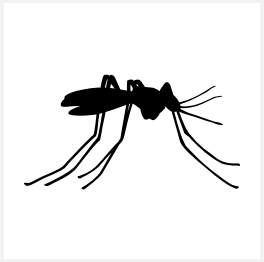
West Nile Virus: Prediction & Prevention

Serene, Evan, Yu Fung

Birds

- Host



Mosquitoes

- Carrier

Humans

- Infected

What is West Nile Virus (WNV)

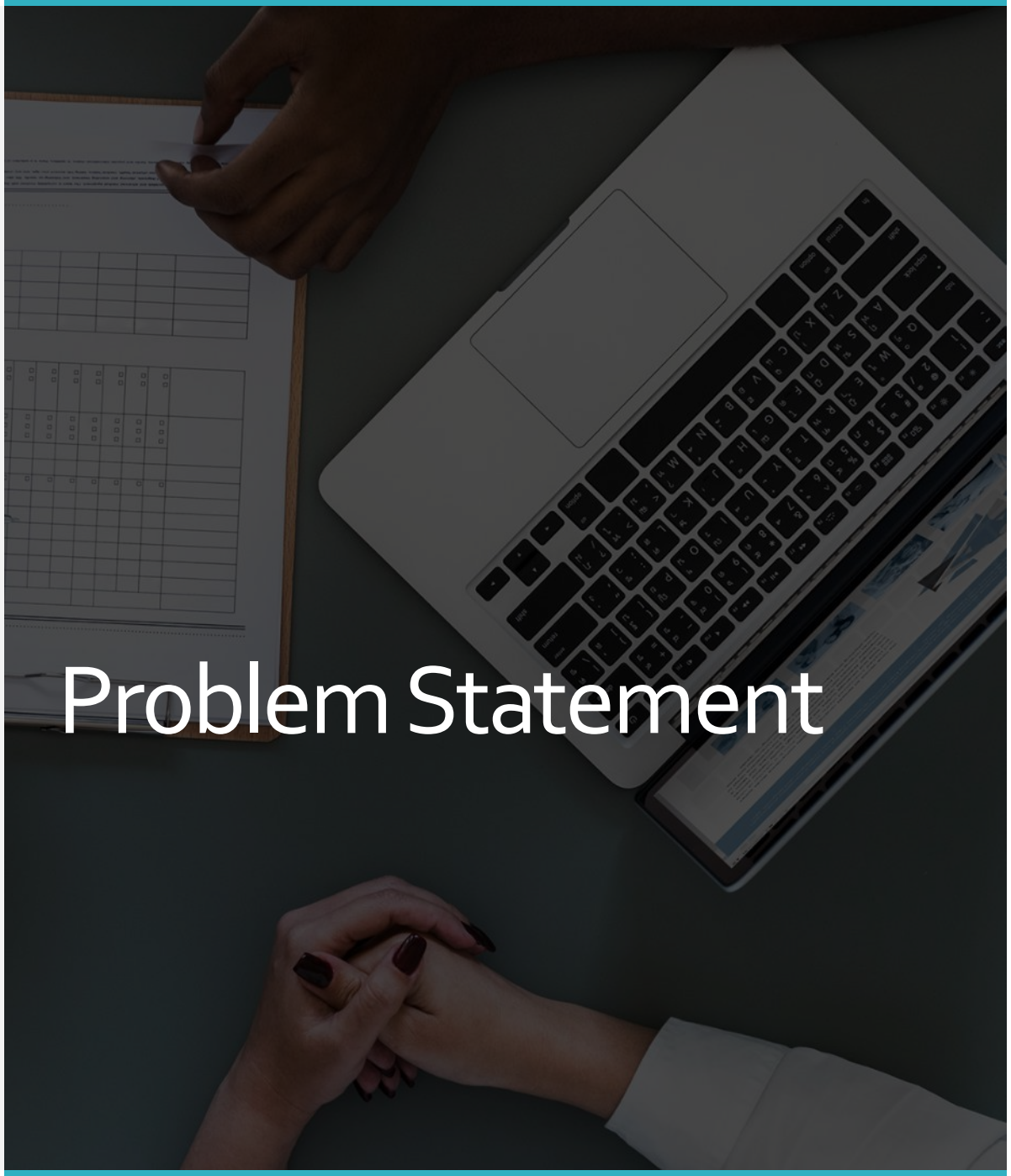
Transmission Path

Public Health Impact

- 8 in 10 are asymptomatic
- 1 in 5 develop symptoms
  - Fever, headache, tiredness, body aches, nausea, vomiting, rash
- 1 in 150 develop serious complications that can result in death
- No vaccine available

Problem Statement

- As mosquito control is an expensive exercise, the state would like our team to propose a cost-effective plan for pesticide deployment.

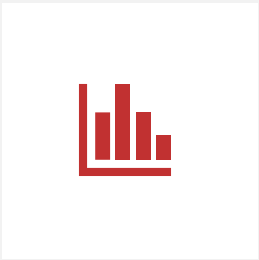


Data

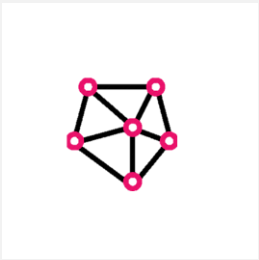
- Trap
  - Locations and WNV test results
  - May to October; 2007, 2009, 2011, 2013
- Spray
  - GIS data for spray efforts
  - 2011 and 2013
- Weather
  - Weather conditions from 2 weather stations
  - 2007 to 2014

Our Problem-Solving Process

Data Preprocessing

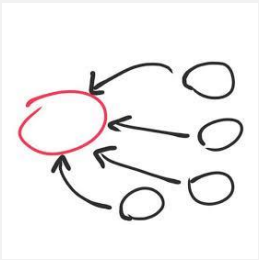


EDA



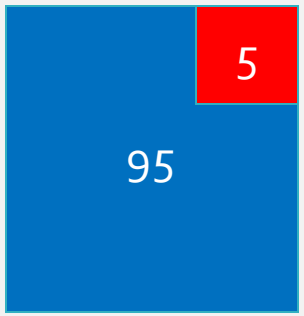
Modelling

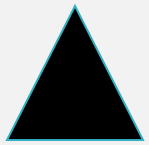
Cost-benefit analysis



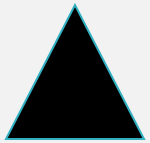
Conclusion & Recommendation

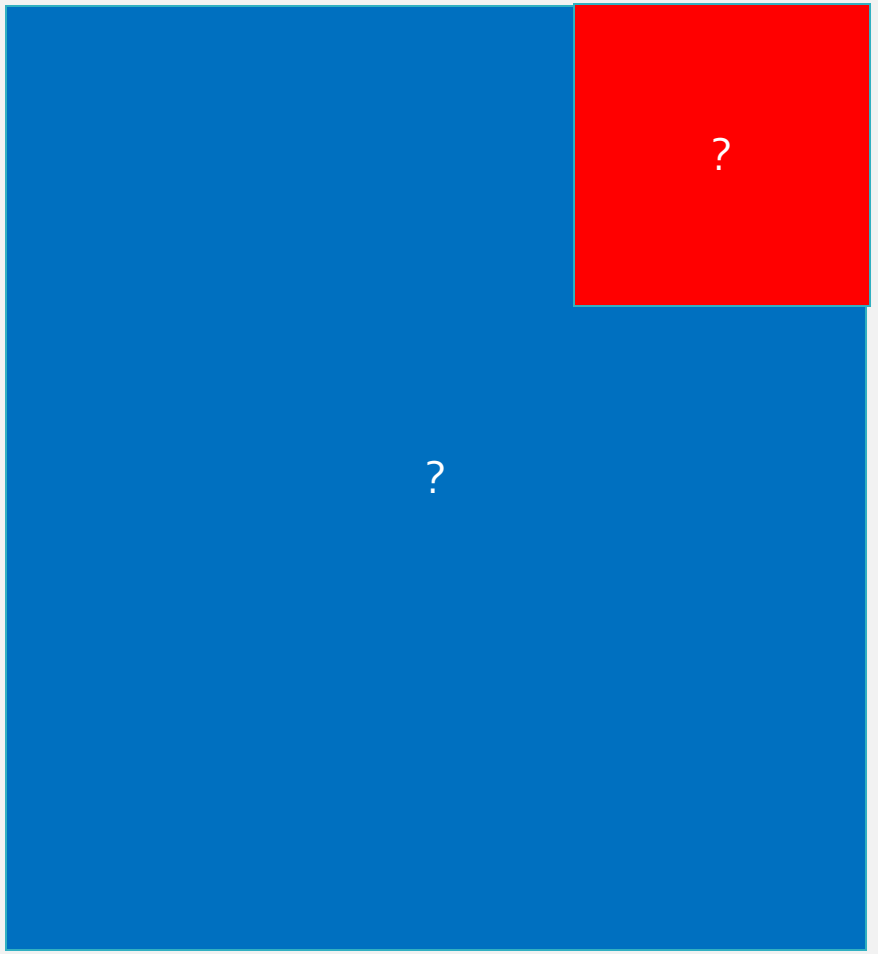
Limitation – Imbalanced Dataset



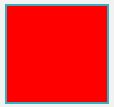


Holdout



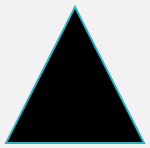


** After Train-test-split*

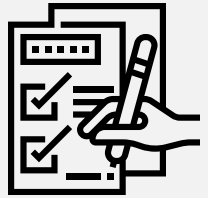


- Virus (Positive)

- No Virus (Negative)

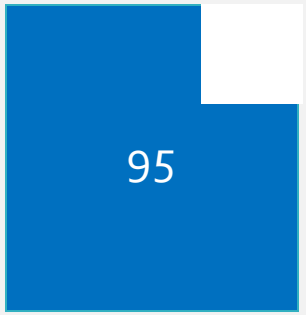


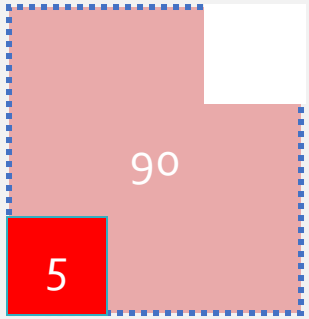
- Model

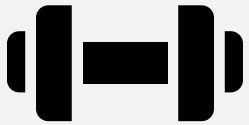


Test

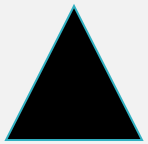
Limitation – Possible solutions (**Oversampling**)



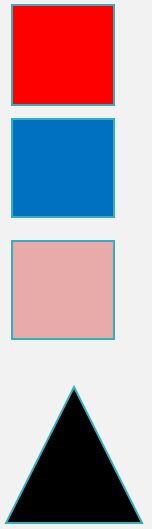




** Before Train-test-split*



Holdout



- Virus (Positive)
- No Virus (Negative)
- Virus - *generated* (Positive)
- Model

Limitation – Possible solutions (**Emphasize features**)

Based on research about ideal conditions:

- 1. Higher humidity ->
- 2. Stagnant water is ideal ->
- 3. Weather conditions ->
- 4. Virus thrives in the sun ->
- 5. Temperature (70-80F) ->

Limitation – Possible solutions (**Emphasize features**)

Based on research about ideal conditions:

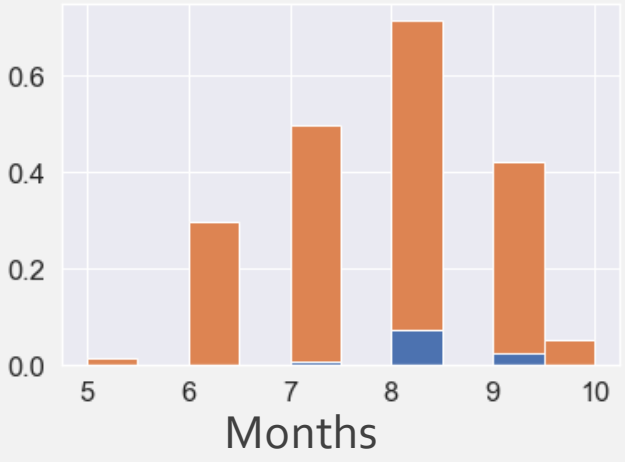
- 1. Higher humidity -> WetBulb, DewPoint
- 2. Stagnant water is ideal -> Wind speeds
- 3. Weather conditions -> Seasons -> months/weeks
- 4. Virus thrives in the sun -> Daytime
- 5. Temperature (70-80F) -> Tavg

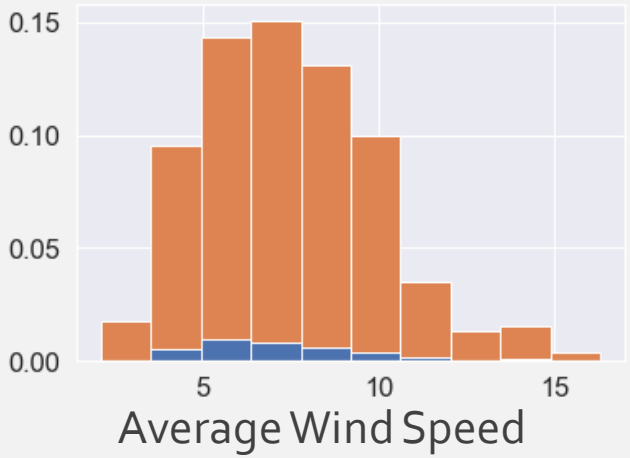
Limitation – Possible solutions (**Emphasize features**)

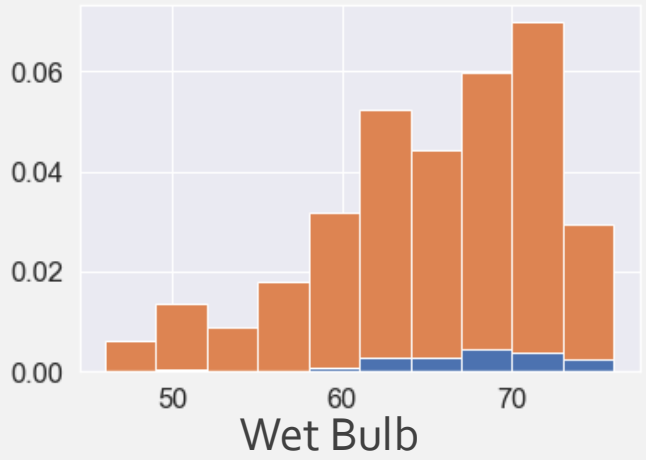
Based on research about ideal conditions:

- 1. Higher humidity -> WetBulb, DewPoint
- 2. Stagnant water is ideal -> Wind speeds
- 3. Weather conditions -> Seasons -> months/weeks
- 4. Virus thrives in the sun -> Daytime
- 5. Temperature (70-80F) -> Tavg

- Virus (Positive)
- No Virus (Negative)

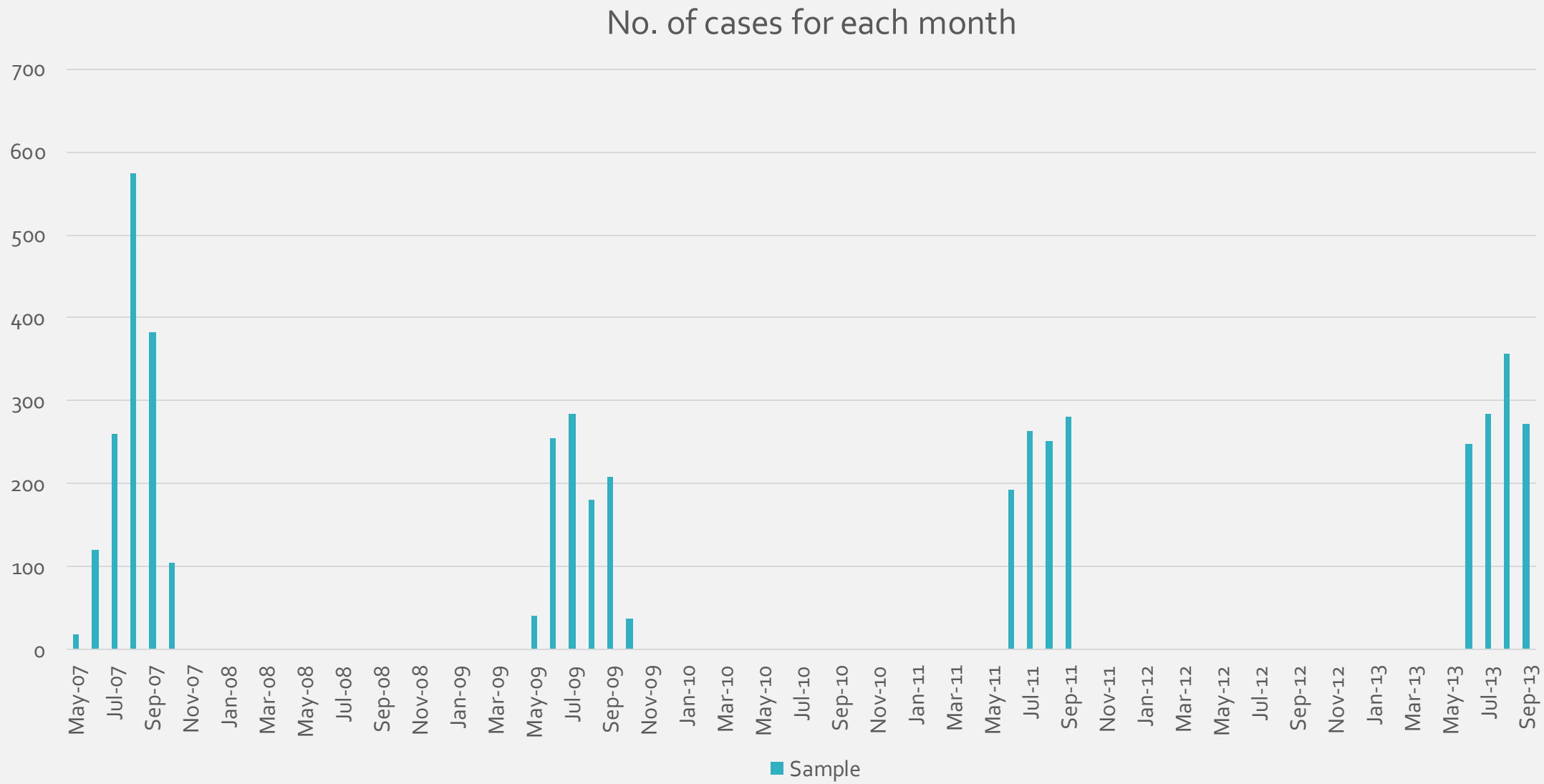






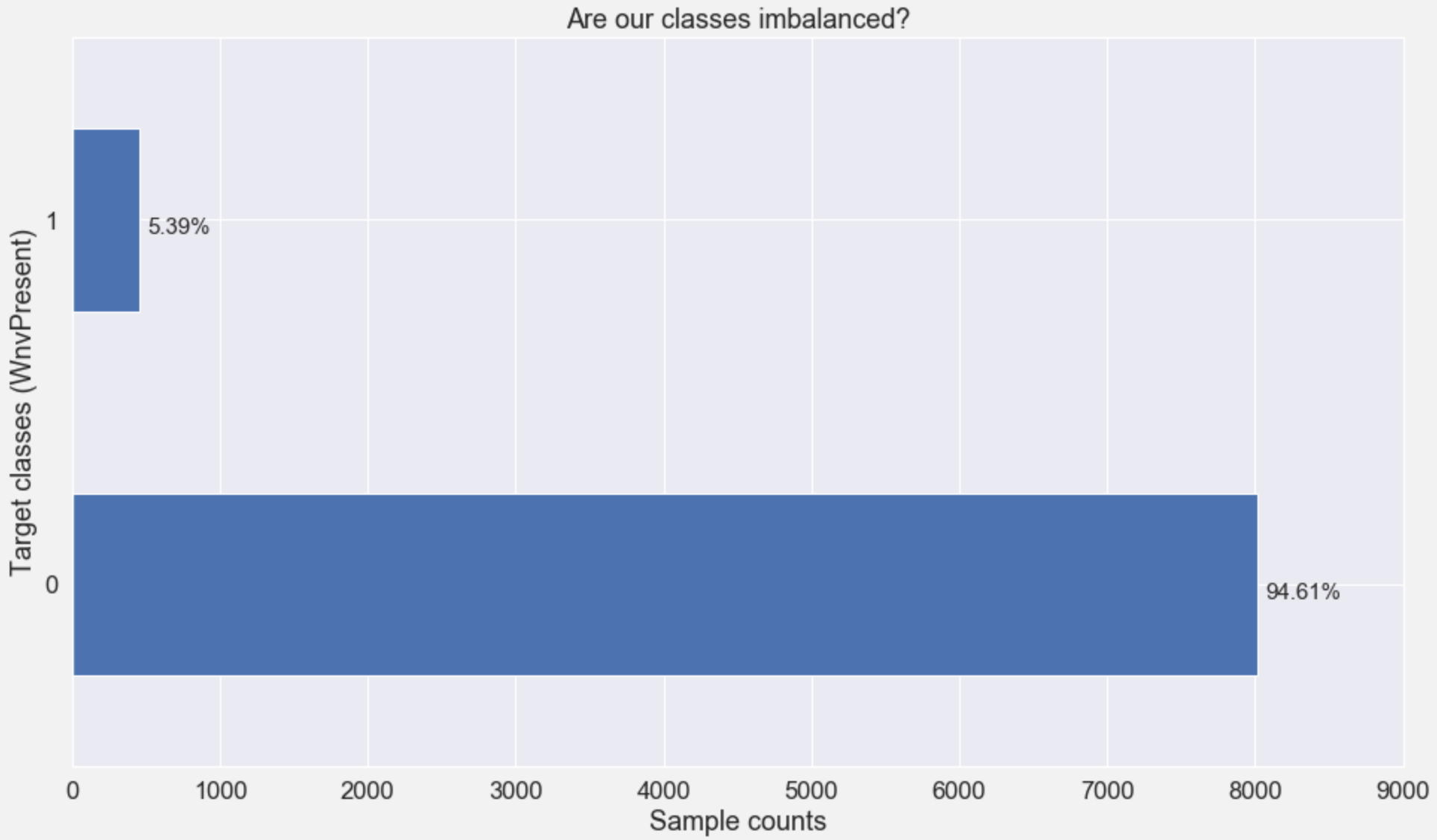
Limitation – Train dataset measured periods and sample size

Practice caution when analysing data due to skewed dataset.



Limitation – Train dataset measured periods and sample size

Practice caution when analysing data due to skewed dataset.

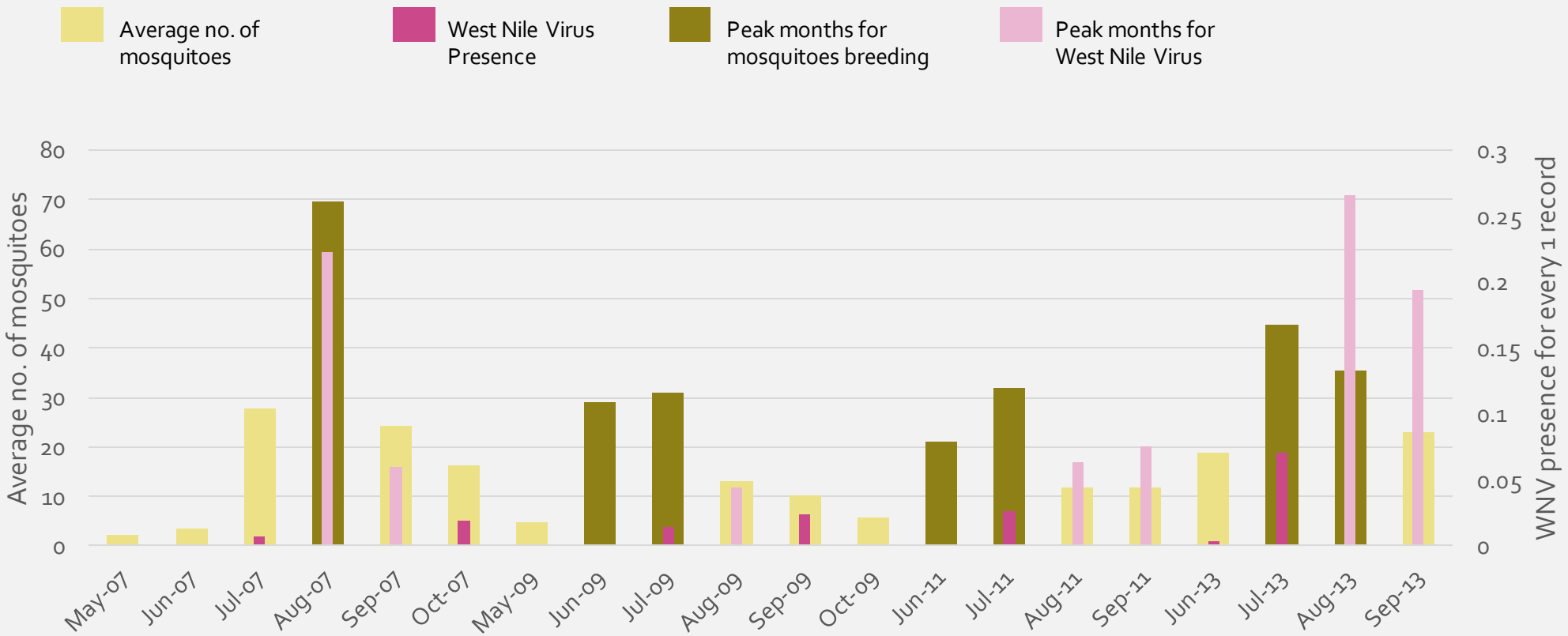


What we know

Exploratory Data Analysis

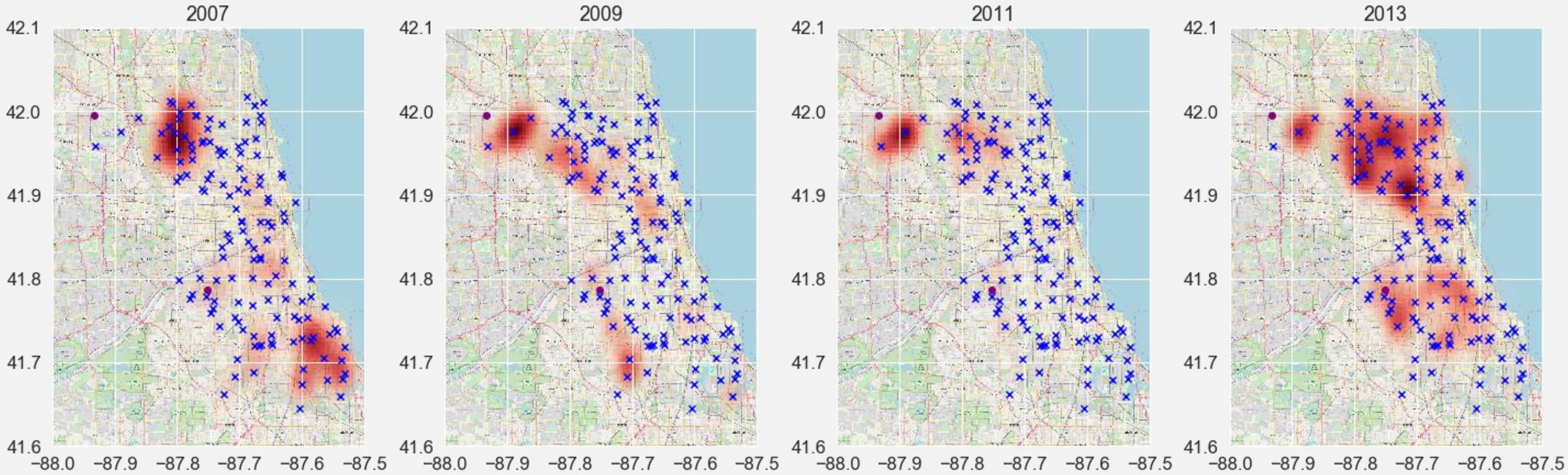
Through June to August were the peak months for mosquitoes breeding, while August to September were the peak for WNV.

Chart: Periods with highest average no of mosquitoes & WNV



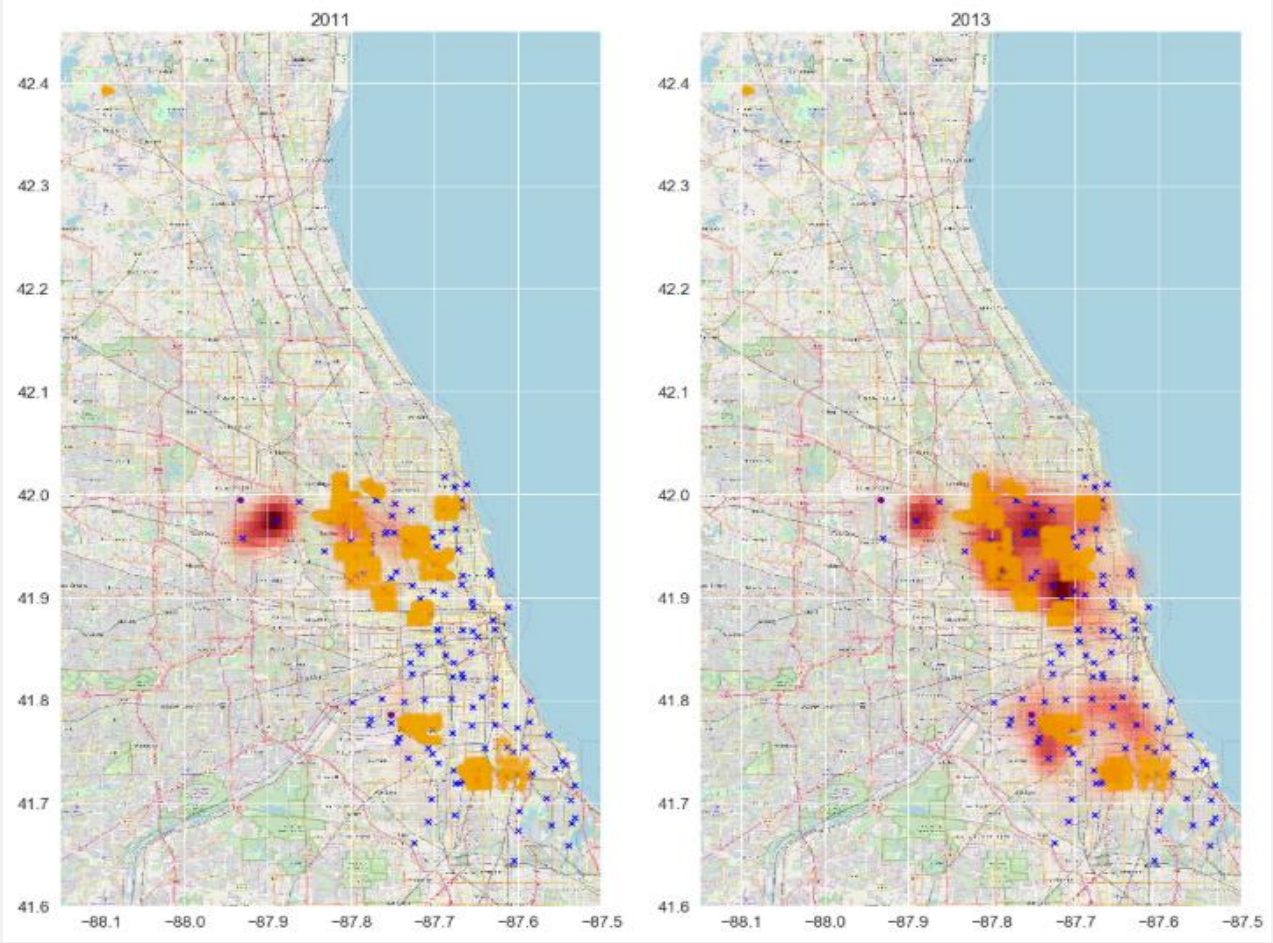
WMV was widely spread in 2007 and 2009. The situation was controlled in 2011 but got worse in 2013

Chart: Location heatmap of WNV in Y2007, 2009, 2011 and 2013



Spraying doesn't seem to be very effective in controlling the spread of WNV.

Chart: Location heatmap of spraying activities in Y2011 and 2013



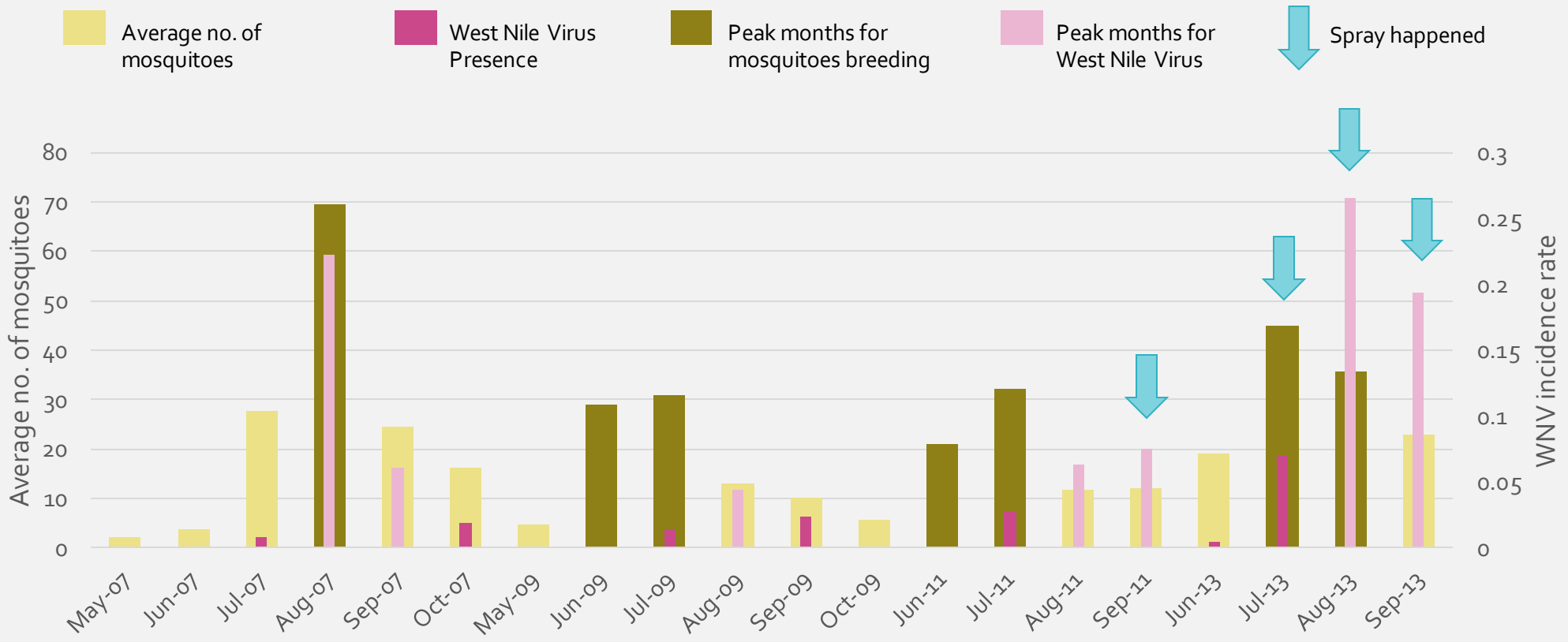
X Trap location

Presence of West Nile Virus

Spray location

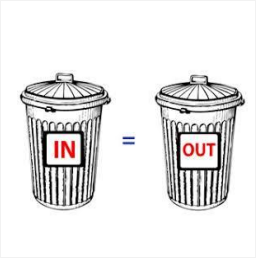
Average no. of mosquitoes drop after Jul-13 after the start of spraying activities. This suggests time of spray could be crucial to prevent mosquitoes breeding.

Chart: Periods with highest average no of mosquitoes & WMV



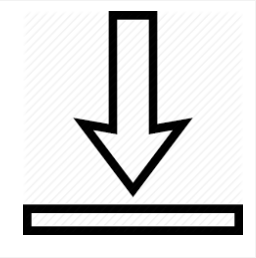
Modelling

Our Modelling Strategy



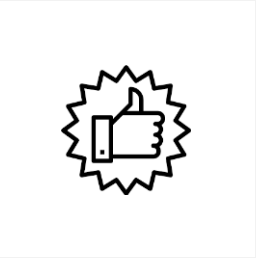
Avoid GIGO Situation

By doing necessary data cleaning and feature engineering



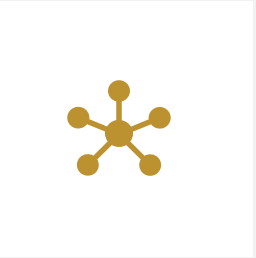
Baseline

Have a baseline model on hand for comparison



Best Parameters

Finding the best parameters for each model



Ensemble

Combine the force of multiple models with their best parameters

Our Evaluation Metric

ROC-AUC and

F1 score

Our Models

Logistic Regression

```
True Negatives: 1791
False Positives: 703
False Negatives: 417
True Positives: 2067
Precision: 0.7462093862815884
Recall: 0.8321256038647343
F1 score: 0.7868290826037306
roc_auc_score: 0.8543011278598427
```

Kaggle: 0.64

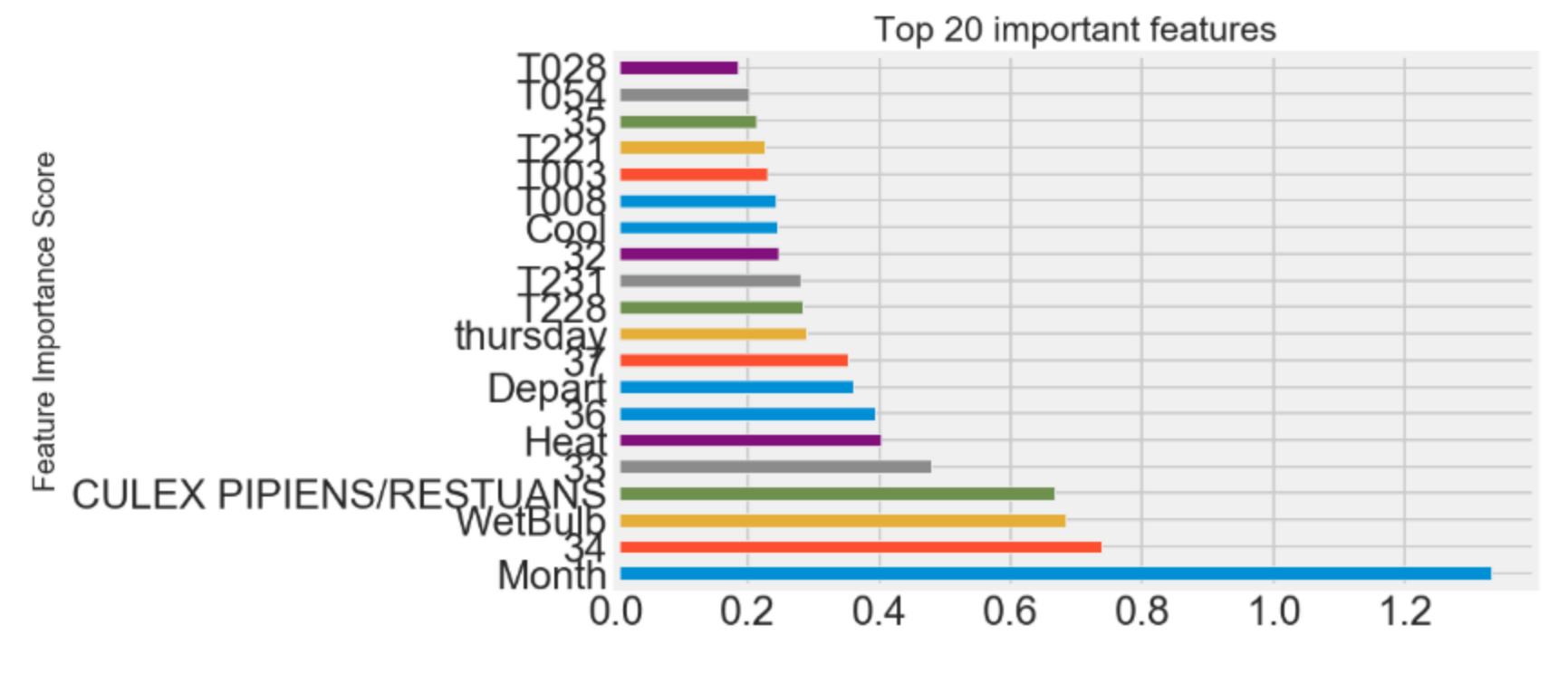
Logistic Regression + SVM

```
True Negatives: 1756
False Positives: 738
False Negatives: 345
True Positives: 2139
Precision: 0.7434827945776851
Recall: 0.8611111111111112
F1 score: 0.7979854504756575
roc_auc_score: 0.8543011278598427
```

Better F1 Score

Kaggle: 0.74

Features Importance



Cost-Benefit Analysis

Costs of Inaccuracy

WNV detected when there is not	No WNV detected when there is
• Ecological impact from unnecessary spraying • Productivity loss from extra clinic visits • Reduced tourism activities	• Risk of WNV epidemic • Strain on healthcare resources • Loss of confidence in public health officials

Cost-Benefit to Spraying

Cost	Benefit

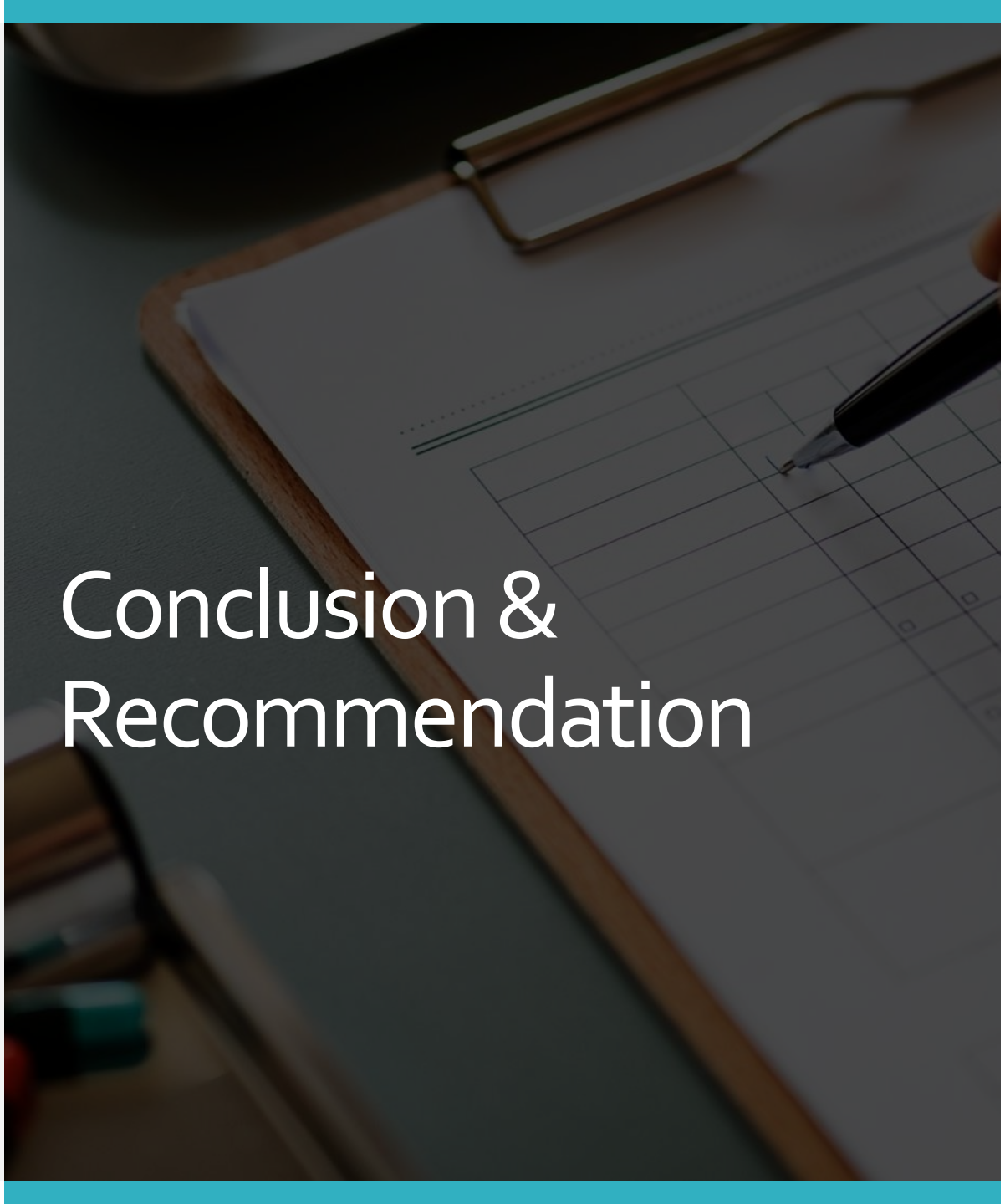
• $ 701k
  • Aerial Spray
	• $16k
	  • Medical care
	• $3k
	  • Lost productivity
	• $300k
	  • Lifetime lost productivity (death)

1 spray : 37 non-death cases or;
1 spray : 2.5 death cases

- WNV remediation efforts would be more effective in limiting illness and death associated with human infection if before the occurrence of WNV.

- What can we do with our model?
  - Identify locations that are more susceptible to mosquitoes breading and WNV and perform preventive measures – such as spraying activities.

- One thing to note is that there are different kind of spraying treatments, hence it would also be good to know the development stages of the mosquitoes collected so that spraying treatments can be more targeted and effective.



Conclusion & Recommendation

Future Developments

How will we scale in the future

Feature Engineer

- Research on weather variables further to use them more effectively
- Add ARMA (Auto Regressive Moving Average)

Gradient Boosting

- Try gradient boosting to optimize model

Spray Data

- Insufficient data to conclude effects of spray

References

Barber, L. M., Schleier, J. J., & Peterson, R. K. (2010). Economic Cost Analysis of West Nile Virus Outbreak, Sacramento County, California, USA, 2005. *Emerging Infectious Diseases*, *16*(3), 480-486. https://dx.doi.org/10.3201/eid1603.090667.

Staples, J. E., Shankar, M. B., Sejvar, J. J., Meltzer, M. I., & Fischer, M. (2014). Initial and long-term costs of patients hospitalized with West Nile virus disease. *The American journal of tropical medicine and hygiene*, *90*(3), 402–409. doi:10.4269/ajtmh.13-0206


Thank You