DSI30 Project 4

Predicting West Nile Virus Infection of Mosquitos in Chicago

Edward | Keziah | Pei Yiing

Agenda

01 Introduction

02 Exploratory Data Analysis (EDA)

03 Modelling

04 Cost-Benefit Analysis

05 Cost-Benefit Analysis

01

Introduction

Background

First detected in North America

1999

Illinois had counted human cases (884) and deaths (64)

2002

There are no vaccines to prevent or medications to treat WNV in people

Present

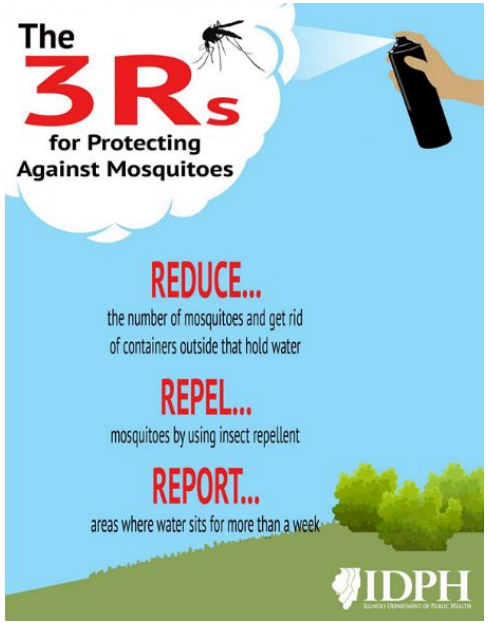
2001

Lab tests confirmed its presence in two dead crows.

2003

The worst years for having the most cases of 9,862

sources: https://dph.illinois.gov/topics-services/diseases-and-conditions/west-nile-virus.html



Problem Statement

To predict the presence of WNV area during the summer.
Establish a vector control plan

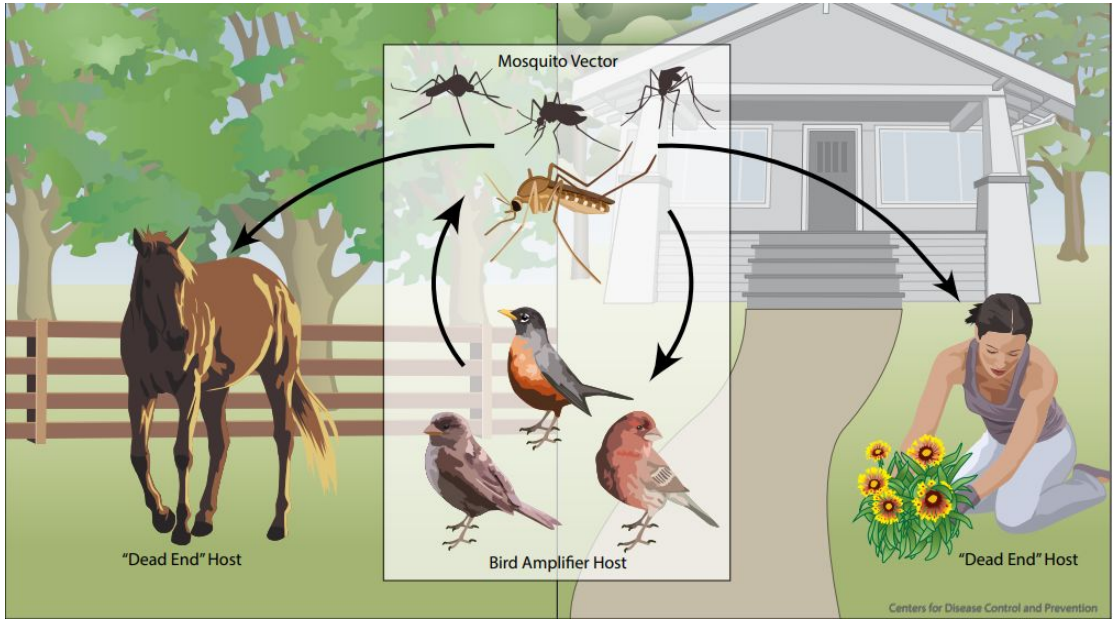
West Nile Virus Transmission Cycle

Infected Birds have the virus in their bloodstream.

Mosquitoes become infected by biting these infected birds

After about a week, infected mosquitoes can pass the virus to more birds.

Mosquitoes also infect people and horses and other mammals.



USD $500

Seasonal spraying from May to September of a ½ an acre property

21 Days

Sprays your yard every three weeks, typical life cycle of the mosquito

USD $8,124

West Nile meningitis ($p = 0.0004$) and $192 for West Nile fever ($p<0.0001$)

02

Exploratory Data Analysis

Our Dataset

Mosquito Trap & Location

- Trap Coordinates
- Mosquitoes Captured Quantity
- WNV Presence

2007 - 2014

Train - 2007, 2009, 2011, 2013
Test - 2008, 2010, 2012, 2014

Spray Schedule

- Spray Dates
- Spray Coordinates

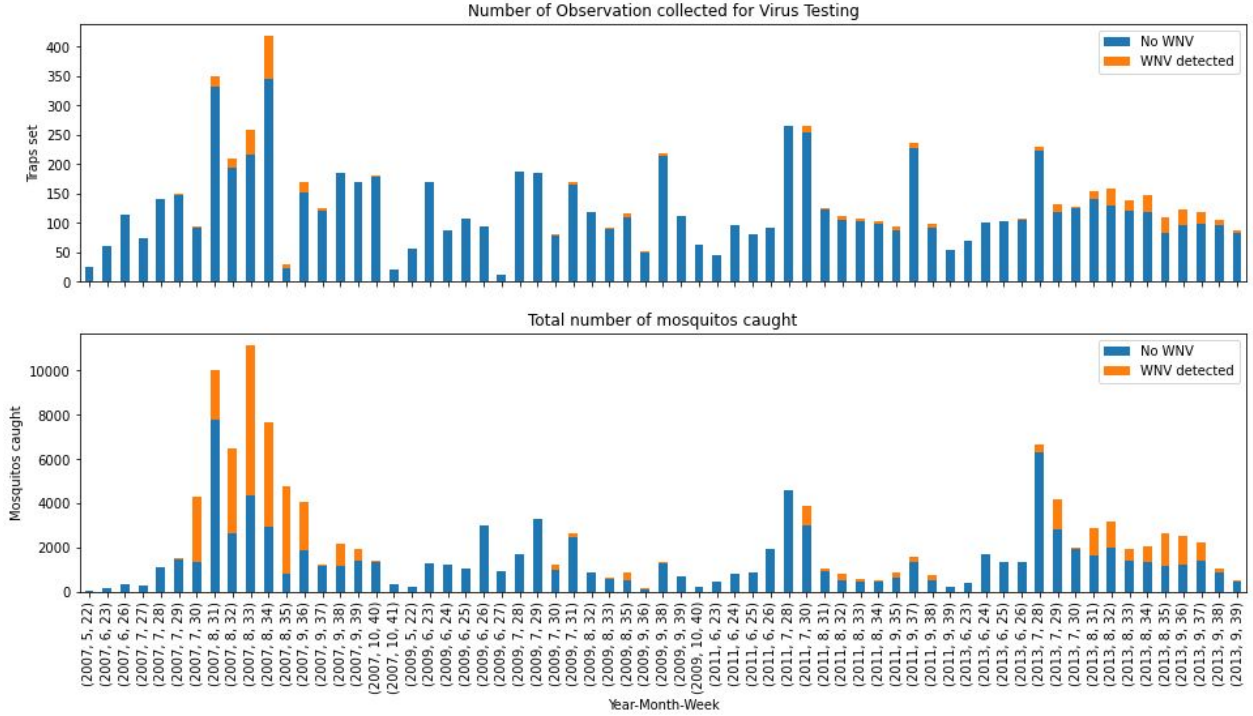
2011 - 2013

Weather Details

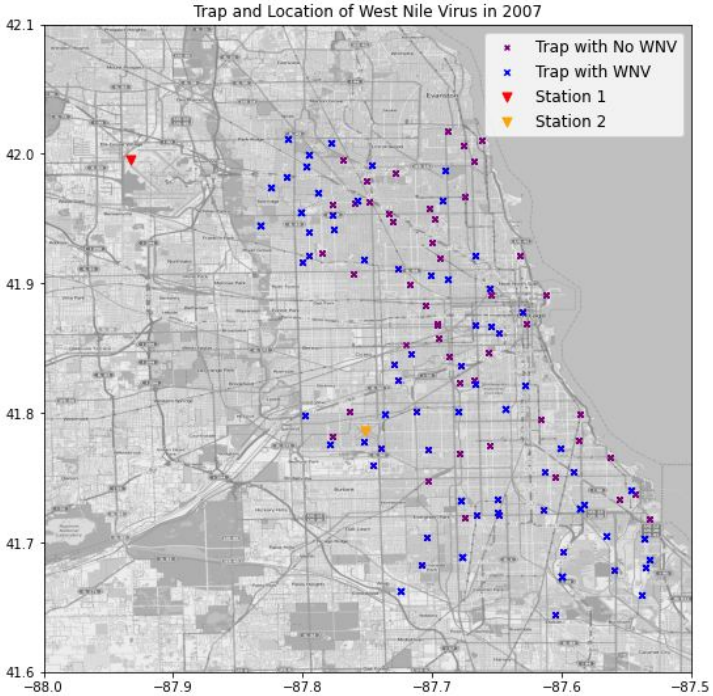
- Temperature
- Humidity
- Precipitation Total
- Code Sum (Weather Code)

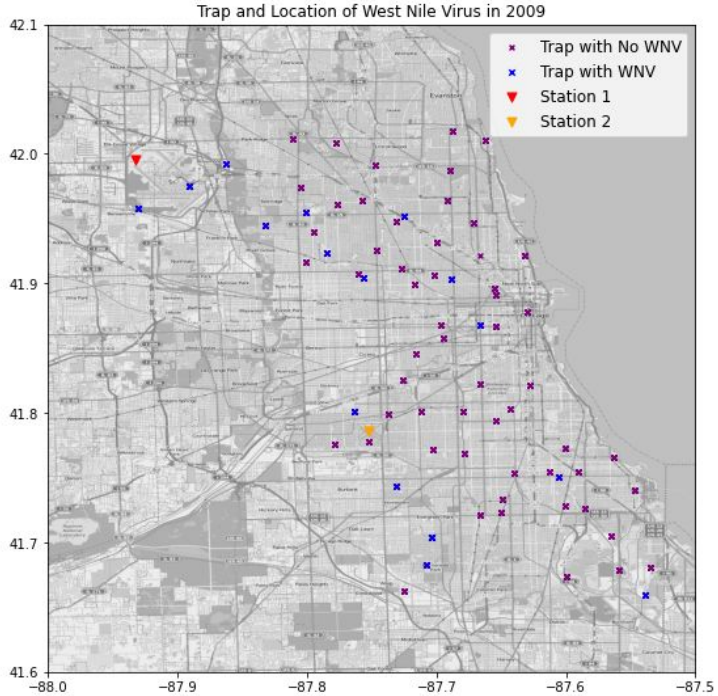
2007 - 2014

Mosquitoes Observations

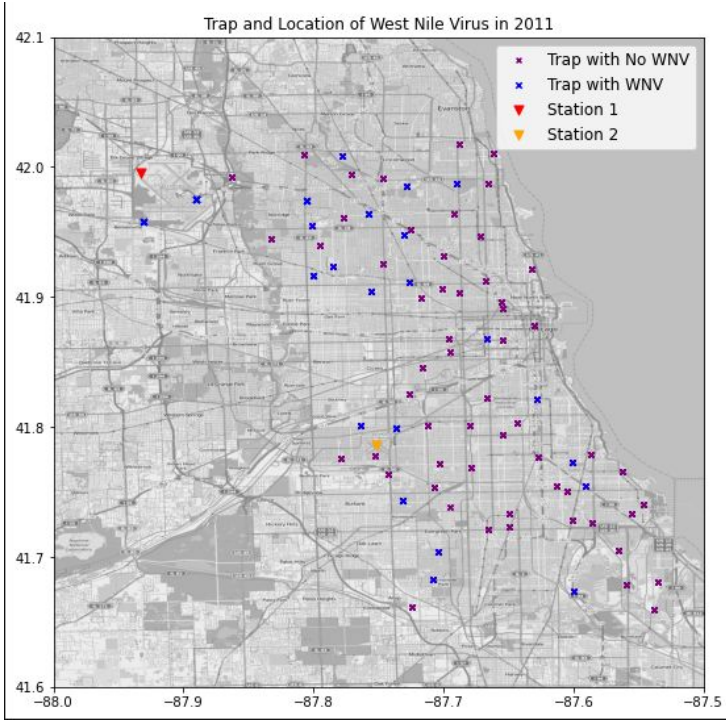


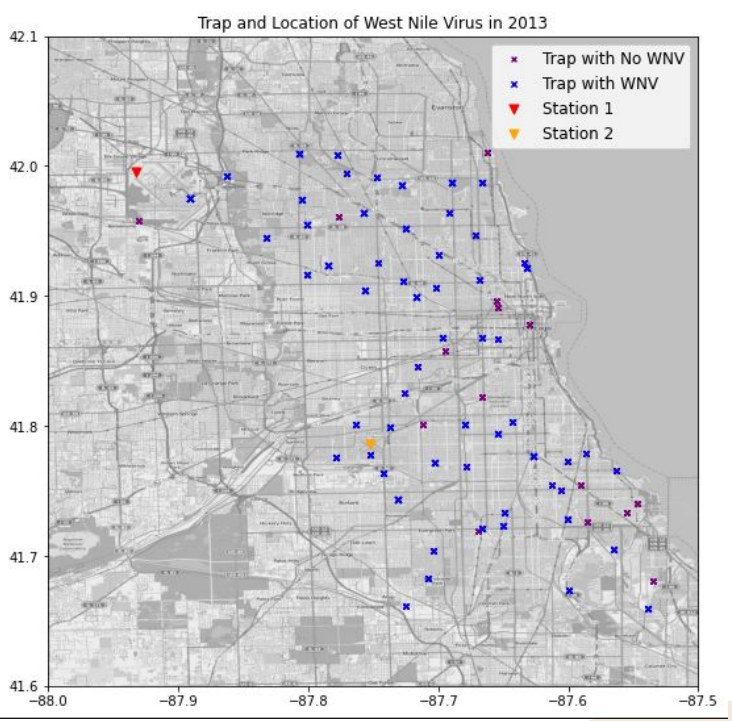
Traps Deployed



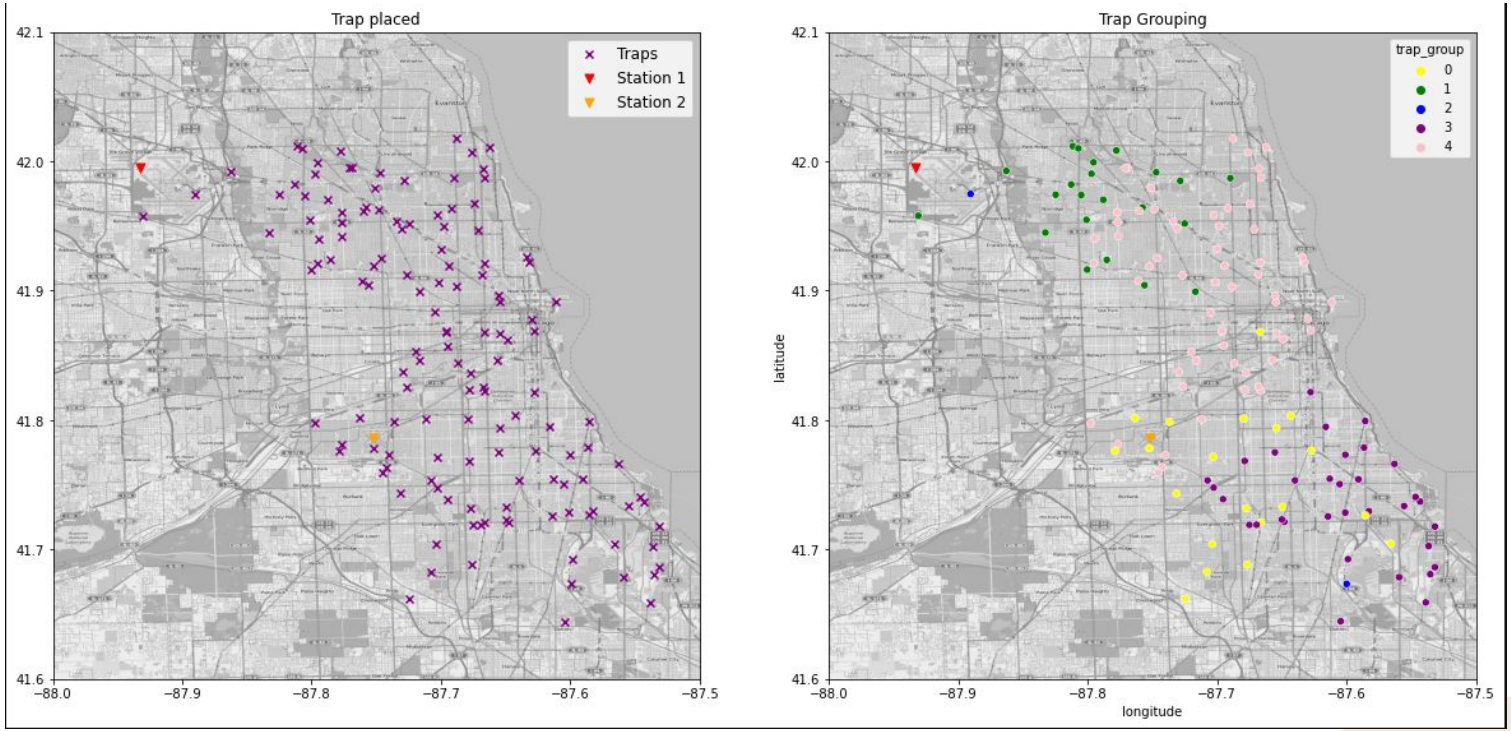


Traps Deployed

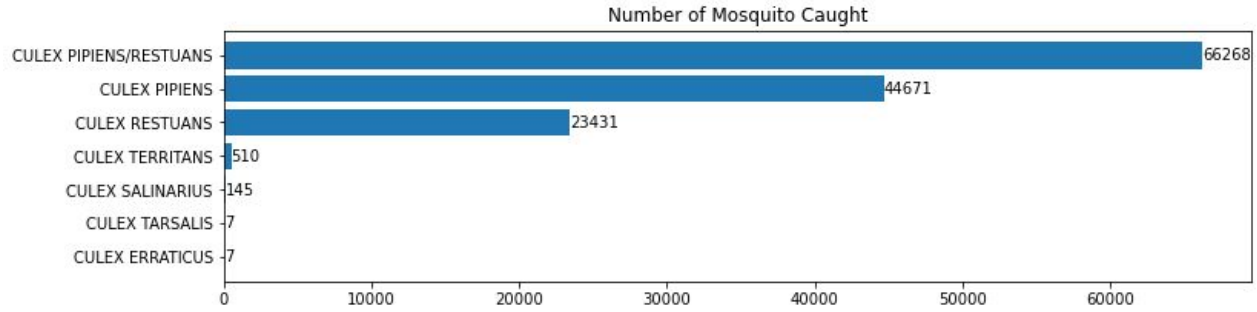


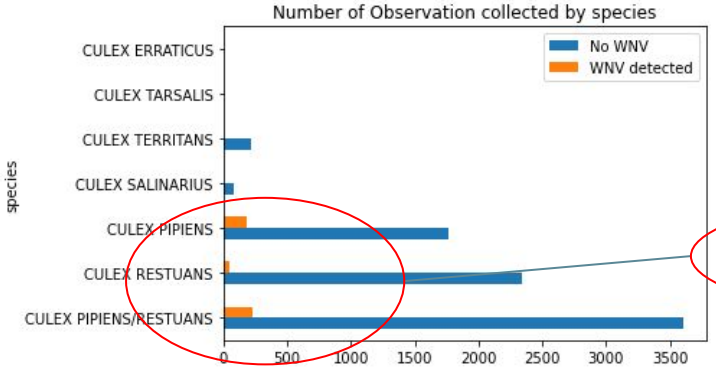


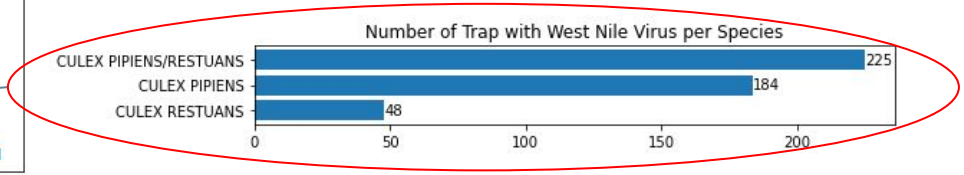
Trap Clusters



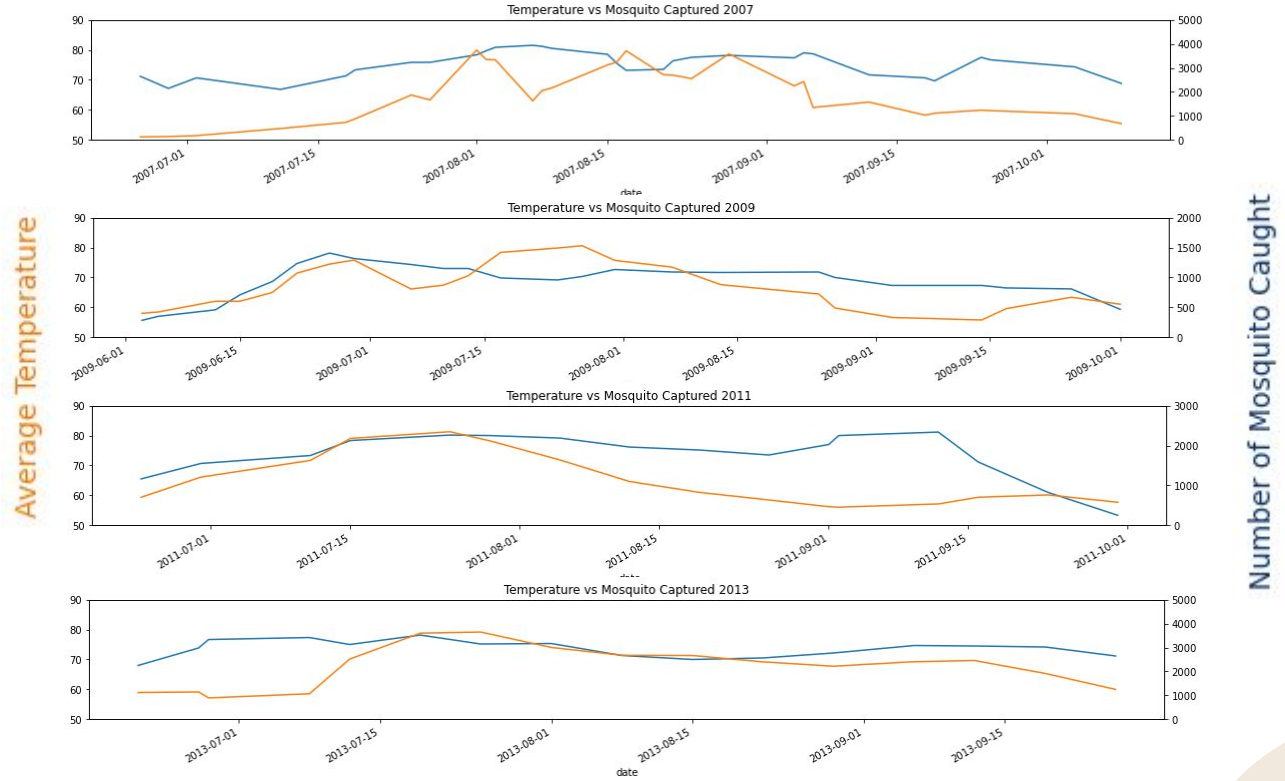
Mosquitoes Species



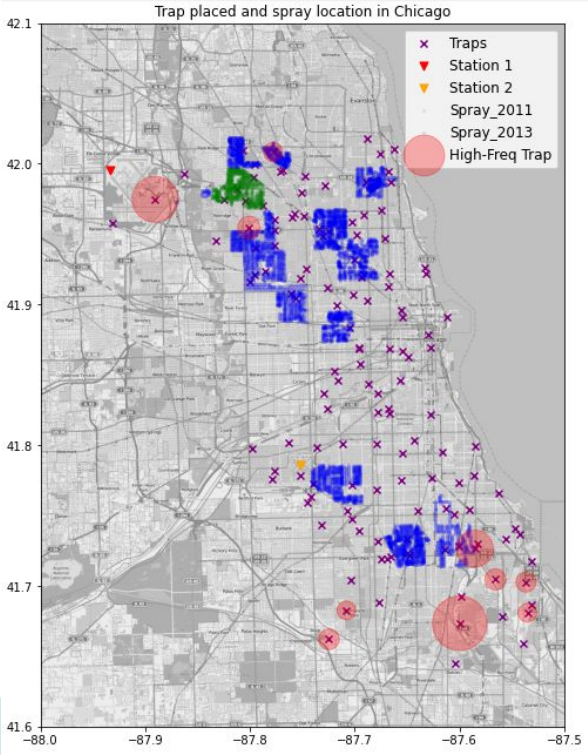


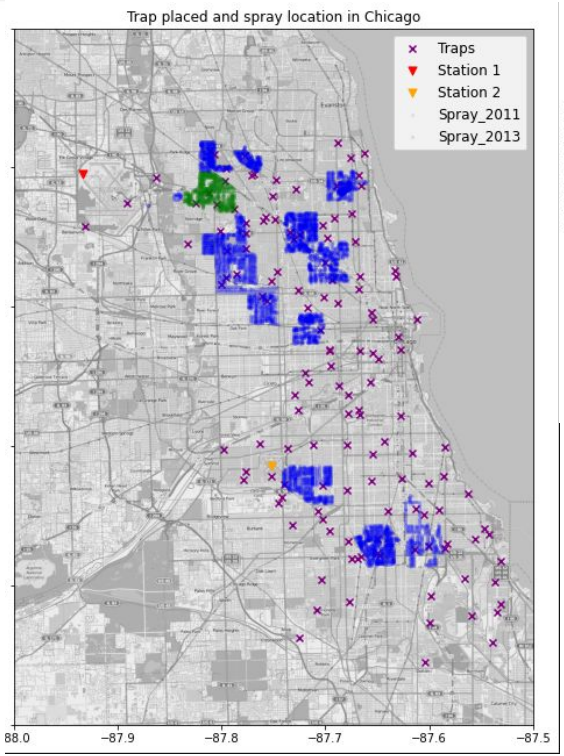


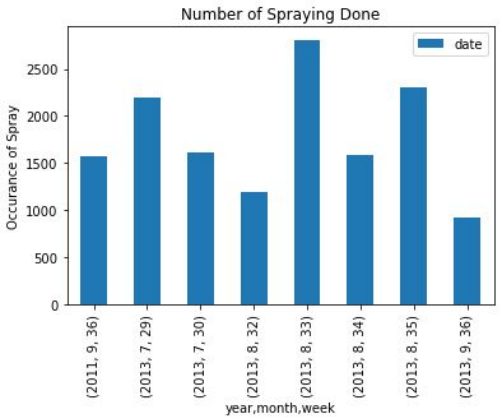
Temperature vs mosquito captured



Spray Location







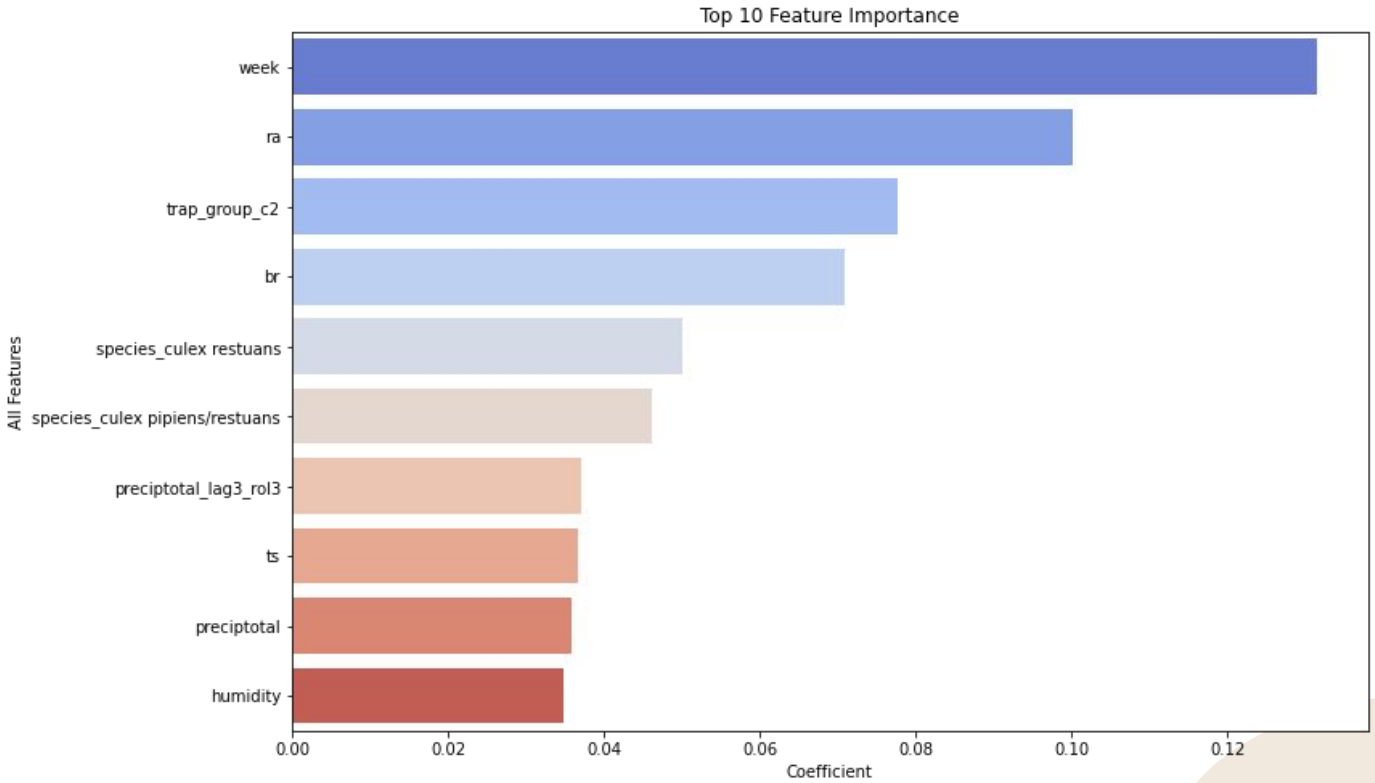
03

Modelling

Model Comparison

Model	ROC AUC Train	ROC AUC Validation	Sensitivity
Logistic Regression	0.793587	0.798697	0.659341
AdaBoost Classifier	0.837229	0.838580	0.846154
Gradient Boost Classifier	0.904576	0.843283	0.604396
XGBoost Classifier	0.953800	0.852166	0.879121
Random Forest Classifier	0.872042	0.832202	0.714286

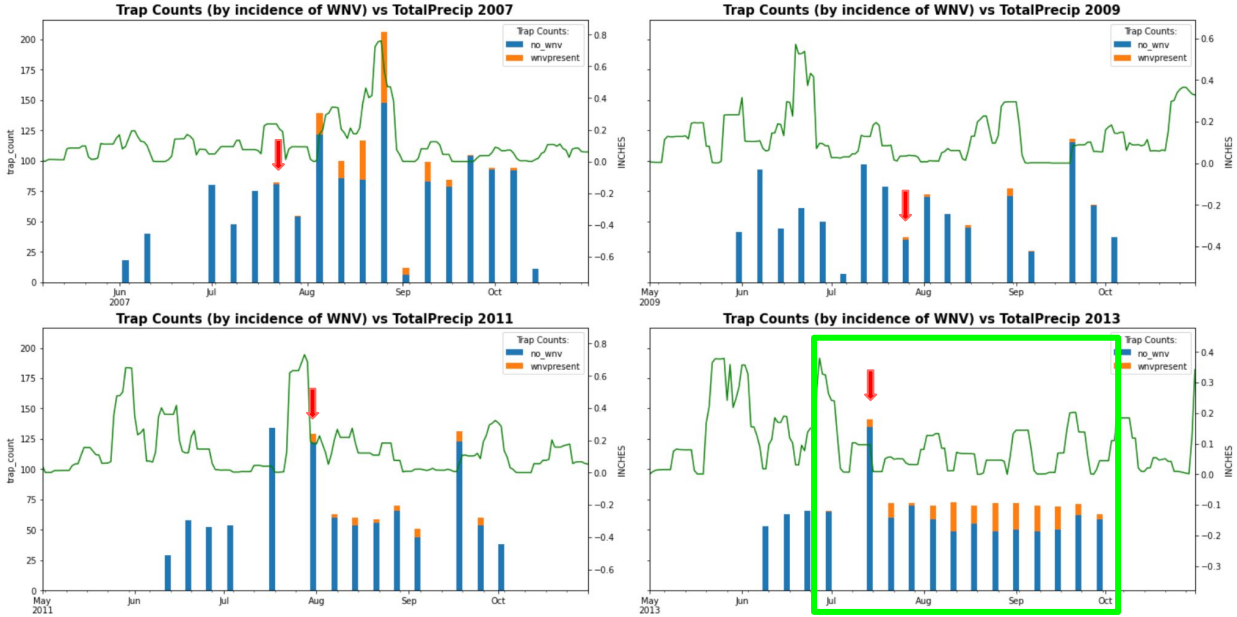
Features Importance



04

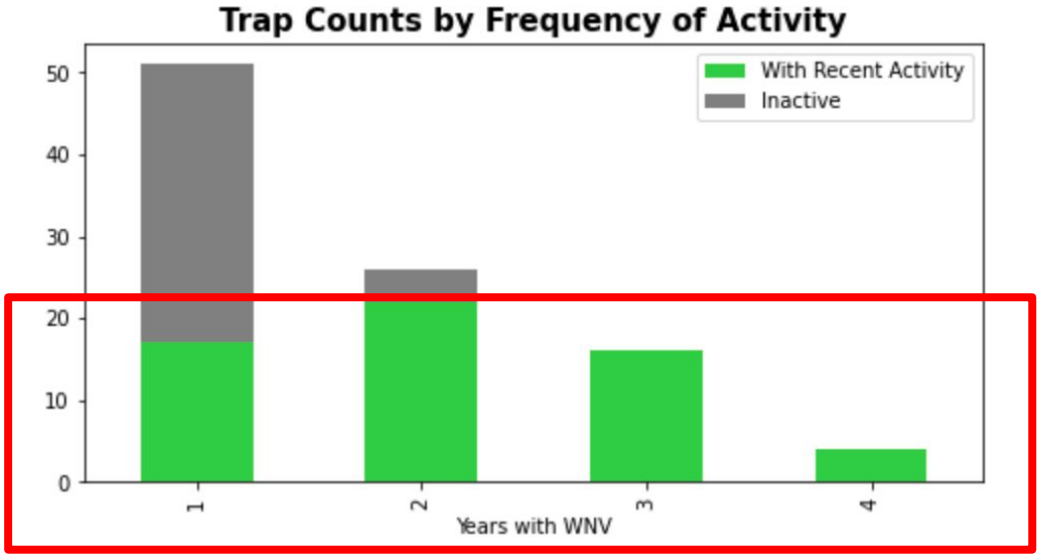
Cost-Benefit Analysis

Time and Duration of Vector Control Application



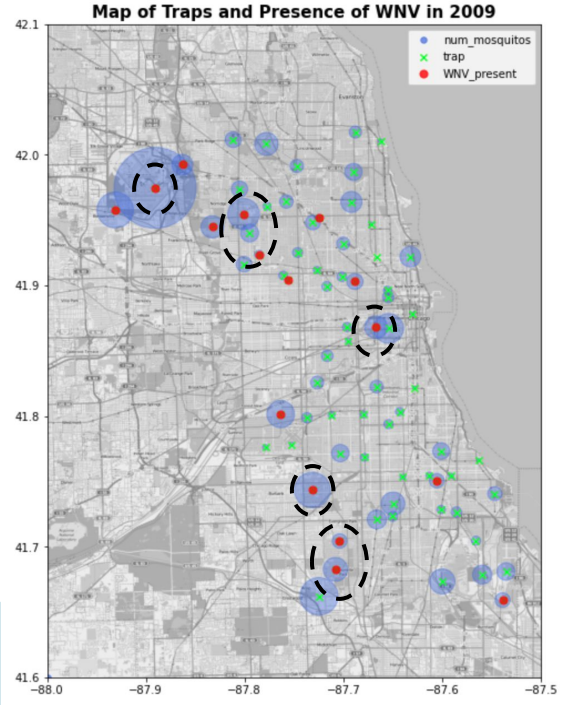
- Begin active surveillance and vector control in **first week of July till the end of September** - subject to weather conditions
- Frequency of control measures would depend on method used.

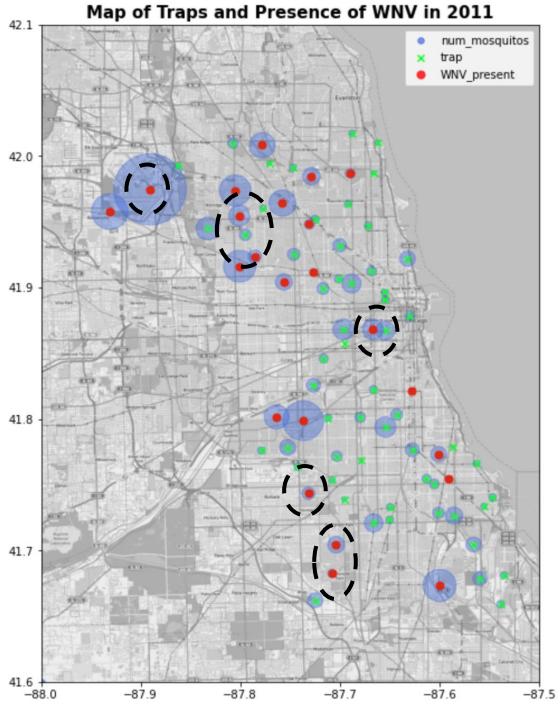
Location of Vector Control Application

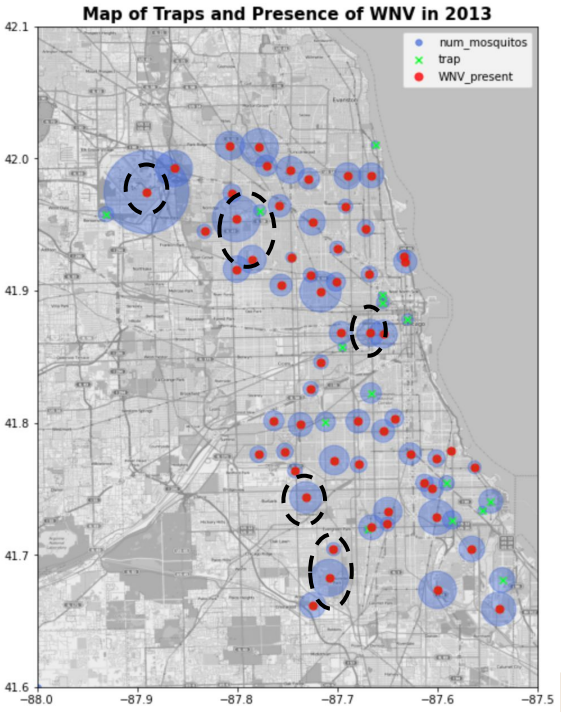


We should increase surveillance and vector control at trap locations that are frequently infected.

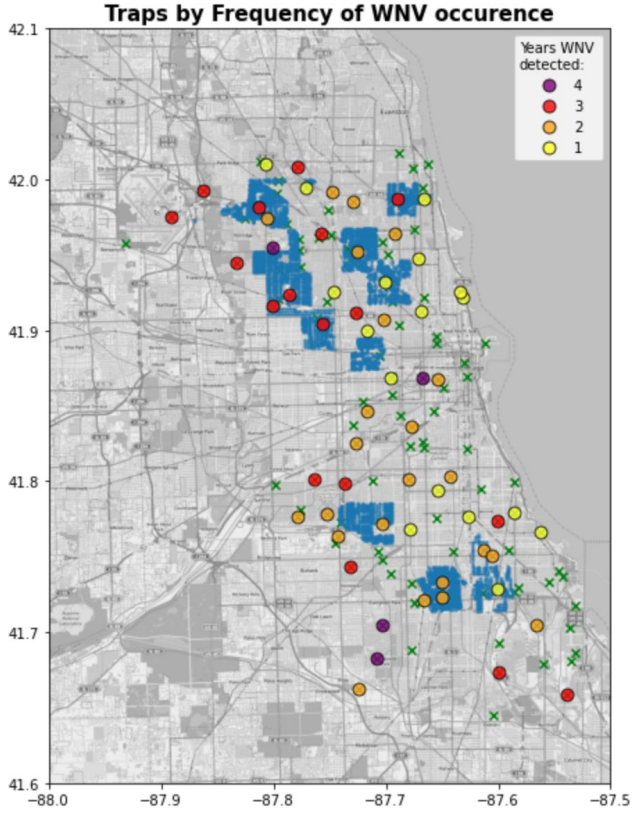
Target: Frequently Infected Clusters

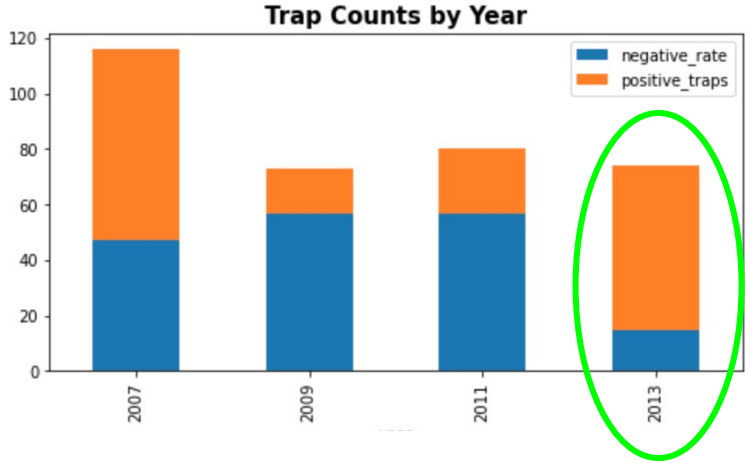






Spray Locations vs Incidence of WNV





- Vector control efforts in 2011 and 2013 appear to have missed these hotspots.
- This could have been one of the factors contributing to the spike in WNV infection in 2013

05

Conclusion

Summary

1) Best Model based on sensitivity: **XG Boost**.

2) **Top Features** that predict incidence of WNV:
   - Week
   - Weather conditions: Rain, Mist
   - Trap cluster
   - Species of Mosquito

3) Strategies to **maximize benefits of vector control** while minimizing cost:
   - Start active vector control in first week **July** instead of May/June
   - **Target hotspots** for increased surveillance/vector control measures

Proposed Future Works

1) **Refine Current Model:**
   - Re-engineer some existing features
   - Add more features:
     - Wind Speed/Direction, Thunderstorms, Mist
     - Geographical Features
     - Density of Vector Host Birds

2) **Build and optimize new model**:
   - Recurrent Neural Network

Thank you!

Any Questions?