

West Nile Virus Prediction

Presented by - Cathy, Miao Er, Janet



Image Source: <https://www.cdc.gov/dotw/westnilevirus/index.html>



FLY THE FLAG

IDYLLIC Weather, Pages 2, 32

WEDNESDAY, AUGUST 7, 2002

Sports Final

WEST NILE VIRUS HITS ILLINOIS



West Nile virus is transmitted to humans through mosquito bites. Symptoms include fever, muscle aches and a rash.



This test strip of mosquitoes collected in the Chicago area came back positive for West Nile virus.

22-year-old woman is first to contract disease in state

By Dave McKenney
and Gary Wessly
First Syndrome

SPRINGFIELD—Perhaps it came during an evening walk from her car to her house or in the parking lot of a grocery or on her back patio. No one knows for sure when a mosquito unleashed its sting on an unwitting 22-year-old student.

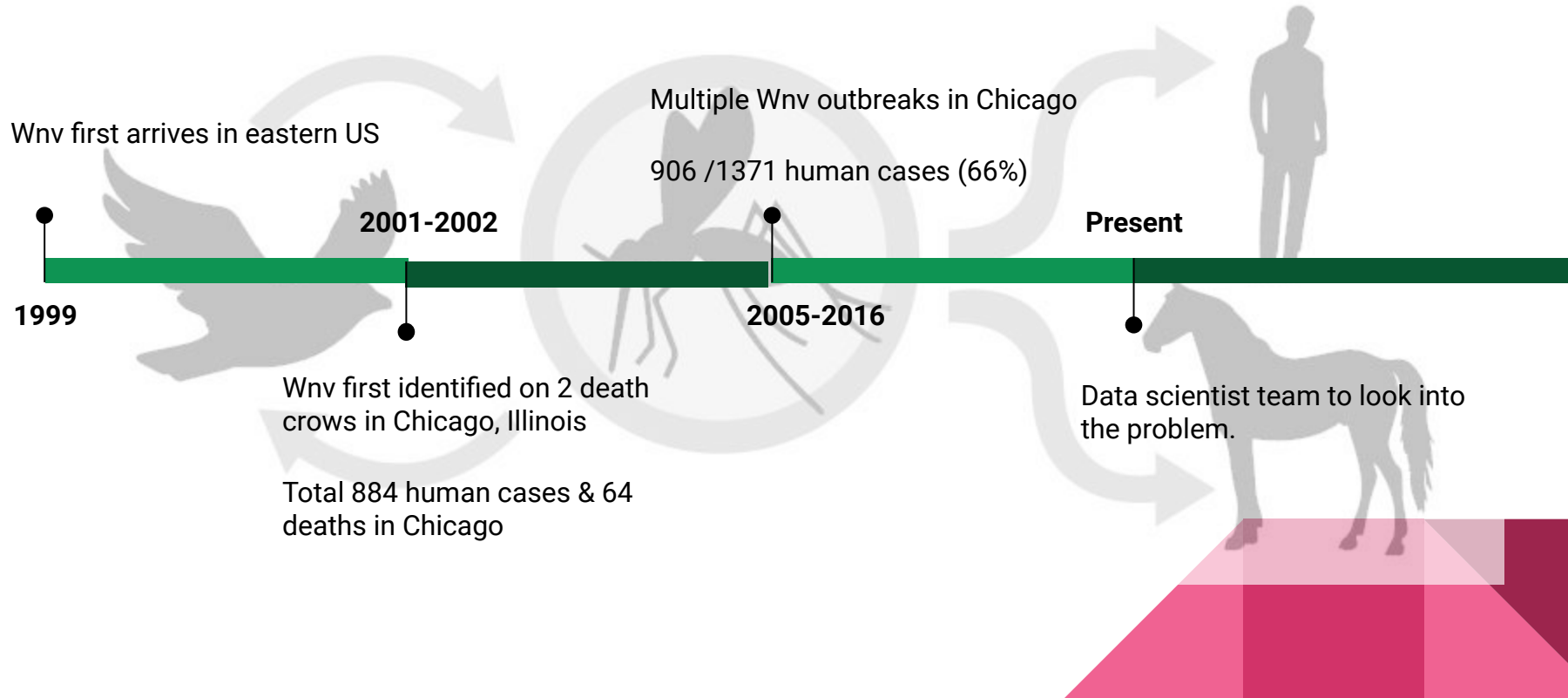
What is known is that nearly two weeks ago, the young woman came down with fever, aching muscles and a slight rash and learned through tests completed Tuesday that she had become the first

Turn to Next Page

Illinois' first human case of West Nile disease is confirmed on August 6, 2002.

Source <http://www.idph.state.il.us/timeline/2000wnvhuman2.jpg>

Timeline of West Nile Virus



25,849 cases*

2,456 deaths

of Wnv in the US 1999-2020

Source : <https://www.cdc.gov/westnile/statsmaps/cumMapsData.html>

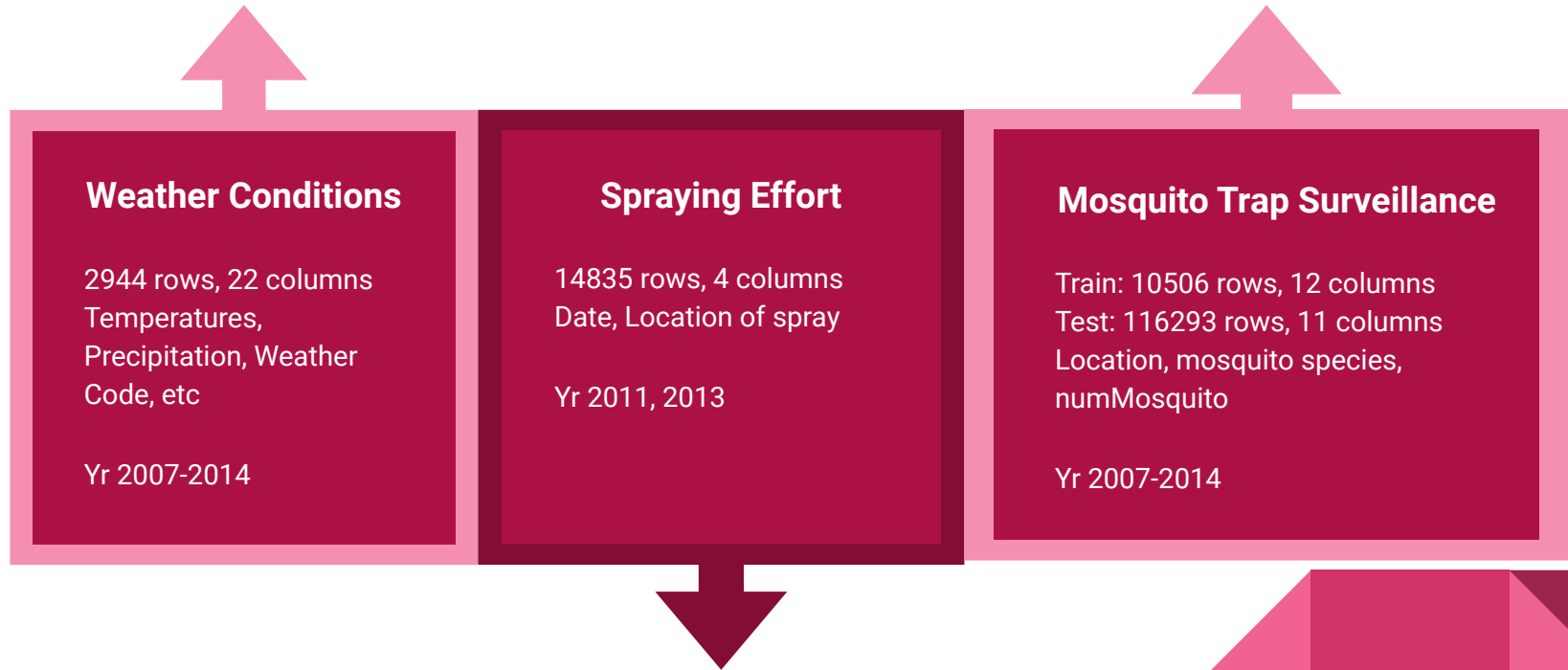
Problem Statement

Being hired by the division of Societal Cures In Epidemiology and New Creative Engineering (DATA-SCIENCE), we're tasked to build a model that predicts outbreaks of the WNV.

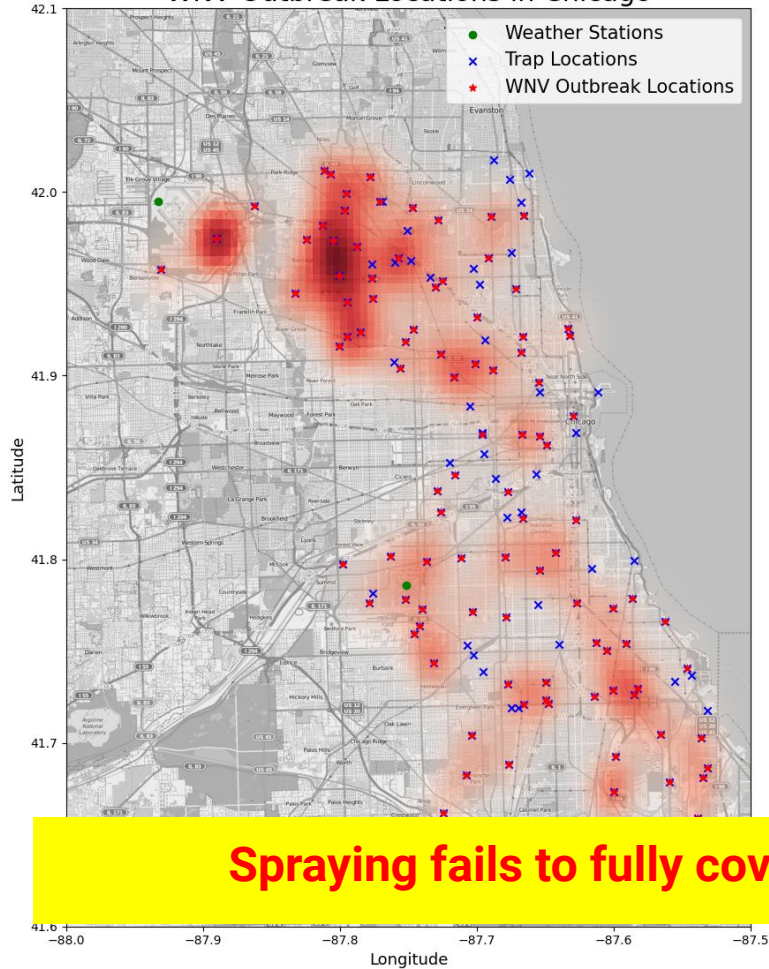
1. Helps the City of Chicago and the Chicago Department of Public Health (CDPH) more efficiently and effectively allocate resources toward preventing the transmission of this potentially deadly virus.
2. Aims to determine the best strategy for controlling the spread of the WNV.



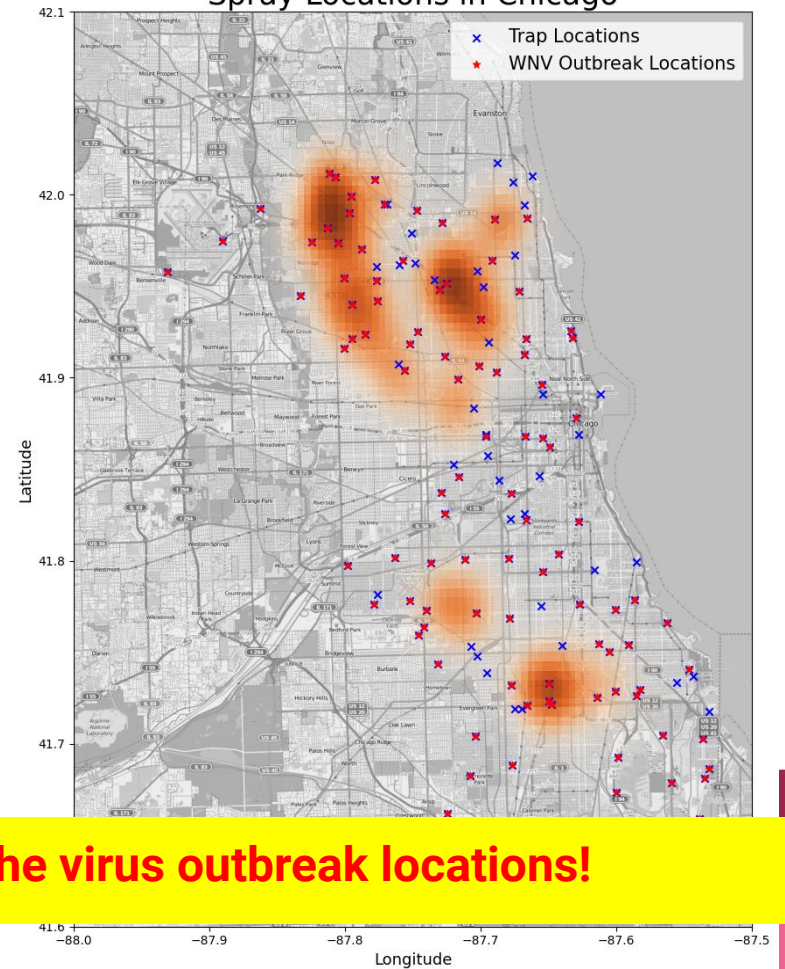
Datasets & Dictionary



WNV Outbreak Locations in Chicago



Spray Locations in Chicago



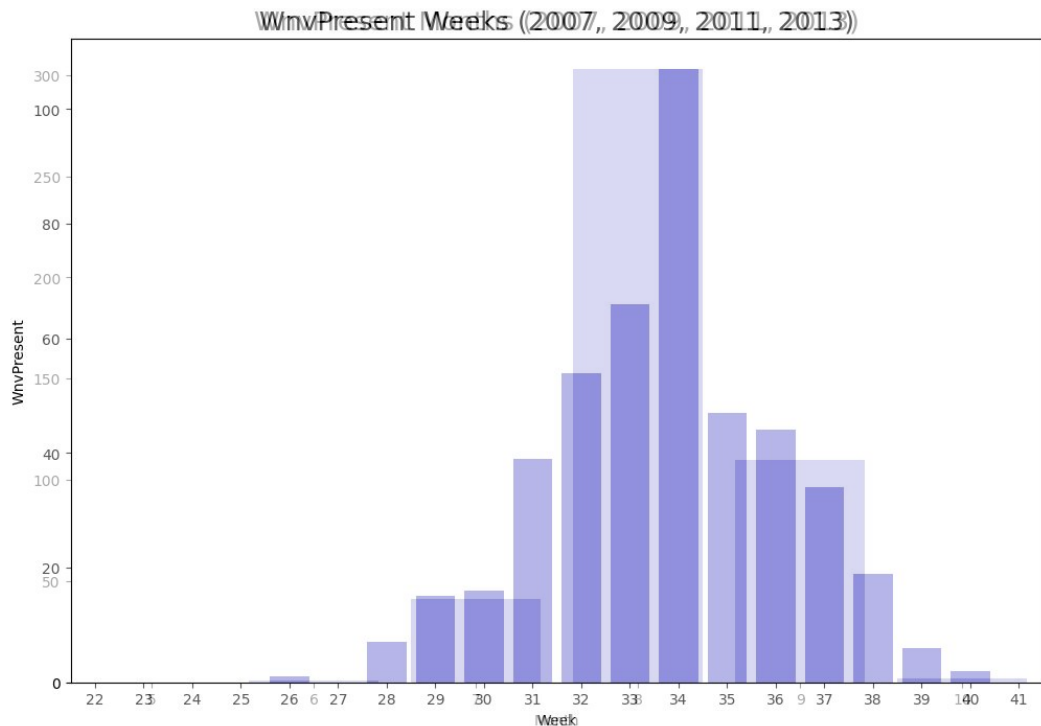
Spraying fails to fully cover the virus outbreak locations!

Data Cleaning

1. Dropping duplicate rows,
2. Imputing missing and zero values,
3. Splitting strings and replace with correct format,
4. Converting to right data types,
5. Dropping columns with high missing values,
6. Creating more interpretable features,
7. etc



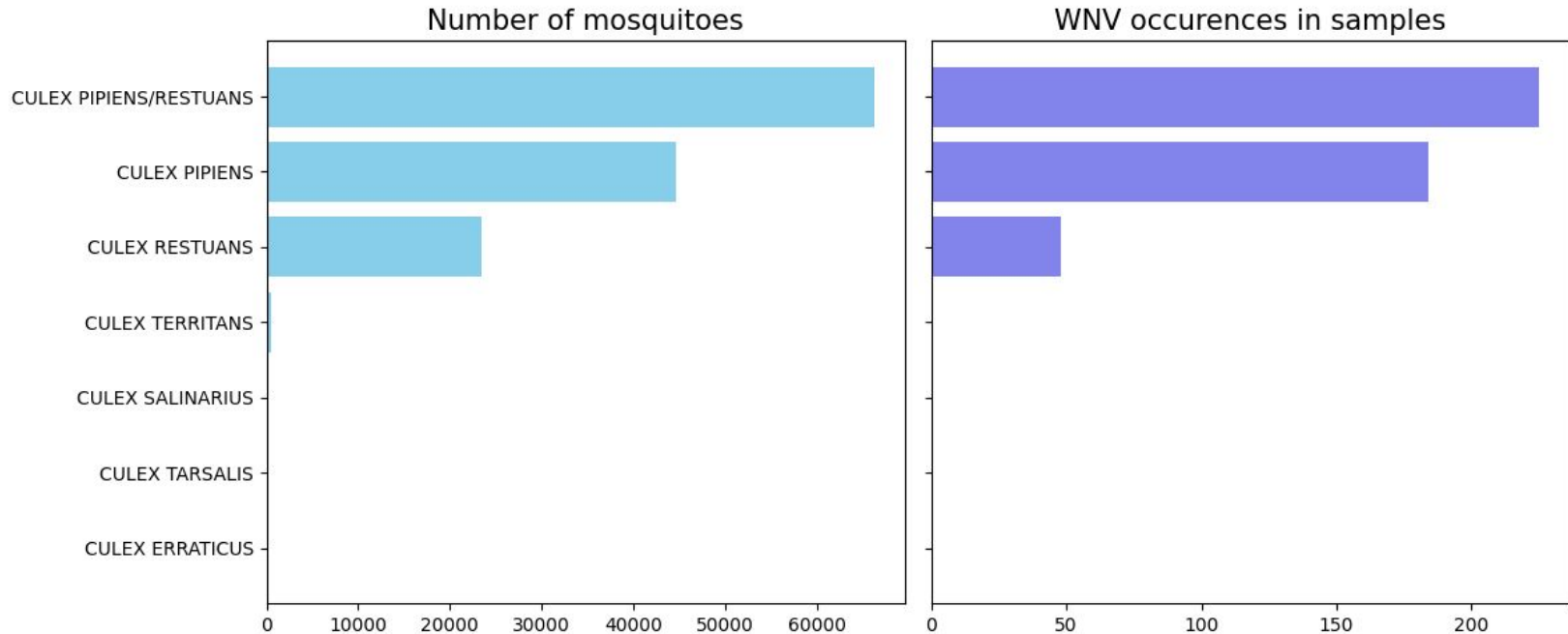
August/wk 34 seems to be the month/week with highest Wnv occurrences



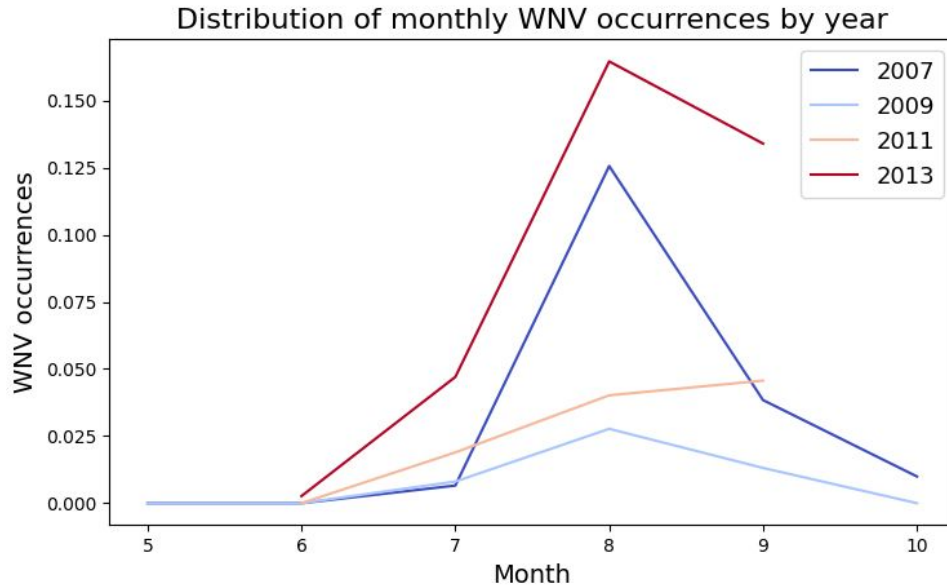
- Wnv started from Jul, peak in Aug and reduced in Sept. This is the summer season in Chicago.
- Equivalent period from Week 26 - 40.
- Summer weather is conducive for mosquito breeding.

Only 2 species sampled are Wnv carrying vectors

=> **Culex Pipiens** | **Culex Restuans**

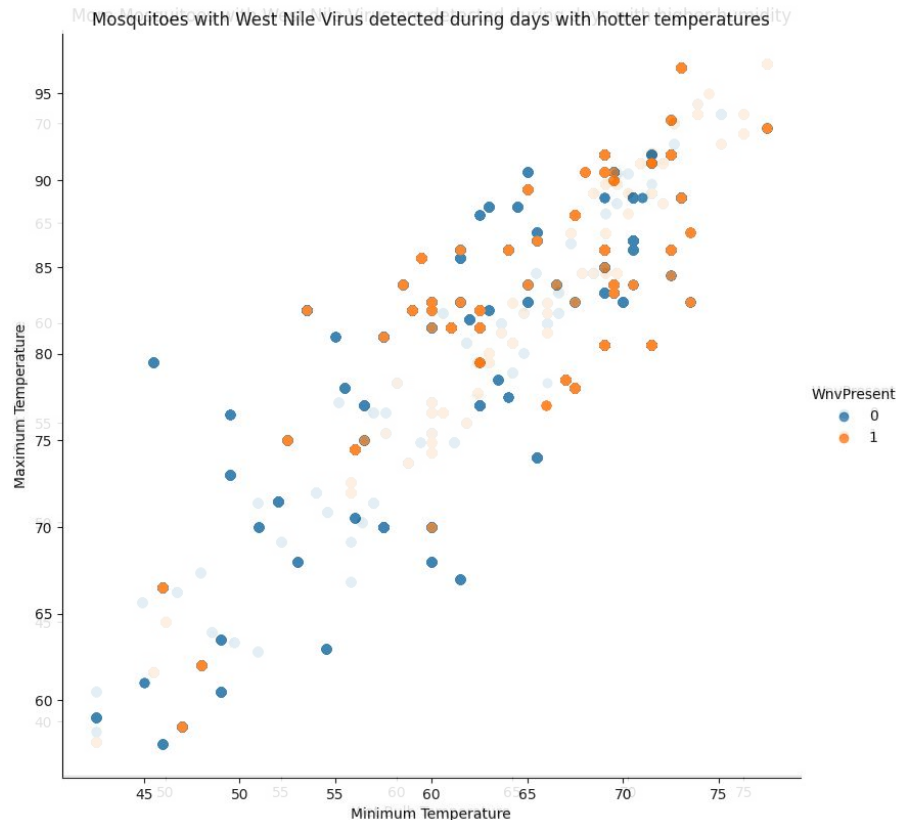


What contributes to higher Wnv outbreak in 2013?



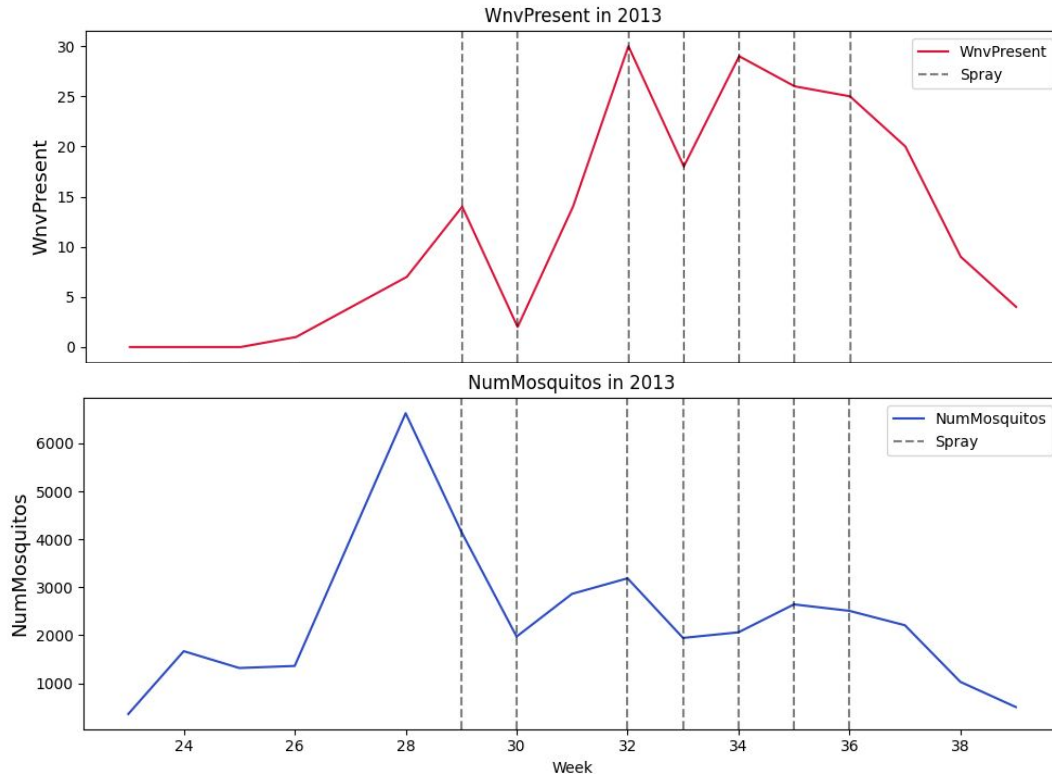
- 2013 stands out with highest Wnv outbreaks, probably due to heat wave in Chicago?
- 2009 is the year with least Wnv outbreaks, probably due to colder temperature on year record?
- Positive correlation between Wnv and temperature.

WnvPresent linearly increases with Humidity & Temp



Higher humidity and hotter temperature lead to more Wnv outbreaks.

Spraying has little effect on reducing Wnv outbreak



Spraying helps to reduce NumMosquitos but not Wnv outbreak.

Possible reasons:

- Insufficient coverage of spray
- Time lag factor in infection vs spray timing

Key Questions and Understanding

What are the contributing factors leading to Wnv spread?

Which mosquito species carry Wnv?

How can we reduce Wnv cases?

Where in Chicago is Wnv most concentrated in?

How Wnv spreads over time ?

What are the strategies for controlling Wnv spread?

Feature Engineering

Weather -- Sunhour Feature

To capture the duration of the day, or number of hours when the Sun is out:

$\text{SunHours} = \text{Sunset} - \text{Sunrise}$

Weather -- Codesum Feature

Replace all empty rows with '0' due to uneventful weather (quiet weather pattern)

Create dummy variables - Rain, Snow, Windy and Misty - as follows:

Rain = 1 if CodeSum contains ['TS', 'RA', 'DZ', 'SH'] else = 0

Snow = 1 if CodeSum contains ['SN', 'SG', 'GS', 'PL', 'IC', 'DR', 'BC', 'GR'] else = 0

Windy = 1 if CodeSum contains ['SQ', 'DS', 'SS', 'PO', 'BL'] else = 0

Misty = 1 if CodeSum contains ['BR', 'HZ', 'FG', 'FU'] else = 0



Feature Engineering

Weather -- Humidity Feature

High humidity conditions tend to favour mosquito activity

Create Humidity column using DewPoint and Tavg (Average temperature)

[Reference](<https://www.calcunation.com/calculator/humidity-calculator.php>)

```
weather['Humidity'] = np.round(100-25/9*(weather['Tavg']-weather['DewPoint']),1)
```

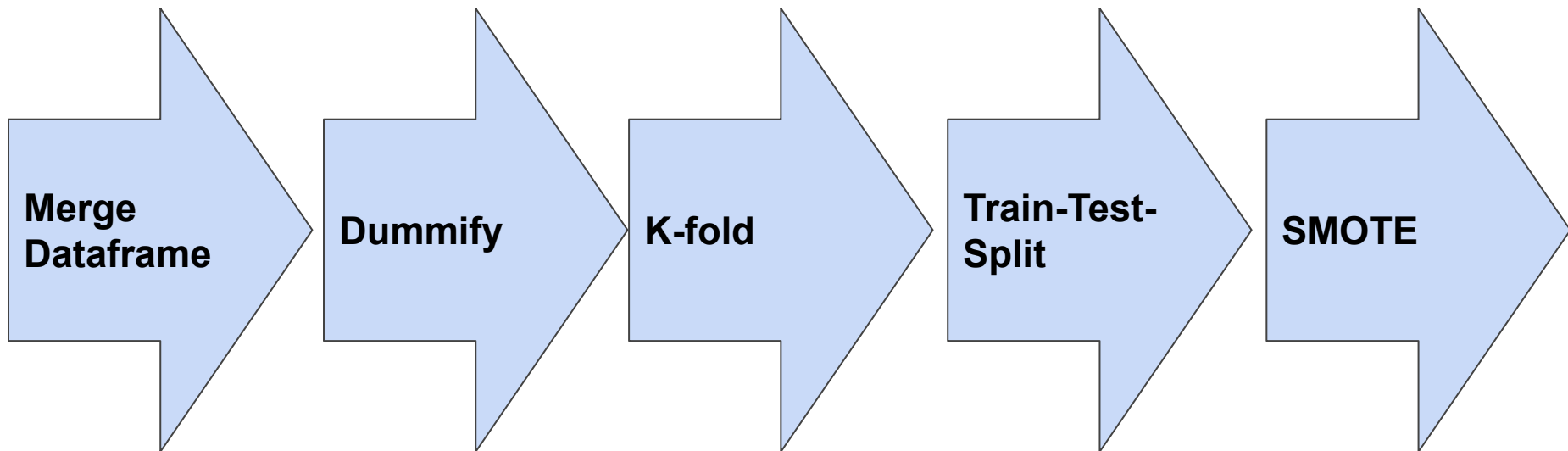
Train – Mosquito Species

Only species with 'Restuans' or 'Pipiens' - had tested positive for the virus

Dummify 'Restuans' or 'Pipiens' - add weightage



Preprocessing



Modelling

Build pipeline for StandardScaler and models with parameters

- Logistic Regression
- K Nearest Neighbors Classifier
- Random Forest
- Extra Trees
- Gradient Boosting

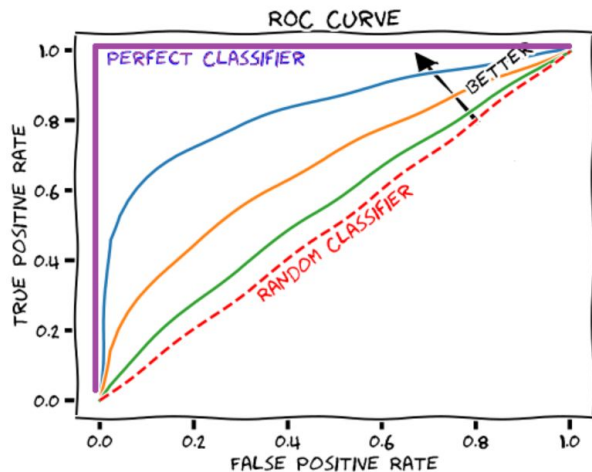
Use Gridsearch CV to look for the best parameter



Modelling Metrics

Receiver Operating Characteristic (ROC) Area Under Curve (AUC)

Plot the True Positive Rate (**sensitivity**) vs. False Positive Rate (**1 - specificity**)



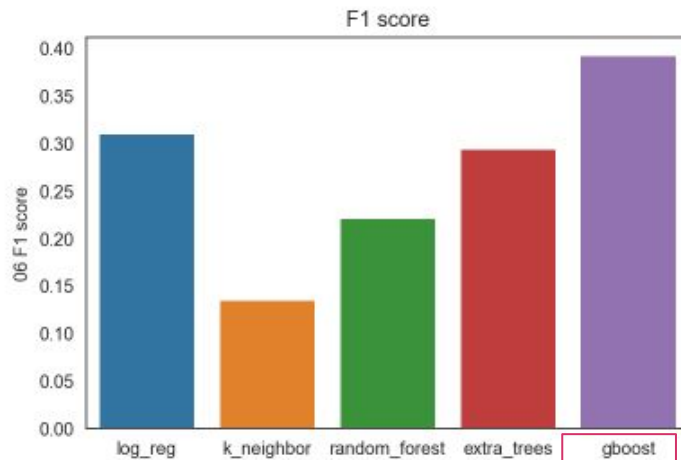
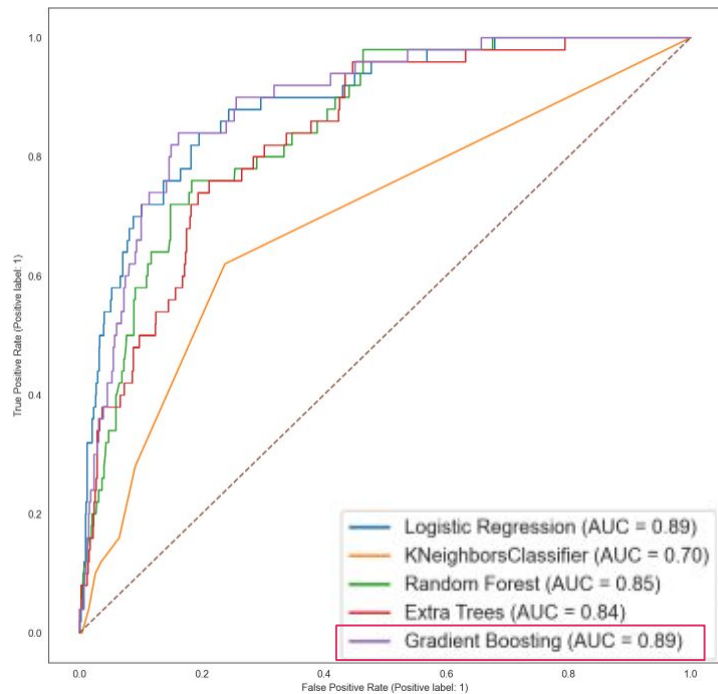
F1 Score

The model's balanced ability to both capture positive cases (recall) and be accurate with the cases it does capture (precision)

$$\text{F1 score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

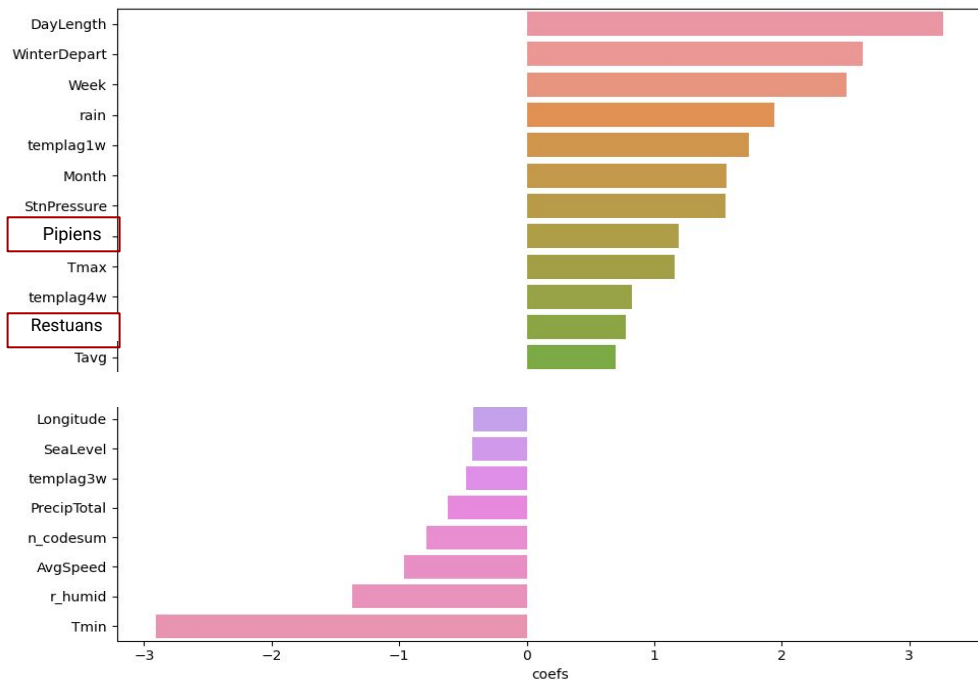
Model Evaluation

- Gradient Boosting has the highest AUC score, and also highest F1 score.



Results Interpretation

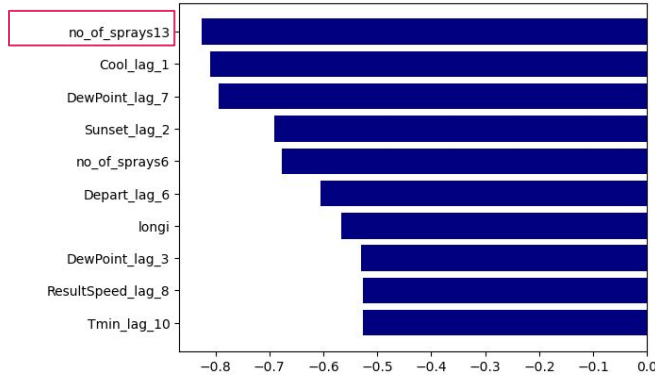
Best Estimator Coefficient



- Weather conditions
- Two species that carry WNV

Cost Benefit Analysis

Top suppression factors for WNV



- For cases when sprays were done 13 days ago, probability of being WNV positive decreases to a 0.25% the probability without sprays 13 days ago.
- Sprays effectively eradicates 99.75% WNV carriers for a 2-week period.



From research we also know ...

- Adulticide sprays do not stop immature stage mosquitoes from developing and breeding.
- It takes 4 days to a month for mosquitoes to develop from egg to adult Culex mosquitoes.



To continually exterminate the WNV carriers, sprays need to be at 2 week intervals.

Cost Benefit Analysis

Costs

- Total land area 145,500 acres in Chicago
- Spray in use by the Chicago Department of Public Health (CDPH): **Zenivex E4**
- Spray cost **\$0.92 per acre**



- Three levels of spray effort (Jun to Nov ~26 weeks)

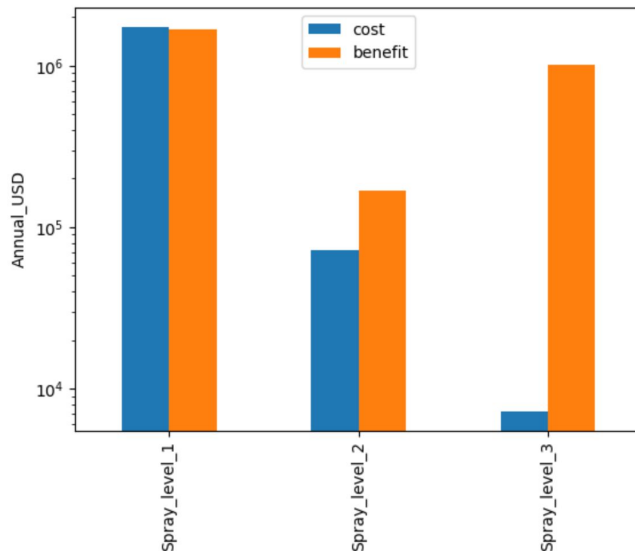
Level 1: Whole city spray at 2 weeks interval

Level 2: Spray for **more populated communities** (population density over 10k/km2) at 2 weeks interval (adds to 6,055 acres)

Level 3: Spray with info of **WNV surveillance** system and our **prediction** system (adds to 606 acres, 25 cases annually /13 intervals * 35 acres)

Benefits:

- Average costs from fever and hospitalization is \$7,500
- Annual hospitalization cases 225
- **Recall 0.6**
- \Rightarrow **Direct healthcare costs avoided = Annual hospitalization cost x Recall ~ \$1,000,000**



Conclusion and Recommendations

- Developed a model that successfully predicts WNV with a AUC score 0.89.
- Identified top weather conditions that contribute WNV incubation.
- Confirmed the effectiveness of large-scale pesticide sprays on extermination of WNV.
- Performed cost-benefit analysis on three different levels of spray effort and evaluated the economic value of our prediction model.
- Future work: include longer period of spray data in the modeling; include geospatial info in spray data; reduce percentage of class 0

