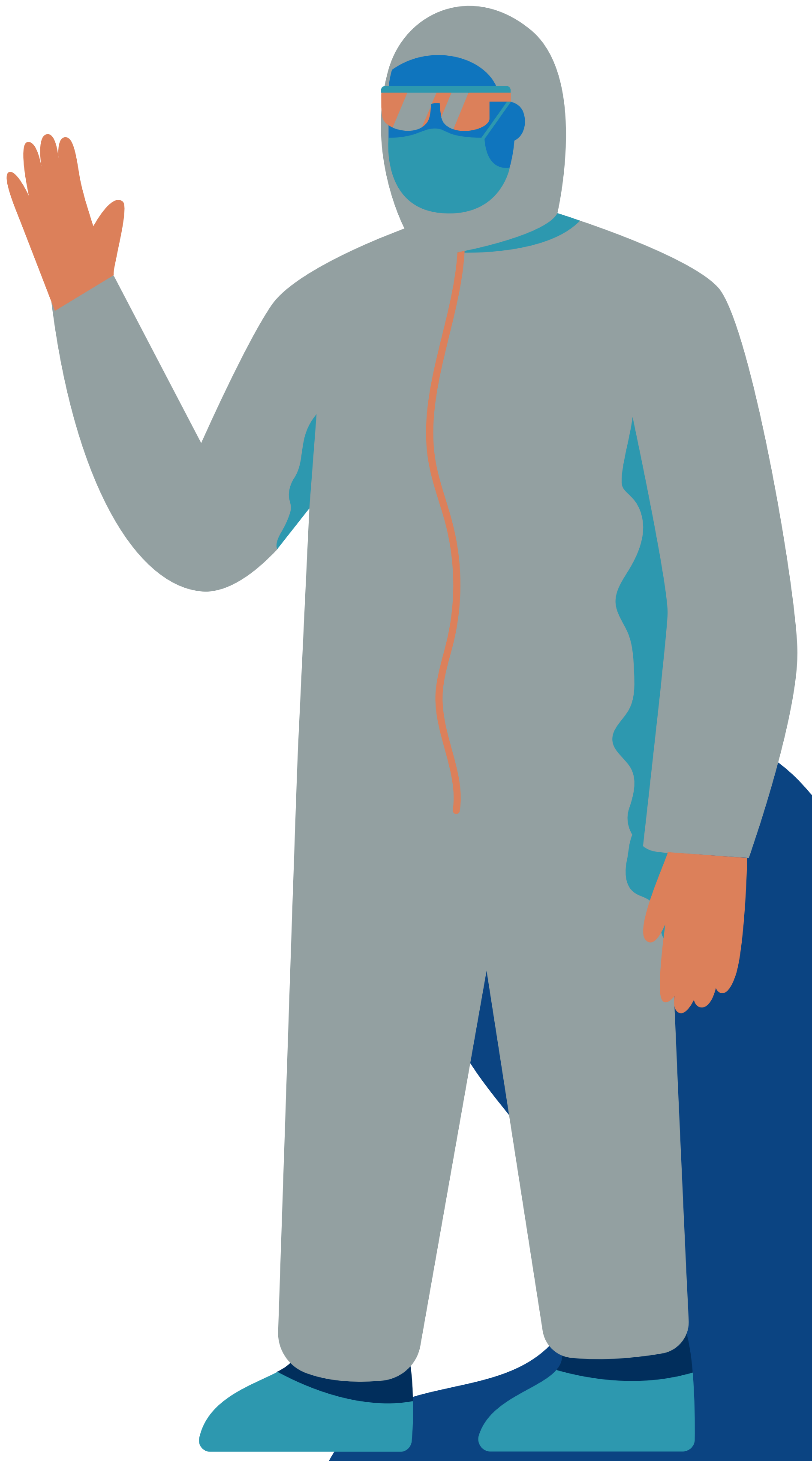
CSMODEL S12

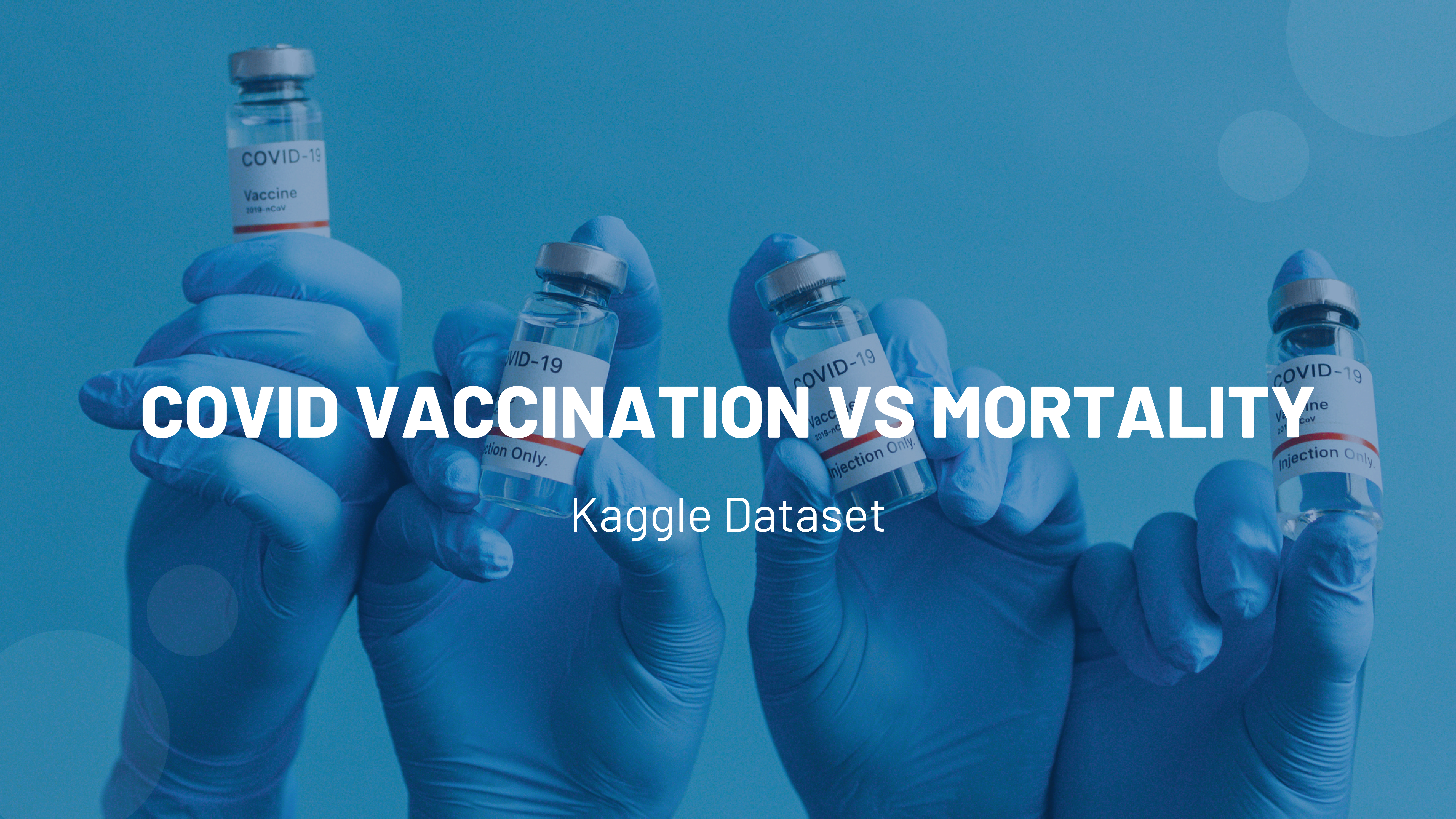
Investigating COVID Vaccination vs Mortality Data

Members: Ong, Teves, Yu



COVID VACCINATION VS MORTALITY

Kaggle Dataset



01.
Description of Data & Method of Collection

Data Sources for COVID vaccination vs. mortality

1. COVID-19 World Vaccination Progress

2. WHO COVID-19 Global Data

3. 2021 World Population

Dataset Structure

a. Variables

1. Unnamed
2. country
3. iso_code
4. date
5. total_vaccinations
6. people_vaccinated
7. people_fully_vaccinated
8. New_deaths
9. population
10. ratio

b. Observations

- 32,911 Observations
- Each Observation Contains Data for: Country, Date, Vaccination Information (Partial and Complete), New Deaths Reported

02.
Data Cleaning

DATA CLEANING PROCESS

1. Identified the number of entries per country
2. Decided on the scope of the research
3. Filtered the dataset by the ISO code
4. Selected countries for exploratory data analysis
5. Filtered the dataset using Indonesia, Malaysia, & Singapore ISO codes
6. Ready for exploratory data analysis

```
[4]  covid_df['country'].value_counts()
```

```
Israel                        466
Estonia                       458
Germany                       458
Argentina                     456
Denmark                       455
                             ...
Lesotho                        12
Nauru                           7
Ethiopia                        5
Niue                            4
Falkland Islands (Malvinas)     2
Name: country, Length: 136, dtype: int64
```

```python
# An array containing the ISO code of every Southeast Asian country.
se_asia = ["BRN", "MMR", "KHM", "TLS", "IDN", "LAO", "MYS", "PHL", "SGP", "THA", "VNM"]
# Creating a filtered dataset containing only SEA countries from the original dataset.
covid_se_df = covid_df[covid_df["iso_code"].isin(se_asia)]
covid_se_df
```

```
covid_se_df['country'].value_counts()
```

```
Singapore                           448
Malaysia                            399
Cambodia                            385
Indonesia                           373
Thailand                            345
Viet Nam                            172
Brunei Darussalam                   154
Lao People's Democratic Republic     88
Philippines                          82
Myanmar                              37
Name: country, dtype: int64
```

03.
Exploratory Data Analysis (EDA)

1. Which country has the most dates recorded?

Objective:

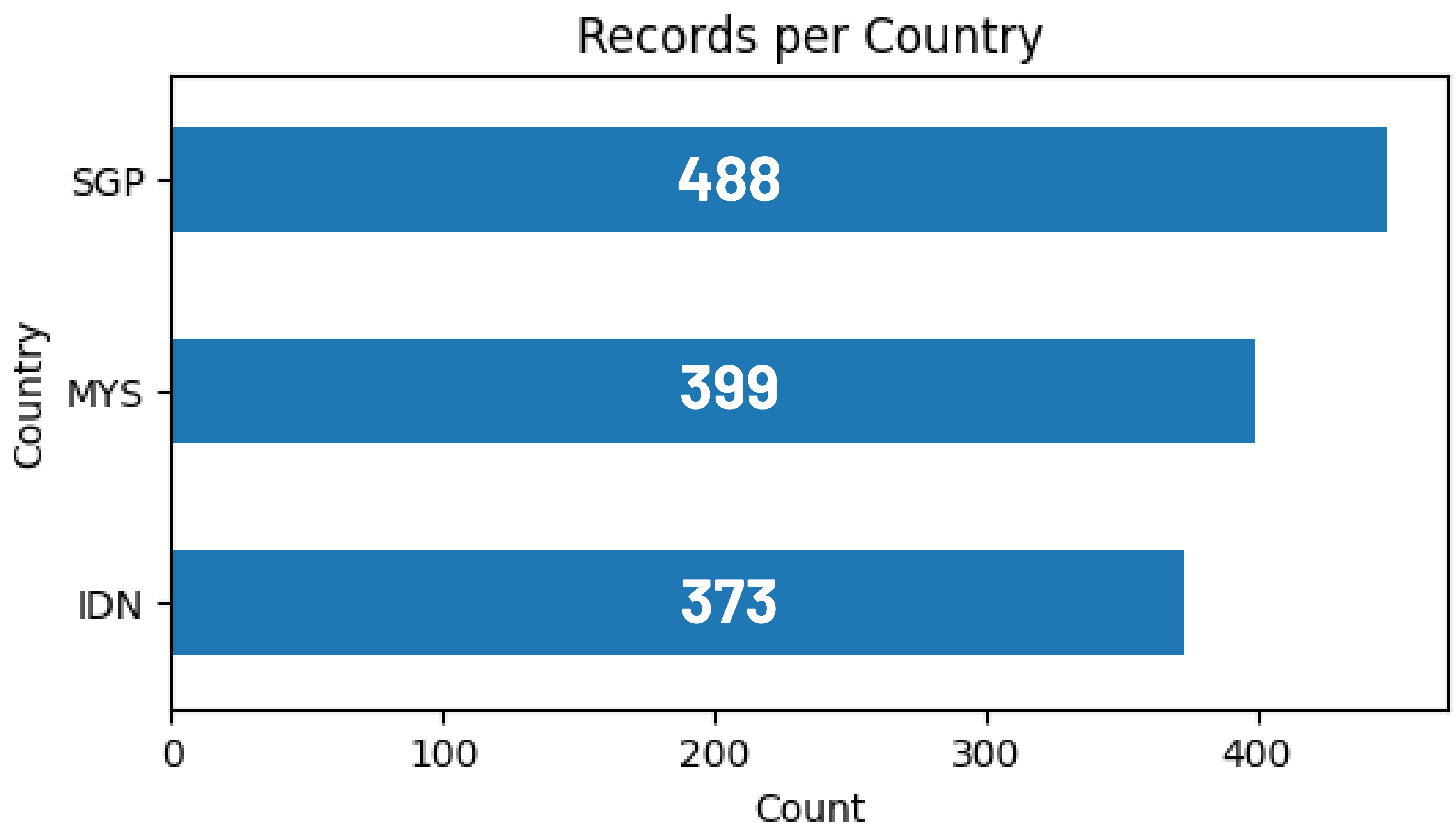
Identify the country with the most extensive COVID-19 data records.

Method:

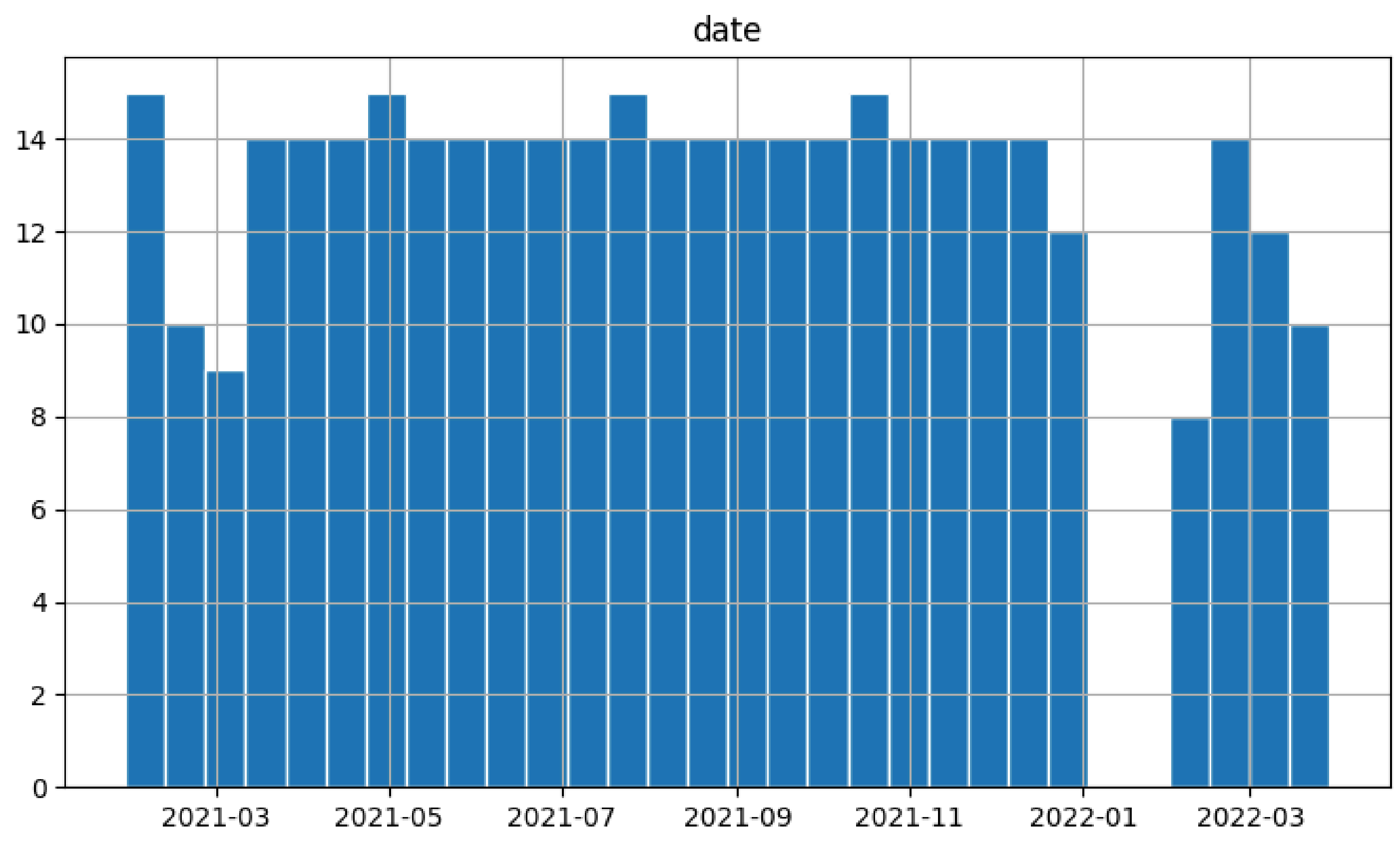
- Utillized *value_counts* tio count entries based on either *country* or *iso_code*
- Visualized data distribution using a bar chart

Results:

Singapore having the most entries or dates recorded, followed by Malaysia and Indonesia.

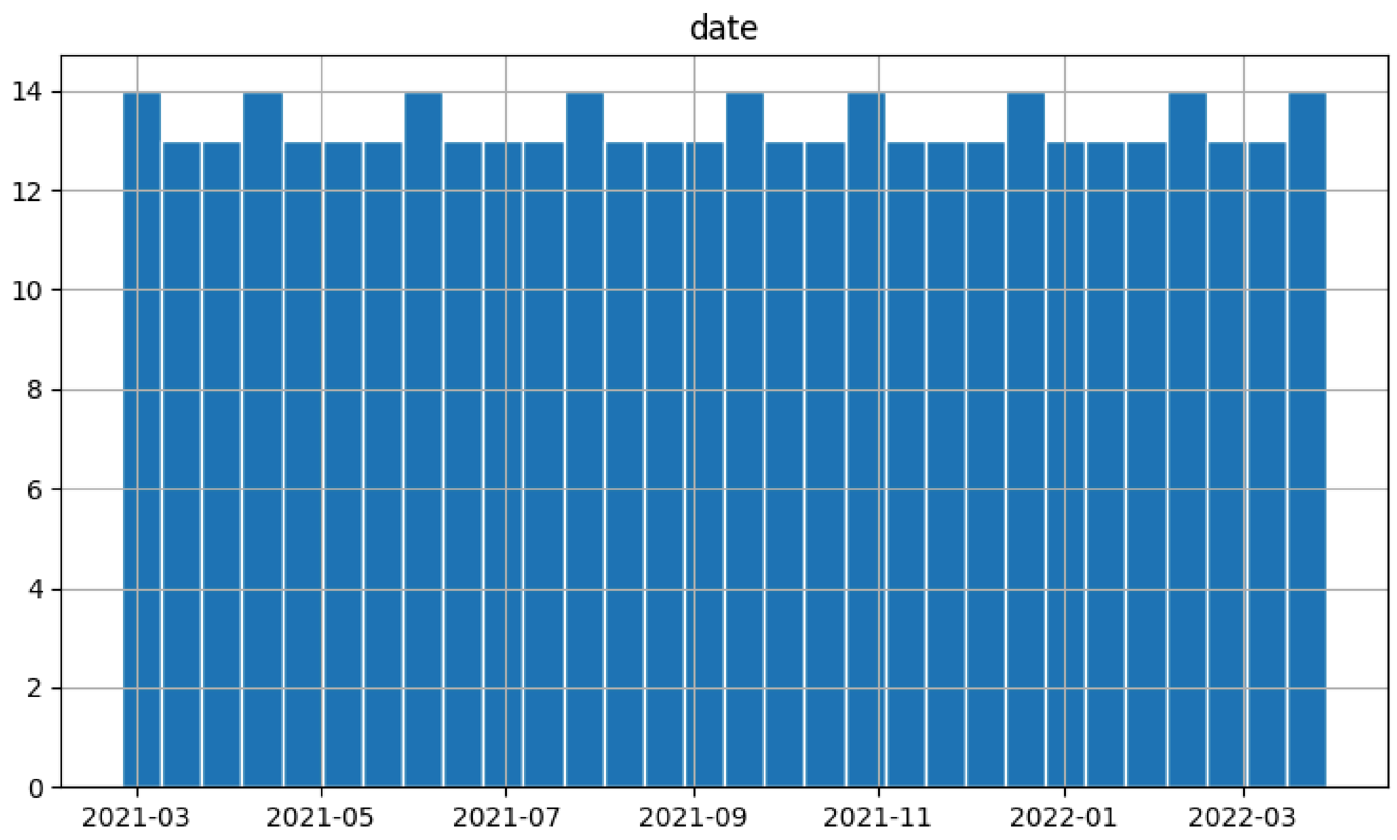


2. What is the distribution of logs recorded for each country throughout the months?



Indonesia

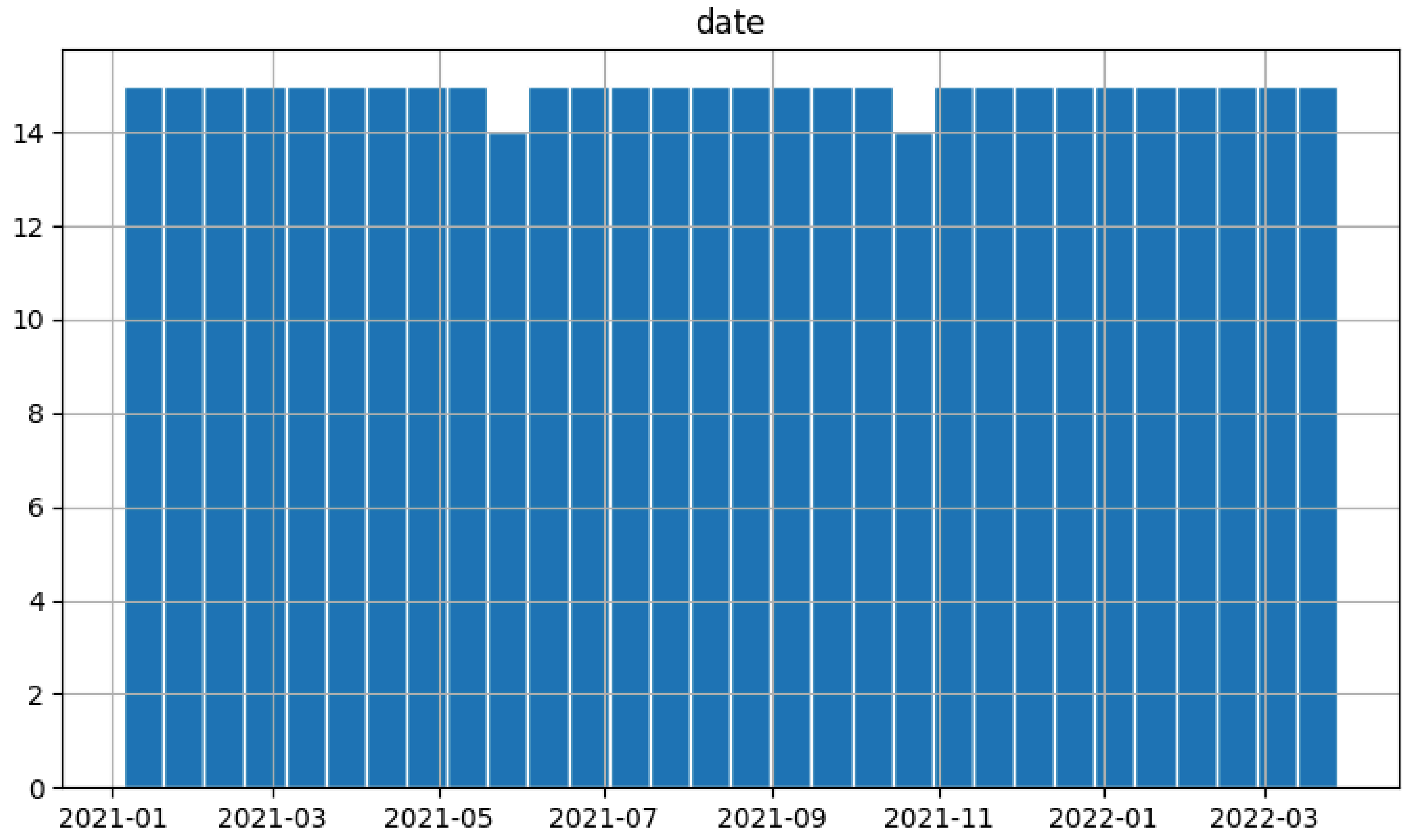
Mean: 26.64
Median: 30
Standard Deviation: 7.35



Malaysia

Mean: 28.50
Median: 30.50
Standard Deviation: 6.82

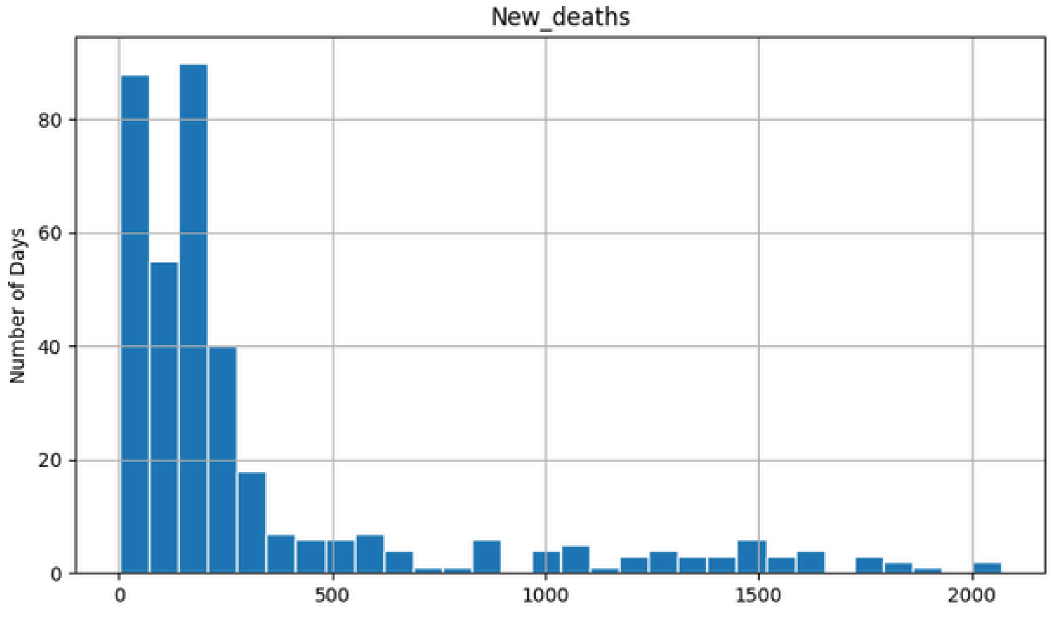
2. What is the distribution of logs recorded per country throughout the months?



Singapore

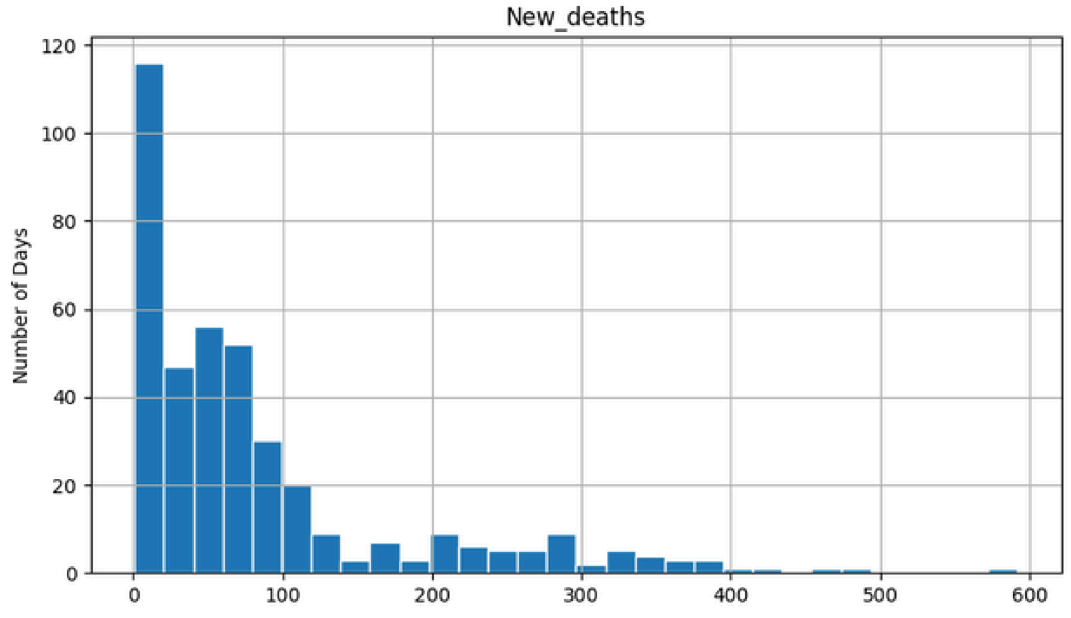
Mean: 29.87
Median: 30
Standard Deviation: 1.41

3. What is the distribution of new deaths per country?



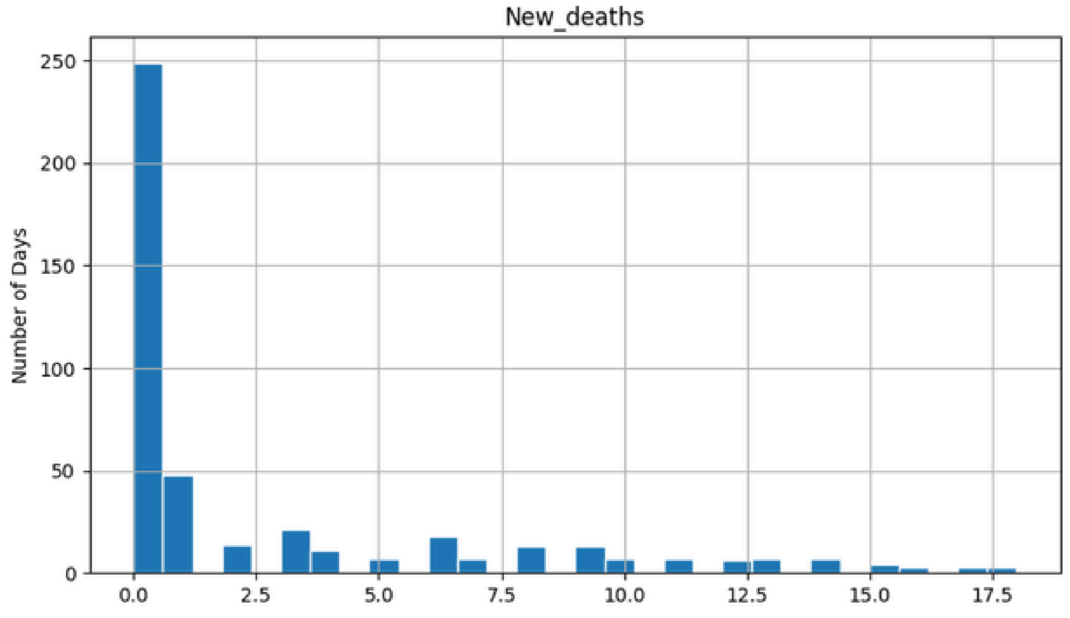
Indonesia

Average number of deaths
= 382.24



Malaysia

Average number of deaths
= 84.66



Singapore

Average number of deaths
= 2.72

4. What is the trend of new deaths due to COVID-19 in 2021?

Indonesia:

Mean = 0.000124

Median = 0.000060

Standard deviation = 0.000168

Malaysia:

Mean = 0.000296

Median = 0.000185

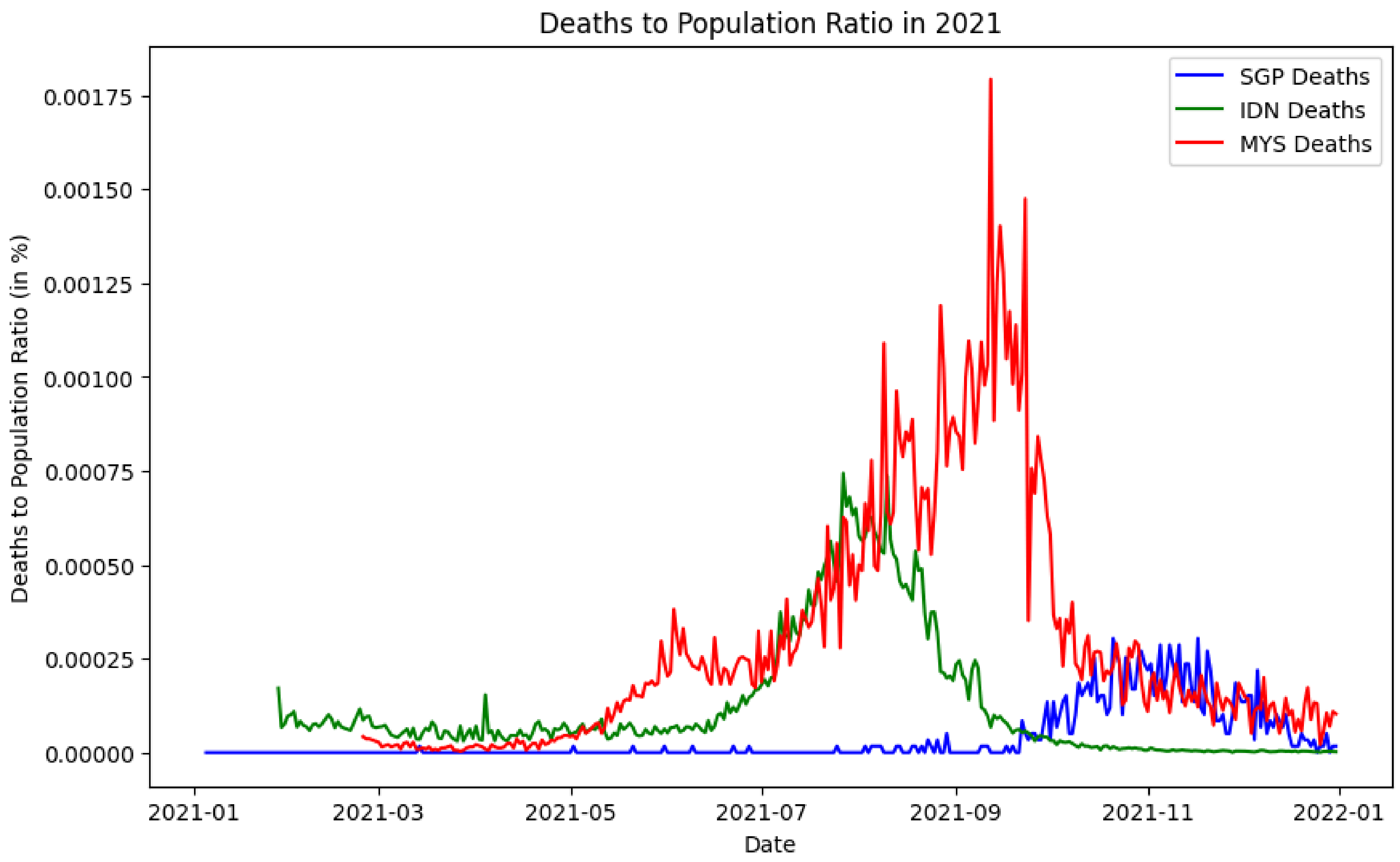
Standard deviation = 0.000324

Singapore:

Mean = 0.000037

Median = 0

Standard deviation = 0.000072



```
[32] # Calculate the ratio of deaths to population for each country
     sgp_df_2021['Deaths_to_Population'] = sgp_df_2021['New_deaths']*100 / sgp_df_2021['population']
     idn_df_2021['Deaths_to_Population'] = idn_df_2021['New_deaths']*100 / idn_df_2021['population']
     mys_df_2021['Deaths_to_Population'] = mys_df_2021['New_deaths']*100 / mys_df_2021['population']

     # Display a portion of sgp_df_2021 as an example proving it has been inserted
     sgp_df_2021.head()
```

```
[33] # Aggregating Indonesian deaths to population ratio in 2021
agg_idn = idn_df_2021.agg({"Deaths_to_Population": ["mean", "median", "std"]})
agg_idn
```

	Deaths_to_Population
mean	0.000124
median	0.000060
std	0.000168

```
[34] # Aggregating Malaysian deaths to population ratio in 2021
agg_mys = mys_df_2021.agg({"Deaths_to_Population": ["mean", "median", "std"]})
agg_mys
```

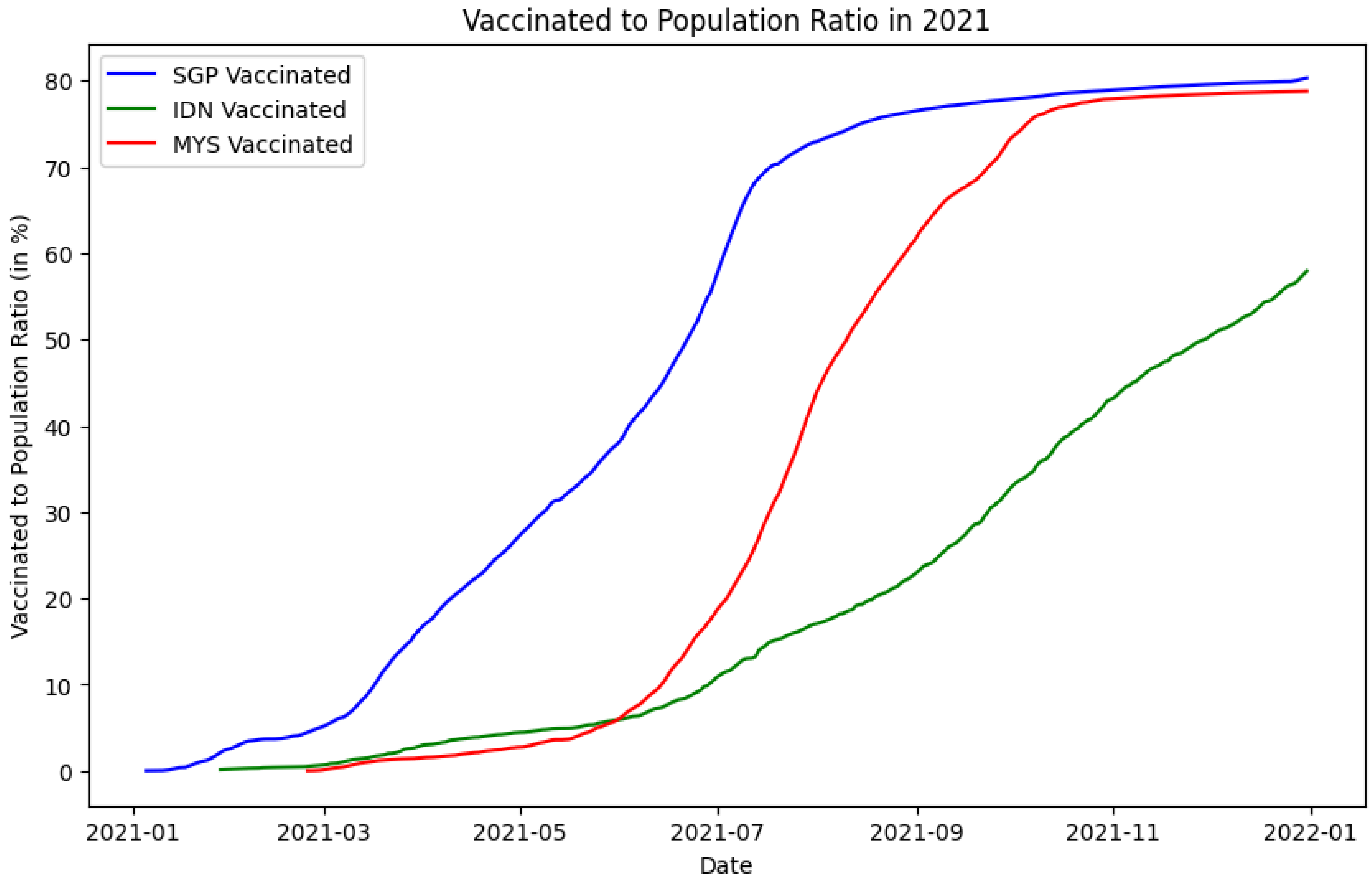
	Deaths_to_Population
mean	0.000296
median	0.000185
std	0.000324

```
# Aggregating Singaporean deaths to population ratio in 2021
agg_sgp = sgp_df_2021.agg({"Deaths_to_Population": ["mean", "median", "std"]})
agg_sgp
```

	Deaths_to_Population
mean	0.000037
median	0.000000
std	0.000072

5. What is the trend of vaccinations due to COVID-19 in 2021?

Vaccinated to Population



Indonesia

Mean: 20.331808

Standard deviation: 18.298695

Average vaccination rate: 20.33%

Malaysia

Mean: 39.559622

Standard deviation: 32.605549

Average vaccination rate: 39.56%

Singapore

Mean: 48.558395

Standard deviation: 30.706552

Average vaccination rate: 48.56%

```
✓ 1s [95] # Aggregating Indonesian vaccinated to population ratio in 2021
agg_idn = idn_df_2021.agg({"ratio": ["mean", "std"]})
agg_idn
```

	ratio
mean	20.331808
std	18.298695

```
✓ 0s [96] # Aggregating Malaysian vaccinated to population ratio in 2021
agg_mys = mys_df_2021.agg({"ratio": ["mean", "std"]})
agg_mys
```

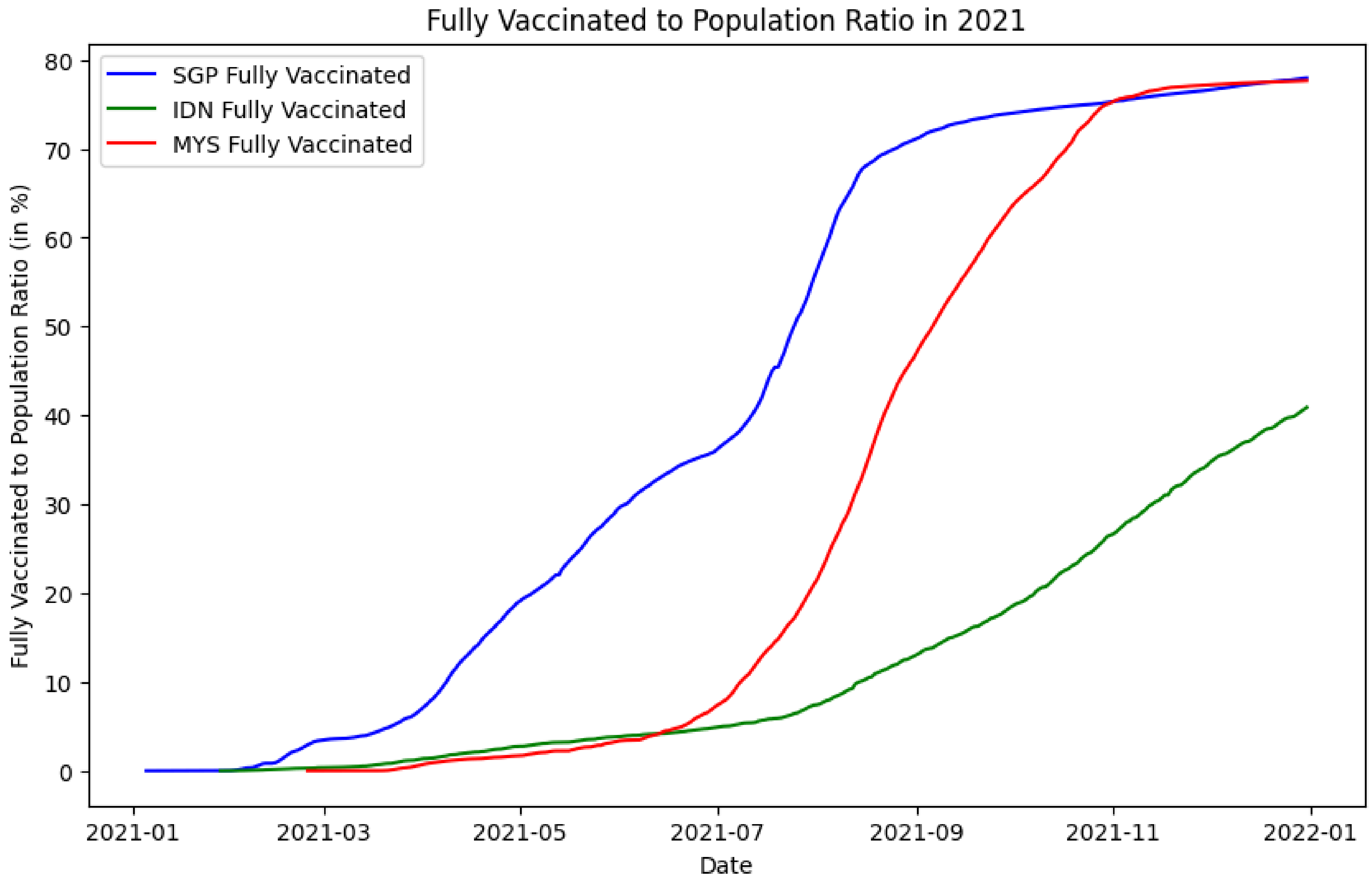
	ratio
mean	39.559622
std	32.605549

```
✓ 0s [97] # Aggregating Singaporean vaccinated to population ratio in 2021
agg_sgp = sgp_df_2021.agg({"ratio": ["mean", "std"]})
agg_sgp
```

	ratio
mean	48.558395
std	30.706552

5. What is the trend of vaccinations due to COVID-19 in 2021?

Fully Vaccinated to Population



Indonesia

Mean: 12.348227

Standard deviation: 12.453458

Malaysia

Mean: 32.966865

Standard deviation: 31.731440

Singapore

Mean: 41.274693

Standard deviation: 30.287835

```python
[100] # Aggregating Indonesian fully vaccinated to population ratio in 2021
      agg_idn = idn_df_2021.agg({"Fully_Vaccinated_to_Population": ["mean", "std"]})
      agg_idn
```

	Fully_Vaccinated_to_Population
mean	12.348227
std	12.453458

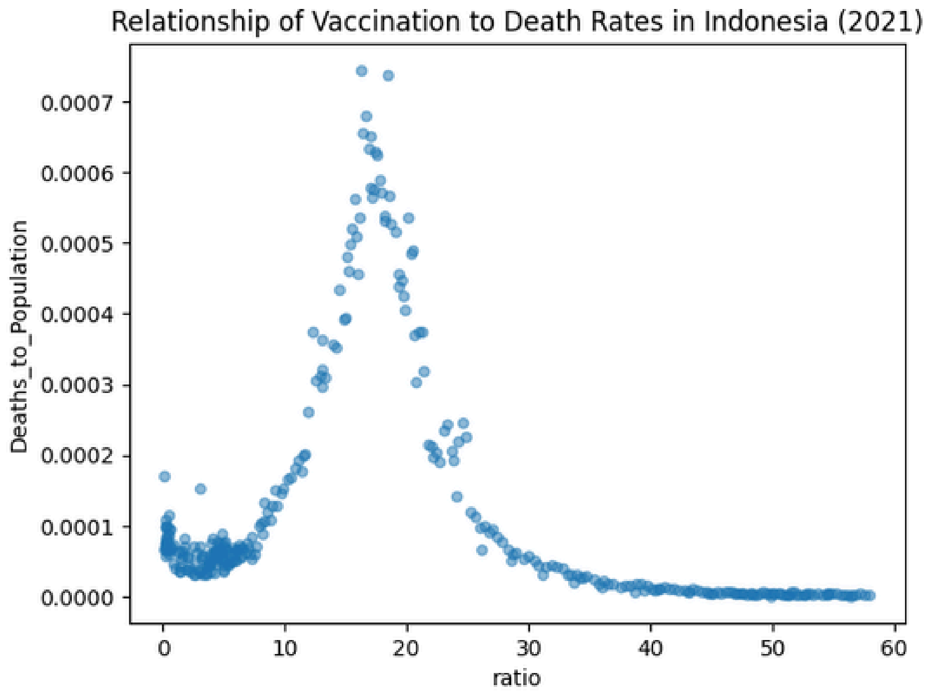
```python
[101] # Aggregating Malaysian fully vaccinated to population ratio in 2021
      agg_mys = mys_df_2021.agg({"Fully_Vaccinated_to_Population": ["mean", "std"]})
      agg_mys
```

	Fully_Vaccinated_to_Population
mean	32.966865
std	31.731440

```python
# Aggregating Singaporean fully vaccinated to population ratio in 2021
agg_sgp = sgp_df_2021.agg({"Fully_Vaccinated_to_Population": ["mean", "std"]})
agg_sgp
```

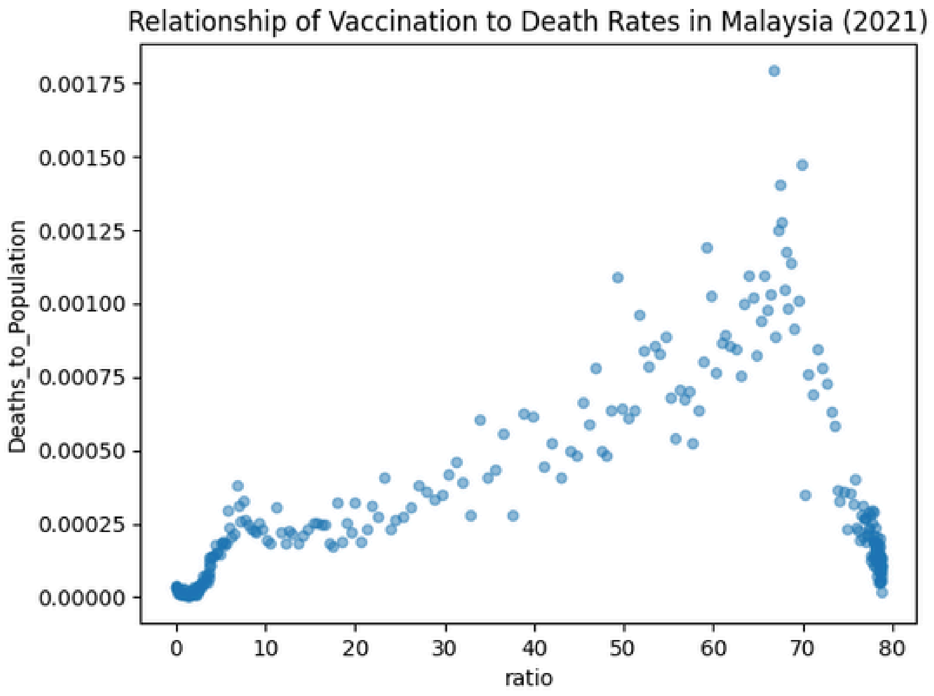
	Fully_Vaccinated_to_Population
mean	41.274693
std	30.287835

6. What is the correlation between vaccination rates and deaths due to COVID-19 in 2021?



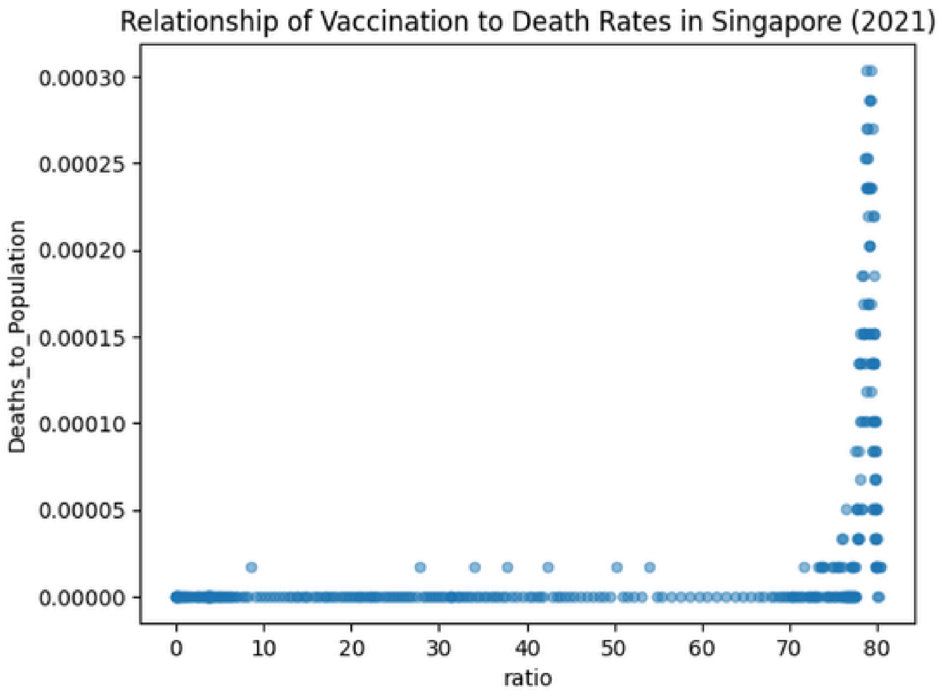
Indonesia

Correlation = -0.21
(Weak Negative)



Malaysia

Correlation = 0.39
(Weak Positive)



Singapore

Correlation = 0.51
(Moderate Positive)

```
[104] # Correlation between vaccination rate to ratio of new deaths to population in Indonesia (2021)
      ratio_deathtopop_idn = idn_df_2021[['ratio', 'Deaths_to_Population']]
      ratio_deathtopop_idn.corr()
```

	ratio	Deaths_to_Population
ratio	1.000000	-0.213342
Deaths_to_Population	-0.213342	1.000000

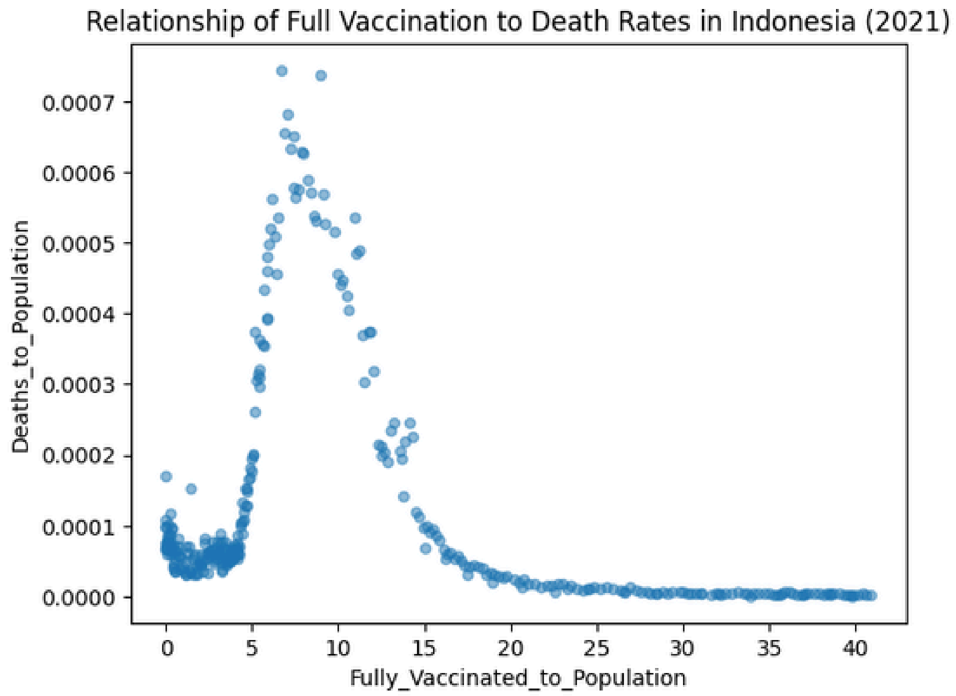
```
[105] # Correlation between vaccination rate to ratio of new deaths to population in Malaysia (2021)
      ratio_deathtopop_mys = mys_df_2021[['ratio', 'Deaths_to_Population']]
      ratio_deathtopop_mys.corr()
```

	ratio	Deaths_to_Population
ratio	1.000000	0.389743
Deaths_to_Population	0.389743	1.000000

```
[106] # Correlation between vaccination rate to ratio of new deaths to population in Singapore (2021)
      ratio_deathtopop_sgp = sgp_df_2021[['ratio', 'Deaths_to_Population']]
      ratio_deathtopop_sgp.corr()
```

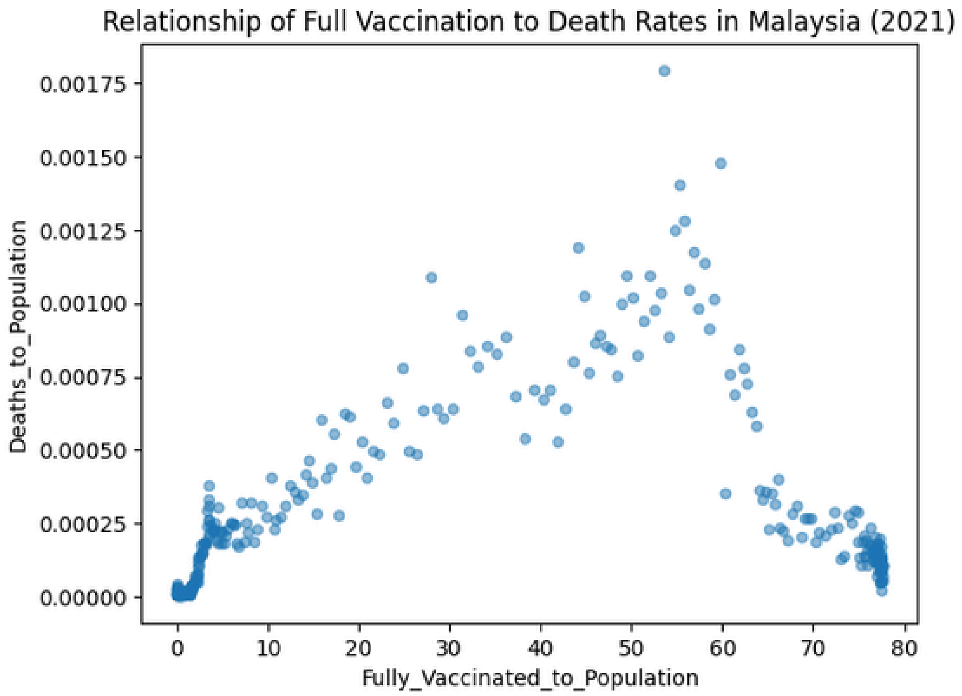
	ratio	Deaths_to_Population
ratio	1.000000	0.507243
Deaths_to_Population	0.507243	1.000000

7. What is the correlation between fully vaccinated rates and deaths due to COVID-19 in 2021?



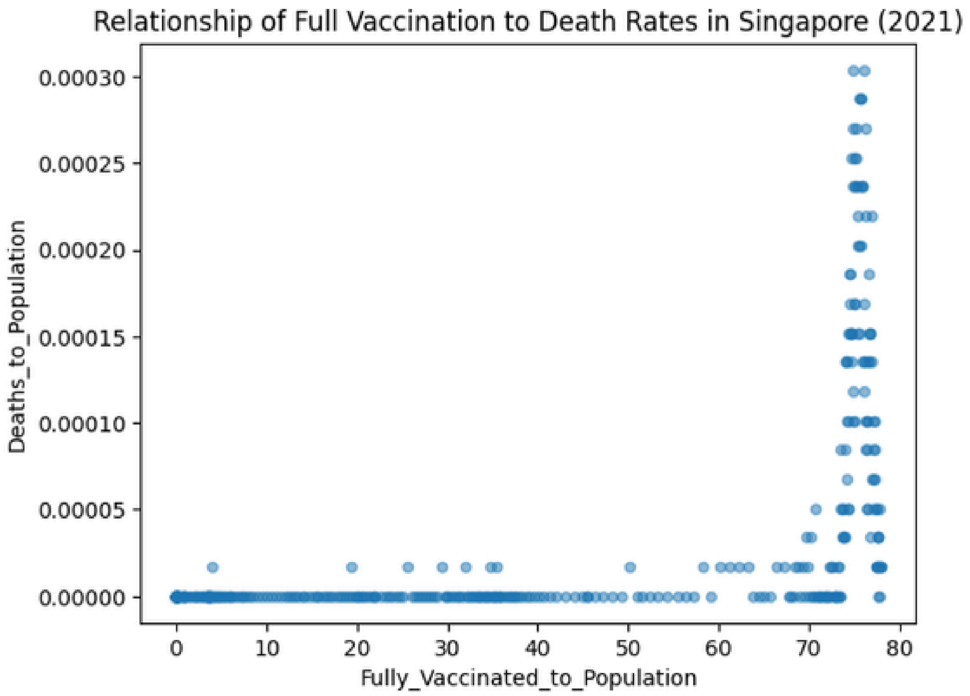
Indonesia

Correlation = -0.29
(Weak Negative)



Malaysia

Correlation = 0.24
(Weak Positive)



Singapore

Correlation = 0.58
(Moderate Positive)

```python
[110] # Correlation between full vaccination rate to ratio of new deaths to population in Indonesia (2021)
ratio_deathtopop_idn = idn_df_2021[['Fully_Vaccinated_to_Population', 'Deaths_to_Population']]
ratio_deathtopop_idn.corr()
```

	Fully_Vaccinated_to_Population	Deaths_to_Population
Fully_Vaccinated_to_Population	1.000000	-0.291724
Deaths_to_Population	-0.291724	1.000000

```python
[111] # Correlation between full vaccination rate to ratio of new deaths to population in Malaysia (2021)
ratio_deathtopop_mys = mys_df_2021[['Fully_Vaccinated_to_Population', 'Deaths_to_Population']]
ratio_deathtopop_mys.corr()
```

	Fully_Vaccinated_to_Population	Deaths_to_Population
Fully_Vaccinated_to_Population	1.000000	0.243903
Deaths_to_Population	0.243903	1.000000

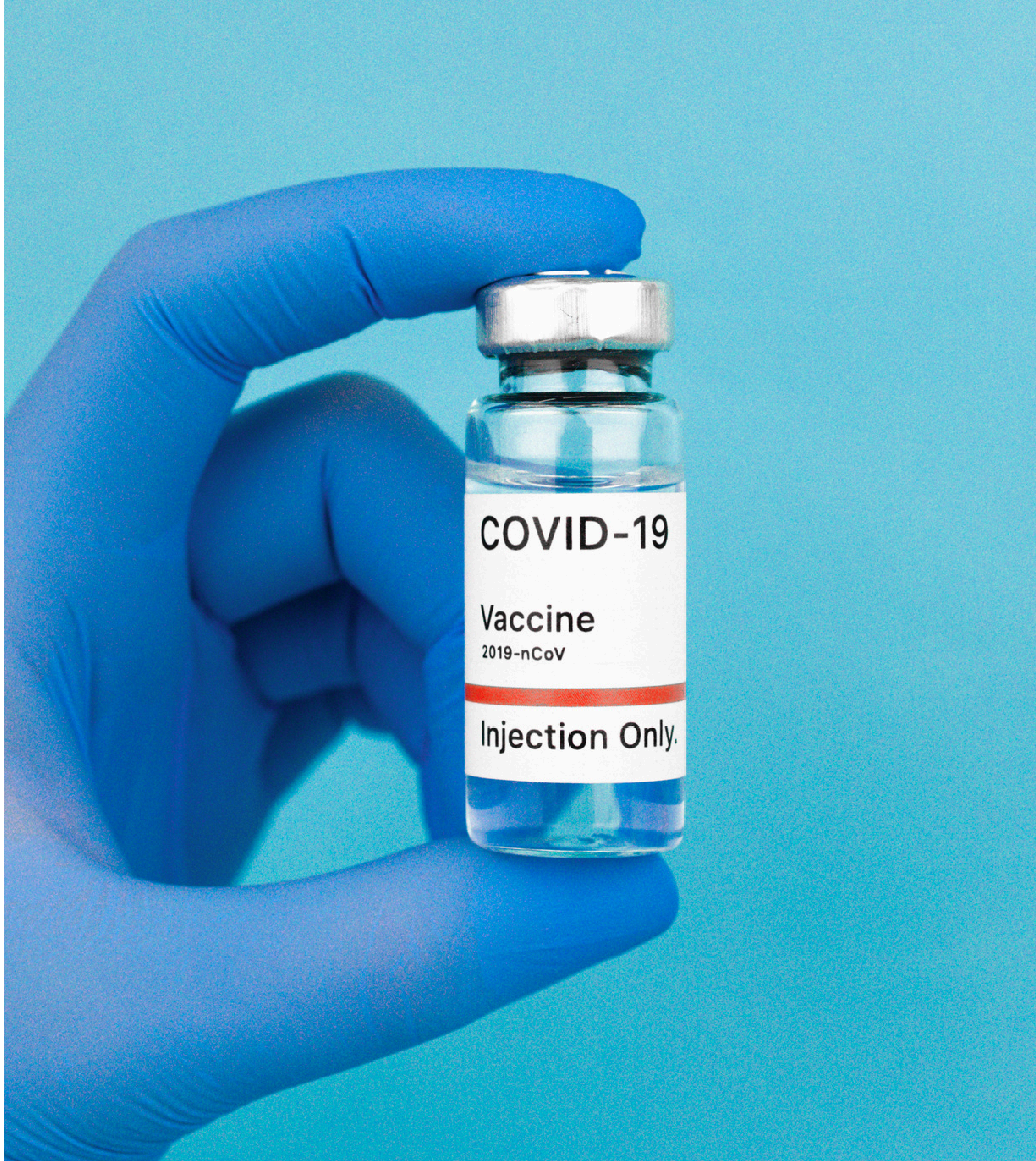
```python
[112] # Correlation between full vaccination rate to ratio of new deaths to population in Singapore (2021)
ratio_deathtopop_sgp = sgp_df_2021[['Fully_Vaccinated_to_Population', 'Deaths_to_Population']]
ratio_deathtopop_sgp.corr()
```

	Fully_Vaccinated_to_Population	Deaths_to_Population
Fully_Vaccinated_to_Population	1.000000	0.576939
Deaths_to_Population	0.576939	1.000000

04.
Research Question

Research Question

How much does the vaccination rate affect the death rate due to COVID-19 in Southeast Asian countries?



CONCLUSION

KEY FINDINGS

- Positively Skewed Distributions: COVID-19 deaths in 2021 in Indonesia, Malaysia, and Singapore show varied mortality rates with a positive skew.
- Threshold Effect: The relationship between vaccination rates and new deaths displays a distinct threshold, shifting from positive to negative correlation.

CHALLENGES

- Inconsistent Data: Data gaps in Indonesia for January-March 2022 hindered a full-year analysis.
- Non-Linear Patterns: Complex non-linear patterns, like the vaccination rate and new deaths correlation, pose analytical challenges.

Thank you!

This conclude our presentation.

CSMODEL S12

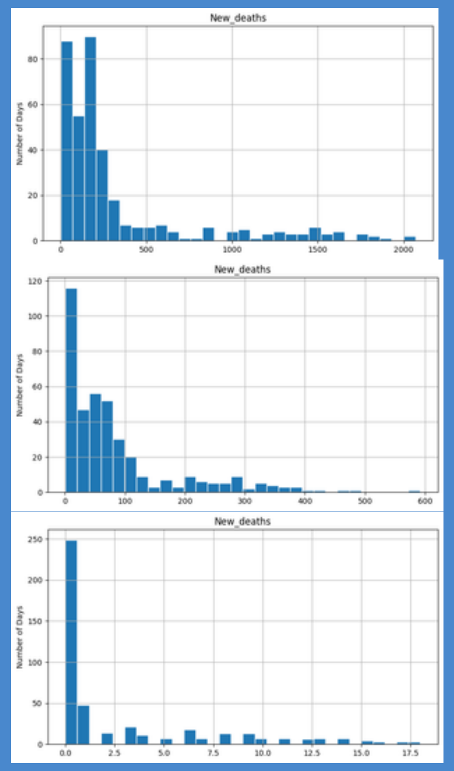
Investigating COVID Vaccination vs Mortality Data

Members: Ong, Teves, Yu

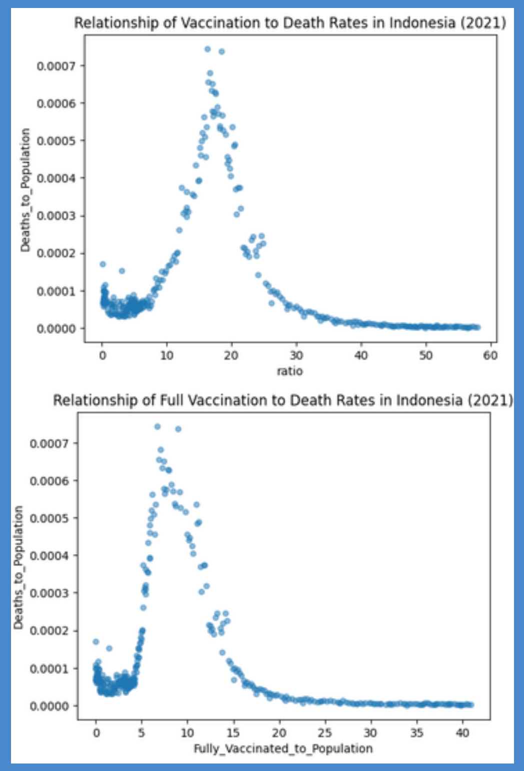
COVID vaccination vs. mortality dataset

- Collected from:
    1. COVID-19 World Vaccination Progress
    2. WHO COVID-19 Global Data
    3. 2021 World Population
- 32,911 observations, each for **Country**, **Date**, **Vaccination** information (Partial and Complete), **New Deaths** reported
- For this research, scope limited to **2021** data of **Indonesia**, **Malaysia**, and **Singapore**

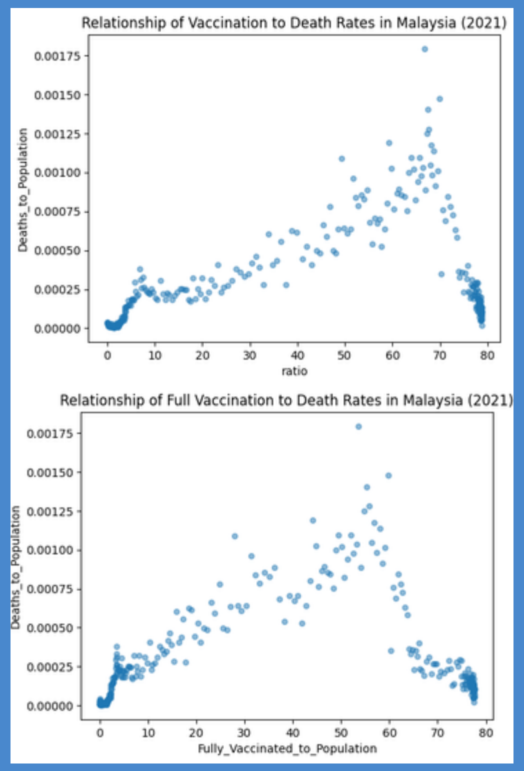
EDA & Research Question



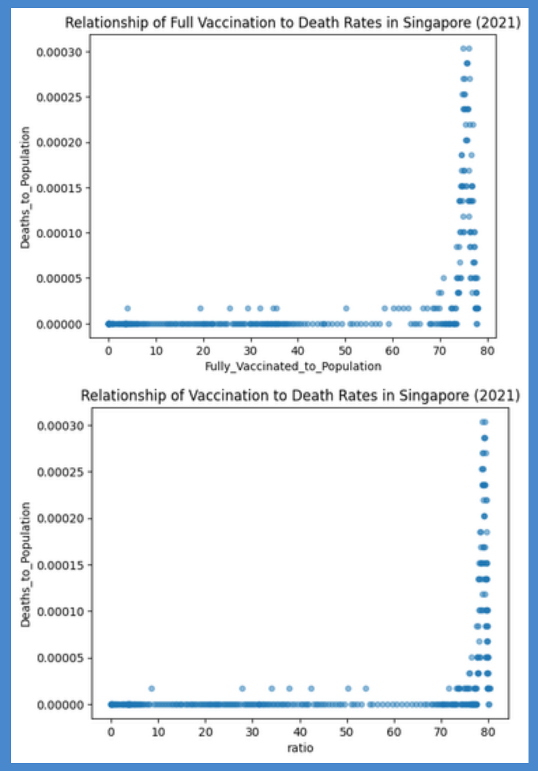
Distribution of new death rate in IDN, MLY, SGP



Correlation of Vaccination to Death in Indonesia



Correlation of Vaccination to Death in Malaysia



Correlation of Vaccination to Death in Singapore

How much does the COVID-19 death rate affect the COVID-19 vaccination rate in Southeast Asian countries?

Data Modeling

Preprocessing: Handle Missing Data

Use pandas **dataframe.isnull().sum().sum()**: returns the total count of missing values across the dataset

```python
# Print the total number of null values for each country's dataframe
print("Singapore: ", sgp_df_2021.isnull().sum().sum())
print("Indonesia: ", idn_df_2021.isnull().sum().sum())
print("Malaysia: ", mys_df_2021.isnull().sum().sum())
```

✓ 0.0s Python

```
Singapore:  0
Indonesia:  0
Malaysia:  0
```

Preprocessing: Outlier Detection

```
# Set threshold
threshold = 3

# Calculate z-scores for ratio, fully vaccinated to population, and deaths to population
z_scores_ratio = zscore(sgp_df_2021['ratio'])
z_scores_fullvacc = zscore(sgp_df_2021['Fully_Vaccinated_to_Population'])
z_scores_mortality = zscore(sgp_df_2021['Deaths_to_Population'])

# Identifying outliers
outliers_ratio = (abs(z_scores_ratio) > threshold)
outliers_fullvacc = (abs(z_scores_fullvacc) > threshold)
outliers_mortality = (abs(z_scores_mortality) > threshold)

# Create a new column to mark outliers
sgp_df_2021['outlier_ratio'] = outliers_ratio
sgp_df_2021['outlier_fullvacc'] = outliers_fullvacc
sgp_df_2021['outlier_mortality'] = outliers_mortality

# Drop rows with outliers
sgp_df_2021_processed = sgp_df_2021.drop(sgp_df_2021[sgp_df_2021['outlier_ratio'] == True].index)
sgp_df_2021_processed = sgp_df_2021.drop(sgp_df_2021[sgp_df_2021['outlier_fullvacc'] == True].index)
sgp_df_2021_processed = sgp_df_2021.drop(sgp_df_2021[sgp_df_2021['outlier_mortality'] == True].index)
```

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Data Modeling: Association Rule Mining

- **t**he most effective method for revealing associations and dependencies among the options available
- identifies potential **causal relationship**
- more appropriate compared to clustering and recommender systems

Preprocessing: Binning

Heuristic to determine the number of bins:

bins = √ (total data points)

```
# Calculate the square root of the number of data points
sgp_bins = int(np.sqrt(len(sgp_df_2021_processed)))
print(sgp_bins)
```

✓ 0.0s Python

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

ratio:

```
# Use pd.cut to discretize the data into bins
sgp_discretized_r = pd.cut(sgp_df_2021_processed['ratio'], bins=sgp_bins, labels=False)
sgp_discretized_r

sgp_df_2021_processed['ratio_Binned'] = sgp_discretized_r
sgp_df_2021_processed['ratio_Binned'] = sgp_df_2021_processed['ratio_Binned'].astype('category')
```

Fully_Vaccinated_to_Population:

```
# Use pd.cut to discretize the data into bins
sgp_discretized_fvp = pd.cut(sgp_df_2021_processed['Fully_Vaccinated_to_Population'], bins=sgp_bins,
sgp_discretized_fvp

sgp_df_2021_processed['Fully_Vaccinated_to_Population_Binned'] = sgp_discretized_fvp
sgp_df_2021_processed['Fully_Vaccinated_to_Population_Binned'] = sgp_df_2021_processed['Fully_Vaccina
```

Deaths_to_Population:

```
# Use pd.cut to discretize the data into bins
sgp_discretized_dtp = pd.cut(sgp_df_2021_processed['Deaths_to_Population'], bins=sgp_bins, labels=Fal
sgp_discretized_dtp

sgp_df_2021_processed['Deaths_to_Population_Binned'] = sgp_discretized_dtp
sgp_df_2021_processed['Deaths_to_Population_Binned'] = sgp_df_2021_processed['Deaths_to_Population_Bi
```

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Association Rule Mining

Support Threshold: 0.1
Confidence Threshold: 0.8

Convert categorical columns to numeric labels

```
# Convert categorical columns to numeric labels
sgp_df_2021_processed['ratio_Binned'] = sgp_df_2021_processed['ratio_Binned'].cat.codes
sgp_df_2021_processed['Fully_Vaccinated_to_Population_Binned'] = sgp_df_2021_processed['Fully_Vaccinated_to_Population_Binned'].cat.codes
sgp_df_2021_processed['Deaths_to_Population_Binned'] = sgp_df_2021_processed['Deaths_to_Population_Binned'].cat.codes
```

Dataframe grouped by binned ratio, Fully_Vaccinated_to_Population & Deaths_to_Population

```
sgp_transactions = sgp_df_2021_processed.groupby(['ratio_Binned', 'Fully_Vaccinated_to_Population_Binned', 'Deaths_to_Population_Binned']).size().reset_index(name='count')
print(sgp_transactions)
```

✓ 0.0s

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Extract frequent itemsets from the dataframe

```
sgp_frequent_itemsets = rule_miner.get_frequent_itemsets(sgp_transactions)
print(sgp_frequent_itemsets)
```

✓ 0.0s Python

```
[['Deaths_to_Population_Binned', 'Fully_Vaccinated_to_Population_Binned', 'count', 'ratio_Binned']]
```

Derive the association rules from the transaction data

```
sgp_association_rules = rule_miner.get_association_rules(sgp_transactions)
print(sgp_association_rules)
```

✓ 0.0s Python

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Association Rules

Common association rules in Indonesia, Malaysia, and Singapore:

1. {Deaths_to_Population_Binned, Fully_Vaccinated_to_Population_Binned, count} → {ratio_Binned}
2. {Deaths_to_Population_Binned, Fully_Vaccinated_to_Population_Binned, ratio_Binned} → {count}
3. {Deaths_to_Population_Binned, count, ratio_Binned} → {Fully_Vaccinated_to_Population_Binned}
4. {Deaths_to_Population_Binned} → {Fully_Vaccinated_to_Population_Binned, count, ratio_Binned}

Association Rules

Malaysia-specific association rules

1. {**ratio_Binned}**
   → {Deaths_to_Population_Binned, Fully_Vaccinated_to_Population_Binned, count}

2. **{count}**
   → {Deaths_to_Population_Binned, Fully_Vaccinated_to_Population_Binned, ratio_Binned}

3. **{Fully_Vaccinated_to_Population_Binned}**
   → {Deaths_to_Population_Binned, count, ratio_Binned}

4. **{Fully_Vaccinated_to_Population_Binned, count, ratio_Binned}**
   → {Deaths_to_Population_Binned}

Statistical Inference

Obtaining Samples

- Slovin's Formula: $n = \dfrac{N}{1+Ne^2}$

```python
N = len(sgp_df_2021_processed)
e = 0.05
n = int(N/(1+N*(e**2)))
```

- Using 5% margin of error, ideal sample size: Indonesia: 178
  Malaysia: 173
  Singapore: 186

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Obtaining Samples

```
sgp_df_2021_processed_sample = sgp_df_2021_processed.sample(n, random_state=8)
sgp_df_2021_processed_sample.head()
```

```
agg = sgp_df_2021_processed_sample.agg({"ratio_Binned": ["mean", "median", "std"]})

sgp_ratio_sample_mean = agg.loc["mean"][0]
sgp_ratio_sample_median = agg.loc["median"][0]
sgp_ratio_sample_std = agg.loc["std"][0]

print('Ratio of Vaccinated:')
print('Sample Mean: {:.2f}'.format(sgp_ratio_sample_mean))
print('Sample Median: {:.2f}'.format(sgp_ratio_sample_median))
print('Sample Standard Deviation: {:.2f}'.format(sgp_ratio_sample_std))
```

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Obtaining Samples

	Sample Mean	Median	Standard Deviation
IDN Vax Rate	5.73	4.00	5.39
Full Vax Rate	4.97	2.00	5.23
New Deaths Rate	2.51	1.00	3.89
MYS Vax Rate	8.12	9.00	6.94
Full Vax Rate	6.88	4.00	6.82
New Deaths Rate	3.47	2.00	4.11
SGP Vax Rate	10.16	11.50	6.79
Full Vax Rate	8.82	8.00	6.89
New Deaths Rate	1.69	0.00	3.53

Confidence Interval

- Formula: $\bar{x} \pm z^* \frac{s}{\sqrt{n}}$

```
# Find the CRITICAL VALUE for 95% confidence interval
z_star_95 = norm.ppf(0.975)

#Compute for the MARGIN OF ERROR
print("Margins of Error:\n--------------------")

print("Singapore")
sgp_ratio_moe = z_star_95*((sgp_ratio_sample_std)/n**(1/2))
print('Vaccination Rate: {:.2f}'.format(sgp_ratio_moe))
sgp_fully_vaccinated_moe = z_star_95*((sgp_fully_vaccinated_sample_std)/n**(1/2))
print('Full Vaccination Rate: {:.2f}'.format(sgp_fully_vaccinated_moe))
sgp_deaths_moe = z_star_95*((sgp_deaths_sample_std)/n**(1/2))
print('Deaths: {:.2f}\n'.format(sgp_deaths_moe))
```

- Z-score used: 1.96 (95% confidence level)

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Confidence Interval

```python
# SINGAPORE

sgp_ratio_min = sgp_ratio_sample_mean - sgp_ratio_moe
sgp_ratio_max = sgp_ratio_sample_mean + sgp_ratio_moe
print('Singapore Sample Mean: {:.2f}'.format(sgp_ratio_sample_mean))
print('Singapore Ratio Range: ({:.2f}'.format(sgp_ratio_min) + ", " + '{:.2f})\n'.format(sgp_ratio_max))

sgp_fully_vaccinated_min = sgp_fully_vaccinated_sample_mean - sgp_fully_vaccinated_moe
sgp_fully_vaccinated_max = sgp_fully_vaccinated_sample_mean + sgp_fully_vaccinated_moe
print('Singapore Sample Fully Vaccinated Mean: {:.2f}'.format(sgp_fully_vaccinated_sample_mean))
print('Singapore Fully Vaccinated Range: ({:.2f}'.format(sgp_fully_vaccinated_min) + ", " + '{:.2f})\n'.format(sgp_fully_vaccinated_max))

sgp_death_min = sgp_deaths_sample_mean - sgp_deaths_moe
sgp_death_max = sgp_deaths_sample_mean + sgp_deaths_moe
print('Singapore Sample Deaths Mean: {:.2f}'.format(sgp_deaths_sample_mean))
print('Singapore Deaths Range: ({:.2f}'.format(sgp_death_min) + ", " + '{:.2f})\n\n'.format(sgp_death_max))
```

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Confidence Interval

	Sample Mean	Margin of Error	Confidence Interval
IDN Vax Rate	5.73	0.80	(4.93, 6.53)
Full Vax Rate	4.97	0.78	(4.19, 5.75)
New Deaths Rate	2.51	0.58	(1.93, 3.08)
MYS Vax Rate	8.12	1.03	(7.09, 9.16)
Full Vax Rate	6.88	1.02	(5.87, 7.90)
New Deaths Rate	3.47	0.61	(2.86, 4.08)
SGP Vax Rate	10.16	1.01	(9.15, 11.17)
Full Vax Rate	8.82	1.03	(7.80, 9.85)
New Deaths Rate	1.69	0.53	(1.17, 2.22)

Hypothesis Testing

H_0 (Null Hypothesis): COVID-19 death rates have no notable effect on COVID-19 vaccination rates in Southeast Asian Countries. ($\mu = 0$)

H_A (Alt. Hypothesis): COVID-19 death rates have a notable effect on COVID-19 vaccination rates in Southeast Asian Countries. ($\mu \neq 0$)

```
sgp_low_deaths = sgp_df_2021_processed[sgp_df_2021_processed['Deaths_to_Population_Binned'] < sgp_deaths_sample_mean]
sgp_high_deaths = sgp_df_2021_processed[sgp_df_2021_processed['Deaths_to_Population_Binned'] >= sgp_deaths_sample_mean]
```

Note: The code snippet pertains to Singapore, but the same process is applied to Indonesia and Malaysia.

Hypothesis Testing

	Sample Mean	Median	Standard Deviation
IDN Vax Rate (Low D)	5.89	2.00	6.05
Vax Rate (High D)	4.29	4.00	1.68
Full Vax Rate (Low D)	5.29	1.00	5.77
Full Vax Rate (High D)	2.92	3.00	1.49
MYS Vax Rate (Low D)	6.69	2.00	7.50
Vax Rate (High D)	10.71	12.00	4.05
Full Vax Rate (Low D)	6.23	0.00	7.59
Full Vax Rate (High D)	7.63	8.00	4.35
SGP Vax Rate (Low D)	8.01	7.00	6.48
Vax Rate (High D)	17.00	17.00	0.00
Full Vax Rate (Low D)	6.28	5.00	5.93
Full Vax Rate (High D)	16.94	17.00	0.24

Hypothesis Testing

	Sample Mean	Margin of Error	Confidence Interval
IDN Vax Rate (Low D)	5.89	0.90	(4.99, 6.80)
Vax Rate (High D)	4.29	0.25	(4.03, 4.54)
Full Vax Rate (Low D)	5.29	0.86	(4.43, 6.15)
Full Vax Rate (High D)	2.92	0.22	(2.70, 3.14)
MYS Vax Rate (Low D)	6.69	1.12	(5.58, 7.81)
Vax Rate (High D)	10.71	0.60	(10.10, 11.31)
Full Vax Rate (Low D)	6.23	1.13	(5.10, 7.36)
Full Vax Rate (High D)	7.63	0.65	(6.98, 8.28)
SGP Vax Rate (Low D)	8.01	0.96	(7.04, 8.97)
Vax Rate (High D)	17.00	0.00	(17.00, 17.00)
Full Vax Rate (Low D)	6.28	0.88	(5.40, 7.17)
Full Vax Rate (High D)	16.94	0.04	(16.90, 16.98)

Hypothesis Testing

```
sgp_diff = sgp_high_deaths_ratio_sample_mean - sgp_low_deaths_ratio_sample_mean
print('{:.2f}'.format(sgp_diff))

idn_diff = idn_high_deaths_ratio_sample_mean - idn_low_deaths_ratio_sample_mean
print('{:.2f}'.format(idn_diff))

mys_diff = mys_high_deaths_ratio_sample_mean - mys_low_deaths_ratio_sample_mean
print('{:.2f}'.format(mys_diff))
```

```
column_name = 'ratio_Binned'

sgp_low_deaths_values = sgp_low_deaths[column_name]
sgp_high_deaths_values = sgp_high_deaths[column_name]

t_statistic, p_value = ttest_ind(sgp_low_deaths_values, sgp_high_deaths_values, equal_var=False)

print("T-Statistic:", t_statistic)
print("P-Value:", p_value)

alpha = 0.05
if p_value < alpha:
    print("Reject the null hypothesis. There is a significant difference.")
else:
    print("Fail to reject the null hypothesis. There is no significant difference.")
```

Hypothesis Testing

	μ_D	T-Statistic	P-Value	Reject or Accept
IDN Vax Rate	-1.61	3.73	2.30×10^{-4}	Reject
Full Vax Rate	-2.37	5.83	1.32×10^{-8}	Reject
MYS Vax Rate	4.01	-6.05	4.36×10^{-9}	Reject
Full Vax Rate	1.40	-2.04	4.25×10^{-2}	Reject
SGP Vax Rate	8.99	-22.69	3.74×10^{-64}	Reject
Full Vax Rate	10.66	-29.27	1.45×10^{-85}	Reject

Insights and Conclusion

Insights

1. Strong association between vaccination, full vaccination, and death rates due to a high frequency of association.
2. Potential relationships exist between partial and full vaccination rates with death due to heavy association between them for Indonesia, Malaysia, and Singapore.
3. Fluctuations in vaccination trend expected in Malaysia and Indonesia due to the sample mean and median being significantly different.
4. All sample means can represent the true population mean 95% of the time.
5. High rate of deaths could be a factor that made people vaccinate.

Conclusion

How much does the COVID-19 death rate affect the COVID-19 vaccination rate in Southeast Asian countries?

Answer:

The death rate has a significant effect in the vaccination rates of Southeast Asian countries, as evidenced by the frequent association of the death rate to the partial and full vaccination rates and the differences in vaccinations on days with low and high deaths.