Project 4: West Nile Virus

Kwek Jun Hong

Nicholas Lim

Malcolm Lau

TABLE OF CONTENTS

01
Introduction and Problem Statement

Problem Statement

- City of Chicago has been having to deal with seasonal upticks of incidences of the West Nile virus
- Predict the prevalence of the West Nile virus amongst mosquitoes within the Chicago city area
- Cost-benefit analysis for application of pesticide

02
Data Cleaning, Feature Engineering & EDA

Data Cleaning

- Column Name Standardization
- Missing Values Imputation
  - *Weather* dataset
    - M, T, *, -: Missing & Trace info
    - No Sunrise/Sunset info for Station 2

Imputed with '0'

	Station	Depth	Water1	SnowFall	Sunrise	Sunset
0	1	0	M	0.0	0448	1849
1	2	M	M	M	-	-
2	1	0	M	0.0	0447	1850
3	2	M	M	M	-	-
4	1	0	M	0.0	0446	1851

Data Cleaning

- Missing Values Imputation
  - *spray* dataset
    - Missing Time info for '2011-09-07'

Removed 'Time' column'

	date	time	latitude	longitude
1030	2011-09-07	NaN	41.987092	-87.794286
1031	2011-09-07	NaN	41.987620	-87.794382
1032	2011-09-07	NaN	41.988004	-87.794574
1033	2011-09-07	NaN	41.988292	-87.795486
1034	2011-09-07	NaN	41.988100	-87.796014
...	...	...	...	...
1609	2011-09-07	NaN	41.995876	-87.811615
1610	2011-09-07	NaN	41.995972	-87.810271
1611	2011-09-07	NaN	41.995684	-87.810319
1612	2011-09-07	NaN	41.994724	-87.810415
1613	2011-09-07	NaN	41.993092	-87.810415

Data Cleaning

- Removing Duplicate Values
  - *spray* dataset
    - Duplicate Values for records in “2011-09-07”

Removed duplicates

	date	latitude	longitude
484	2011-09-07	41.983917	-87.793088
485	2011-09-07	41.983917	-87.793088
489	2011-09-07	41.986460	-87.794225
490	2011-09-07	41.986460	-87.794225
491	2011-09-07	41.986460	-87.794225

543 rows × 3 columns

Feature Selection

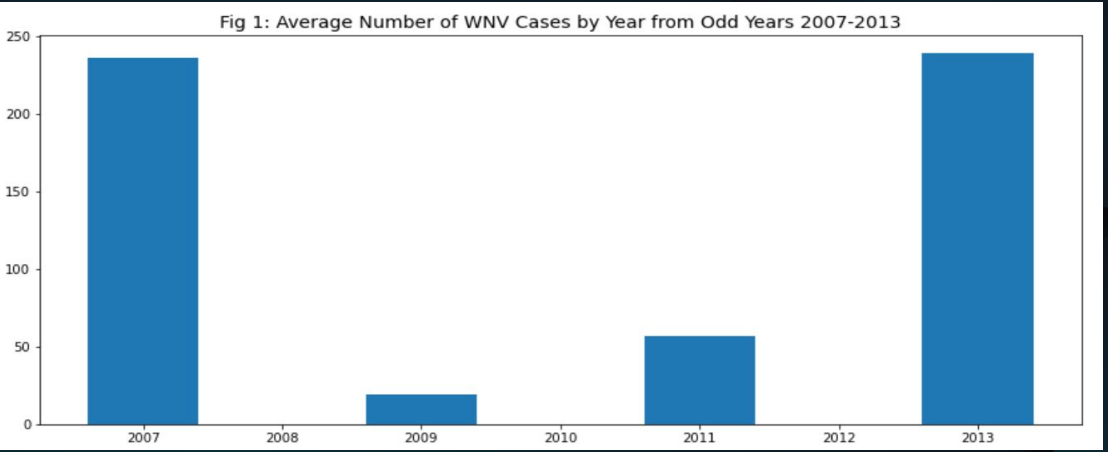
- train features
  - date of observation
  - species of mosquito
  - latitude of trap
  - longitude of trap
  - number of mosquitoes in each trap
- weather features
  - station of observation
  - date of observation
  - average temperature for day
  - dew point for day
  - total precipitation for day
  - time of sunrise
  - time of sunset
  - air pressure at station
  - average wind speed

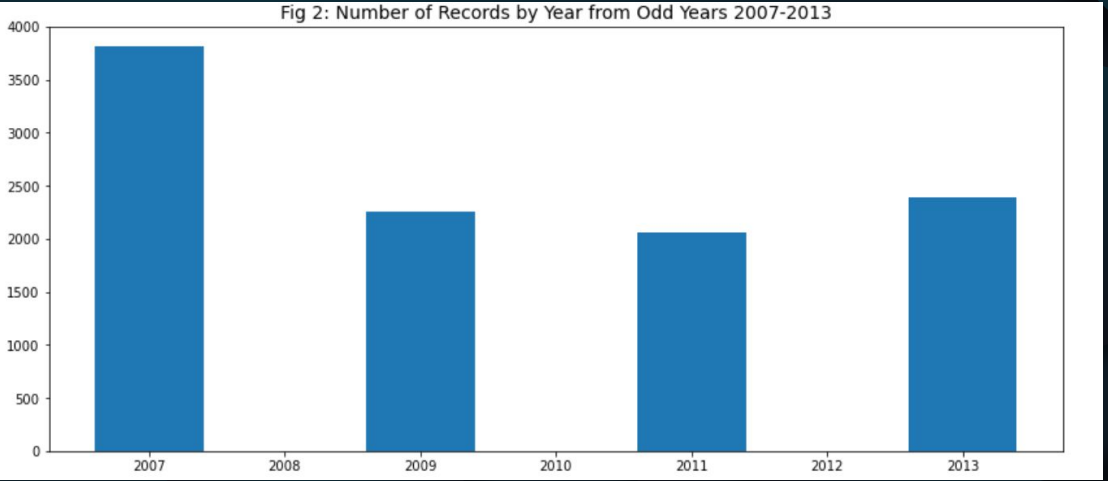
EDA

Macro Trends

- **WNV Cases vs Testing Efforts**

Q: Are we recording more WNV cases because we are testing more?



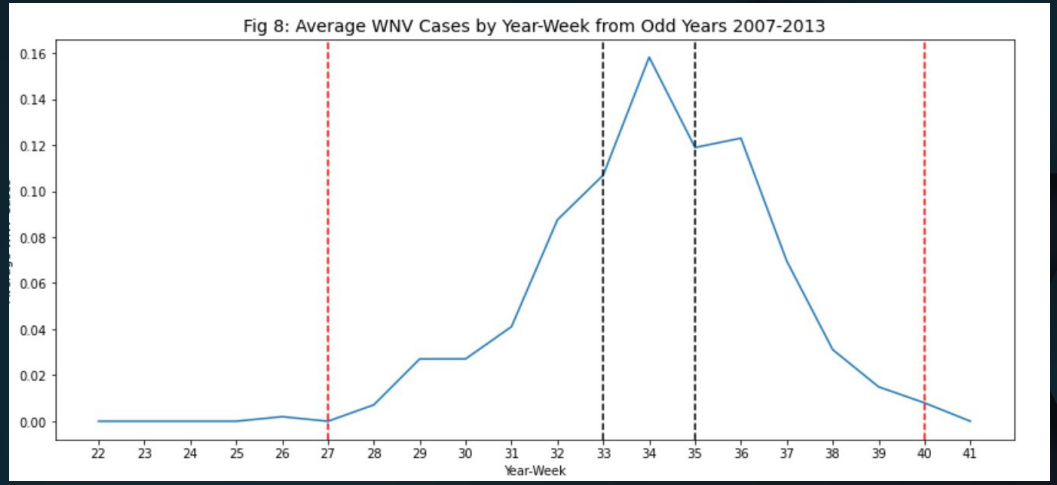


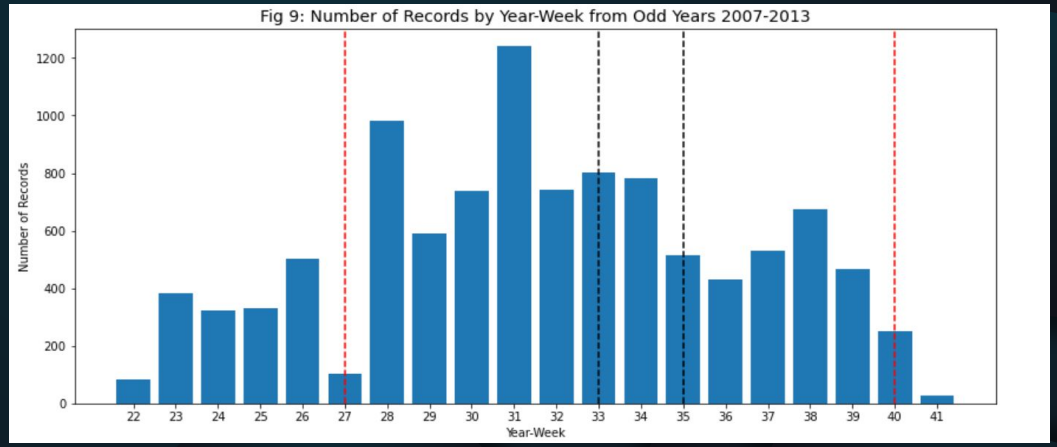
EDA

Seasonal Trends

- **WNV Cases by Year-week**

Q: Are there certain weeks of the year where WNV peaks?





Feature Engineering

- **Daylight Hours**
- **Relative Humidity**
  - Using Tavg, Dew Point
- **Aggregating Station 1 & 2 Information**

Daylight Hours

Sunrise Hours - Sunset Hours

Relative Humidity

$$E_s = 6.11 \times e^{\left(\frac{17.67 \times T}{243.5 + T}\right)}$$

$$E = 6.11 \times e^{\left(\frac{17.67 \times T_{dew}}{243.5 + T_{dew}}\right)}$$

$$\text{Relative Humidity } \% = \frac{E}{E_s} \times 100$$

T = Ambient Temperature in Celsius
T_{dew} = Dew Point in Celsius
E_s = Saturation Vapor Pressure
E = Actual Vapor Pressure

Aggregating Station 1 & 2 Information

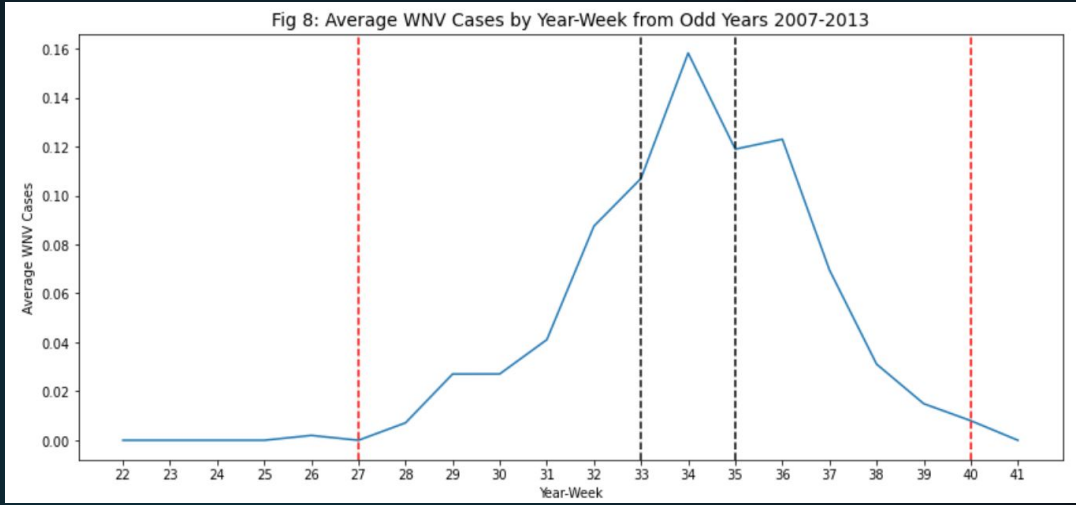
- Taking the mean of all weather features

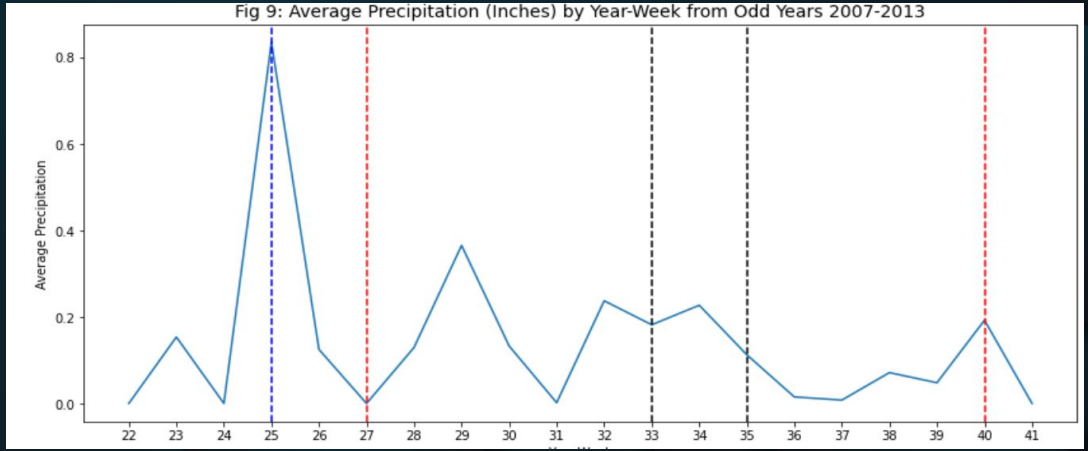
EDA

Seasonal Trends

- **WNV Cases by Year-week**
- **Precipitation**

Q: Are the WNV peaks facilitated by certain weather conditions?



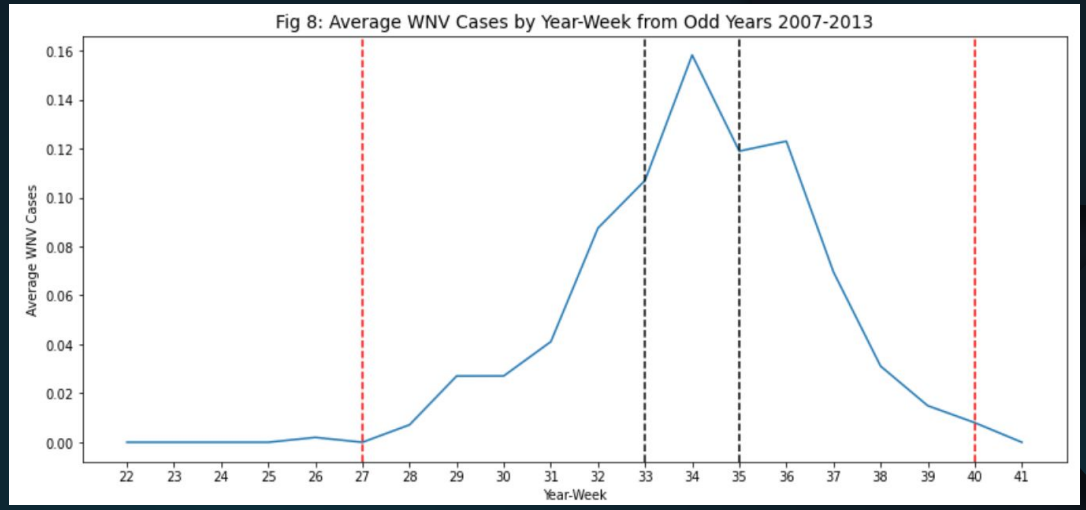


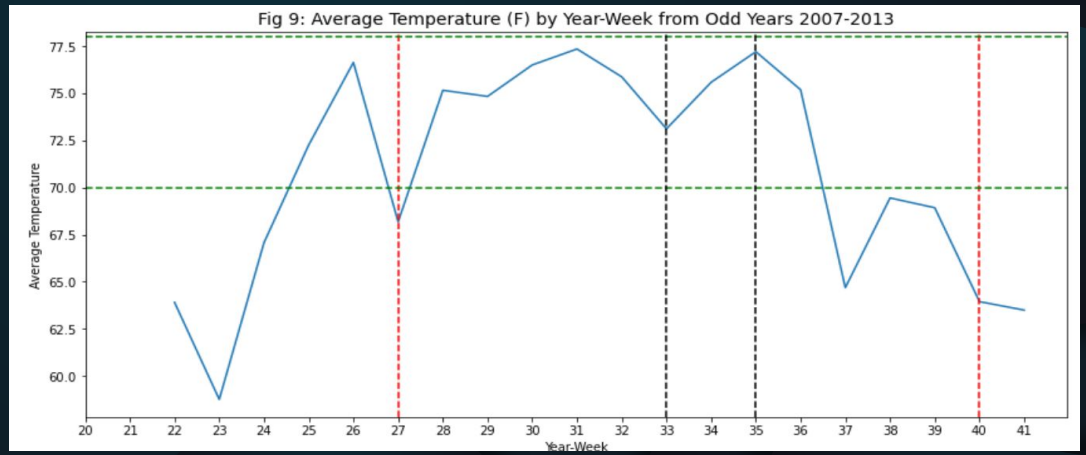
EDA

Seasonal Trends

- WNV Cases by Year-week
- Avg Temperature

Q: Are the WNV peaks facilitated by certain weather conditions?



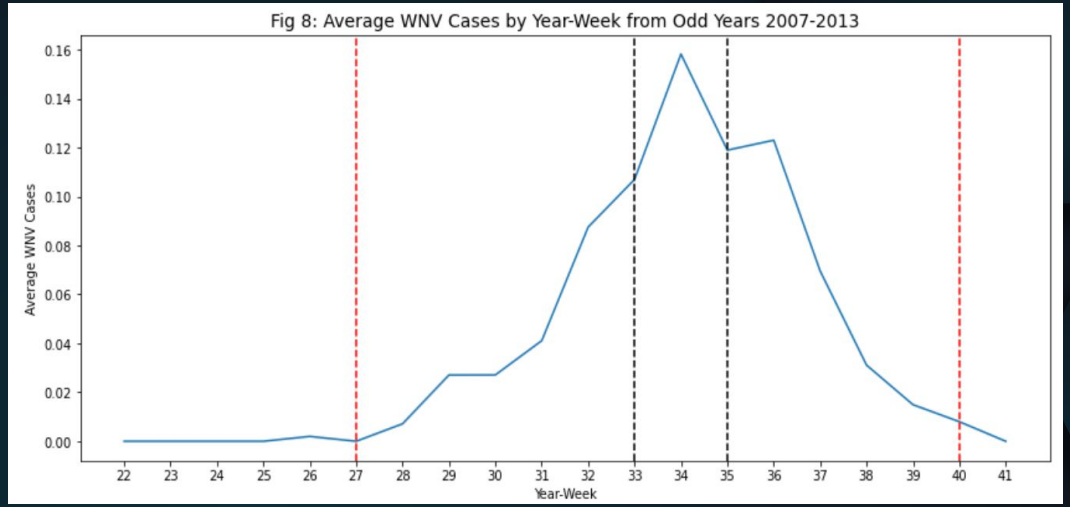


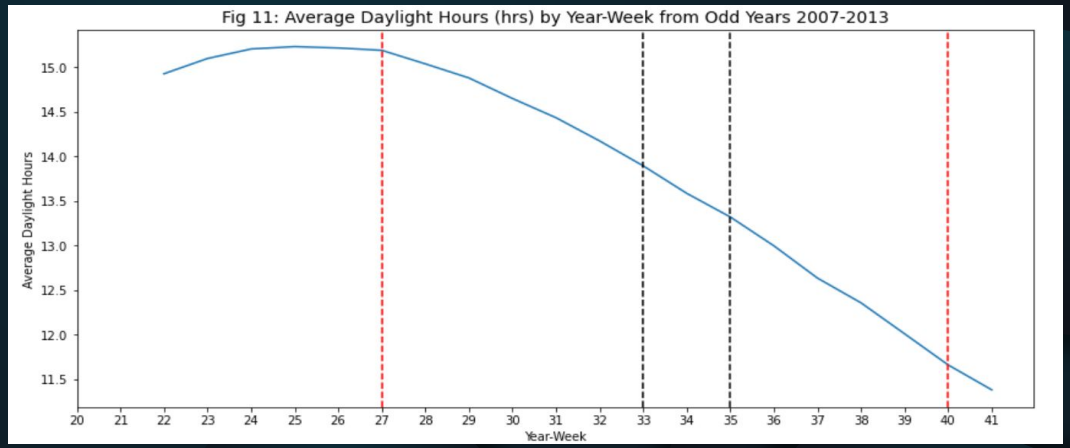
EDA

Seasonal Trends

- **WNV Cases by Year-week**
- **Daylight Hours**

Q: Are the WNV peaks facilitated by certain weather conditions?



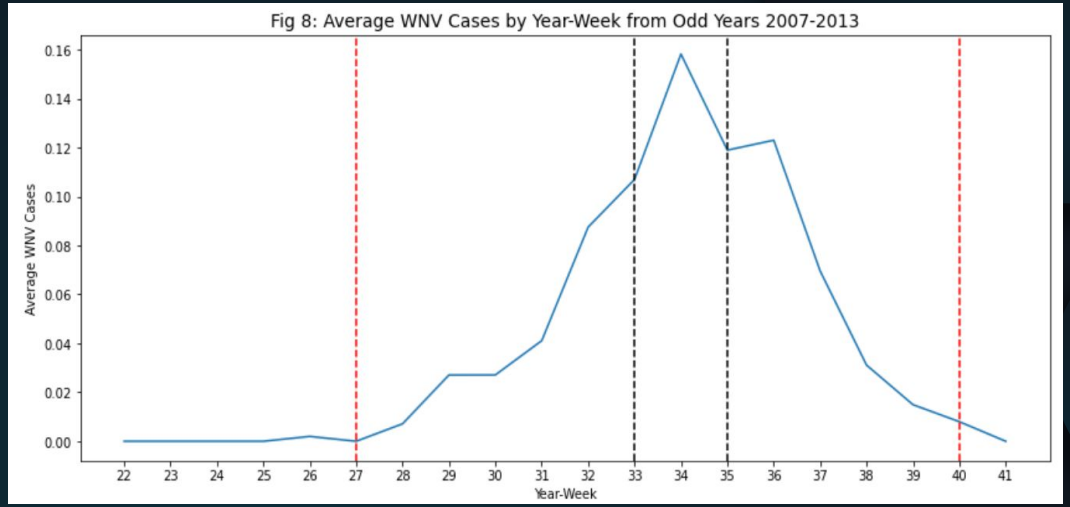


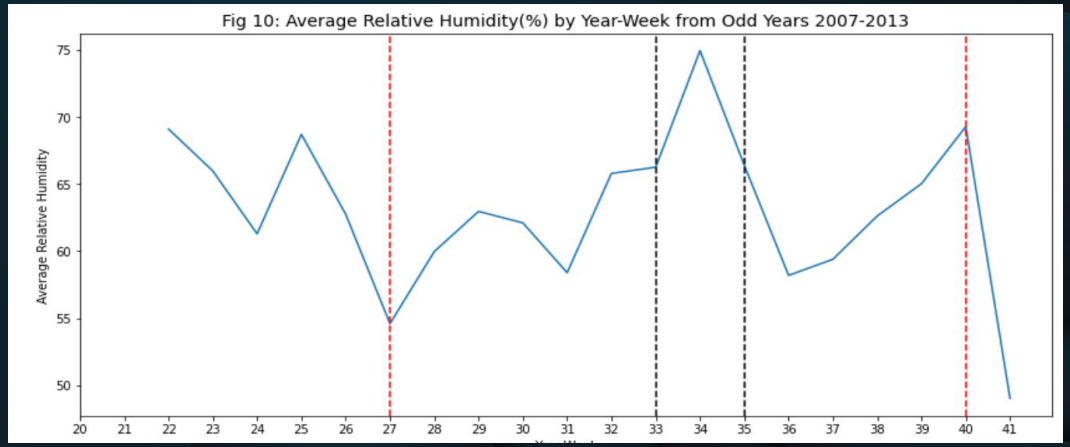
EDA

Seasonal Trends

- **WNV Cases by Year-week**
- **Relative Humidity**

Q: Are the WNV peaks facilitated by certain weather conditions?





03
Modelling

Modeling

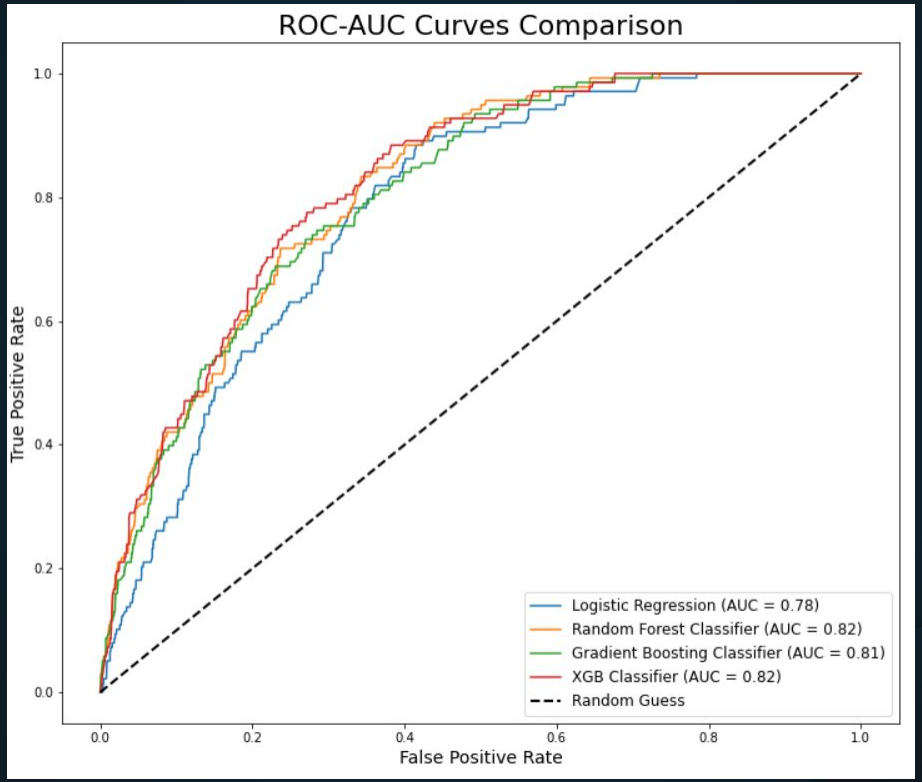
Baseline Model	BernoulliNB Model
Model 1	Logistic Regression Model
Model 2	Random Forest Classifier Model
Model 3	Gradient Boosting Classifier Model
Model 4	XGBoost Classifier Model

SMOTE will be used to oversample the data as there is an imbalance in the dependent variable of the presence of WNV.

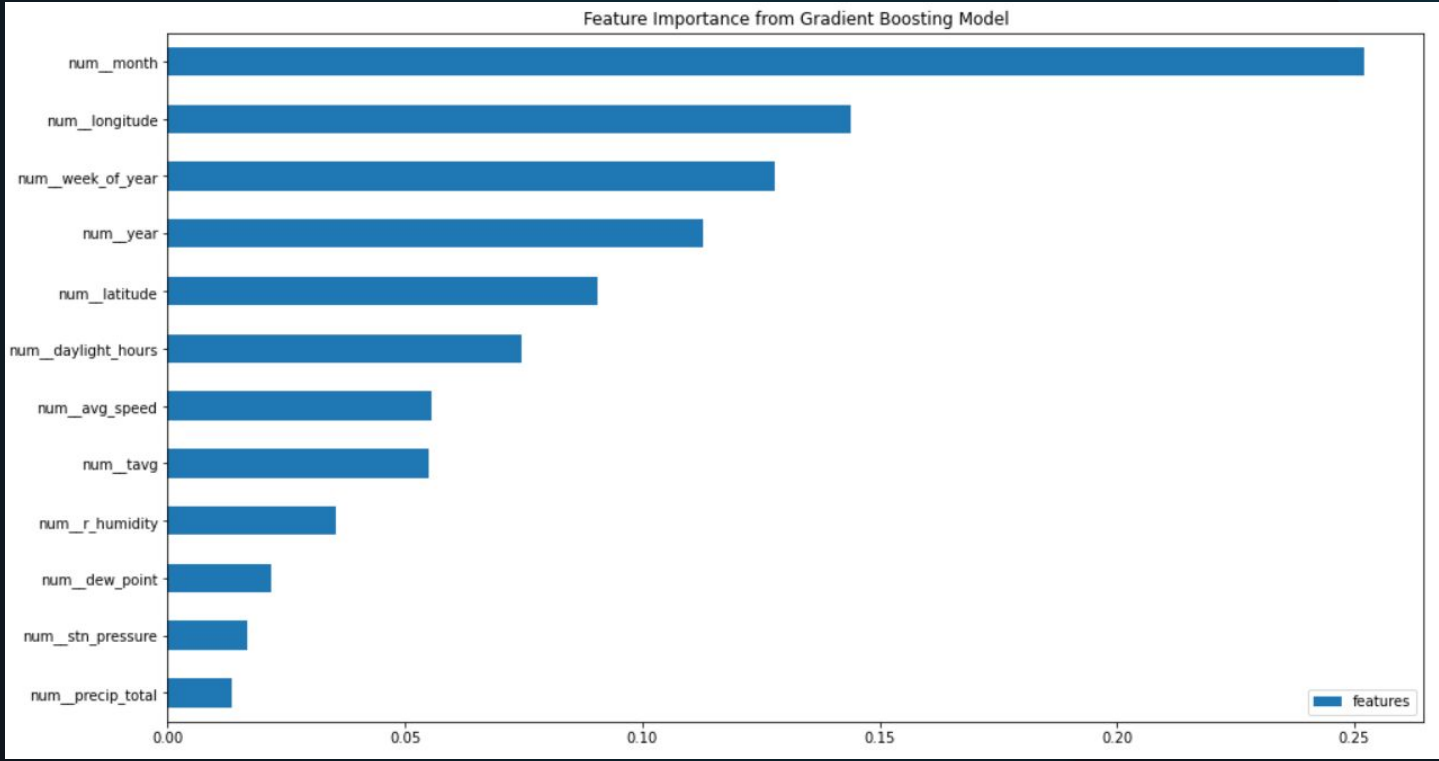
Modeling Results

	Models	Train AUC	Test AUC	F-score	Gridsearch AUC
Baseline Model	BernoulliNB Model	0.744	0.721	0.154	N/A
Model 1	Logistic Regression Model	0.773	0.782	0.197	0.769
Model 2	Random Forest Classifier Model	0.890	0.817	0.238	0.832
Model 3	Gradient Boosting Classifier Model	0.891	0.808	0.249	0.828
Model 4	XGBoost Classifier Model	0.873	0.823	0.240	0.833

Modeling Results



Best Model – Gradient Boosting Features



04
Cost-Benefit Analysis

Costs of Pesticide Spraying

$1.16/acre

Cost of Zenivex (Projected 2022 price*)

145,863 acres

Size of the City of Chicago

~$1.69 Million

Cost for fortnightly spraying during Summer

Weeks 20-40

** Assuming inflation of 3% p.a.*

Productivity Losses due to WNV

1 in 5 will develop fever, 1 in 150 will develop more severe illnesses.
Using 2012 data, 60 people contracted the WNV.

($733x1) + (1x $71,546)

= **~$80,000**

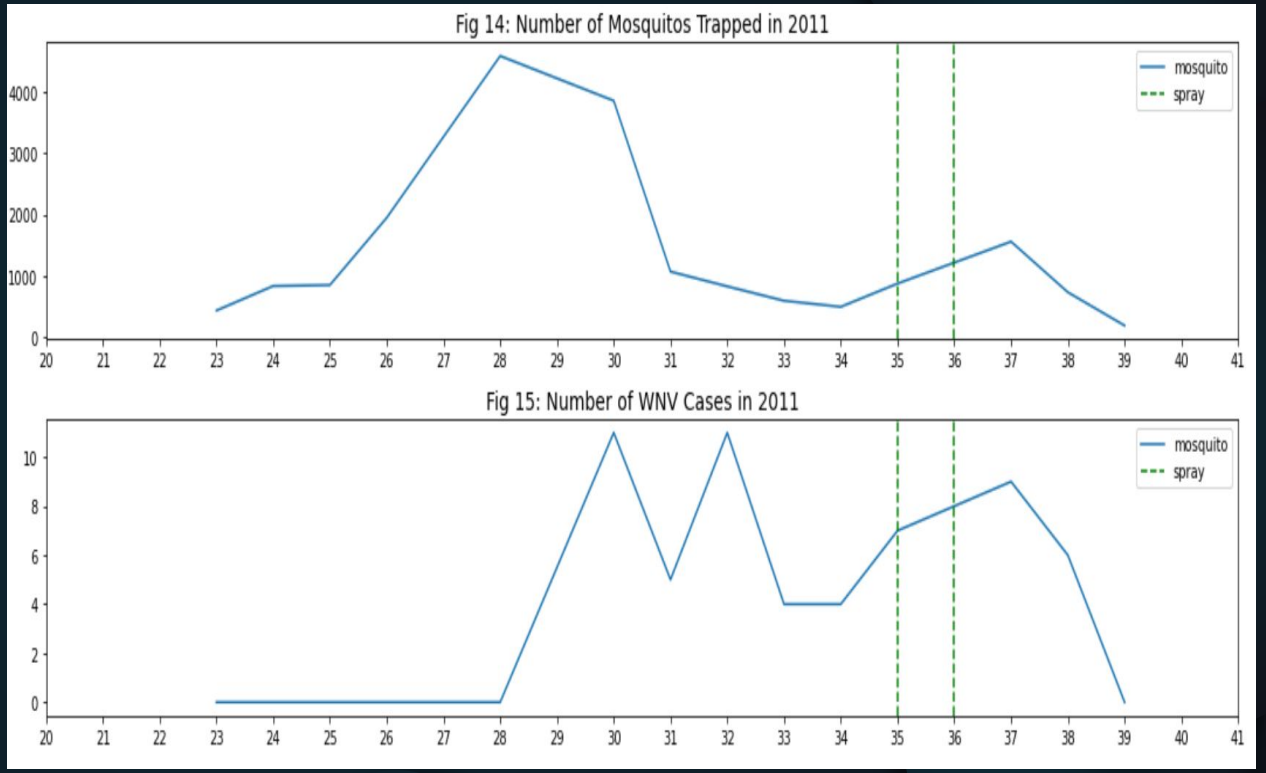
Recall that fortnightly spraying costs
= ~$1.01 million!

Disease	Lost Productivity in 2022 prices ($)
Fever	733
Meningitis	919
Encephalitis	71,546
Acute Flaccid Paralysis	16,607

EDA

2011 Spray Effectiveness

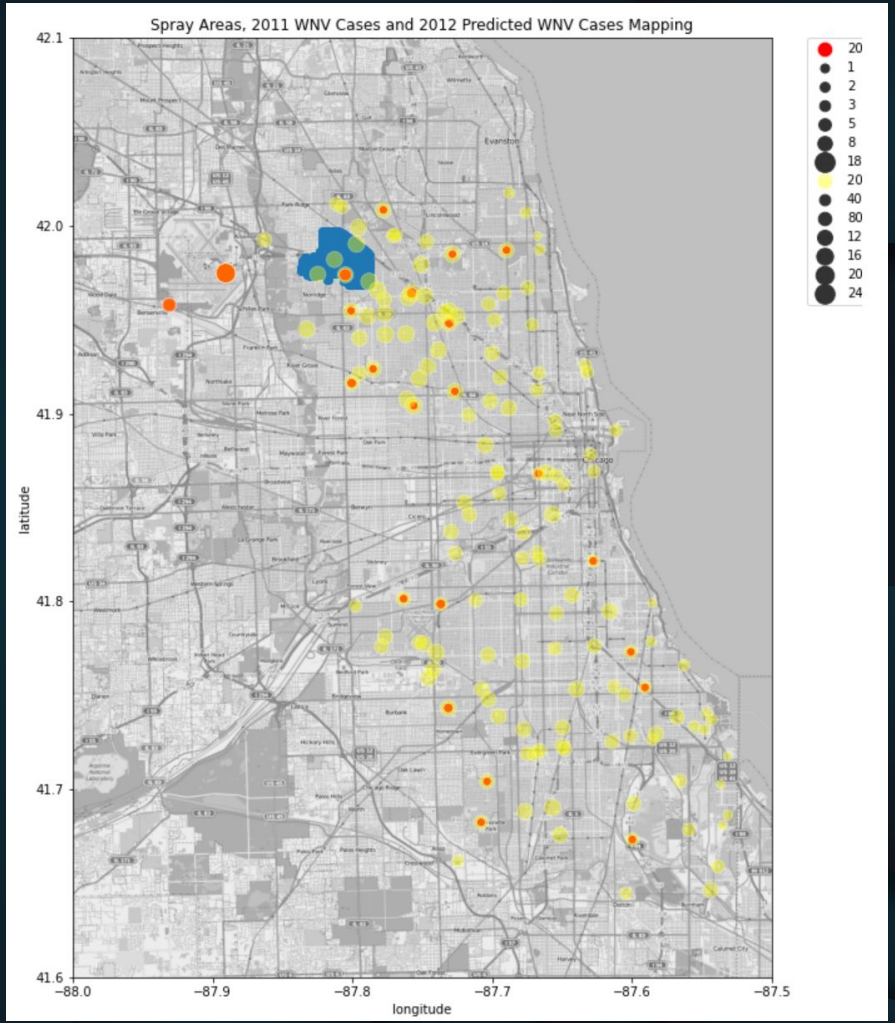
Q: Did spraying reduce the number of mosquitos & WNV cases?



EDA

2011 Spray Effectiveness on 2012 Cases

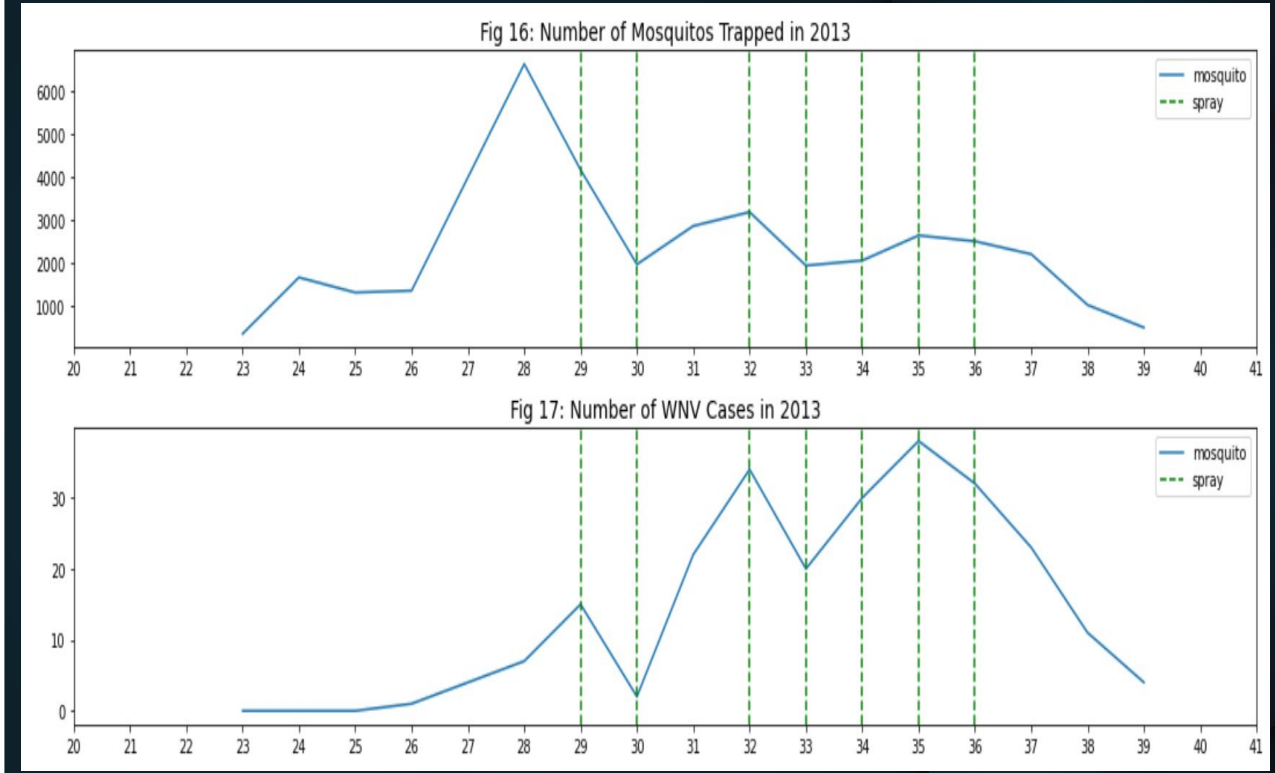
Q: Did spraying reduce the number of mosquitos & WNV cases?



EDA

2013 Spray Effectiveness

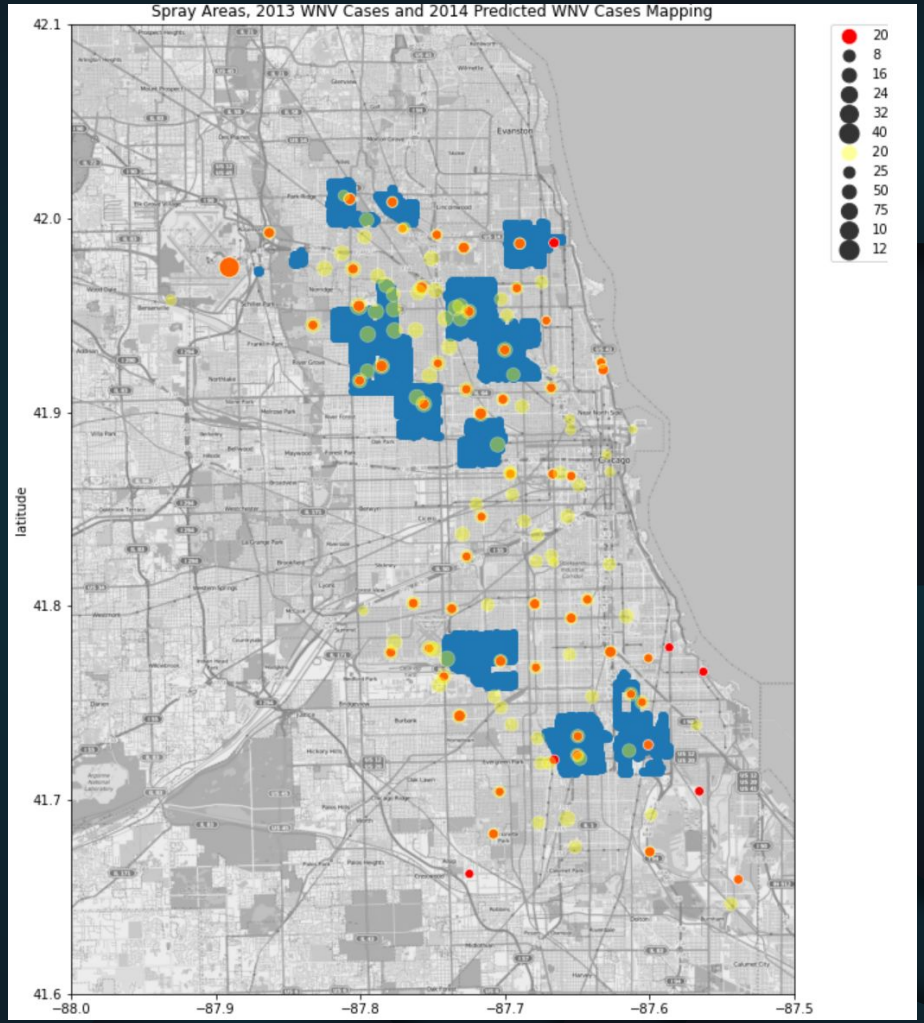
Q: Did spraying reduce the number of mosquitos & WNV cases?



EDA

2013 Spray Effectiveness on 2014 Cases

Q: Did spraying reduce the number of mosquitos & WNV cases?



05

Conclusion & Recommendations

Recommendations

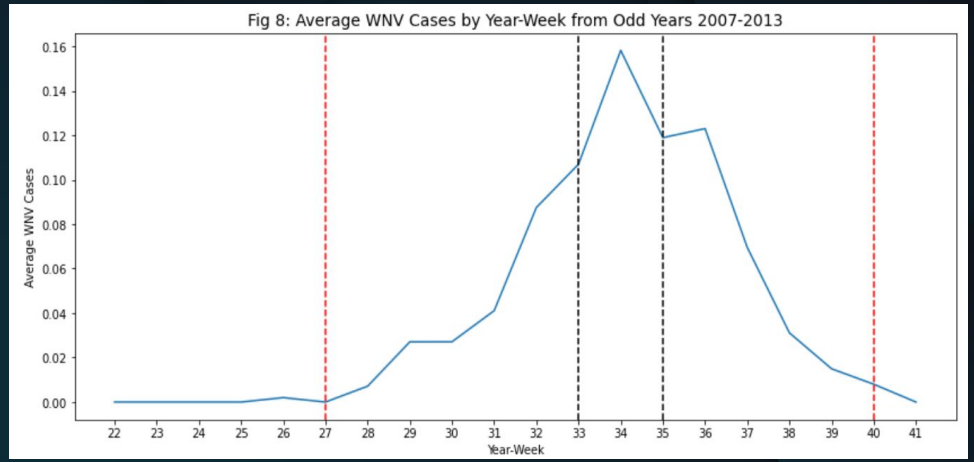
Spraying 2 weeks prior to Week 25 to reduce breeding opportunities

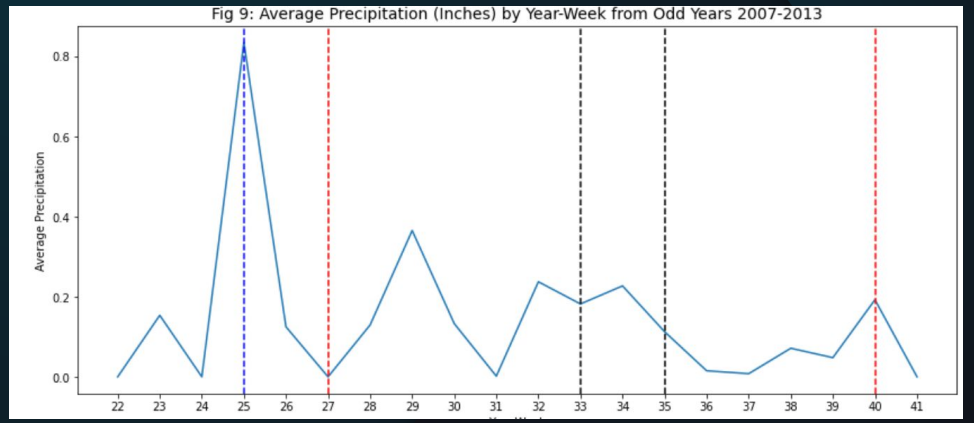
- Heavy rains at week 25 may be driving the mosquito population

- Fortnightly spraying can be done in weeks 24-35 instead of fortnightly spraying

Cost: ~$850,000

50% Cost Reduction!

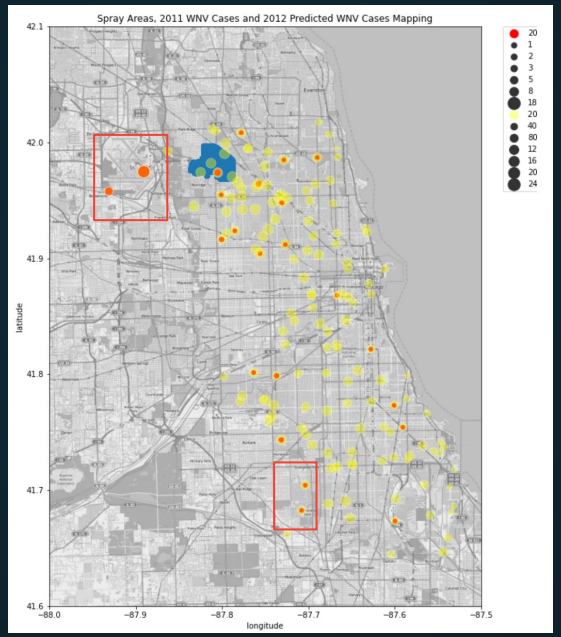


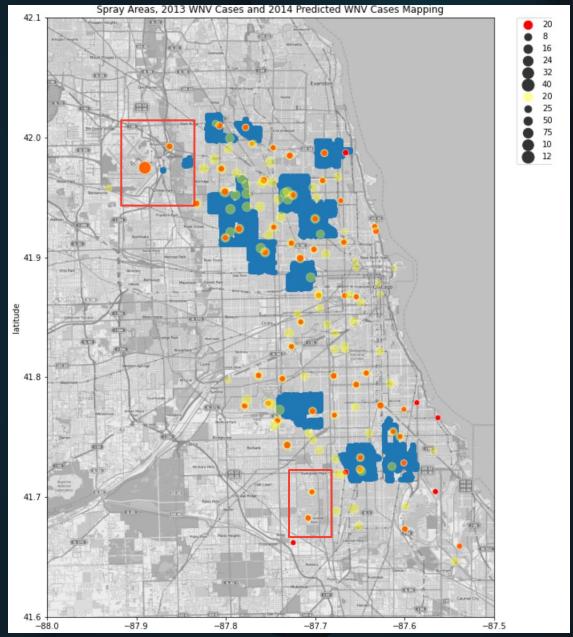


Recommendations

Target large clusters that consistently appear every year

- May be more cost effective to target larger clusters to give more ROI to drive down costs further to make it viable to invest in a spraying regime

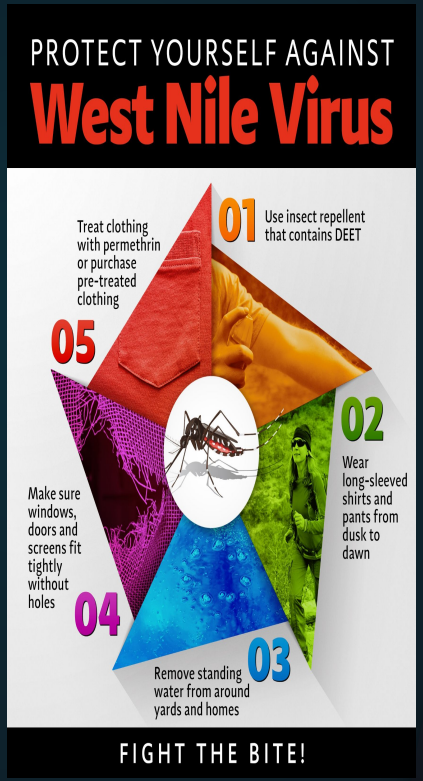




Recommendations

Citizen Outreach

- Conduct outreach and education efforts to inform citizens on the best practices to prevent mosquito breeding, periods of 'mosquito season' and the measures to prevent getting bitten



CONCLUSION

Recommended Approach

- Conduct fortnightly-spraying from weeks 24-35 (prior to the potential mass hatching from the week 25 rains till the traditional peak of WNV season at week 35)
- Conduct targeted spraying on large, consistent clusters from historical data.
- Prevention is the best medicine: Citizen education to deny mosquitoes of breeding grounds and feeding opportunities

Above approach enables the city to reap at least 50% savings on pesticide costs.

THANKS!


slidesgo