Mosquito Control:

West Nile Virus

Group 2 (Edmund, Marc, Mingzi, Rohazeanti)
SG DSIF5

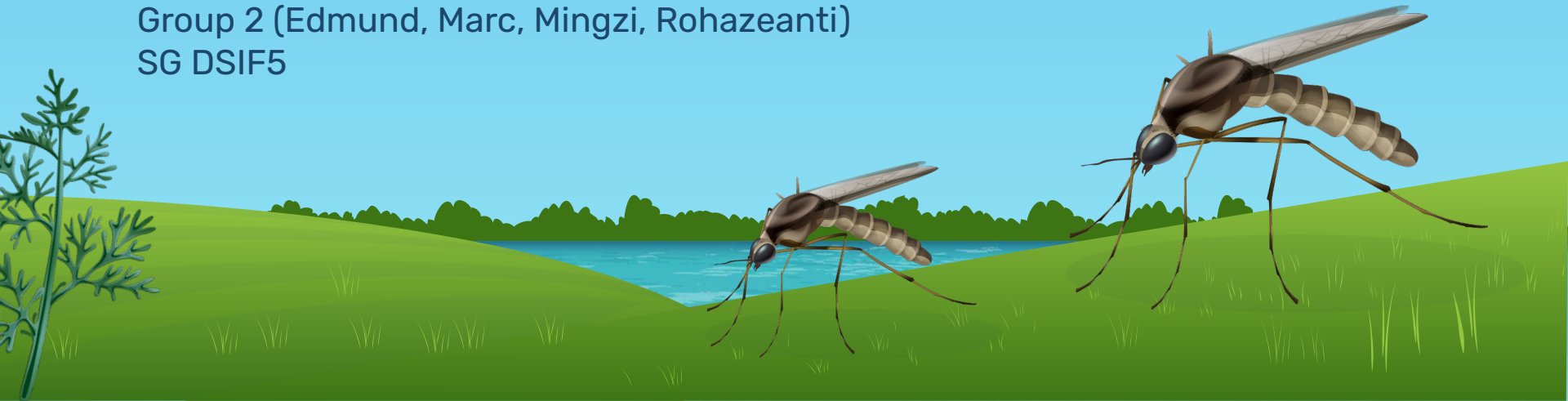
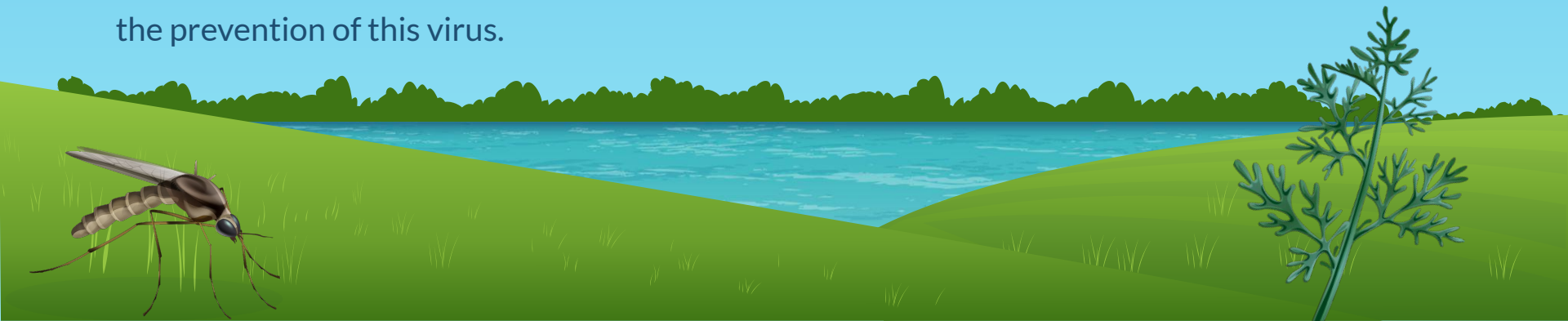


Table of Contents

Problem Statement

West Nile virus (WNV) is the leading cause of mosquito-borne disease in the continental United States. It is most commonly spread to people by the bite of an infected mosquito. Cases of WNV occur during mosquito season, which starts in the summer and continues through fall. There are no vaccines to prevent or medications to treat WNV in people.

In this project, we are provided with mosquito trapping, weather, location and fumigation datasets to predict WNV outbreaks in the City of Chicago for effective resources spending on the prevention of this virus.



Introduction

Viral

West Nile Virus (WNV) as leading cause of mosquito borne disease in the continental United States

Asymptomatic

Usually asymptomatic, but 1 in 150 develop serious symptoms that may be sometimes fatal

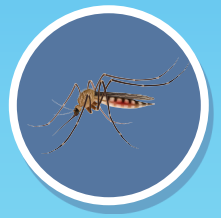
Trapping

Chicago has a surveillance programme where mosquito traps are set up across the city from May to Oct

Test lab

Capture mosquitos and test for presence of WNV

Datasets



Training

Trap locations and Date
Species and WNV presence

Weather

Meteorological Conditions

Spray

Spray location, date & time

Data Cleaning

Train dataset

- Noted Class imbalance with only 5% West Nile Virus occurrences
- Combine mosquito cap records together

Test dataset

- Additional traps not present in Train dataset
- New Mosquito Species not sampled in Train dataset

Weather dataset

- Dropped irrelevant columns
- Filled missing values with mean values for each feature

Exploratory Data Analysis

Spray

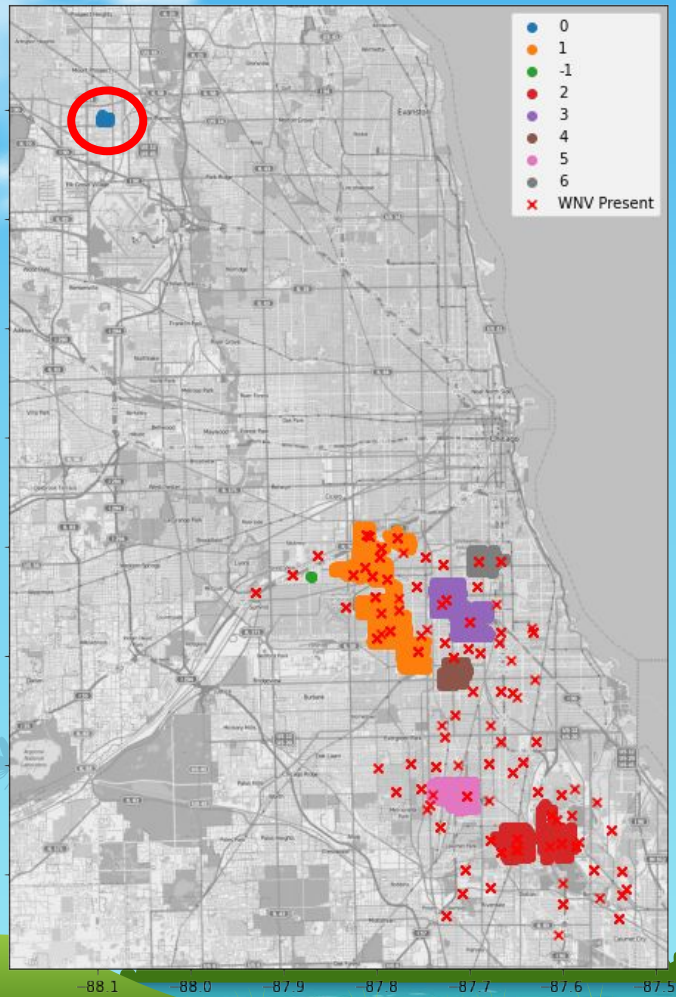
- Overall spray location
- Spray location vs WNV location

Weather

- Stations Meteorological Data
- Weather effects on Mosquitos

Virus and Traps

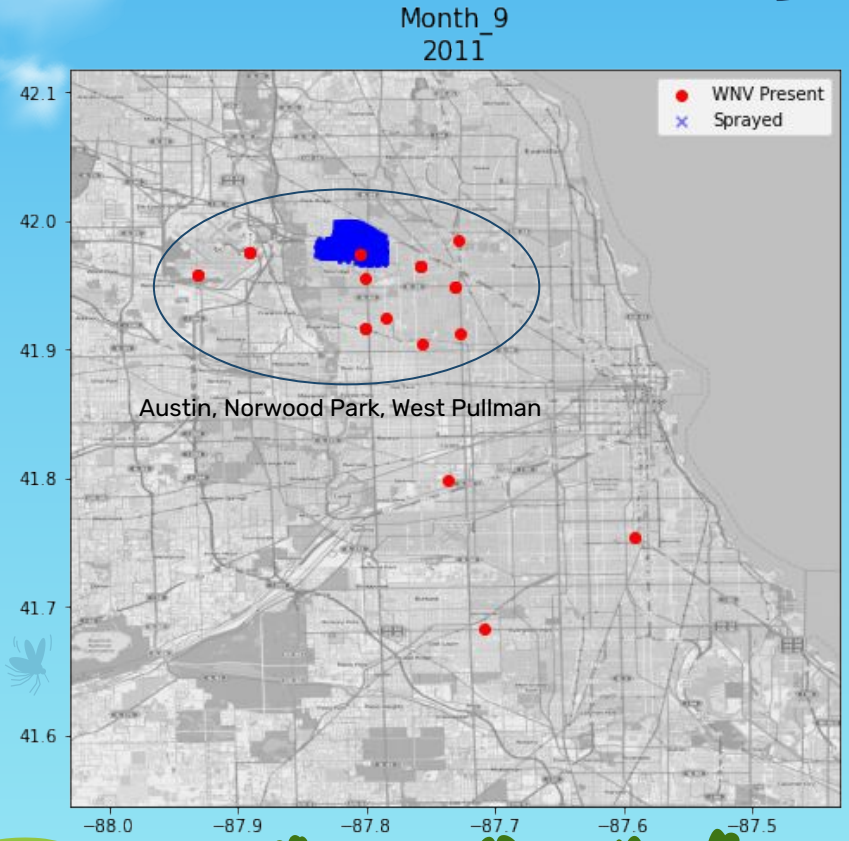
- Data collection heat map
- Interactive Plotly charts
- Species analysis
- WNV presence (Year/Month)
- Spray location vs WNV location

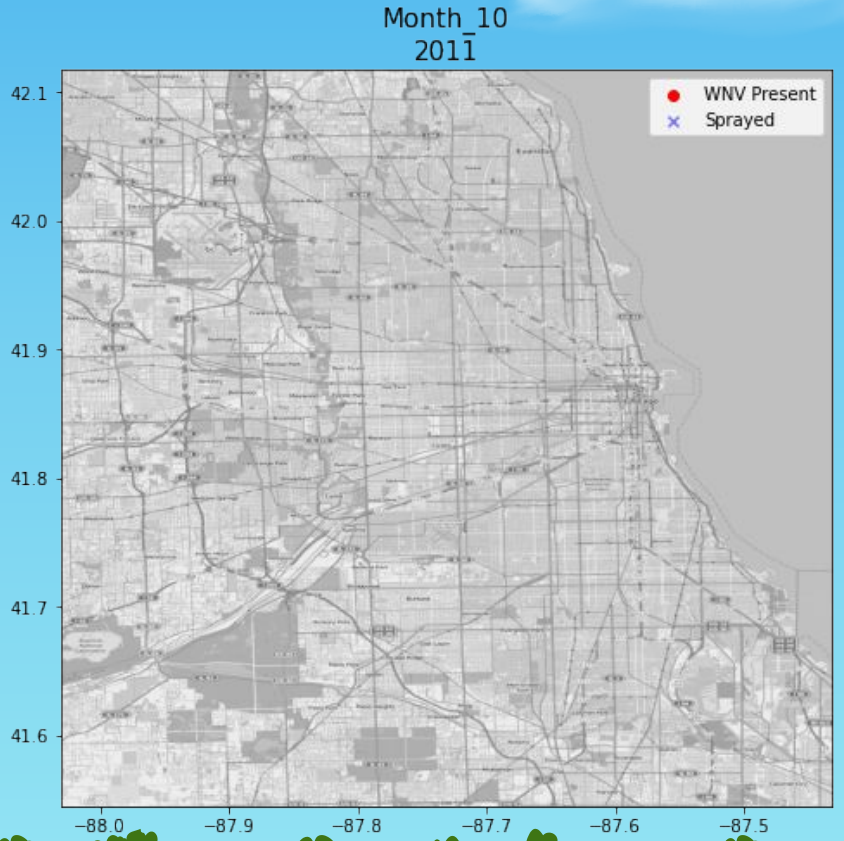


Spray Data EDA

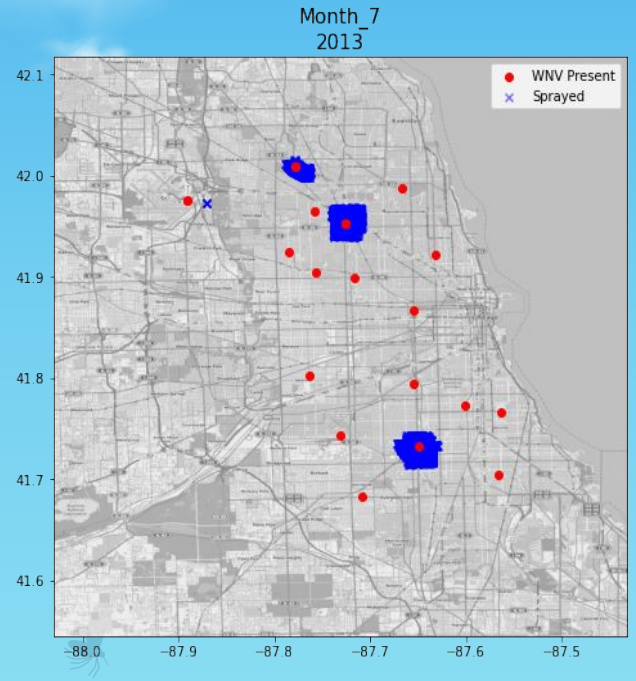
- Only exist data for 2011 and 2013
- July - September highest prevalence of WNV
- Spraying occured in the evenings between 7pm to 9pm
- Random spray cluster at High Ridge Knolls Park (August, 2011)

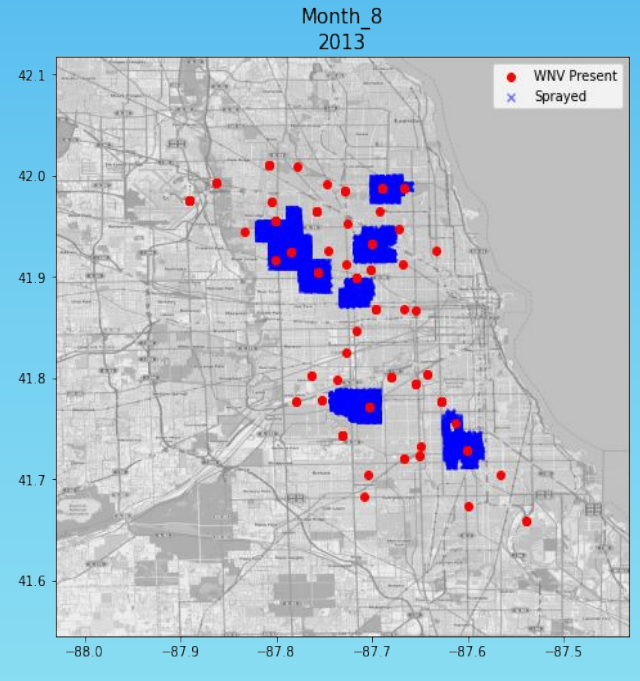
Spray Data EDA

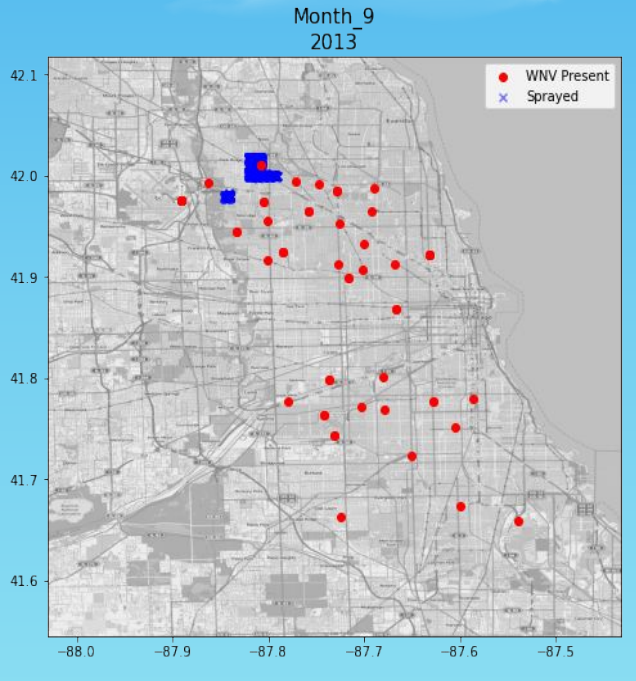




Spray Data EDA

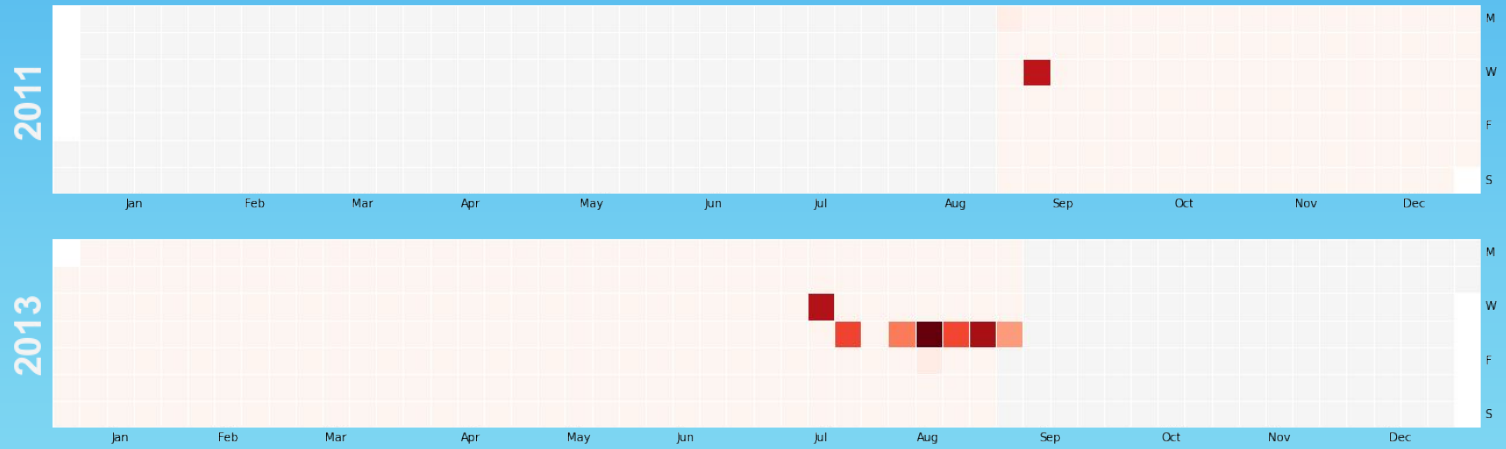






No sufficient evidence to support whether spraying of pesticide is effective thus cannot conclude that the spray data can help to predict WNV presence.

Spray Dataset EDA

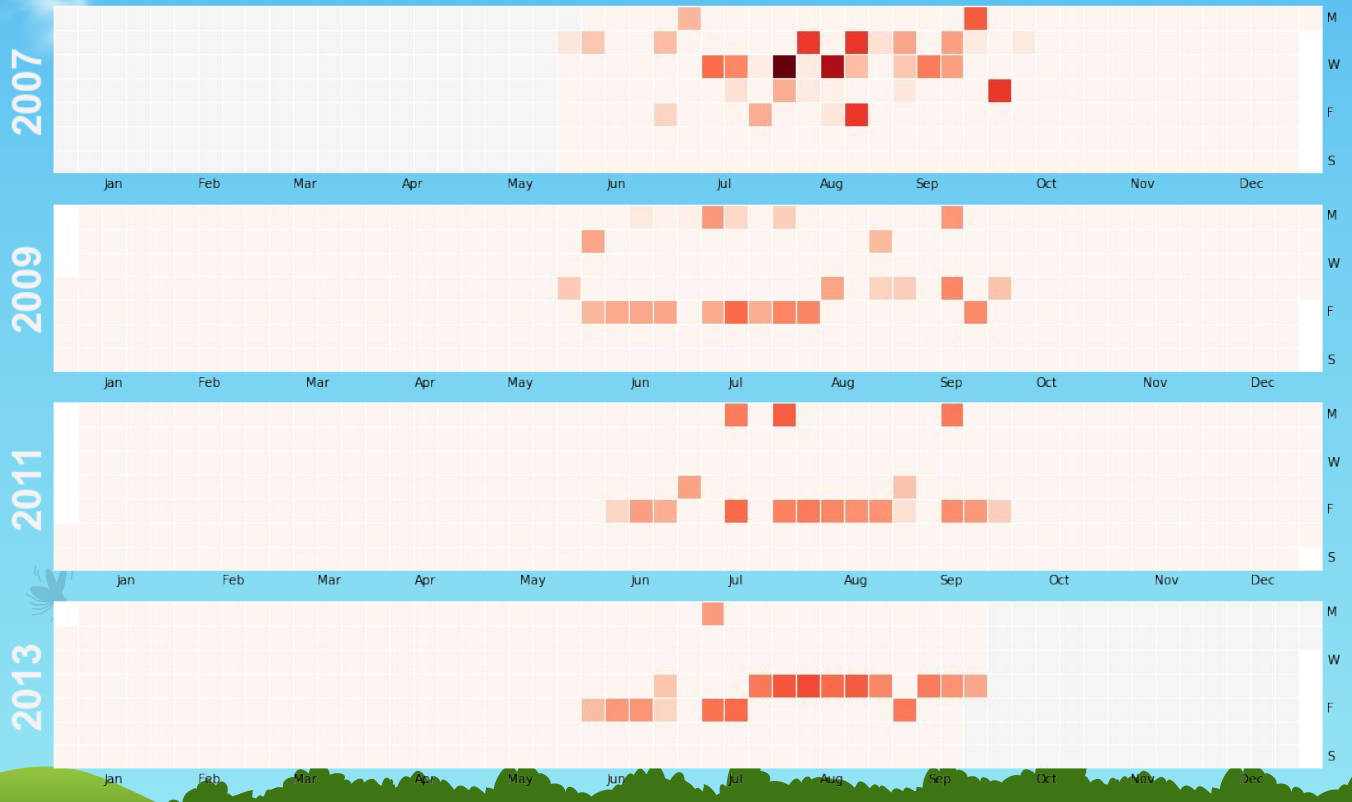


We have only 10 days to data available for mosquito spray. Out of those days only one day sprayed in 2011 and rest of the 9 days are in 2013. There is no data available in other years.

We will not use spray data for modelling because:

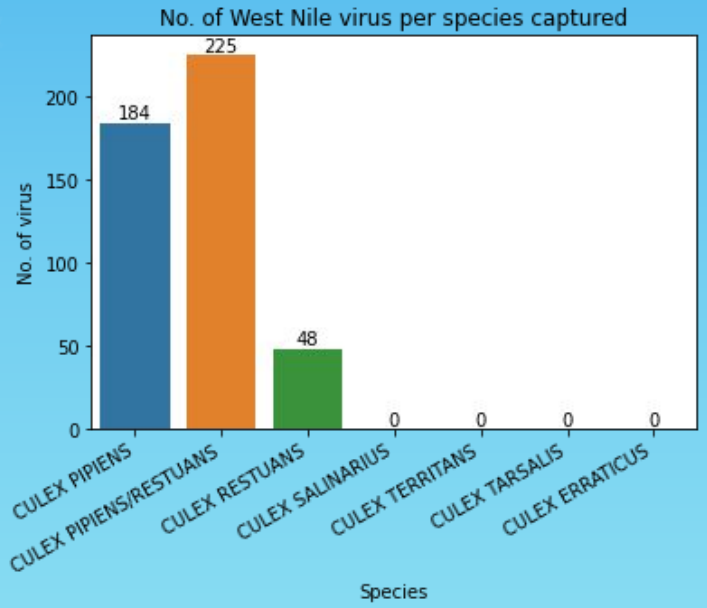
1. Spray data is only available for 2011 and 2013
2. No data is available in the time frame where we need to predict the presence of WNV
3. May not be useful to predict WNV presence

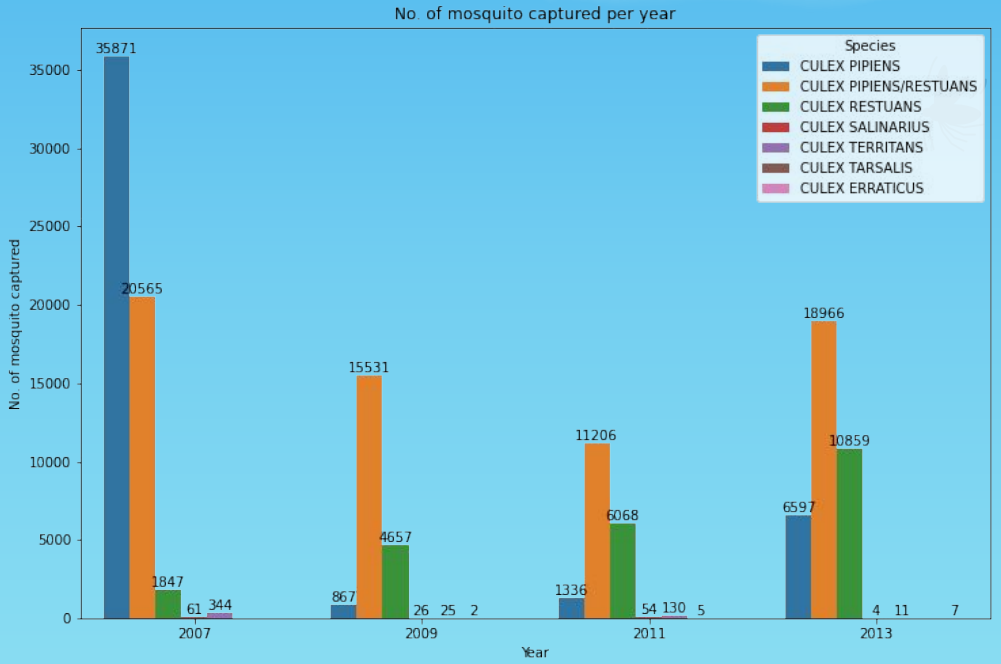
Train Dataset EDA



- Majority of the data points noted to be collected for periods starting in mid May to late Sep
- “Warm Season” in Chicago runs from early June to late Sep. Hottest month of the year being July

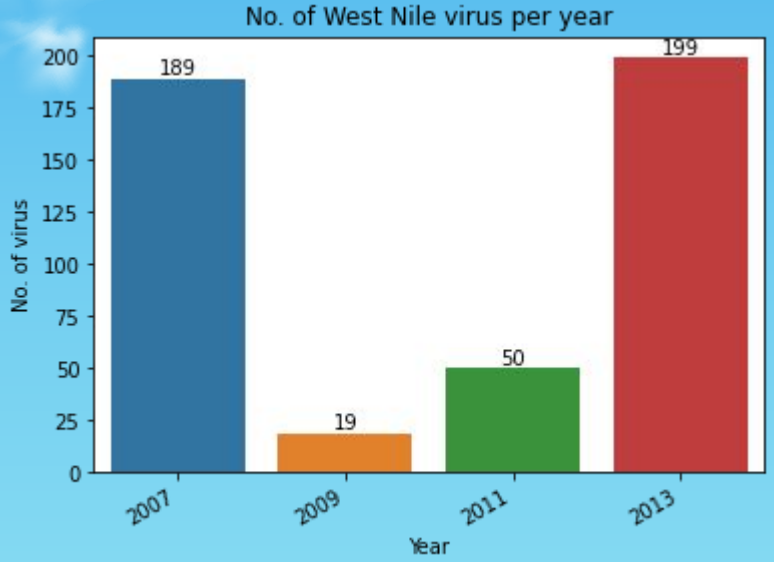
Train Dataset EDA

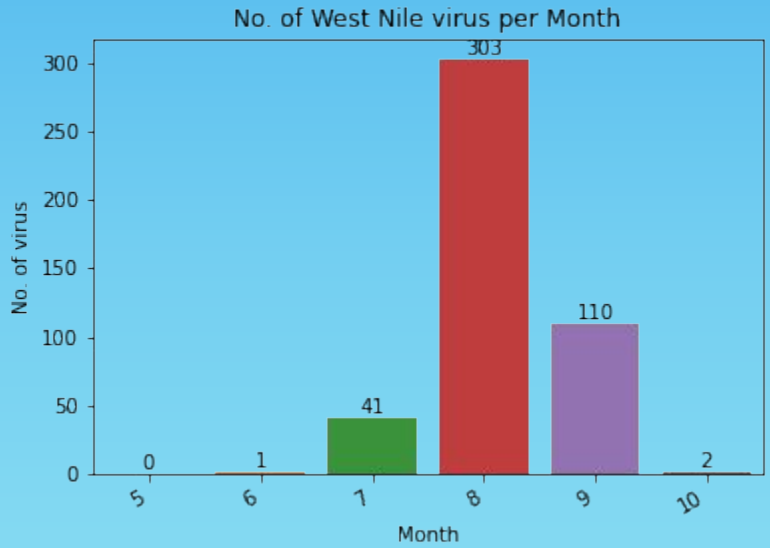




It was identified that there were mainly 2 main species that spread WNV which is namely: Culex Pipiens and Culex Restuans

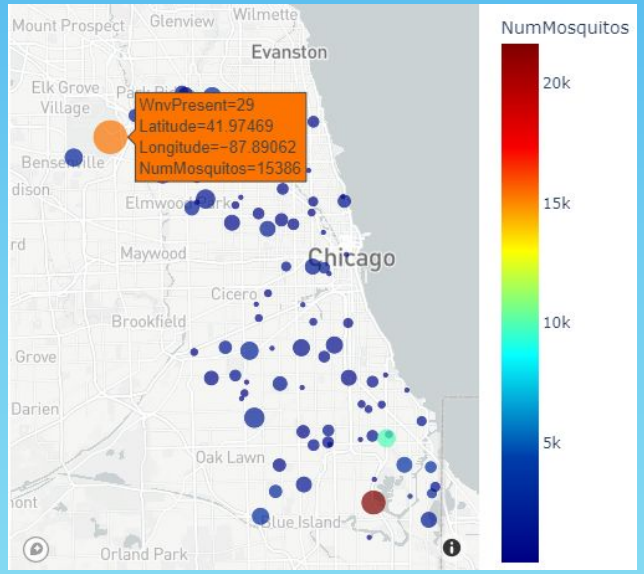
Train Dataset EDA





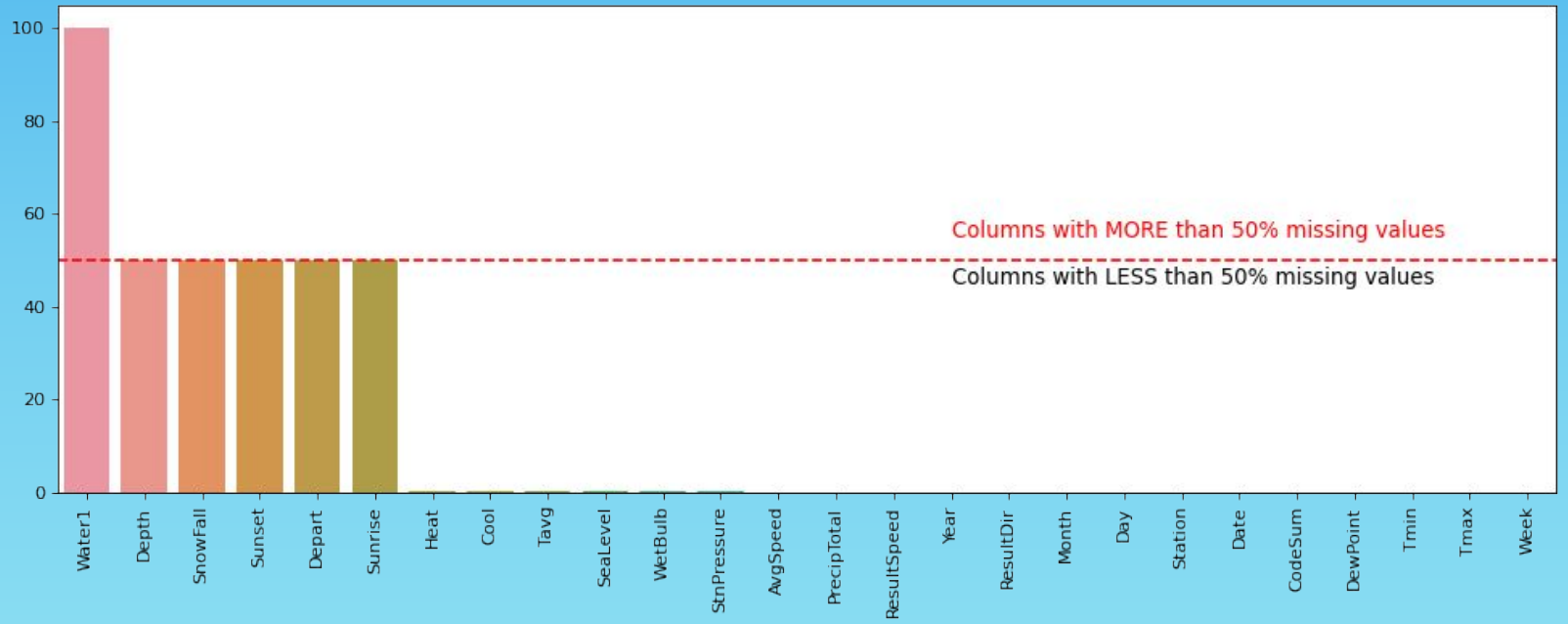
As there were more WNV species mosquito captured during 2007 and 2013, it is also expected that there are higher presence of mosquito with WNV virus captured as evidently shown from the chart. The peak for the virus is during Summertime which has an increase in mosquito population due to a higher rainfall.

Train Dataset EDA



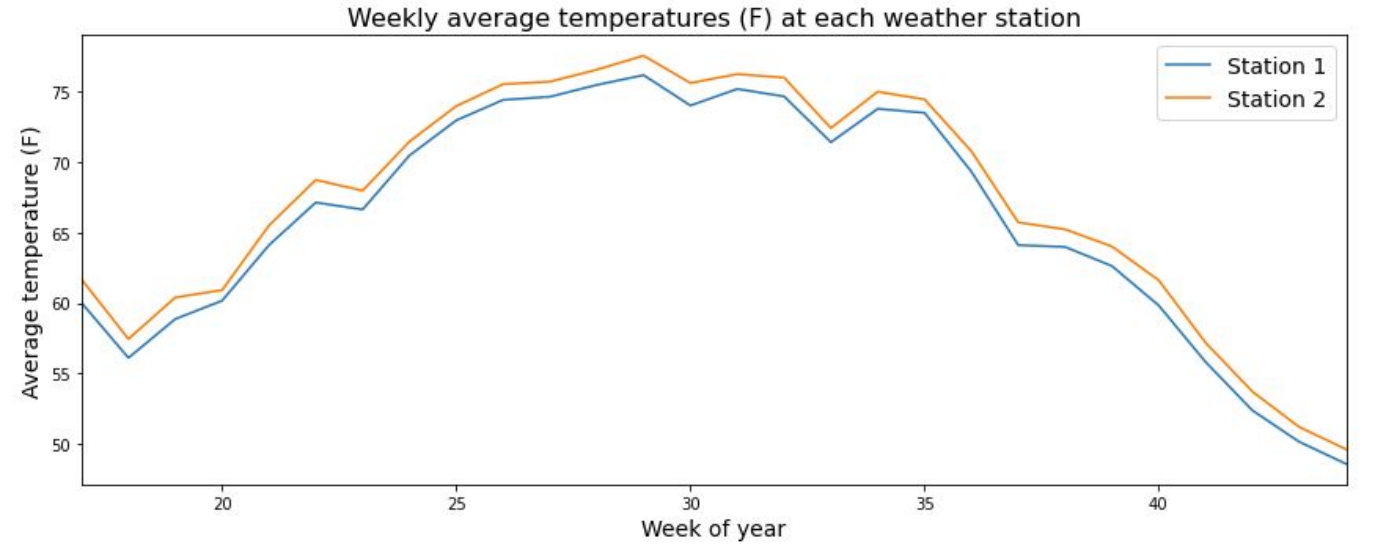
WNV mosquito vs
Total no. of mosquitos

Weather Dataset EDA



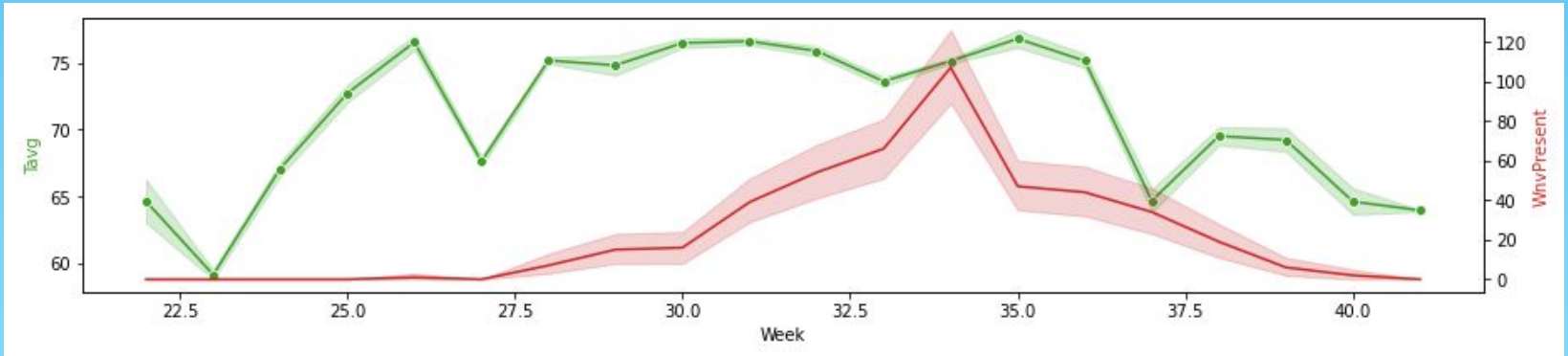
Noted columns above with missing values of > 50%. We have decided to drop them except for Sunset & Sunrise (explained later)

Weather Dataset EDA



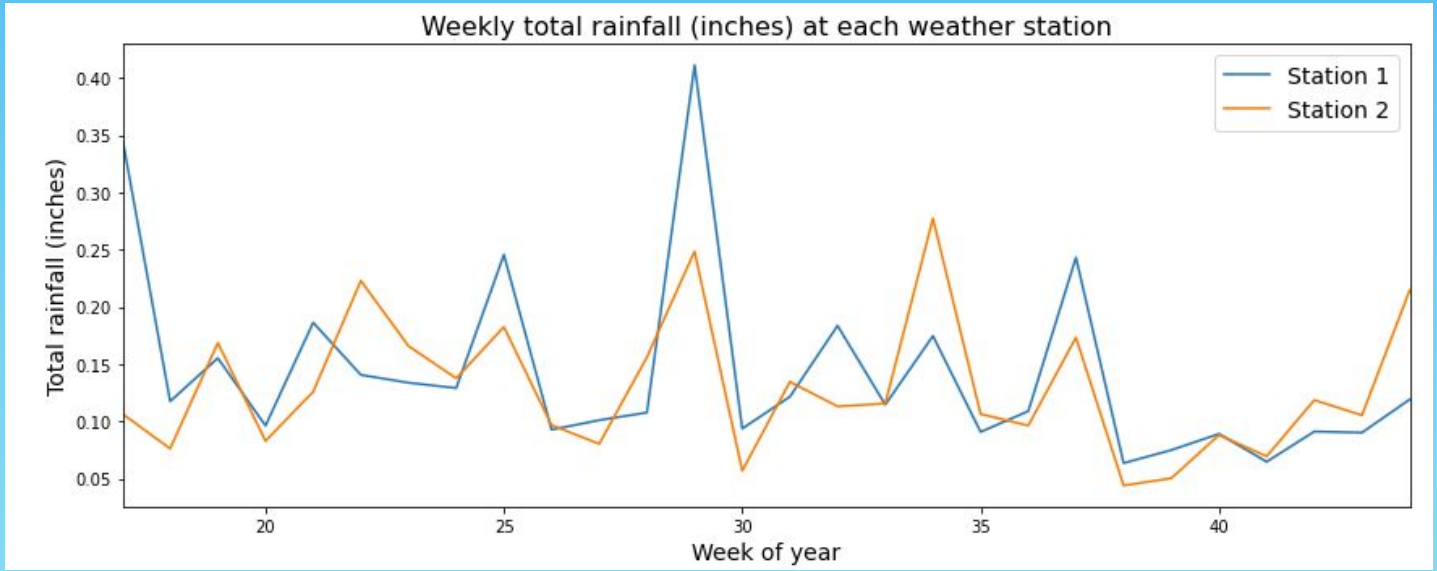
Station 1 generally cooler temperatures than station 2 throughout the year

Weather Dataset EDA



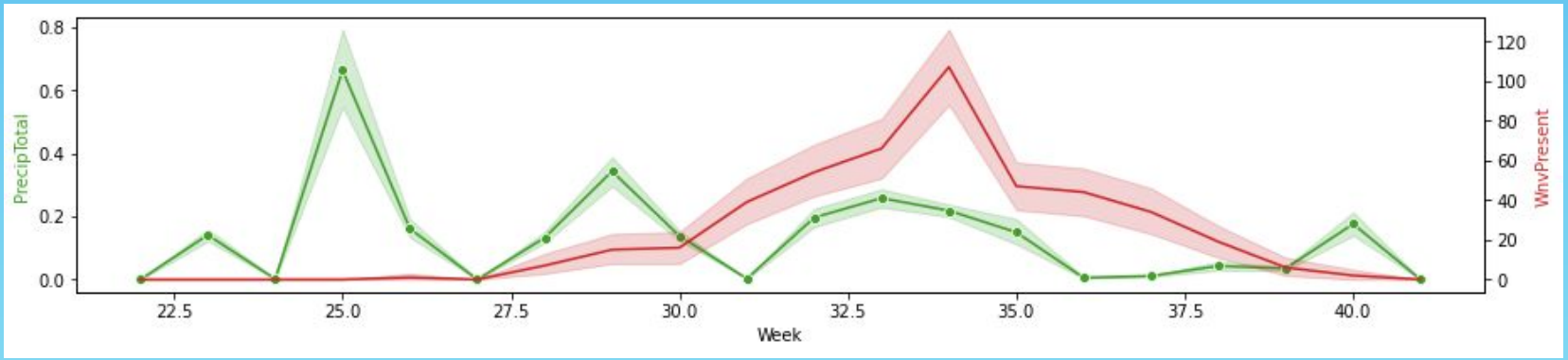
Mosquitoes becomes active and begin breeding when temperature are consistently 50 Fahrenheit or higher.

Weather Dataset EDA



Erratic rainfall noted for both stations over the weeks of the year

Weather Dataset EDA



Flooding rains create ideal breeding conditions for mosquitoes.

Feature Engineering

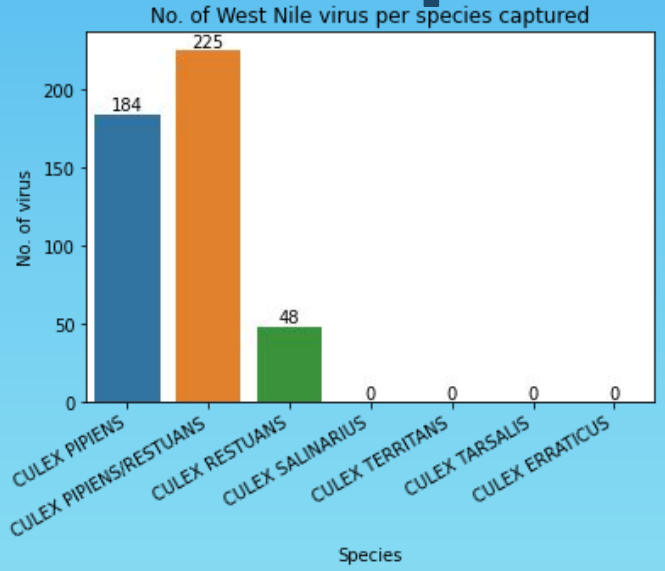
Train Dataset

- Mosquito Species
- Weather conditions of Traps
- Month, Week

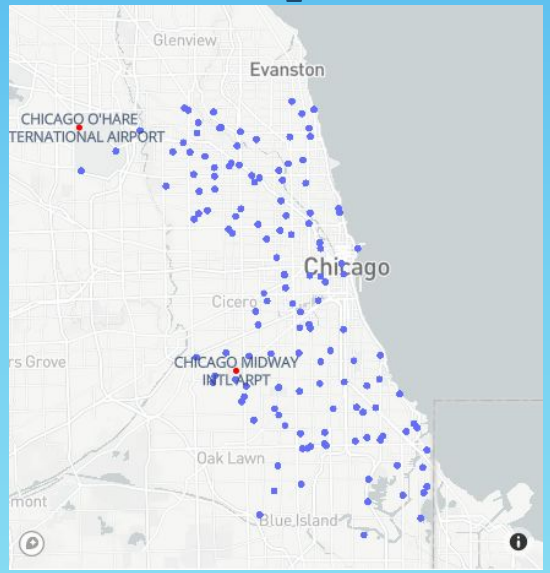
Weather

- Sunset/Sunrise and Sun Hours
- Wet Dry feature from CodeSum
- Relative Humidity
- Temperature range
- Delayed weather measurements (Time lag)

Feature Engineering: Mosquito Species and Traps



Map Non-WNV species as Others
Performed categorical encoding on Species



Map weather station to each trap according to proximity

Feature Engineering: Weather

Sunrise/Sunset , Sunhours

Culex mosquitoes are highly active at dusk and dawn. Sunhours feature can be an important feature in modelling

Wet Dry from CodeSum

Simplify the CodeSum feature from different weather conditions to just Wet or Dry condition

Relative humidity and Temperature range

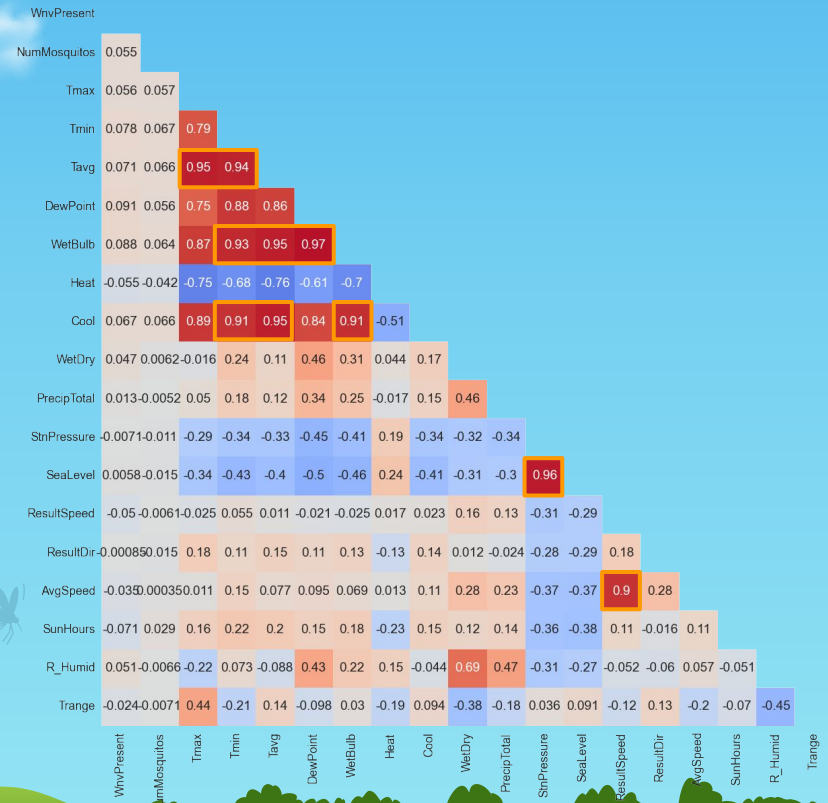
Humidity increases mosquito activities therefore increases the likelihood of getting bitten

Delayed weather features (7 days)

Time lag weather features created to account for the development from egg to adult mosquitoes.

Feature Selection: Multicollinearity

Pearson Correlation Heatmap



Variance Inflation Factor

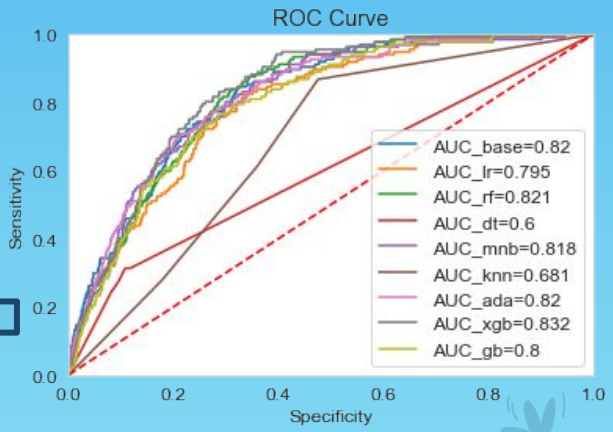
Features with correlation > 0.90 were identified

With the pairs identified, the VIF of these features were calculated and the features with VIF > 10 were manually removed one at a time

Tmin	5.94
DewPoint	5.93
AvgSpeed	5.90
ResultSpeed	5.76
Tmax	2.84
StnPressure	1.53

Model Evaluation

	Train score	Test score	Generalisation	Accuracy	Precision	Recall	Specificity	F1	ROC AUC
Logistic Regression (no SMOTE)	0.946	0.947	-0.106	0.947	0.500	0.015	0.999	0.029	0.8203
Logistic Regression	0.957	0.928	3.030	0.928	0.253	0.182	0.970	0.212	0.7949
Random Forest Classification	0.951	0.866	8.938	0.866	0.171	0.394	0.892	0.238	0.8208
Decision Tree Classification	0.985	0.878	10.863	0.878	0.142	0.255	0.913	0.182	0.5997
Multinomial Naive Bayes Classification	0.834	0.785	5.875	0.785	0.152	0.664	0.791	0.247	0.8181
k-Nearest Neighbour Classification	0.901	0.638	29.190	0.638	0.088	0.613	0.640	0.154	0.6813
AdaBoost Classification	0.960	0.913	4.896	0.913	0.224	0.255	0.950	0.238	0.8195
XGBoost Classification	0.969	0.889	8.256	0.889	0.187	0.321	0.921	0.236	0.8319
Gradient Boosting Classification	0.981	0.915	6.728	0.915	0.199	0.197	0.955	0.198	0.8000

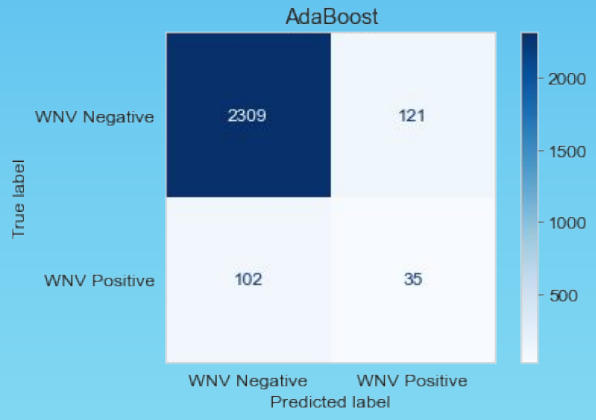


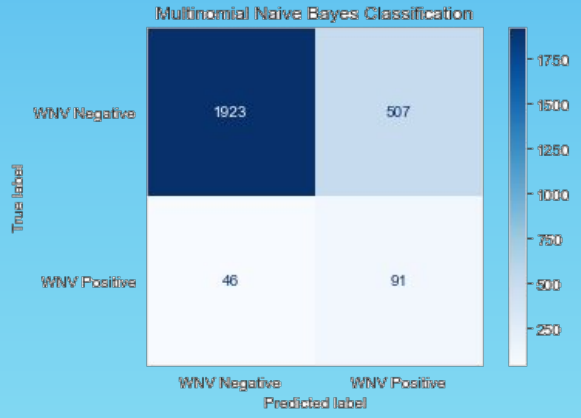
Chosen Model: AdaBoost Model

- One of the best AUC Score
- Generalisation < 5%
- F1 Score also tend to be one of the highest

Model Evaluation

$$\text{Recall} = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

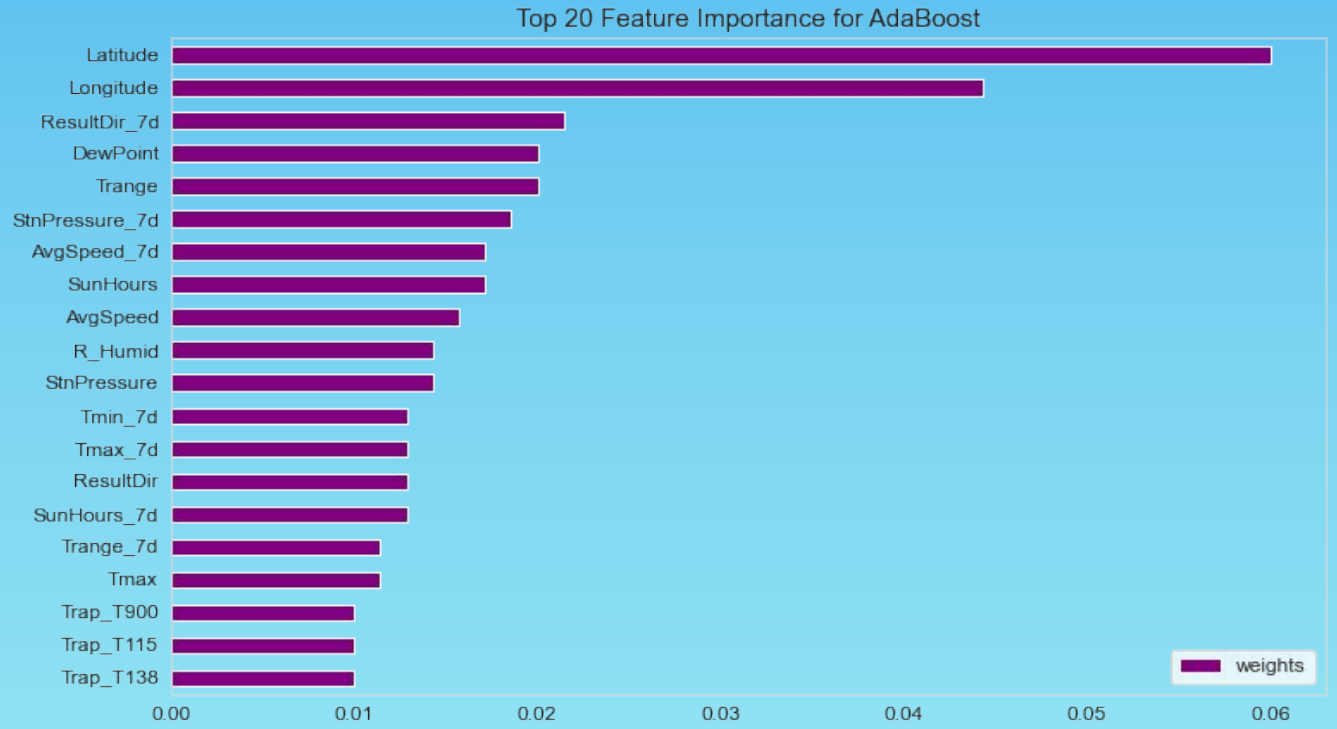




	Train score	Test score	Generalisation	Accuracy	Precision	Recall	Specificity	F1	ROC AUC
Multinomial Naive Bayes Classification	0.834	0.785	5.875	0.785	0.152	0.664	0.791	0.247	0.8181
AdaBoost Classification	0.960	0.913	4.896	0.913	0.224	0.255	0.950	0.238	0.8195

- MNB has a higher Recall value as compared to AdaBoost but as the False Positive is too high, this model could be costly

Feature Importance: AdaBoost



Cost-Benefit Analysis

Cost

- Currently no vaccine to prevent WNV infection
- Pesticide Zenivex E4 (etofenprox) used as an Adult mosquito control product
- Cost incurred to conduct the spraying exercise

Benefits

- Lives saved! (For cases of fatality)
- Medical cost incurred to treat infections

Cost Calculation

Cost of pesticide

Application of 1.5 oz / Acre
Size of Chicago = 60,000 hectares (~ 148,263 acres)
Cost of Zenivex E4 RTU 16 oz = $57.99
Amount of Zenivex E4 RTU required : 98,842 oz
Avg cost = **$358,240**

Application rate pound A.I. per acre	Flow rates Undiluted Oz/Acre	Flow rates Undiluted Oz/Minute	Vehicle Speed
0.00175	0.75	2.25	5
		4.50	10
		7.00	15
0.00350	1.5	4.50	5
		9.00	10
		13.50	15
0.00700	3.0	9.00	5
		18.00	10

Cost for manpower

Avg salary for pest control technician in Chicago, IL = $19.81/hr
Time needed = ~ 184 hours
Assuming the avg technician works for 6 hours, we will require a team of roughly 30 individuals.
Total manpower cost = **$3,600**

Total cost required per spray exercise = $361,840

Benefit Calculation

Median cost of Hospitalization in Illinois for acute viral infections : $33,951

Cost incurred for spraying (per exercise) : $361,840

Spraying exercises per year : 2 times per month over 3 months (Jun - Sep)

Total cost = $2,171,040

Cost of spray exercise in terms of individuals hospitalized = ~ 63.94

Based on our assumptions, it will be financially justified to conduct a spraying as long as it prevents more than 11 individuals from being hospitalized per exercise or 64 individuals from being hospitalized per year.

Conclusion

- Using ADABoost (our best performing model), we achieved an ROC_AUC score of 0.8328 and F1 score of 0.242.
- Feature importance of our model showed that location features (Latitude & Longitude) as well as weather features (DewPoint, ResultDir, Temperature & SunHours) ranked the highest. This indicates that WNV is most likely to occur at given locations and under certain weather conditions.
- Our interpretation for these features to score high could be attributed to denser locations which gives the mosquitoes more opportunities for breeding as well as seasonal cycles where temperatures are ideal for the Culex species to thrive such as Summer.
- Therefore spray efforts should be concentrated at these locations when weather conditions are right.

Recommendations

- Through our cost-benefit analysis, the projected cost of spraying would be financially justified as long as it prevents more than 64 individuals from being hospitalized due to the West Nile Virus.

- Though our costing assumptions are completely straightforward, we believe that there are other more cost-effective techniques that may be applied in conjunction to spraying such as creating awareness amongst the community. These initiatives may be performed through campaigns, education programs and home visits/checks.

- Explore further in detail on deploying targeted spray areas from our model predictions. This will in turn help directly reduce the cost of spraying efforts across Chicago (such as the random spray cluster at High Ridge Knolls Park). However, as the current spray datasets does not substantially quantify the spraying efforts, more evidence (from a better designed and documented spraying regime) would be recommended.

Recommendations

BREAK up hardened soil

LIFT and empty flowerpot plates

OVERTURN pails and wipe their rims

CHANGE water in vases

KEEP roof gutters clear and place BTI insecticide

SPRAY insecticide in dark corners around the house

APPLY insect repellent regularly

WEAR long-sleeve tops and long pants

ption
ge}

2021 Illinois West Nile Virus

Numbers at a Glance

Year: 2021

64 human cases

5 human deaths

65 years - median age of human cases*

11 years - youngest human case*

89 years - oldest human case*

48 counties with positive humans, birds, mosquitoes and/or horses

27 positive birds

2662 positive mosquito batches

5 positive horses and other animals

* This data will be reported when 10 cases have been identified.

Thanks!

Do you have any questions?