Project 4: West Nile Virus Detection

A presentation to Biostatisticians, Epidemiologists, and Decision Makers of Chicago

Overview

Overview

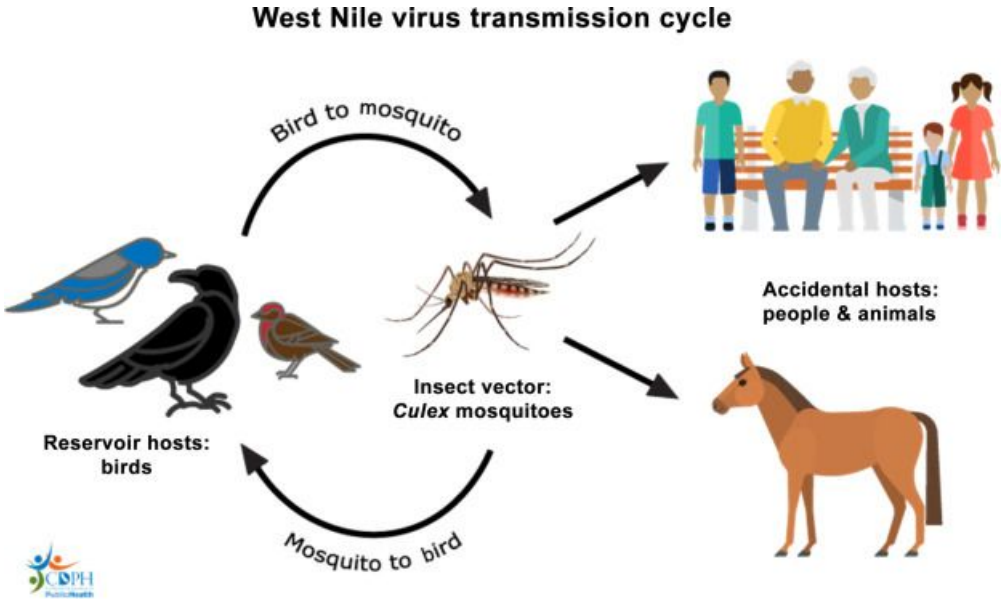
- Context
- Data Exploration: EDA on the train and weather data sets
- Date Modelling: Structure of Train and Test sets
- Date Modelling: Models tested and results
- Cost Benefit Analysis
- Recommendations

Context

The West Nile Virus (WNV) is a dangerous disease that can cause severe illness and even death in humans.

The recent epidemic of WNV in the Chicago has raised concerns about the public health and safety of the citizens.

The Department of Public Health has set up a surveillance and control system to collect data over time to learn from the mosquito population.



Problem Statement

The objective of this project is to create a model that can predict the presence of WNV in mosquitoes in the city of Chicago.

Metric: AUC ROC

Data Sets

Weather data (1/5/2007 - 31/10/2014)

Spray effort data (2011 and 2013)

Train data (2007,2009,2011,2013)

This model will help the city officials to decide where to spray pesticides,
which will not only save lives and
also reduce the cost of pesticide deployment.

The model will be based on the data collected from the surveillance and control system set up by the Department of Public Health.

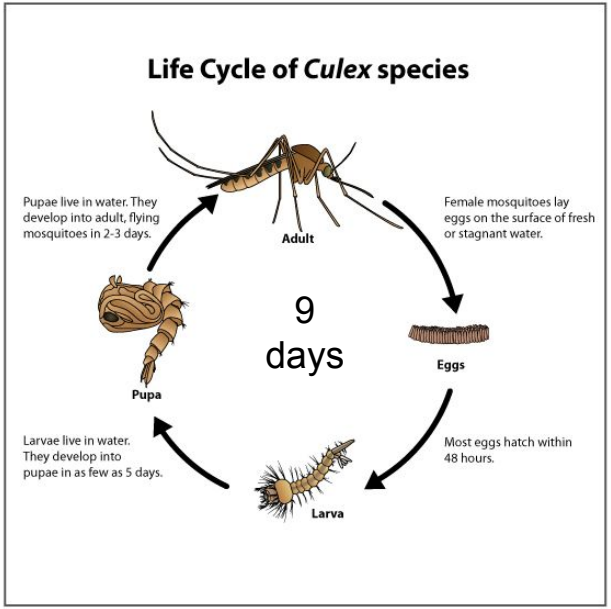
Impact on humans

The majority of individuals (about 80%) who are bitten by an infected mosquito develop no symptoms.

Roughly 20% of people develop the flu-like symptoms of West Nile fever including:

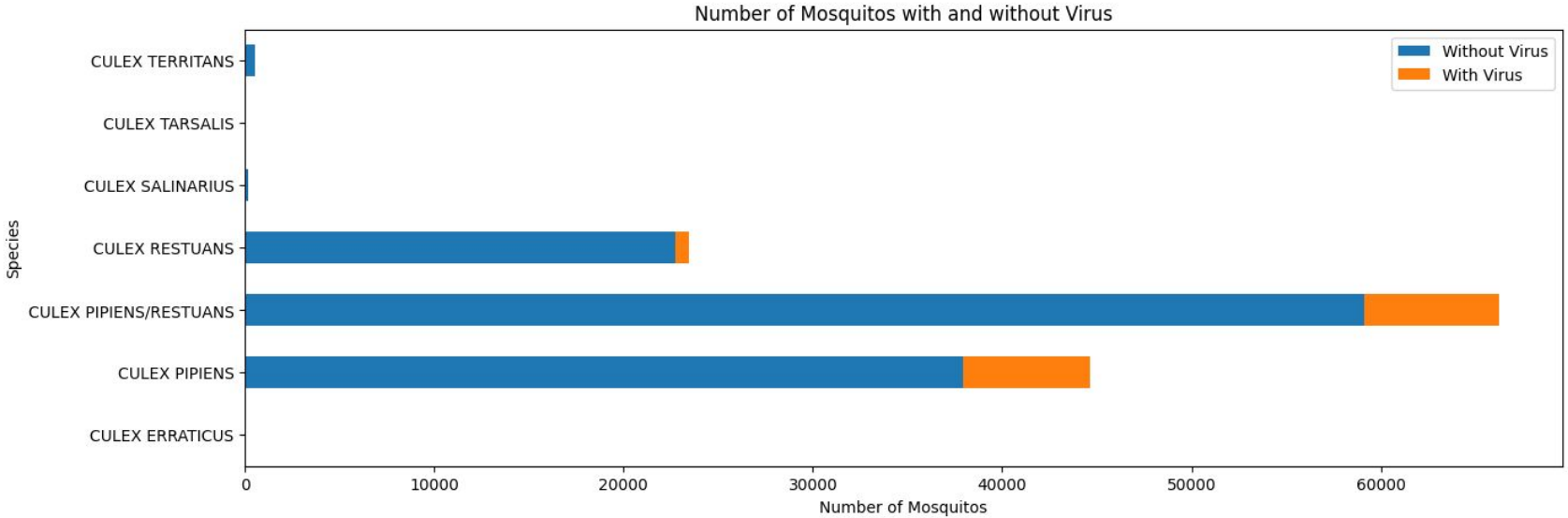
- Fever
- Headache
- Fatigue
- Nausea
- Vomiting
- Body aches
- Occasional skin rash
- Eye pain or swollen lymph glands

1% of infected gets neural invasive disease → Can lead to death

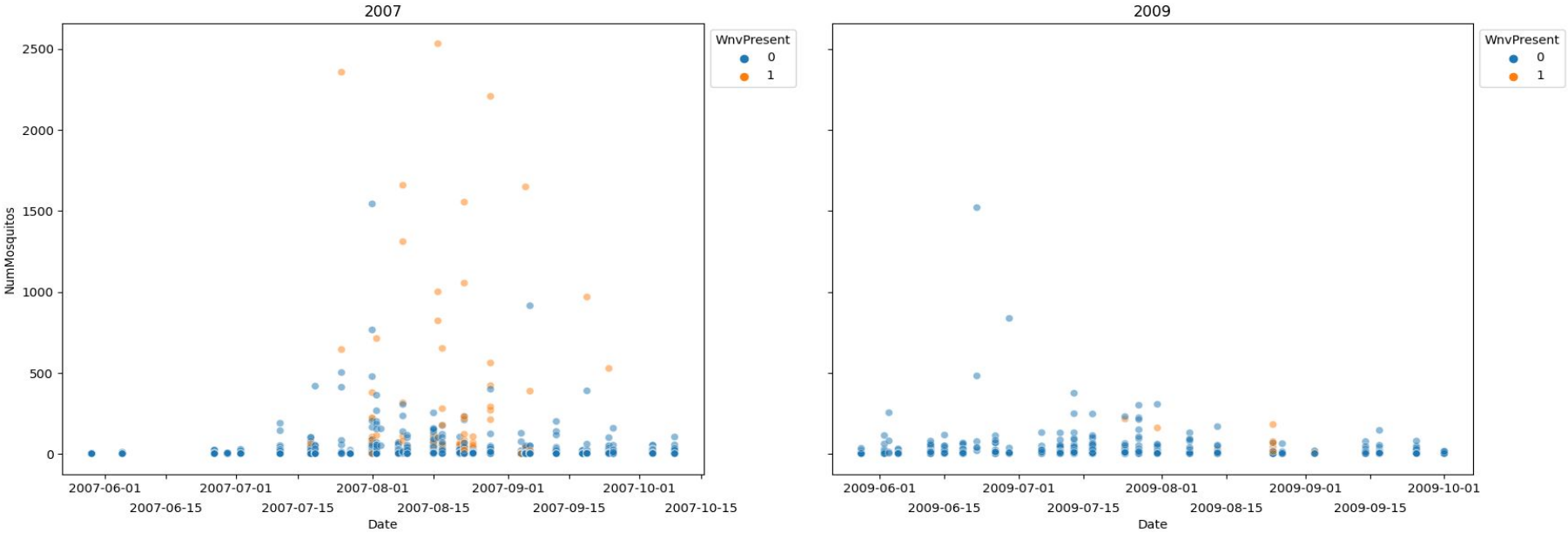


EDA

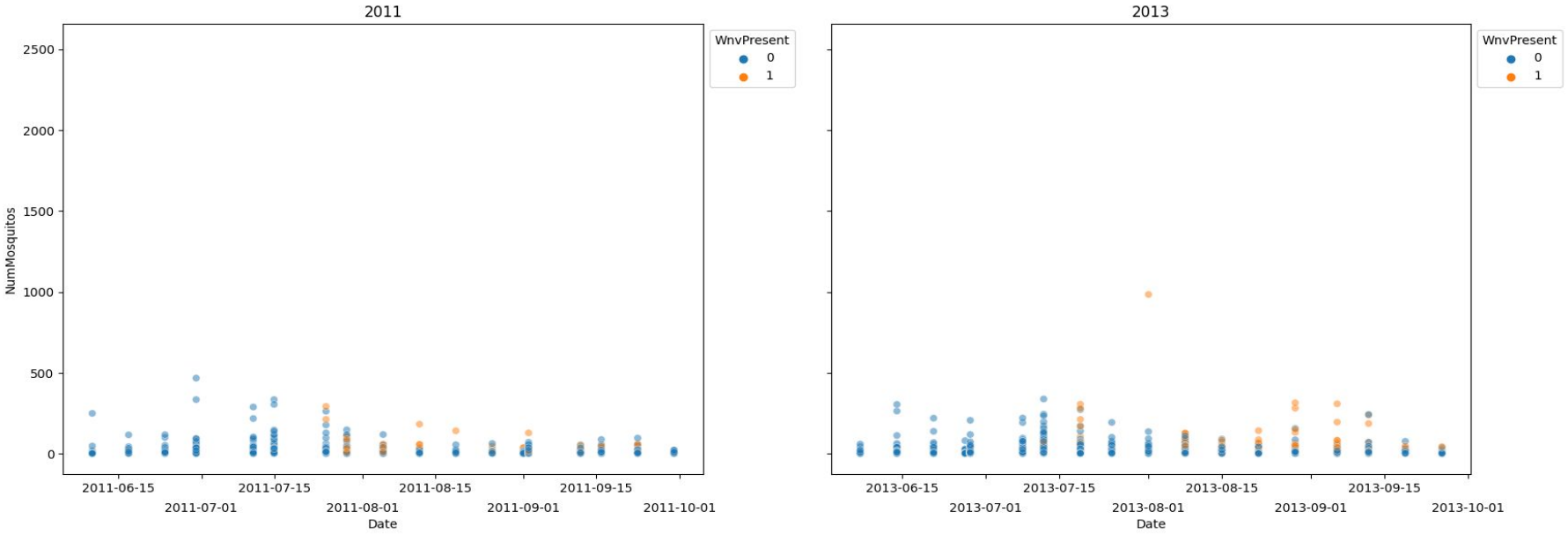
Distribution of Mosquito Species in Chicago City



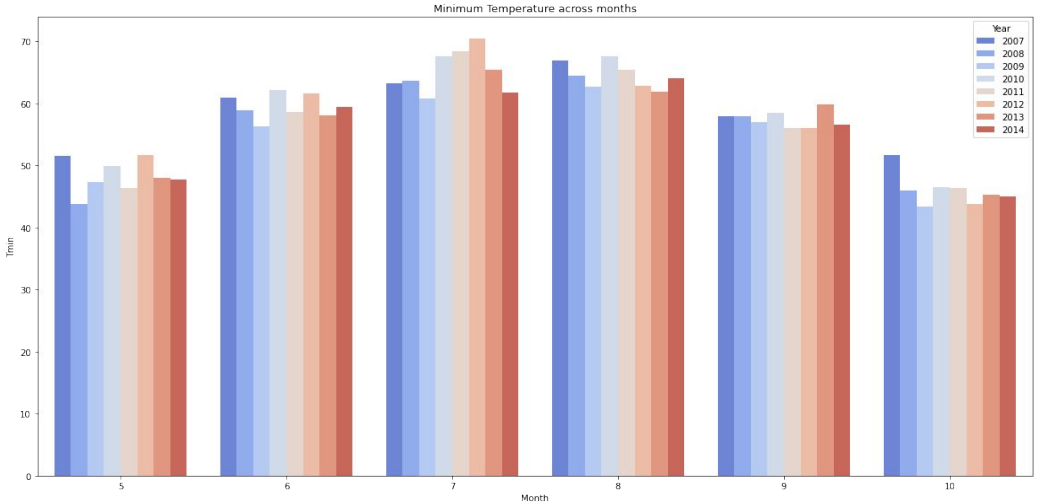
Number of Mosquitos vs Time



Number of Mosquitos vs Time



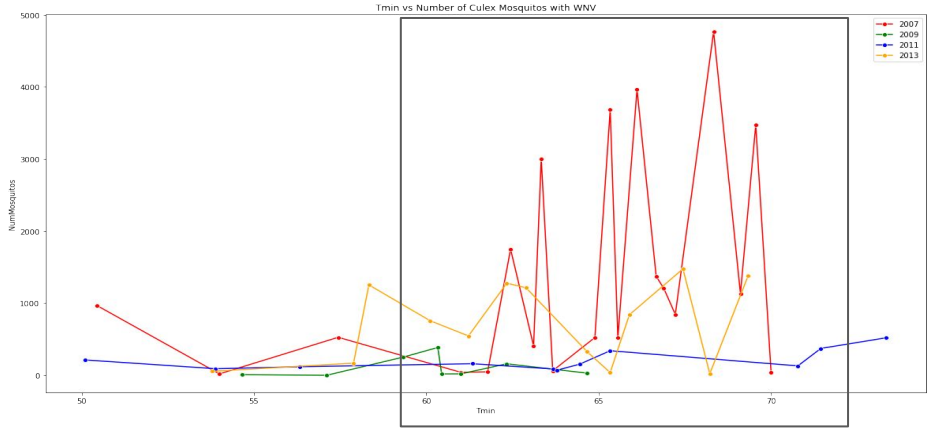
Weather trends in Chicago City



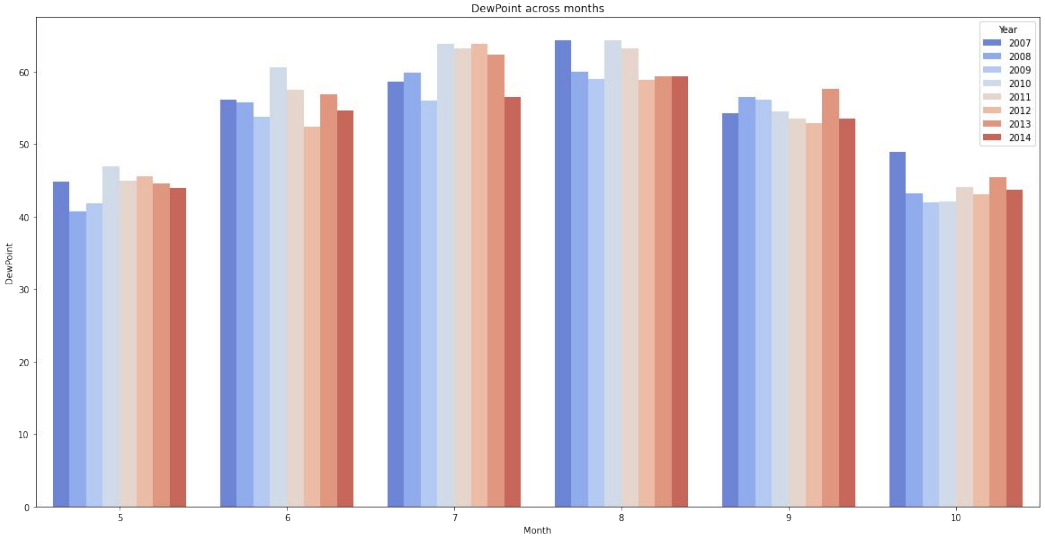
Air temperatures in Chicago peaked at 70°F

Ideal transmission range of WNV: 57°F - 70°F

West Nile Virus ‘activates’ at 73.4°F



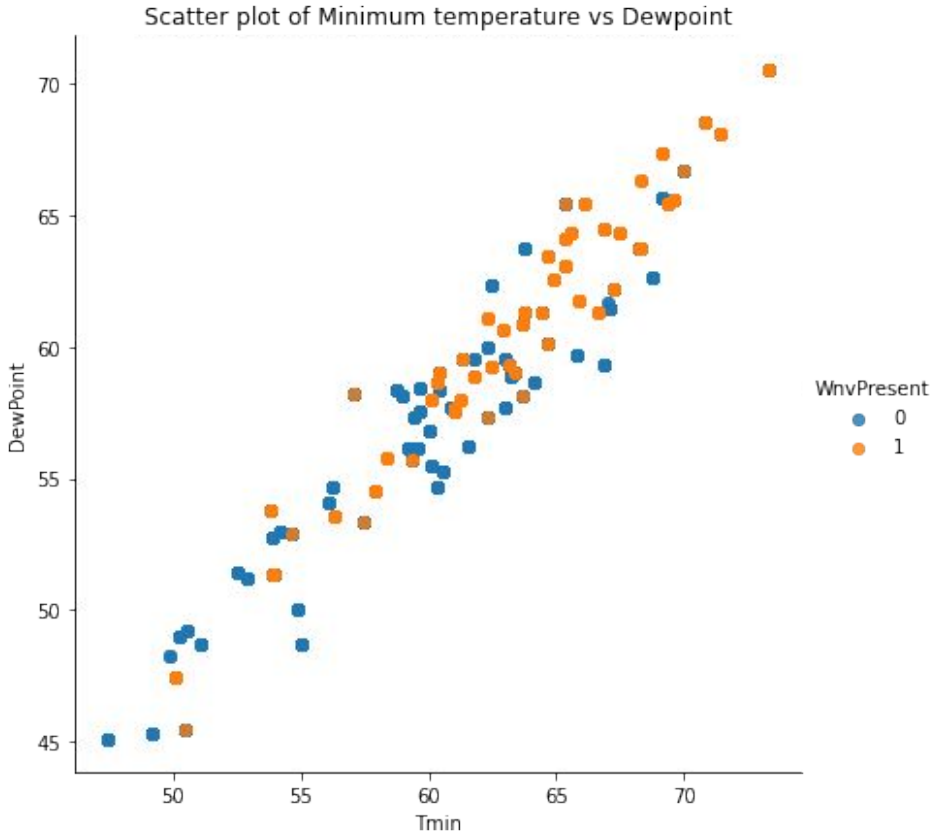
Weather trends in Chicago City



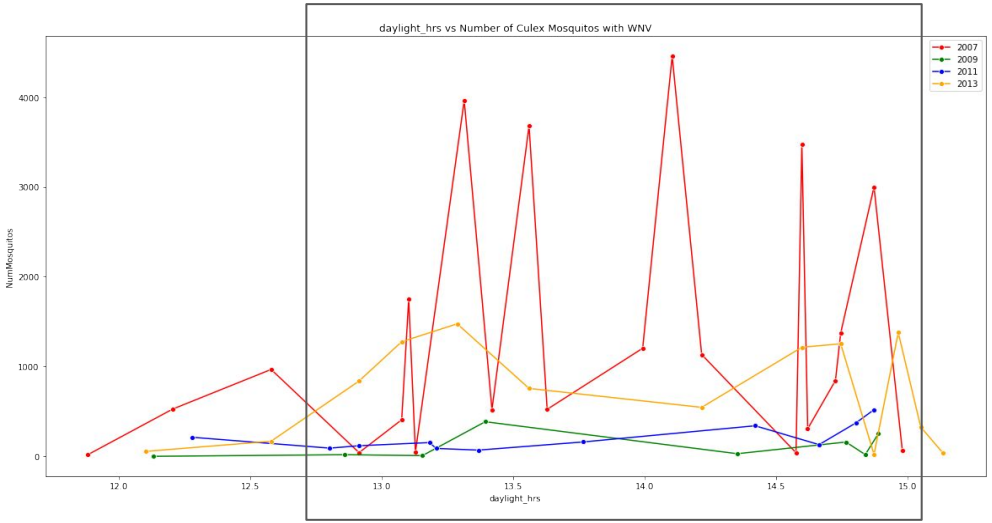
Dew point temperatures in Chicago peaked at 65°F

Higher chance of stagnant water
Increase viable breeding grounds for mosquitos

Weather trends and association with West Nile Virus



Weather trends and association with West Nile Virus



Culex Mosquitoes are nocturnal and feed between dusk and dawn

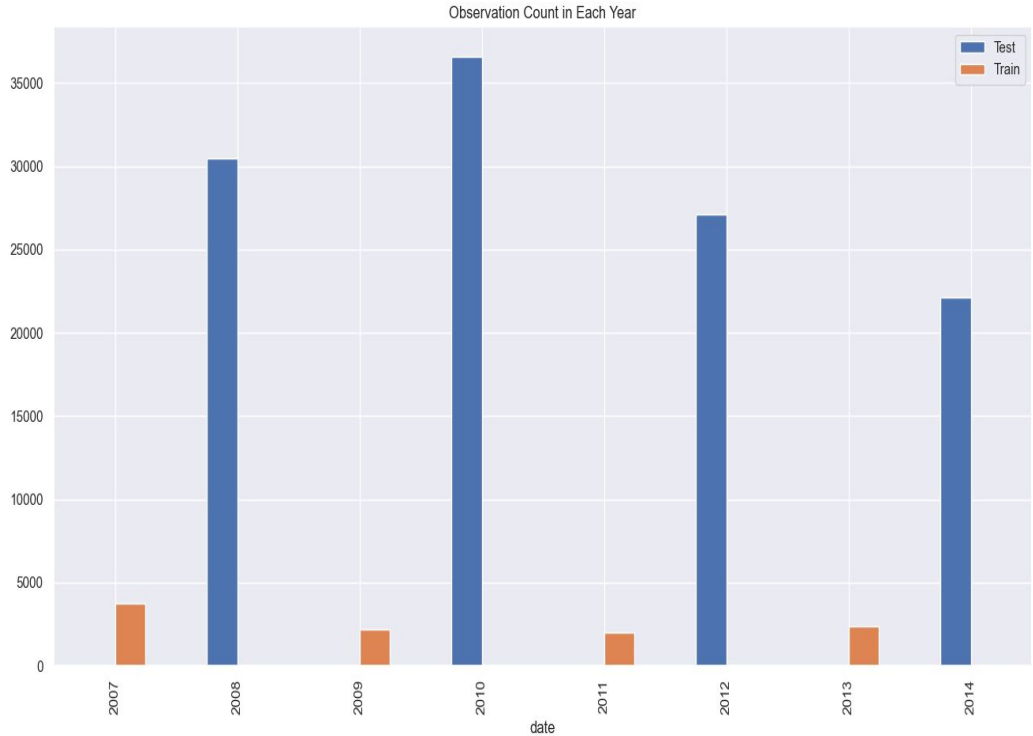
Above a certain threshold, transmission activity of the WNV between Culex mosquitoes increases with number of daylight hours

Surrounding air temperature is still the key factor to the transmission activity

Models & Metrics

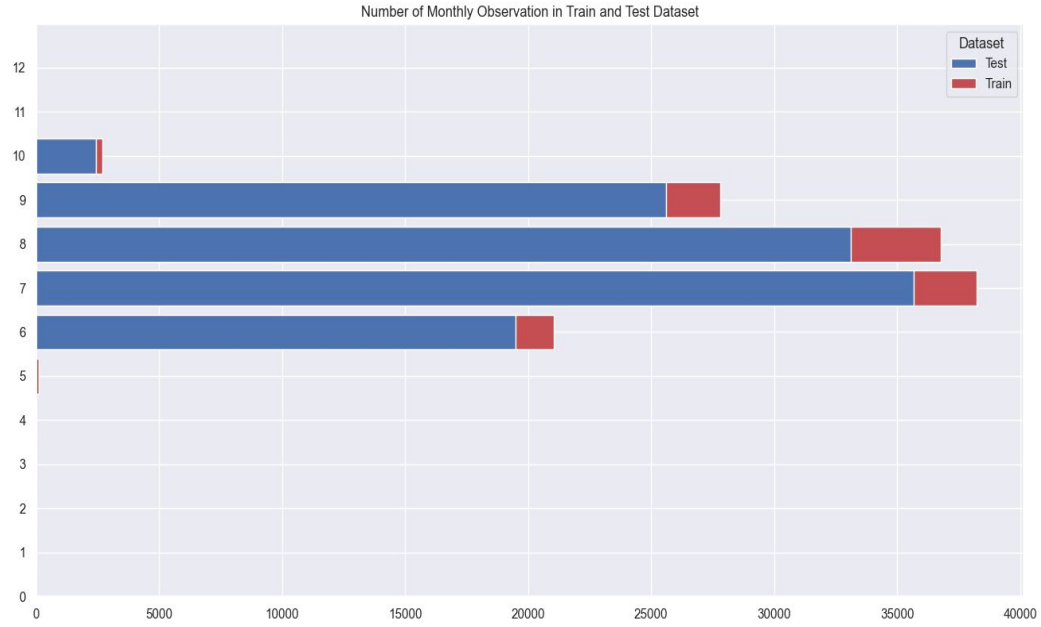
Structure of train and test data

	Train	Test
Year	2007,2009, 2011,2013	2008,2010, 2012,2014
Observation row	10,298	116,293



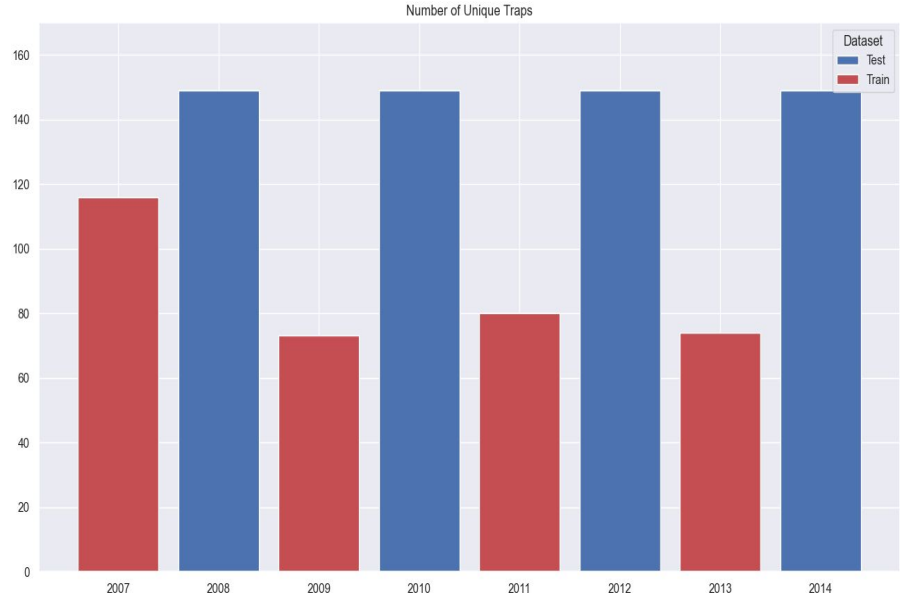
Structure of train and test data

	Train	Test
Year	2007,2009, 2011,2013	2008,2010, 2012,2014
Observation row	10,298	116,293
Months	May-Oct	June-Oct



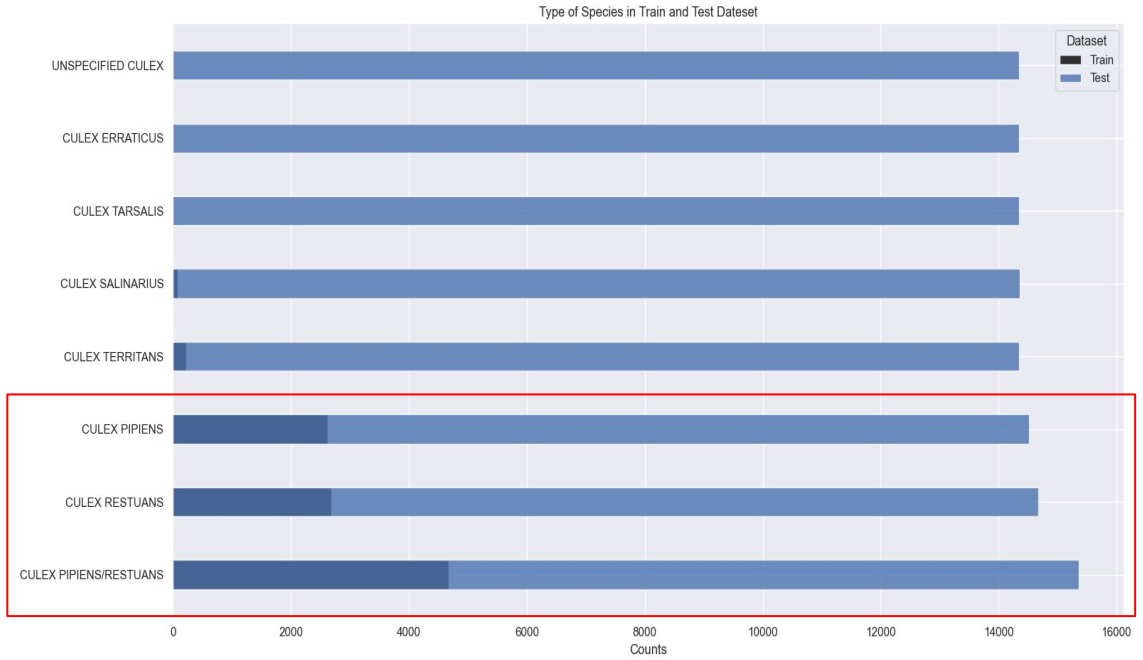
Structure of train and test data

	Train	Test
Year	2007,2009, 2011,2013	2008,2010, 2012,2014
Observation row	10,298	116,293
Months	May-Oct	June-Oct
Traps	136	149



Structure of train and test data

	Train	Test
Year	2007,2009, 2011,2013	2008,2010, 2012,2014
Observation row	10,298	116,293
Months	May-Oct	June-Oct
Traps	136	149
Species	6 Categories	Additional ‘unspecified Culex’



Data Cleaning

Most of the missing values were from Weather Station 2.

Impute using Weather Station 1 data.

The Weather Code Summary column was parsed and made into dummy variables

As part of our EDA, we also parsed the dates into ‘Year’, ‘Month’ to identify patterns between time periods

Model Feature Preprocessing

- Weather Data: Rolling mean for 9 days
  - This is to account for the mosquito breeding cycle
  - The dummy variables for Weather Codes can be interpreted as the chance that the weather event has occurred in the last 9 days and were treated as continuous variables
- Date
  - We converted Date to day of year, since we observed similar weather patterns over the same period between years
- One Hot Encoding
  - ‘Trap’
  - ‘Species’
- Class imbalance
  - Class weights {0:1, 1:10}

Models Used

- Null
  - Probability of WnvPresent in each trap from the train data set
  - Satellite Traps used Main Trap probability [e.g. T002A → T002]
- Logistics Regression:
  - Features used: ['Trap', 'DewPoint', 'Tmin', 'TS', 'RA', 'Longitude', 'dayofyear', 'Species']
  - Class weights = 1:10
  - Regularization parameter = 0.5
  - Ridge
- Decision Tree:
  - Features used: ['Trap','Species','DewPoint','Tmin','TS','RA','Longitude']
    *limited by computational power

Scoring Metric and Scores

Metric: ROC AUC (the expected true positive rate, averaged over all false positive rates)

Model	Null	Log Reg	Decision Tree
Train/Test split	-	0.843 / 0.802	0.807 / 0.790
Test data (Kaggle)	0.602	0.706	0.645

Data Insights

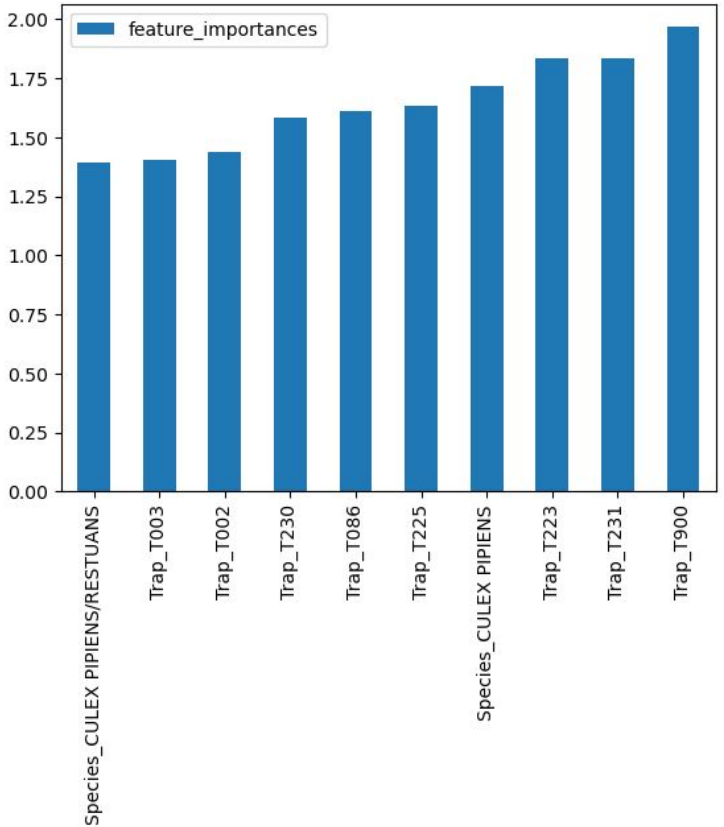
From the logistic regression model:

- Species
- Traps

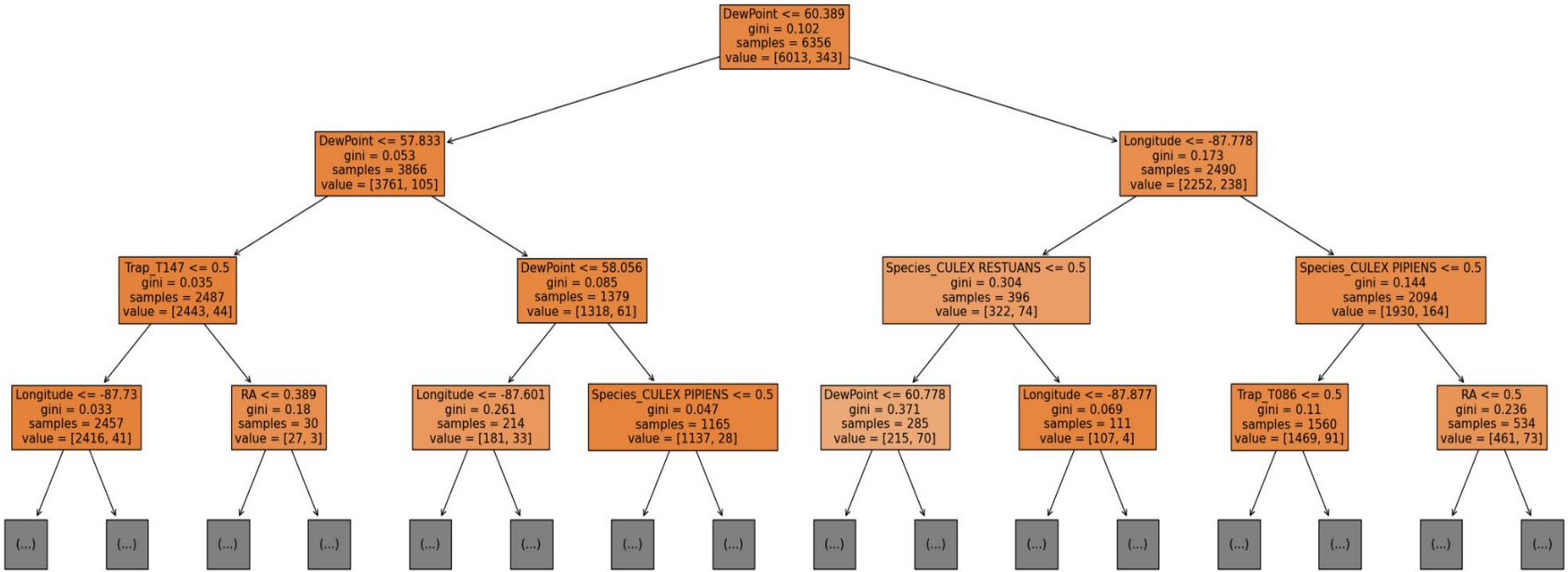
Out of 149 traps:

36 had a probability of 80% or greater

110 had a probability of 50% or greater



Dew Point, Longitude, Species are important



Improving the model

- More data
  - Density of people
  - WNV spread rates
  - Bird deaths by WNV

- More computational power
  - Increase features
  - Include polynomial features

- More subject matter expertise
  - Feature engineering

CBA

Cost Benefit Analysis

Costs and Benefits

COST

- **Sprays**
- **Manpower**
- **Research and development**

BENEFIT

- **Improved Public Health**
- **Improving Productivity**
- **Environment**

Costs and Benefits

	Broad-based Approach Across Chicago	**Targeted Approach** Predicted Traps
Cost ● Spray Cost	$390,604	$14,113 (149 unique traps)
Benefit ● Reduced Medical Costs ● Improved Productivity	$354,216 ● West-Nile Fever: $1,946 per case ● West-Nile Neuroinvasive Disease: $54,532 per case	
Benefit Cost Ratio	**0.907**	**25.1**

Cost of spray → 4 sprays between late May to Early October

390,000

149 unique traps → 14,410

Zenivex → 30 days

adulticides in conjunction with larvicides

Zenivex® E20 and Zenivex® E4 are oil-based formulations for use in Ultra Low Volume (ULV) applications and do not require dilution.

Recommendations

Benefits of spraying outweigh costs

- Prevention of 1 case of West-Nile Neuroinvasive Disease already covers the cost

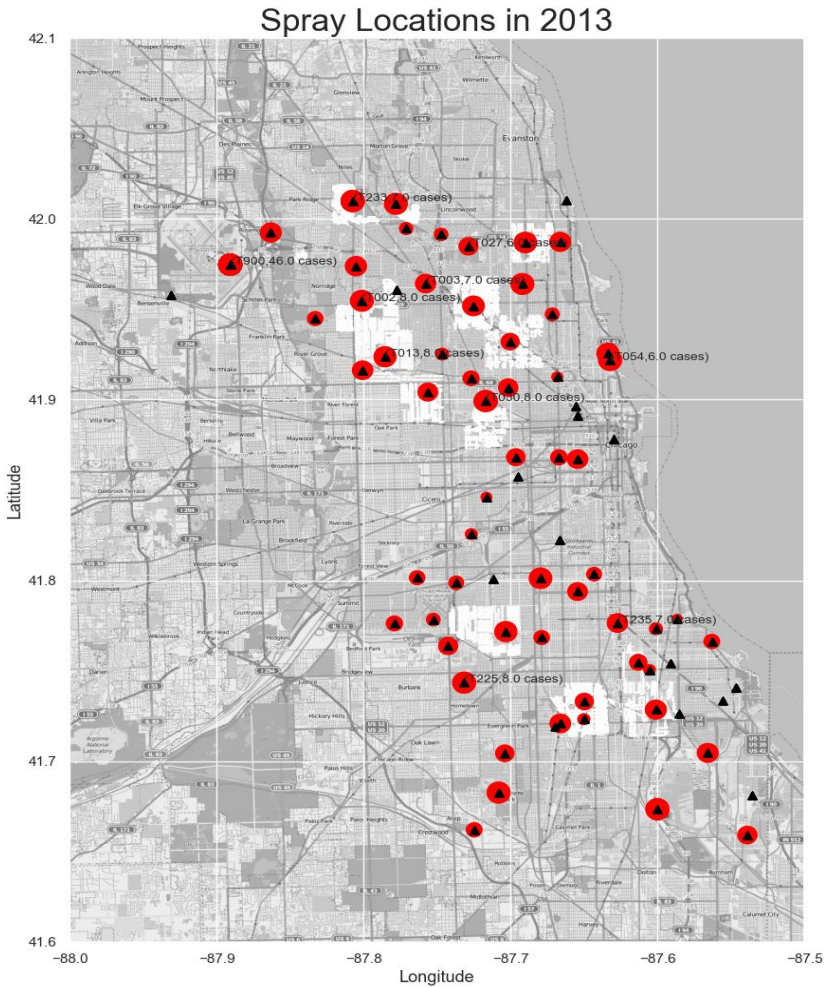
- Reduction of other mosquito-borne diseases (Zika, Malaria, Dengue)

Where to spray?

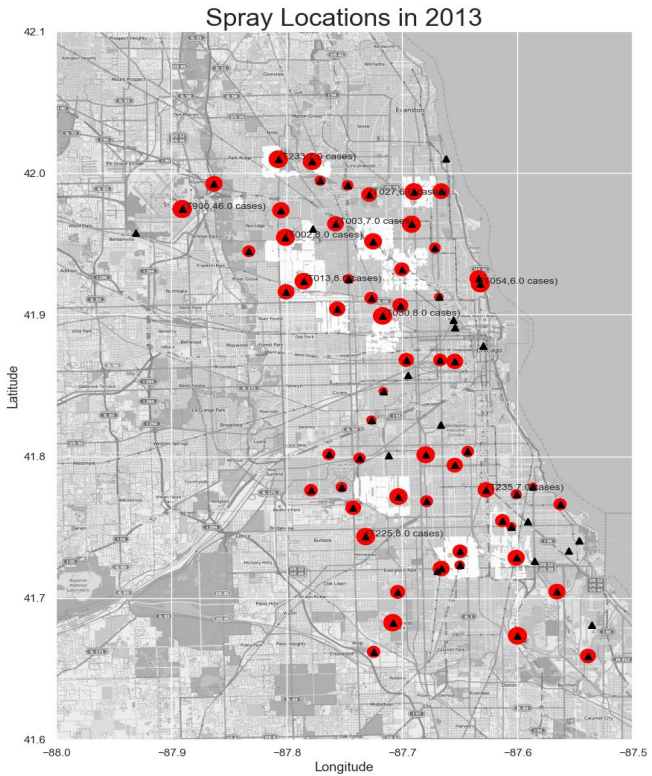
Spray is effective in reducing the virus carrier and halt the transmission to host.

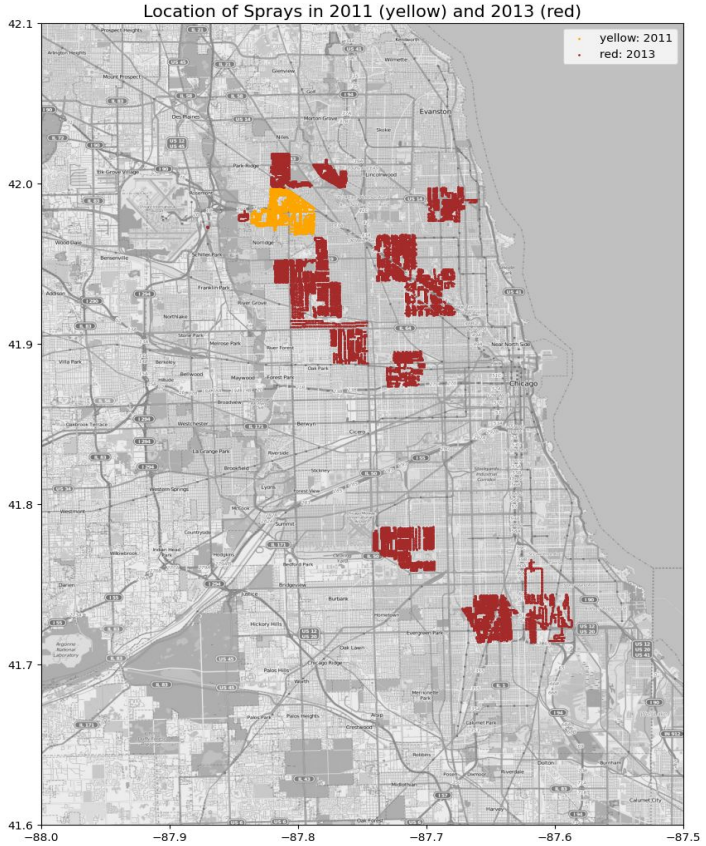
However, timing and spray location are important.

Our model can predict which traps are more likely to harbour WNV.

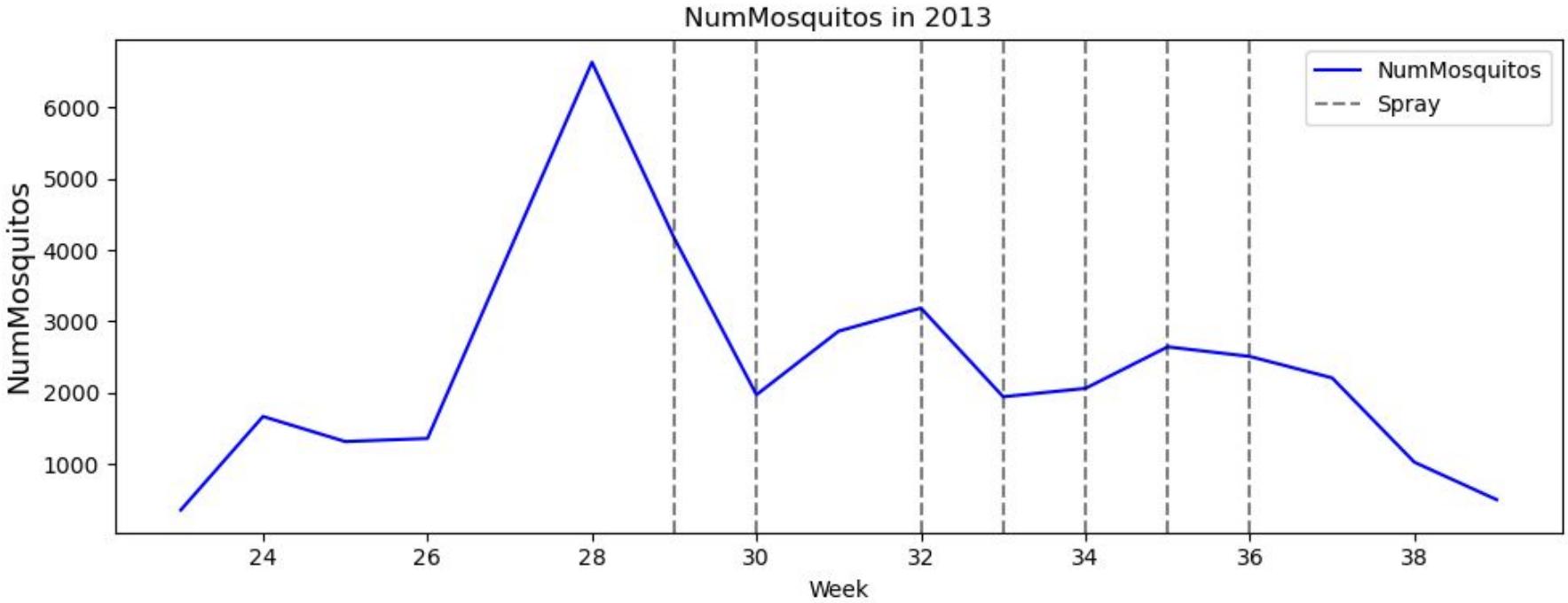


Where to spray?

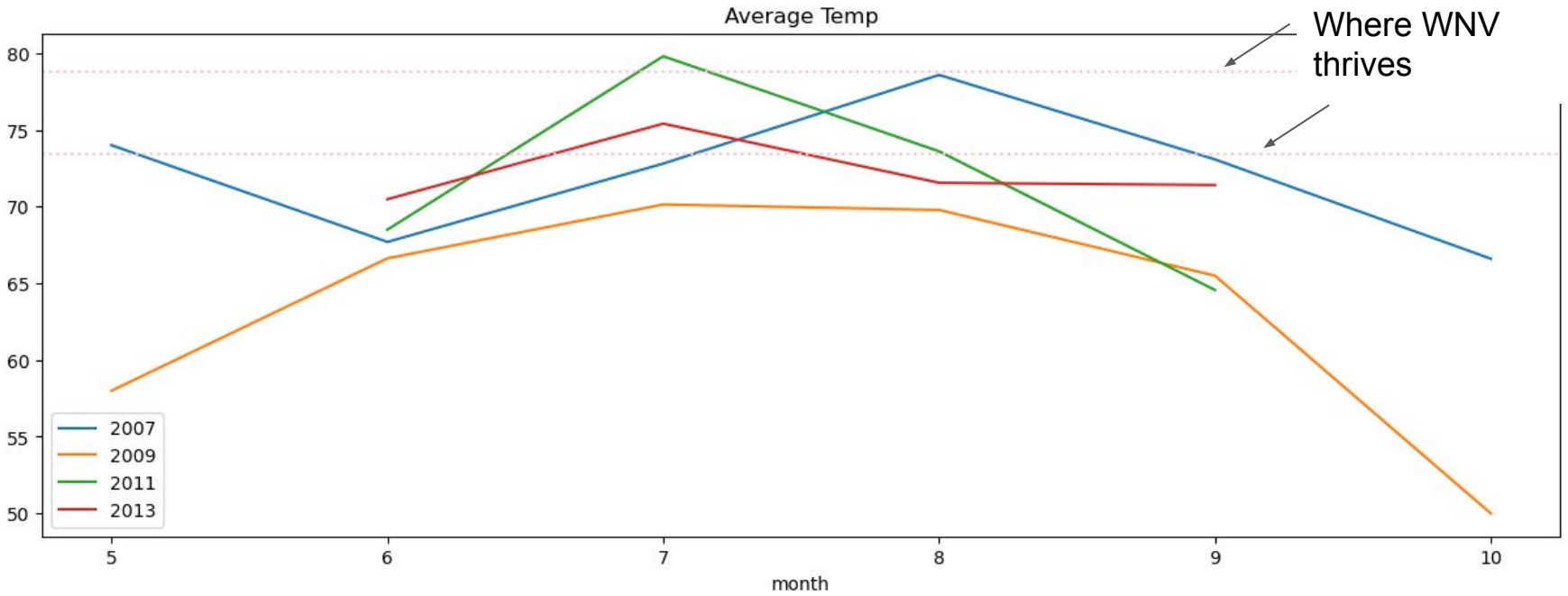




Spray at least 2 weeks earlier



Look at weather conditions



However…

While number of mosquitos reduced significantly in 2013, the proportion of WNV cases reported is the highest among other years.