warehouse* penyebaran vaksinasi di dunia. Yang bertujuan untuk menyampaikan informasi vaksinasi setiap negara berupa nama negara, kode negara, tanggal, total vaksinasi, jumlah orang yang telah di vaksin, vaksin yang digunakan dan vaksin yang telah dilakukan perhari yang sebelumnya terlebih dahulu diolah menggunakan cara *slicing*, *dicing*, *roll up* dan *drill down* yang merupakan cara pengolahan dari teknik OLAP.

1.3 Manfaat

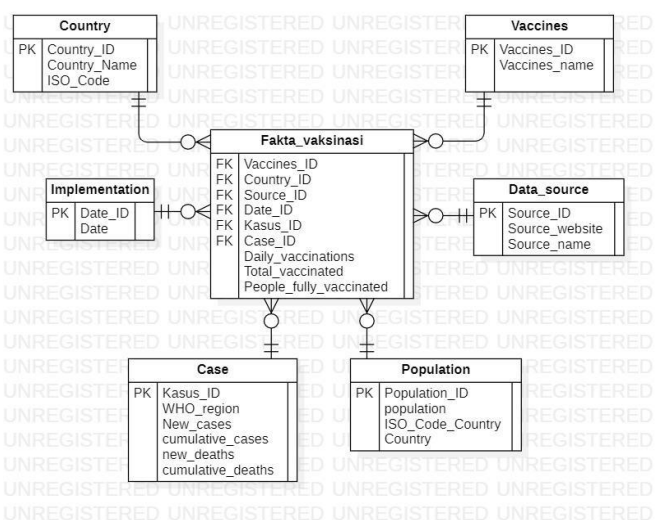
Dalam perancangan ini diharapkan dapat memetakan persebaran vaksinasi yang sudah terjadi di seluruh belahan dunia ini, terlebih pada setiap negara yang sudah melakukan vaksinasi, yang dimana dapat diketahui sudah seberapa masif dan meratanya penyebaran vaksinasi pada setiap bagian negara.

BAB II

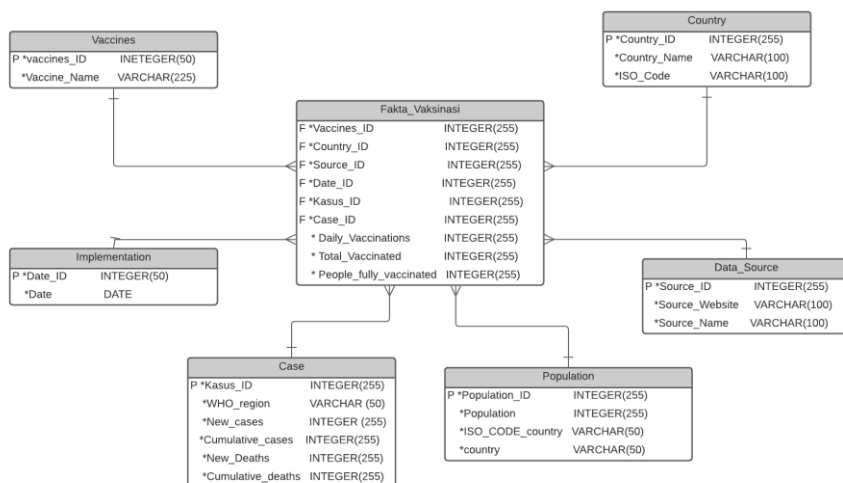
PERANCANGAN

2.1 Pemodelan Data Warehouse

Untuk pemodelan Data Warehouse menggunakan pendekatan dengan skema *Star*, dimana Tabel fakta akan menyimpan data-data utama sementara tabel dimensi mendeskripsikan setiap nilai dari suatu dimensi dan dapat direlasikan ke tabel fakta jika diperlukan dan terdapat satu tabel fakta (fact table) di pusat bintang dengan beberapa tabel dimensi (dimensional tables) yang mengelilinginya diantaranya *Vaccines*, *Country*, *Implementation*, *Data_Source*, *Case*, *Population*.



Gambar 2.1.1 Entity Relationship Diagram (ERD)



Gambar 2.1.2 Entity Relationship Model (ERM)

2.2 Aplikasi

Untuk aplikasinya disini kami menggunakan Google Collaboratory untuk mengolah data yang ada menggunakan algoritma linear regression dengan tujuan mendapatkan data yang valid antara hubungan Vaccination rate dengan kasus angka kematian (new deaths) dengan akurat dan mengurangi human error.

Berikut tutorial pengolahan data menggunakan Google Collaboratory :



Gambar 2.2.1 Tutorial pengolahan data menggunakan Google Colaboratory

Pertama kali kita akan menganalisa sebuah permasalahan yaitu Wabah COVID-19 telah membuat seluruh negara bertekuk lutut. Lebih dari 4,5 juta orang telah meninggal berdasarkan dataset yang ada, dan satu-satunya jalan keluar yang dapat dilakukan dari bencana ini adalah dengan memvaksinasi semua bagian masyarakat. Terlepas dari kenyataan bahwa manfaat vaksinasi telah terbukti di seluruh negara, meskipun ada kelompok anti-vaksin bermunculan di berbagai negara. Setelah itu kumpulkan data-data yang dibutuhkan diantaranya data vaksinasi, data populasi manusia di seluruh negara dan data angka kematian yang disebabkan oleh COVID-19, setelah datanya telah dikumpulkan maka setelah itu data-data tersebut digabungkan berdasarkan tiap-tiap negaranya, setelah itu dilakukan proses modelling menggunakan algoritma regresi linear yaitu dengan mengambil 2 variabel X yaitu Vaccination rate dan Y yaitu new deaths dan tahap terakhir yaitu melakukan evaluasi yaitu untuk mengukur seberapa besar akurasi dari model yang dibuat dengan mengambil salah satu contoh negara.

BAB III

PEMBAHASAN

3.1 Proses ETL

ETL merupakan singkatan dari *Extract Transform Loading* yaitu dimulai dari *extract data* yang merupakan pengambilan data mentah dari sumber data yang telah diambil lalu *transform* data adalah merubah data sesuai dengan kebutuhan seperti merubah data ke numerik dengan proses perhitungan lalu langkah terakhir *loading* data guna menampung sistem data ke data *warehouse*. Ke-3 hal tersebut merupakan pondasi awal data *analytics* dan *machine learning* yang bertujuan untuk menganalisis dan mengorganisir data.

3.1.1 Extract

Awal langkah proses ETL yaitu ekstraksi . Data harus diekstrak terlebih dahulu dari sumbernya sebelum dipindahkan ke tempat yang lain. Pada langkah pertama proses ETL ini, data terstruktur dan tidak terstruktur diimpor dan dikonsolidasikan ke dalam satu wadah penyimpanan.

	A	B	C	D	E	F	G	H	I	J	K
1	country	iso_code	Date_ID	date	total_vaccinated	people_fully_vaccinated	daily_vaccinations	vaccines_id	vaccines-name	source_name	source_website
2	Afghanistan	AFG	1	2/22/2021				1	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
3	Afghanistan	AFG	2	2/23/2021			1367	2	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
4	Afghanistan	AFG	3	2/24/2021			1367	3	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
5	Afghanistan	AFG	4	2/25/2021			1367	4	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
6	Afghanistan	AFG	5	2/26/2021			1367	5	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
7	Afghanistan	AFG	6	2/27/2021			1367	6	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
8	Afghanistan	AFG	7	2/28/2021	8200		1367	7	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
9	Afghanistan	AFG	8	3/1/2021			1580	8	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
10	Afghanistan	AFG	9	3/2/2021			1734	9	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
11	Afghanistan	AFG	10	3/3/2021			2008	10	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
12	Afghanistan	AFG	11	3/4/2021			2221	11	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
13	Afghanistan	AFG	12	3/5/2021			2435	12	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
14	Afghanistan	AFG	13	3/6/2021			2649	13	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
15	Afghanistan	AFG	14	3/7/2021			2862	14	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
16	Afghanistan	AFG	15	3/8/2021			2862	15	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
17	Afghanistan	AFG	16	3/9/2021			2862	16	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
18	Afghanistan	AFG	17	3/10/2021			2862	17	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
19	Afghanistan	AFG	18	3/11/2021			2862	18	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
20	Afghanistan	AFG	19	3/12/2021			2862	19	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
21	Afghanistan	AFG	20	3/13/2021			2862	20	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
22	Afghanistan	AFG	21	3/14/2021			2862	21	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
23	Afghanistan	AFG	22	3/15/2021			2862	22	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
24	Afghanistan	AFG	23	3/16/2021	54000		2862	23	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
25	Afghanistan	AFG	24	3/17/2021			2862	24	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
26	Afghanistan	AFG	25	3/18/2021			2902	25	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
27	Afghanistan	AFG	26	3/19/2021			2921	26	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
28	Afghanistan	AFG	27	3/20/2021			2941	27	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
29	Afghanistan	AFG	28	3/21/2021			2961	28	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
30	Afghanistan	AFG	29	3/22/2021			2980	29	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
31	Afghanistan	AFG	30	3/23/2021			3000	30	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
32	Afghanistan	AFG	31	3/24/2021			3000	31	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
33	Afghanistan	AFG	32	3/25/2021			3000	32	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
34	Afghanistan	AFG	33	3/26/2021			3000	33	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
35	Afghanistan	AFG	34	3/27/2021			3000	34	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
36	Afghanistan	AFG	35	3/28/2021			3000	35	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/
37	Afghanistan	AFG	36	3/29/2021			3000	36	Johnson&Johnson, OxfordAstraZeneca, PfizerBioNTech, SinopharmBeijing	World Health Organization	https://covid19.who.int/

Gambar 3.1.1.1 Data Extracting data vaksinasi

	A	B	C	D	E	F	G	H	I
1	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths	
2	1/3/2020	AF	Afghanistan	EMRO	0	0	0	0	
3	1/4/2020	AF	Afghanistan	EMRO	0	0	0	0	
4	1/5/2020	AF	Afghanistan	EMRO	0	0	0	0	
5	1/6/2020	AF	Afghanistan	EMRO	0	0	0	0	
6	1/7/2020	AF	Afghanistan	EMRO	0	0	0	0	
7	1/8/2020	AF	Afghanistan	EMRO	0	0	0	0	
8	1/9/2020	AF	Afghanistan	EMRO	0	0	0	0	
9	1/10/2020	AF	Afghanistan	EMRO	0	0	0	0	
10	1/11/2020	AF	Afghanistan	EMRO	0	0	0	0	
11	1/12/2020	AF	Afghanistan	EMRO	0	0	0	0	
12	1/13/2020	AF	Afghanistan	EMRO	0	0	0	0	
13	1/14/2020	AF	Afghanistan	EMRO	0	0	0	0	
14	1/15/2020	AF	Afghanistan	EMRO	0	0	0	0	
15	1/16/2020	AF	Afghanistan	EMRO	0	0	0	0	
16	1/17/2020	AF	Afghanistan	EMRO	0	0	0	0	
17	1/18/2020	AF	Afghanistan	EMRO	0	0	0	0	
18	1/19/2020	AF	Afghanistan	EMRO	0	0	0	0	
19	1/20/2020	AF	Afghanistan	EMRO	0	0	0	0	
20	1/21/2020	AF	Afghanistan	EMRO	0	0	0	0	
21	1/22/2020	AF	Afghanistan	EMRO	0	0	0	0	
22	1/23/2020	AF	Afghanistan	EMRO	0	0	0	0	
23	1/24/2020	AF	Afghanistan	EMRO	0	0	0	0	
24	1/25/2020	AF	Afghanistan	EMRO	0	0	0	0	
25	1/26/2020	AF	Afghanistan	EMRO	0	0	0	0	
26	1/27/2020	AF	Afghanistan	EMRO	0	0	0	0	
27	1/28/2020	AF	Afghanistan	EMRO	0	0	0	0	
28	1/29/2020	AF	Afghanistan	EMRO	0	0	0	0	
29	1/30/2020	AF	Afghanistan	EMRO	0	0	0	0	

Gambar 3.1.1.2 Data Extracting data WHO-Covid-19

	A	B	C	D	E	F	G	H	I	J	K
1	iso_code	country	2021_last_updated	2020_population	area	density_sq_km	growth_rate	world_%	rank		
2	CHN	China	1,447,065,329	1,439,323,776	9,706,961 sq_km	149/sq_km	0.34%	18.34%	1		
3	IND	India	1,401,310,563	1,380,004,385	3,287,590 sq_km	424/sq_km	0.97%	17.69%	2		
4	USA	United States	334,058,426	331,002,651	9,372,610 sq_km	36/sq_km	0.58%	4.23%	3		
5	IDN	Indonesia	278,037,263	273,523,615	1,904,569 sq_km	145/sq_km	1.04%	3.51%	4		
6	PAK	Pakistan	227,724,796	220,892,340	881,912 sq_km	255/sq_km	1.95%	2.86%	5		
7	BRA	Brazil	214,832,901	212,559,417	8,515,767 sq_km	25/sq_km	0.67%	2.72%	6		
8	NGA	Nigeria	214,507,696	206,139,589	923,768 sq_km	229/sq_km	2.55%	2.68%	7		
9	BGD	Bangladesh	167,247,491	164,689,383	147,570 sq_km	1,127/sq_km	0.98%	2.11%	8		
10	RUS	Russia	145,899,956	145,934,462	17,098,242 sq_km	9/sq_km	-0.02%	1.85%	9		
11	MEX	Mexico	131,046,075	128,932,753	1,964,375 sq_km	66/sq_km	1.03%	1.65%	10		
12	JPN	Japan	125,802,521	126,476,461	377,930 sq_km	334/sq_km	-0.34%	1.60%	11		
13	ETH	Ethiopia	119,590,501	114,963,588	1,104,300 sq_km	107/sq_km	2.53%	1.50%	12		
14	PHL	Philippines	111,913,102	109,581,078	342,353 sq_km	324/sq_km	1.34%	1.41%	13		
15	EGY	Egypt	105,390,688	102,334,404	1,002,450 sq_km	104/sq_km	1.88%	1.32%	14		
16	VNM	Vietnam	98,655,916	97,338,579	331,212 sq_km	296/sq_km	0.85%	1.25%	15		
17	TUR	Turkey	85,484,777	84,339,067	783,562 sq_km	109/sq_km	0.83%	1.08%	17		
18	IRN	Iran	85,627,052	83,992,949	1,648,195 sq_km	52/sq_km	1.23%	1.08%	18		
19	DEU	Germany	83,975,691	83,783,942	357,114 sq_km	235/sq_km	0.14%	1.07%	19		
20	THA	Thailand	70,039,646	69,799,978	513,120 sq_km	136/sq_km	0.22%	0.89%	20		
21	GBR	United Kingdom	68,401,087	67,886,011	242,900 sq_km	281/sq_km	0.47%	0.87%	21		
22	FR	France	65,520,147	65,273,511	551,695 sq_km	119/sq_km	0.23%	0.83%	22		
23	TZA	Tanzania	62,539,416	59,734,218	945,087 sq_km	65/sq_km	2.95%	0.78%	23		
24	ITA	Italy	60,320,493	60,461,826	301,336 sq_km	200/sq_km	-0.16%	0.77%	24		
25	ZAF	South Africa	60,474,550	59,308,690	1,221,037 sq_km	49/sq_km	1.24%	0.76%	25		
26	KEN	Kenya	55,703,607	53,771,296	580,367 sq_km	95/sq_km	2.26%	0.70%	26		
27	MMR	Myanmar	55,037,700	54,409,800	676,578 sq_km	81/sq_km	0.73%	0.70%	27		

Gambar 3.1.1.3 Data Extracting data 2021_Population

3.1.2 Transform

Kemudian transform yaitu pembersihan dan mempersiapkan agregasi untuk analisis. Langkah ini sangat penting dalam proses ETL karena membantu memastikan data yang akan diolah sepenuhnya siap dan kompatibel.

	A	B	C	D	E	F	G
1	country_id	Date_ID	total_vaccinated	people_fully_vaccinated	daily_vaccinations	vaccines_id	source_id
2	1	1	0			1	1
3	2	2			1367	2	2
4	3	3			1367	3	3
5	4	4			1367	4	4
6	5	5			1367	5	5
7	6	6			1367	6	6
8	7	7	8200		1367	7	7
9	8	8			1580	8	8
10	9	9			1794	9	9
11	10	10			2008	10	10
12	11	11			2221	11	11
13	12	12			2435	12	12
14	13	13			2649	13	13
15	14	14			2862	14	14
16	15	15			2862	15	15
17	16	16			2862	16	16
18	17	17			2862	17	17
19	18	18			2862	18	18
20	19	19			2862	19	19
21	20	20			2862	20	20
22	21	21			2862	21	21
23	22	22			2862	22	22
24	23	23	54000		2862	23	23
25	24	24			2862	24	24
26	25	25			2902	25	25
27	26	26			2921	26	26
28	27	27			2941	27	27
29	28	28			2961	28	28
30	29	29			2980	29	29
31	30	30			3000	30	30
32	31	31			3000	31	31
33	32	32			3000	32	32
34	33	33			3000	33	33

Gambar 3.1.2.1 Data transform data vaksinasi menjadi tabel fakta_vaksinasi

A		B
1	ix_vaccines_id	vaccines-name
2	1	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
3	2	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
4	3	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
5	4	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
6	5	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
7	6	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
8	7	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
9	8	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
10	9	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
11	10	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
12	11	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
13	12	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
14	13	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
15	14	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
16	15	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
17	16	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
18	17	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
19	18	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
20	19	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
21	20	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
22	21	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
23	22	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
24	23	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
25	24	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
26	25	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
27	26	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
28	27	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
29	28	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
30	29	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
31	30	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
32	31	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
33	32	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing
34	33	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing

Gambar 3.1.2.2 Data *transform* data vaksinasi menjadi tabel vaccines

A		B
1	ix_date_ID	date
2	1	2021-02-22
3	2	2021-02-23
4	3	2021-02-24
5	4	2021-02-25
6	5	2021-02-26
7	6	2021-02-27
8	7	2021-02-28
9	8	2021-03-01
10	9	2021-03-02
11	10	2021-03-03
12	11	2021-03-04
13	12	2021-03-05
14	13	2021-03-06
15	14	2021-03-07
16	15	2021-03-08
17	16	2021-03-09
18	17	2021-03-10
19	18	2021-03-11
20	19	2021-03-12
21	20	2021-03-13
22	21	2021-03-14
23	22	2021-03-15
24	23	2021-03-16
25	24	2021-03-17
26	25	2021-03-18
27	26	2021-03-19
28	27	2021-03-20
29	28	2021-03-21
30	29	2021-03-22
31	30	2021-03-23
32	31	2021-03-24
33	32	2021-03-25
34	33	2021-03-26

Gambar 3.1.2.3 Data *transform* data vaksinasi menjadi tabel Implementations

A		B	C
1	ix_country_id	country	iso_code
2	1	Afghanistan	AFG
3	2	Afghanistan	AFG
4	3	Afghanistan	AFG
5	4	Afghanistan	AFG
6	5	Afghanistan	AFG
7	6	Afghanistan	AFG
8	7	Afghanistan	AFG
9	8	Afghanistan	AFG
10	9	Afghanistan	AFG
11	10	Afghanistan	AFG
12	11	Afghanistan	AFG
13	12	Afghanistan	AFG
14	13	Afghanistan	AFG
15	14	Afghanistan	AFG
16	15	Afghanistan	AFG
17	16	Afghanistan	AFG
18	17	Afghanistan	AFG
19	18	Afghanistan	AFG
20	19	Afghanistan	AFG
21	20	Afghanistan	AFG
22	21	Afghanistan	AFG
23	22	Afghanistan	AFG
24	23	Afghanistan	AFG
25	24	Afghanistan	AFG
26	25	Afghanistan	AFG
27	26	Afghanistan	AFG
28	27	Afghanistan	AFG
29	28	Afghanistan	AFG
30	29	Afghanistan	AFG
31	30	Afghanistan	AFG
32	31	Afghanistan	AFG
33	32	Afghanistan	AFG
34	33	Afghanistan	AFG

Gambar 3.1.2.4 Data *transform* data vaksinasi menjadi tabel country

	A	B	
1	ix_source_id	source_name	source_website
2	1	World Health Organization	https://covid19.who.int/
3	2	World Health Organization	https://covid19.who.int/
4	3	World Health Organization	https://covid19.who.int/
5	4	World Health Organization	https://covid19.who.int/
6	5	World Health Organization	https://covid19.who.int/
7	6	World Health Organization	https://covid19.who.int/
8	7	World Health Organization	https://covid19.who.int/
9	8	World Health Organization	https://covid19.who.int/
10	9	World Health Organization	https://covid19.who.int/
11	10	World Health Organization	https://covid19.who.int/
12	11	World Health Organization	https://covid19.who.int/
13	12	World Health Organization	https://covid19.who.int/
14	13	World Health Organization	https://covid19.who.int/
15	14	World Health Organization	https://covid19.who.int/
16	15	World Health Organization	https://covid19.who.int/
17	16	World Health Organization	https://covid19.who.int/
18	17	World Health Organization	https://covid19.who.int/
19	18	World Health Organization	https://covid19.who.int/
20	19	World Health Organization	https://covid19.who.int/
21	20	World Health Organization	https://covid19.who.int/
22	21	World Health Organization	https://covid19.who.int/
23	22	World Health Organization	https://covid19.who.int/
24	23	World Health Organization	https://covid19.who.int/
25	24	World Health Organization	https://covid19.who.int/
26	25	World Health Organization	https://covid19.who.int/
27	26	World Health Organization	https://covid19.who.int/
28	27	World Health Organization	https://covid19.who.int/
29	28	World Health Organization	https://covid19.who.int/
30	29	World Health Organization	https://covid19.who.int/
31	30	World Health Organization	https://covid19.who.int/
32	31	World Health Organization	https://covid19.who.int/
33	32	World Health Organization	https://covid19.who.int/
34	33	World Health Organization	https://covid19.who.int/

Gambar 3.1.2.5 Data *transform* data vaksinasi menjadi tabel data_source

	A	B	C	D	E	F	G
1	Kasus_ID	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths	
2	1	EMRO	0	0	0	0	
3	2	EMRO	0	0	0	0	
4	3	EMRO	0	0	0	0	
5	4	EMRO	0	0	0	0	
6	5	EMRO	0	0	0	0	
7	6	EMRO	0	0	0	0	
8	7	EMRO	0	0	0	0	
9	8	EMRO	0	0	0	0	
10	9	EMRO	0	0	0	0	
11	10	EMRO	0	0	0	0	
12	11	EMRO	0	0	0	0	
13	12	EMRO	0	0	0	0	
14	13	EMRO	0	0	0	0	
15	14	EMRO	0	0	0	0	
16	15	EMRO	0	0	0	0	
17	16	EMRO	0	0	0	0	
18	17	EMRO	0	0	0	0	
19	18	EMRO	0	0	0	0	
20	19	EMRO	0	0	0	0	
21	20	EMRO	0	0	0	0	
22	21	EMRO	0	0	0	0	
23	22	EMRO	0	0	0	0	
24	23	EMRO	0	0	0	0	
25	24	EMRO	0	0	0	0	
26	25	EMRO	0	0	0	0	
27	26	EMRO	0	0	0	0	
28	27	EMRO	0	0	0	0	
29	28	EMRO	0	0	0	0	

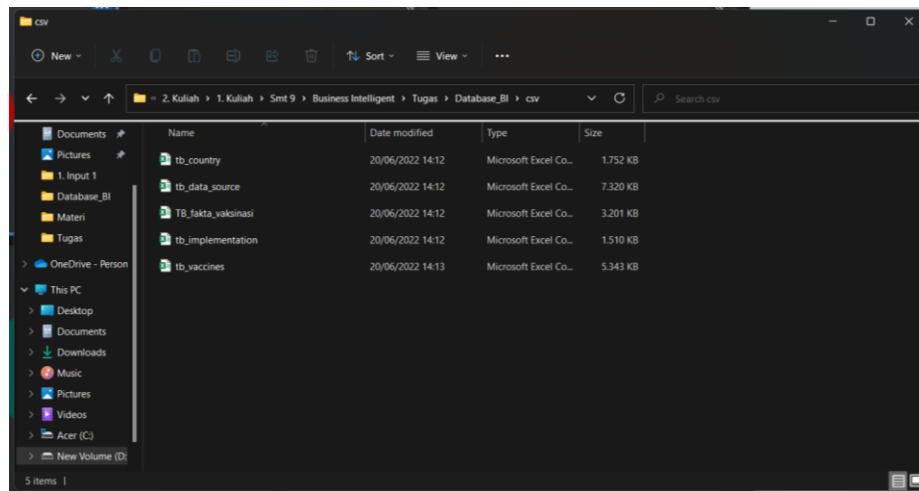
Gambar 3.1.2.6 Data *transform* data vaksinasi menjadi tabel case

	A	B	C	D	E	F	G	H
1	Population_ID	iso_code	country	2021_last_updated				
2	1	CHN	China	1,447,065,329				
3	2	IND	India	1,401,310,563				
4	3	USA	United States	334,058,426				
5	4	IDN	Indonesia	278,037,263				
6	5	PAK	Pakistan	227,724,796				
7	6	BRA	Brazil	214,832,901				
8	7	NGA	Nigeria	214,507,696				
9	8	BGD	Bangladesh	167,247,491				
10	9	RUS	Russia	145,899,956				
11	10	MEX	Mexico	131,046,075				
12	11	JPN	Japan	125,802,521				
13	12	ETH	Ethiopia	119,590,501				
14	13	PHL	Philippines	111,913,102				
15	14	EGY	Egypt	105,390,688				
16	15	VNM	Vietnam	98,655,916				
17	16	TUR	Turkey	85,484,777				
18	17	IRN	Iran	85,627,052				
19	18	DEU	Germany	83,975,691				
20	19	THA	Thailand	70,039,646				
21	20	GBR	United Kingdom	68,401,087				
22	21	FRA	France	65,520,147				
23	22	TZA	Tanzania	62,539,416				
24	23	ITA	Italy	60,320,493				
25	24	ZAF	South Africa	60,474,550				
26	25	KEN	Kenya	55,703,607				
27	26	MMR	Myanmar	55,037,700				
28	27	KOR	South Korea	51,331,264				
29	28	COL	Colombia	51,504,213				

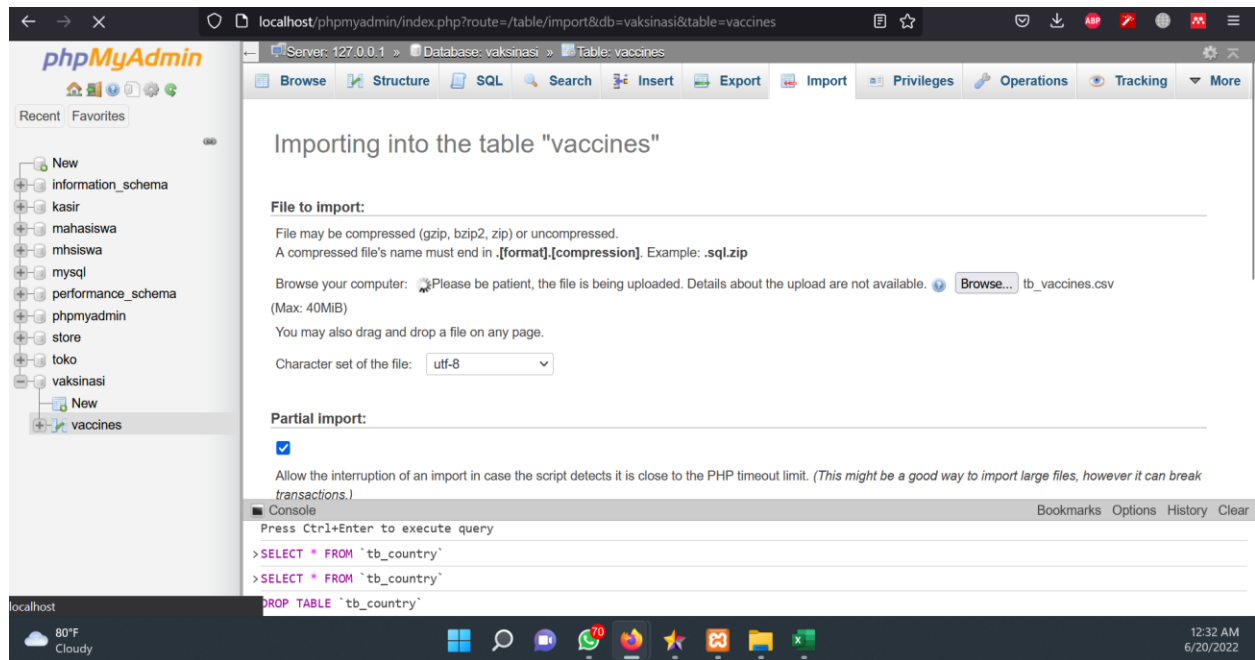
Gambar 3.1.2.7 Data *transform* data 2021_population menjadi tabel Population

3.1.3 Loading

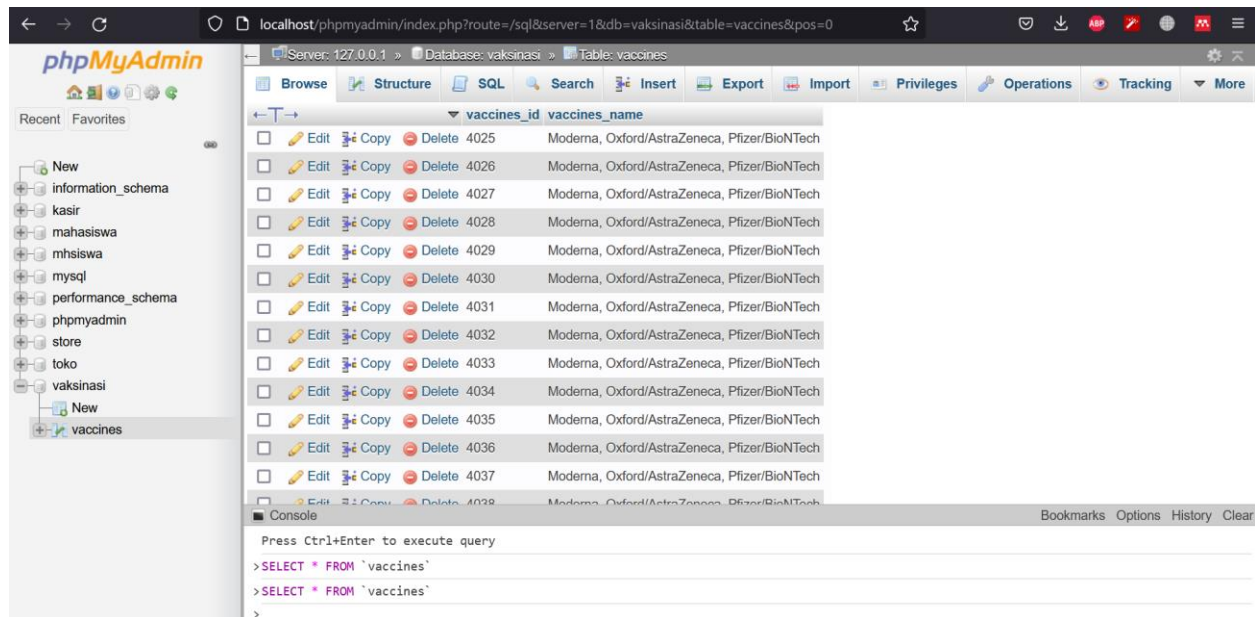
Terakhir adalah loading yaitu data yang sudah diubah kemudian dipindah dari area *staging* ke data *warehouse* target. Biasanya, langkah ini mencakup *loading* awal data keseluruhan, diikuti dengan *loading* berkala terhadap perubahan data hingga *refresh* data untuk menghapus atau mengganti data dalam *warehouse*.



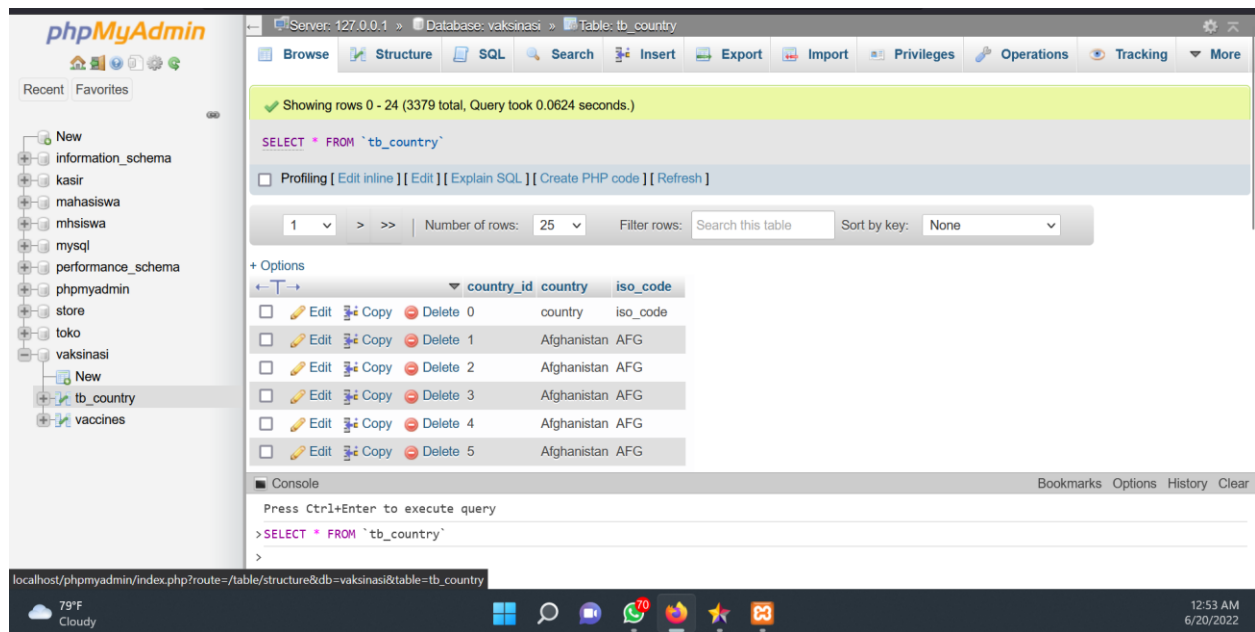
Gambar 3.1.3.1 Proses *extract* data menjadi data berbentuk .csv



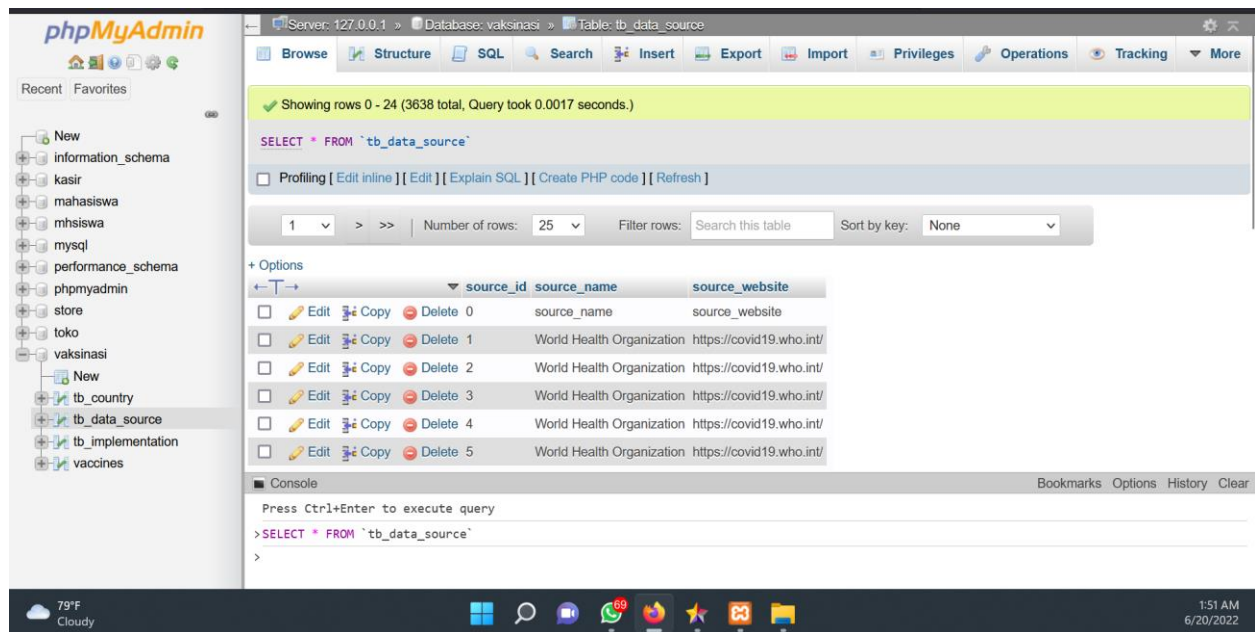
Gambar 3.1.3.2 Proses memasukkan *extract* data ke dalam data warehouse



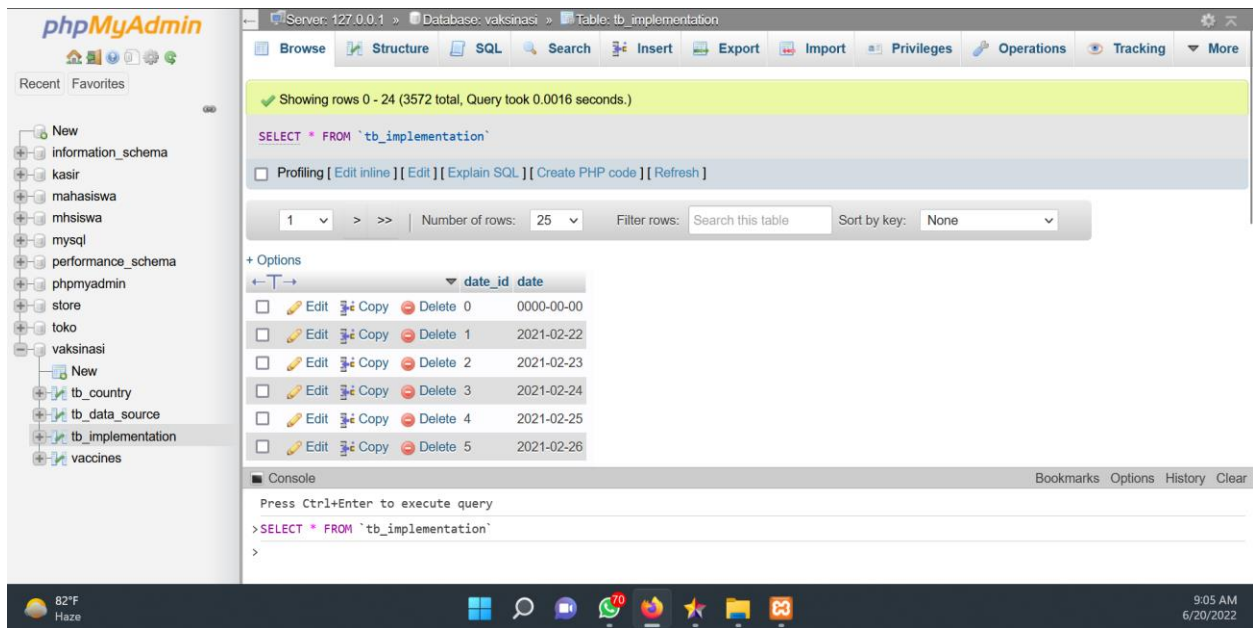
Gambar 3.1.3.3 Hasil view tabel vaccines dalam data warehouse



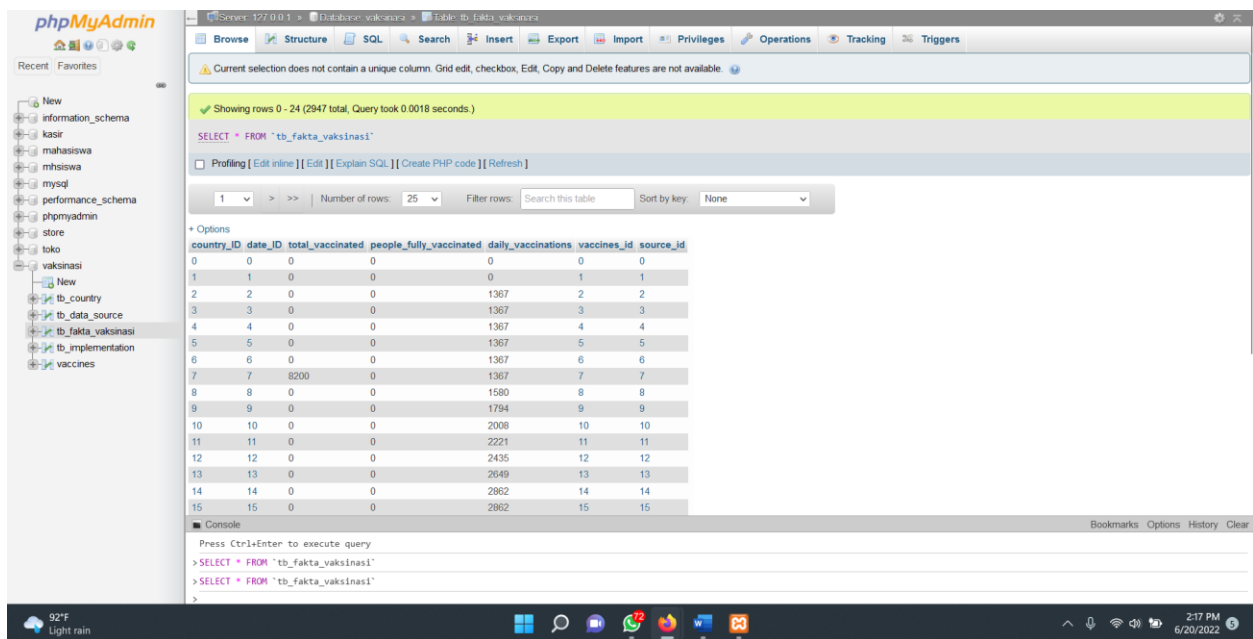
Gambar 3.1.3.4 Hasil view tabel country dalam data warehouse



Gambar 3.1.3.15 Hasil view tabel data_source dalam data warehouse



Gambar 3.1.3.6 Hasil view tabel implementations dalam data warehouse



Gambar 3.1.3.7 Hasil view tabel data_fakta_vaksinasi dalam data warehouse

Server: 127.0.0.1 » Database: vaksinasi » Table: case

[Browse](#)
[Structure](#)
[SQL](#)
[Search](#)
[Insert](#)
[Export](#)
[Import](#)
[Privileges](#)
[Operations](#)

[Table structure](#)
[Relation view](#)

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/> 1	kasus_ID	bigint(255)			No	None			Change Drop More
<input type="checkbox"/> 2	WHO_region	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
<input type="checkbox"/> 3	New_kasus	bigint(255)			No	None			Change Drop More
<input type="checkbox"/> 4	cumulative_kasus	bigint(255)			No	None			Change Drop More
<input type="checkbox"/> 5	New_deaths	bigint(255)			No	None			Change Drop More
<input type="checkbox"/> 6	cumulative_deaths	bigint(255)			No	None			Change Drop More

Gambar 3.1.3.8 Hasil *view* tabel case dalam data *warehouse*

[Browse](#)
[Structure](#)
[SQL](#)
[Search](#)
[Insert](#)
[Export](#)
[Import](#)
[Privileges](#)
[Operations](#)

[Table structure](#)
[Relation view](#)

#	Name	Type	Collation	Attributes	Null	Default	Comments	Extra	Action
<input type="checkbox"/> 1	population_ID	bigint(255)			No	None			Change Drop More
<input type="checkbox"/> 2	population	bigint(255)			No	None			Change Drop More
<input type="checkbox"/> 3	ISO_code_country	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More
<input type="checkbox"/> 4	country	varchar(50)	utf8mb4_general_ci		No	None			Change Drop More

Gambar 3.1.3.9 Hasil *view* tabel population dalam data *warehouse*

3.2 OLAP

OLAP Data Cube adalah representasi data yang dapat dilihat dari berbagai dimensi. Bentuk kubus adalah representasi yang digunakan untuk view data yang dapat dilihat dari 3 dimensi yang berbeda. OLAP Data Cube memiliki beberapa operasi yang dapat dilakukan untuk merubah view data. Beberapa operasi tersebut antara lain :

3.2.1 Slicing & Dicing

Slice adalah operasi OLAP yang membentuk sub-cube dari sebuah OLAP data cube dengan memilih satu *dimension* yang spesifik. Sedangkan *Dice* adalah operasi OLAP yang membentuk sub-cube dari OLAP data cube dengan memilih lebih dari satu *dimension* yang spesifik.

```
SELECT tb_country.country,tb_implementation.date,tb_vaccines.vaccines FROM ((tb_country JOIN tb_implementation ON tb_country.country_id = tb_implementation.date_ID)JOIN tb_vaccines ON tb_country.country_id = tb_vaccines.vaccines_ID);
```

Gambar 15. Kueri 1 Slicing

+ Options

country	date	vaccines
Afghanistan	2021-02-22	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-02-23	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-02-24	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-02-25	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-02-26	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-02-27	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-02-28	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-01	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-02	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-03	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-04	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-05	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-06	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-07	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-08	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-09	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-10	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-11	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...
Afghanistan	2021-03-12	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNte...

Gambar 16. Hasil dari Kueri 1 Slicing

Pada proses Slicing, kami menggabungkan table country dengan table implementation dan vaccines untuk melakukan filter subset data dari sebuah OLAP cube. Pada percobaan Slicing kali ini penulis mengambil nama kota pada table country untuk mengidentifikasi atau untuk mengetahui negara-negara yang sedang melakukan vaksinasi dengan melihat tanggal yang diambil dari table implementation kolom date dan melihat jenis vaksin yang digunakan dengan mengambil data dari table vaccines.

```
SELECT country, total_vaccinated FROM tb_fakta_vaksinasi WHERE total_vaccinated IS NOT NULL AND total_vaccinated != 0;
```

Gambar 17. Kueri 2 Slicing

country	total_vaccinated
Afghanistan	8200
Afghanistan	54000
Afghanistan	120000
Afghanistan	240000
Afghanistan	448878
Afghanistan	470341
Afghanistan	476367
Afghanistan	479372
Afghanistan	479574
Afghanistan	480226
Afghanistan	481690
Afghanistan	481800
Afghanistan	482952

Gambar 18. Hasil dari Kueri 2 *Slicing*

Pada proses tersebut penulis ingin mengetahui nama negara yang memiliki nilai total_vaccinated bukan yang bernilai kosong maka dari itu perintah yang digunakan adalah SELECT kolom yang akan digunakan yaitu negara dan total_vaccinated FROM tabel fakta vaksinasi WHERE total_vaksinasi tidak bernilai NULL AND total_vaksinasi tidak sama dengan 0

Kemudian *dicing*. Berikut ini merupakan query sekaligus output dari proses Dicing.

```
SELECT tb_country.country,tb_implementation.date,tb_vaccines.vaccines FROM ((tb_country JOIN tb_implementation ON tb_country.country_id = tb_implementation.date_ID)JOIN tb_vaccines ON tb_country.country_id = tb_vaccines.vaccines_ID) WHERE tb_implementation.date = '2021-02-22';
```

Gambar 19. Kueri 2 *Dicing*

country	date	vaccines
Afghanistan	2021-02-22	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTe...
Albania	2021-02-22	Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sput...
Algeria	2021-02-22	Oxford/AstraZeneca, Sinopharm/Beijing, Sinovac, Sp...
Andorra	2021-02-22	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Anguilla	2021-02-22	Oxford/AstraZeneca, Pfizer/BioNTech
Antigua and Barbuda	2021-02-22	Oxford/AstraZeneca, Pfizer/BioNTech, Sputnik V
Argentina	2021-02-22	CanSino, Moderna, Oxford/AstraZeneca, Pfizer/BioNT...

Gambar 20. Hasil dari Kueri 1 *Dicing*

Pada proses Dicing, kami menggabungkan table country dengan table implementation dan vaccines untuk melakukan filter subset data dari sebuah OLAP cube. Pada percobaan Slicing kali ini kami mengambil nama kota pada table country untuk mengidentifikasi atau untuk mengetahui negara-negara yang sedang melakukan vaksinasi dengan melihat tanggal yang diambil dari table implementation kolom date dan melihat jenis vaksin yang digunakan dengan mengambil data dari table vaccines. Setelah itu data di ambil atau di kelompokkan berdasarkan tabel implementation dengan ketentuan tanggal yang diambil adalah 2021-02-22.

3.2.2 Roll up & Drill down

Roll-up adalah operasi OLAP yang dilakukan untuk meningkatkan informasi yang didapatkan dari data ke level yang lebih abstrak. *Roll-up* dapat dilakukan dengan mengurangi jumlah *dimension* atau meningkatkan hirarki data menjadi level yang lebih abstrak. Sedangkan *Drill-down* adalah operasi OLAP yang dilakukan untuk meningkatkan hirarki data menjadi level yang lebih detail.

DRILL DOWN

Pada fungsi ini kami mengambil dari tabel country dan tabel fakta, dimana akan diurutkan dari paling bawah dengan batas hanya 100 data.

```
SELECT DISTINCT tb_country.country AS 'kota', tb_fakta.total_vaccinations, tb_implementation.date AS 'tanggal' FROM ((tb_country JOIN tb_fakta ON
tb_country.country_id = tb_fakta.country_ID)JOIN tb_implementation ON tb_country.country_id = tb_implementation.date_ID) GROUP BY
tb_fakta.total_vaccinations ORDER BY tb_fakta.total_vaccinations DESC LIMIT 100;
```

Gambar 21. Kueri 1 *Drill Down*

kota	total_vaccinations ▾ 1	tanggal
Angola	17535411	2022-03-25
Angola	17262044	2022-03-17
Angola	16850195	2022-03-08
Angola	16633167	2022-03-02
Angola	16259606	2022-02-22
Angola	15902065	2022-02-15
Angola	15505389	2022-02-09
Angola	15039557	2022-02-02
Angola	14588435	2022-01-28
Angola	13944656	2022-01-23
Algeria	13704895	2022-03-09
Algeria	13631683	2022-02-20

Gambar 22. Hasil dari Kueri 1 *Drill Down*

ROLL UP

Pada fungsi ini kami mengambil dari tabel country dan tabel fakta, dimana akan diurutkan dari paling atas sesuai tanggal.

```
SELECT tb_country.country_id, tb_country.country, tb_implementation.date, tb_fakta.daily_vaccinations FROM ((tb_fakta JOIN tb_country ON
tb_fakta.country_ID = tb_country.country_id) JOIN tb_implementation ON tb_country.country_id = tb_implementation.date_ID);
```

Gambar 21. Kueri 1 *Roll Up*

country_id	country	date	daily_vaccinations
1	Afghanistan	2021-02-22	0
2	Afghanistan	2021-02-23	1367
3	Afghanistan	2021-02-24	1367
4	Afghanistan	2021-02-25	1367
5	Afghanistan	2021-02-26	1367
6	Afghanistan	2021-02-27	1367
7	Afghanistan	2021-02-28	1367
8	Afghanistan	2021-03-01	1580
9	Afghanistan	2021-03-02	1794
10	Afghanistan	2021-03-03	2008

Gambar 23. Hasil dari Kueri 1 *Roll Up*

3.2.3 Pivot

Pivot adalah operasi OLAP yang merubah atau merotasi sumbu data untuk menghasilkan tampilan data yang berbeda.

Query 1

Pada kode dibawah ini, kami menggunakan fungsi COUNT untuk menghitung data sesuai nama dari vaksin yang digunakan di negara Afghanistan, dan mengubah nama tabel dengan nama vaksinnya.

```
SELECT tb_country.country_id, COUNT(IF(tb_vaccines.vaccines = 'Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing',tb_vaccines.vaccines,NULL)) AS 'Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing' , COUNT(IF(tb_vaccines.vaccines = 'Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V',tb_vaccines.vaccines,NULL)) AS 'Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V' , COUNT(IF(tb_vaccines.vaccines = 'Moderna, Oxford/AstraZeneca, Pfizer/BioNTech',tb_vaccines.vaccines,NULL)) AS 'Moderna, Oxford/AstraZeneca, Pfizer/BioNTech' FROM tb_country, tb_vaccines;
```

Gambar 24. Kueri 1 *Pivot*

country	Johnson&Johnson, Oxford/AstraZeneca, Pfizer/BioNTech, Sinopharm/Beijing	Oxford/AstraZeneca, Pfizer/BioNTech, Sinovac, Sputnik V	Moderna, Oxford/AstraZeneca, Pfizer/BioNTech
Afghanistan	1241494	1383289	1301363

Gambar 25. Hasil dari Kueri 1 *Pivot*

Query 2

Pada kode ini, menggunakan fungsi SUM untuk menghitung berapa hari vaksin yang sudah dilakukan di suatu negara. dimana per harinya akan mengeluarkan output 1, tetapi karena lebih dari 1 hari, maka outputnya ditotal dan dimasukkan ke dalam kolom baru.

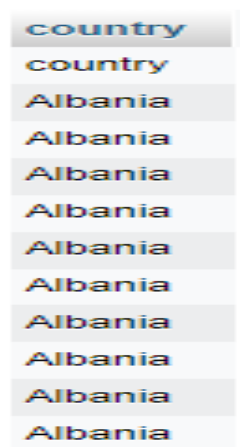
3.2.4 Filtering

Filtering berguna untuk menampilkan data yang telah disaring sebelumnya untuk menampilkan hasil yang lebih spesifik sesuai dengan data yang ingin diketahui

```
SELECT country FROM tb_fakta_vaksinasi WHERE NOT source_website = "https://covid19.who.int/";
```

Gambar 26. Kueri 1 *Filtering*

Untuk mengetahui negara mana saja yang ingin diketahui namun dengan ketentuan tidak berasal dari salah sumber informasi tertentu maka digunakan perintah SELECT pilih kolom tabel yang ingin diketahui yaitu negara yang dipilih FROM tabel fakta vaksinasi WHERE NOT jika tidak berasal dari source website <https://covid19.who.int/>.



country
country
Albania
Albania
Albania
Albania
Albania
Albania
Albania
Albania
Albania
Albania

Gambar 27. Hasil dari Kueri 1 *Filtering*

```
SELECT * FROM tb_fakta_vaksinasi WHERE Date_ID = ' 600' AND vaccines_id = '600';
```

Gambar 28. Kueri 2 *Filtering*

Untuk mengetahui data yang ingin diketahui dengan ketentuan date_id dan vaccines_id = 600 maka digunakan perintah SELECT seluruh kolom tabel dari tabel fakta vaksinasi WHERE date_id dan vaccines_id memiliki nilai 600.

country	Date_ID	total_vaccinated	people_fully_vaccinated	daily_vaccinations	vaccines_id	source_website
Albania	600	683052	538588	8910	600	https://shendetesia.gov.al/vaksinimi-anticovid-275

Gambar 29. Hasil dari Kueri 2 *Filtering*

3.3 Hasil Aplikasi

3.3.1 Perhitungan Aplikasi

Langkah pertama adalah memasukkan dataset yang sudah kami buat sebelumnya dan ditambahkan dataset baru dari WHO terkait data kematian dan dataset dari kaggle terkait data populasi di tiap negara. Dengan metode ini kami ingin menghitung bagaimana keterkaitan dari rasio kematian di tiap negara dengan persentase total dari orang yang sudah di vaksinasi di tiap negara.

Langkah awal yaitu kami mengimpor library yang akan kami gunakan dan memasukkan data dari 3 dataset yang kami peroleh ke dalam collab, beserta hasilnya :

```
from sklearn import linear_model
import pandas as pd
import matplotlib.pyplot as plt
import pylab as pl
import numpy as np
%matplotlib inline
```

Gambar 30. library

```
[22] who_source = pd.read_csv("https://covid19.who.int/WHO-COVID-19-global-data.csv")
who_source = who_source.rename(columns={"Country": "country", "Date_reported": "date" })
who_source.head(2)
```

	date	Country_code	country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
0	2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
1	2020-01-04	AF	Afghanistan	EMRO	0	0	0	0

Gambar 31. Data WHO

```
[23] vaccine_source = pd.read_csv("/content/country_vaccinations.csv")
vaccine_source.head(2)
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccin
0	Afghanistan	AFG	2021-02-22	0.0	0.0	NaN	NaN	NaN	NaN	0.0
1	Afghanistan	AFG	2021-02-23	NaN	NaN	NaN	NaN	1367.0	NaN	NaN

Gambar 32. Data Vaksin

```
[24] population_source = pd.read_csv("/content/2021_population.csv")
population_source = population_source.rename(columns={"2021_last_updated": "population"})
population_source.head(2)
```

	iso_code	country	population	2020_population	area	density_sq_km	growth_rate	world_%	rank
0	CHN	China	1,447,065,329	1,439,323,776	9,706,961 sq_km	149/sq_km	0.34%	18.34%	1
1	IND	India	1,401,310,563	1,380,004,385	3,287,590 sq_km	424/sq_km	0.97%	17.69%	2

Gambar 33. Data populasi

Setelah itu akan dilakukan data understanding dan preparation, dimana dari data populasi dan who akan di cek perbedaan nama negara yang ada. Lalu akan didapatkan mana nama negara yang tersedia, dan seterusnya akan dilakukan perubahan nama menyesuaikan data dari WHO, lalu perubahan tersebut juga akan diterapkan pada data vaksin.

```

set(who_source.country) - set(population_source.country)

('Antigua and Barbuda',
 'Bolivia (Plurinational State of)',
 'Bonaire',
 'Bosnia and Herzegovina',
 'British Virgin Islands',
 'Brunei Darussalam',
 'Cabo Verde',
 'Congo',
 'Curaçao',
 'Czechia',
 'Côte d'Ivoire',
 'Democratic People's Republic of Korea',
 'Democratic Republic of the Congo',
 'Eswatini',
 'Falkland Islands (Malvinas)',
 'French Guiana',
 'Guadeloupe',
 'Guernsey',
 'Guinea-Bissau',
 'Holy See',
 'Iran (Islamic Republic of)',
 'Isle of Man',
 'Jersey',
 'Kosovo[1]',
 'Lao People's Democratic Republic',
 'Micronesia (federated States of)',

```

Gambar 34. Cek nama negara

```

toedit = ['Antigua and Barbuda': 'Antigua and Barbuda',
 'Bolivia': 'Bolivia (Plurinational State of)',
 'Bosnia and Herzegovina': 'Bosnia and Herzegovina',
 'Brunei': 'Brunei Darussalam',
 'Cape Verde': 'Cabo Verde',
 'Curaçao': 'Curaçao',
 'North Korea': 'Democratic People's Republic of Korea',
 'Republic of the Congo': 'Democratic Republic of the Congo',
 'Falkland Islands': 'Falkland Islands (Malvinas)',
 'Guinea-Bissau': 'Guinea-Bissau',
 'Iran': 'Iran (Islamic Republic of)',
 'Isle of Man': 'Isle of Man',
 'Laos': 'Lao People's Democratic Republic',
 'Micronesia': 'Micronesia (Federated States of)',
 'Northern Mariana Islands': 'Northern Mariana Islands (Commonwealth of the)',
 'South Korea': 'Republic of Korea',
 'Moldova': 'Republic of Moldova',
 'Russia': 'Russian Federation',
 'Reunion': 'Reunion',
 'Saint Kitts and Nevis': 'Saint Kitts and Nevis',
 'Saint Pierre and Miquelon': 'Saint Pierre and Miquelon',
 'Saint Vincent and the Grenadines': 'Saint Vincent and the Grenadines',
 'Sao Tome and Principe': 'Sao Tome and Principe',
 'Syria': 'Syrian Arab Republic',
 'United Kingdom': 'The United Kingdom',
 'Timor-Leste': 'Timor-Leste',
 'Trinidad and Tobago': 'Trinidad and Tobago',
 'Turks and Caicos Islands': 'Turks and Caicos Islands',
 'Tanzania': 'United Republic of Tanzania',
 'United States': 'United States of America',
 'Venezuela': 'Venezuela (Bolivarian Republic of)',
 'Vietnam': 'Viet Nam',
 'Wallis and Futuna': 'Wallis and Futuna',
 'Palestine': 'occupied Palestinian territory, including east Jerusalem']
population_source['country'] = population_source['country'].replace(toedit)
vaccine_source['country'] = vaccine_source['country'].replace(toedit)

```

Gambar 35. Ganti nama negara

Setelah itu kami menggabungkan beberapa kolom dari tabel di tiap dataset, untuk mendapatkan data akhir bernama df, data ini berupa contoh. Maka langkah selanjutnya adalah menggabungkan data dari who dan populasi menjadi sebuah data baru bernama who_population_merge. Dan juga kami menggabungkan data vaksin dengan data who_population_merge menjadi data df final. Lalu setelah itu dapat kita lihat kolom nama negara dari data akhir.

```

[27] who = who_source[['date', 'country', 'New_deaths']]
      population = population_source[['iso_code', 'country', 'population']]
      vaccine = vaccine_source[['country', 'iso_code', 'date', 'total_vaccinations', 'people_vaccinated', 'people_fully_vaccinated']]
      # We drop rows with NaN value at this stage from vaccine data frame
      vaccine = vaccine.dropna().reset_index(drop=True)

```

Gambar 36. Data df contoh

```

[28] who_population_merge = pd.merge(who, population)
      who_population_merge.head(2)

```

	date	country	New_deaths	iso_code	population
0	2020-01-03	Afghanistan	0	AFG	40,374,668
1	2020-01-04	Afghanistan	0	AFG	40,374,668

Gambar 37. who_population_merge

```
[29] df = pd.merge(vaccine_who_population_merge).reset_index(drop=True)
df.head(2)
```

	country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	New_deaths	population
0	Afghanistan	AFG	2021-05-11	504502.0	448878.0	55624.0	12	40,374,668
1	Afghanistan	AFG	2021-05-20	547901.0	470341.0	77560.0	10	40,374,668

Gambar 38. Data df final

```
df.country.unique()
array(['Afghanistan', 'Albania', 'Algeria', 'Andorra', 'Angola',
       'Anguilla', 'Antigua and Barbuda', 'Argentina', 'Armenia', 'Aruba',
       'Australia', 'Austria', 'Azerbaijan', 'Bahamas', 'Bahrain',
       'Bangladesh', 'Barbados', 'Belarus', 'Belgium', 'Belize', 'Benin',
       'Bermuda', 'Bhutan', 'Bolivia (Plurinational State of)',
       'Bosnia and Herzegovina', 'Botswana', 'Brazil',
       'Brunei Darussalam', 'Bulgaria', 'Burkina Faso', 'Burundi',
       'Cambodia', 'Cameroon', 'Canada', 'Cabo Verde', 'Cayman Islands',
       'Central African Republic', 'Chad', 'Chile', 'China', 'Colombia',
       'Comoros', 'Cook Islands', 'Costa Rica', 'Croatia', 'Cuba',
       'Curaçao', 'Cyprus', 'Denmark', 'Djibouti', 'Dominica',
       'Dominican Republic', 'Ecuador', 'Egypt', 'El Salvador',
       'Equatorial Guinea', 'Estonia', 'Ethiopia',
       'Falkland Islands (Malvinas)', 'Fiji', 'Finland', 'France',
       'French Polynesia', 'Gabon', 'Gambia', 'Georgia', 'Germany',
       'Ghana', 'Gibraltar', 'Greece', 'Greenland', 'Grenada',
       'Guatemala', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti',
       'Honduras', 'Hungary', 'Iceland', 'India', 'Indonesia',
       'Iran (Islamic Republic of)', 'Iraq', 'Ireland', 'Israel', 'Italy',
       'Jamaica', 'Japan', 'Jordan', 'Kazakhstan', 'Kenya', 'Kiribati',
       'Kuwait', 'Kyrgyzstan', 'Lao People's Democratic Republic',
       'Latvia', 'Lebanon', 'Lesotho', 'Liberia', 'Libya',
       'Liechtenstein', 'Lithuania', 'Luxembourg', 'Madagascar', 'Malawi',
       'Malaysia', 'Maldives', 'Mali', 'Malta', 'Mauritania', 'Mauritius',
       'Mexico', 'Republic of Moldova', 'Monaco', 'Mongolia',
       'Montenegro', 'Montserrat', 'Morocco', 'Mozambique', 'Myanmar',
       'Namibia', 'Nauru', 'Nepal', 'Netherlands', 'New Caledonia',
       'New Zealand', 'Nicaragua', 'Niger', 'Nigeria', 'Niue',
       'North Macedonia', 'Norway', 'Oman', 'Pakistan',
       'occupied Palestinian territory, including east Jerusalem',
       'Panama', 'Papua New Guinea', 'Paraguay', 'Peru', 'Philippines',
       'Poland', 'Portugal', 'Qatar', 'Romania', 'Russian Federation',
       'Rwanda', 'Saint Kitts and Nevis', 'Saint Lucia', 'Samoa',
       'San Marino', 'Sao Tome and Principe', 'Saudi Arabia', 'Senegal',
       'Serbia', 'Seychelles', 'Sierra Leone', 'Singapore', 'Slovakia',
       'Slovenia', 'Solomon Islands', 'Somalia', 'South Africa',
       'Republic of Korea', 'South Sudan', 'Spain', 'Sri Lanka', 'Sudan',
       'Suriname', 'Sweden', 'Switzerland', 'Syrian Arab Republic',
       'Tajikistan', 'United Republic of Tanzania', 'Thailand', 'Togo',
       'Tokelau', 'Tonga', 'Trinidad and Tobago', 'Tunisia',
       'Turkmenistan', 'Turks and Caicos Islands', 'Tuvalu', 'Uganda',
       'Ukraine', 'United Arab Emirates', 'The United Kingdom',
       'United States of America', 'Uruguay', 'Uzbekistan', 'Vanuatu',
       'Venezuela (Bolivarian Republic of)', 'Viet Nam',
       'Wallis and Futuna', 'Yemen', 'Zambia', 'Zimbabwe'], dtype=object)
```

Gambar 39. Kolom negara

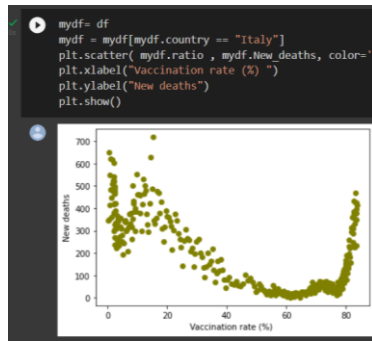
Setelah itu dari data yang ada akan kami hapus tanda koma, dan mengganti tipe datanya menjadi float. Lalu kita menghitung persentase dari vaksinasi di tiap negara, dan dimasukkan ke dalam kolom baru yaitu rasio. Maka selanjutnya akan kami lakukan plotting dari kolom persentase vaksinasi yang dibandingkan dengan kolom jumlah kematian baru dari kasus covid-19, dan yang coba kami plotting yaitu dari negara italia. Dan seperti yang bisa kita lihat, bahwa semakin tinggi vaksinasi yang terjadi di italia, maka akan semakin rendah rasio kematiannya.

```
[31] def f_1(row):
      row.population = row.population.replace(',', '')
      return row
df=df.apply(f_1, axis=1)
df.population = df.population.astype("float64")
df.people_vaccinated = df.people_vaccinated.astype("float64")
```

Gambar 40. Ganti tipe data

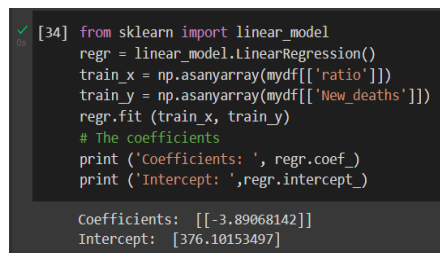
```
[32] df = df.assign(ratio=[0]*len(df))
def f_2(row):
      row.ratio = row.people_vaccinated / row.population * 100
      return row
df=df.apply(f_2, axis=1)
```

Gambar 41. Rasio

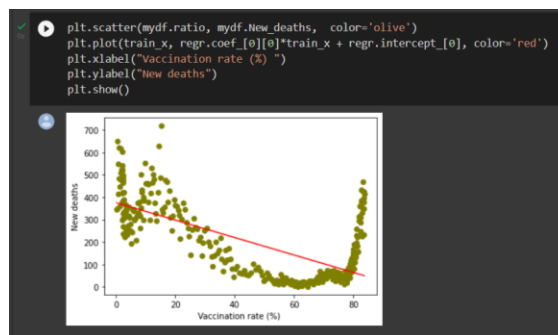


Gambar 42. Plotting negara Italia

Lalu akan kami lakukan modelling dari data yang sudah kami siapkan tadi. Pertama akan kami cari nilai koefisien dan intercept nya. Dan setelah itu kami buat grafik yang menunjukkan garis regresinya.



Gambar 43. Nilai koefisien dan konstanta(intercept)



Gambar 44. Garis regresi

Setelah model didapatkan, maka akan dilakukan evaluasi. Evaluasi ini digunakan untuk melihat seberapa baik model yang kita buat, disini kami mendapat nilai 0,50. Yang mana model yang kami buat sudah cukup baik. Lalu model yang kami buat ini, kami coba pada data negara portugal, dan hasil yang didapatkan yaitu kurang baik, karena nilainya 0,20. Tetapi dari grafik yang ada, bisa kita lihat juga bahwa semakin tinggi persentase vaksinasi yang terjadi, maka akan semakin berkurang rasio kematiannya juga.

```

▶ from sklearn.metrics import r2_score

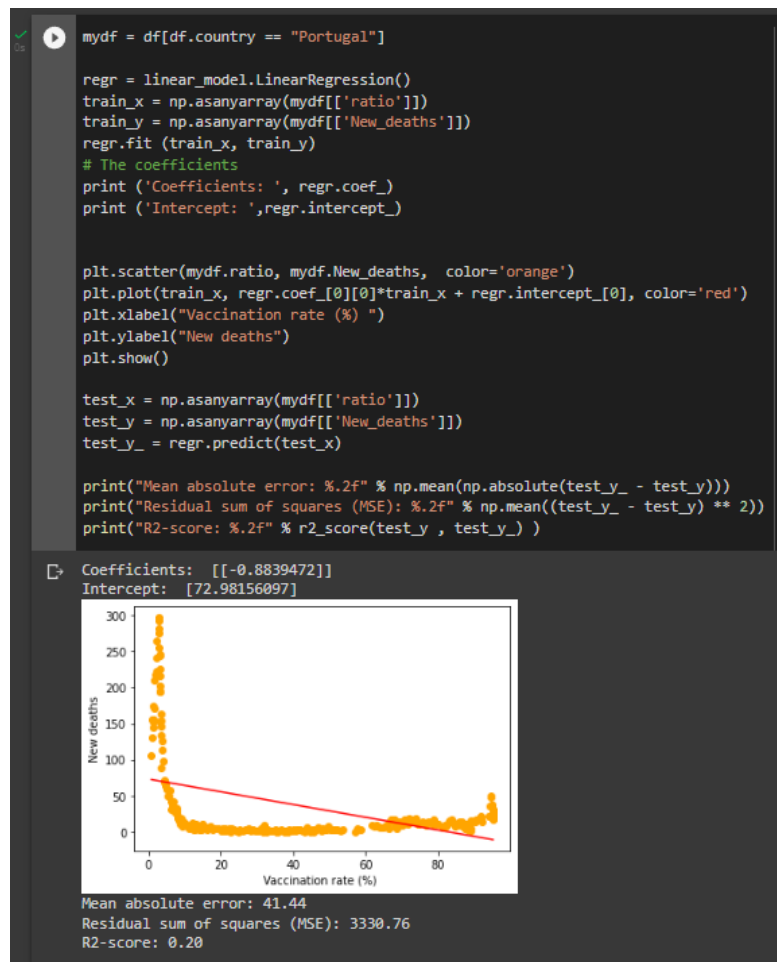
test_x = np.asarray(mydf[['ratio']])
test_y = np.asarray(mydf[['New_deaths']])
test_y_ = regr.predict(test_x)

print("Mean absolute error: %.2f" % np.mean(np.absolute(test_y_ - test_y)))
print("Residual sum of squares (MSE): %.2f" % np.mean((test_y_ - test_y) ** 2))
print("R2-score: %.2f" % r2_score(test_y , test_y_ ) )

```

Mean absolute error: 89.00
Residual sum of squares (MSE): 13971.64
R2-score: 0.50

Gambar 45. Hasil evaluasi



Gambar 46. Mencoba model pada negara Portugal

3.3.2 Perhitungan Manual

Perhitungan Konstanta A dan Koefisien B

Dengan mengambil beberapa contoh negara diketahui sebagai berikut :

No	Negara	Vaksin_rate(x)	new_death(y)	x2	y2	xy
1	afghanistan	10,26	5167	105,2676	26697889	53013,42
2	albania	39,77	2041	1581,6529	4165681	81170,57
3	algeria	15,66	3520	245,2356	12390400	55123,2
4	andorra	73,79	56	5444,9641	3136	4132,24
5	angola	22,23	1352	494,1729	1827904	30054,96
total		161,71	12136	7871,2931	45085010	223494,39

Dengan sudah diketahuinya x2, y2, dan xy maka dapat dicari untuk konstanta a dan koefisien b menggunakan rumus:
$$a = \frac{(\sum y) - (\sum x^2) - (\sum x)(\sum xy)}{n(\sum x^2) - (\sum x)^2}$$
 dan
$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n(\sum x^2) - (\sum x)^2}$$

Dengan menghasilkan hasil perhitungan sebagai berikut ini:

Konstanta A = 4496,683332 dan Koefisien B = -63,98748786

Setelah mendapatkan hasil konstanta a dan koefisien b, langkah selanjutnya ialah membuat model Regresi dengan rumus :

$$Y = a + bX$$

Dengan rumus tersebut dapat dihasilkan

Y = 4496,683332 + -63,98748786 X	
Y = 4496,683332 + -63,98748786 (42%)	
Y =	4469,808587

Prediksikan Jumlah kematian jika rata2 vaksinasi rendah (Variabel X), contohnya : 42%

Selanjutnya kita akan menghitung nilai korelasi yang berguna untuk mengukur kuatnya hubungan antara dua variable atau lebih. Besarnya korelasi berkisar antara -1 0 dan 1. Besaran koefisien -1 dan 1 adalah hubungan sempurna, sedangkan nilai koefisien 0 atau mendekati 0 dianggap tidak berhubungan antara dua variabel yang diuji.

Menghitung korelasi itu sendiri dapat dilakukan dengan menggunakan rumus berikut :

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\{n\sum x^2 - (\sum x)^2\}\{n\sum y^2 - (\sum y)^2\}}}$$

Dengan hasil perhitungan $r = -845040,61/1015862,809$ yang menghasilkan $r = -0,8318452086$
Kemudian kita akan menghitung nilai koefisien determinasi (r^2) :

$$r^2 = -0,8318452086^2 = 0,6919664511$$

Hal ini memberikan informasi bahwa kasus kematian yang terjadi **69%** ditentukan oleh rasio vaksinasi

3.3.3 Rules

1. Menentukan tujuan dari melakukan analisis Regresi Linear yaitu mempelajari hubungan yang diperoleh dan dinyatakan dalam persamaan matematika yang menyatakan hubungan antar variabel.
2. Mengidentifikasi Variabel Faktor Penyebab (X) dan Variabel Akibat (Y)
 - a. Variabel Faktor Penyebab (X) : Vaccination rate
 - b. Variabel Faktor Akibat(Y) : New deaths
3. Melakukan Pengumpulan Data

Data yang digunakan adalah data yang bersumber dari :

 - <https://www.kaggle.com/datasets/rsrishav/world-population>
Data yang diambil merupakan data populasi dari seluruh negara
 - <https://www.kaggle.com/datasets/gpreda/covid-world-vaccination-progress>
Data yang diambil merupakan data proses vaksinasi yang di lakukan di seluruh negara
 - <https://covid19.who.int/data>
Data yang diambil merupakan data angka kematian di setiap negara
4. menghitung X^2 , Y^2 , XY
5. menghitung nilai dari konstanta (a) dan koefisien (b) berdasarkan persamaan yang sudah ditentukan
6. membuat model persamaan regresi linear sederhana
7. langkah yang terakhir dilakukan proses korelasi atau seberapa besar pengaruh variabel X dengan variabel Y

BAB IV

PENUTUP

4.1 Simpulan

Kesimpulan yang diperoleh dari tugas ini adalah dalam membangun data warehouse dibutuhkan tools untuk menunjang hal tersebut seperti PHPMYADMIN sebagai alat yang dapat menangani administrasi dari MySQL dengan melalui web yang penulis gunakan adalah XAMPP, sistem manajemen dengan basis data menggunakan MYSQL dan data yang akan digunakan.

Data warehouse yang telah dibangun yaitu data vaksinasi dunia maka diperlukan teknik menganalisisnya. Dalam hal ini penulis menggunakan teknik OLAP (Online Analytical Processing) untuk memudahkan dalam menganalisis, mengelola dan mengidentifikasi informasi yang dibutuhkan ada didalam data yang lebih efektif dan efisien seperti mengetahui informasi perkembangan data vaksinasi di seluruh negara sesuai dengan jumlah vaksin, orang yang telah divaksin, jenis vaksin dan lainnya.

Kemudian untuk mengetahui suatu informasi berupa hubungan pengaruh rata-rata vaksinasi dengan jumlah kematian setiap negara didalam data warehouse dilakukan dengan menggunakan algoritma regresi linier. Yang hasilnya berupa jumlah vaksinasi semakin banyak maka jumlah kematian akan berkurang begitu pula jika jumlah vaksinasi kurang maka jumlah kematian cukup tinggi dan ke-2 hal tersebut terjadi disetiap negara yang memiliki kondisi pandemi covid nya tersendiri.