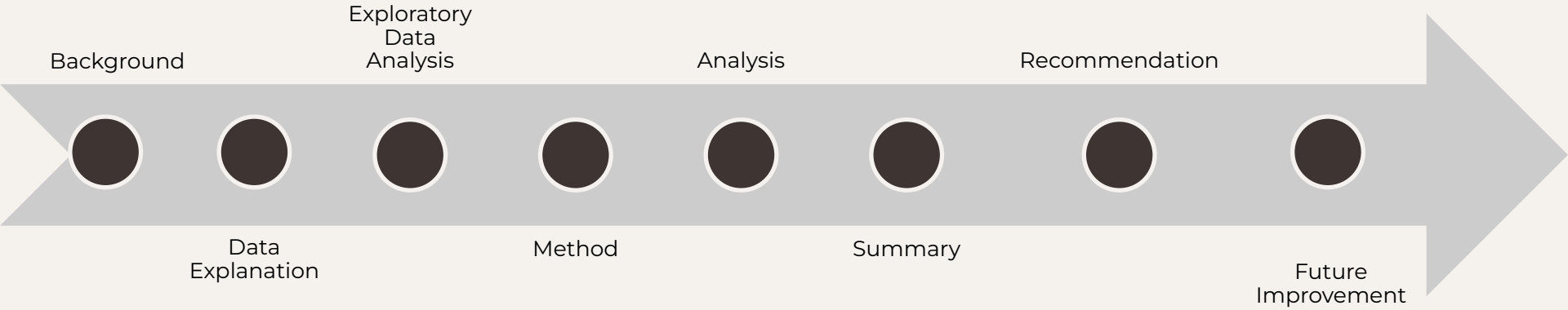
Machine Learning & Data Analysis approach to Predict Vulnerability of Dengue Outbreak due to Climate Change

TERRAVERSE Team

Alfredo Marciano Zola (alfredomarciano98@gmail.com)
Anugrah Muzakki Puar (anugrahmuzakkipuar@gmail.com)
Muhammad Insan Aprilian (insanaprilian50@gmail.com)
Rafif Abdussalam (rafifabdussalam2@gmail.com)
Rahman Fitra Perdana (rahmanfitra.p@gmail.com)

Table of contents



01

Background

Yearly Dengue Cases

4 Billion ⇨ **400 Million** ⇨ **100 Million** ⇨ **40 Thousand**

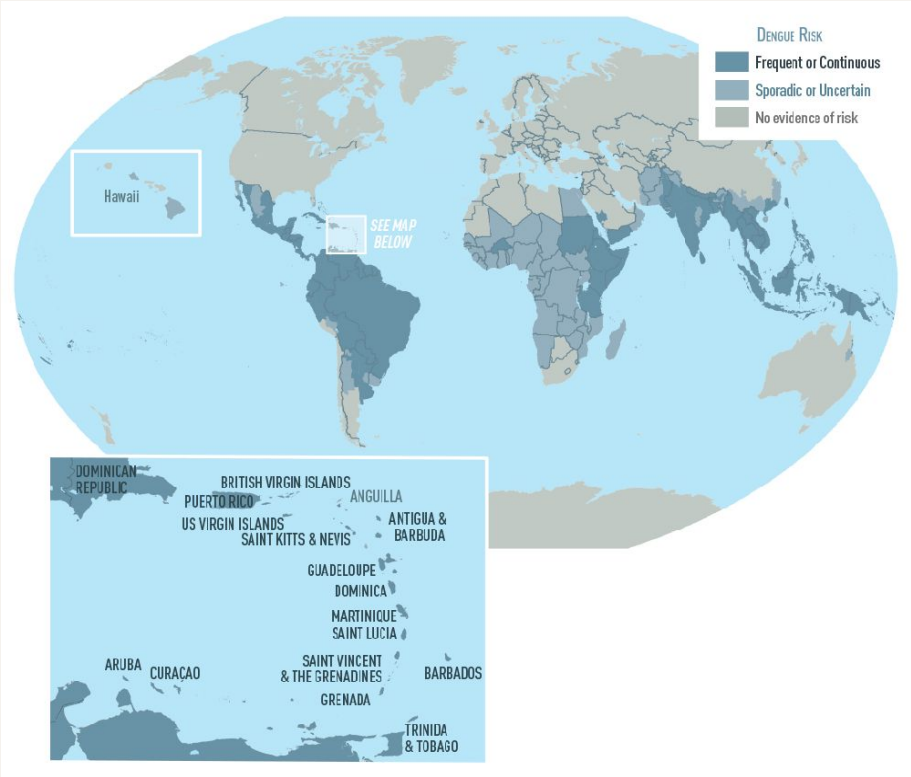
- Live in areas with a **risk** of dengue

- Get **infected** with dengue

- **Sick** from infection

- **Die** from Dengue

Dengue Around the World



- **Frequent/Continuous**:

Either frequent or outbreaks occur or transmission is ongoing.

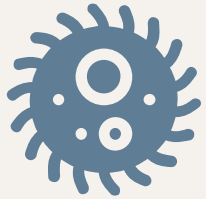
- **Sporadic/uncertain**:

Risk varies and is unpredictable and that country-level data is not available

Dengue Influences Factor

Geographic and **seasonal** distribution of vector

Socioeconomic and cultural factors



Human **response** to disease risk

Pest control

Access to **health care**

Climate change

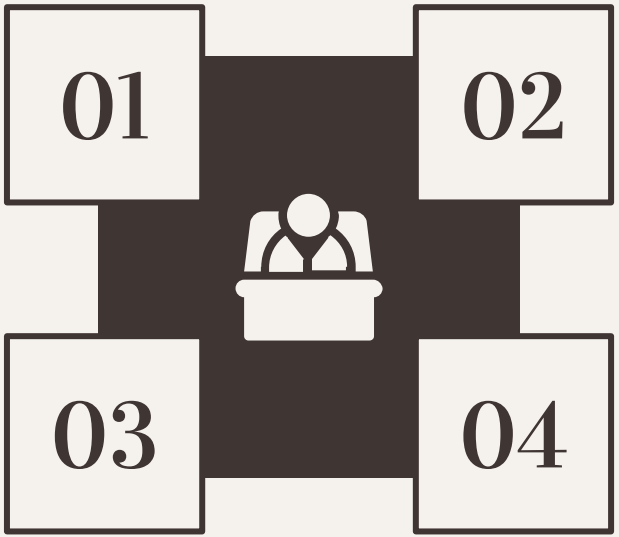
Climate Change Effect on Dengue

Transmission Season

Transmission season become **longer** than usual

Spatial Range

Vector can live in **higher altitude** as it's warmer



Physiological Traits

Affected by **rising temperatures**

Precipitation

Increasing the breeding sites

Objective

Understanding the **vulnerability** of countries facing **dengue** outbreak due to **climate change** for a **sustainable** society

Challenges

Data Coverage

Our dataset does not cover all country

Data Range

Some of the time frame in our dataset is **not** in the same range

Data Granularity

Granularity for all features are not the same on each country

02
Data Explanation

Dataset

Health Facility and Security:

1. Current Health Expenditure (% of GDP)
2. Hospital Beds per 1000 People
3. Nurse and Midwives per 1000 People
4. UHC Service Coverage Index
5. Global Health Security Index
6. Early Detection & Reporting for Epidemic of Potential International Concern
7. Rapid Response to and Mitigation of the Spread of an Epidemic
8. Sufficient and Robust Health Sector to Treat the Sick and Protect Health Workers
9. Commitments to Improving National Capacity, Financing, and Adherence to Norms
10. Overall Risk Environment and Country Vulnerability to Biological Threats

Climate and Geographical conditions:

1. Land Area (km)
2. Length of the Transmission Seasons (LTS)

Socio economic condition:

1. Life Expectancy (Year)
2. Poverty
3. Unemployment Rate
4. GINI Index
5. GDP per Capita (USD)
6. Population
7. Number of Migrant People

Historical case:

1. History of Dengue Case

Data Source

THE WORLD BANK

World Health Organization

https://worldpopulationreview.com/

GHS INDEX
GLOBAL HEALTH SECURITY INDEX

OSF

https://osf.io/hpaey/

Data for Model Development

THE WORLD BANK

GDP per Capita, GINI index

World Health Organization

UHC Service Coverage, Hospital and nurse

GHS INDEX
GLOBAL HEALTH SECURITY INDEX

Epidemic Detection and Mitigation

OSF

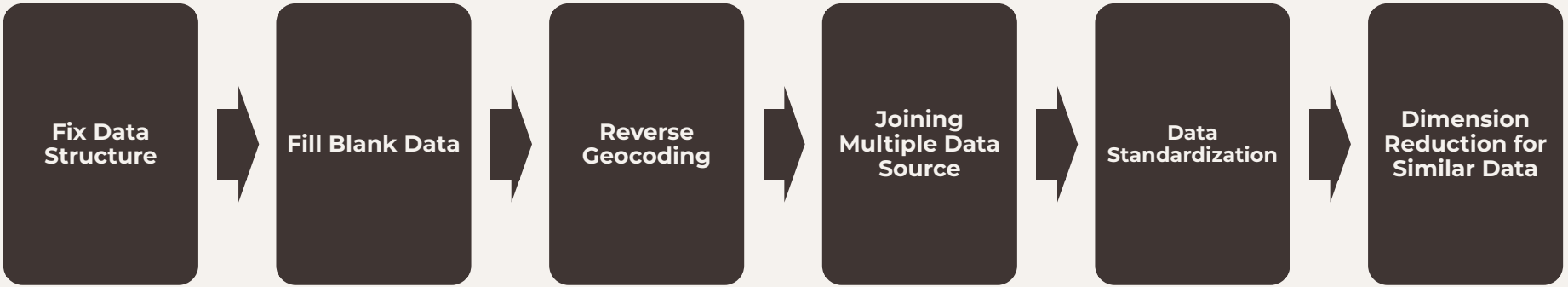
LTS, History Dengue Case

Data Merge and Cleaning

Aggregation for Model Training

Country	country_coc	Current health	Hospital beds (pe	Nurses and midwi	UHC service c	Land Area	LifeExp(years	Poverty	Unemploym	GINI_index	GDP/capi	Populatio	Developed	migration_2
Afghanistan	AFG	13.24220181	0.39	0.4461	37	652,860	62.879	54.5	13.28	29.4	2390	41,128,77	0	54000
Africa Eastern and S	AFE	6.266385633	0.9118708018	0.9156768626	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Africa Western and (	AFW	3.366099661	1.474775782	1.149448148	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Albania	ALB	5.23207188	2.89	5.088	62	27,400	76.833	14.3	11.82	33.2	14218	2,842,321	0	24000
Algeria	DZA	6.24337101	1.9	1.5477	75	2,381,740	77.129	5.5	12.7	27.6	11112	44,903,22	0	34000
American Samoa	ASM	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	2000
Andorra	AND	6.71158504	2.5	4.0128	77	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	3000
Angola	AGO	2.53335953	0.8	0.4075	39	1,246,700	61.929	32.3	8.53	51.3	6932	35,588,96	0	128000
Antigua and Barbud	ATG	4.44295979	2.89	4.5171	72	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	2000
Arab World	ARB	5.085541836	1.380434026	2.240588219	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A
Argentina	ARG	9.5100708	4.99	2.5996	73	2,736,690	76.064	35.5	10.9	42.9	20751	45,510,31	0	346000
Armenia	ARM	11.33526421	4.2	4.3683	69	28,470	73.372	26.4	20.9	29.9	13261	2,780,469	0	13000
Aruba	ABW	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	11000
Australia	AUS	9.91078568	3.84	13.2416	87	7,682,300	83.579	12.4	5.11	34.4	51680	26,177,41	1	633000
Austria	AUT	10.43403149	7.27	7.0891	82	82,409	82.412	13.3	6.3	30.8	55218	8,939,617	1	139000
Azerbaijan	AZE	4.03829002	4.82	6.4338	65	82,658	73.488	6	6.58	26.6	14431	10,358,07	0	28000
Bahamas, The	BHS	5.74977493	2.96	4.5669	70	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	7000
Bahrain	BHR	4.01147509	1.74	2.4944	71	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	59000
Bangladesh	BGD	2.48357606	0.79	0.3916	51	130,170	73.698	24.3	5.23	32.4	5307	171,186,3	0	589000
Barbados	BRB	6.29990387	5.97	3.06	74	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	3000
Belarus	BLR	5.86133957	10.83	11.0027	74	202.910	73.246	5	4.74	25.3	20187	9.534.954	0	46000

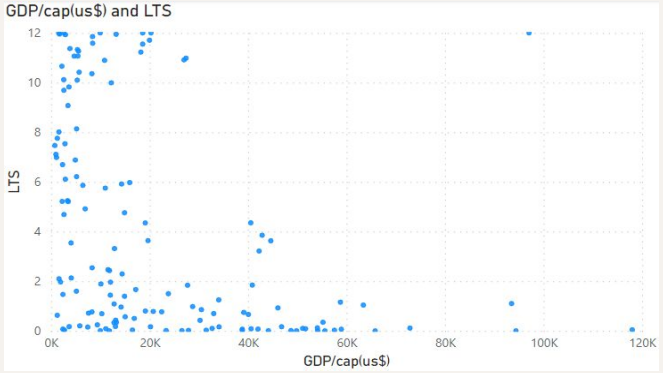
Data Preparation

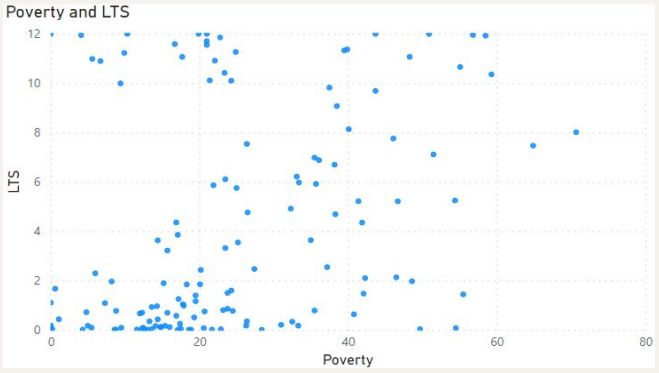


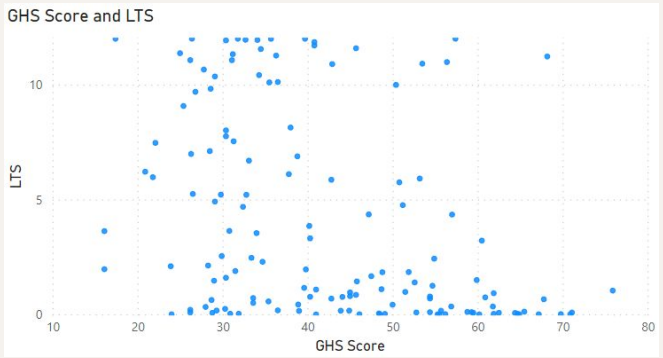
03

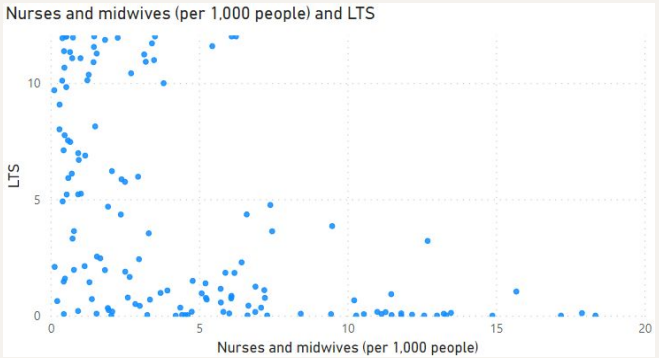
Exploratory Data Analysis

Exploratory Data Analysis









We found that there is **no clear correlation** between **LTS** and some country's aspect like **economy** (GDP/cap), **social** (Poverty), **pandemic readiness**, or **healthcare facilities** (nurses & midwives).

This means high LTS level can be happened in **any** country.

That's why we need to create **unsupervised machine learning** model to get **more information** of vulnerability segmentation

04

Method

Segment Estimation

Development stage:

Iterative Development(LTS, Socio economic, historical case, Health Index)

Feature Scaling

Validation

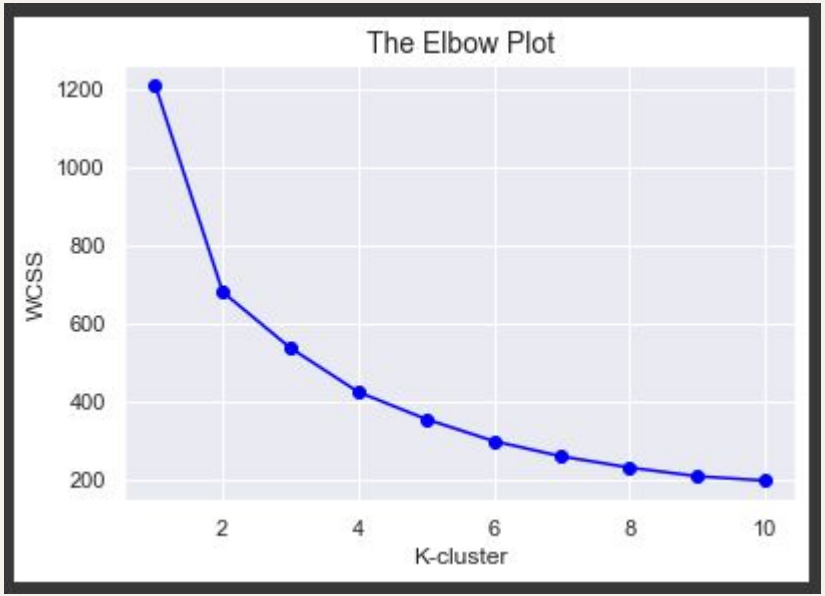
Feature Selection

Machine Learning Model

Model Result

Segment Evaluation

Using K-Means Clustering



1. Find the right # of cluster using **elbow** plot and **WCSS** (Withing-Cluster Sum of Squared) value

2. Use the right # of cluster to create a segmentation based on used data

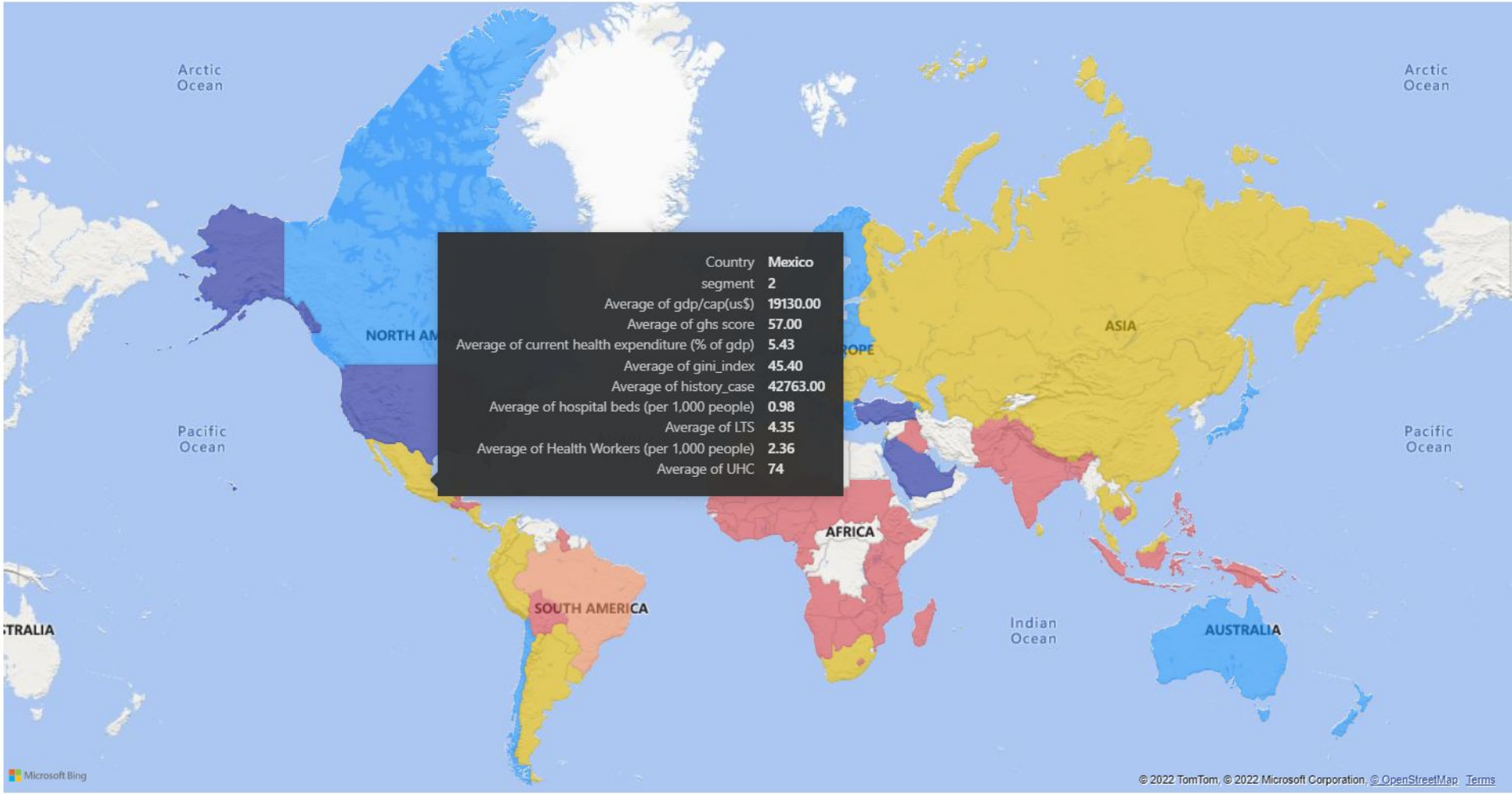
05

Analysis

Segmentation Map and Statistic Info of Each Country

segment 0 1 2 3 4

Access the Power BI dashboard interactively in [this link]



Summary - General Interpretation

Overall, segment **0** has the **lowest** LTS average

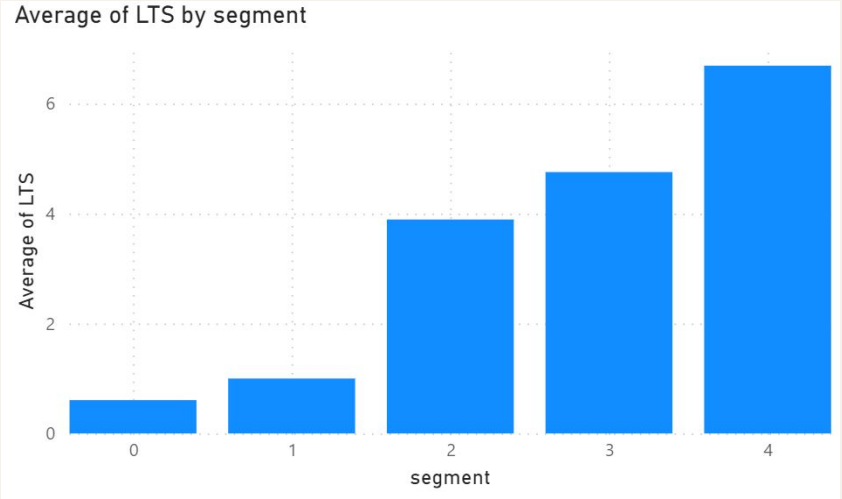
While segment **4** has the **highest** LTS average

So we could interpret :

The higher the segment,

=

the higher the LTS average



our main target of analysis will be segment 2, 3 and 4

Segment 0 Interpretation

From graph in previous slide, we could see that segment 0 has the lowest average LTS and Case Fatality Rate is **almost 0%**, so this segment **won't be** our main analysis target

Country that included in segment 0

Row Labels
0
Iran, Islamic Rep.
Jordan
Saudi Arabia
Turkiye
United Arab Emirates
United States

Historical CFR

Segment	Case Fatality Rate
0	0.00%
1	0.05%
2	0.13%
3	0.04%
4	0.62%

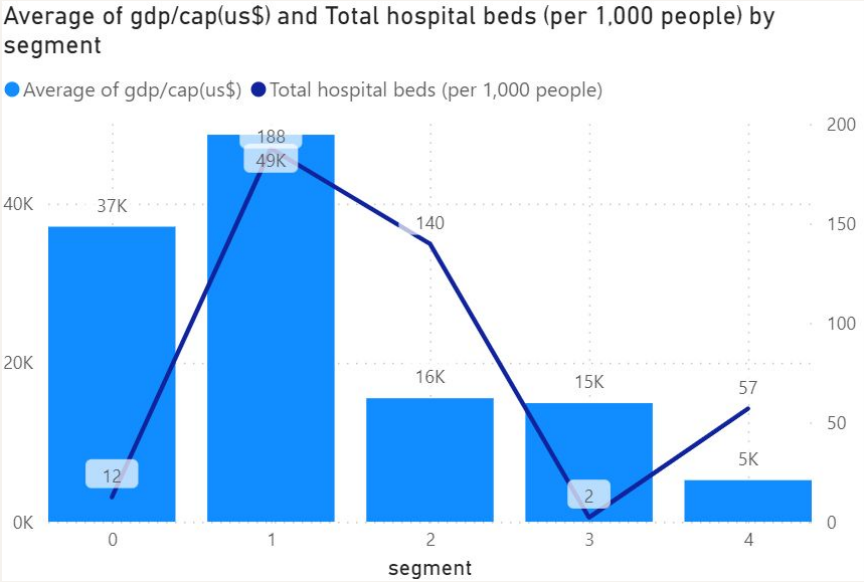
Segment 1 Interpretation

We interpret segment 1 as **prosperous** country and already have **good** health infrastructure.

Graph beside is showing segment 2 has **highest** average GDP per capita and **highest** total hospital beds (per 1000 people)

So, segment 1 **doesn't need** more attention.

We should focus on other segment instead



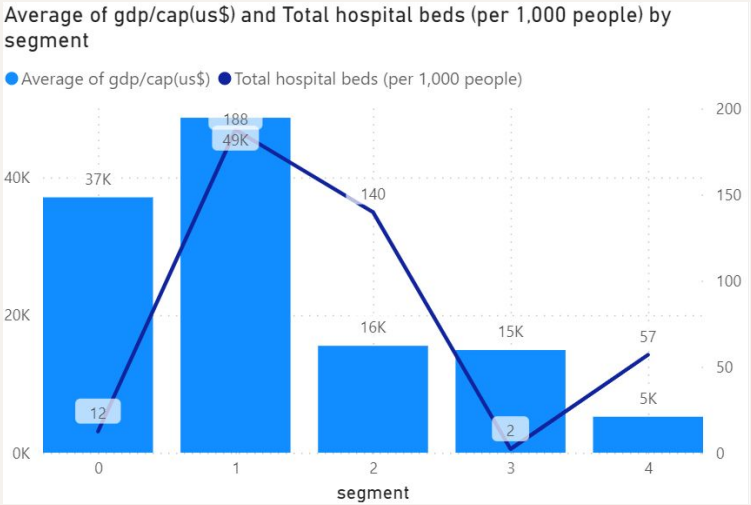
Segment 2 Interpretation

We interpret segment 2 as developing country with lower average GDP per capita, **but** it has **enough** and better health infrastructure, compared to segment **4** which has the highest case fatality rate due to dengue.

We might need to analyze **deeper** to know which country of this segment, is **still vulnerable** and could potentially become like Brazil (a country with similar feature with segment 2 but with extreme case outbreak)

Historical CFR

Segment	Case Fatality Rate
0	0.00%
1	0.05%
2	0.13%
3	0.04%
4	0.62%

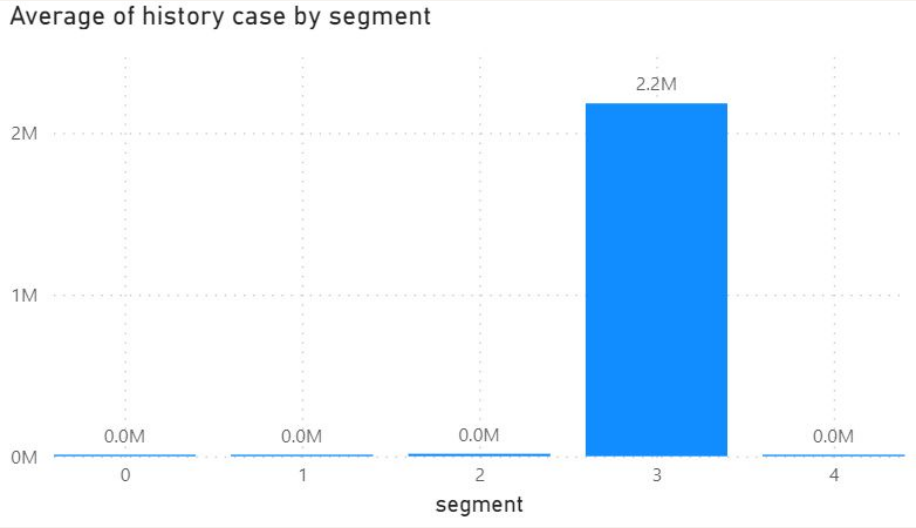


Segment 3 Interpretation

Segment 3 only consist **1 country**, which is Brazil.

Overall Brazil has **similar** characteristic with country in Segment **2**, but this year, there's a **dengue outbreak** in Brazil and the total case is increasing **extremely** and the average history case is much higher than other segment

That's why Brazil couldn't be included into either segment 2 or 4



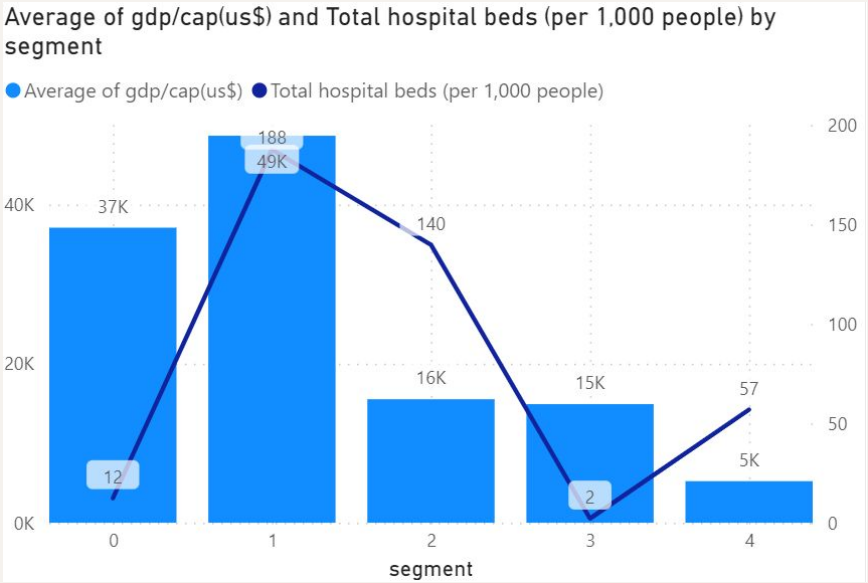
Segment 4 Interpretation

We interpret segment 4 as developing country with lowest average GDP per capita, and it may be the reason they don't have health infrastructure as good as other segment.

ALL of country included in this segment need more attention

Historical CFR

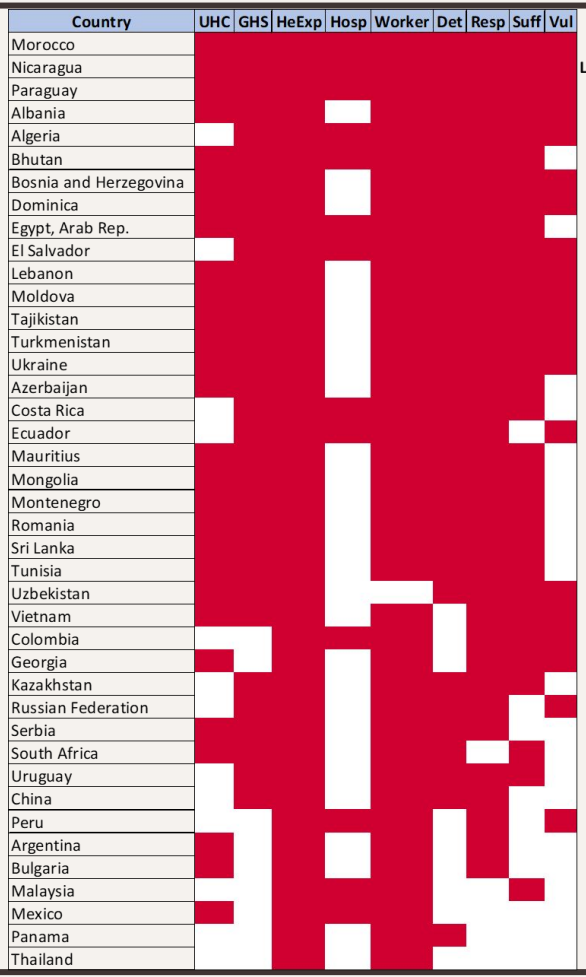
Segment	Case Fatality Rate
0	0.00%
1	0.05%
2	0.13%
3	0.04%
4	0.62%



Findings

We compare value from 9 parameters, with Brazil as the benchmark shown by heatmap on the right side.

Red colour indicates that the parameter value **is lower** than **benchmark**, which means **more vulnerable** and could possibly become same level as Brazil.



Legend =

UHC	=	Universal Health Coverage
GHS	=	Global Health Security
HeExp	=	Health Expenditure
Hosp	=	Total hospital per 1000 people
Worker	=	Total health workers (like nurses) per 1000 people
Det	=	Score of Early Detection & reporting for potential epidemics
Resp	=	Score of rapid response & epidemic spread mitigation
Suff	=	Score of sufficient & robust health sector
Vul	=	Score of overall risk environment & country vulnerability to biological threat

06

Summary

Summary

1. We can use historical case outbreak, health facility and security, climate and geographical conditions, and socio economic condition to **create a vulnerability map** for each country.

1. There's an **extreme dengue case in Brazil** caused by climate change, urbanization, and mobility.

1. Vulnerability measurement result could be used to determined which country **need** to act as soon as possible or help to handle dengue outbreak so we could create more **sustainable society** facing this global warming.

07

Recommendation

Recommendation

1. **Brazil** need to find an **effective and suitable solution** to reduce the number of active dengue cases.

1. Countries in **cluster 2** need to **prepare for a better health facility** just in case there's any dengue outbreak, especially countries **below Brazil** in the 9 parameters rank (visualized by heatmap in page 31)

2. Countries in **cluster 4** need to **increase their health facility and budget** as it's really vulnerable to dengue cases. Due to their low income and low healthcare facility, maybe UN can give a hand so they could be prepared if there is any dengue outbreak in the future.

3. Countries in **cluster 4** are also **have very high case fatality rate** caused of dengue compared to other clusters. This shows these countries also **need to be helped to improve**, mainly in their healthcare workforce as they are very shortage compared to others as shown below

Segment	Average of nurses and midwives (per 1,000 people)
0	5.94
1	9.82
2	3.79
3	7.40
4	1.22

08

Future Improvement

Future Improvement

1. As dengue's transmission is not limited in the country, it will be better to have a travelling or **trade data** between countries so that we can **track** the cases.

1. Find **more detailed** data (e.g., province, district, sub-district) to get deeper analysis of country that needs **more attention**

1. Use **geospatial** tool like ArcGIS to visualize the map better and also could analyze the **coverage** of health infrastructure (**not** only analyze the **number**), so it could be distributed more effectively.

09

Reference

Reference

1. Projecting the risk of mosquito-borne diseases in a warmer and more populated world: a multi-model, multi-scenario intercomparison modelling study
2. https://osf.io/hpaey/
3. https://www.ghsindex.org/
4. https://worldpopulationreview.com/country-rankings/gdp-per-capita-by-country
5. https://data.worldbank.org/indicator/SI.POV.GINI
6. https://worldpopulationreview.com/country-rankings/life-expectancy-by-country
7. https://www.indexmundi.com/g/r.aspx?v=69
8. https://data.humdata.org/dataset/world-bank-indicators-of-interest-to-the-covid-19-outbreak
9. https://ntdhq.shinyapps.io/dengue5/
10. https://data.worldbank.org/indicator/SH.XPD.CHEX.GD.ZS

Appendix : Deliverables Description

Script	What's in it?
[PYTHON - PRIMARY] Clustering for Vulnerability Mapping	Main python script, consist **data preparation** process (like dimensional reduction) and **the clustering** itself
[PYTHON - ADDITIONAL] Scraping Script for Some Data Collection	Python script for some data **collection** process since some data couldn't be downloaded easily from website
[PYTHON - ADDITIONAL] Script for Reverse Geocoding LTS Model data	Python script to complete **3rd** process of data preparation, which is Reverse Geocoding
[GOOGLE SHEETS - ADDITIONAL] Dataset Compilation	to complete **4th** process of data preparation, which is Joining multiple data source
[POWER BI - ADDITIONAL] Exploratory Data Analysis	Power BI file that consist some graph that we created when conduct exploratory data analysis. Some graph result is already **copy** into this **ppt** file as **explanatory** analysis
[POWER BI - PRIMARY] Vulnerability Map with Statistic Info per Country	Power BI file that consist **final** dashboard. It shows **vulnerability** map and **statistic info** from each country as **tooltips**