

DENGUE WITH DATA

01

02

03

04

05

Introduction

- Situation
- Business Problem
- Data Project

01

Introduction

- Situation
- Business Problem
- Data Project

02

Data Understanding

- Purpose
- Source
- Quality

03

04

05

01

Introduction

- Situation
- Business Problem
- Data Project

02

Data Understanding

- Purpose
- Source
- Quality

03

Data Preparation

- Cleaning
- Relationship

04

05

01

Introduction

- Situation
- Business Problem
- Data Project

02

Data Understanding

- Purpose
- Source
- Quality

03

Data Preparation

- Cleaning
- Relationship

04

Analysis

Dashboard



05

01 Introduction

- Situation
- Business Problem
- Data Project

02

Data Understanding

- Purpose
- Source
- Quality

03

Data Preparation

- Cleaning
- Relationship

04

Analysis

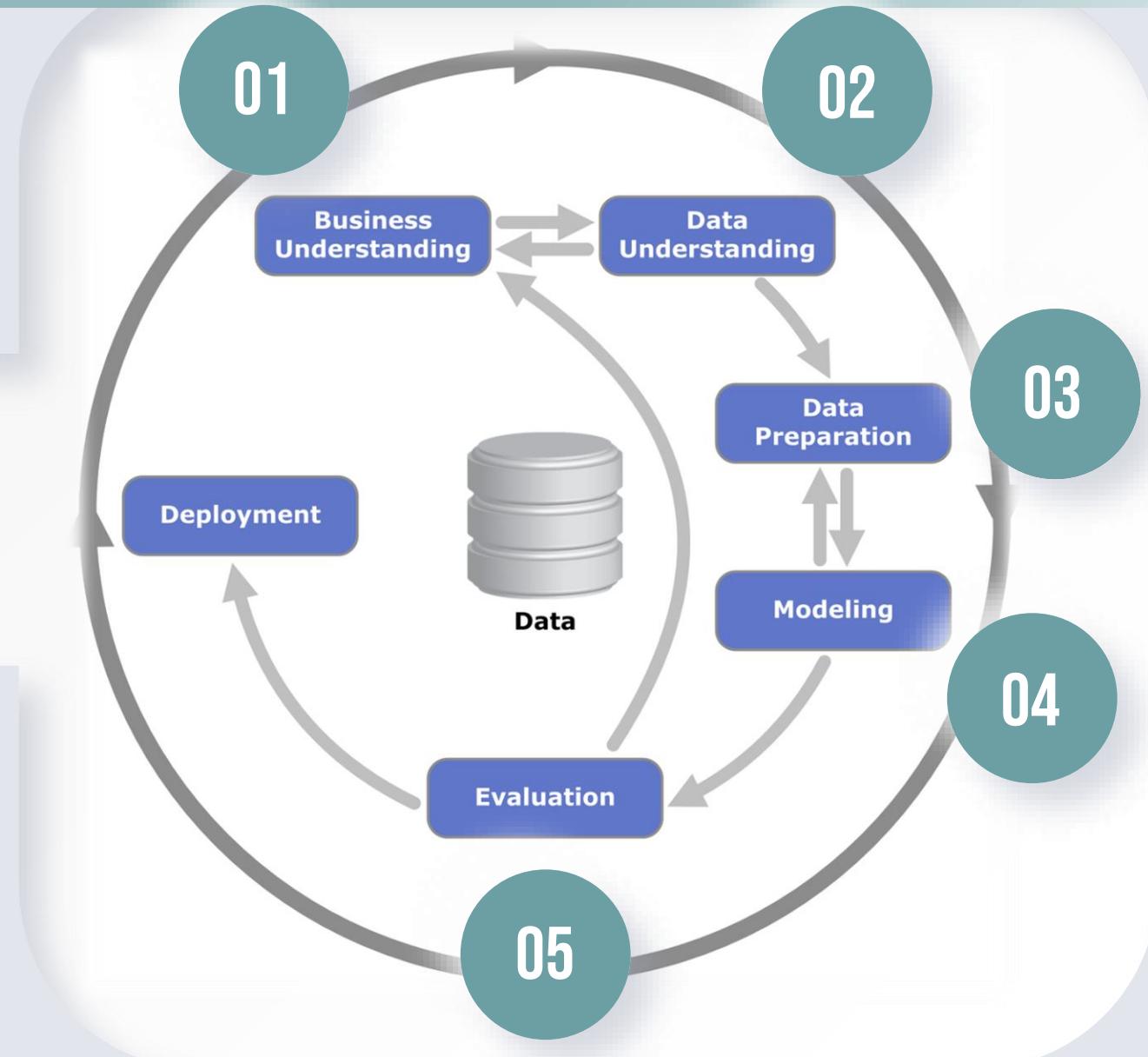


05

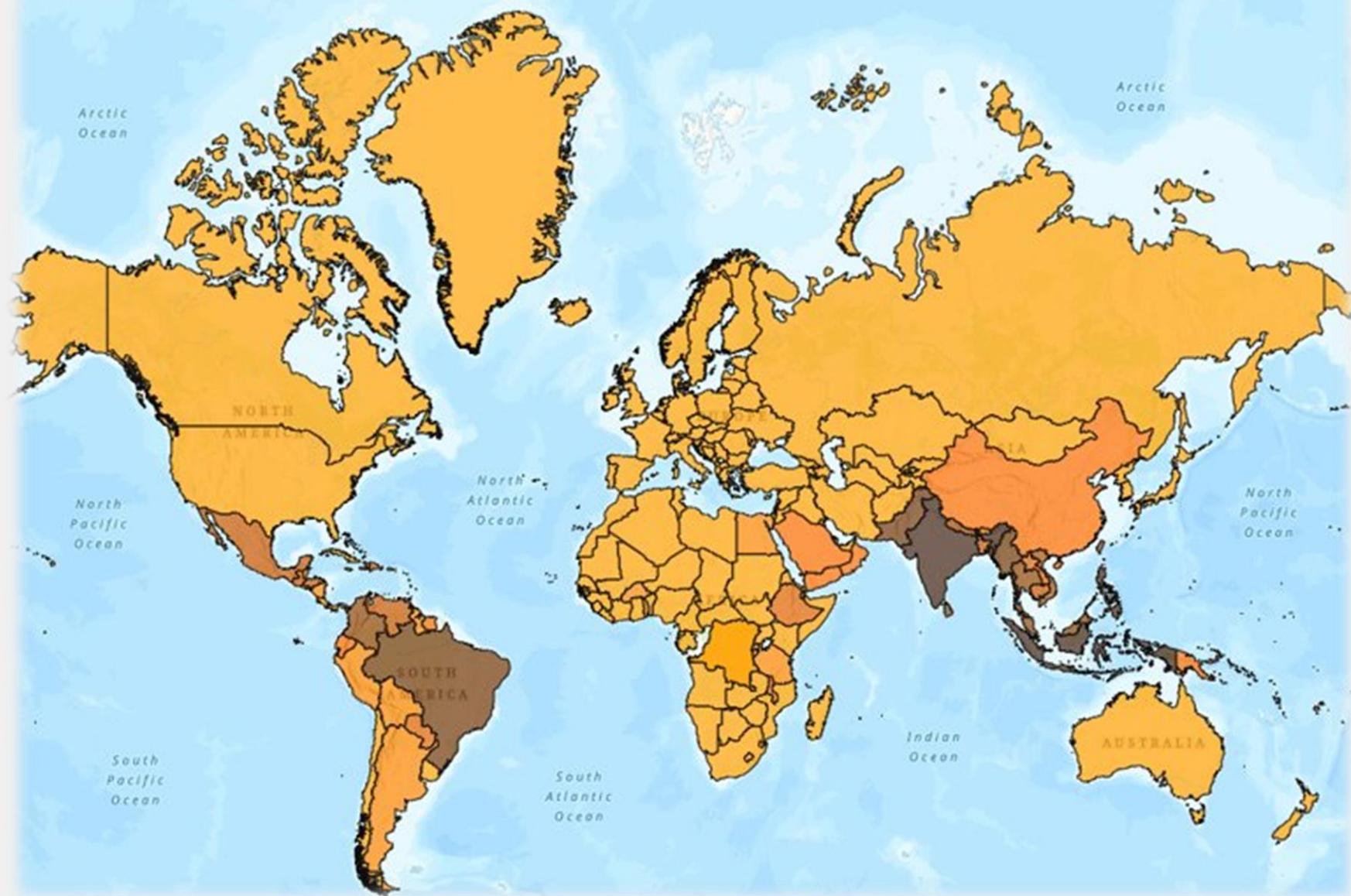
Dashboard

- ### Development
- Evaluation
 - Conclusion
 - Reflection

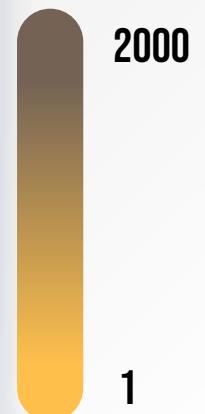
CRISP-DM Methodology



SYMPTOMS



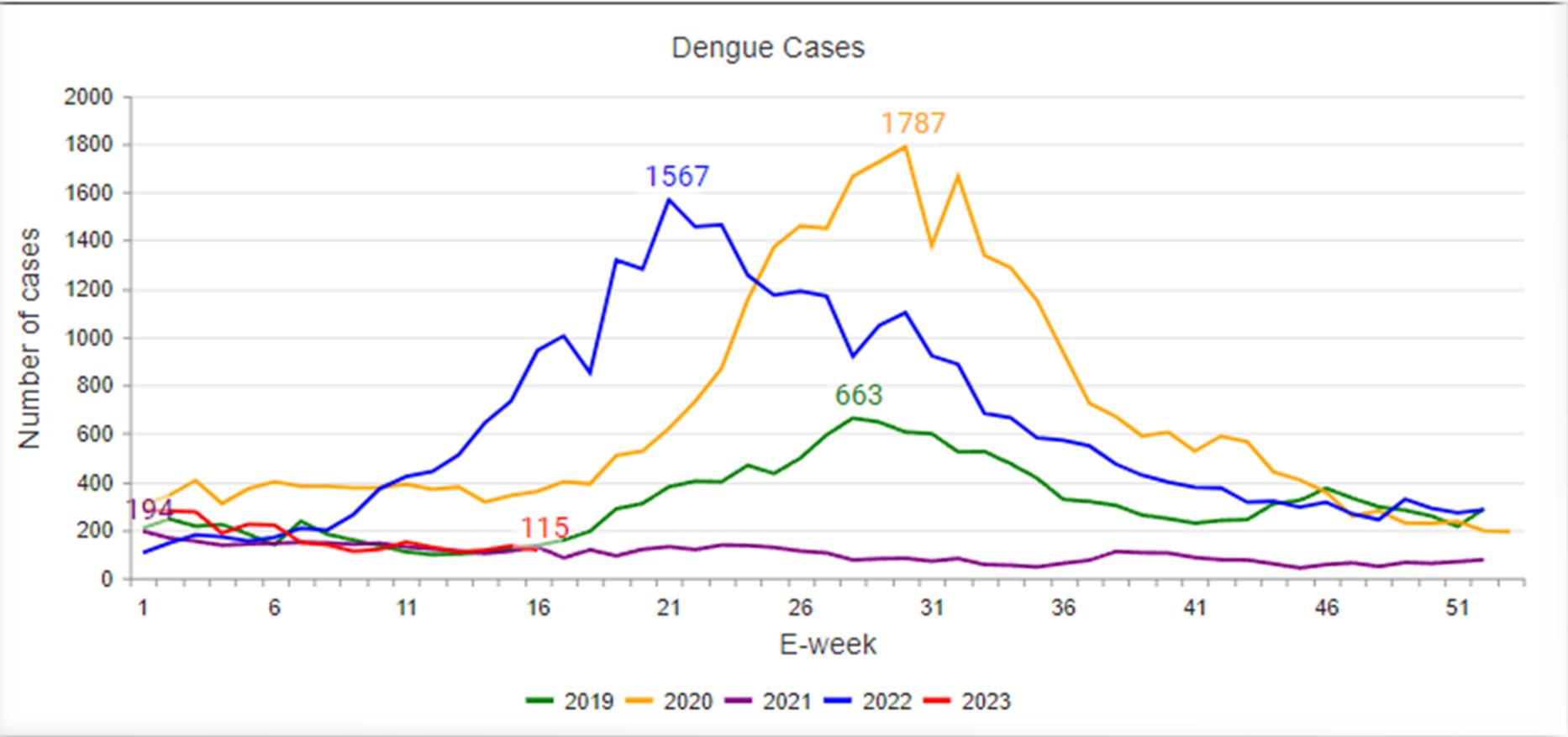
Yearly
Average
Dengue
Death
2013-2019



DENGUE

Cumulative No. of cases for 2023 (First 16 E-weeks) 2725

Compiled by Communicable Diseases Division, Ministry of Health



~32,000
in 2022

Singapore

NEA warns of another dengue outbreak in 2023 as cases remain high in January

A total of 279 dengue cases were reported last week, almost double the number in the same period last year.



An Aedes aegypti mosquito. (File photo: AFP/Miguel SCHINCARIOL)

SINGAPORE: The number of weekly dengue cases has remained high going into the new year, prompting the National Environment Agency (NEA) to warn of the risk of another outbreak in 2023.

More than 32,000 dengue cases were reported in 2022, [about six times the total number of cases recorded in the whole of 2021](#).

19 Jan 2023

Business understanding

Singapore

More than 100 dengue cases reported weekly, risk of transmission remains high: NEA

25 Mar 2023 06:01PM
(Updated: 25 Mar 2023 11:09PM)



The number of weekly dengue cases in Singapore has remained above 100 since the start of the year, with more than 2,000 cases reported as of Mar 24. Lauren Ong reports.

SINGAPORE: The number of weekly dengue cases in Singapore has remained above 100 since the start of the year, with more than 2,000 cases reported as of Mar 24.

Related Topics

National Environment Agency

25 Mar 2023



MINISTRY OF HEALTH
SINGAPORE



BLOCK & SAW

STAKEHOLDERS



National
Environment
Agency

Safeguard • Nurture • Cherish



Gravitraps

PROJECT
WOLBACHIA
SINGAPORE



2016



Ground Inspections

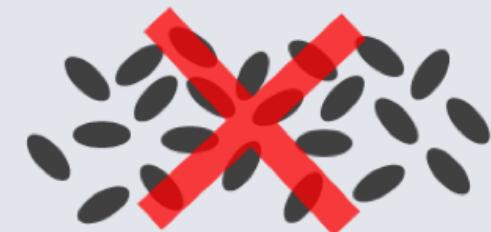
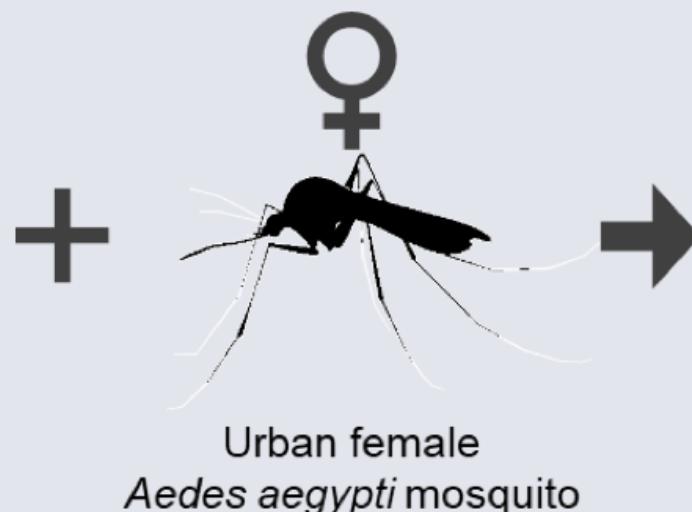
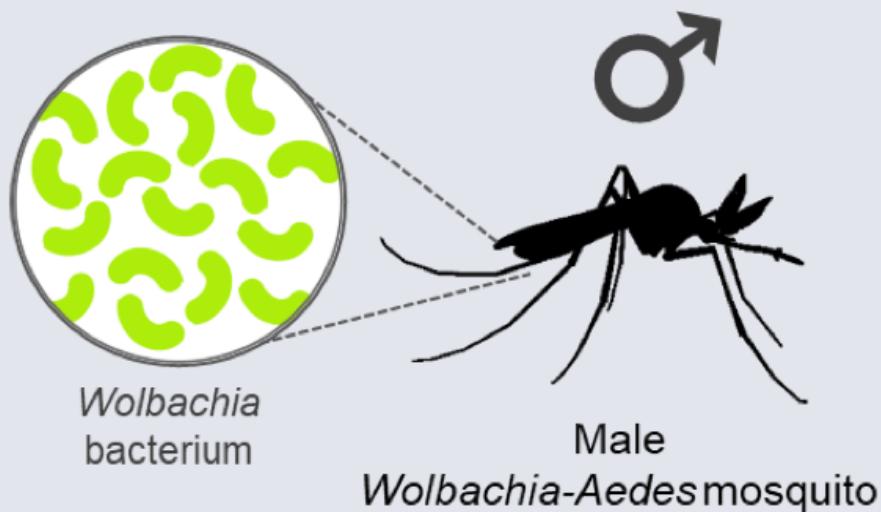


STAKEHOLDERS



National
Environment
Agency

PROJECT **WOLBACHIA** SINGAPORE





STAKEHOLDERS



MINISTRY OF HEALTH
SINGAPORE



National
Environment
Agency

Safeguard • Nurture • Cherish



STAKEHOLDERS



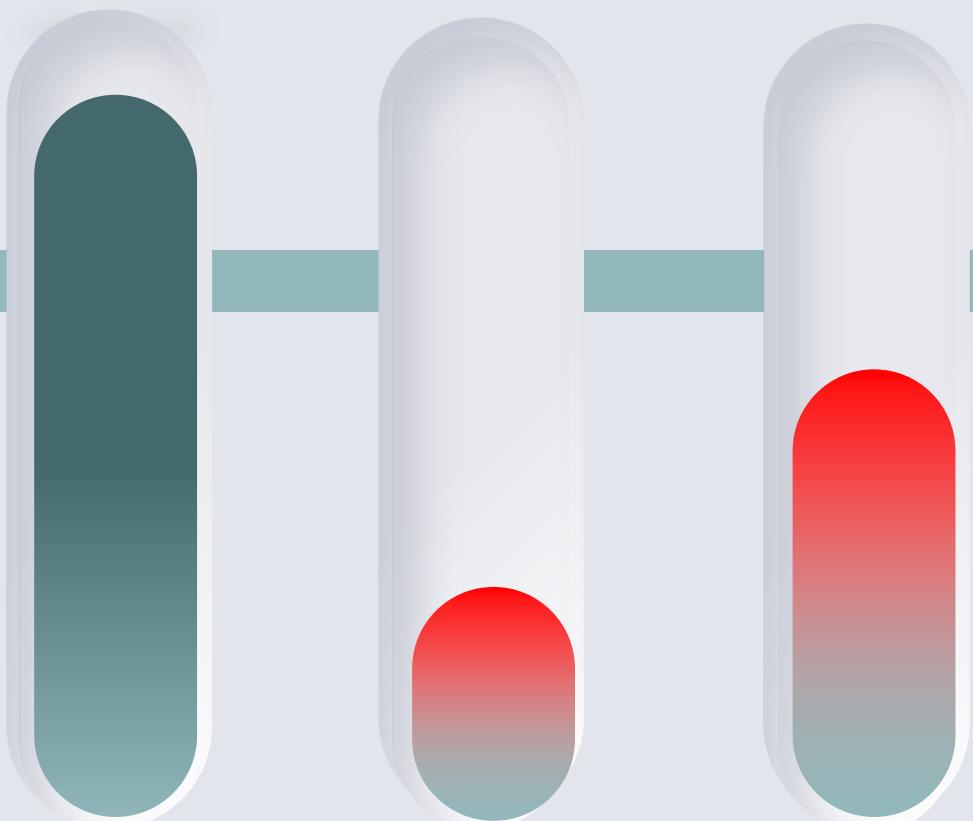
MINISTRY OF HEALTH
SINGAPORE



National
Environment
Agency

Safeguard • Nurture • Cherish





STAKEHOLDERS



MINISTRY OF HEALTH
SINGAPORE



National
Environment
Agency
Safeguard • Nurture • Cherish



Target Action Plan Dashboard



Location



Timeframe

High probability/possibility of Dengue

GAMEPLAN

Business understanding

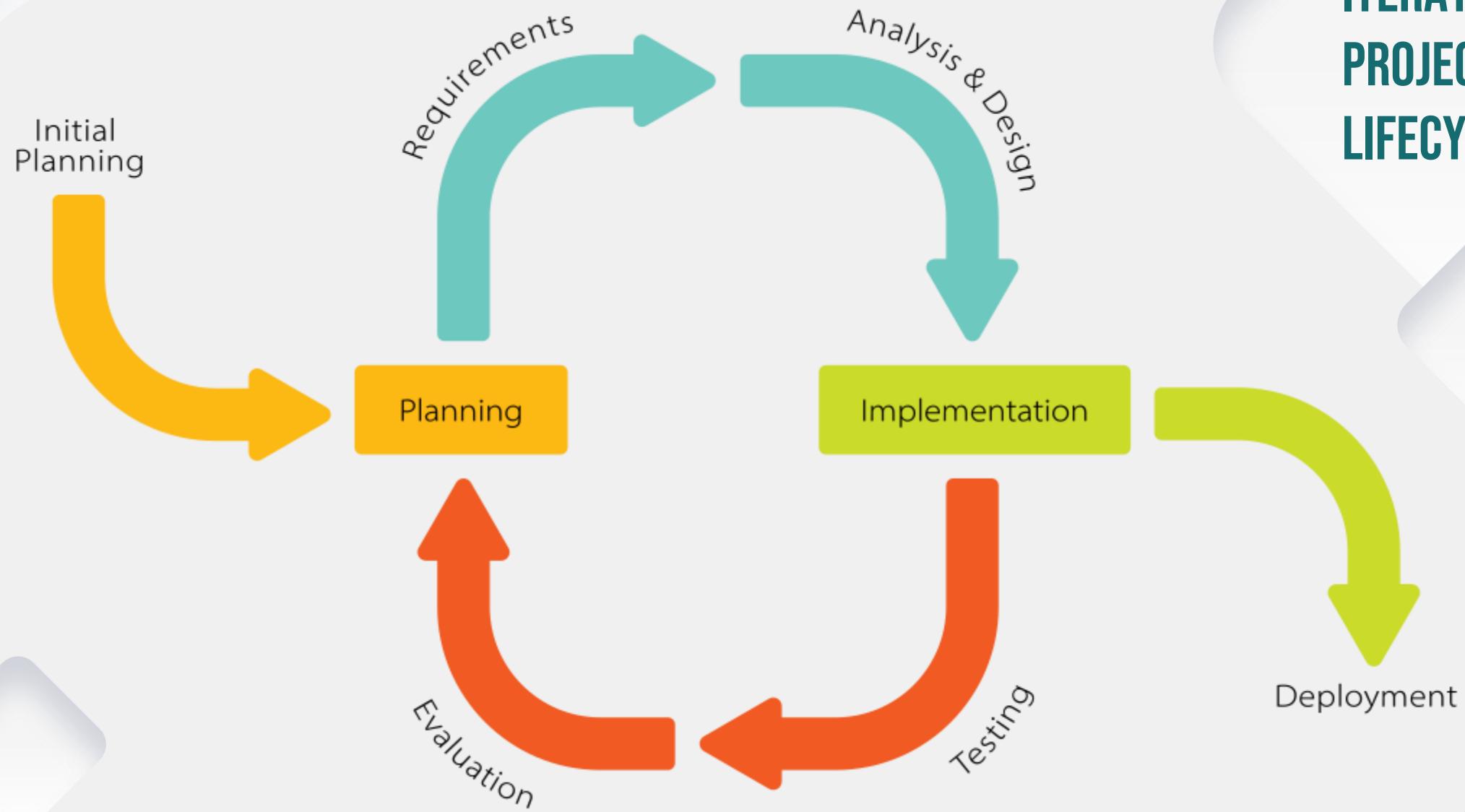
Explore Factors

Analyse Patterns and
Associations

Discover insight for targeted
Vector action plan

Evaluate for next iteration

ITERATIVE PROJECT LIFECYCLE





Data Sources



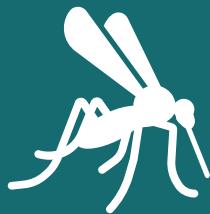
Data Understanding

SGCharts

2013 - 2020

Dengue Count

Date
Address



Temperature
Humidity
Rainfall
Date



Address
Latitude/Longitude
Planning Area

Urbanization Infra (*Hawker Centre, Parks and water bodies*)

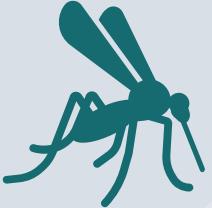


Population
Type of Dwelling
Date
Planning Area

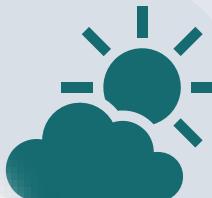
SGCharts
2013 - 2020

Dengue Count

Date
Address



Temperature



**METEOROLOGICAL
SERVICE
SINGAPORE**

Data Exploration

**One
map**
Singapore

ngitu

Planning Area

Urban
Centre,



**URBAN
REDEVELOPMENT
AUTHORITY**



DEPARTMENT OF
**STATISTICS
SINGAPORE**

—Empowering You with Trusted Data—



Date
Planning Area



Data Exploration & Transformation

Data Preparation

SGCharts



WIKIPEDIA



Data Preparation

Remove duplicates,
Remove empty data,
Joining via common keys



Data Preparation

Data Understanding

Scatterplot, Descriptive Statistics,
Histograms, Barcharts,
Correlation, Inspection



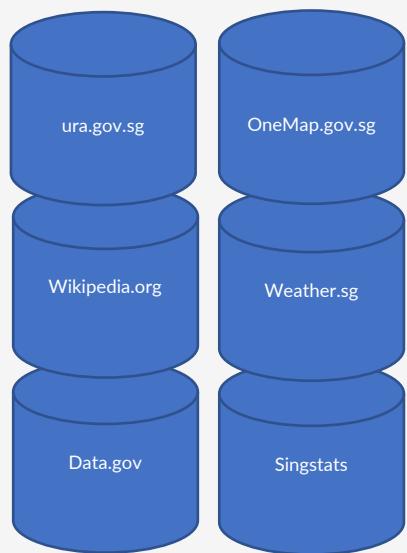
DATA WAREHOUSE



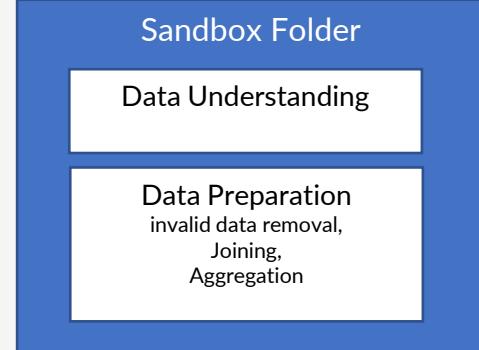
DASHBOARD



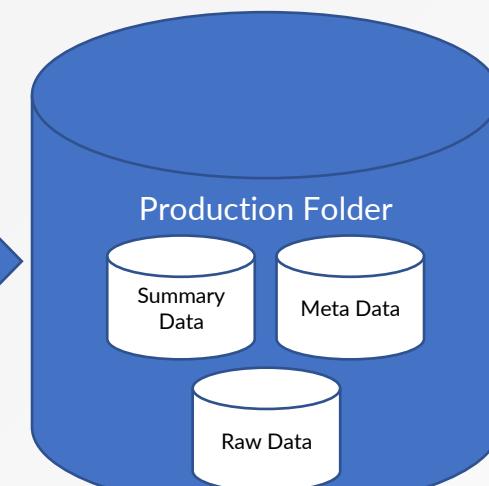
DATA SOURCE



ETL



PRODUCTION



Latitude/
Longitude

Dengue Case Count				
latitude	longitude	case_count	Type	year
1.264620	103.8144	2	Dengue Cases	2020
1.267379	103.8167	2	Dengue Cases	2016
1.268166	103.8145	2	Dengue Cases	2015
1.268866	103.8246	2	Dengue Cases	2016
1.269953	103.6206	3	Dengue Cases	2016
1.270483	103.8231	1	Dengue Cases	2014
1.270483	103.8231	1	Dengue Cases	2016
1.270958	103.8229	1	Dengue Cases	2013
1.270958	103.8229	2	Dengue Cases	2016
1.270958	103.8229	1	Dengue Cases	2019

Rainfall		
Planning Area	Date (M/D/YYYY)	Sum Daily Rainfall (mm)
Ang Mo Kio	1/1/2010	1.2
Ang Mo Kio	2/1/2010	0
Ang Mo Kio	3/1/2010	10.2
Ang Mo Kio	4/1/2010	0.2
Ang Mo Kio	5/1/2010	0
Ang Mo Kio	6/1/2010	22.4
Ang Mo Kio	7/1/2010	0
Ang Mo Kio	8/1/2010	0
Ang Mo Kio	9/1/2010	0
Ang Mo Kio	10/1/2010	0
Ang Mo Kio	11/1/2010	0
Ang Mo Kio	12/1/2010	0
Ang Mo Kio	13/1/2010	2.4
Ang Mo Kio	14/1/2010	0.2
Ang Mo Kio	15/1/2010	0

Planning Area

Address
Postal Code

Hawker Centers		
Address	Postal Code	Planning Area
1, Jalan Dato' Dr. Mahathir (S) 289760	289896	NOVENA
1, Jalan Dato' Dr. Mahathir (S) 289760	289896	DOWNTOWN CORE
1, Raffles Quay (S) 049972	049911	RAFFLES QUAY
38A, Bee Crescent, S(169982)	169982	BUKIT MERAH
166, Jalan Besar, S(208877)	208877	ROCHOR
51, Upper Bukit Timah Road, S(588215)	588215	BUKIT MERAH
20, Kensington Park Road, S(557269)	557269	SERangoon
31, Commonwealth Avenue, S(149644)	149644	QUEENSTOWN
271, Clem Road, S(242768)	242768	MARINE PARADE
1220, East Coast Parkway, S(488960)	488960	BEDOK
Geylang Serai (S) 402001	402001	GEYLANG
569, Branks Road, S(199583)	199583	LAZANG
1, Lorong Bahru (S) 277700	277700	QUEENSTOWN
17, Old Airport Road, S(142972)	142972	GEYLANG
38A, Margaret Drive, S(142038)	142038	QUEENSTOWN
86, Market Street, Capital Building, S(048947)	048947	DOWNTOWN CORE
4, Woodlands Street (S) 199522	199522	WOODLANDS
1, Kadamban Street (S) 090184	090184	DOWNTOWN CORE
500, Clemenceau Ave North, S(229495)	229495	NEWTON
861, North Bridge Road, S(198783)	198783	KALLANG

Data Table (Others)

Feature 1, Feature 2, ...

Original datasets do not have a common identifier that could be used to integrate the data tables.

No common
identifier tag

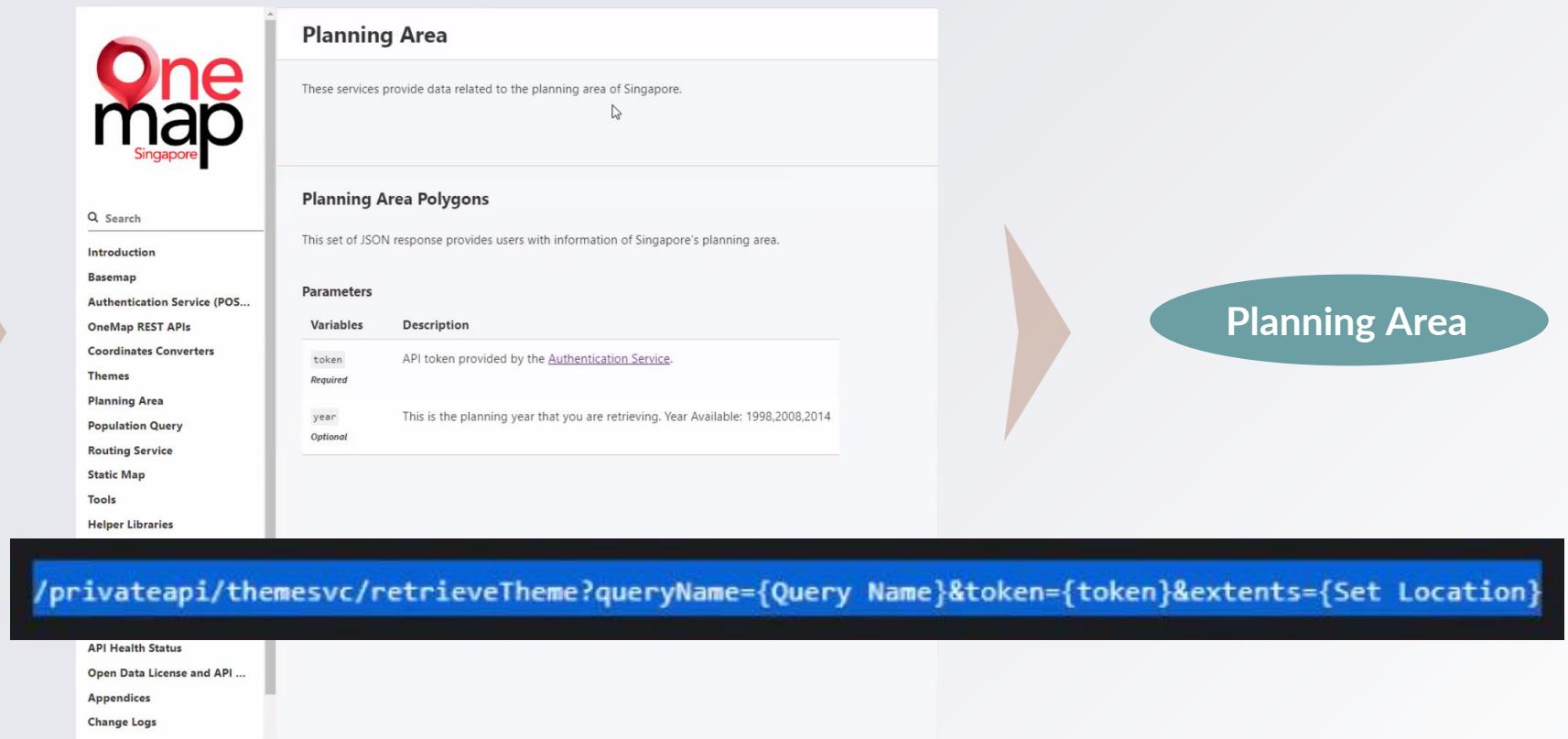
Dengue Case Count				
latitude	longitude	case_count	Type	year
1.264620	103.8144	2	Dengue Cases	2020
1.267379	103.8167	2	Dengue Cases	2016
1.268166	103.8145	2	Dengue Cases	2015
1.268866	103.8246	2	Dengue Cases	2016
1.269953	103.6206	3	Dengue Cases	2016
1.270483	103.8231	1	Dengue Cases	2014
1.270483	103.8231	1	Dengue Cases	2016
1.270958	103.8229	1	Dengue Cases	2013
1.270958	103.8229	2	Dengue Cases	2016
1.270958	103.8229	1	Dengue Cases	2019

Rainfall		
Planning Area	Date (M/D/YYYY)	Sum Daily Rainfall (mm)
Ang Mo Kio	1/1/2010	1.2
Ang Mo Kio	2/1/2010	0
Ang Mo Kio	3/1/2010	10.2
Ang Mo Kio	4/1/2010	0.2
Ang Mo Kio	5/1/2010	0
Ang Mo Kio	6/1/2010	22.4
Ang Mo Kio	7/1/2010	0
Ang Mo Kio	8/1/2010	0
Ang Mo Kio	9/1/2010	0
Ang Mo Kio	10/1/2010	0
Ang Mo Kio	11/1/2010	0
Ang Mo Kio	12/1/2010	0
Ang Mo Kio	13/1/2010	2.4
Ang Mo Kio	14/1/2010	0.2
Ang Mo Kio	15/1/2010	0

UNIQUE KEYS

- Address
- Postal Code
- Latitude/
Longitude

WEB API



COMMON IDENTIFIER

- Planning Area

DATA SOURCE

Station	Year	Month	Day	Daily Rainf
Paya Lebar	2010	1	1	0
Paya Lebar	2010	1	2	0
Paya Lebar	2010	1	3	20.9
Paya Lebar	2010	1	4	0.1
Paya Lebar	2010	1	5	0.1
Paya Lebar	2010	1	6	2.3
Paya Lebar	2010	1	7	3.1
Paya Lebar	2010	1	8	2.1
Paya Lebar	2010	1	9	0
Paya Lebar	2010	1	10	0
Paya Lebar	2010	1	11	0
Paya Lebar	2010	1	12	5.9
Paya Lebar	2010	1	13	0.5
Paya Lebar	2010	1	14	0
Paya Lebar	2010	1	15	0
Paya Lebar	2010	1	16	0
Paya Lebar	2010	1	17	0
Paya Lebar	2010	1	18	0.4
Paya Lebar	2010	1	19	6.9
Paya Lebar	2010	1	20	0
Paya Lebar	2010	1	21	2.5
Paya Lebar	2010	1	22	11.3
Paya Lebar	2010	1	23	0
Paya Lebar	2010	1	24	0
Paya Lebar	2010	1	25	0
Paya Lebar	2010	1	26	0
Paya Lebar	2010	1	27	0
Paya Lebar	2010	1	28	5.9
Paya Lebar	2010	1	29	2
Paya Lebar	2010	1	30	0
Paya Lebar	2010	1	31	0

Total rows = 273,498

PARSING SCRIPT

```

temp.py x grabWeatherDataSg.py x appendPlanningArea.py
1 # -*- coding: utf-8 -*-
2 """
3 Created on Sat Mar 11 21:06:43 2023
4
5 This script is to parse the data in Weather Data Extraction.xlsx
6 If executing from a command line, enter: python grabWeatherDataSg.py palist.csv [start] [end]
7 Note: Install Ananconda and use ananconda shell. Run script file in residing folder
8
9 grabWeatherDataSg.py: script to get weather data of Singapore from MSS
10 palist.csv: file that lists the planned area code
11 start: start year and month in yyyy-mm, e.g. 2010-01
12 end: end year and month in yyyy-mm, e.g. 2023-03
13
14 Default, without any start and end entered, script would extract data from 2010-01 onwards to now
15
16 Script would save extracted data into sgWeatherData.csv, in the script's residing folder
17
18 Note: earliest data starts from 1980 in MSS website
19
20 @author: Wee Yang
21 """
22
23 import pandas as pd
24 import requests as rq
25 import sys, os, re, time
26
27 from datetime import datetime
28
29
30 def chkDateFormatOk(string):
31     return(re.match('^\d{4}-\d{2}$', string) is not None)
32
33 def parseDate(string):
34     return(string.split('-'))
35
36 def grabData(year, month, palist, homePath, url):
37     palist = pd.read_csv(os.path.join(homePath, palist))
38     year = str(year)
39     month = str(month)
40     data = pd.DataFrame()
41
42     if len(month) < 2:
43         month = "0" + month
44
45     for i in range(len(palist)):
46         print(palist.iloc[i, 1])
47
48         arealNum = str(palist.iloc[i, 0])
49
50         if len(arealNum) < 2:
51             arealNum = "0" + arealNum
52
53         csvFile = url + "/DAILYDATA_S" + arealNum + "_" + year + str(month) + ".csv"
54
55         time.sleep(1) # suspect MSS ftp server download unable to respond in time
56
57         temp = rq.get(csvFile).content.decode('unicode_escape')
58         temp = temp.replace('\xc2', '')
59         temp = temp.replace('\xa0', '')

```

grabWeatherDataSg.py appendPlanningArea.py

OUTPUT

Planning Area

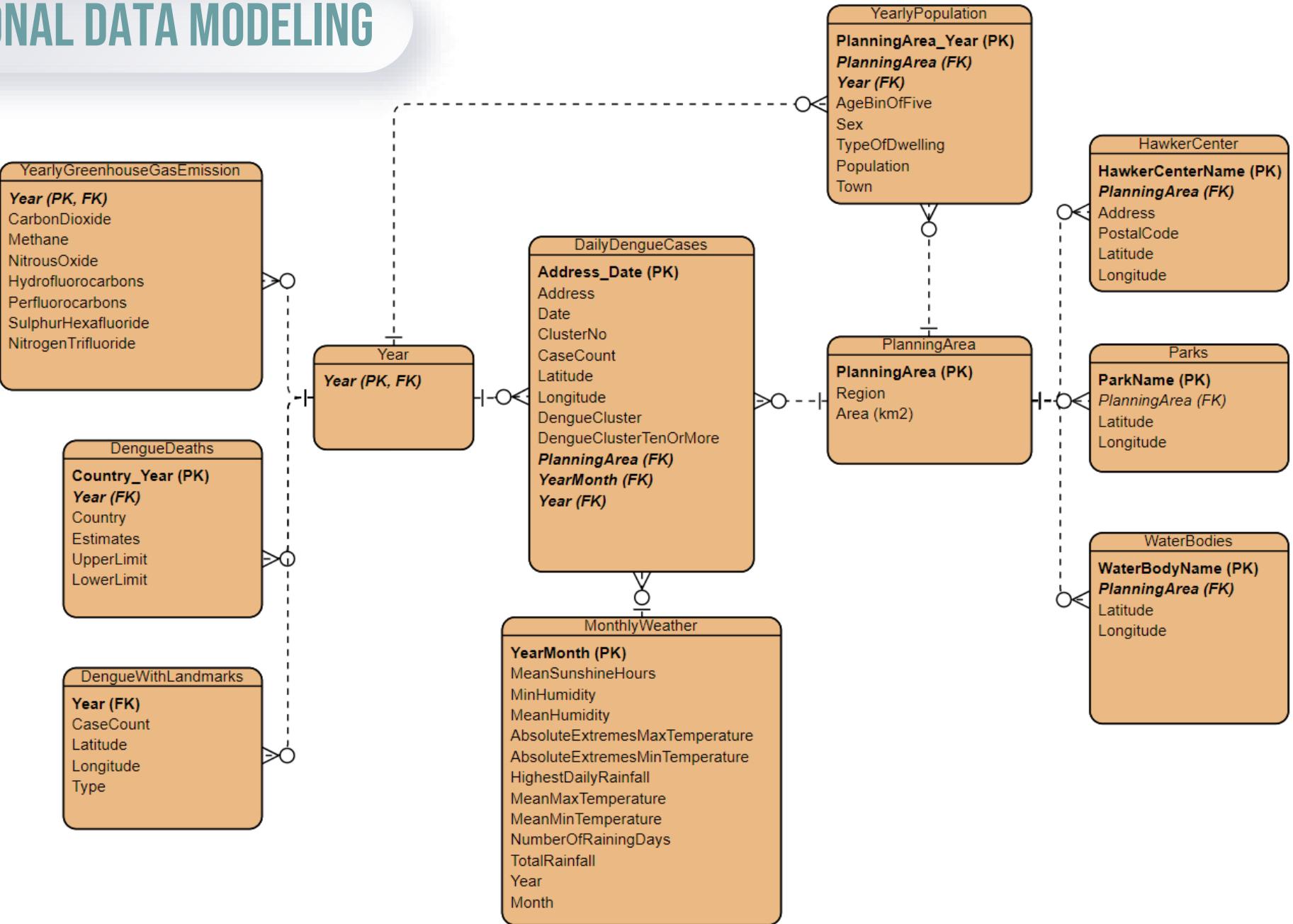
Planning Area	Station	Year	Month	Day	Daily Rainf
Central Water Catchment	Lower Peirce Reservoir	2010	1	2	6.6
Central Water Catchment	Lower Peirce Reservoir	2010	1	3	22
Central Water Catchment	Lower Peirce Reservoir	2010	1	4	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	5	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	6	18.6
Central Water Catchment	Lower Peirce Reservoir	2010	1	7	6.2
Central Water Catchment	Lower Peirce Reservoir	2010	1	8	8.8
Central Water Catchment	Lower Peirce Reservoir	2010	1	9	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	10	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	11	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	12	2
Central Water Catchment	Lower Peirce Reservoir	2010	1	13	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	14	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	15	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	16	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	17	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	18	0.6
Central Water Catchment	Lower Peirce Reservoir	2010	1	19	6.6
Central Water Catchment	Lower Peirce Reservoir	2010	1	20	0
Central Water Catchment	Lower Peirce Reservoir	2010	1	21	51.4
Central Water Catchment	Lower Peirce Reservoir	2010	1	22	12.4
Central Water Catchment	Lower Peirce Reservoir	2010	1	23	0

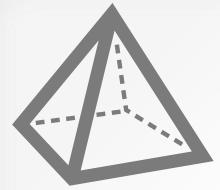
Planning Area

DATA DICTIONARY

Table	Field/ Column Name	Key (PK/FK)	Unique Value?	Is required?	Data Type	Domain	Sample data value	Description
Year	Year	PK/FK	Y	Y	int	[0-9][0-9][0-9][0-9]	2013	Year
YearlyGreenhouseGasEmission	Year	PK/FK	Y	Y	int	[0-9][0-9][0-9][0-9]	2013	Year
	CarbonDioxide	-	N	N	float	[0-9][0-9].[0-9]	46.7	Gas emissions in CO2 megatons equivalent
	Methane	-	N	N	float	[0-9].[0-9]	0.1	Gas emissions in CO2 megatons equivalent
	NitrousOxide	-	N	N	float	[0-9].[0-9]	0.6	Gas emissions in CO2 megatons equivalent
	Hydrofluorocarbons	-	N	N	float	[0-9].[0-9]	0.5	Gas emissions in CO2 megatons equivalent
	Perfluorocarbons	-	N	N	float	[0-9].[0-9]	1.3	Gas emissions in CO2 megatons equivalent
	SulphurHexafluoride	-	N	N	float	[0-9].[0-9]	0.1	Gas emissions in CO2 megatons equivalent
	NitrogenTrifluoride	-	N	N	float	[0-9].[0-9]	0.4	Gas emissions in CO2 megatons equivalent
PlanningArea	PlanningArea	PK	Y	Y	varchar	[A-Z]...	Jurong	District demarcated for administration and development
	Region	-	N	N	varchar	[A-Z]...	Northwest	Geographical regional district
	Area (km ²)	-	N	N	int	[0-9][0-9][0-9]	54	Land area in kilometer square
YearlyPopulation	PlanningArea_Year	PK	Y	Y	varchar	[A-Z]..._[0-9][0-9][0-9][0-9]	Jurong_2013	Composite key made up of planning areas and year
	PlanningArea	FK	N	Y	varchar	[A-Z]...	Jurong	District demarcated for administration and development
	Year	FK	N	Y	int	[0-9][0-9][0-9][0-9]	2013	Year
Parks	ParkName	PK	Y	Y	varchar	[A-Z]...	Windsor Nature Park	Name of park
	PlanningArea	FK	N	Y	varchar	[A-Z]...	Jurong	District demarcated for administration and development
	Latitude	-	N	Y	float	[0-9][0-9][0-9].[0-9]...	103.836995927634	Geographic coordinates
	Longitude	-	N	Y	float	[0-9][0-9][0-9].[0-9]...	103.846320994422	Geographic coordinates

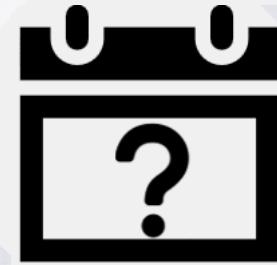
OLAP DIMENSIONAL DATA MODELING





Data Analysis + Dashboard

Key Objectives



WHEN



WHERE

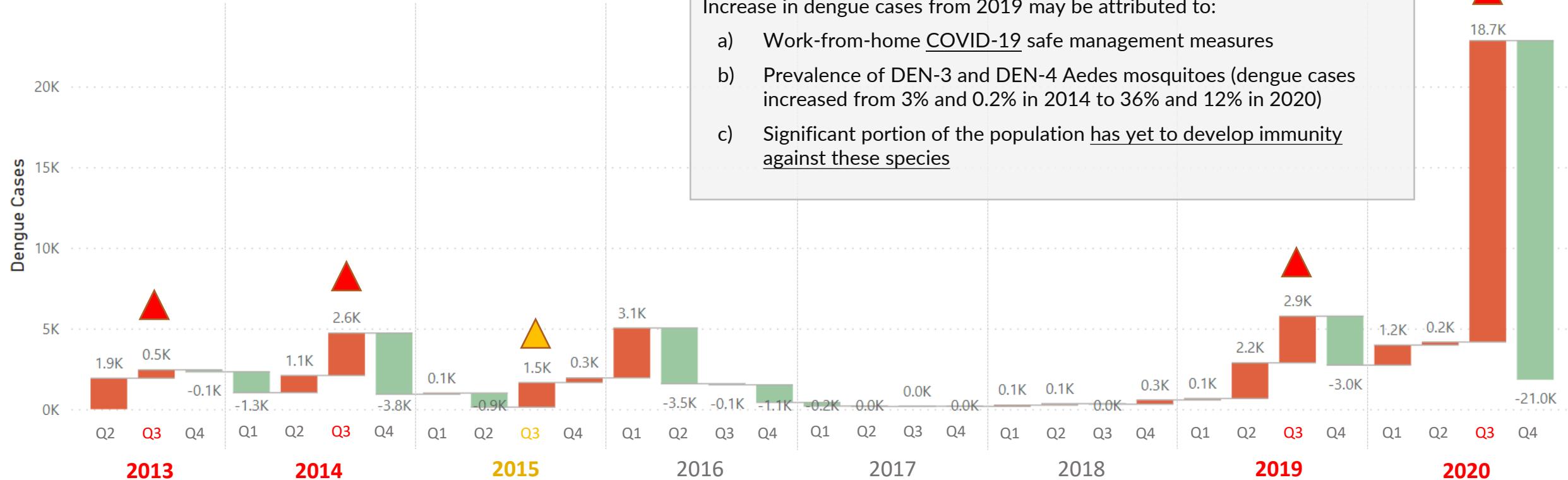


WHEN

Historical Trends on Dengue Cases

Most of the Q3 hit HIGHEST Dengue Cases

● Increase ● Decrease



Increase in dengue cases from 2019 may be attributed to:

- Work-from-home COVID-19 safe management measures
- Prevalence of DEN-3 and DEN-4 Aedes mosquitoes (dengue cases increased from 3% and 0.2% in 2014 to 36% and 12% in 2020)
- Significant portion of the population has yet to develop immunity against these species

NEA started Wolbachia-Aedes suppression program since 2016, which managed to control the Dengue Outbreak till 2019.

WHY Q3?

Studies have suggested that the **warmer** and **greater humid** months between **June and October** are contributing to a peak in dengue cases

Environmental Factors



WEATHER



GREENHOUSE
GASES

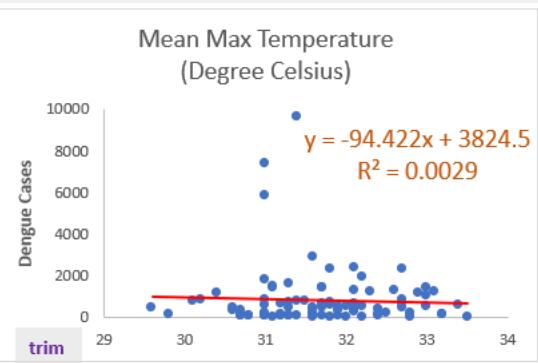


Poor correlation coefficient

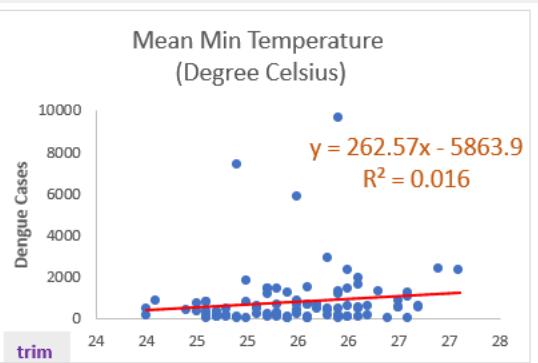
Monthly Weather Reading	Correlation Coefficient
Temperature (Degree Celsius) <ul style="list-style-type: none">- Means Daily Maximum- Means Daily Minimum- Absolute Extremes Maximum- Absolute Extremes Minimum	(0.05) 0.13 (0.16) 0.01
Humidity (%) <ul style="list-style-type: none">- Minimum Relative Humidity- 24 Hours Mean Relative Humidity	(0.02) (0.10)
Rainfall (mm) <ul style="list-style-type: none">- Total Rainfall- Highest Daily Rainfall Total	(0.06) (0.06)
Number Of Rainy Days (Number)	0.01
Bright Sunshine Daily Mean (Hour)	0.01



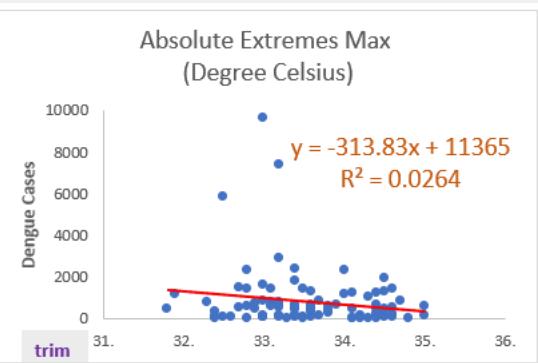
Mean Max Temperature
(Degree Celsius)



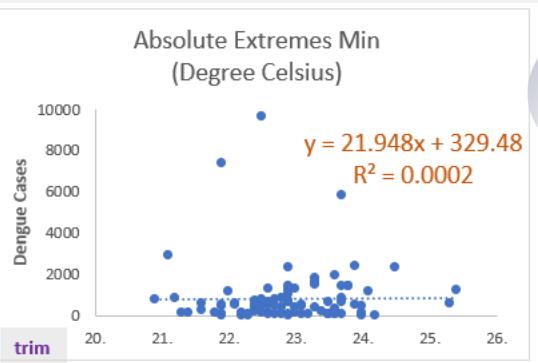
Mean Min Temperature
(Degree Celsius)



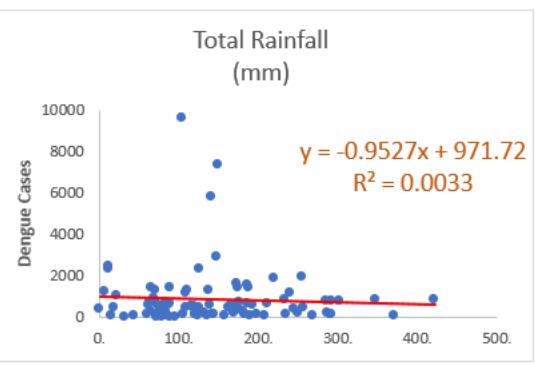
Absolute Extremes Max
(Degree Celsius)



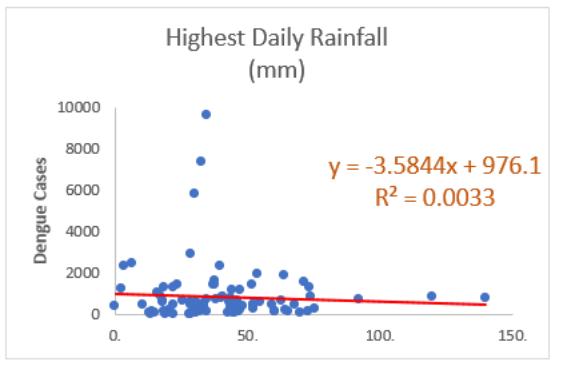
Absolute Extremes Min
(Degree Celsius)



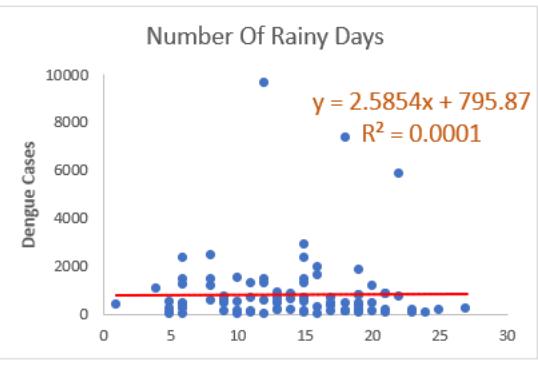
Total Rainfall
(mm)



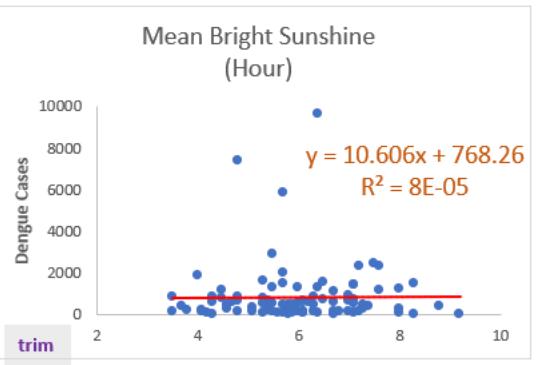
Highest Daily Rainfall
(mm)



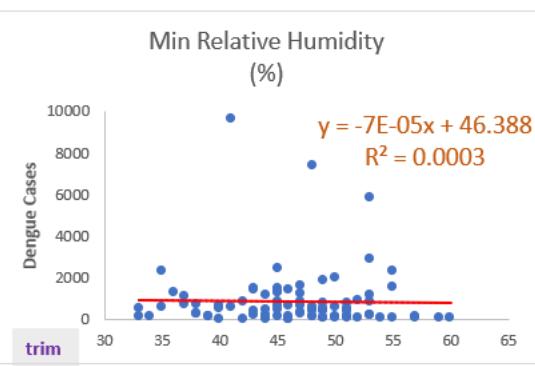
Number Of Rainy Days



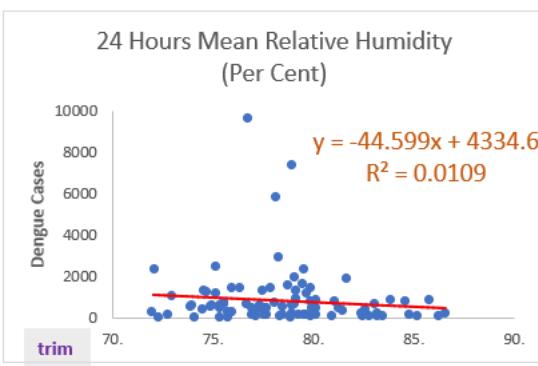
Mean Bright Sunshine
(Hour)



Min Relative Humidity
(%)



24 Hours Mean Relative Humidity
(Per Cent)



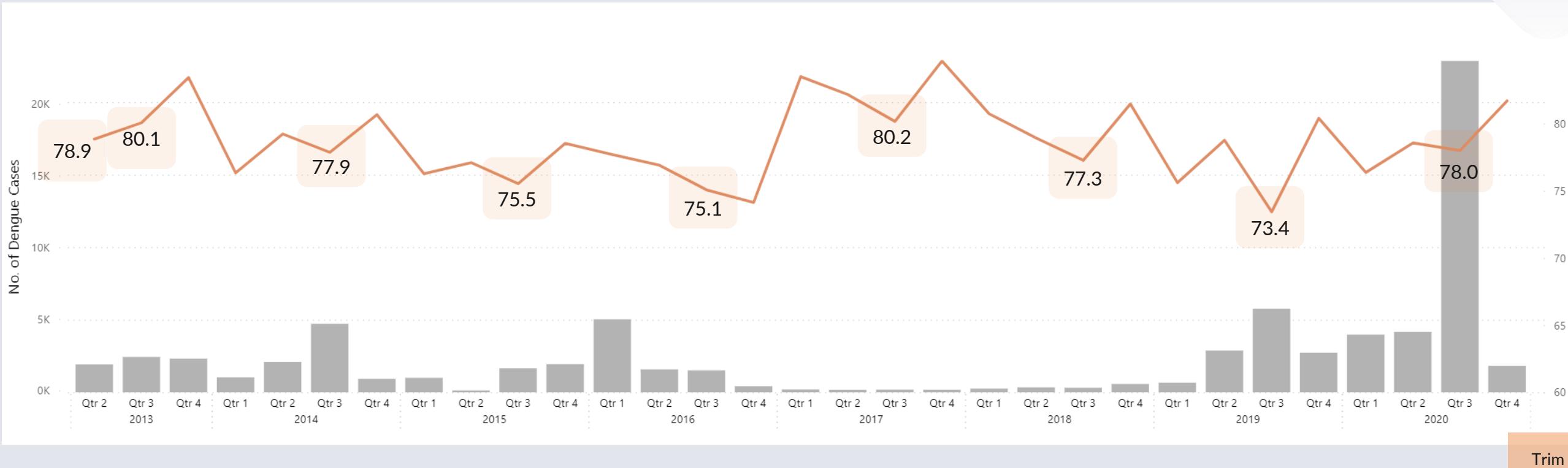
Poor coefficient of determination

Absolute Extreme Max Temperature



From 2013 – 2020, the absolute max temperature dropped in Q3 as compared to preceding quarters.

Mean Humidity



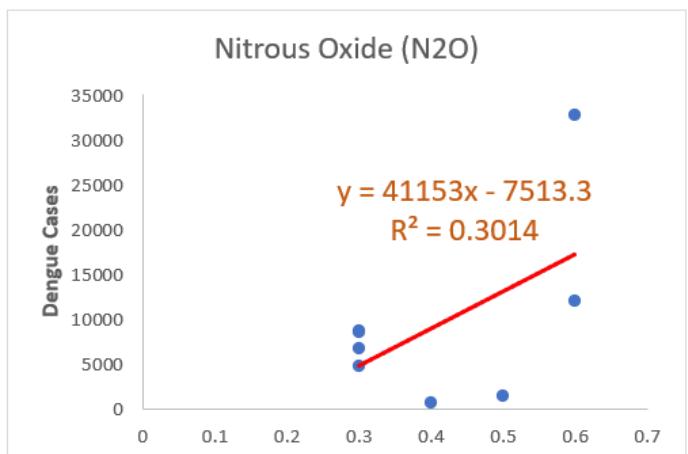
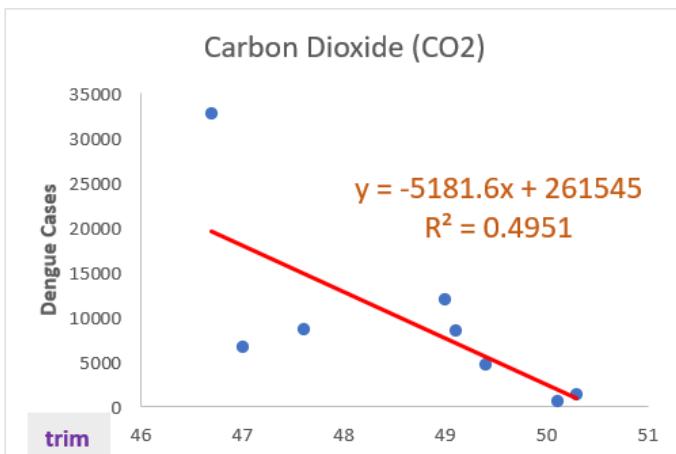
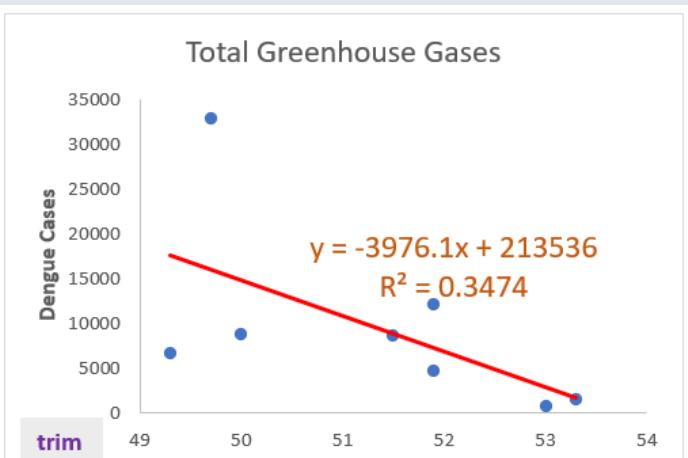
Mean humidity also declined in Q3 from 2014 – 2020.

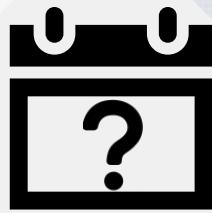




- Moderate to strong negative correlation between total greenhouse gases and carbon dioxide with the number of dengue cases

Greenhouse Gas Reading	Correlation Coefficient
Total Greenhouse Gase	(0.59)
Carbon Dioxide (CO2)	(0.70)
Methane (CH4)	0.00
Nitrous Oxide (N2O)	0.55
Hydrofluorocarbons (HFCs)	0.15
Perfluorocarbons (PFCs)	(0.18)
Sulphur Hexafluoride (SF6)	(0.03)
Nitrogen Trifluoride (NF3)	0.24





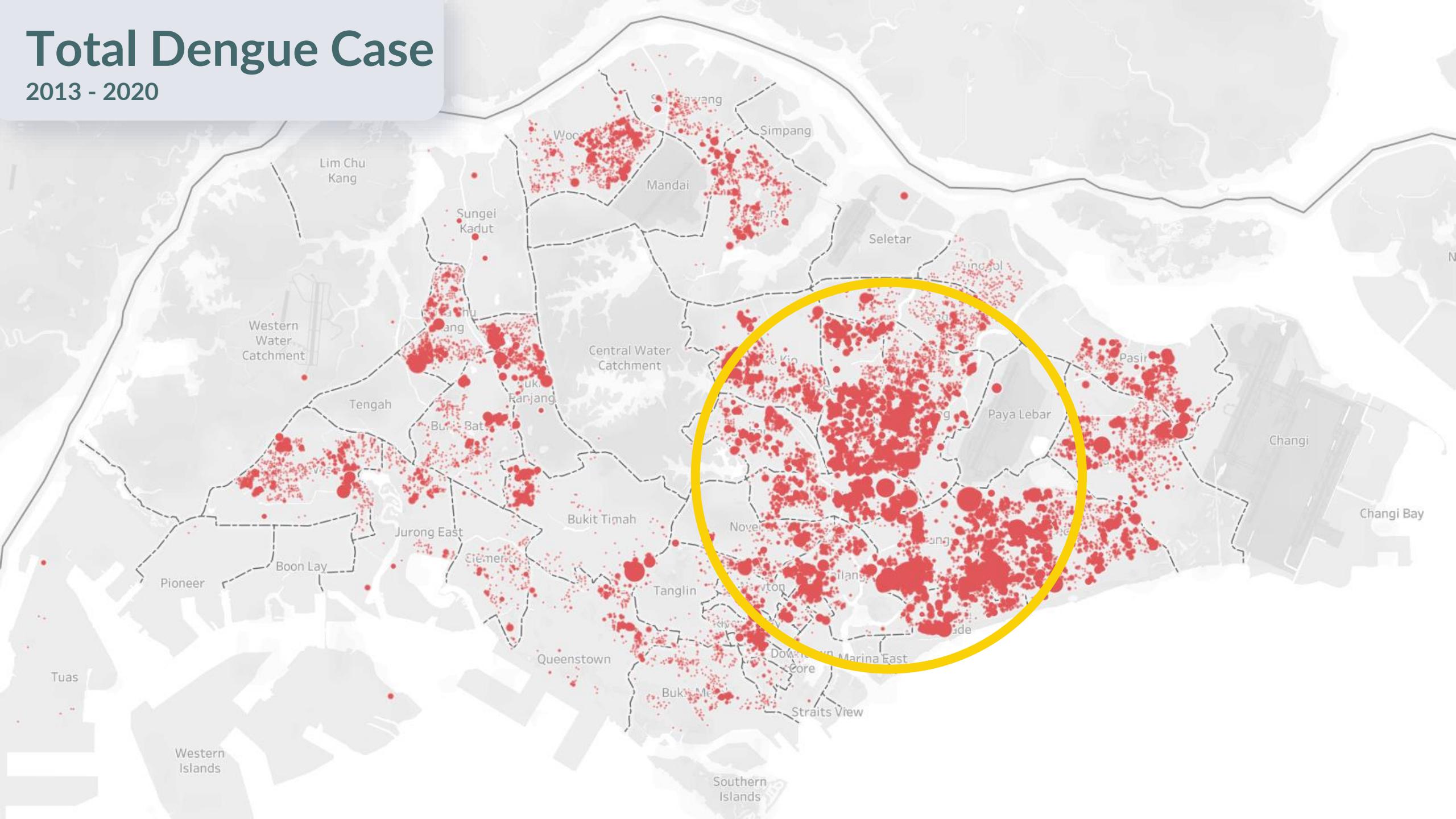
- **Unable to find a clear correlation** between weather patterns and high dengue cases in Singapore
- However, based on historical trends, the **number of cases tends to be higher** during the **third quarter**
- Therefore, it is recommended that authorities focus on dengue prevention measures during this period



WHERE

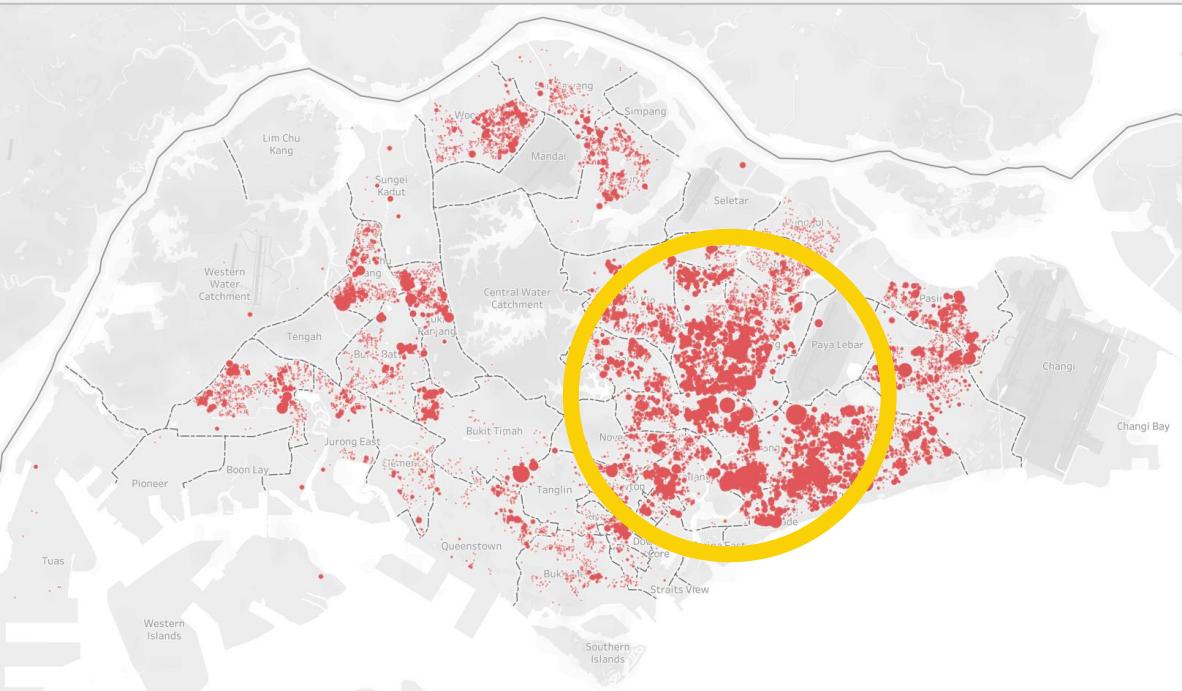
Total Dengue Case

2013 - 2020



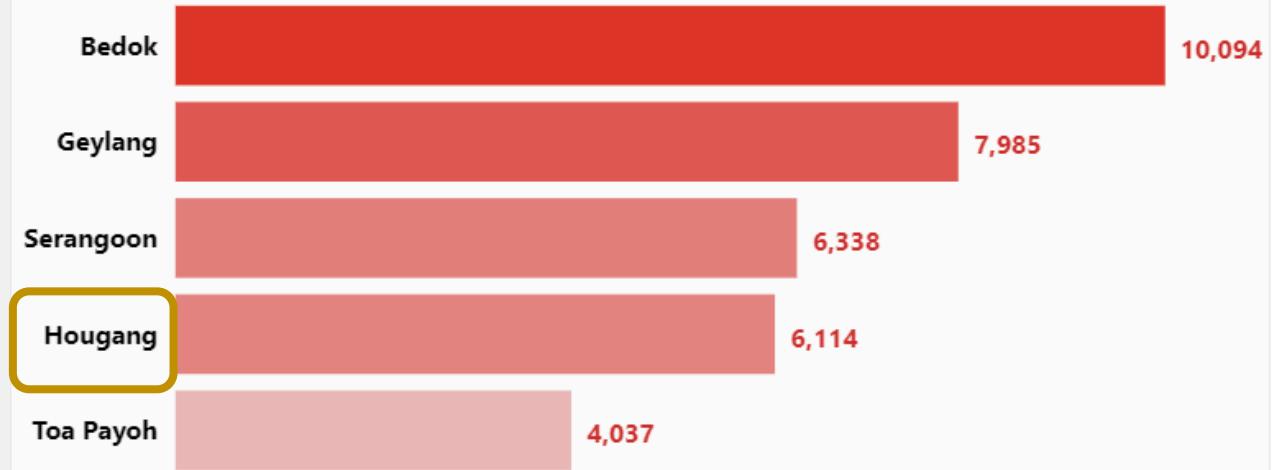
Total Dengue Case

2013 - 2020



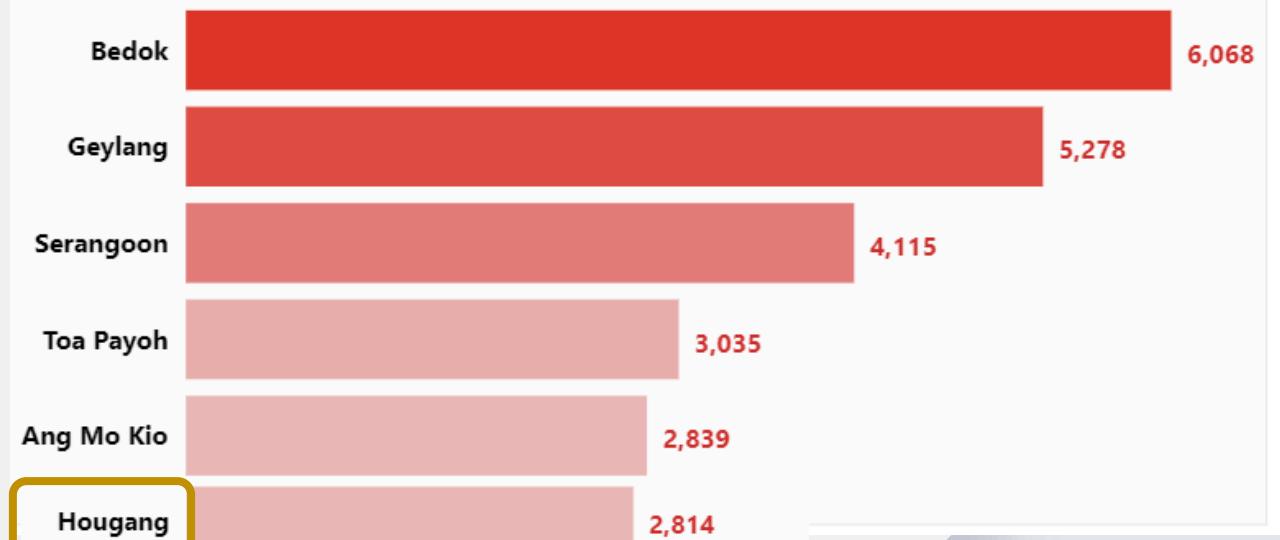
Top 5 Dengue Hotspots

From 2013 - 2020:



Top 5 Dengue Hotspots

From 2019 - 2020:

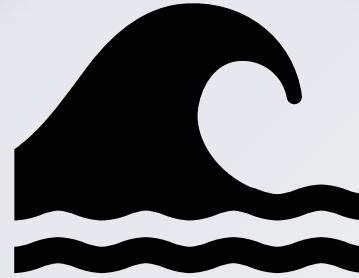




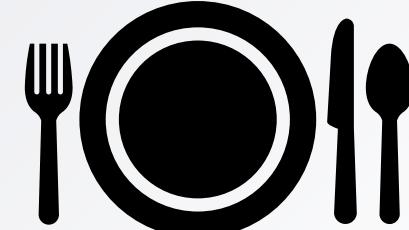
Population



Parks

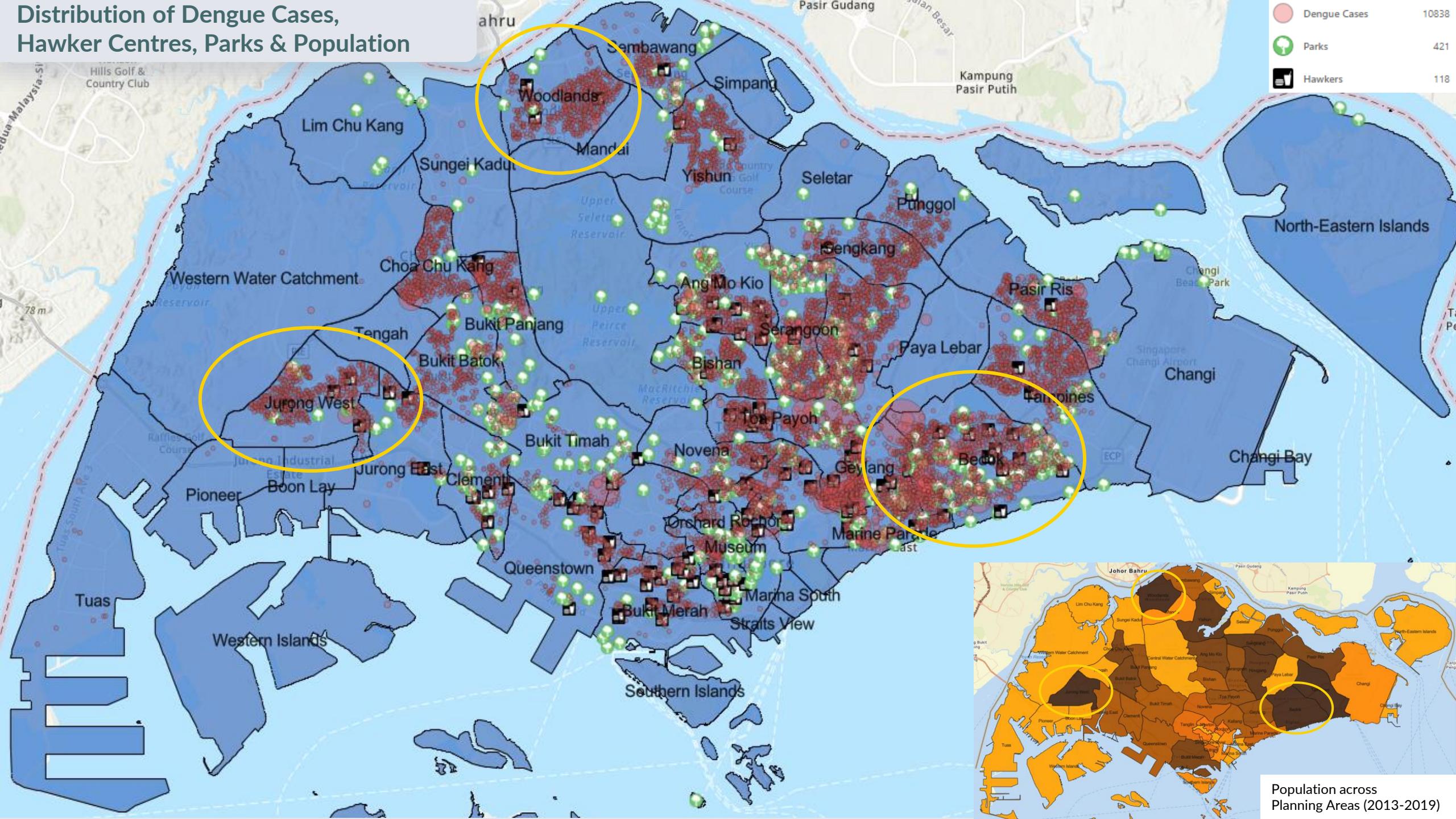


Water
Bodies



Hawker
Centres

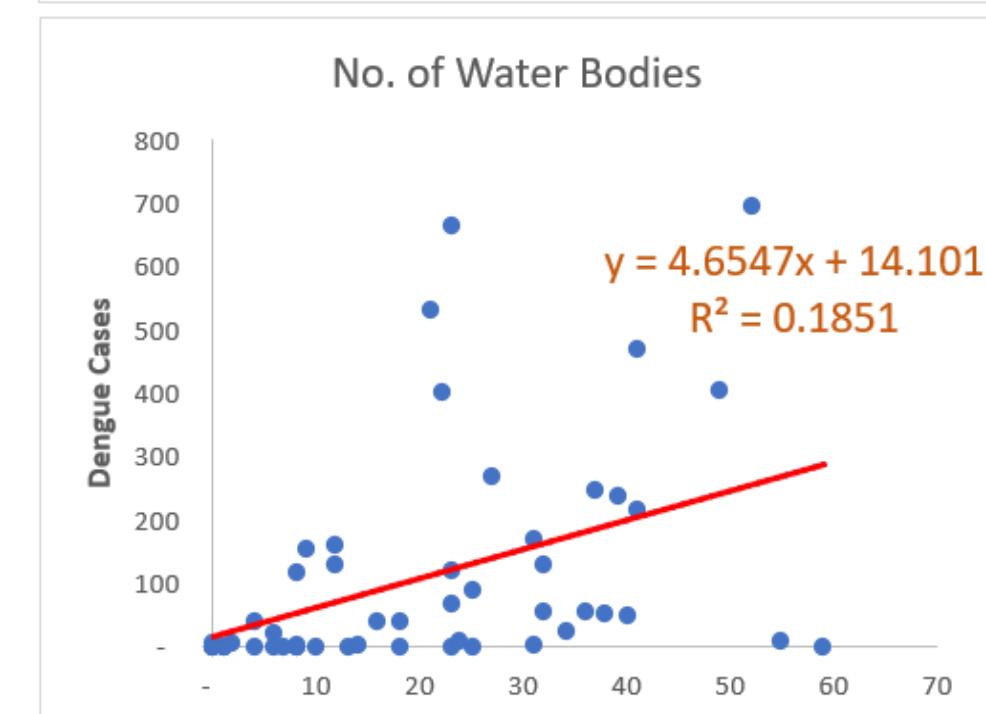
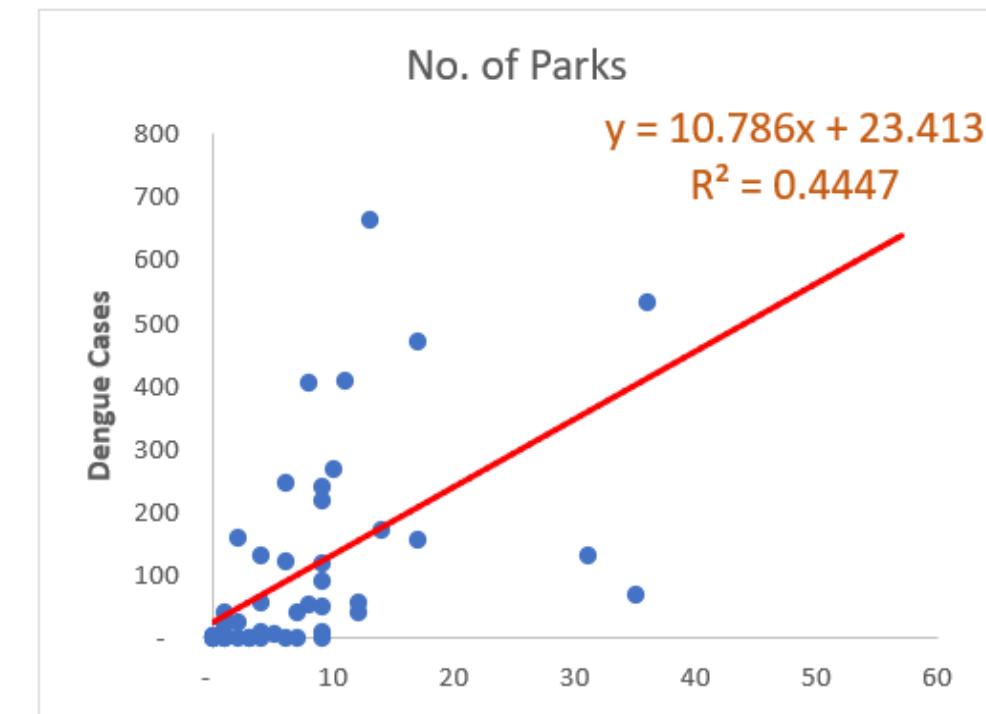
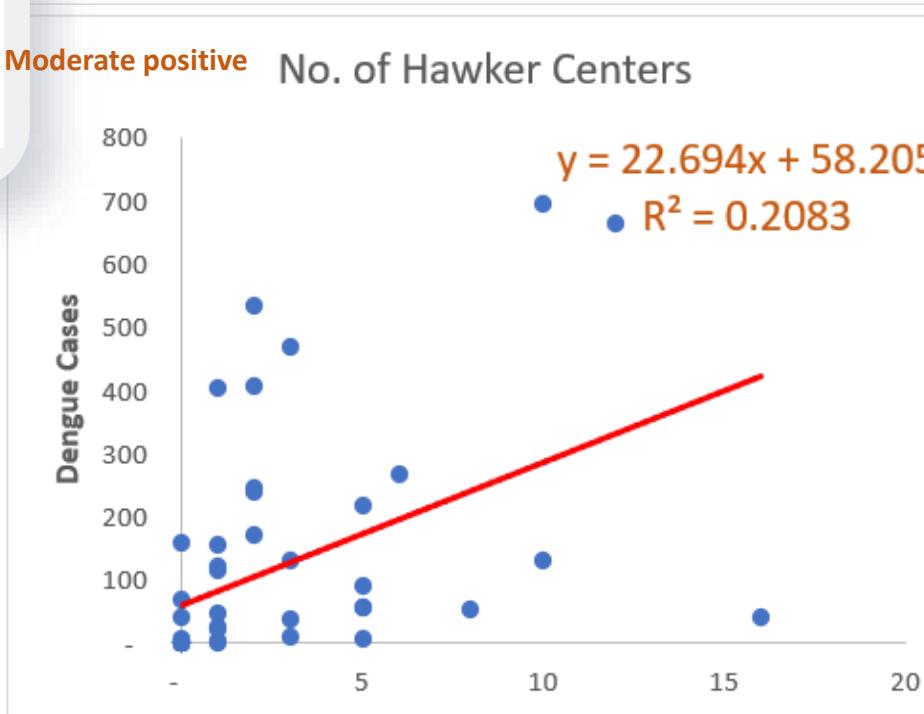
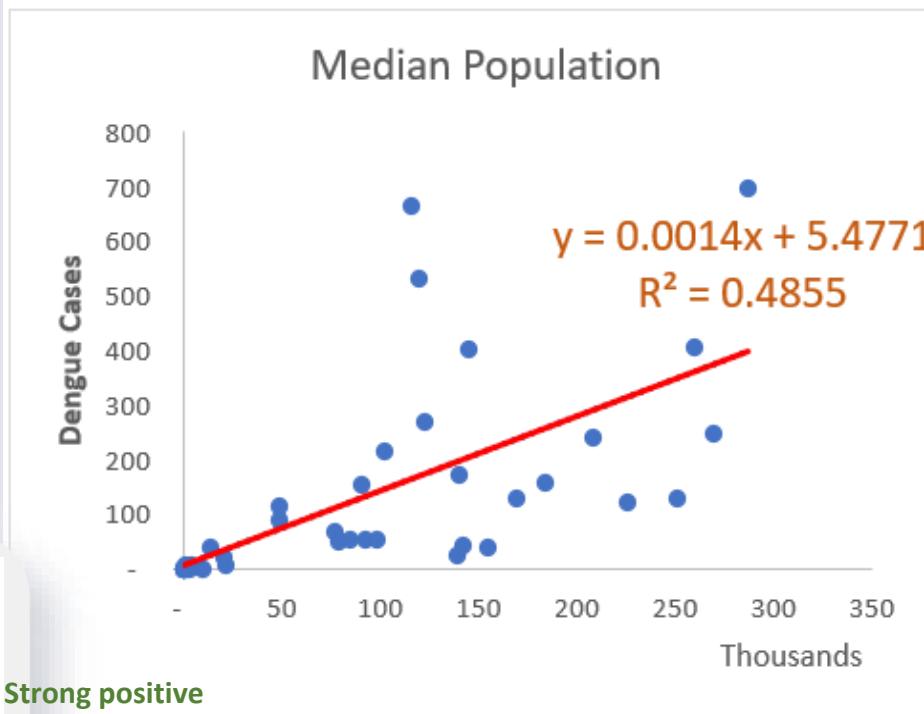
Distribution of Dengue Cases, Hawker Centres, Parks & Population



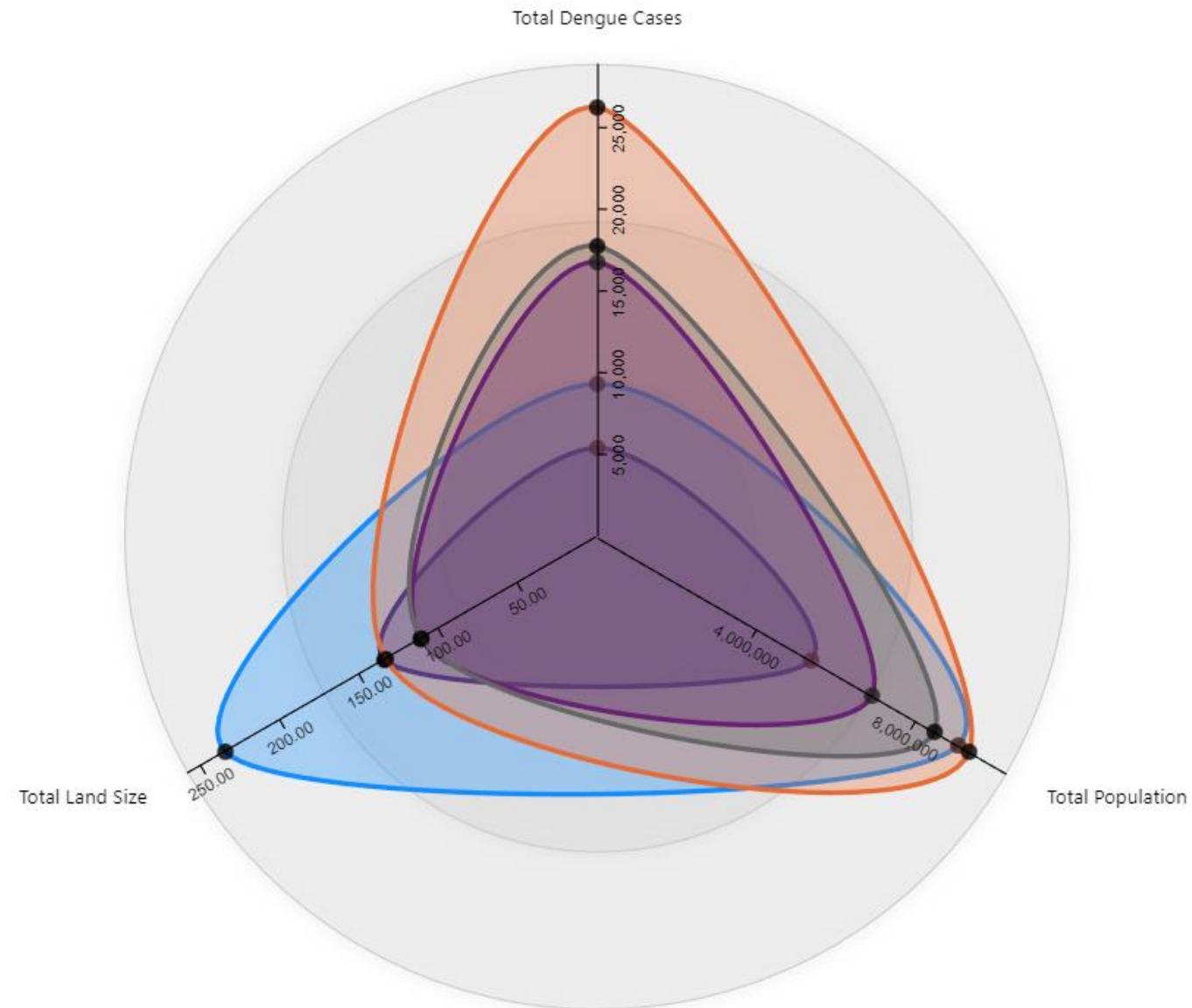
Population across
Planning Areas (2013-2019)

	Correlation Coefficient
Median Population	0.70
No. of parks	0.67
No. of hawker centers	0.46
No. of water bodies	0.43

* Correlation Coefficient is calculated at Planning Area aggregation level.



Region • West • North • Central • East • North-East



Clustering using K-Means Algorithm

*(To group clusters of Planning Areas together
based on distinct characteristics)*

Pre-Processing Steps

- 5 features used in the model:
 - Dengue Cases, Population, Hawker Centres, Parks and Size of Planning Area
- Group by Planning Area granularity:
 - Dengue Cases & Population – Sum per Year and Aggregate by Median
 - Hawker Centres & Parks – Aggregate by Count
- Apply Min-Max Normalisation to all features to normalise values between 0-1 so that each Feature contributes equally to the clustering process

K-Means Clustering

The screenshot shows a user interface for a machine learning model configuration. At the top, there is a section titled "Features (5)" with a "Filter" button. Below it is a list of five features, each preceded by a red square icon with a white letter "N": Median Dengue Cases, No. of Hawker Centers, Median Population, No. of Parks, and Area. Below this is a large empty white area. At the bottom, there is a "Target" section containing a single-line input field, and a "Metas (1)" section containing a single item: Planning Area, preceded by a blue square icon with a white letter "S".

Features used for model training :

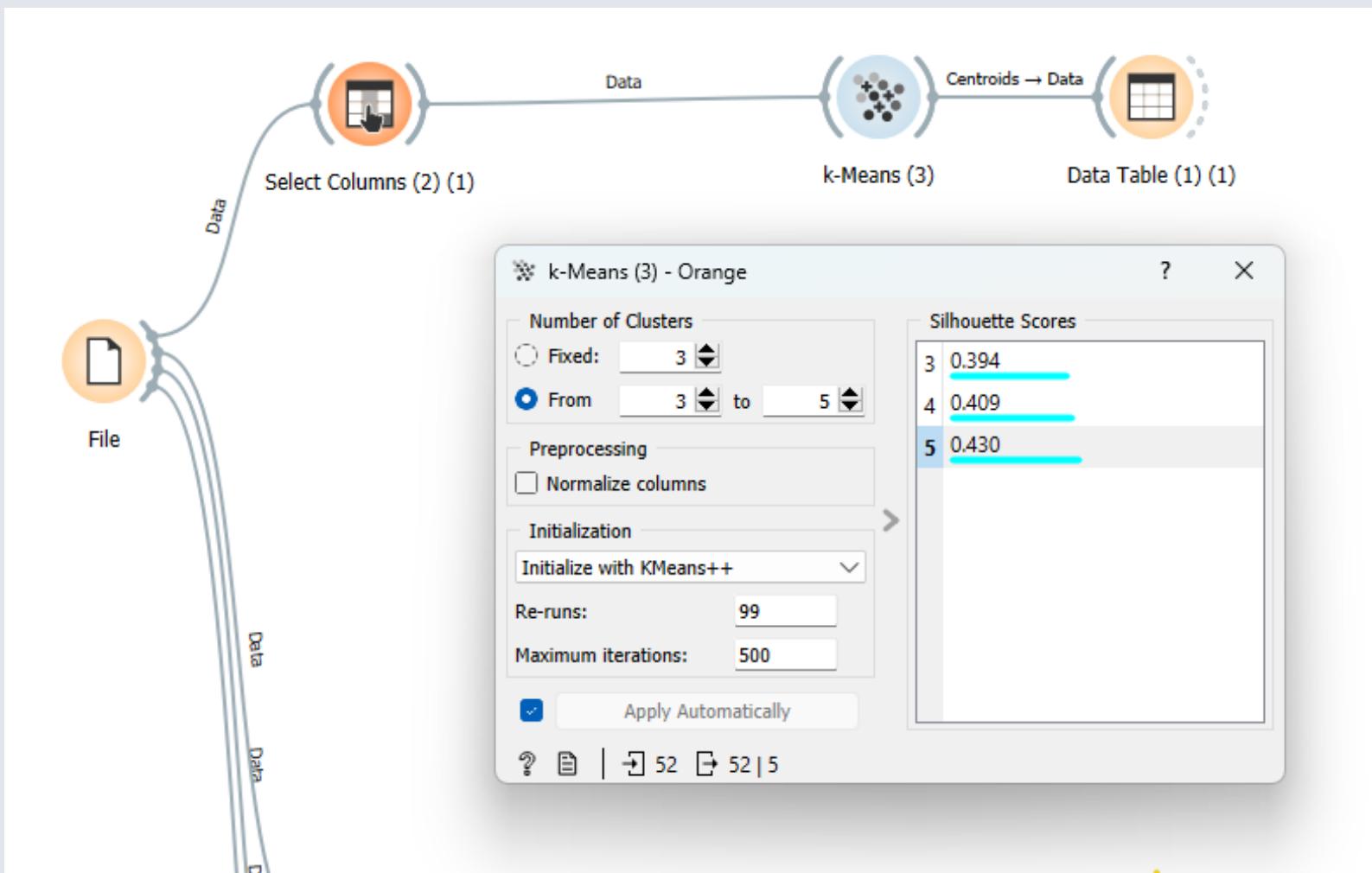
- Dengue Cases
- Population
- Parks
- Hawker Centres
- Planning Area Size

Features used to validate results:

- Planning Area

*All Features have been Normalised

K-Means Clustering



A 5 Cluster solution was chosen as the Silhouette Score was the highest (0.430 – Moderate Acceptance)

K-Means Clustering

	Silhouette	Area	No. of Hawker Centers	No. of Parks	Median Population	Median Dengue Cases
Cluster 1:	0.647291	0.629433	0.0125	0.0561404	0.00219136	0.00143678
Cluster 2:	0.687377	0.0784775	0.039773	0.0414673	0.0207082	0.013878
Cluster 3:	0.533447	0.16411	0.290179	0.224311	0.383086	0.145577
Cluster 4:	0.623383	0.199293	0.109375	0.138158	0.769691	0.390894
Cluster 5:	0.520412	0.18913	0.5000	0.619883	0.606643	0.90613

We will use the term '**amenities**' to describe
Parks and Hawker Centres as they are
used as recreational and social space

K-Means Clustering

	Silhouette	Area	No. of Hawker Centers	No. of Parks	Median Population	Median Dengue Cases
Cluster 1:	0.647291	0.629433	0.0125	0.0561404	0.00219136	0.00143678
Cluster 2:	0.687377	0.0784775	0.039773	0.0414673	0.0207082	0.013878
Cluster 3:	0.533447	0.16411	0.290179	0.224311	0.383086	0.145577
Cluster 4:	0.623383	0.199293	0.109375	0.138158	0.769691	0.390894
Cluster 5:	0.520412	0.18913	0.5000	0.619883	0.606643	0.90613

Key description:

- Cluster 1: **big** area, **low** amenities, **low** population, **low** dengue counts
- Cluster 2: **low** area, **low** amenities, **low** population, **low** dengue counts
- Cluster 3: moderate area, moderate amenities, moderate population, **low** dengue counts
- Cluster 4: moderate area, moderate amenities, **high** population, moderate dengue counts
- Cluster 5: moderate area, **high** amenities, **high** population, **high** dengue counts



Vast non-populated space



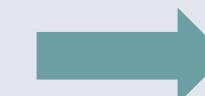
Small non-populated space



Low population density with **some** amenities



High population density with **some** amenities

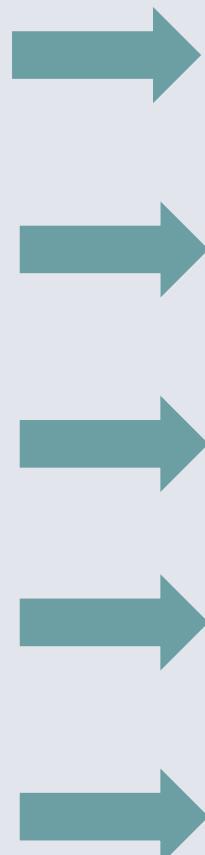


High population density with **many** amenities

K-Means Clustering

Key description:

Vast non-populated space
Small non-populated space
Low population density with some amenities
High population density with some amenities
High population density with many amenities



Label
Large nature reserves/ factory spaces
Small nature spots/ factory units/ shopping districts
Residential Districts
Dense residential districts
Dense residential districts with many amenities

K-Means Clustering

Key description:

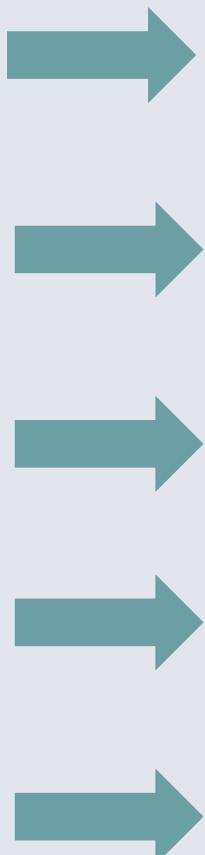
Vast non-populated space

Small non-populated space

Low population density with some amenities

High population density with some amenities

High population density with many amenities



Label	Dengue Cases
Large nature reserves/ factory spaces (e.g. Central Water Catchment, Tuas)	Low
Small nature spots/ factory units/ shopping districts (e.g. Lim Chu Kang, Tanglin)	Low
Residential Districts (e.g. Bishan, Sembawang)	Low - Medium
Dense residential districts (e.g. Choa Chu Kang, Woodlands)	Medium
Dense residential districts with many amenities (e.g. Bedok, Serangoon)	High

K-Means Clustering

Label	Dengue Cases
Large nature reserves/ factory spaces (e.g. Central Water Catchment, Tuas)	Low
Small nature spots/ factory units/ shopping districts (e.g. Lim Chu Kang, Tanglin)	Low
Residential Districts (e.g. Bishan, Sembawang)	Low - Medium
Dense residential districts (e.g. Choa Chu Kang, Woodlands)	Medium
Dense residential districts with many amenities (e.g. Bedok, Serangoon)	High

Conclusion:

1. Population Density has high influence on Dengue Cases
2. Among places with high Population Density, prevalence of recreational/social spaces makes a significant difference in Dengue Cases

Actions:

1. Implement more Gravitraps/ Wolbachia interventions in places with high Population Density
2. Conduct more Dengue Checks/ Surveillance/ Public Education Advertisements in places with many amenities (recreational/ social space)

K-Means Clustering

Label	Dengue Cases	Actions
Large nature reserves/ factory spaces (e.g. Central Water Catchment, Tuas)	Low	No Interventions, No Checks/Surveillance/Public Education Advertisements
Small nature spots/ factory units/ shopping districts (e.g. Lim Chu Kang, Tanglin)	Low	No Interventions, No Checks/Surveillance/Public Education Advertisements
Residential Districts (e.g. Bishan, Sembawang)	Low - Medium	Some Interventions, Some Checks/Surveillance/Public Education Advertisements
Dense residential districts (e.g. Choa Chu Kang, Woodlands)	Medium	Moderate Interventions, Moderate Checks/Surveillance/Public Education Advertisements
Dense residential districts with many amenities (e.g. Bedok, Serangoon)	High	Frequent Interventions, Frequent Checks/Surveillance/Public Education Advertisements

K-Means Clustering

	Planning Area	Silhouette	Median Dengue Cases	No. of Hawker Centers	Median Population	No. of Parks	Area
Cluster 1 (Large nature reserves/ factory spaces)	Tuas	0.573074	0.00215517	0.0000	0	0	0.425616
	Central Water Catchment	0.664365	0	0.0000	0	0.105263	0.529215
	Changi	0.672038	0	0.0625	0.00825249	0.122807	0.57963
	North-Eastern Islands	0.681516	0	0.0000	0	0.0526316	0.612706
	Western Water Catchment	0.645464	0.00502874	0.0000	0.00270431	0	1
Cluster 2 (Small nature spots/ factory units/ shopping districts)	Museum	0.688446	0	0.0000	0.00134343	0.157895	0
	Orchard	0.710211	0	0.0000	0.00289622	0.0350877	0.00189422
	Singapore River	0.710669	0.00359195	0.0000	0.0094389	0.0175439	0.00189422
	Outram	0.610272	0.0100575	0.3125	0.0740631	0.0877193	0.00786828
	River Valley	0.709812	0	0.0000	0.0329053	0.0175439	0.00947108
	Marina South	0.710597	0	0.0000	0	0	0.011511
	Rochor	0.670236	0.0560345	0.1875	0.0480145	0.0175439	0.011511
	Marina East	0.711912	0	0.0000	0	0.0175439	0.0144252
	Newton	0.70765	0.00718391	0.0625	0.0250017	0.0175439	0.0180679
	Downtown Core	0.660273	0.0122126	0.1875	0.0111138	0.157895	0.0511438
	Simpang	0.711882	0	0.0000	0	0	0.0626548
	Southern Islands	0.711364	0	0.0000	0.00226813	0	0.0763514
	Marine Parade	0.60595	0.167385	0.0625	0.167998	0.157895	0.07708
	Tengah	0.709402	0	0.0000	0	0	0.0957307
	Tanglin	0.701399	0.0308908	0.0625	0.0698409	0.0175439	0.099082
	Boon Lay	0.707558	0	0.0000	0	0	0.107825
	Seletar	0.702093	0.00359195	0.0000	0.000907251	0.0175439	0.137258
	Paya Lebar	0.694307	0.00215517	0.0000	0	0	0.15824
	Mandai	0.693403	0.000718391	0.0000	0.00722311	0.0526316	0.159406
	Pioneer	0.691873	0	0.0000	0	0	0.164214
	Sungei Kadut	0.659194	0.0114943	0.0000	0.00246005	0.0701754	0.220895
	Lim Chu Kang	0.643798	0	0.0000	0.000104683	0.0701754	0.239983
Cluster 3 (Residential Districts)	Bishan	0.532799	0.221983	0.0625	0.315566	0.298246	0.0989363
	Toa Payoh	0.544856	0.385057	0.3750	0.42831	0.175439	0.10695
	Novena	0.474952	0.130029	0.3125	0.169063	0.157895	0.118753
	Bukit Panjang	0.518901	0.246408	0.1250	0.486793	0.245614	0.118898
	Kallang	0.588874	0.311063	0.3125	0.35545	0.157895	0.121521
	Punggol	0.514534	0.0344828	0.0625	0.484367	0.0350877	0.123998
	Clementi	0.575139	0.0797414	0.3125	0.321673	0.210526	0.126184
	Bukit Batok	0.532326	0.0596264	0.0000	0.495778	0.210526	0.15008
	Sembawang	0.434609	0.0696839	0.0625	0.274391	0.157895	0.167711
	Ang Mo Kio	0.548325	0.186063	0.6250	0.591301	0.54386	0.191024
	Bukit Merah	0.55495	0.058908	1.0000	0.538087	0.122807	0.196853
	Bukit Timah	0.521693	0.0984195	0.0000	0.265842	0.614035	0.243334
	Jurong East	0.537705	0.0797414	0.3125	0.294089	0.0701754	0.247705
	Queenstown	0.588591	0.0768678	0.5000	0.342487	0.140351	0.285589
Cluster 4 (Dense residential districts)	Choa Chu Kang	0.582263	0.229167	0.0000	0.639438	0.0350877	0.0769343
	Sengkang	0.622768	0.173132	0.0625	0.788593	0.105263	0.142212
	Woodlands	0.629246	0.188218	0.1875	0.874276	0.0701754	0.185925
	Hougang	0.620379	0.674569	0.1875	0.780114	0.298246	0.190879
	Jurong West	0.660286	0.354885	0.1250	0.939877	0.105263	0.201952
	Pasir Ris	0.58136	0.579023	0.0625	0.503455	0.140351	0.206761
	Tampines	0.65291	0.584052	0.1250	0.904477	0.192982	0.292292
	Yishun	0.637855	0.344109	0.1250	0.727301	0.157895	0.297392
Cluster 5 (Dense residential districts with many amenities)	Geylang	0.52548	0.953305	0.7500	0.401633	0.22807	0.12837
	Serangoon	0.466744	0.765086	0.1250	0.418295	0.631579	0.135072
	Bedok	0.569013	1	0.6250	1	1	0.303949

Dengue Monitoring Dashboard

Period

Region

2020

AI

Total Dengue Cases:

Total No. of Clusters:

32,802

6,184



2,592

Clusters with 0 new cases

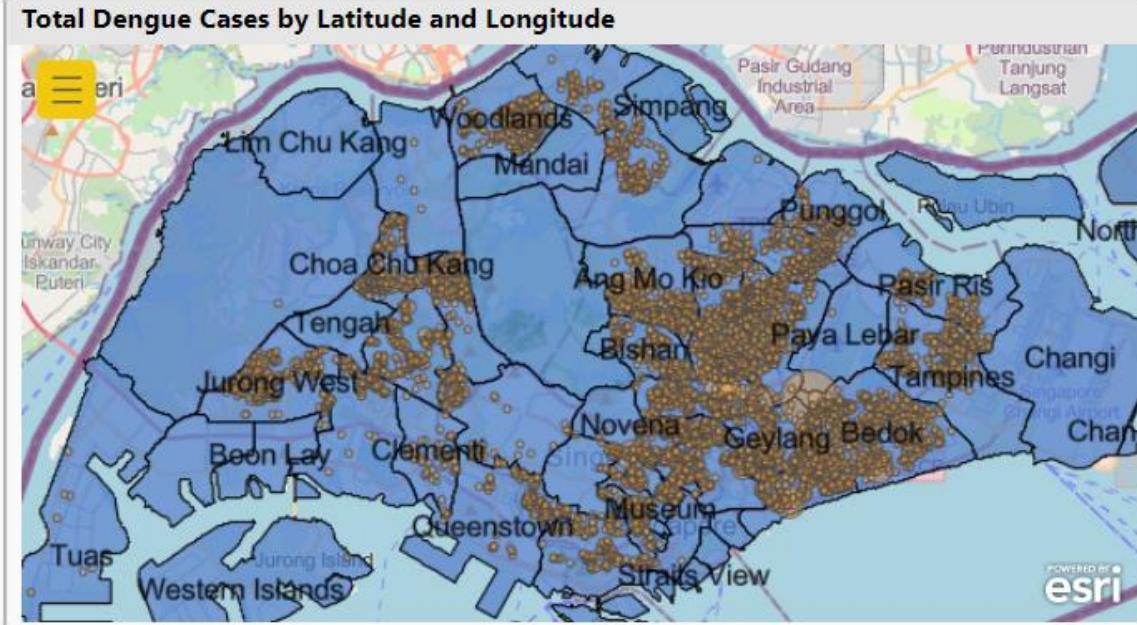
3,302

Clusters with 1 to 9 cases

3,124

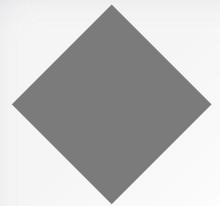
Clusters with 10 or more cases

Planning Area	Address	Date	Total Dengue Cases	Total Cluster Cases
Bedok	Kaki Bukit Industrial Terrace	18/9/2020	104	0
Bedok	Kaki Bukit Industrial Terrace	7/8/2020	99	104
Toa Payoh	Shaw Road	6/11/2020	85	1
Toa Payoh	Youngberg Terrace (Avon Park)	3/7/2020	84	0
Bedok	Jalan Sempadhan (Villa Marina)	18/9/2020	65	5
Bedok	Kaki Bukit Industrial Terrace	3/7/2020	63	203
Toa Payoh	Bidadari Park Drive	3/7/2020	60	11
Serangoon	Kensington Park Drive (Kensington Park Condominium)	7/8/2020	60	5
Ang Mo Kio	Thomson Green	18/9/2020	56	8
Toa Payoh	Woodleigh Link	3/7/2020	56	0
Newton	Bukit Timah Road (Goodwood Residence)	18/9/2020	55	18
Newton	Bukit Timah Road (Goodwood Residence)	7/8/2020	51	88
Bedok	Jalan Sempadhan (Villa Marina)	7/8/2020	51	74
River Valley	River Valley Road (Aspen Heights)	18/9/2020	50	28



When using this dashboard, there are 2 points to note:

- According to NEA's definition , each cluster case is defined as any case within 150 metres from the point of origin, not including its own location case, within the past 14 days.
 - Each latitude and longitude corresponds to 1 address and each address is considered as 1 cluster, regardless of the number of dengue or cluster cases.



Retrospective & Conclusion

Dengue Cases Spread Over Singapore (2013 - 2020)

10/6/2020 11/6/2020

Region

Central

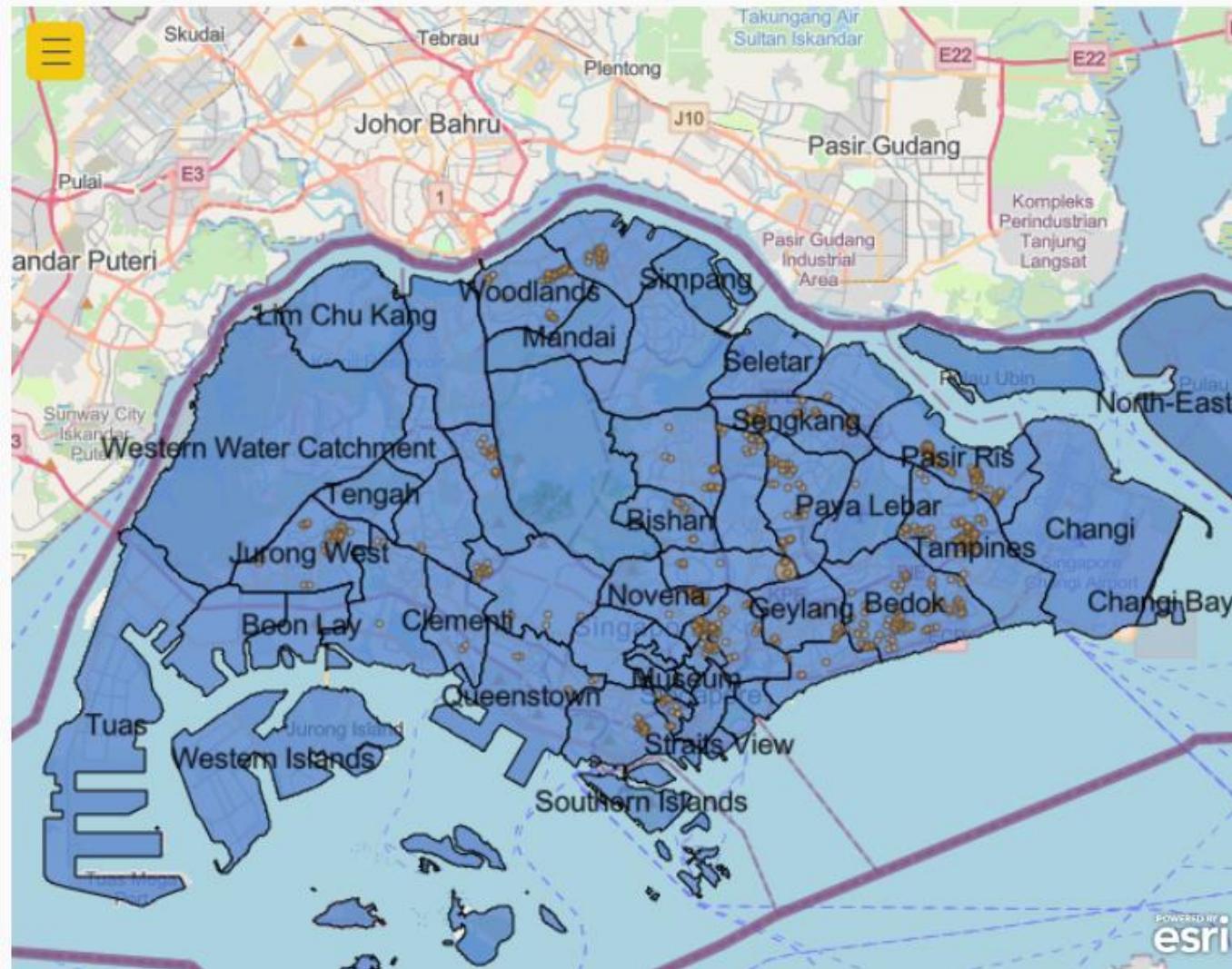
East

North

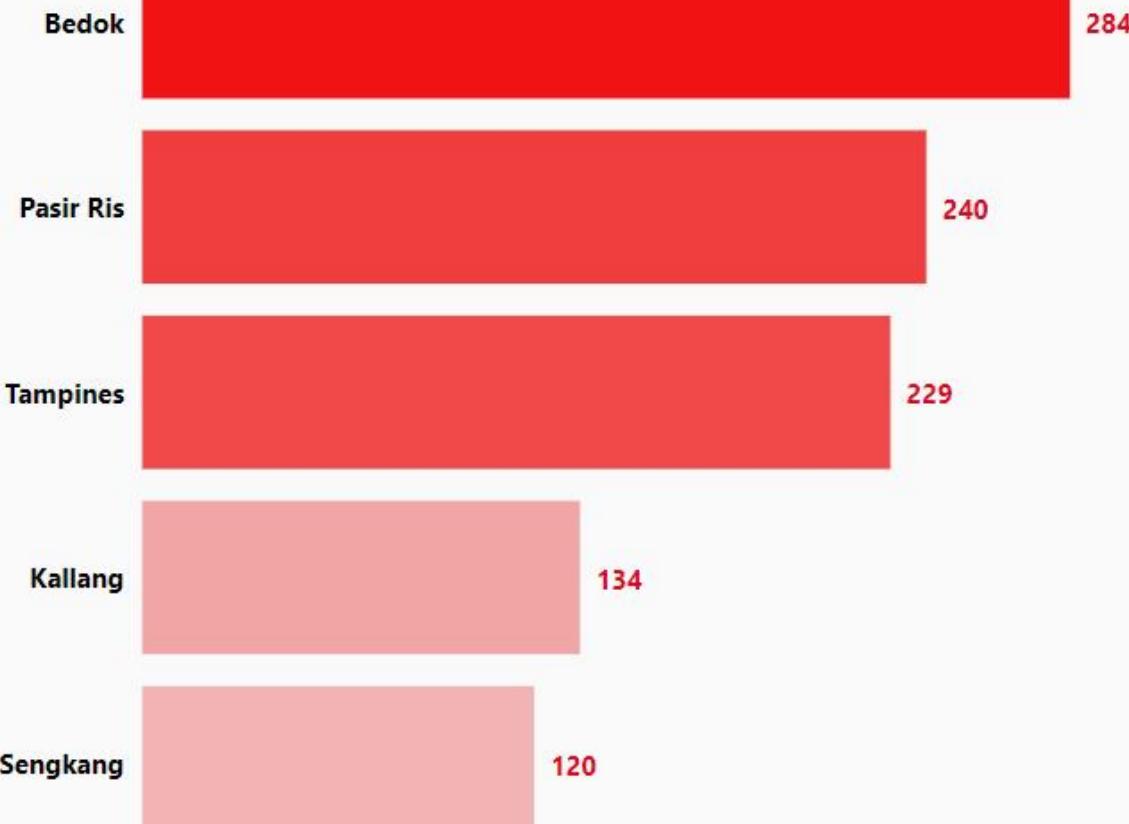
North-East

West

Total Dengue Cases by Latitude and Longitude



Top 5 Dengue Hotspots





DATA SECURITY

Data Anonymization/Masking

Dengue addresses should be anonymized to protect the privacy of individuals

Data Access and Lineage

Designated visuals for different stakeholders: Admin and management to view detailed data, while the public can only view aggregated data.

Upstream (Live sources) to Downstream (Dashboard/Analysis)

Regulation Compliance

Ensure that the data security setup and data analysis practices comply with Singapore government regulations such as the Personal Data Protection Act (PDPA) and the Computer Misuse and Cybersecurity Act

GANNT CHART

				January				February				March				April				May	
Action Item	Start Date	End Date	Man Day efforts	9 Jan	16 Jan	23 Jan	30 Jan	6 Feb	13 Feb	20 Feb	27 Feb	6 Mar	13 Mar	20 Mar	27 Mar	3 Apr	10 Apr	17 Apr	24 Apr	1 May	8 May
Project topic research & gathering topics	14/1/2023	19/3/2023	22																		
Data collection	21/1/2023	4/2/2023	5																		
Data preparation & transformation	28/1/2023	11/3/2023	14																		
Exploratory Data Analysis	4/2/2023	11/3/2023	12																		
Predictive Model Development	11/3/2023	25/3/2023	5																		
Model Evaluation	25/3/2023	1/4/2023	3																		
Visualisation and Dashboard	4/2/2023	19/3/2023	15																		
Maintenance and refinement	18/3/2023	15/4/2023	10																		
Project Demo Slide Preparation & Rehearsal	25/3/2023	29/4/2023	12																		
Project Demo	29/4/2023	29/4/2023	1																		

AGILE RETROSPECTIVE

What Worked Well?

- ✓ Weekly meeting
- ✓ Teamwork
- ✓ Delivery on time



What Puzzled Us?

- ❖ Inception Phase
- ❖ Data Quality & Preparation
- ❖ Scope Creep



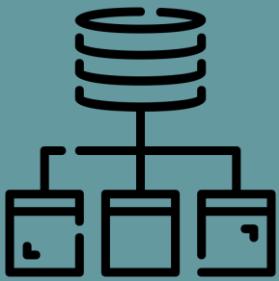


FUTURE WORK

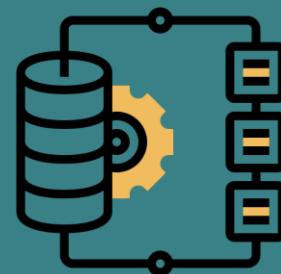
Data Quality



Modelling



Integration



Complexity



Data Quality



Limited publicly accessible data from online, analysis accuracy can be improved from a more reliable data source

Modelling



Integration

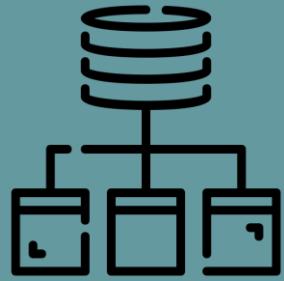


Complexity



Modelling

Data Quality



More advanced machine learning algorithm, run prediction for future dengue cases

Integration



Complexity



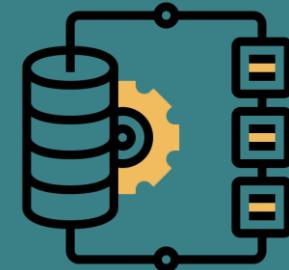
Data Quality



Modelling



Integration



Complexity



Form a data integration
data lake, data warehouse,
for a ETL process Flow

Data Quality



Modelling



Integration



Complexity



Expand the Analytic work to other countries in the region and even leverage to worldwide, to have a more comprehensive analysis of dengue cases

CONCLUSION

- Identified the trends and key contributors for rising dengue cases in Singapore
- Data sourced from NEA and MOH
- Involved various data analysis techniques
 - Data collection & cleaning
 - Exploratory data analysis
 - Data modelling: K-Means clustering, correlation/regression analysis
 - Dictionary & profiling
 - Data visualization using BI tools: PowerBI, Tableau
 - Data security
- Provided valuable insights for government and healthcare organizations to take proactive measures for controlling the spread of dengue
- Established a performance tracker for past initiatives evaluation