

COVID 19 VACCINATION BY COUNTRY

Members:

William Powell, Jungkyu Kim, Kaixin Yang,
Hannah Slay, Alex Gainer



ETL PROJECT

01

EXTRACT

02

TRANSFORM

03

LOAD

04


CONCLUSIONS





⁺ Goal⁺

Have clean data that allows us to compare multiple variables and how they relate to Covid vaccine rollouts.





Finding Data



We found all of our data on Kaggle. The following are links to the specific data sets.

Covid World Vaccination Progress:

- <https://www.kaggle.com/gpreda/covid-world-vaccination-progress>

World Happiness Data:

- <https://www.kaggle.com/unsdsn/world-happiness>

UN Country Statistics:

- <https://www.kaggle.com/sudalairajkumar/undata-country-profiles>
- 



Project Process

Create
Postgres
database &
tables

01

Read in
CSV/JSON

02

Remove
unnecessary
columns

03

Retrieve most
recent date
from individual
datasets

04

Set DataFrame
equal to most
recent date.

05



Project Process

Standardized
formatting in
preparation for
Postgresql



Load into
tables



Create database
schema



Create project
report



The type of final production database to load the data into

We chose relational based on Country.





WORLD DATA SCHEMA

www.quickdatabasediagrams.com

vaccine	
country	varchar(50)
date	date
total_vaccinations	bigint
people_vaccinated	bigint
people_fully_vaccinated	bigint
daily_vaccination	bigint
total_vaccinations_per_hundred	bigint
people_vaccinated_per_hundred	bigint
people_fully_vaccinated_per_hundred	bigint
daily_vaccinations_per_million	bigint
vaccines	varchar(50)

UN	
country	varchar(50)
Region	varchar(50)
Population in thousands (2017)	bigint
Population density (per km2, 2017)	bigint
Sex ratio (m per 100 f, 2017)	int
GDP: Gross domestic product (million current US\$)	bigint
GDP per capita (current US\$)	bigint
Unemployment (% of labour force)	int
Population growth rate (average annual %)	int
Urban population (% of total population)	int
Urban population growth rate (average annual %)	int
Life expectancy at birth (females/males, years)	varchar(10)
Population age distribution (0-14 / 60+ years, %)	varchar(10)
Health: Total expenditure (% of GDP)	int
Health: Physicians (per 1000 pop.)	int

happiness	
Country or region	varchar(50)
Score	int
Social support	int
Healthy life expectancy	int
Freedom to make life choices	int
Generosity	int
Perceptions of corruption	int



THANKS!

CREDITS: This presentation template was created by
Slidesgo, including icons by **Flaticon**, and
infographics & images by **Freepik**.