

WEST NILE VIRUS ANALYSIS

DONE BY:

Jeffrey Sim, Katherine Wang, Koh Yong Tat, Luan Yuanfeng





DATA CLEANING & EDA



MODELLING



SPRAY & COST BENEFIT ANALYSIS



INTRODUCTION

West Nile Virus is a potentially fatal mosquito-borne disease, which is spread to people by the bite of an infected mosquito.

GOAL OF THIS PROJECT



**REDUCE WNV
CASES**

**COST BENEFIT
ANALYSIS**

**PREDICT WNV
LOCATIONS**

**PESTICIDE SPRAY
STRATEGY**

DATA CLEANING & EDA



1. TRAIN

Datapoints: 10,506

Number of features: 12

Time frame: 2007, 2009, 2011, 2013

3. WEATHER

Datapoints: 2,944

Number of features: 22

Time frame: 2007 - 2014 (all years)

DATASETS

2. TEST

Datapoints: 116,293

Number of features: 11

Time frame: 2008, 2010, 2012, 2014

4. SPRAY

Datapoints: 14,835

Number of features: 4

Time frame: 2011 and 2013

FEATURES AVAILABLE

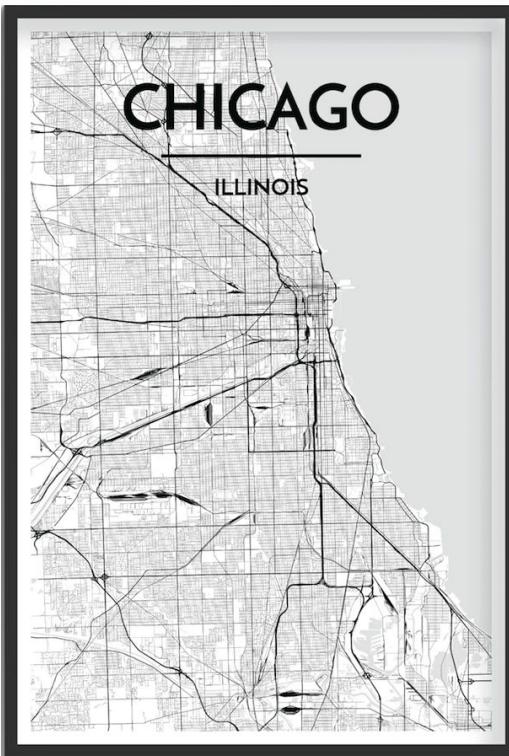
Location



Time



Species



Temperature



Humidity



Precipitation



Wind

DATA CLEANING

IMPUTE SPECIAL SYMBOLS



'M', 'T', '-'

IMPUTE WEATHER DATA



- Previous day
- From other station

CORRECT DATA TYPE



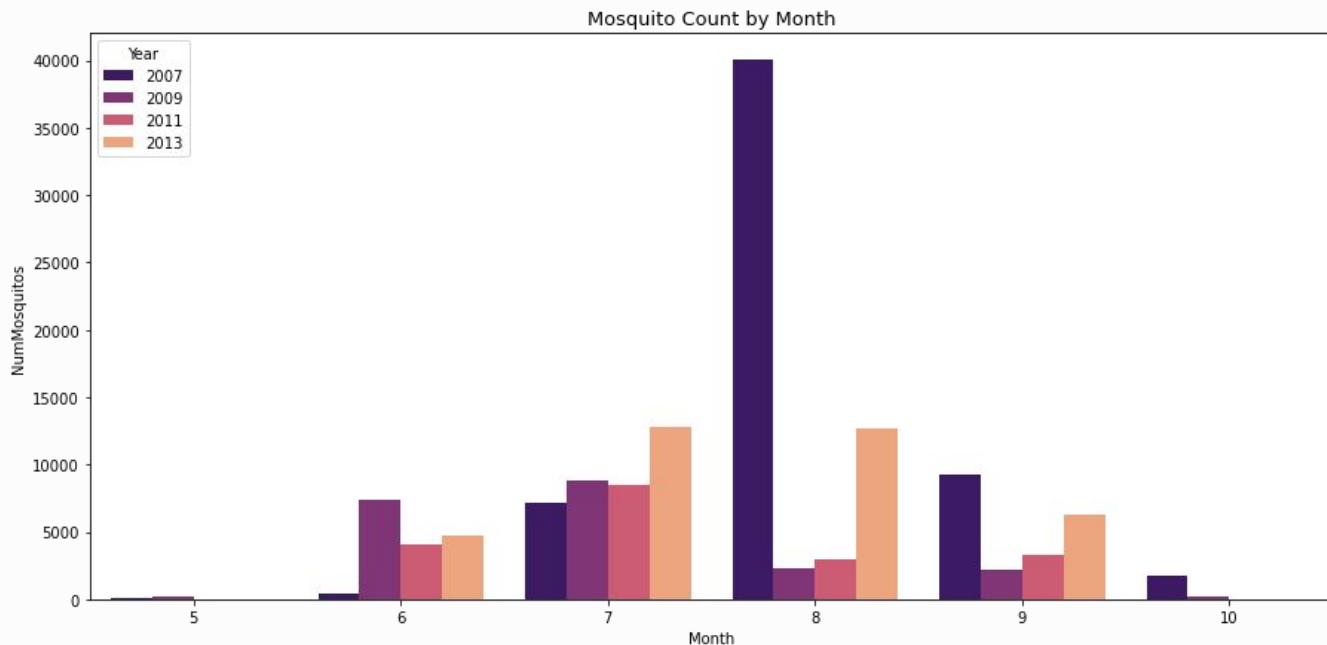
Object & float type

TIME

June to September

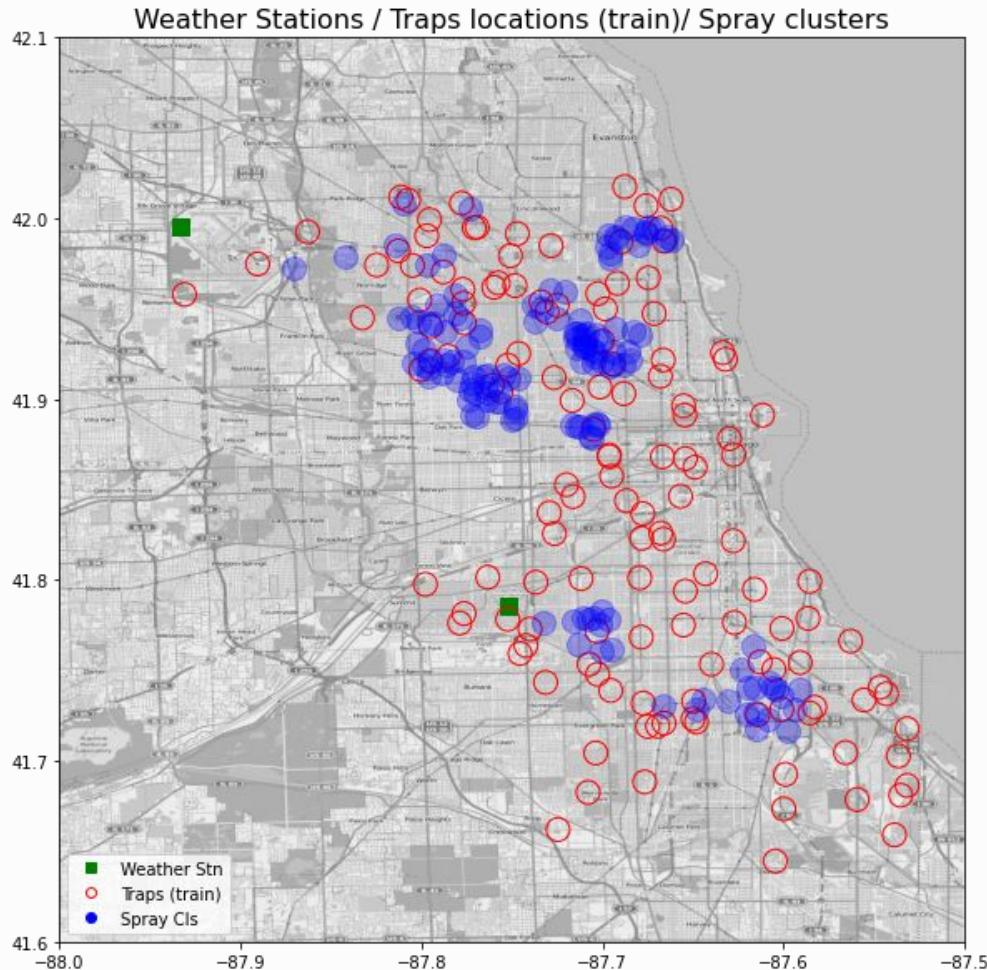
	NumMosquitos	WnvPresent
Year		
2007	58688	236
2009	21108	19
2011	18799	57
2013	36444	239

	Tavg	DewPoint
Year		
2007	26.316	66.769
2009	22.059	58.575
2011	24.150	61.724
2013	22.869	58.144



LOCATION

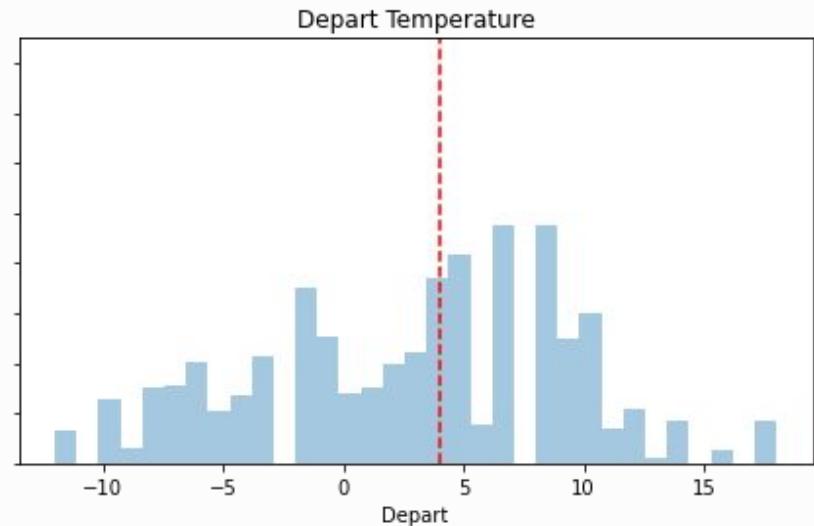
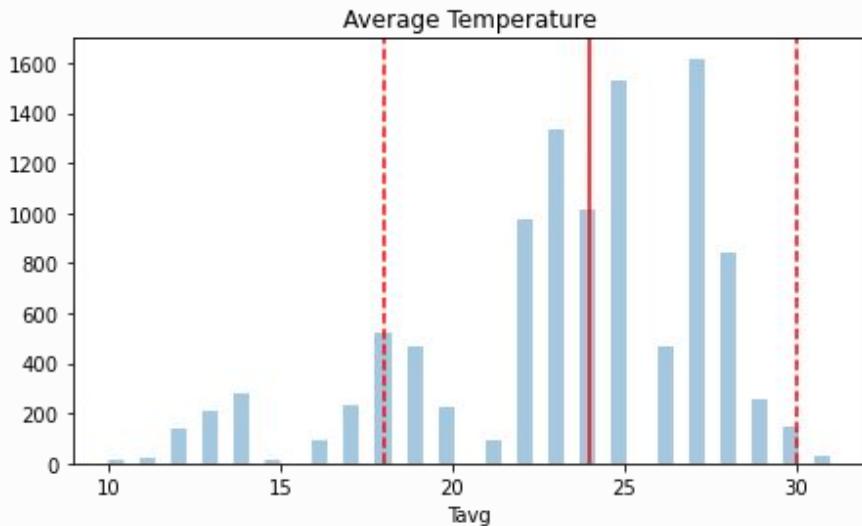
```
CULEX PIPiens/RESTUANS    262  
CULEX PIPiens               240  
CULEX RESTUANS              49  
Name: Species, dtype: int64
```



TEMPERATURE

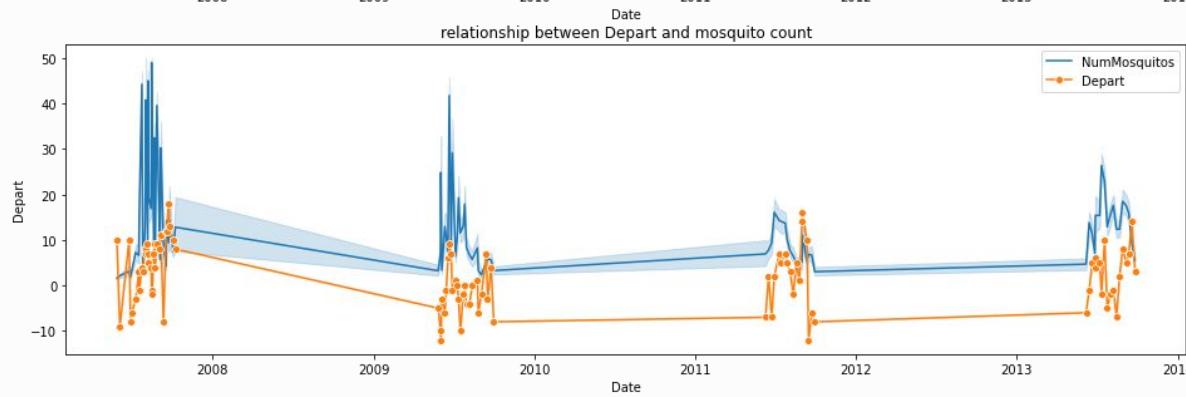
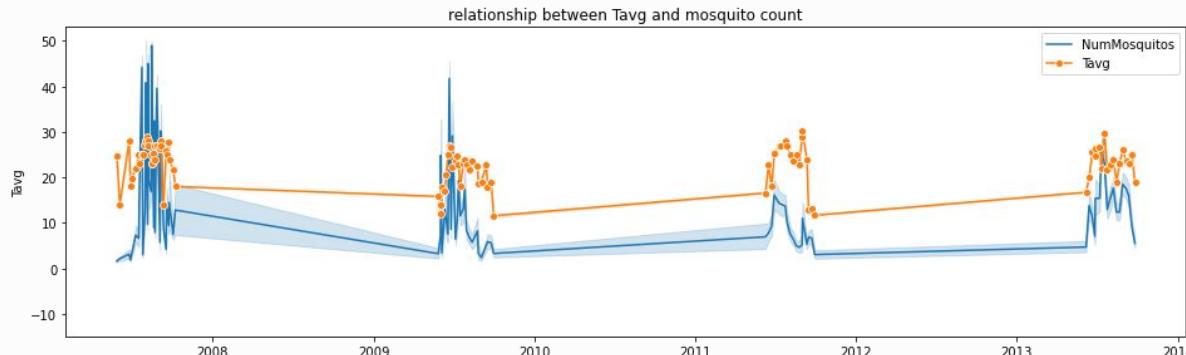
- Transmission peaks at 23-26 Degrees Celsius
- Mosquito is able to emerge from larvae between 18 to 30 Degree Celsius

Temperature Distribution



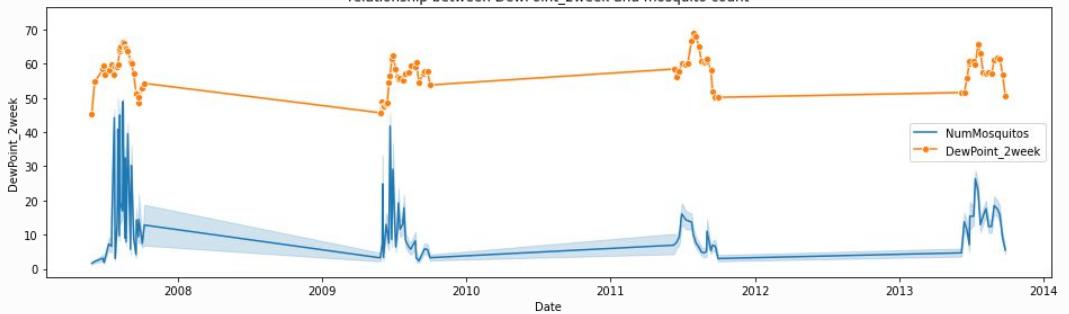
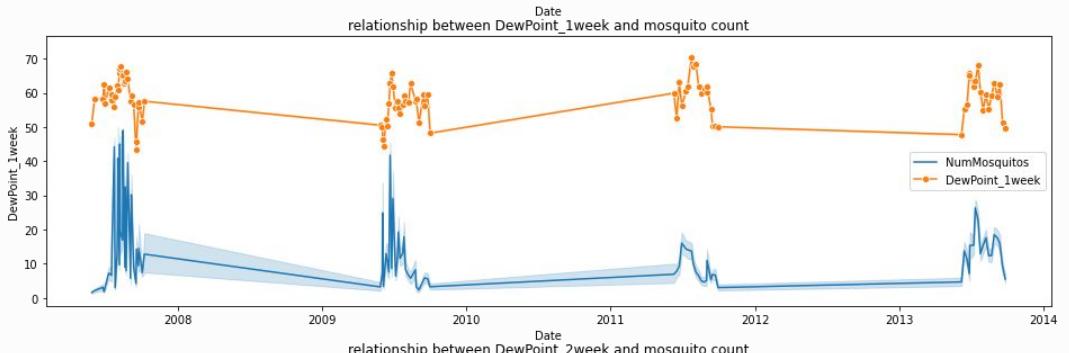
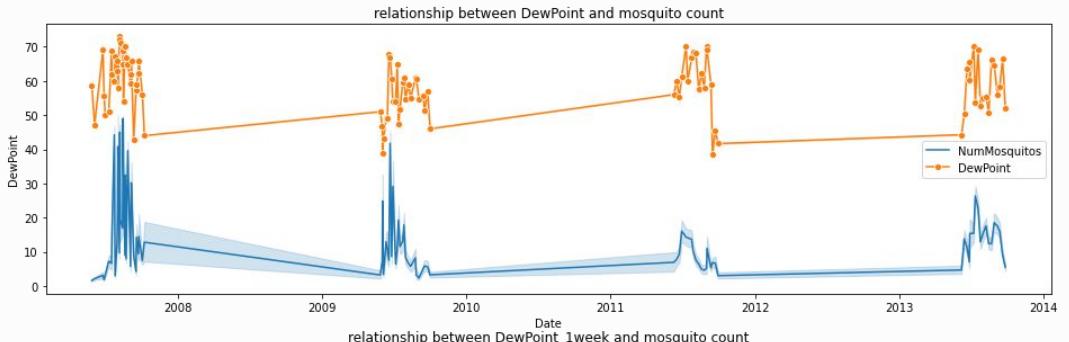
TEMPERATURE

■ Increase in temperature = Increase in Mosquitoes



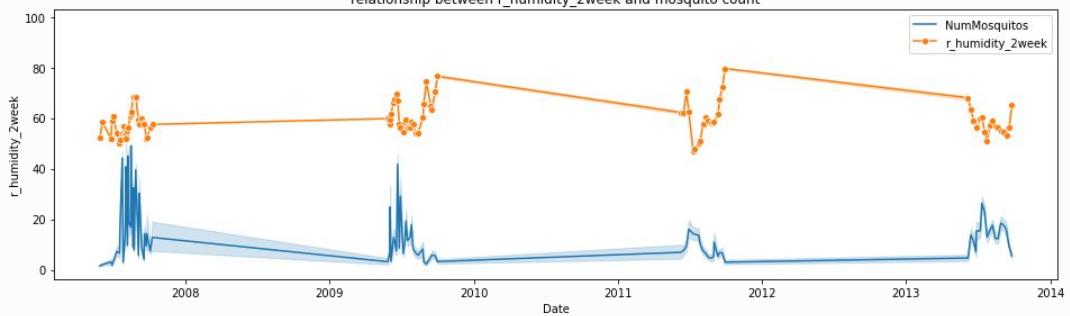
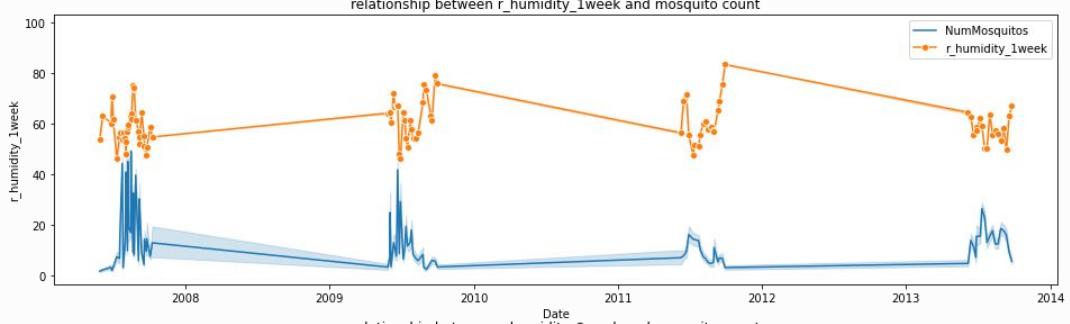
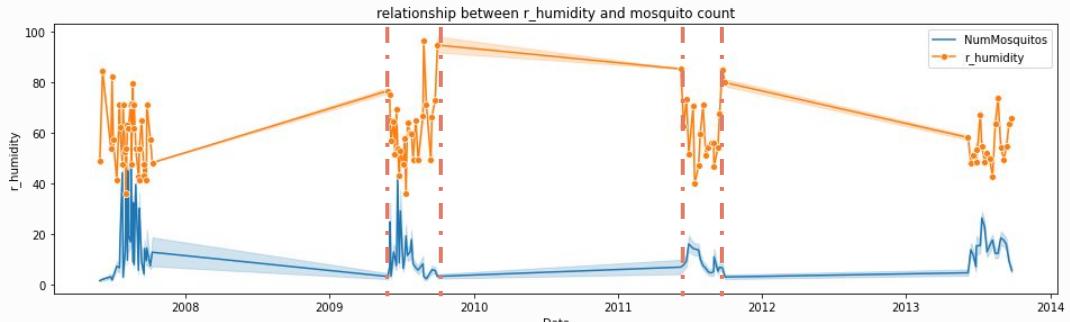
DEWPOINT

Positive relationship



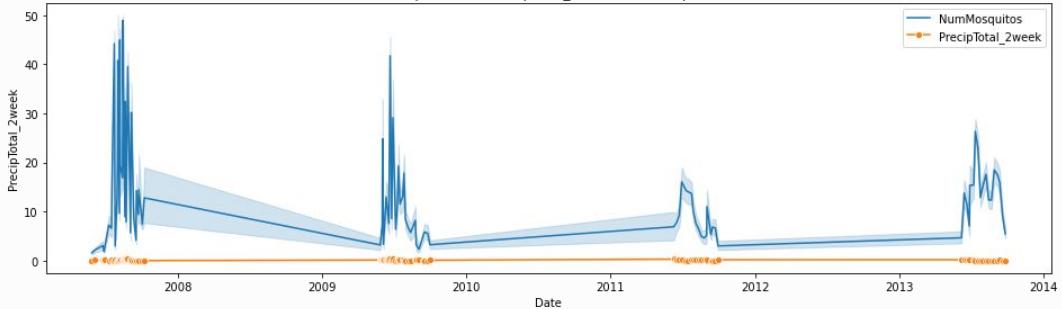
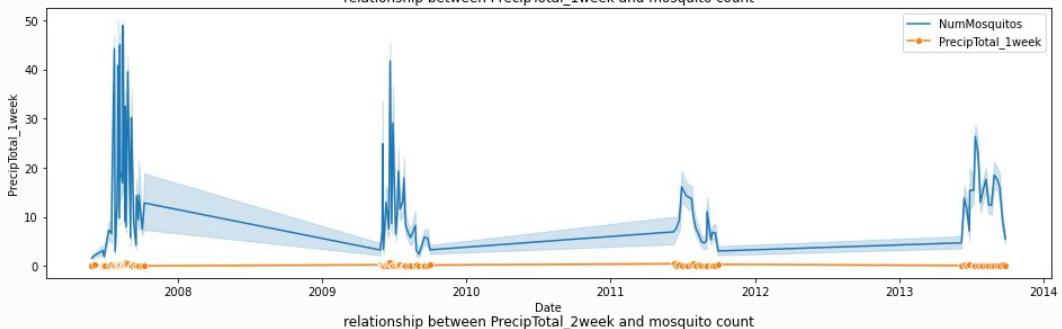
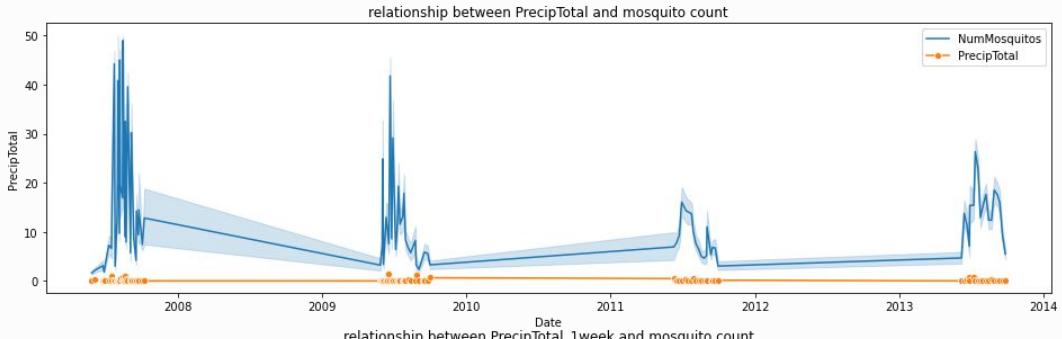
RELATIVE HUMIDITY

Positive relationship



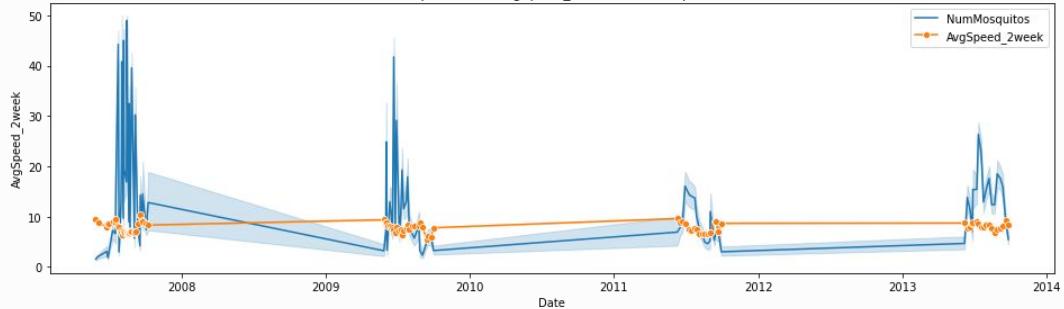
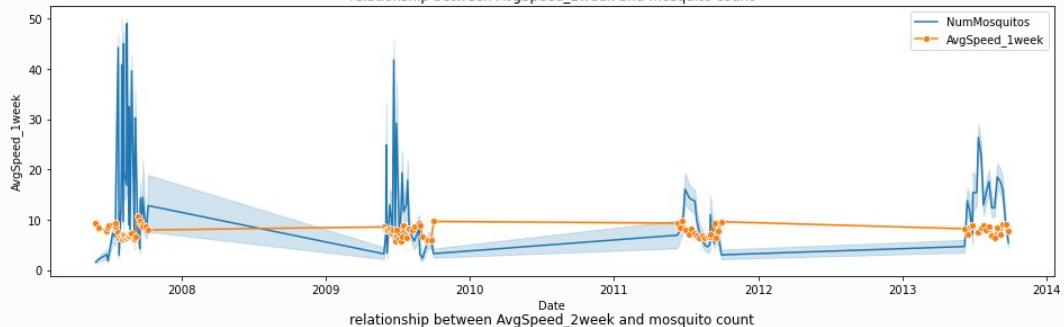
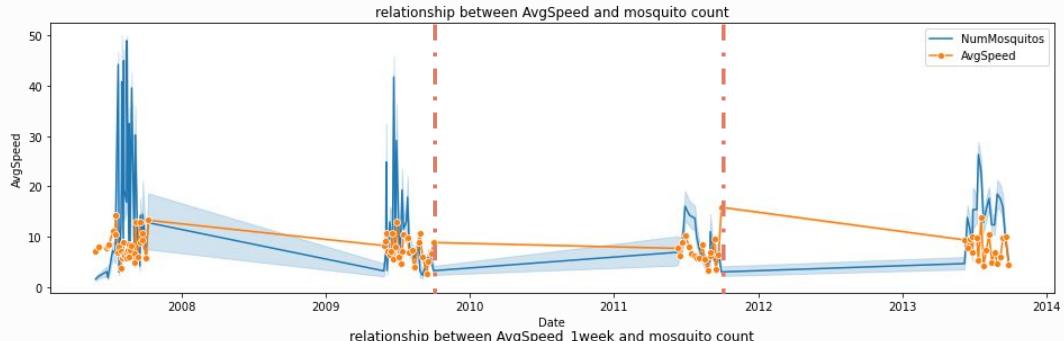
PRECIPITATION

No clear relationship



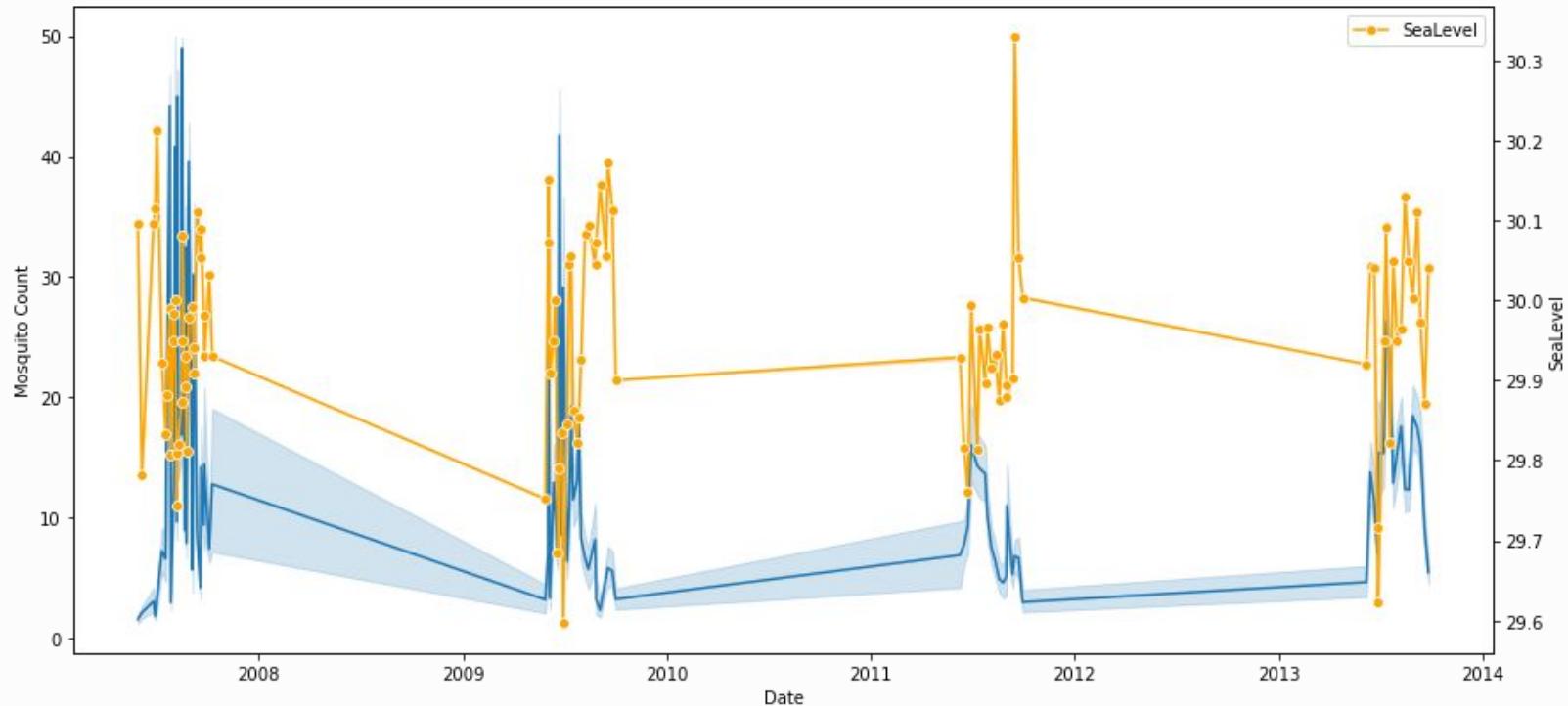
WIND SPEED

Negative relationship



SEALEVEL

■ Decrease in sea level pressure = Increase in Mosquitoes



FEATURES FOR MODELLING

LOCATION & TIME

Latitude, longitude, time

TEMPERATURE

Average temperature, rolling average for temperature, Depart,

SPECIES

Species

WIND

Result direction, average speed, and their rolling averages

WEATHER

Dewpoint, precipitation, relative humidity and their rolling averages

OTHERS

Sea level

MODELING



1 WEEK VS 2 WEEKS ROLLING AVERAGE

Rolling average for:

- Average Temperature
- Dewpoint
- Total Precipitation
- Average Speed
- Relative Humidity

Empirical results show that data with 2 weeks rolling average performed better

It also matches the typical life cycle of mosquito

⇒ Use 2 weeks rolling average

IMBALANCED DATA

Severe imbalanced in dataset

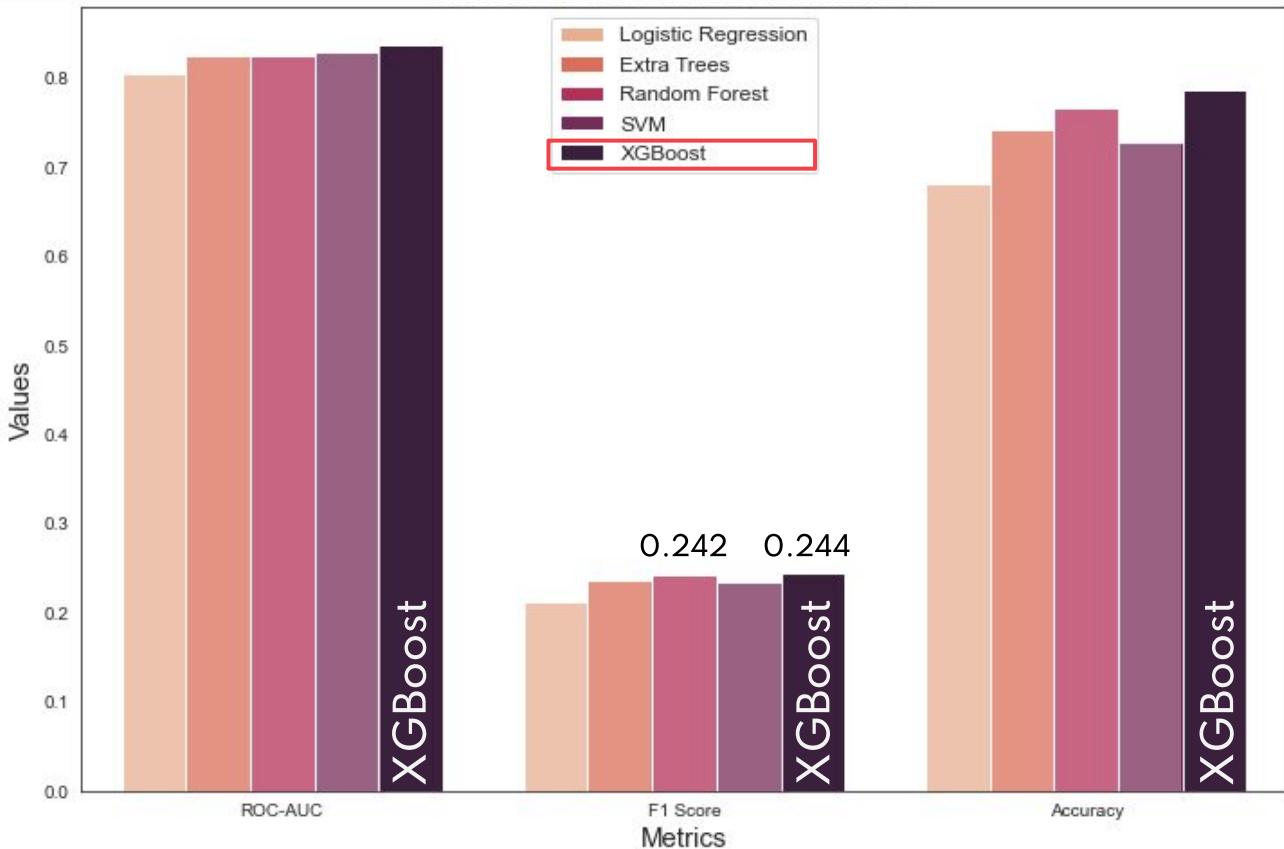
Possible Solutions:

- SMOTETomek
- Class Weight

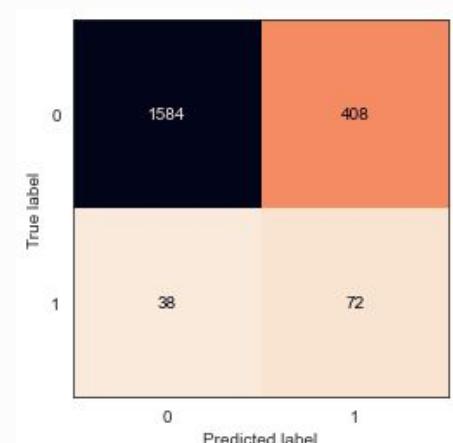
Empirical results show that there is no noticeable difference between the two solutions

⇒ Class Weight was chosen as it is less resource intensive

TUNED MODEL - SCORE METRICS



XGBoost
Confusion Matrix



XGBOOST - FEATURE IMPORTANCE

Most important features

1. Location

- Latitude and longitude

2. Weather

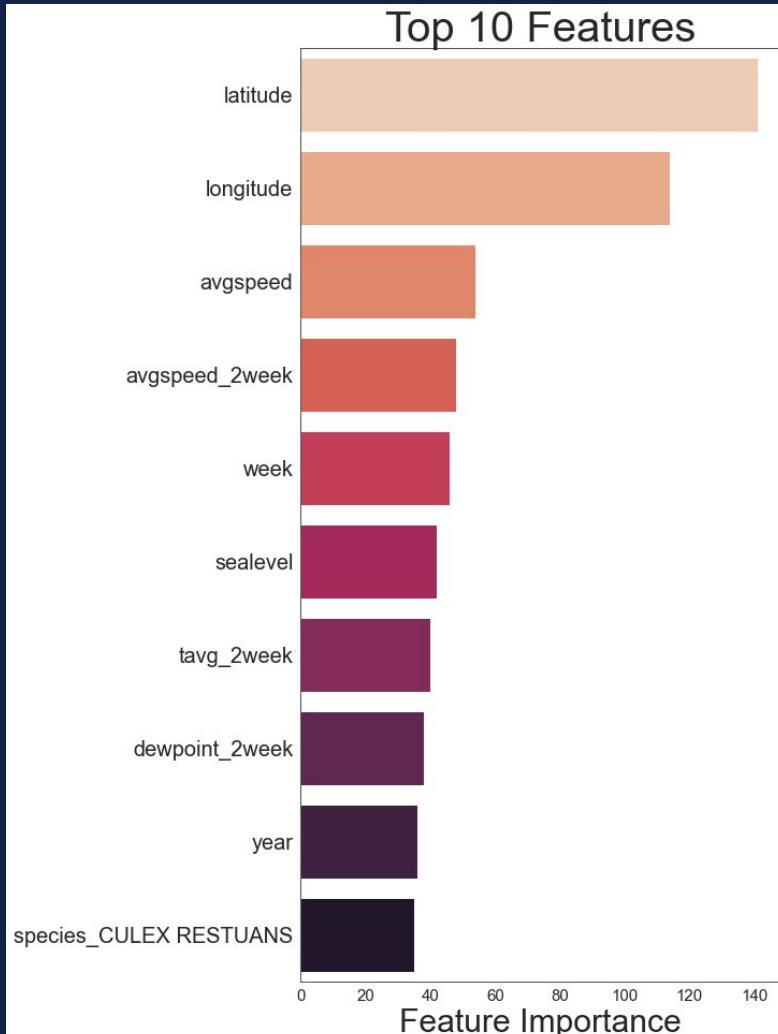
- Average wind speed (day & past 2 weeks)
- Sea level
- Average Temperature (past 2 weeks)
- Average Dewpoint (past 2 weeks)

3. Time

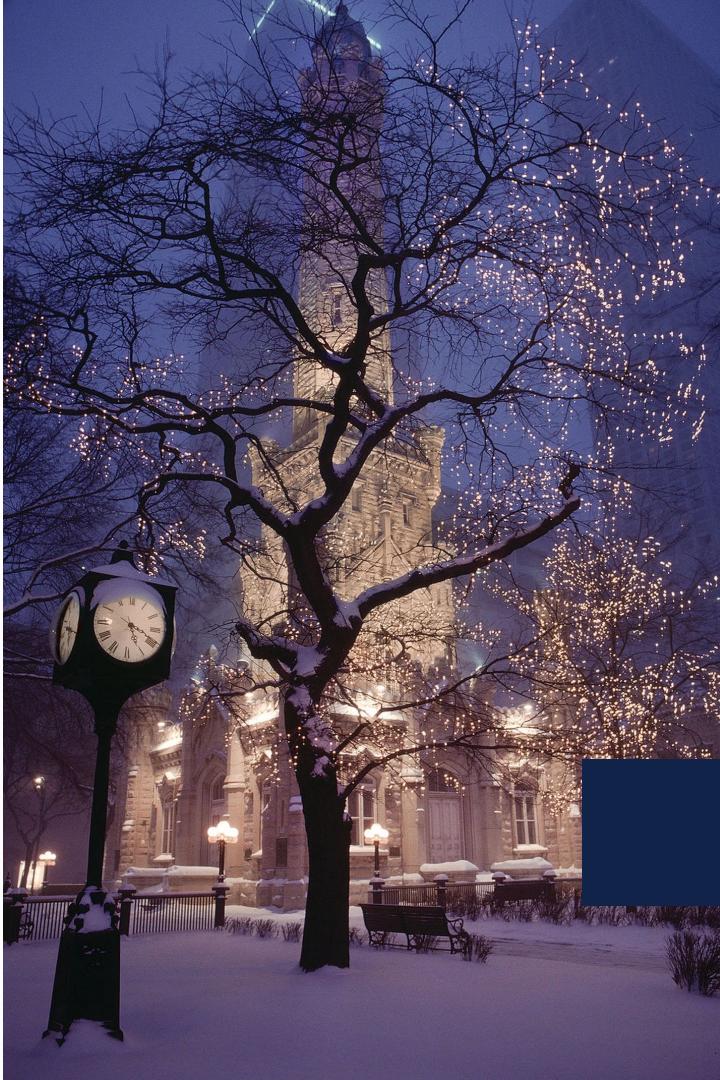
- Week, Year

4. Species

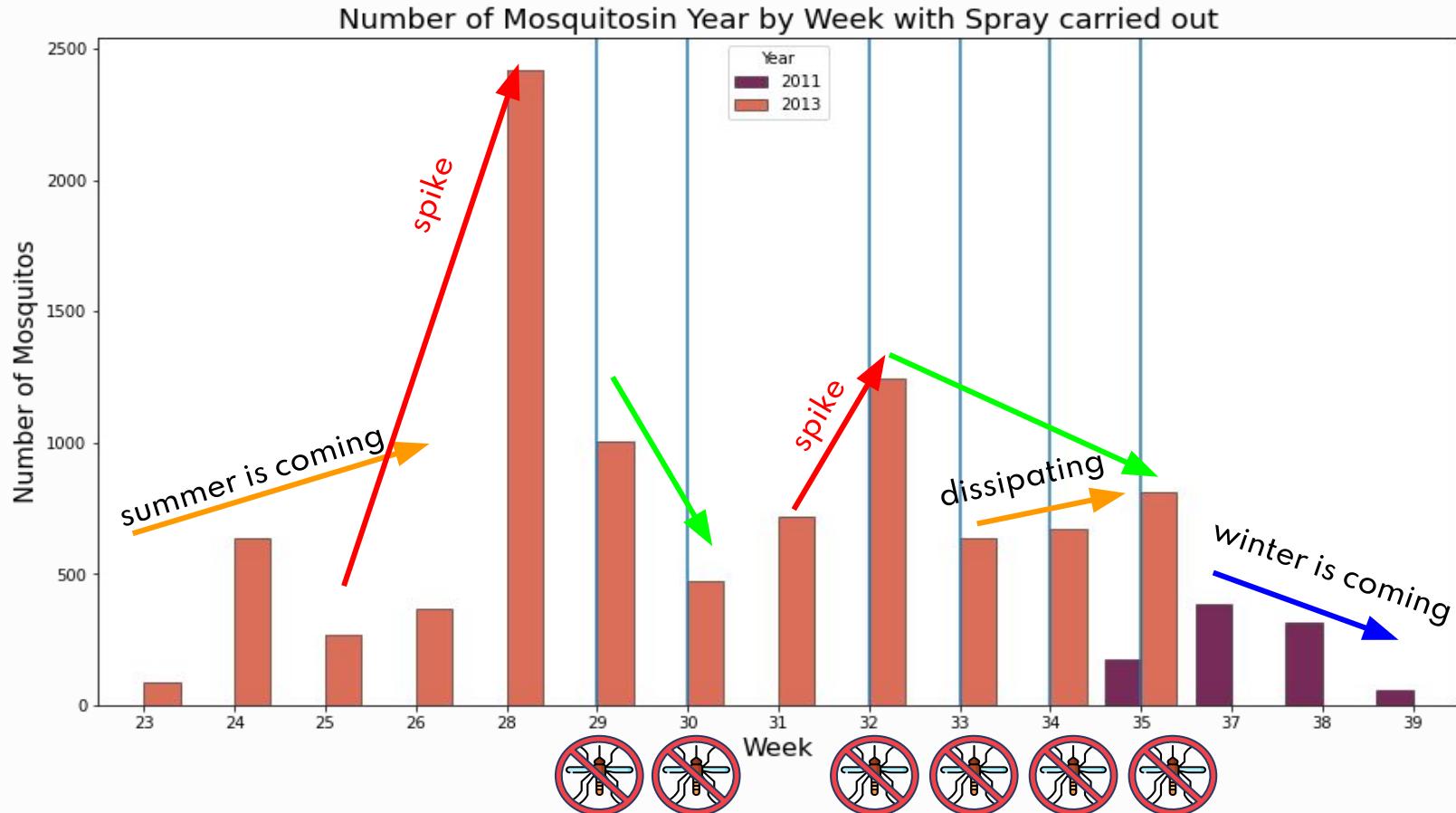
- Culex Restuans



SPRAY EDA & COST-BENEFIT ANALYSIS

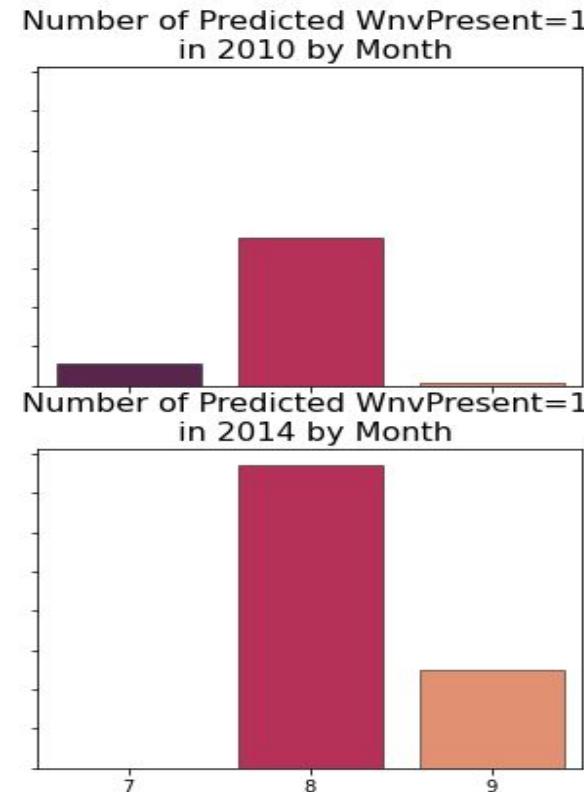
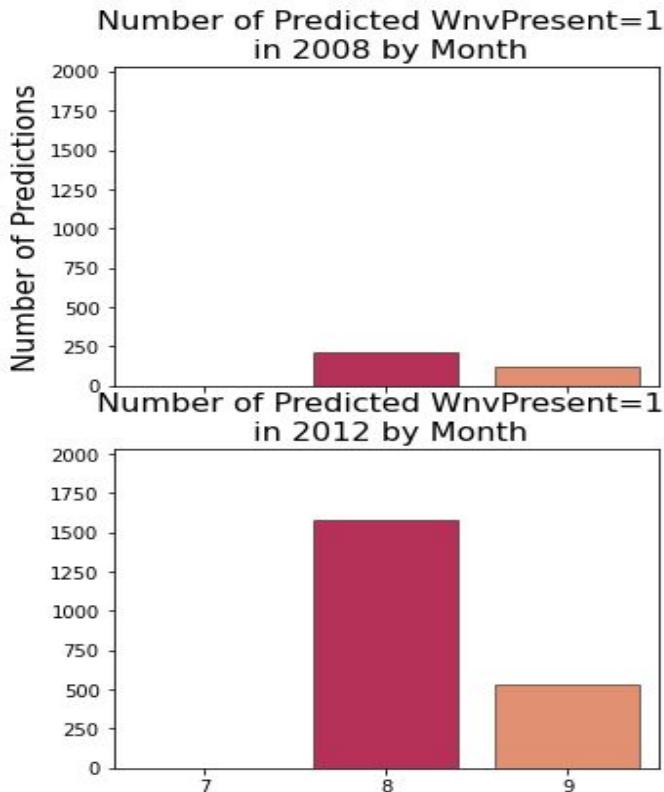
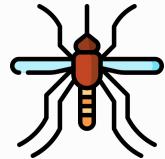


DOES SPRAYING REDUCES NUMBER OF MOSQUITOS?



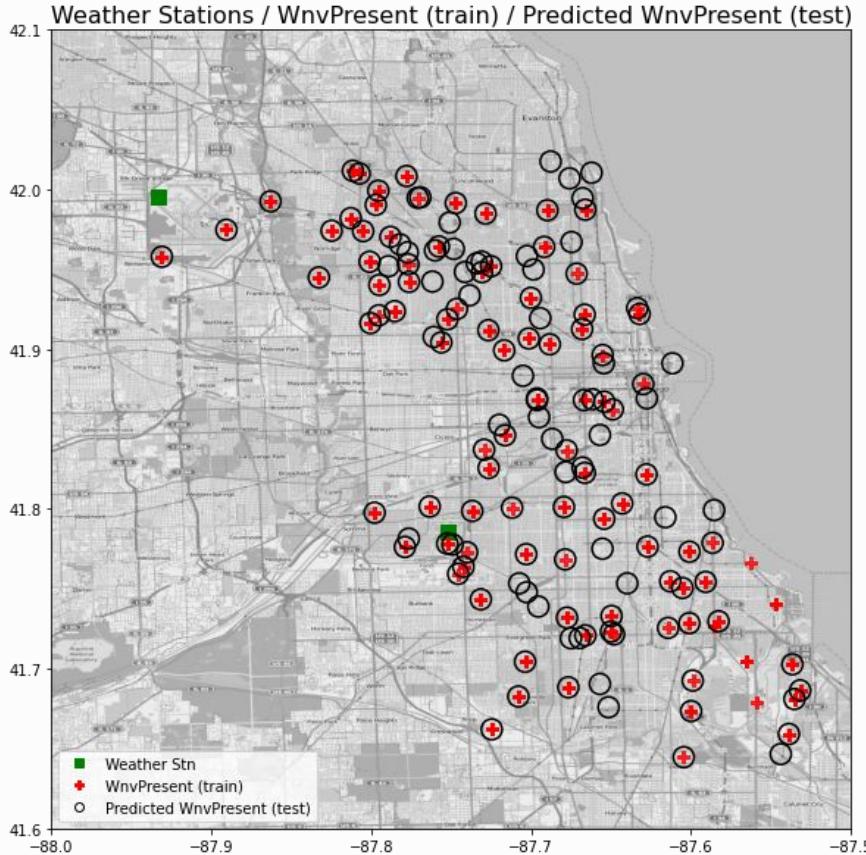
POST MODELING ANALYSIS

WHEN IS THE PEAK SEASON FOR WNVPRESENT?



POST MODELING ANALYSIS

HOW DOES THE WNPRESENT TEST PREDICTIONS COMPARE WITH THE TRAIN DATA?



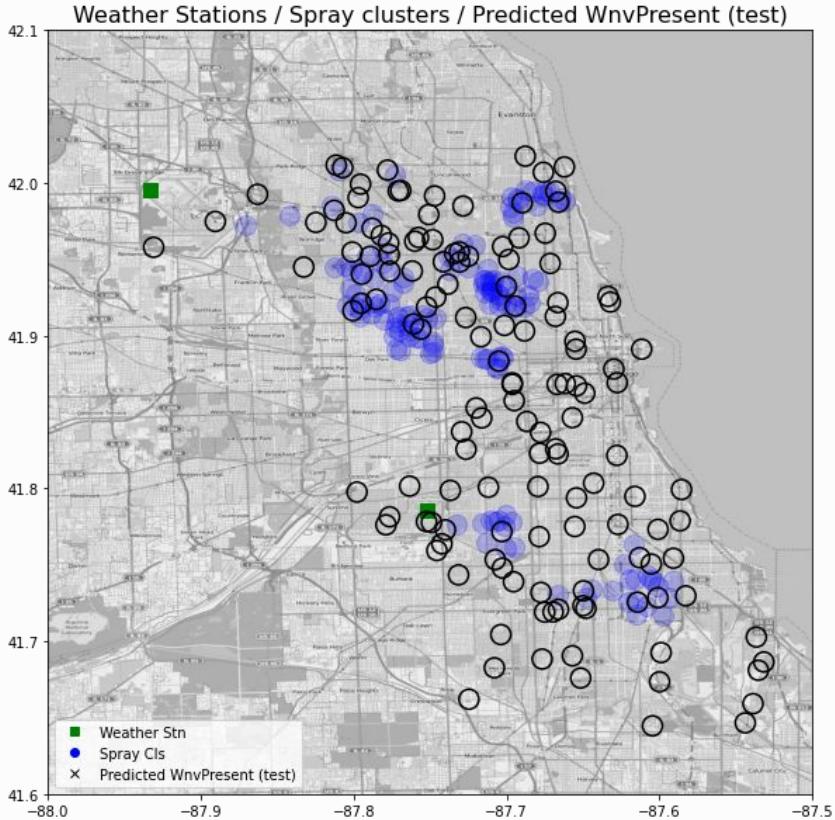
No of unique WnvPresent (train) locations = 99

No of unique WnvPresent (test) locations = 141

- around 10 times more records in test data as compared to train data
- predictions churn out locations that were not in the train data and vice versa
- majority of the predicted WnvPresent locations matched with the actual WnvPresent locations in the train data set

POST MODELING ANALYSIS

DOES CONTINUING WITH THE PREVIOUS SPRAY LOCATIONS WORKS?



No of clustered spray locations (100m range) = 145

No of unique predicted WnvPresent locations = 141

- if we continue with the past spray clusters, it will not be sufficient to cover the predicted WnvPresent locations
- this was also observed with the train data in the EDA section

PRELIMINARY SOCIO-ECONOMIC COST ANALYSIS

AVERAGE 115 CASES / YEAR

1,371 total human WNV cases reported
in Illinois from 2005 to 2016

**PRODUCTIVITY LOSS USD 191
PER WORK DAY MISSED**

Total cost of Productivity loss over 50 days = USD 1,098,250 per year

**TREATMENT COST
USD 33,143 PER CASE**

Total cost of Treatment
USD 3,811,445 per year

**50 WORK DAYS MISSED
PER CASE**

**Total socio-economic loss per year
USD 4,909,695**

PRELIMINARY SPRAY COST ANALYSIS

NO. OF UNIQUE LOCATIONS TO SPRAY	4 MONTHS (JULY TO OCT)	COST PER SPRAY	TOTAL COST OF SPRAY
99	SPRAY 16 WEEKS PER YEAR	USD 26I	USD 413,424
Total cost of spray over 16 wks period per year USD 413,424			

All cost expressed as a percentage of socio-economic cost
 $413,424 / 4,909,695$
(8.4%)

POST MODELING ANALYSIS

WHAT ARE THE OPTIONS AND THEIR COST BENEFITS ANALYSIS?

OPTIONS	SOCIO-ECO COST (USD)	COST OF SPRAY (USD)	%	RISK
OPTION 1 - SPRAY ALL PREDICTED LOCATIONS FOR ALL FOUR MONTHS.		\$588,816	11.99%	LOW
OPTION 2 - SPRAY ALL PREDICTED LOCATIONS BUT ONLY FOR AUG AND SEP.	\$4,909,695	\$294,408	6.00%	MEDIUM
OPTION 3 - SPRAY ONLY PREDICTED LOCATIONS FOR EACH MONTH FOR ALL FOUR MONTHS.		\$285,012	5.81%	MEDIUM
OPTION 4 - SPRAY ONLY PREDICTED LOCATIONS FOR EACH MONTH BUT ONLY FOR AUG AND SEP.		\$263,088	5.36%	HIGH

POST MODELING ANALYSIS

WHAT ARE THE OPTIONS AND THEIR COST BENEFITS ANALYSIS?

OPTIONS	SOCIO-ECO COST (USD)	COST OF SPRAY (USD)	%	RISK
OPTION 1 - SPRAY ALL PREDICTED LOCATIONS FOR ALL FOUR MONTHS.	\$4,909,695	\$588,816	11.99%	LOW

- Spray all predicted locations for all four months is predicted to give the maximum benefits.
- Maximizes the coverage of all predicted WnvPresent locations.
- By covering all predicted locations, lower the risk of mosquitos changing their breeding hotspots to locations that are not spray.
- Ensure consistent and continuous application of sprays that had proven effect of keeping mosquitos population in control.

POST MODELING ANALYSIS

WHAT ARE THE OPTIONS AND THEIR COST BENEFITS ANALYSIS?

OPTIONS	SOCIO-ECO COST (USD)	COST OF SPRAY (USD)	%	RISK
OPTION 4 - SPRAY ONLY PREDICTED LOCATIONS FOR EACH MONTH BUT ONLY FOR AUG AND SEP.	\$4,909,695	\$263,088	5.36%	HIGH

- If cost is a major concern to City of Chicago and CPHD, Option 4 is the cheapest but also comes with the highest risk.

The background image shows a wide-angle aerial view of the Chicago city skyline. The buildings are silhouetted against a bright blue sky filled with large, white, fluffy clouds. In the foreground, the calm, turquoise-colored waters of Lake Michigan are visible, with a few small boats scattered across the surface.

THANK YOU

Let's Work Towards a WNV Free Tomorrow!

The background features a complex, three-dimensional grid of red hexagons on a dark blue surface. The perspective is from below, looking up at the grid, which creates a sense of depth and motion.

BACKUP SLIDES

Table 2

Annual human WNV cases, average seasonal mosquito infection rate (MIR), and mosquito testing from 2005 to 2016 in Cook and DuPage counties.

Year	Number of human cases	Average MIR	Number of pools tested	Number of positive pools	Total number of mosquitoes tested
2005	181	5.33	7,165	1,939	271,235
2006	129	5.35	9,428	1,984	318,386
2007	43	2.65	12,131	1,259	375,520
2008	10	1.91	9,024	587	298,995
2009	1	1.14	9,450	298	311,220
2010	47	5.19	11,491	2,086	393,279
2011	24	3.10	8,911	939	287,774
2012	229	7.35	10,162	3,182	323,497
2013	66	4.26	11,078	1,967	407,326
2014	31	2.97	9,273	990	333,489
2015	36	3.57	7,725	1,046	314,363
2016	108	6.34	6,144	1,687	219,909

Table 2

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

Item	Cost per case†	No. cases to which cost applies‡	% Cases to which cost applies§	Total cost for all cases	Total cost if treatment/service were used in all cases			
Inpatient treatment costs	\$33,143	46	100	\$1,524,570	\$1,524,570			
Table 3								
Estimated economic costs of WNND cases due to productivity loss, Sacramento County, California, 2005*								
Productivity loss	Value of day missed†	Value of nonwork day missed‡	No. work days missed	No. nonwork days missed	No. patients <60	No. patients ≥60	% Cases	Total costs for all cases
For patients <60 y	\$191	\$125	50	10	31	100	100	\$334,800

Verification Required

Project: Bug & Insect Control

Before we connect you to your service pros we need to verify your contact information. A customer care representative will follow up with you to verify your information.

Want to speed up the verification process?

Call (888) 846-0966 or

[Start Live Chat](#)

Need Help? Call 1-888-846-0966 (more)

HIRE AN INSECT CONTROL SERVICE

CHANGE VIEW

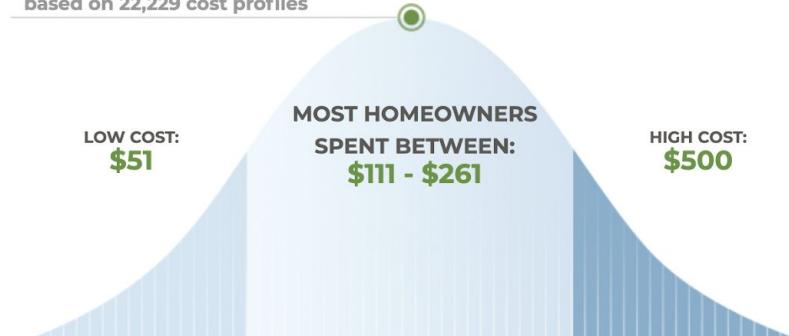


LOCATION: CHICAGO, ILLINOIS

AVERAGE REPORTED COSTS:

\$177

based on 22,229 cost profiles



Cost data is based on actual project costs as reported by HomeAdvisor members.

How do we get this data?