

# Plan4Demand

*Provide impactful & effective solutions*



Richelle-Joy Chia, Er Jie Yong, Meriky Lo



# Our team



**Richelle-Joy Chia**

Team Lead Officer  
(TLO)



**Er Jie Yong**

Team Modeling Officer  
(TMO)



**Meriky Lo**

Team Financial Officer  
(TFO)

# Problem Statement



To help the City of Chicago Public Health Department optimize measures to tackle financial burdens caused by WNV related diseases

# Table of contents

01

Background

04

Modeling

02

Data Cleaning  
and EDA

05

Cost Benefit  
Analysis

03

Feature Engineering

06

Conclusion and  
Recommendations



# 01



## Background

# WNV history

Mosquito-borne virus - Incurable  
1 in 5 develop mild symptoms  
1 in 150 develop severe symptoms



1937

First identified in  
Uganda (Eastern  
Africa)

1999

First  
discovered in  
New York  
(USA)

2001

First detected  
in Chicago

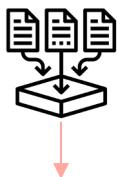


# WNV history

2005

Started peaking in Chicago

2007-2014



2022

Still exists!

2012

Number of cases surpassed by this season

# Our data



Train Test

2007	2008
2009	2010
2011	2012
2013	2014



Weather

2007-2014



Spray

2011  
2013

*Note: our train data set is imbalanced (95% virus absent, 5% virus present)*

# Understanding our subject



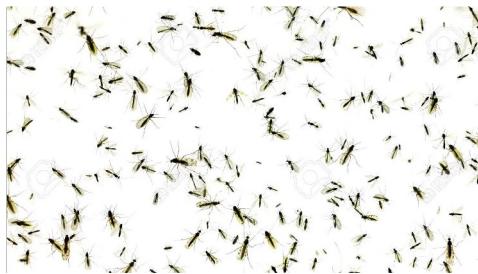
- Life cycle of 8 - 10 days
- Lay eggs in water
- Culex mosquitoes have been known to fly up to 3.2 km
- Temperature, Precipitation, Humidity and Wind correlates positively with predicting WNV



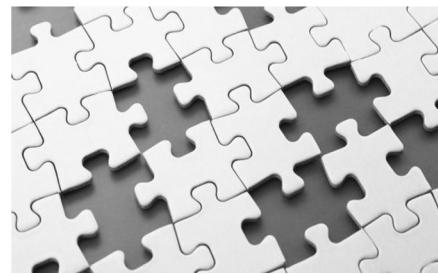
# 02

## Data Cleaning and EDA

# Data Cleaning



Merged rows with  
more than 50  
mosquitos

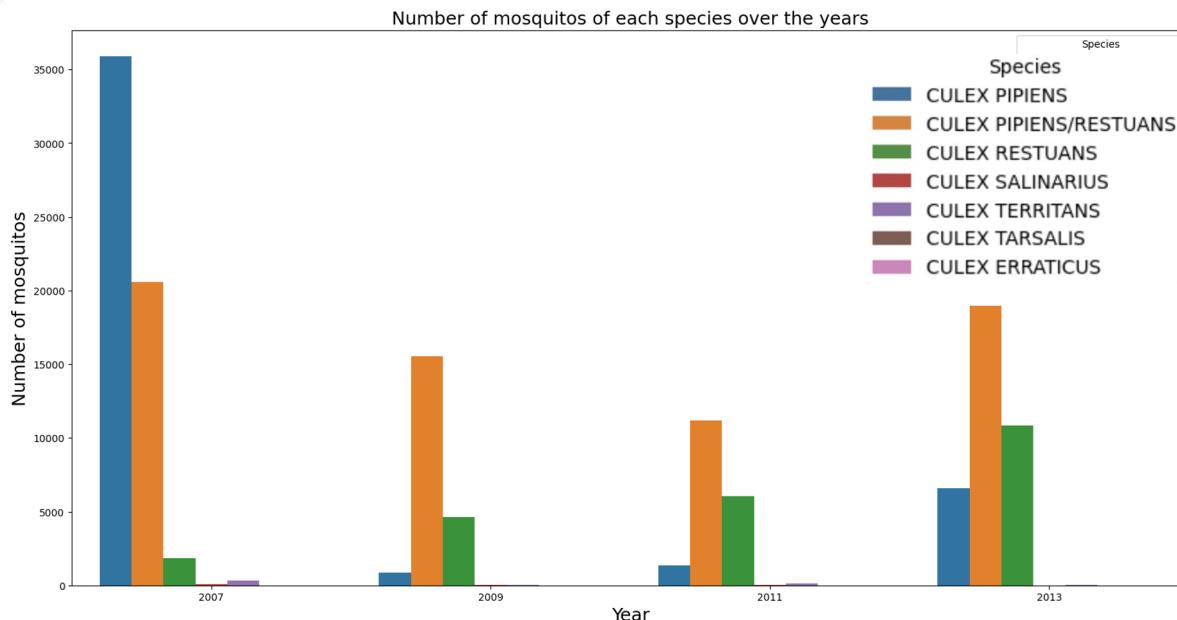


Removed columns  
with more than  
50% missing data



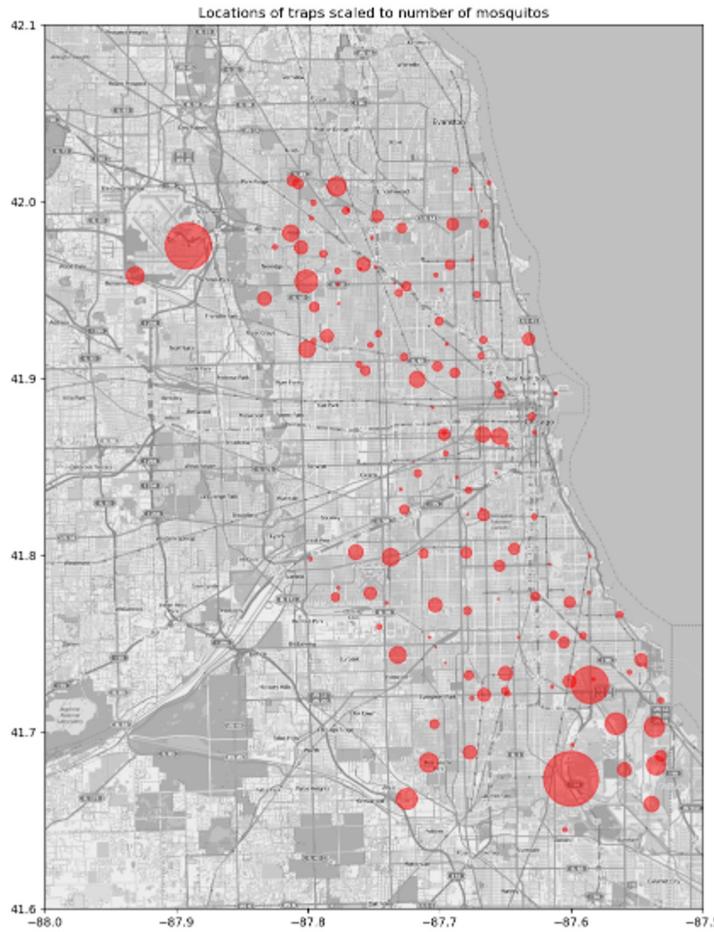
Impute means for  
those with lesser  
missing values

# Number of mosquitoes

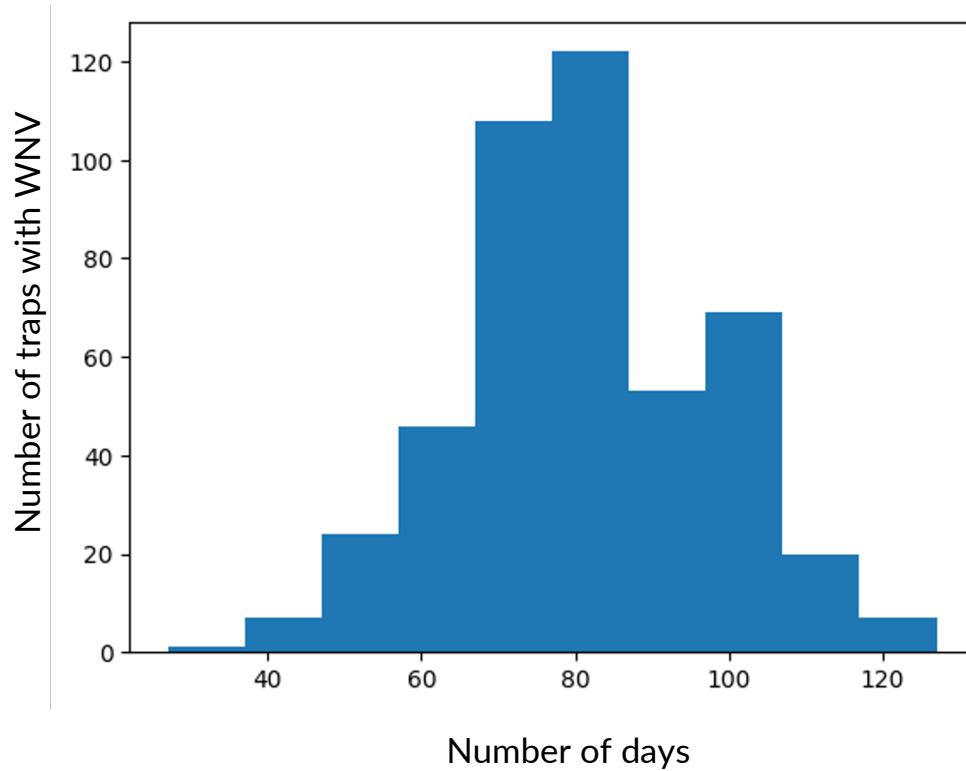


WNV Mosquito Species	
Culex Pipiens/Restuans	262
Culex Pipiens	240
Culex Restuans	49

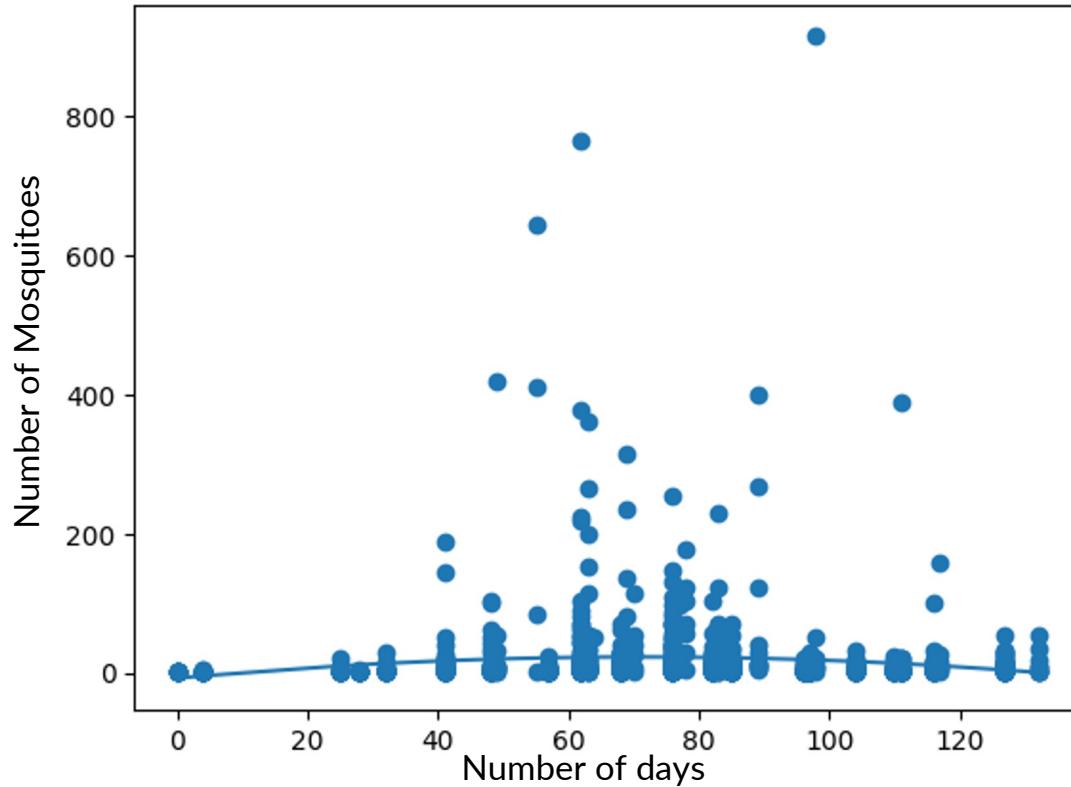
# Location of traps scaled to number of mosquitoes



# Number of traps with WNV detected

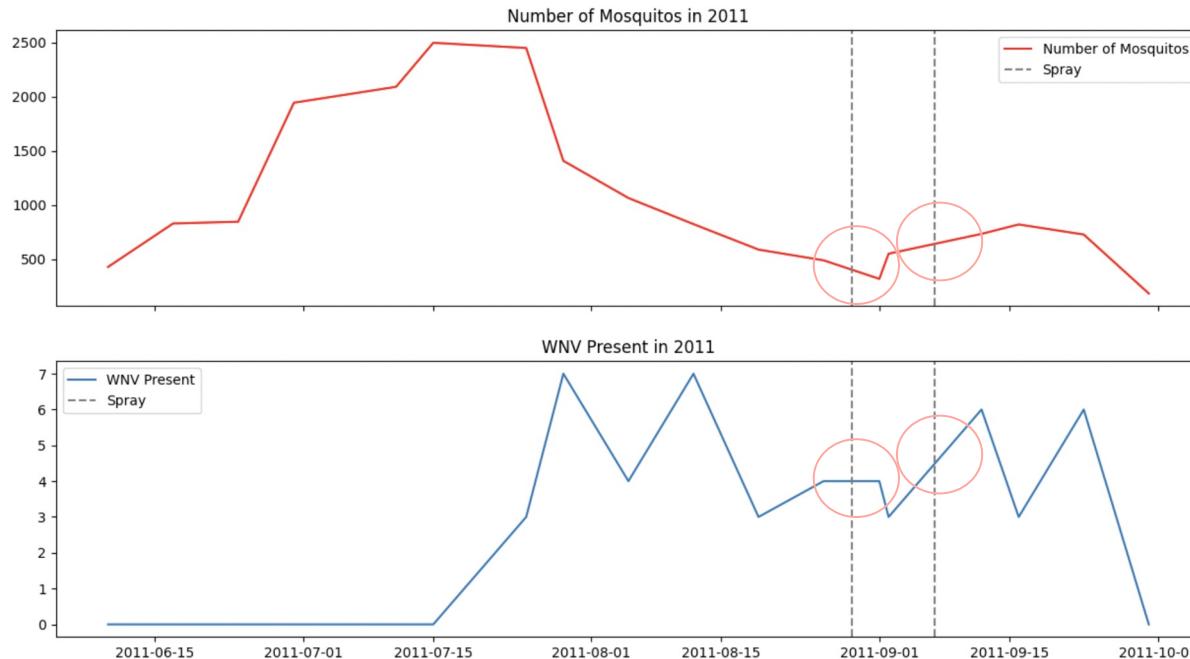


# Number of mosquitoes created for test data

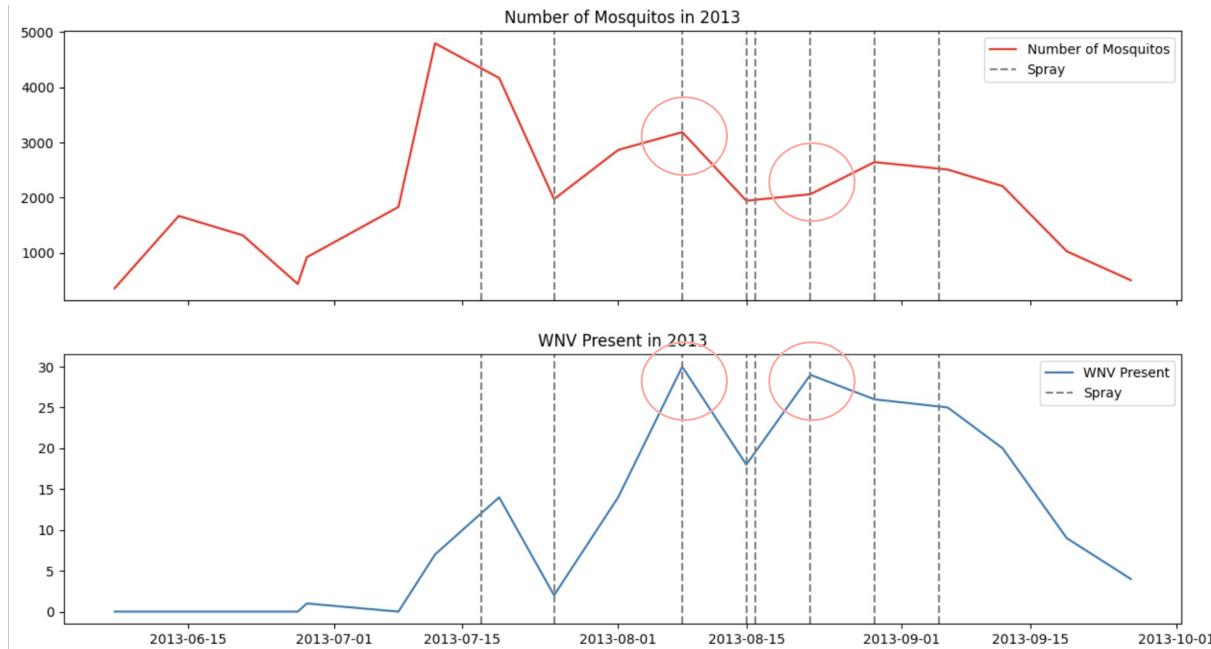


*Note: an unsuccessful attempt due to other unforeseen outbreaks*

# Is the spray helpful in 2011?



# Is the spray helpful in 2013?





# 03

## Feature Engineering

# Feature Engineering

## WetDry



Binary Classification  
based on codesum

## Humidity

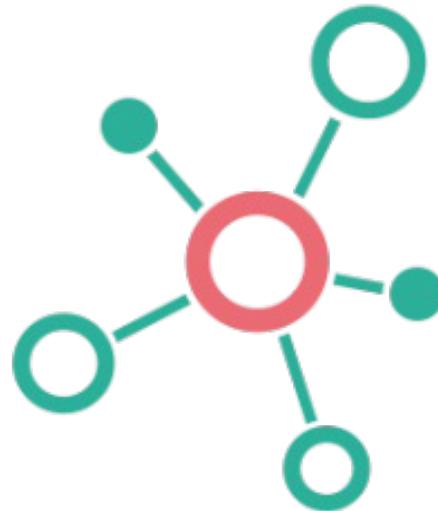


Derived from  
temperature & Dewpoint

## Trap Sprayed



Identify traps being  
sprayed within 300 ft and  
up to 30 days



## Weather Rolling Avg

9 days rolling average of  
weather data



## Days since June 1

Cumulative count  
since start of data  
set (June 1)



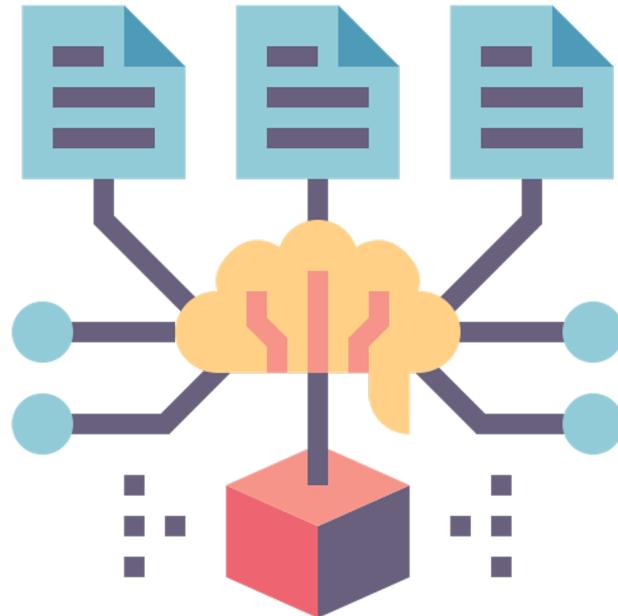
## Trap Matrix

Weighted matrix  
of traps within  
3km



## Normal PDF

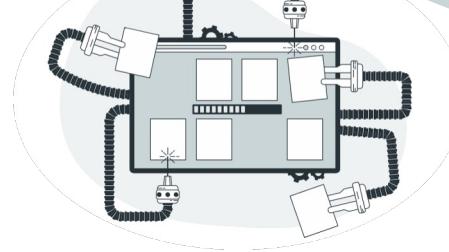
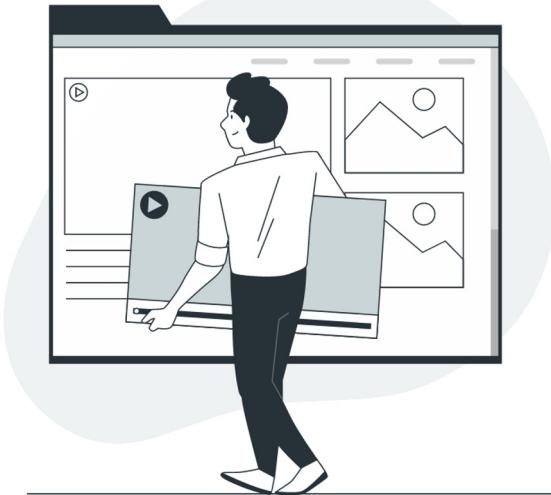
Approximate PDF of  
WNV per year based  
on normal distribution



# 04

## Modeling

# Modeling



## Base parameters

- Fix Imbalance
- Remove multicollinearity
- Polynomial Features

# Modeling

- Standard Scaler
- Feature Selection

- PowerTransformer
- Feature Selection

- Standard Scaler
- PCA

- PowerTransformer
- PCA



Create Cluster



Remove  
Outliers



Remove  
Polynomial

2nd weighted  
classification

Spatial Network  
Trap Matrix

Tune &  
Ensemble

# Results

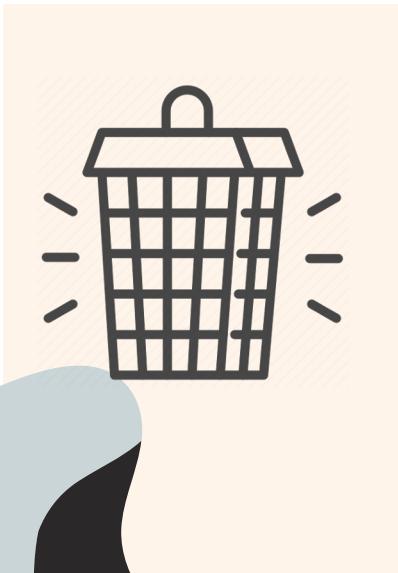
Hyperparameter	Best Model	Train/Test Accuracy	Train/Test AUC	Private Score	Public Score
•Standard Scaler •Feature Selection	Linear Discriminant Analysis	0.6863	0.7508	0.67711	0.70093
•PowerTransformer •Feature Selection	Logistic Regression	0.6944	0.7614	0.69228	0.71435
•Standard Scaler •PCA	Linear Discriminant Analysis	0.6681	0.7456	0.65416	0.67500
•PowerTransformer •PCA	Linear Discriminant Analysis	0.6544	0.7704	0.67743	0.69191
•No polynomial feat •Tune/Ensemble	Logistic Regression	0.6795	0.7908	0.70854	0.72452
•Weighted •Tune/Ensemble	Logistic Regression	0.7218	0.9305	0.70935	0.72172

# Important Features

Based on model using feature\_selection instead of PCA (black box)

01

Traps  
(T071, T138, T082)



02

Species Type



03

Week, Month

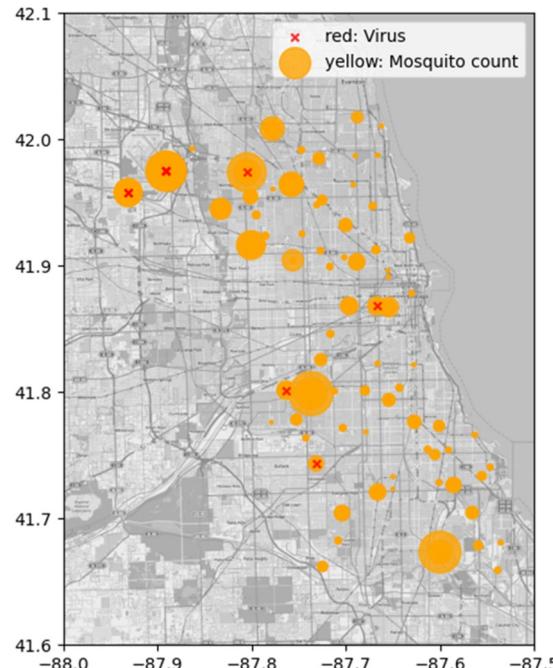




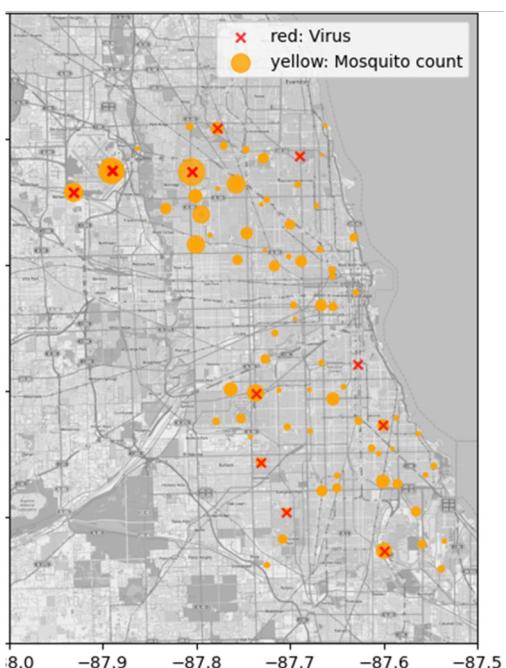
# 05

## Cost Benefit Analysis

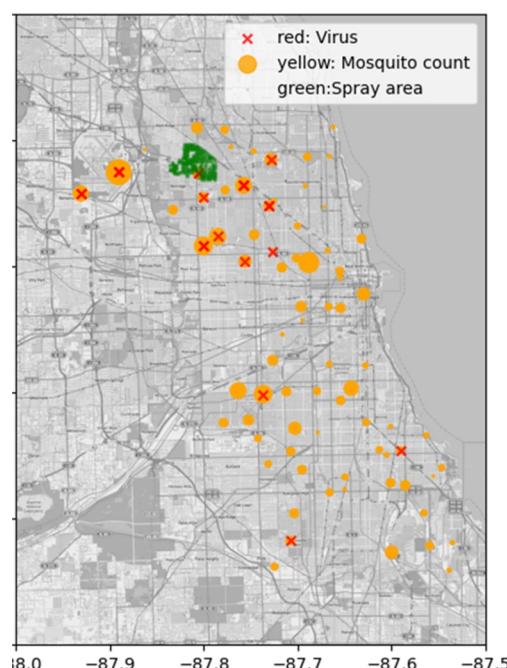
# Mosquito count, WNV Presence, Spray 2011



JULY

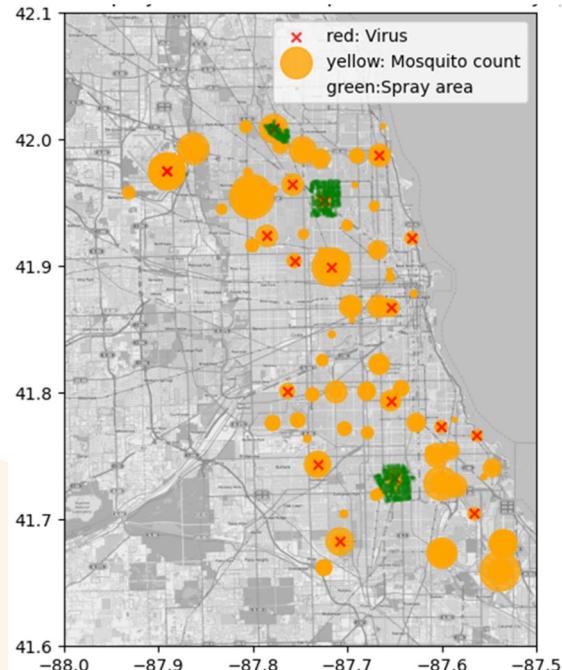


AUGUST

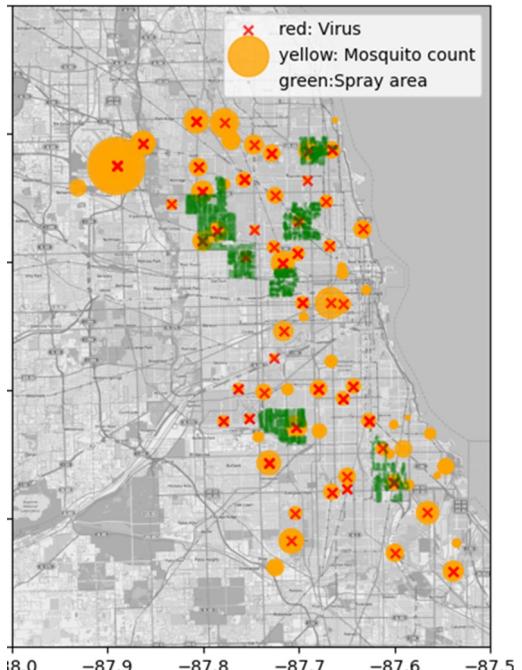


SEPTEMBER

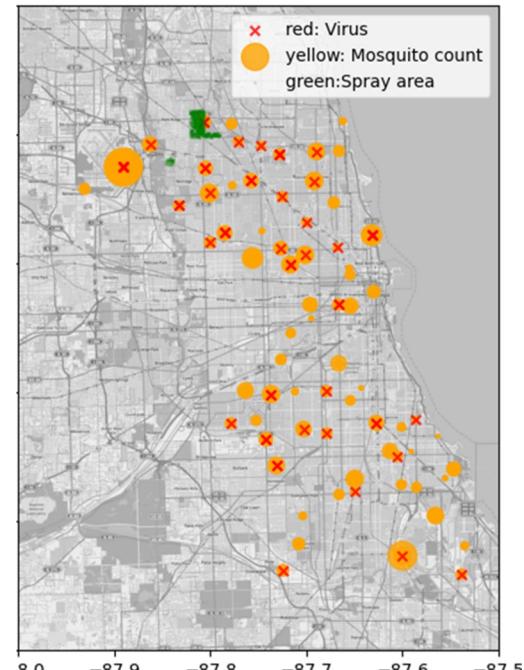
# Mosquito count, WNV Presence, Spray 2013



JULY



AUGUST

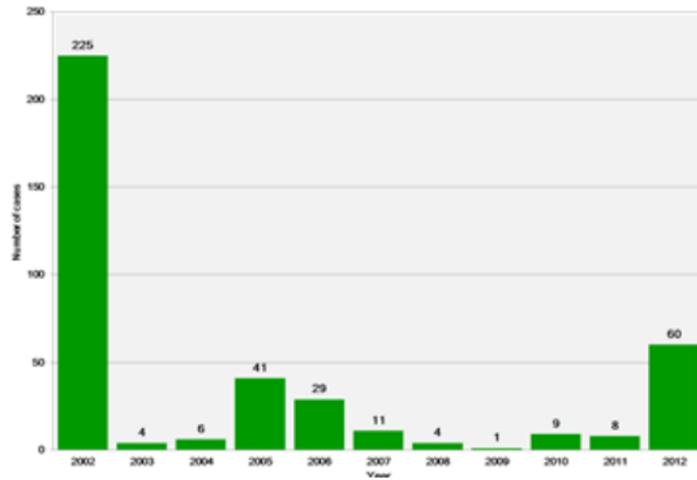


SEPTEMBER

# ISSUES

- Public health
- Economic burden

**Figure 1.** Number of reported confirmed and probable cases of West Nile virus among Chicago residents by year, 2002-2012.



# All Spray

Table 2

Estimated inpatient and outpatient economic costs of WNND cases, Sacramento County, California, 2005\*

Item	Cost per case†	which cost applies‡	which cost applies§	Total cost for all cases	treatment/service were used in all cases
Inpatient treatment costs	\$33,143	46	100	\$1,524,570	\$1,524,570

Price (gallon)	Pounds AI/gallon	Price per Pound	Application Rate/Acre	Cost/Acre	Annual Acres Treated	Annual Cost
----------------	------------------	-----------------	-----------------------	-----------	----------------------	-------------

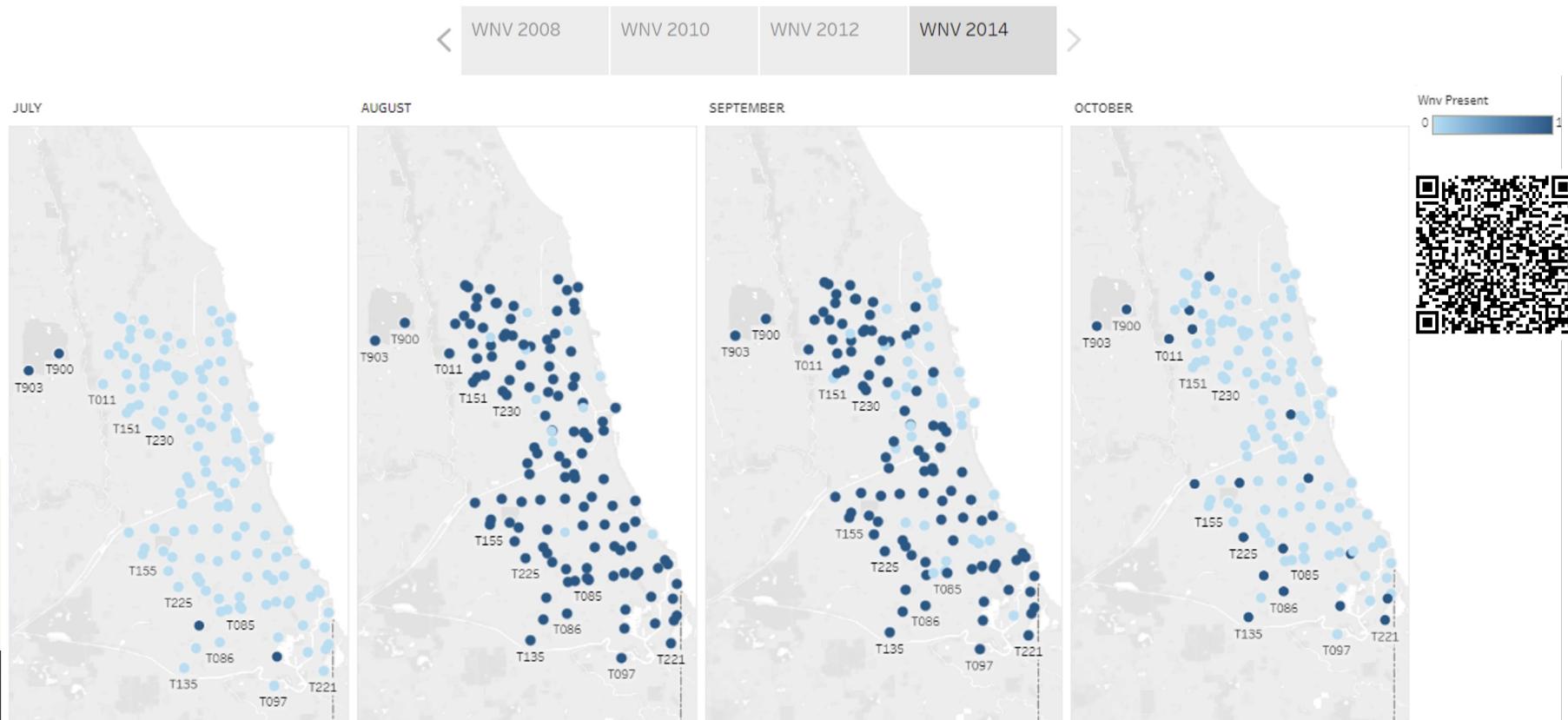
275 gal Zenivex® E20	\$282.00*	1.48	\$190.54	.0035	\$0.67	20,000	\$13,338
275 gal Zenivex® E4	\$78.85*	.3	\$262.83	.0035	\$0.92	20,000	\$18,398
2.5 gal Zenivex® E20	\$296.00*	1.48	\$200.00	.0035	\$0.70	20,000	\$14,000
2.5 gal Zenivex® E4	\$80.75*	.3	\$269.17	.0035	\$0.94	20,000	\$18,842

Mean number of cases in Chicago over 10 years	17
Mean medical cost per case	\$33,143
Total mean medical cost per year	\$563,431

Chicago City Area (acres)	150,000
Pesticide cost per acre	\$0.92
Total cost per month	\$138,000
Assuming 3 sprays/year (Jul, Aug, Sep)	\$414,000

BENEFIT: \$149,431 (26.5%)

# WNV Prediction 2014



# Targeted Spray

MONTH 2014	PREDICTED NO OF TRAPS WITH VIRUS	COST PER MONTH (number of traps x (acre/trap) x (cost/acre))	
JULY	4	$4 \times 1119 \times \$0.92 =$	\$ 4,118
AUGUST	123	$123 \times 1119 \times \$0.92 =$	\$ 126,626
SEPTEMBER	97	$97 \times 1119 \times \$0.92 =$	\$ 99,860
OCTOBER	19	$19 \times 1119 \times \$0.92 =$	\$ 19,560
<b>TOTAL COST PER YEAR:</b>		<b>\$ 250,164</b>	

BENEFIT: \$313,267 (55.6%)

# All Spray vs Targeted Spray

	All Spray	Targeted Spray*
Total mean medical cost per year	\$563,431	\$563,431
Total Spray Cost per year	\$414,000	\$250,164
<b>BENEFIT</b>	<b>\$149,431 (26.5%)</b>	<b>\$313,267 (55.6%)</b>

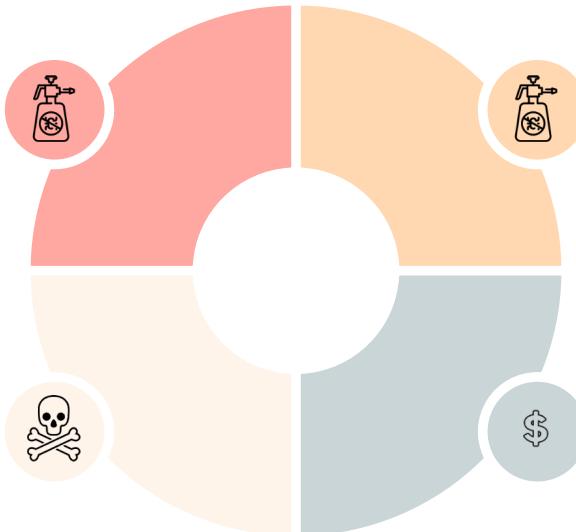
\*Assumption that our model is 100% accurate at predicted trap with virus



# Limitations

Other costs involved  
in spraying

Costs arising from lost  
of productivity due to  
serious illness/deaths

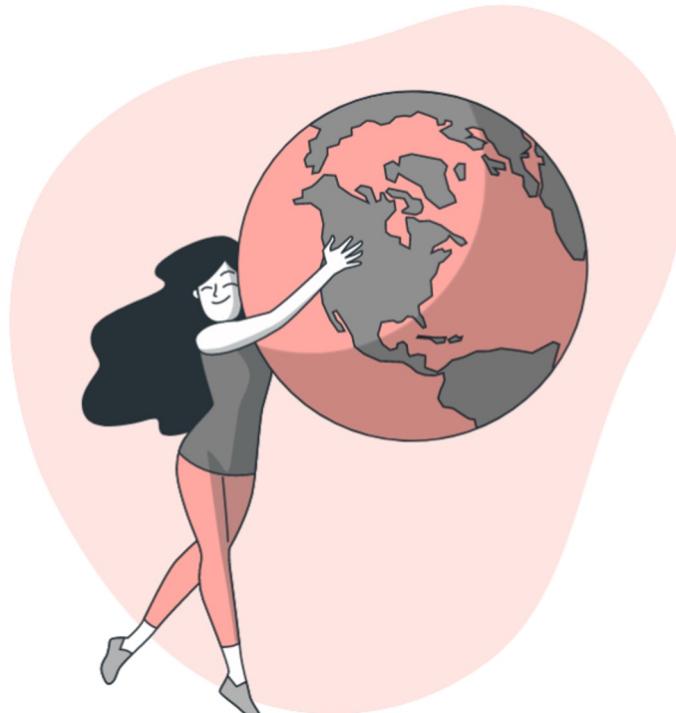


Pesticide is 100%  
effective

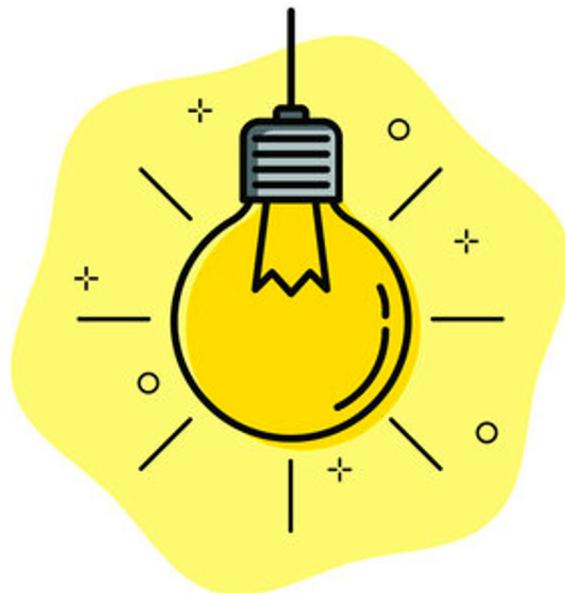
No price for human's  
health/life

# 06

## Conclusion & Recommendations



# Conclusion



## Most important features

Location and date

## Observations from past data

Current spraying is not optimum  
in reducing number of  
mosquitoes/virus

# Recommendations



## Targeted spray

Specific dates and locations  
based on our model  
predictions

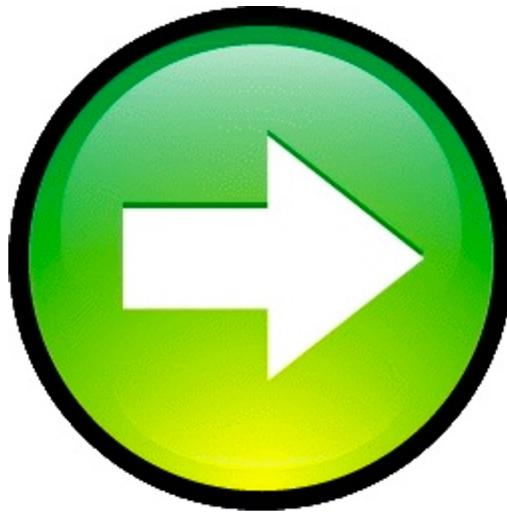
## Identify

Look for more effective  
pesticides

## Preventive measures

Spraying on high risk areas,  
public education

# Next steps



Predict outbreaks timeframe and location

Predict number of mosquitoes

More feature engineering/EDA to identify higher correlated features



**Thank You !**