Group 3 Project 4

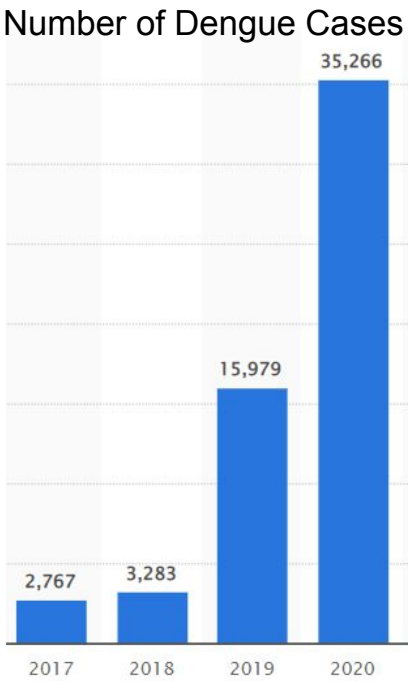
Jance Ng

Jin Min Wood

Adrian Kong

Problem Statement

- Number of Dengue cases increased rapidly from 2017 to 2020.
- Mild dengue fever → high fever and flu-like symptoms.
- Dengue hemorrhagic fever → serious bleeding, a sudden drop in blood pressure (shock) and death.



Assumptions

- Aedes mosquitoes breed in clean, stagnant water.
- Breeding of Aedes mosquito influenced by weather.
- Google trend might be good predictors.

Dataset

Singapore daily climate historical data

(http://www.weather.gov.sg/climate-historical-daily/)

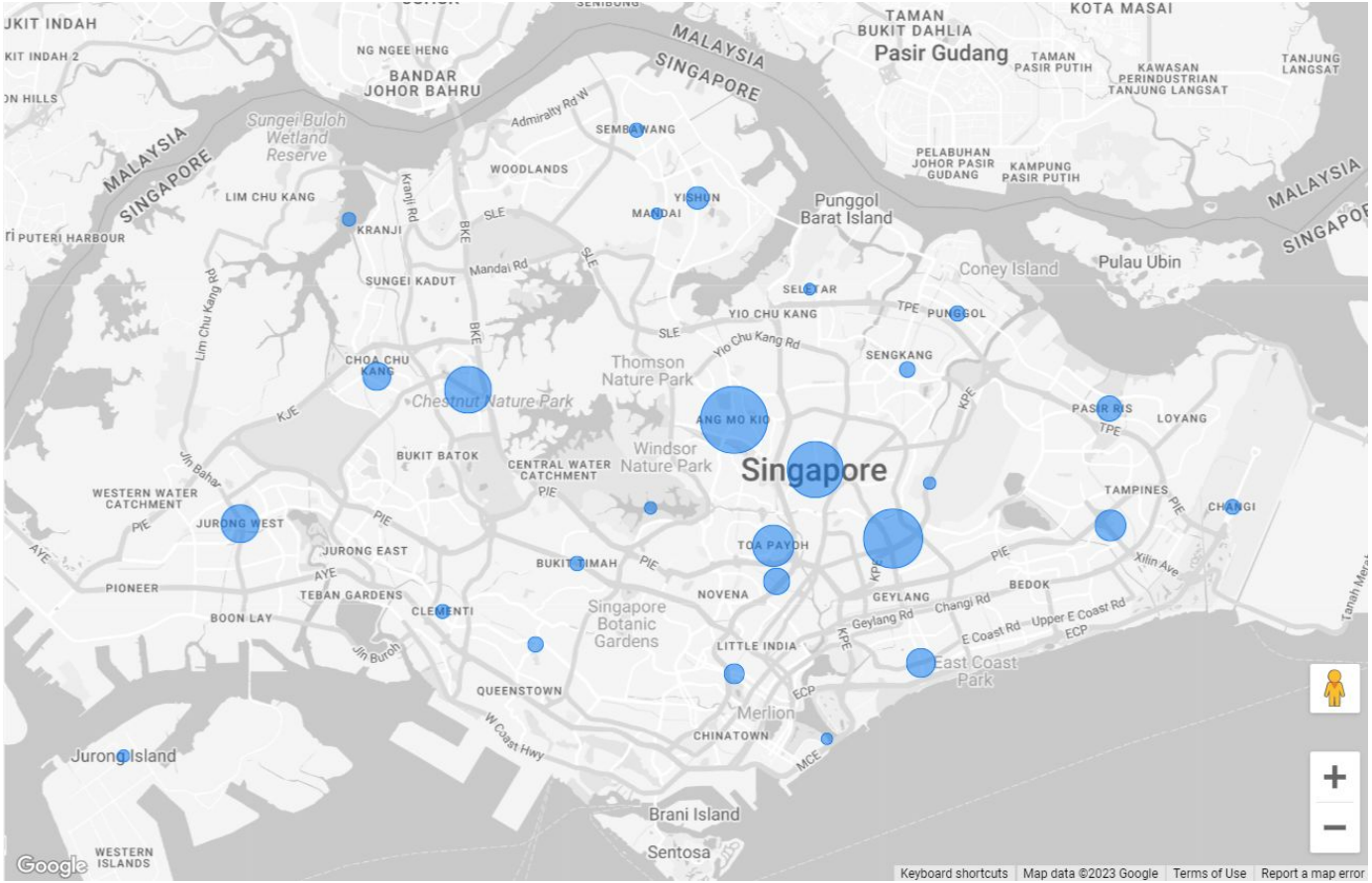
Google trend of various search term

(https://trends.google.com/trends/explore?date=today%205-y&geo=SG&q=%2Fm%2F09wsg)

Dengue clusters in Singapore

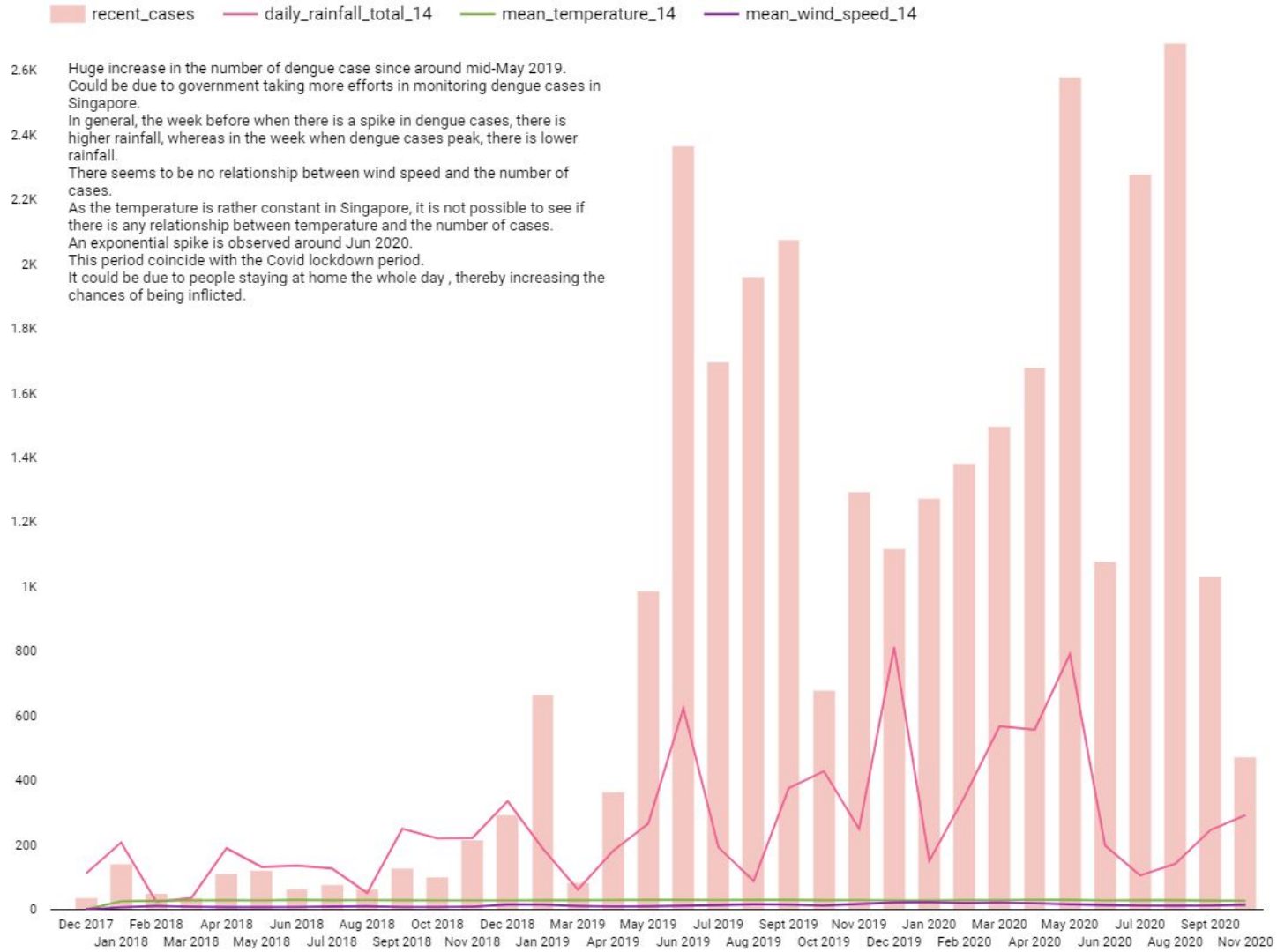
(https://outbreak.sgcharts.com/data)

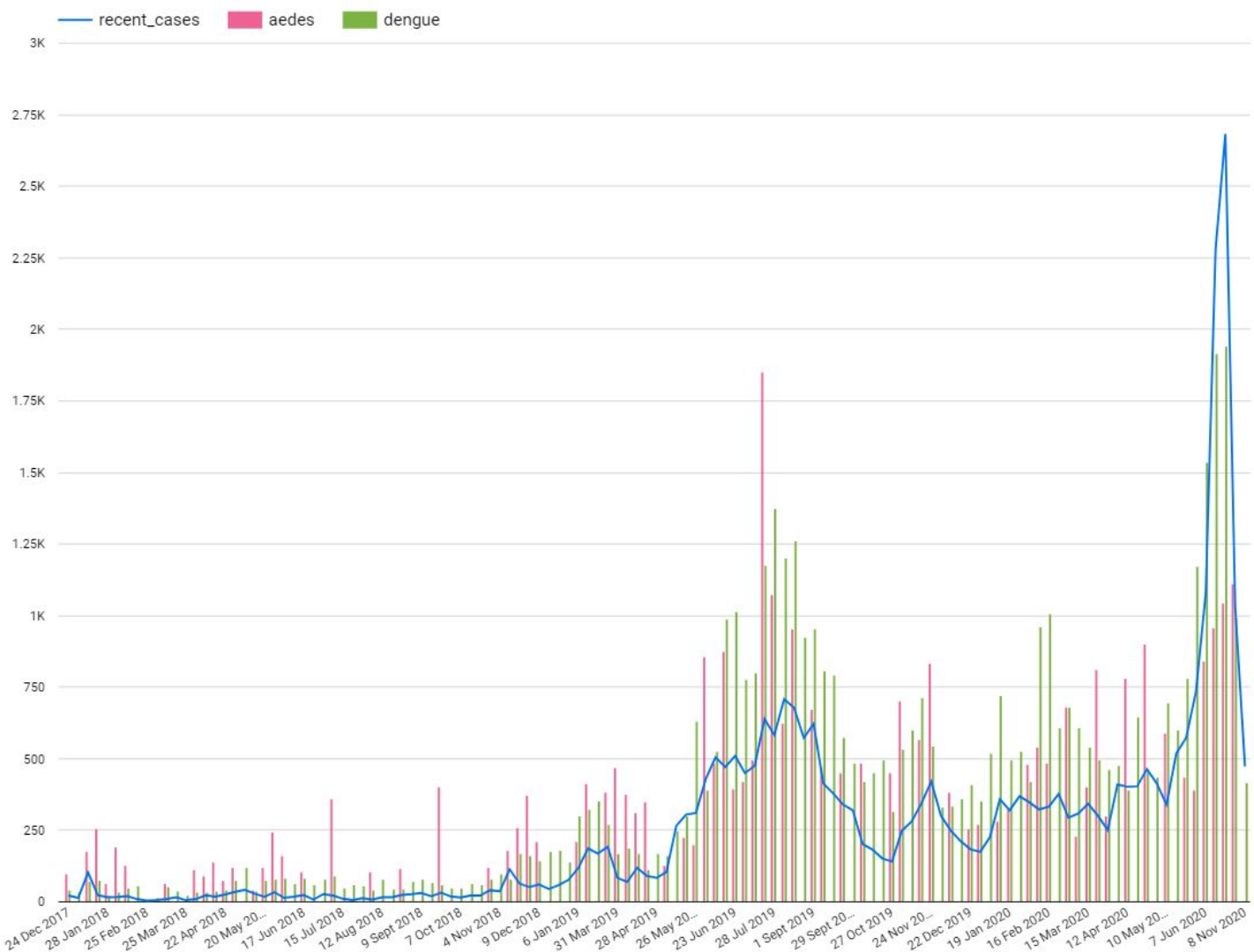
EDA


Singapore
MALAYSIA
SINGAPORE
BANDAR JOHOR BAHRU
Sungei Buloh Wetland Reserve
WOODLANDS
SEMBAWANG
YISHUN
MANDAI
LIM CHU KANG
KRANJI
SUNGEI KADUT
Punggol Barat Island
Coney Island
Pulau Ubin
PUNGGOL
SELETAR
YIO CHU KANG
SENGKANG
Thomson Nature Park
Chestnut Nature Park
ANG MO KIO
CHOA CHU KANG
BUKIT BATOK
Windsor Nature Park
CENTRAL WATER CATCHMENT
WESTERN WATER CATCHMENT
JURONG WEST
JURONG EAST
BUKIT TIMAH
TOA PAYOH
NOVENA
GEYLANG
BEDOK
PASIR RIS
TAMPINES
LOYANG
CHANGI
PIONEER
TEBAN GARDENS
BOON LAY
CLEMENTI
Singapore Botanic Gardens
LITTLE INDIA
East Coast Park
QUEENSTOWN
CHINATOWN
Merlion
Jurong Island
Brani Island
Sentosa
WESTERN ISLANDS
PELABUHAN JOHOR PASIR GUDANG
KAMPUNG PASIR PUTIH
TAMAN BUKIT DAHLIA
Pasir Gudang
TAMAN PASIR PUTIH
KAWASAN PERINDUSTRIAN TANJUNG LANGSAT
TANJUNG LANGSAT
KOTA MASAI
Google
Keyboard shortcuts
Map data ©2023 Google
Terms of Use
Report a map error

Dengue Cases - Top 10 Regions

region	
Chai Chee	4793.0
Ang Mo Kio	3341.0
Tai Seng	2859.0
Serangoon	2730.0
Bukit Panjang	2127.0
Admiralty	1837.0
Toa Payoh	1804.0
Jurong (West)	1616.0
Simei	1229.0
Yishun	1154.0

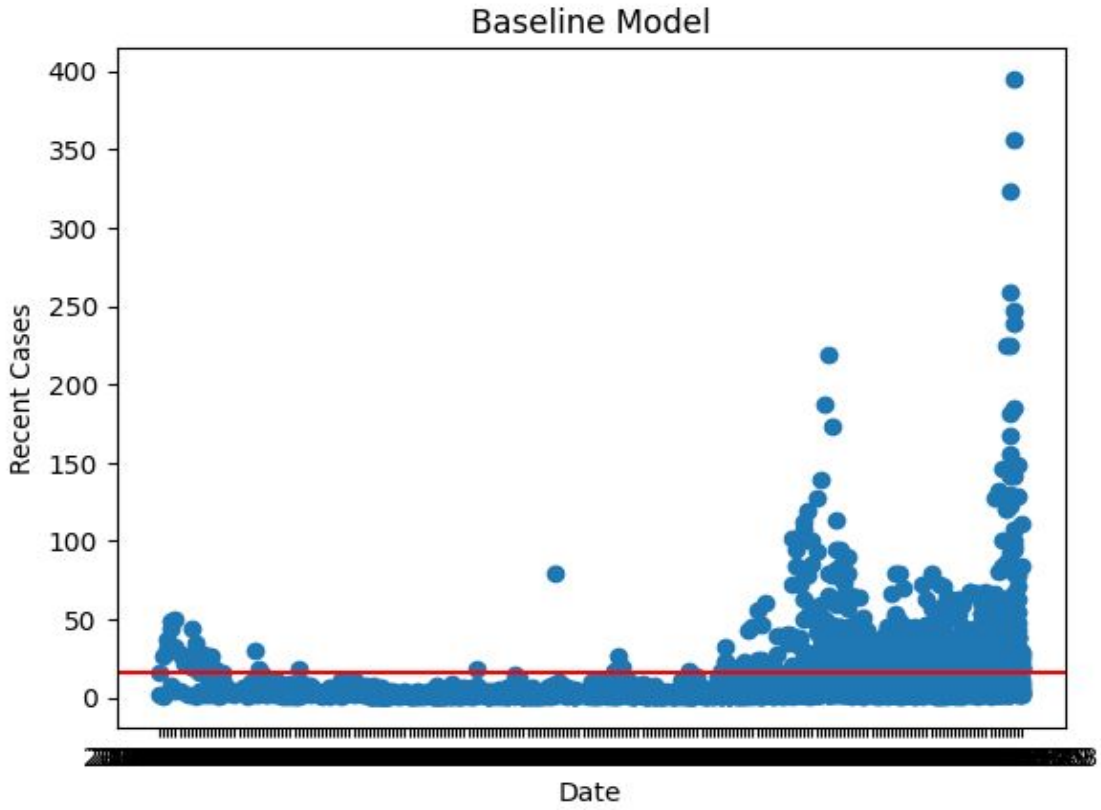

recent_cases
daily_rainfall_total_14
mean_temperature_14
mean_wind_speed_14
Huge increase in the number of dengue case since around mid-May 2019.
Could be due to government taking more efforts in monitoring dengue cases in Singapore.
In general, the week before when there is a spike in dengue cases, there is higher rainfall, whereas in the week when dengue cases peak, there is lower rainfall.
There seems to be no relationship between wind speed and the number of cases.
As the temperature is rather constant in Singapore, it is not possible to see if there is any relationship between temperature and the number of cases.
An exponential spike is observed around Jun 2020.
This period coincide with the Covid lockdown period.
It could be due to people staying at home the whole day , thereby increasing the chances of being inflicted.
2.6K
2.4K
2.2K
2K
1.8K
1.6K
1.4K
1.2K
1K
800
600
400
200
0
Dec 2017
Jan 2018
Feb 2018
Mar 2018
Apr 2018
May 2018
Jun 2018
Jul 2018
Aug 2018
Sept 2018
Oct 2018
Nov 2018
Dec 2018
Jan 2019
Mar 2019
Apr 2019
May 2019
Jun 2019
Jul 2019
Aug 2019
Sept 2019
Oct 2019
Nov 2019
Dec 2019
Jan 2020
Feb 2020
Mar 2020
Apr 2020
May 2020
Jun 2020
Jul 2020
Aug 2020
Sept 2020
Nov 2020


recent_cases
aedes
dengue
3K
2.75K
2.5K
2.25K
2K
1.75K
1.5K
1.25K
1K
750
500
250
0
24 Dec 2017
28 Jan 2018
25 Feb 2018
25 Mar 2018
22 Apr 2018
20 May 20...
17 Jun 2018
15 Jul 2018
12 Aug 2018
9 Sept 2018
7 Oct 2018
4 Nov 2018
9 Dec 2018
6 Jan 2019
31 Mar 2019
28 Apr 2019
26 May 20...
23 Jun 2019
28 Jul 2019
1 Sept 2019
29 Sept 20...
27 Oct 2019
24 Nov 20...
22 Dec 2019
19 Jan 2020
16 Feb 2020
15 Mar 2020
12 Apr 2020
10 May 20...
7 Jun 2020
8 Nov 2020

Data Preparation Steps

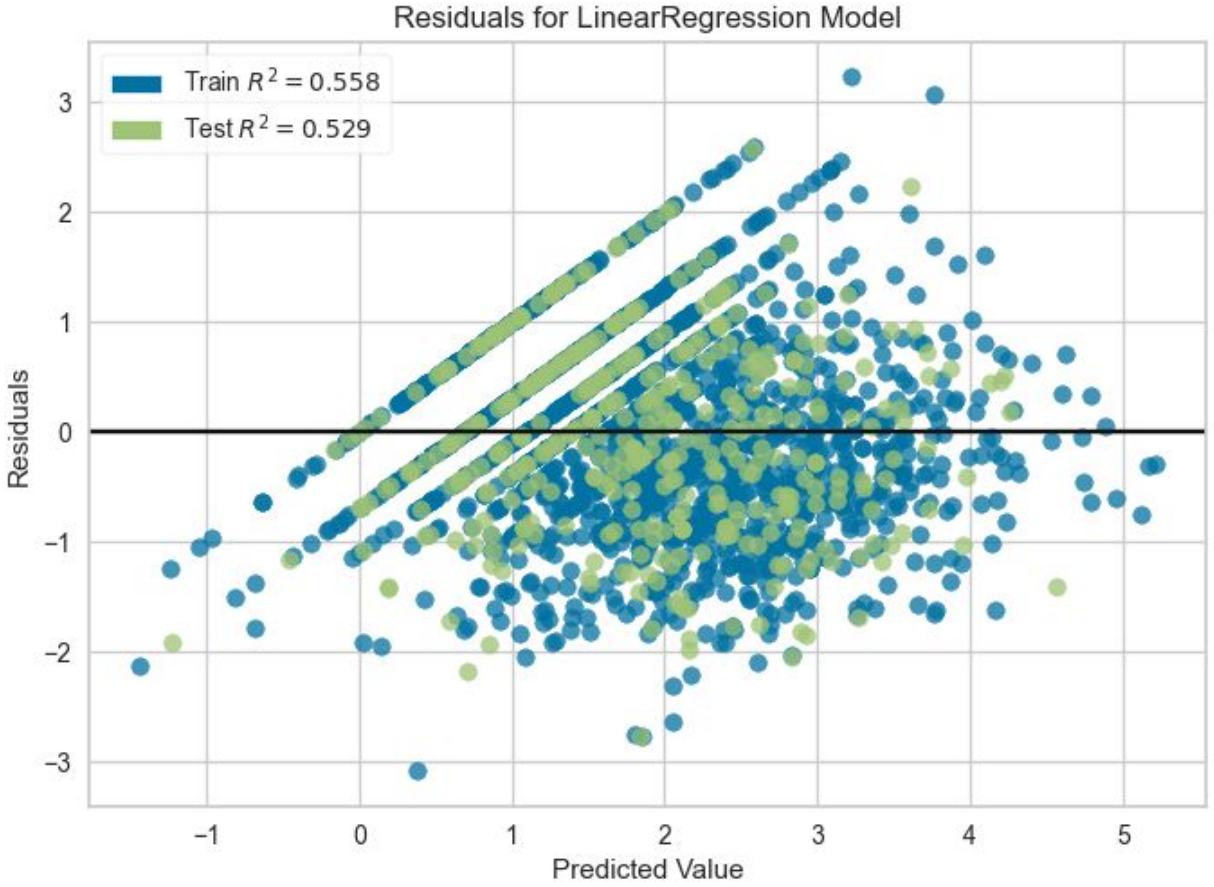
- Remove rows where recent cases is null. New shape: (2027, 28).
- Impute missing values with mean.
- One-hot-encoded categorical data.

Baseline Model



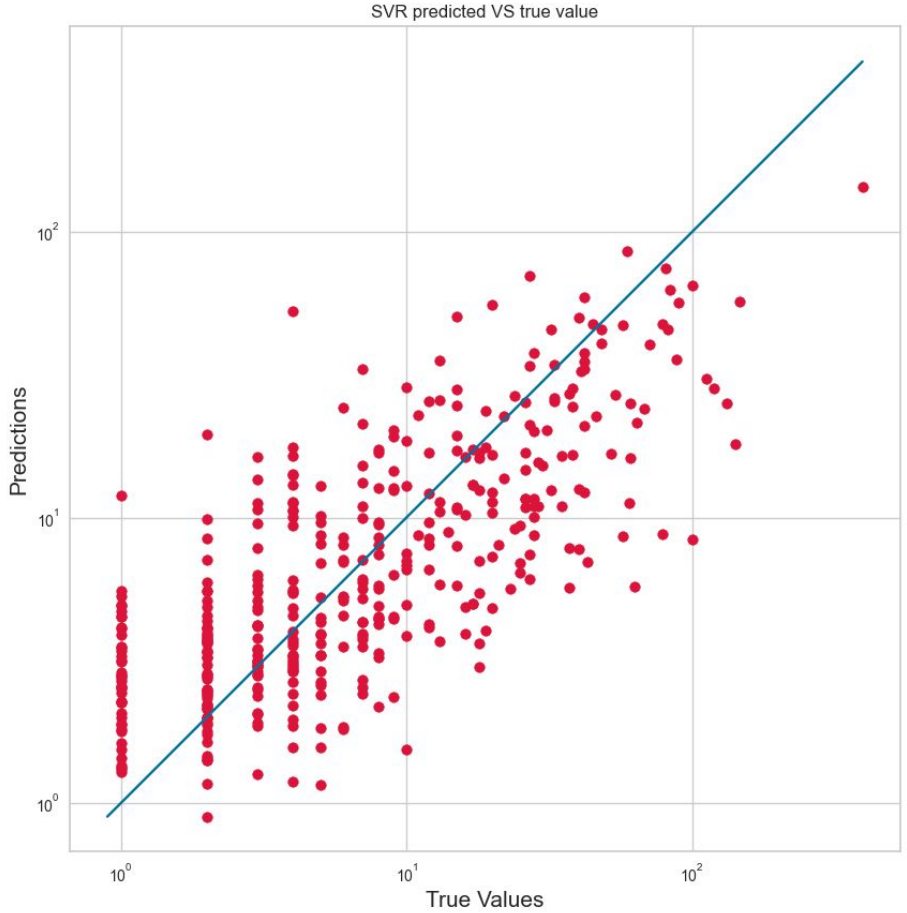
RMSE: 28.92

Linear Regression



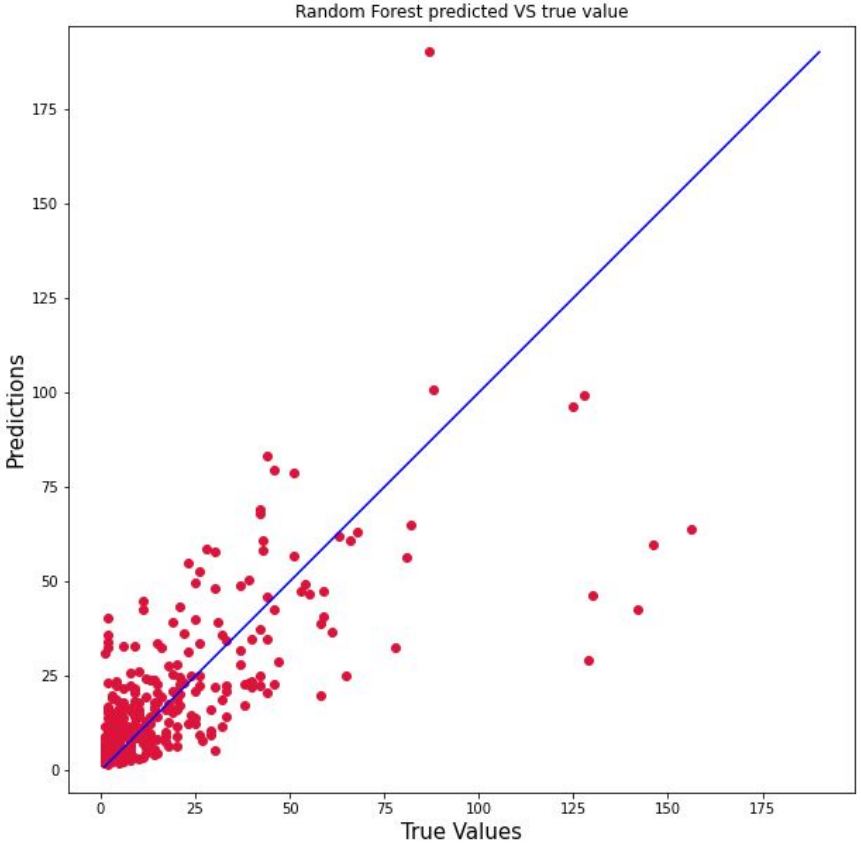
RMSE: 23.70
Training R2 score: 0.558
Testing R2 score: 0.529

Support Vector Regression



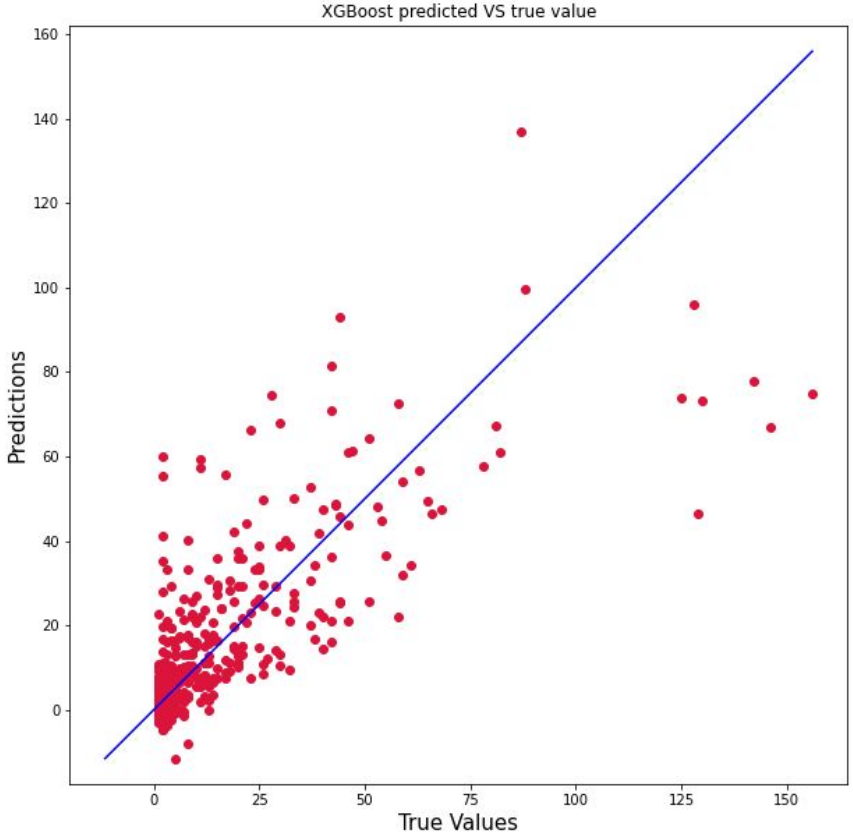
RMSE: 21.32
Training R2 score: 0.703
Testing R2 score: 0.552

Random Forest Regression



RMSE: 15.90
Training R2 score: 0.93
Testing R2 score: 0.52

XGBoost Regression



RMSE: 15.20
Training R2 score: 0.76
Testing R2 score: 0.56

Evaluation

Model	RMSE
Baseline	28.92
Linear Regression	23.70
Support Vector Regression	21.32
Random Forest Regression	15.90
XGBoost Regression	15.20

Cost-benefit Analysis

Each Town Council 3-years Mosquito Contract Average Cost	SGD 1,000,000
Each town council (GRC) consists on average 5 estates. Hence, cost for each estate	SGD 1,000,000 / 5 = SGD 200,000
Assume fumigation twice a month (fortnightly) Cost per fumigation session	SGD 200,000 / 36 / 2 = SGD 2,777,78

Type of Clinic	Dengue Blood Test Cost (SGD)
General Practitioner (GP)	70 - 80
Polyclinic	50 - 60

Treatment Service	Cost (SGD)
Blood Test	75
Consultation	37.5
Total	**112.5**

Items	Cost (SGD)
Fumigation (every two weeks)	2,778
Treatment	112.5
Breakeven number of cases	**25**

Recommendation:

To engage fumigation service only when there are more than 25 cases per fortnight

Cost-benefit Analysis

#	region	total_cases	avg_case
1	Ang Mo Kio	3345.0	35.59
2	Chai Chee	4793.0	34.24
3	Serangoon	2732.0	31.77
4	Tai Seng	2860.0	28.04
5	Toa Payoh	1804.0	23.43
6	Bukit Panjang	2131.0	20.1
7	Buangkok	2764.0	18.07
8	Admiralty	1837.0	16.85
9	Somerset (Road)	176.0	16.0
10	Marine Parade	1058.0	15.79

Recommendation:

To spray pesticides at the top four estate where the average case per fortnight is more than 25