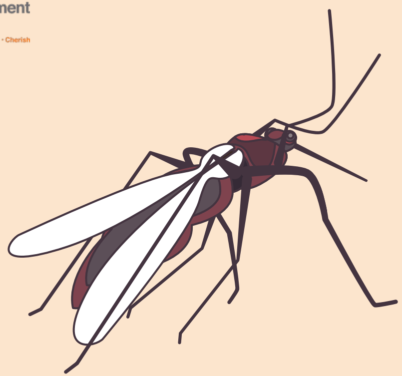
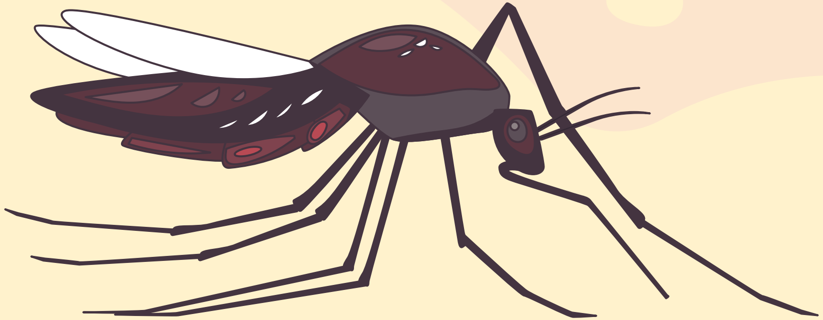
National Environment Agency
Safeguard • Nurture • Cherish



Project 04

Predict Dengue Cases

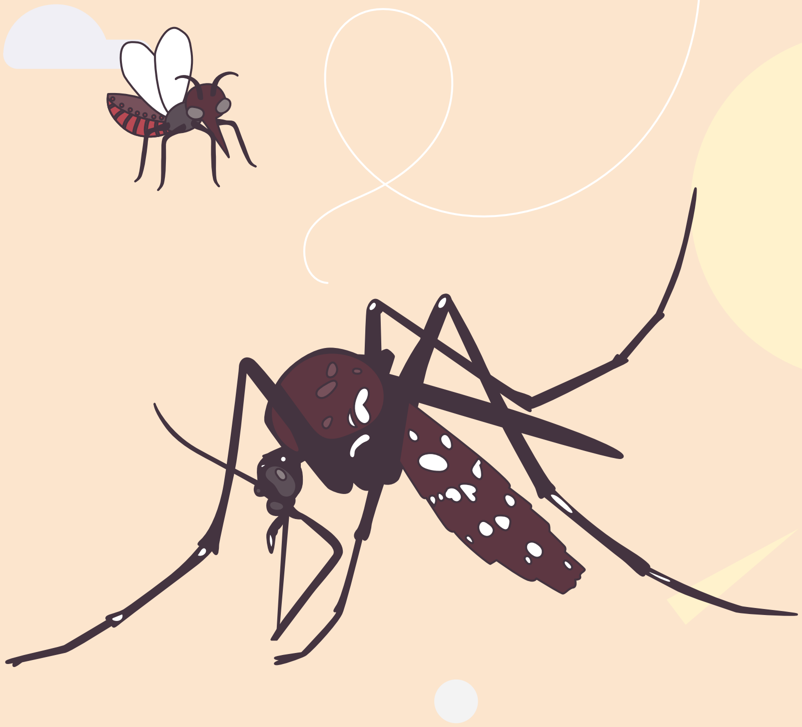


By Cui Cheng,
Zi Ming,
& Junny

National Environment Agency
Safeguard • Nurture • Cherish

Table of contents

National Environment Agency
Safeguard • Nurture • Cherish



01

Background & Problem Statement

Background

- Dengue: A major health threat in Singapore with periodic outbreaks.

- NEA’s response: ’Project Wolbachia’

- Challenge: Timing the implementation right due to factors like weather and operational challenges.

- NEA’s Vector Biology and Control Division’s role:
  - Develop a predictive model for dengue cases for the next four months.

  - Combine research with initiatives like Project Wolbachia for a dengue-safe Singapore.

National Environment Agency
Safeguard • Nurture • Cherish

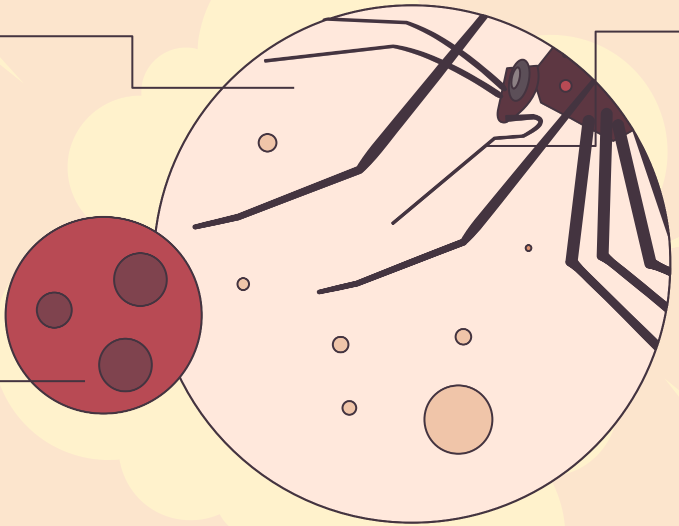
Project Wolbachia

Effectiveness

Cost-effective at ≥ 40% intervention effectiveness.

DALYs

Cost-effectiveness improves over time.



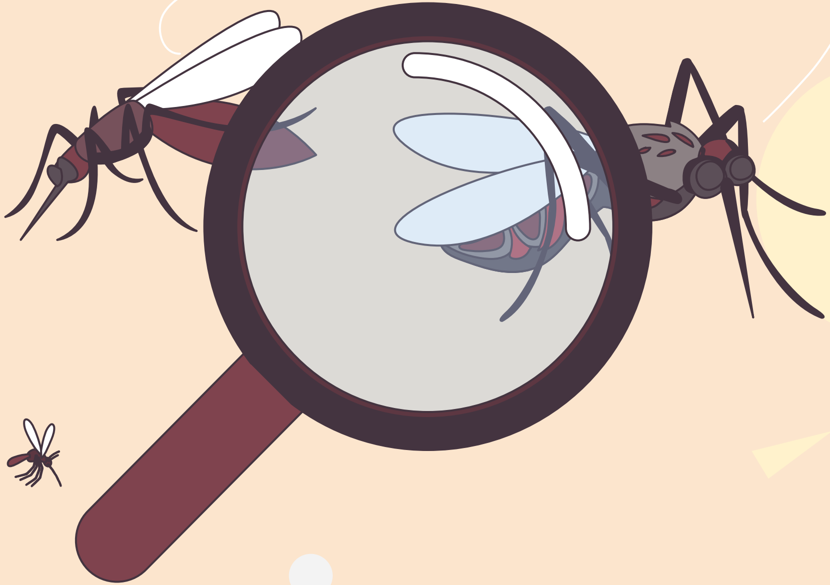
Costs Averted (2010-2020):

- **40% effectiveness:** US$329.40M
- **80% effectiveness:** US$658.79M

Dengue fever remains a significant health concern in Singapore. While Project Wolbachia seeks to counter this through the release of Wolbachia-infected mosquitoes, it grapples with steep expenses. Our objective is to devise a predictive model that can anticipate dengue outbreaks and discern their patterns, thereby optimizing the budget for Project Wolbachia.

—Problem Statement

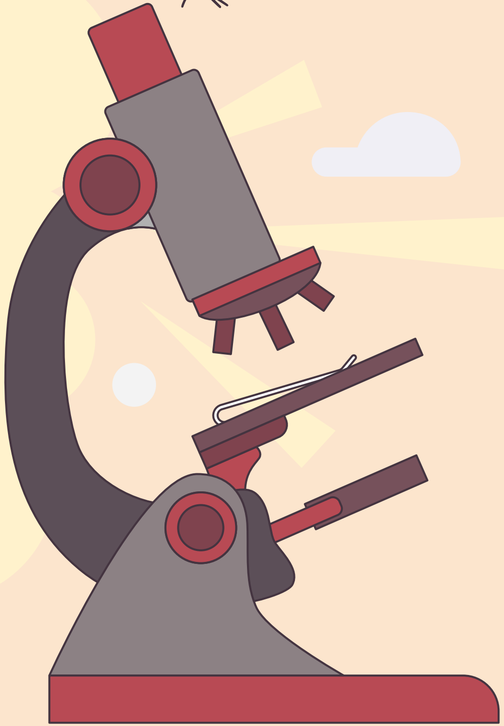
National Environment Agency
Safeguard • Nurture • Cherish



02 Models Approach

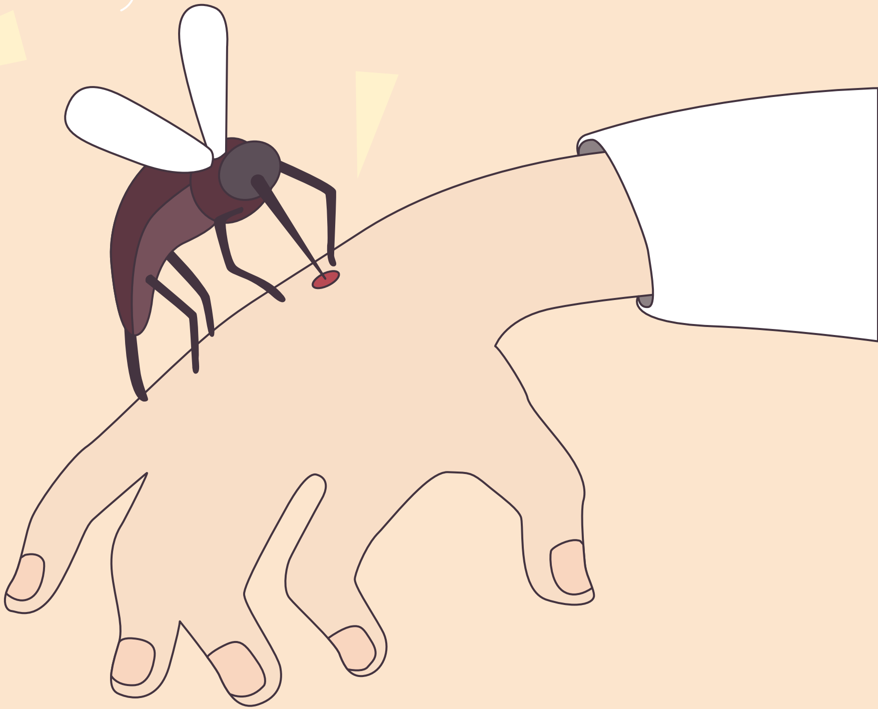
Models

ARIMA	A class of model that captures a suite of different standard temporal structures in time series data
SARIMA	Extension of ARIMA that explicitly supports univariate time series data with a seasonal component
SARIMAX	Extends SARIMA to include external variables
Pycaret	A Python library for machine learning, offers an implementation of SARIMA



National Environment Agency
Safeguard • Nurture • Cherish

03 Data Collection & EDA



Weather Data

Station	Position		Period of Daily Rain Records	Period of 30,60,120-Min Rain Records	Period of Mean Temperature	Period of Max and Min Temperature	Period of Mean Wind Speed	Period of Max Wind Speed
	Lat.(N)	Long. (E)						
Paya Lebar	1.3524	103.9007	Jan 1980-current	-	Sep 2017-current	Jan 1981-current	Jan 1981- current	Jan 2010-current
Tengah	1.3858	103.7114	Jan 1980-current	-	Aug 1986-current	Jan 1985-current	Jan 1985-current	Jan 2010-current
Changi	1.3678	103.9826	Jan 1981-current	Jan 2014-current	Jan 1982-current	Jan 1982-current	Jan 1983-current	Jan 1983-current
Seletar	1.4166	103.8654	Jan 1980-current	-	Aug 1986-current	Jan 1985-current	Jan 1985-current	Jan 2010-current

282,636 Rows from 64 Stations → **605 Rows from 1 Station**

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

National Environment Agency
Safeguard • Nurture • Cherish

Dengue Data

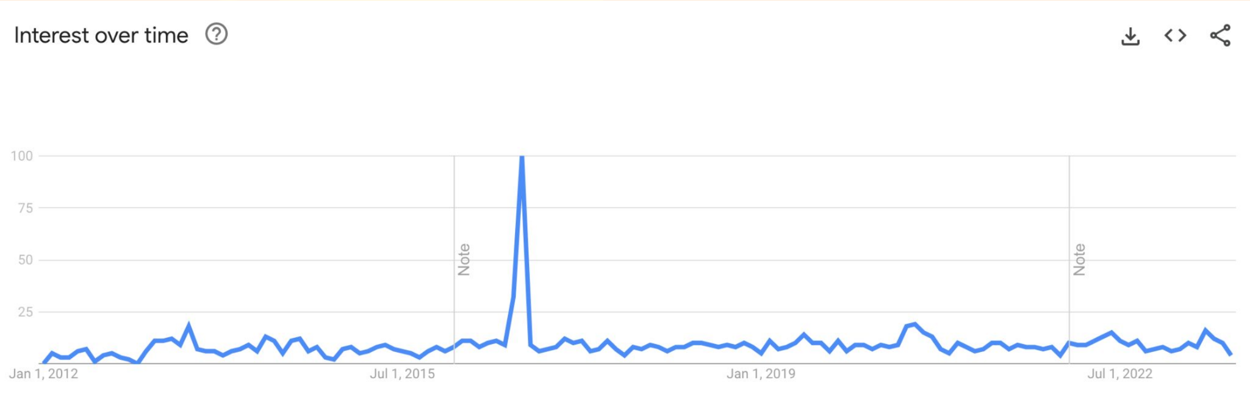
Epidemiological Week	Disease	No. of Cases (No.)
2012-W01	Acute Viral hepatitis B	0
2012-W01	Acute Viral hepatitis C	0
2012-W01	Avian Influenza	0
2012-W01	Campylobacterenterosis	6
2012-W01	Chikungunya Fever	0

605 Rows of Dengue Fever and Dengue Hemorrhagic Fever

https://beta.data.gov.sg/datasets/508/view

Google Trends Data

‘Dengue Fever’

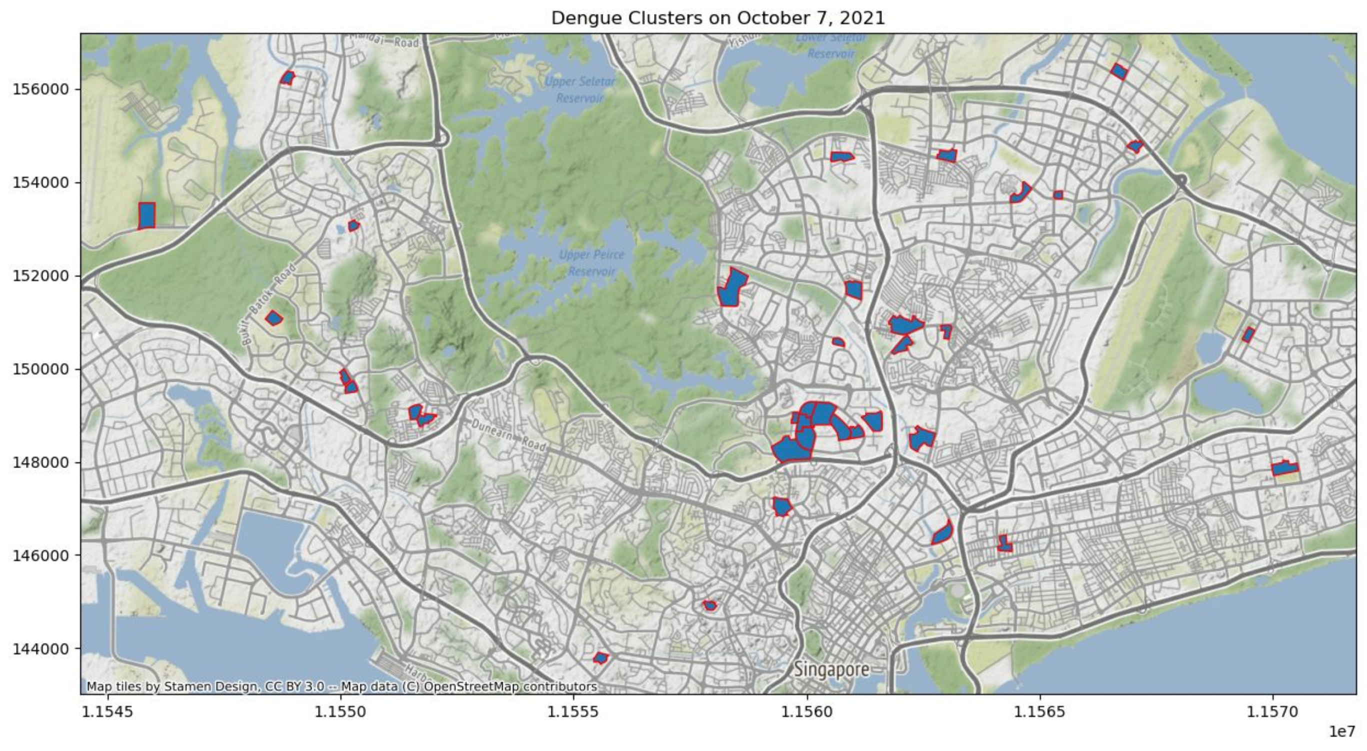


‘Insect Repellent’

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

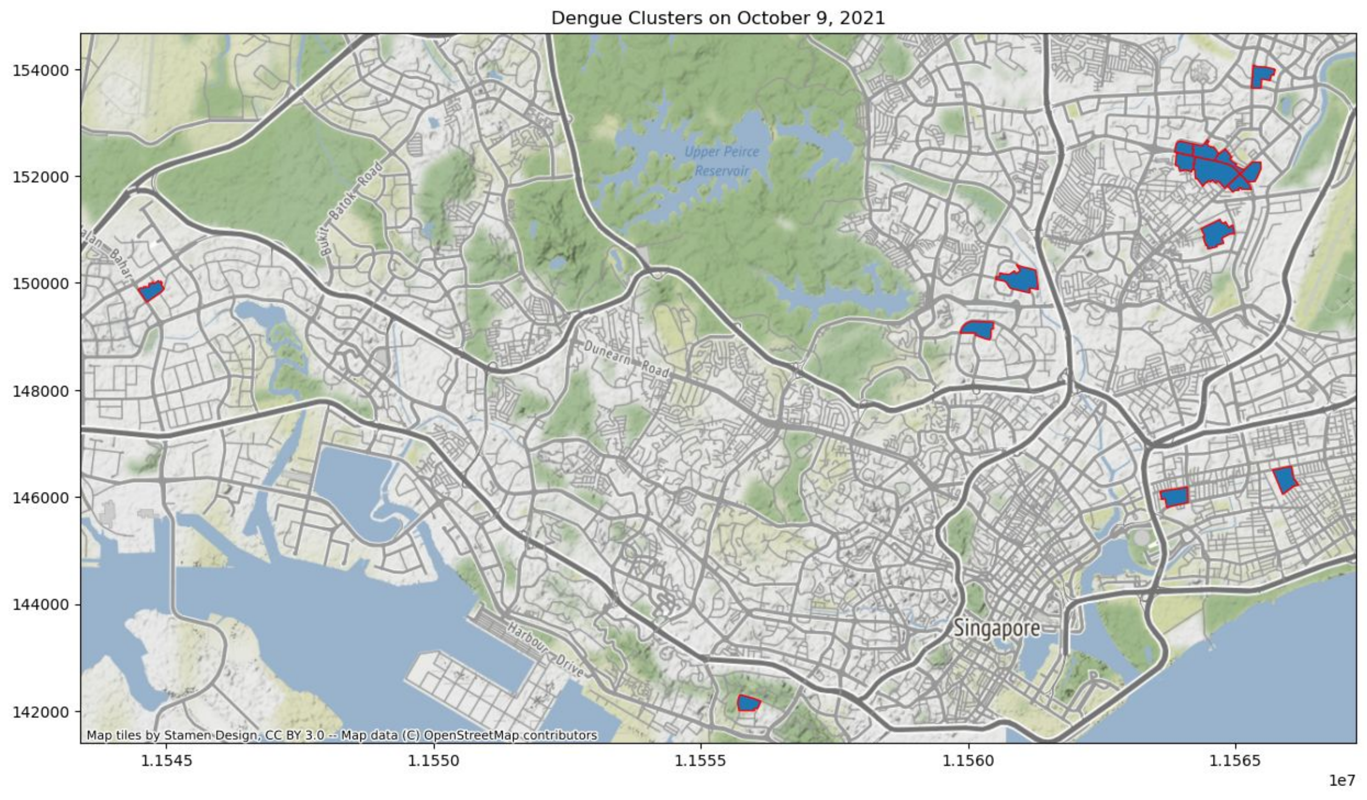
Dengue Cluster Change

7th October 2021

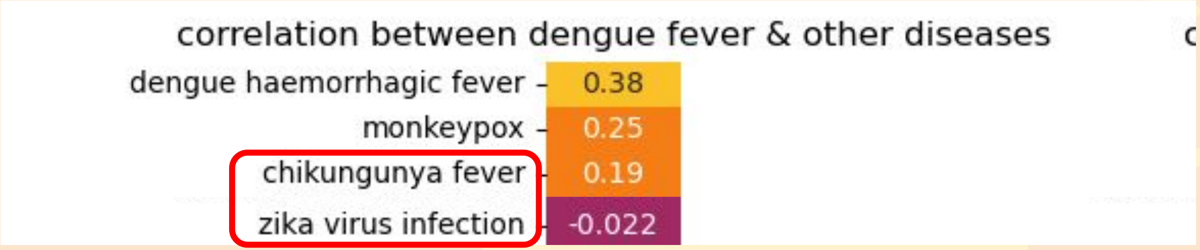


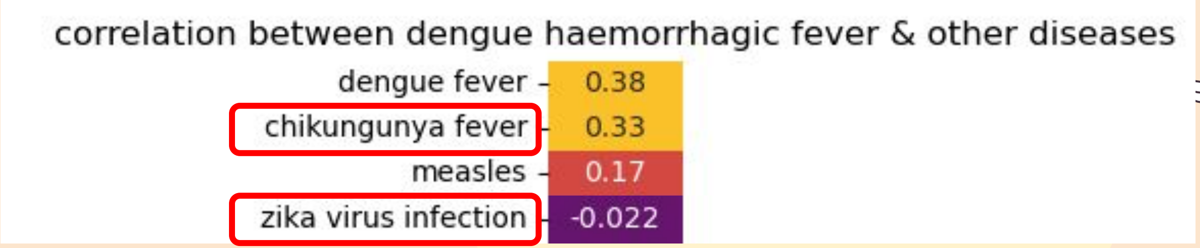
Dengue Cluster Change

9th October 2021

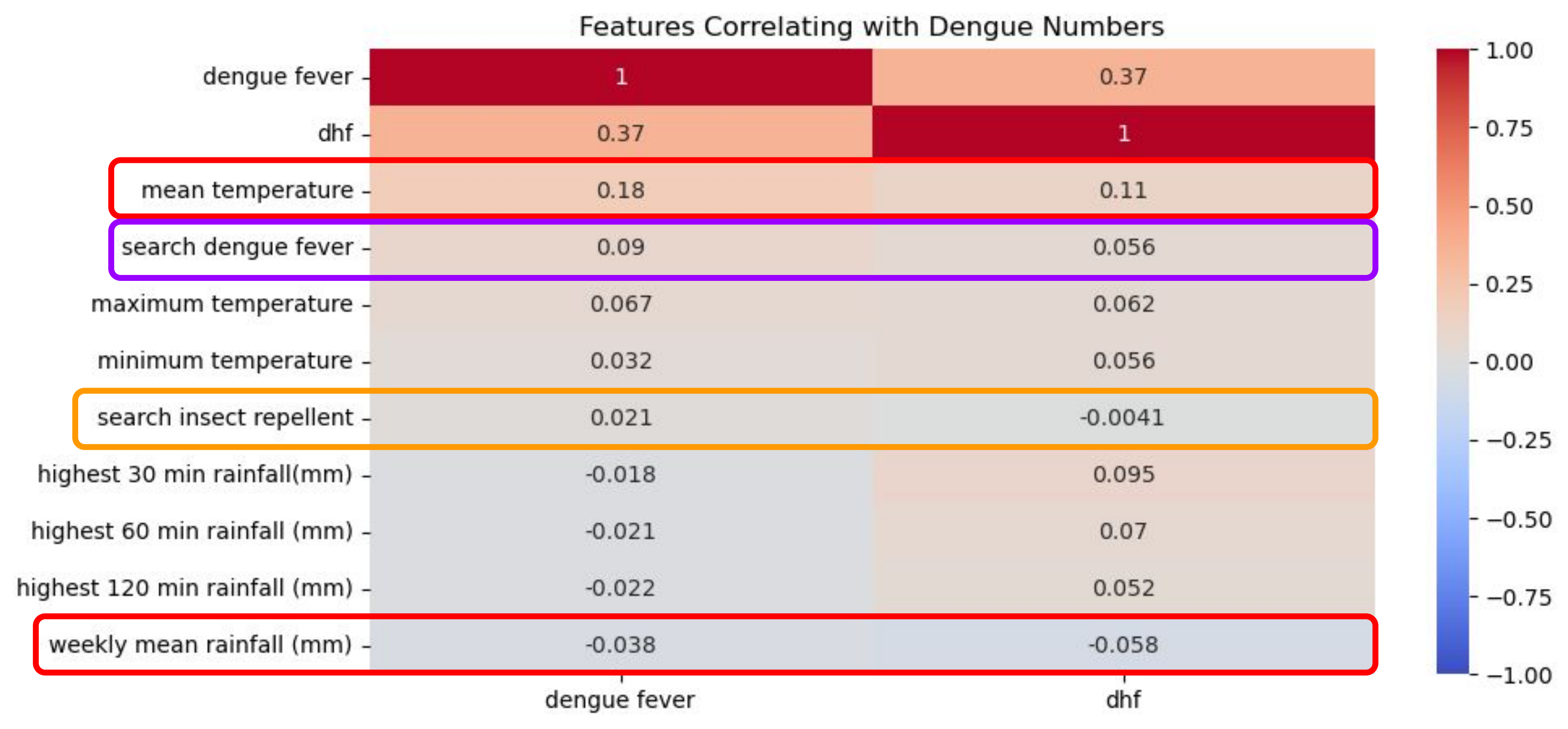


Dengue & Other Diseases

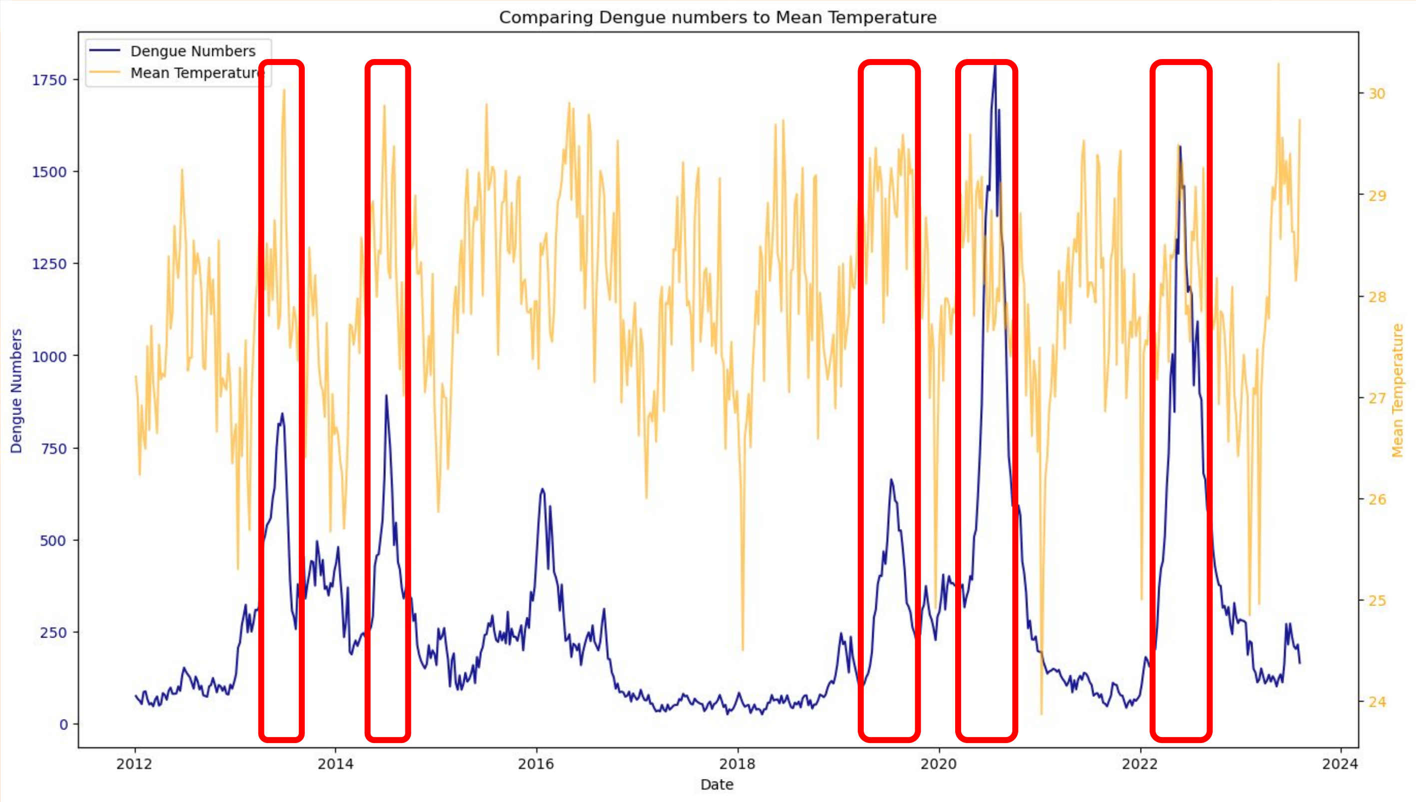




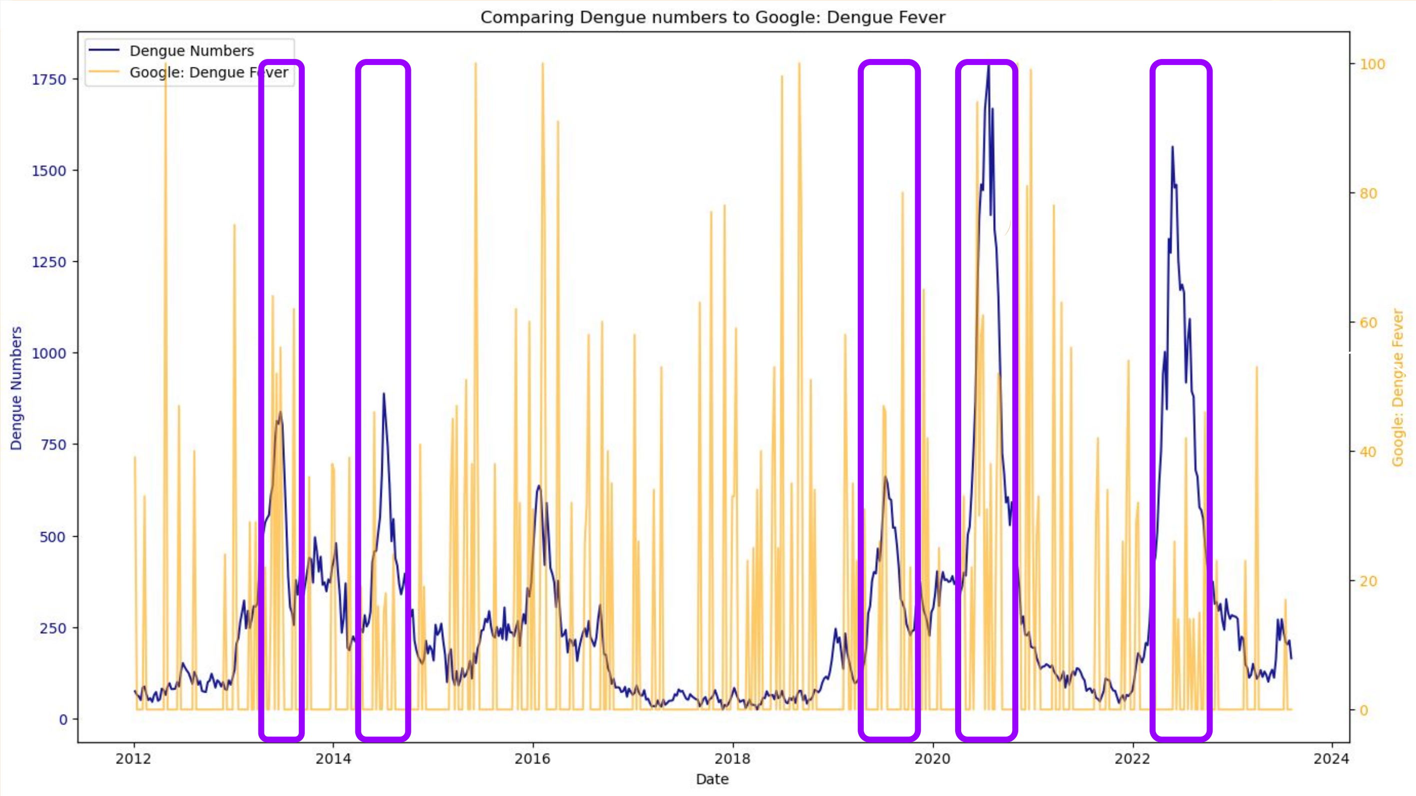
Dengue & Weather



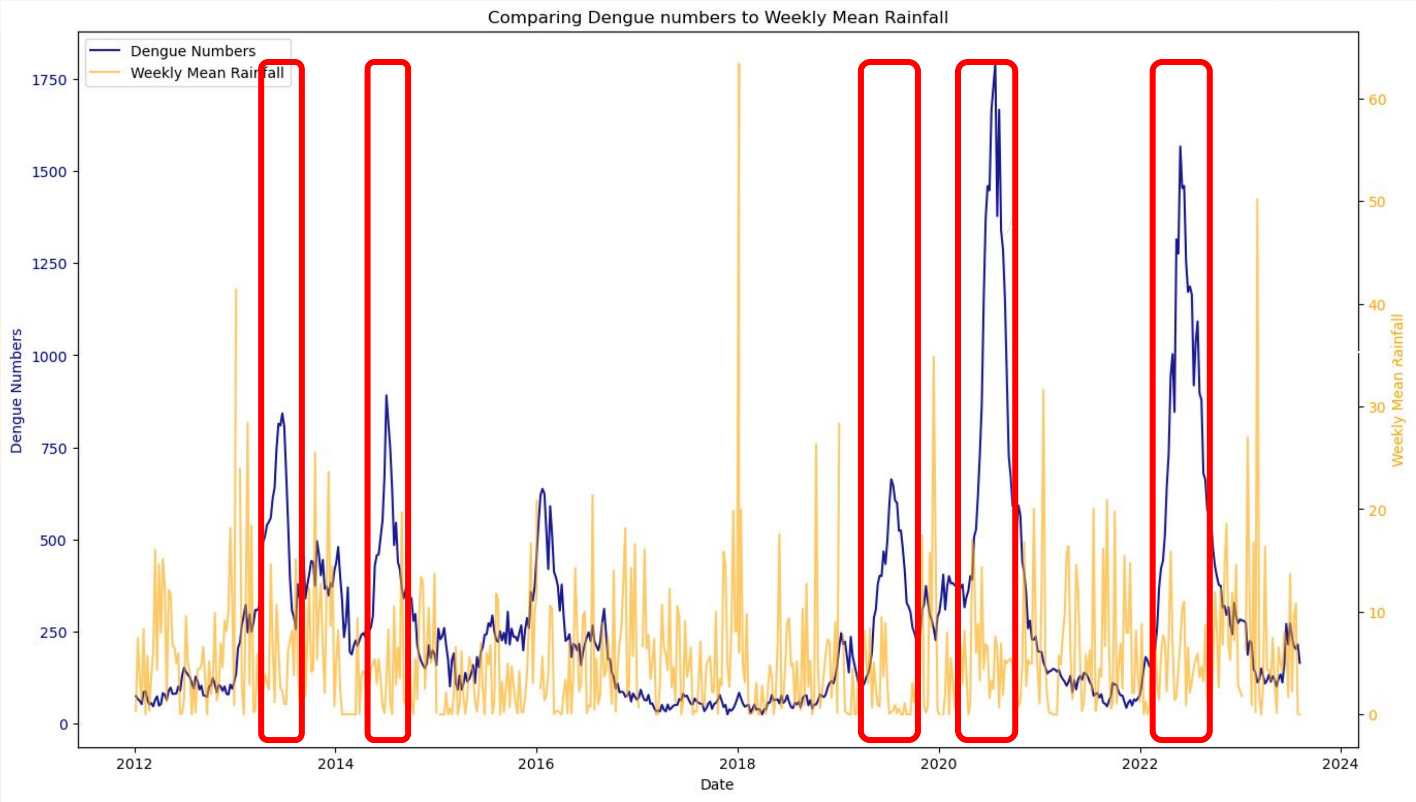
Dengue x Temperature



Dengue & Google Trend: Dengue Fever



Dengue & Other Diseases



National Environment Agency
Safeguard • Nurture • Cherish

04
MODELLING

Modelling

ARIMA/SARIMA/SARIMAX

Small sample size

AR: number of dengue case is influenced by number of dengue case in the past

- Immunity
- Infectious nature of dengue

MA: number of dengue cases is influenced by shocks to the system

- destruction/creation of mosquito habitat can be very random
- Chance occurrence to be bitten + show symptoms

Model summary

	Manual	Pycaret autoARIMA
ARIMA	0.2530	
SARIMA	0.2502	0.250
SARIMAX	0.2382	0.188

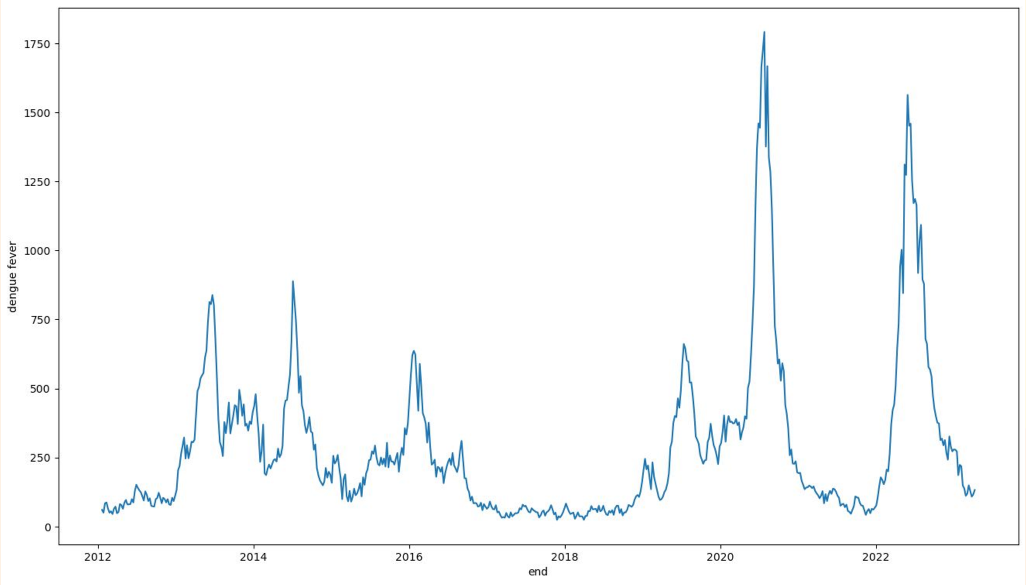
Barely any improvement for SARIMA and SARIMAX

MAPE hovers around 25%

Fitting an ARIMA (p,d,q) model

Determining parameter d for I

Check stationarity of time series



Checking for stationarity/ Determining d:

ADF test: 0.01

- Reject H0: presence of a unit root

KPSS test: 0.0257

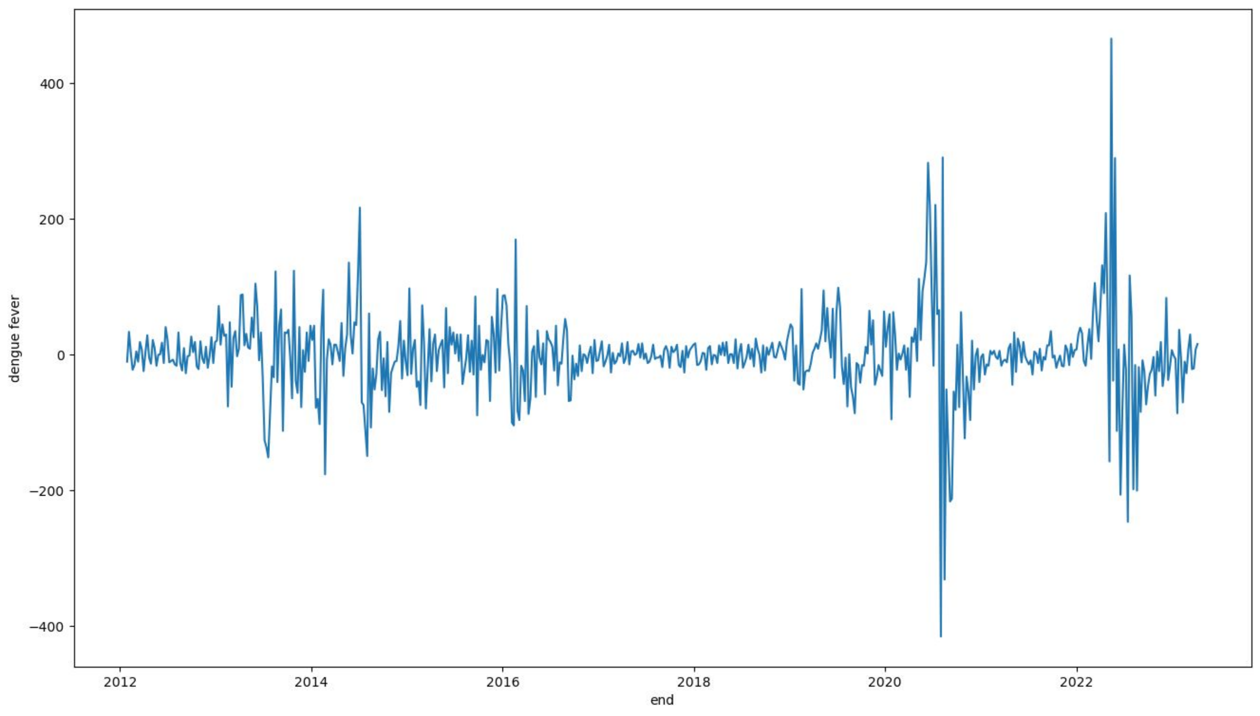
- Reject H0: time series is stationary

Caveat: ADF only tests for the presence of a single unit root;

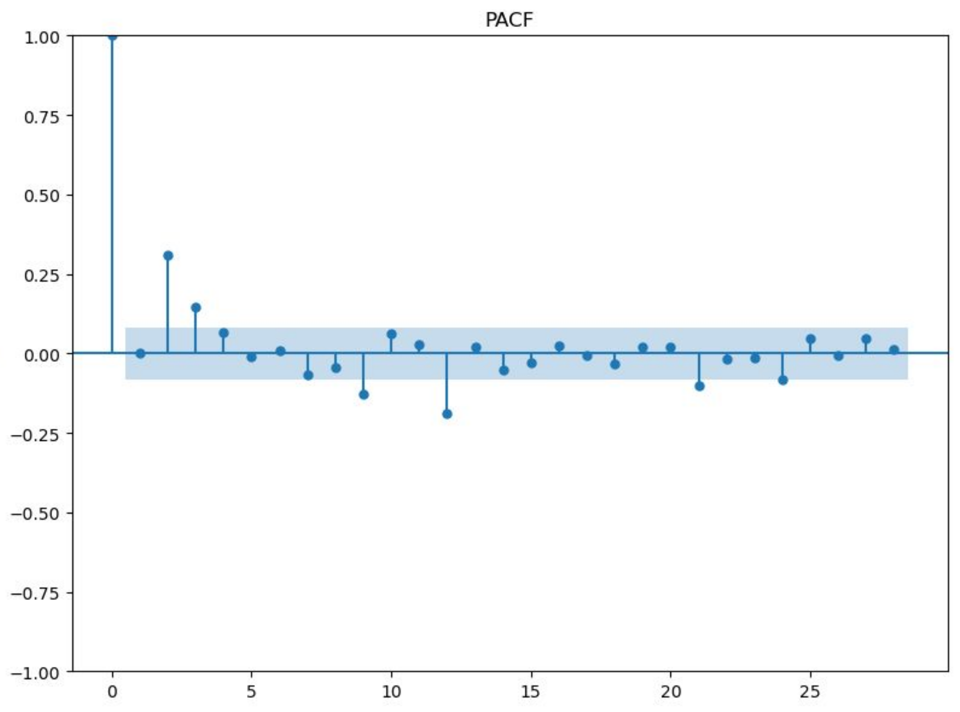
Conclusion: need to take the first order difference

d = 1

ADF = 0.01 KPSS = 0.1



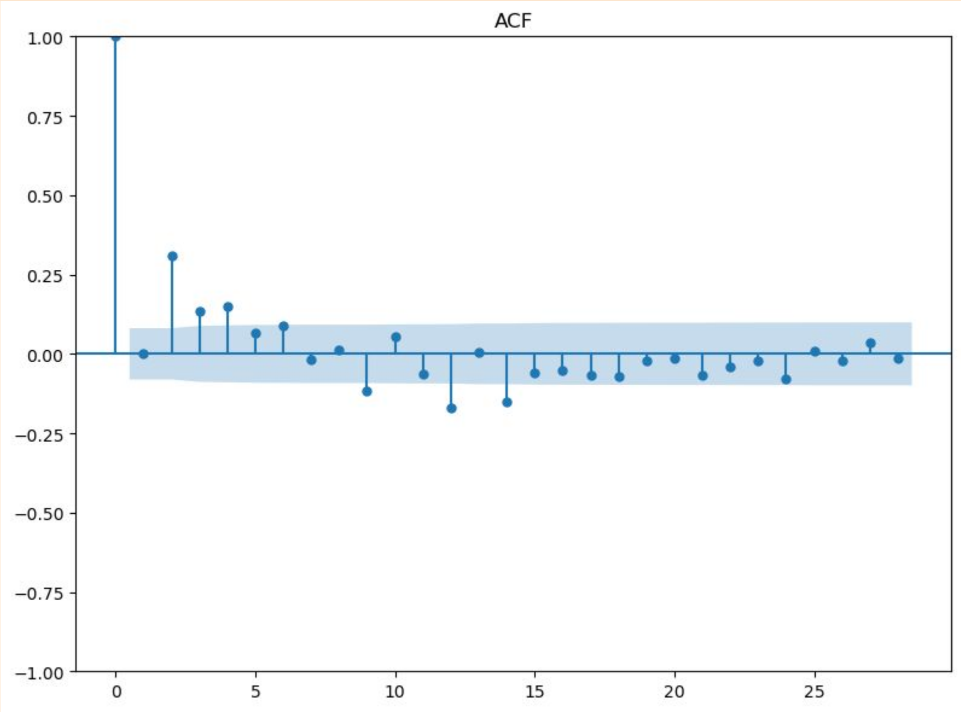
Determining p



Significant correlations at lag p = 2, 3, 9, 12

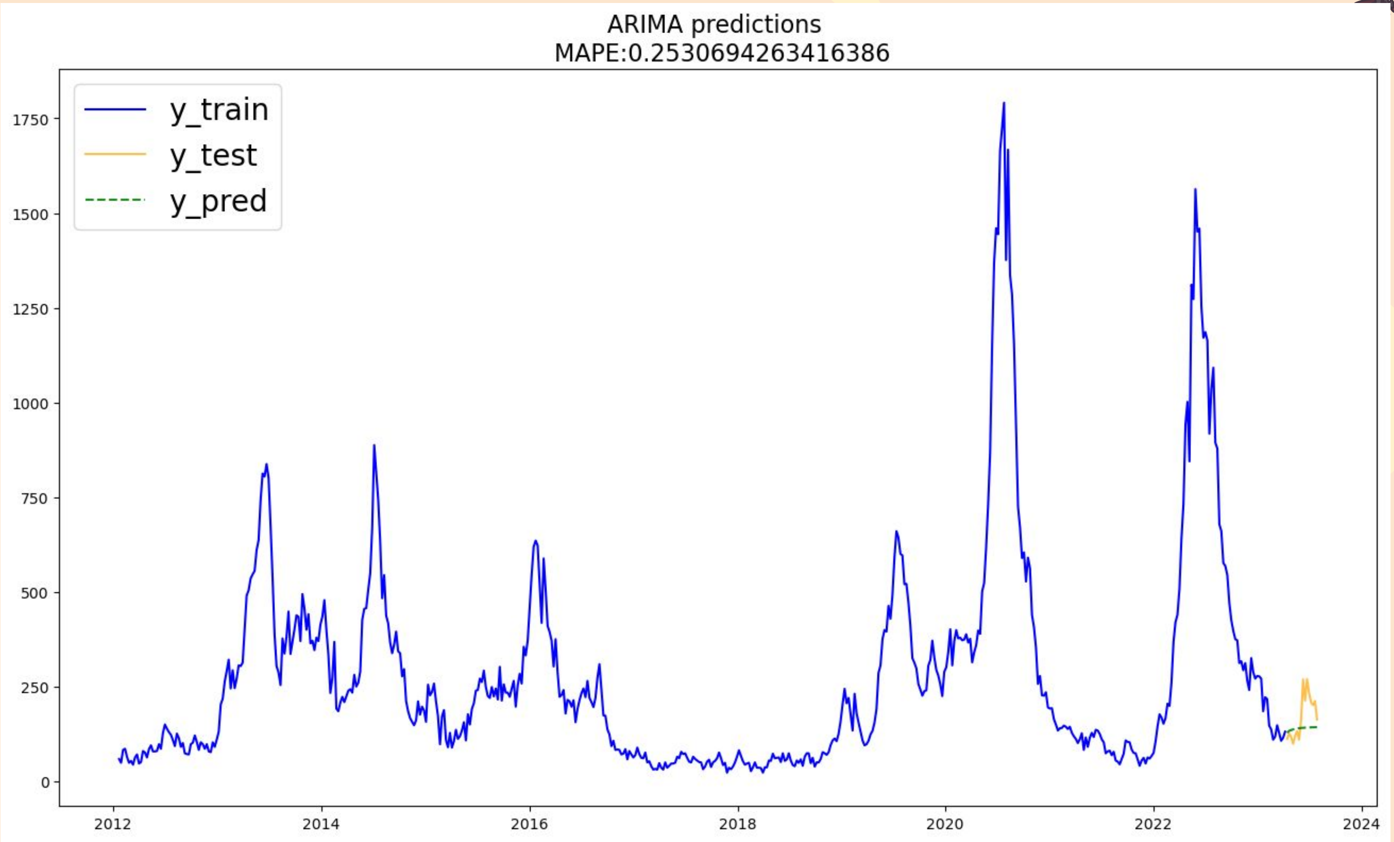
National Environment Agency
Safeguard • Nurture • Cherish

Determining q

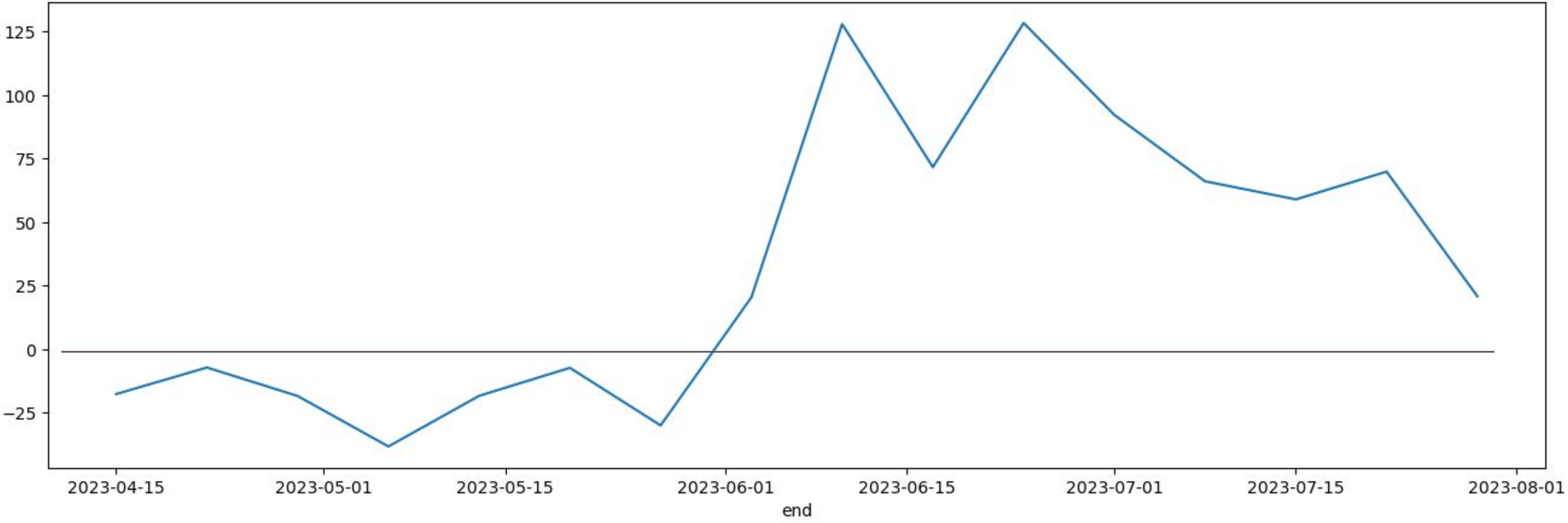


Significant correlations at lag p = 2, 3, 4, 9, 12, 14

ARIMA(2,1,2)


ARIMA predictions
MAPE:0.2530694263416386
y_train
y_test
y_pred
1750
1500
1250
1000
750
500
250
0
2012
2014
2016
2018
2020
2022
2024

Residual plot



Fitting SARIMA model:

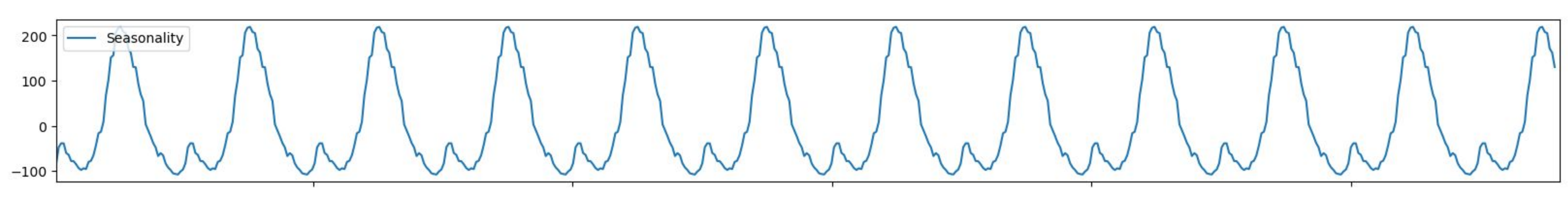
Why SARIMA?

- Accounts for seasonality in dengue cases
- Climate Events
- Temperature
- Rainfall

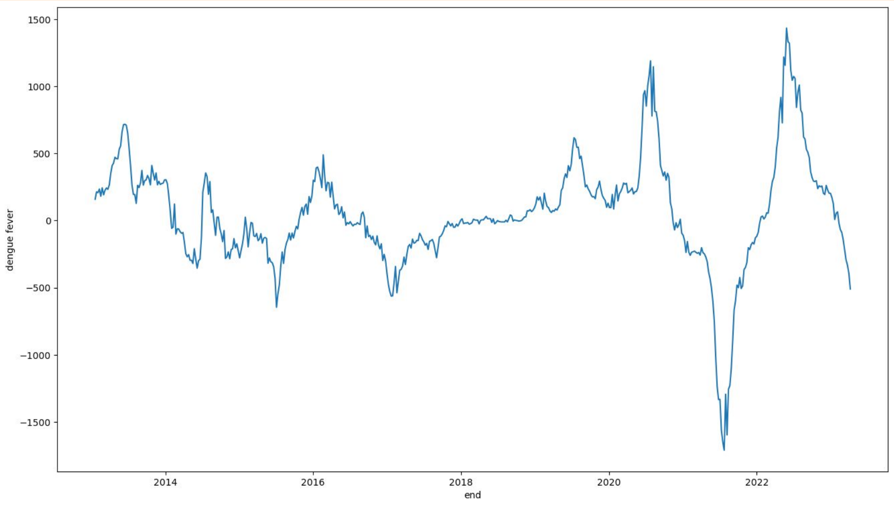
Fitting SARIMA: determining m

11 peaks in 11 years

m = 52



determining D

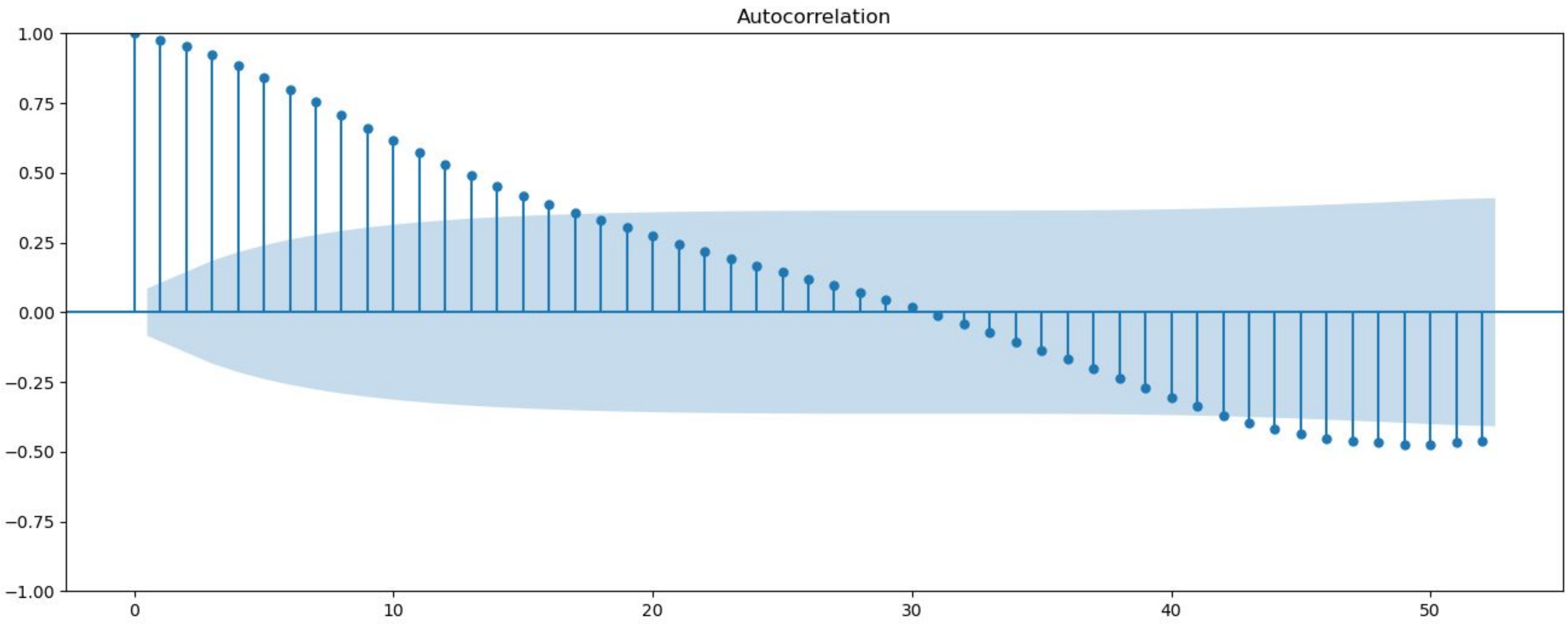


ADF: 0.02 reject H0

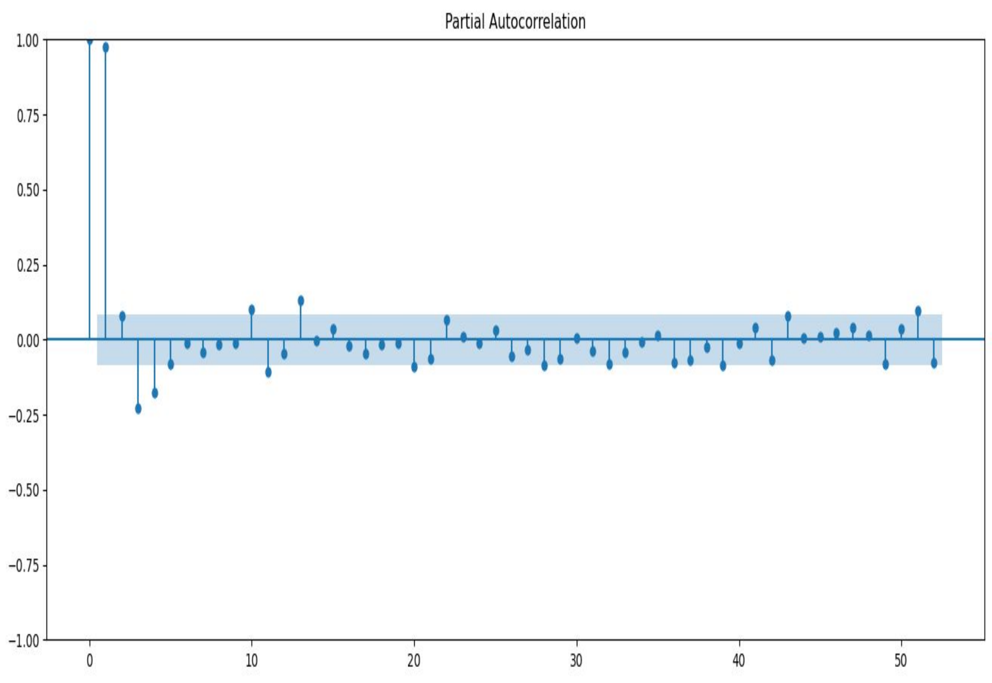
KPS: 0.1 don’t reject H0

D = 1

Determining P

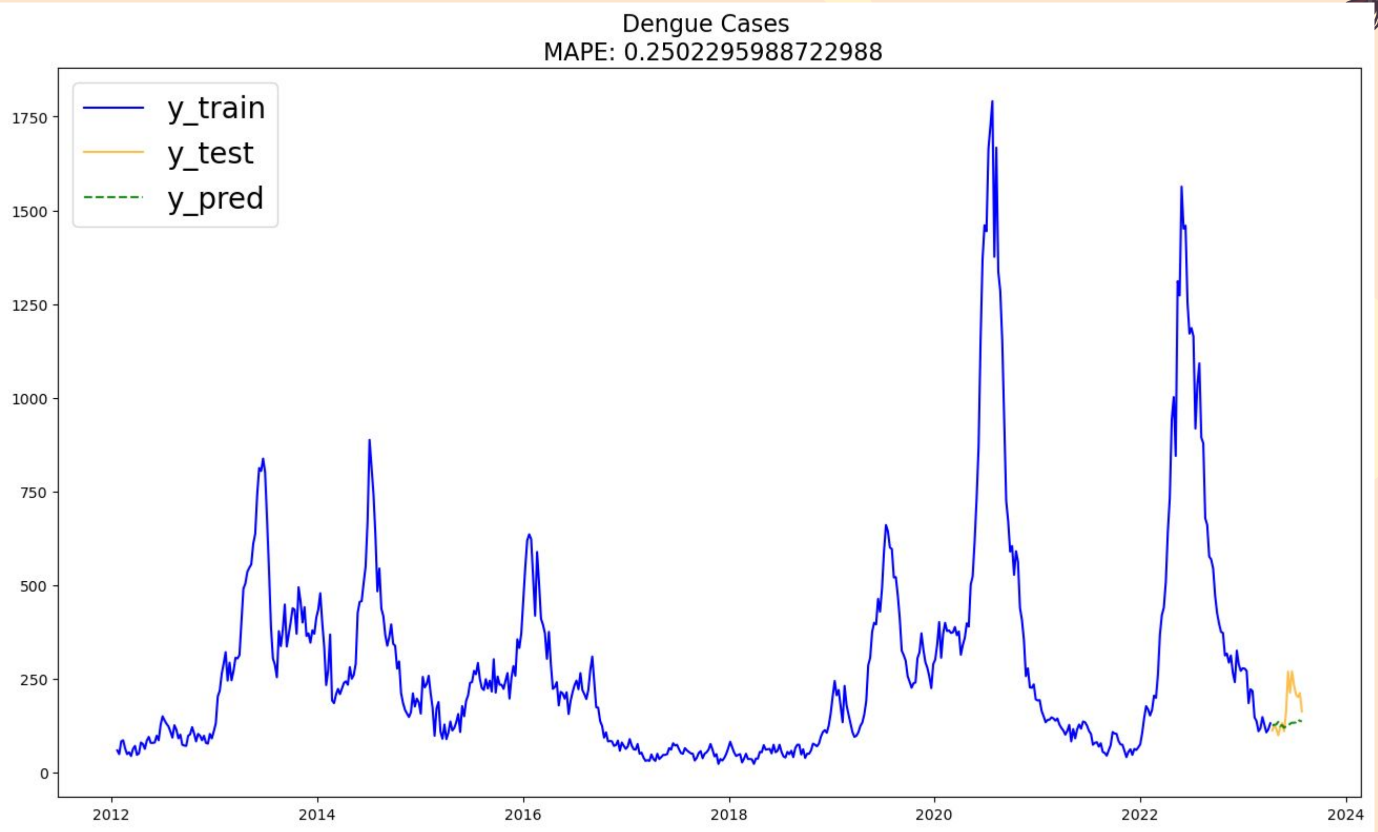


Determining P



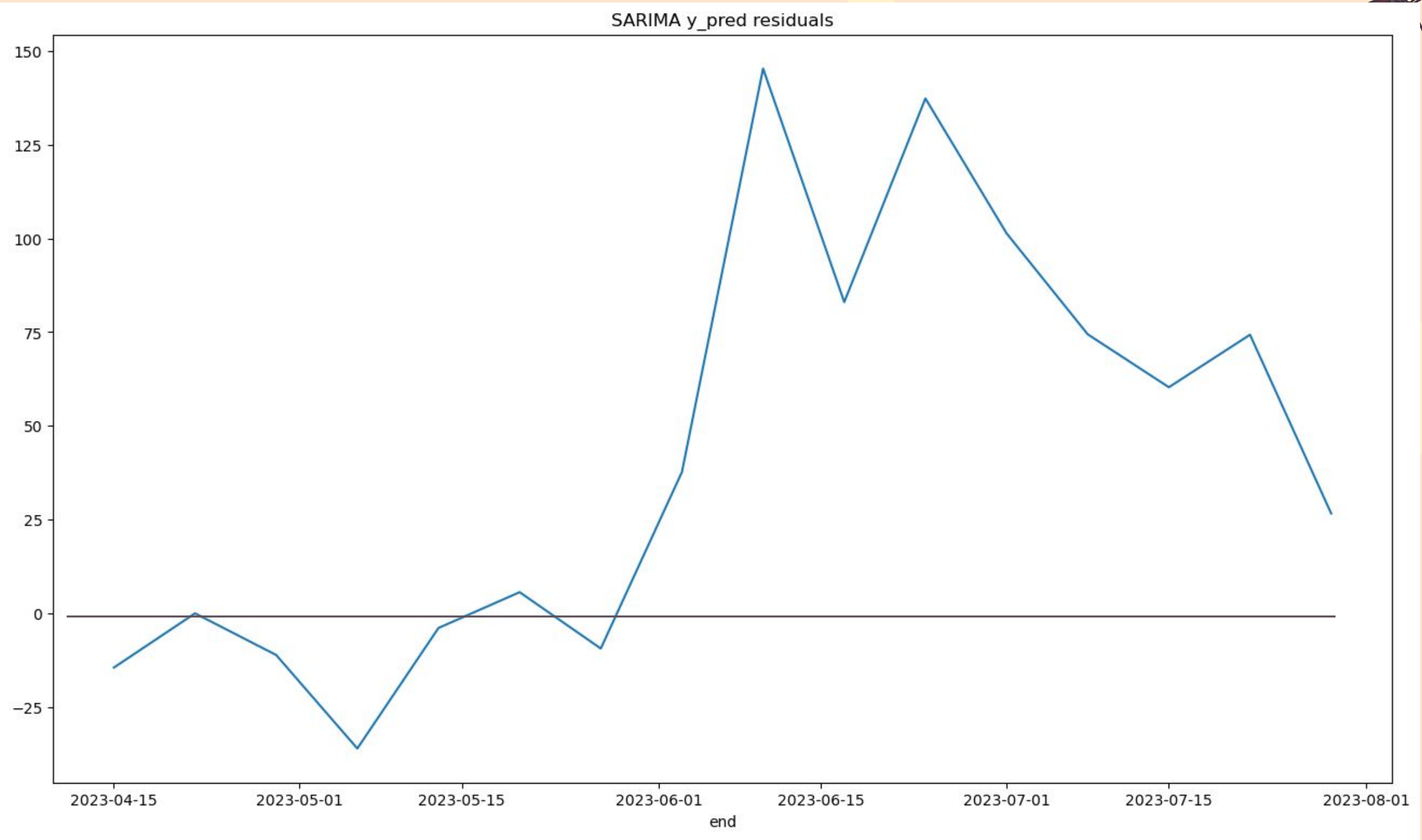
P = 1, 3, 4

SARIMA 2,1,2 1, 0, 0, 52



National Environment Agency
Safeguard • Nurture • Cherish

SARIMA residuals



SARIMA X

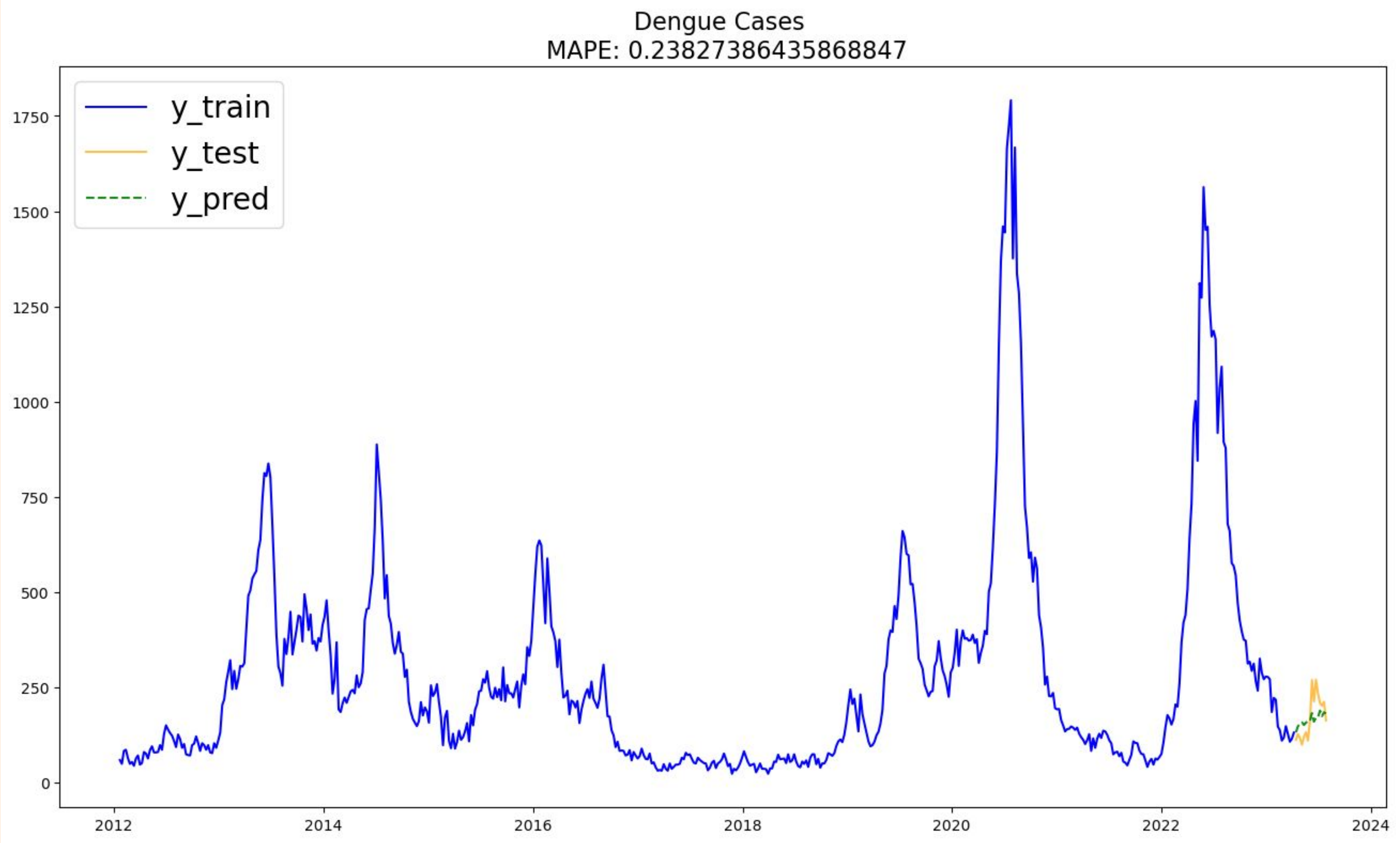
exogenous factors introduced:

Other diseases transmitted by the vector mosquito

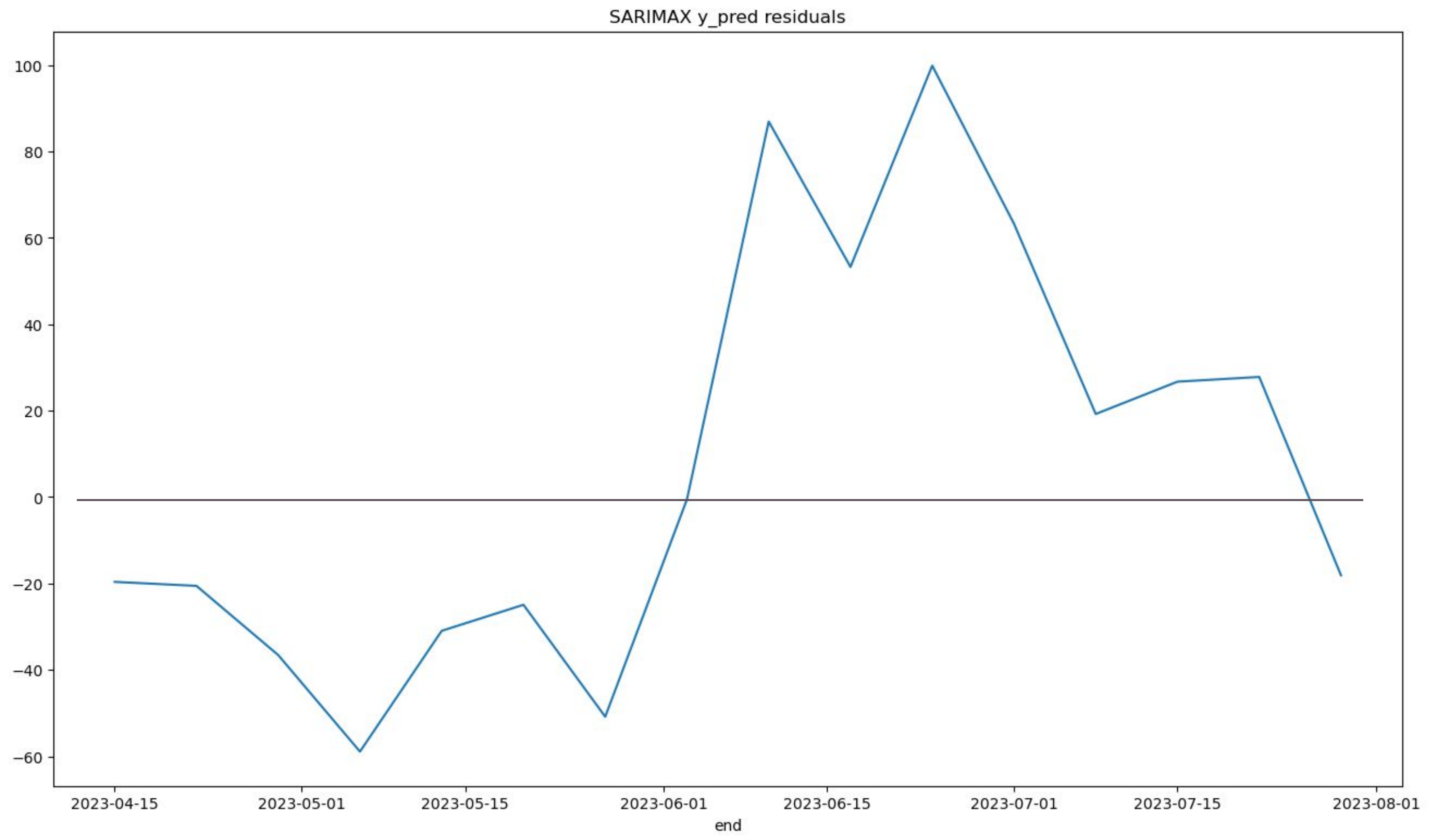
Climate data with lag of 2 weeks

Google Trends on ‘Dengue Fever’ and ‘Insect repellent’ topics

SARIMA X (1,1,2)(0,0,0)[52]



SARIMA X Residuals

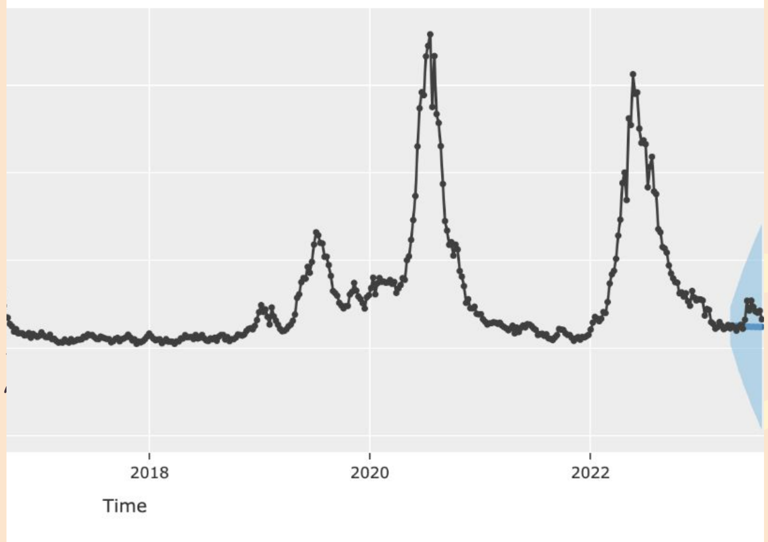


Pycaret

SARIMA

SARIMAX Results

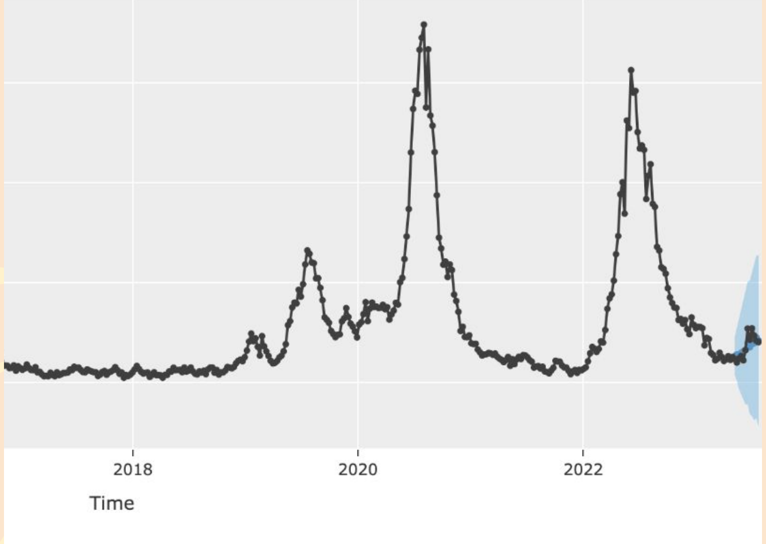
ariable:	y
Model:	SARIMAX(1, 0, 0)x(0, 1, [1], 2)



SARIMA X

SARIMAX Results

Variable:	y
Model:	SARIMAX(1, 1, 0)x(2, 0, [1, 2], 2)



National Environment Agency
Safeguard • Nurture • Cherish

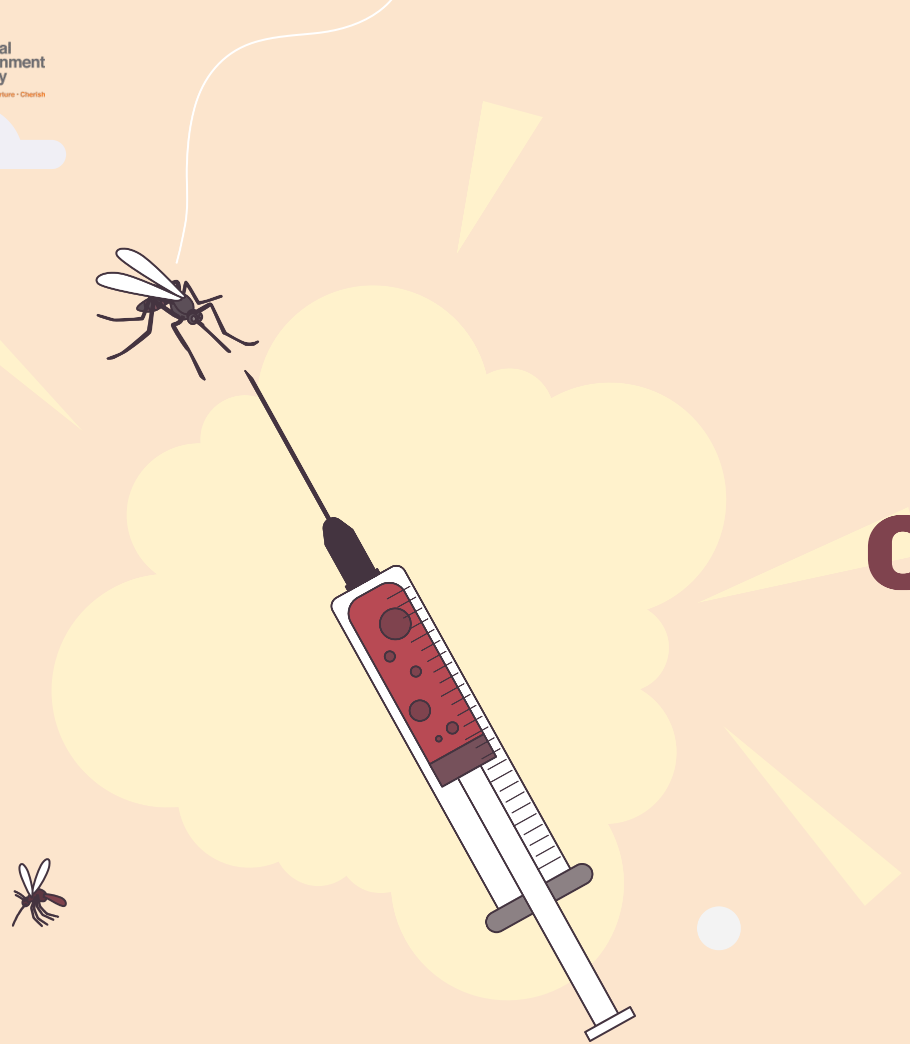
Model summary

	Manual	Pycaret autoARIMA
ARIMA	0.2530	
SARIMA	0.2502	0.250
SARIMAX	0.2382	0.188

Barely any improvement for SARIMA and SARIMAX

MAPE hovers around 25%

National Environment Agency
Safeguard • Nurture • Cherish



05

Cost-Benefits Analysis

National Environment Agency
Safeguard • Nurture • Cherish

Cost Benefit Analysis

Project Wolbachia

- 22.7Million 2010USD Steady-State cost
- Already 'cost effective' at 40% efficacy
- NEA reported 60%-80% efficacy

Cost Concerns:

- Equipment/Labor Cost
- Suppression requires constant maintenance costs

Cost Benefit Analysis

Classification problem

Based on our forecasted dengue cases 4 months into the future,

Should/should not NEA implement project wolbachia?

Cost Benefit Analysis

TS Forecast -> Classification

Model → Prediction → Number of predicted dengue cases calls for Wolbachia implementation?

Yes

Should have implemented:
Cost: 22.7 Million
Benefit: 55.93 Million (saved)

Should not have implemented:
Cost: 22.7Million

No

Should have implemented:
Cost: 55.93 Million

Should not have implemented:
Cost: N/A

Cost Benefit Analysis

We are interested in achieving a high prediction accuracy (1-MAPE)

80% Efficacy	Predicted Positive (action)	Predicted Negative (no action)
Actual Positive (Wolbachia implementation was required)	Gain: $33.23 million TP	Loss: $55.93 million FN
Actual Negative (Wolbachia implementation was not required)	Loss: $22.7 million FP	No Change TN

Expected cost/benefit = (Probability of TP×Gain if TP) + (Probability of FN×Loss if FN) + (Probability of FP×Loss if FP) + (Probability of TN×Gain if TN)

The actual/predicted positive rate and actual negative rates

depend on how/when NEA deems Wolbachia implementation is required

National Environment Agency
Safeguard • Nurture • Cherish

06

Conclusion & Continuous Improvements

Conclusion

Model that predicts with MAPE of 18% at 4 months notice

The National Environment Agency (NEA)

- Project Wolbachia
- Public Health Campaigns
- Plan for fumigation

Ministry of Health:

- Prepare healthcare facilities

Future Works

Explore other models

- Dynamic Harmonic Regression
- Random Forests

Collect more data:

- Look into more correlated Data:
- Try to find data at higher frequencies than weekly

Conclusion and Continuous Improvements

Model Accuracy

PyCaret SARIMAX MAPE : 0.18

Project Wolbachia Benefit

Anticipate, Intervene, Suppress

Implications

NATIONAL ENVIRONMENTAL AGENCY

- Early warning prevention measures
- Reduced medical costs, better health outcomes

MINISTRY OF HEALTH

- Prepare for patient surges
- Allocate funds and resources to dengue research & prevention.

Data Collection

- Expand Data Sources
- Data Frequency

Feedback Mechanism

- Real-time Feedback
- Error Analysis

Stakeholder Collaboration

- Collaborate with Healthcare Professionals
- Engage with MOH and NEA

Public Engagement

- Awareness Campaigns
- Public Feedback

National Environment

Thanks