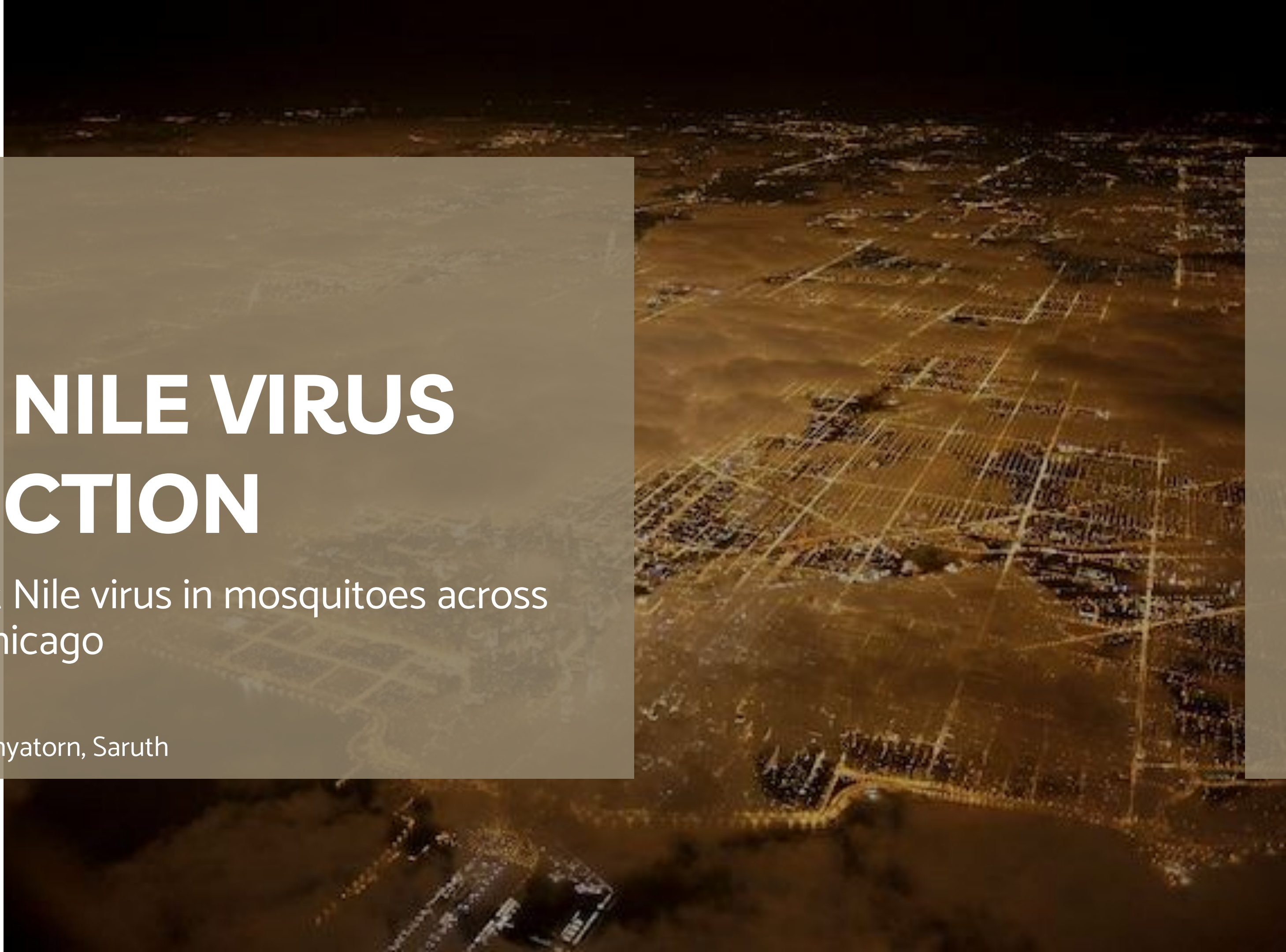


WEST NILE VIRUS PREDICTION

Predict West Nile virus in mosquitoes across the city of Chicago

By Waya, Patchara, Thanyatorn, Saruth

PROBLEM STATEMENT

West Nile virus (WNV) is mosquito-borne disease. It is most commonly spread to people by the bite of an infected mosquito. As data scientists hired by CDC, We want to understand the factors driving the spread of WNV and suggest a cost-efficient method to handle with it.

DATA EXPLANATION

1 WEATHER

Data from 2 weather stations in Chicago included wind, temperature, rainfall amounts: monthly minimum, maximum, average values.

2 SPRAY

Location and dates of pesticide sprays.

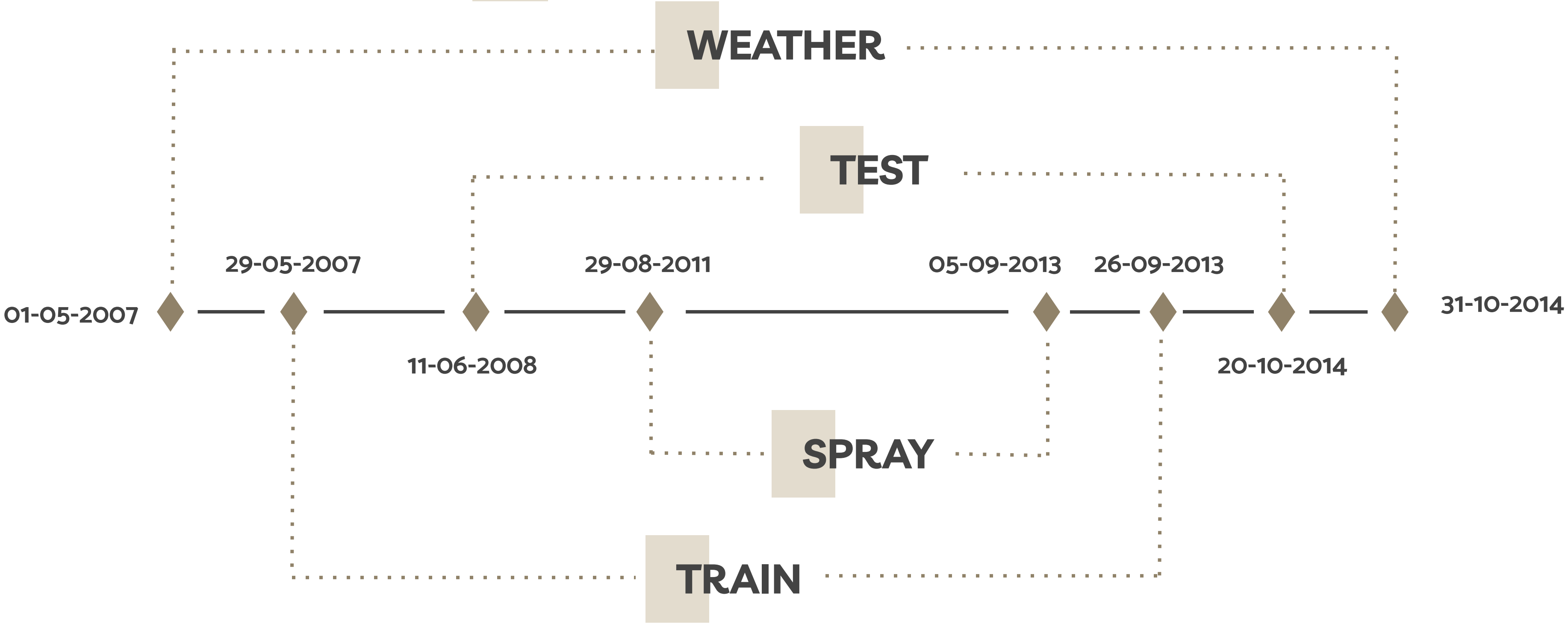
3 TRAIN

Data of date and location of mosquitoes traps with number of mosquitoes and flag of WNV present

4 TEST

Data of date and location of mosquitoes traps

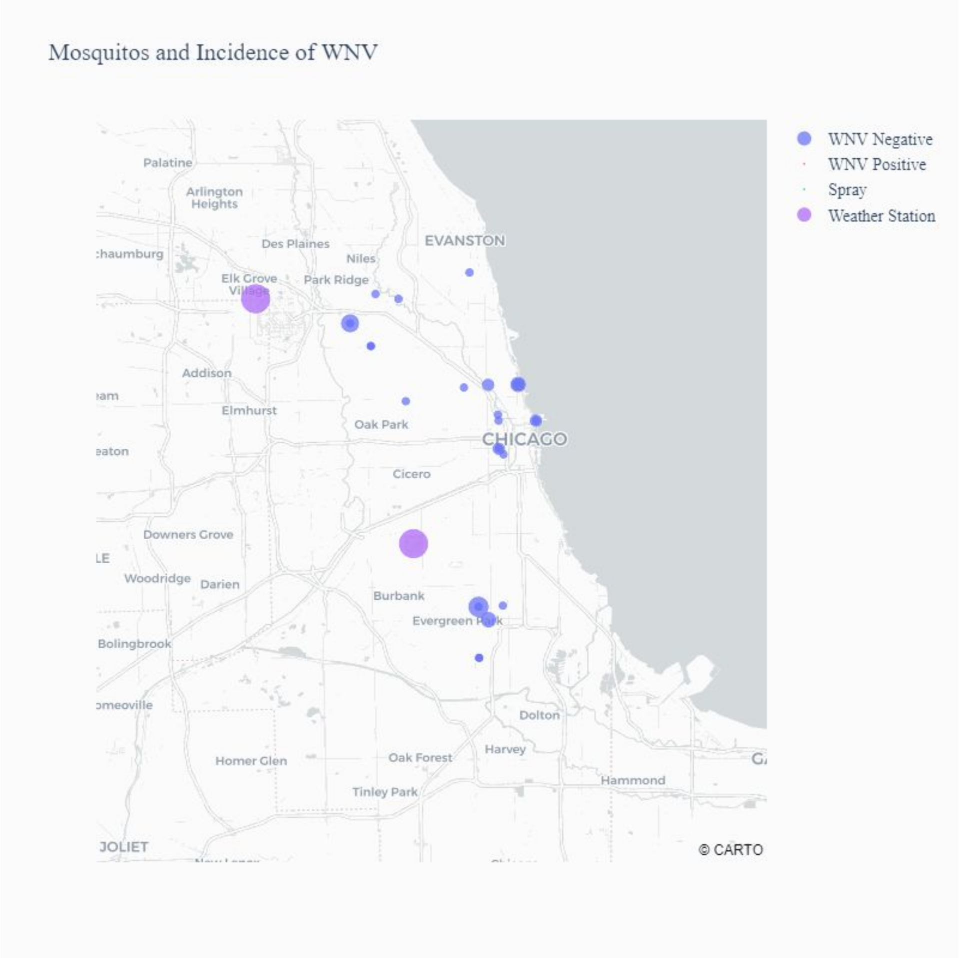
DATA TIMELINE



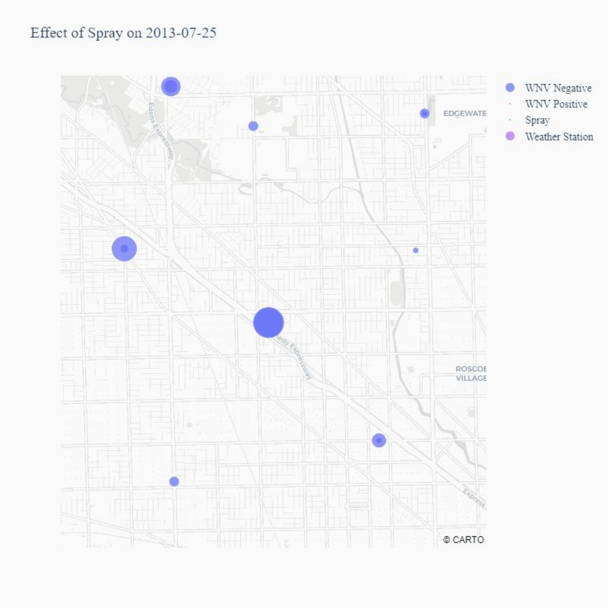
EDA AND FEATURE ENGINEERING

WNV INCIDENCE

- Occurrences are highly related to time.
- The locations also play a part, but not much.
- There are spraying attempts, but the amount, scopes, and frequencies are too small.

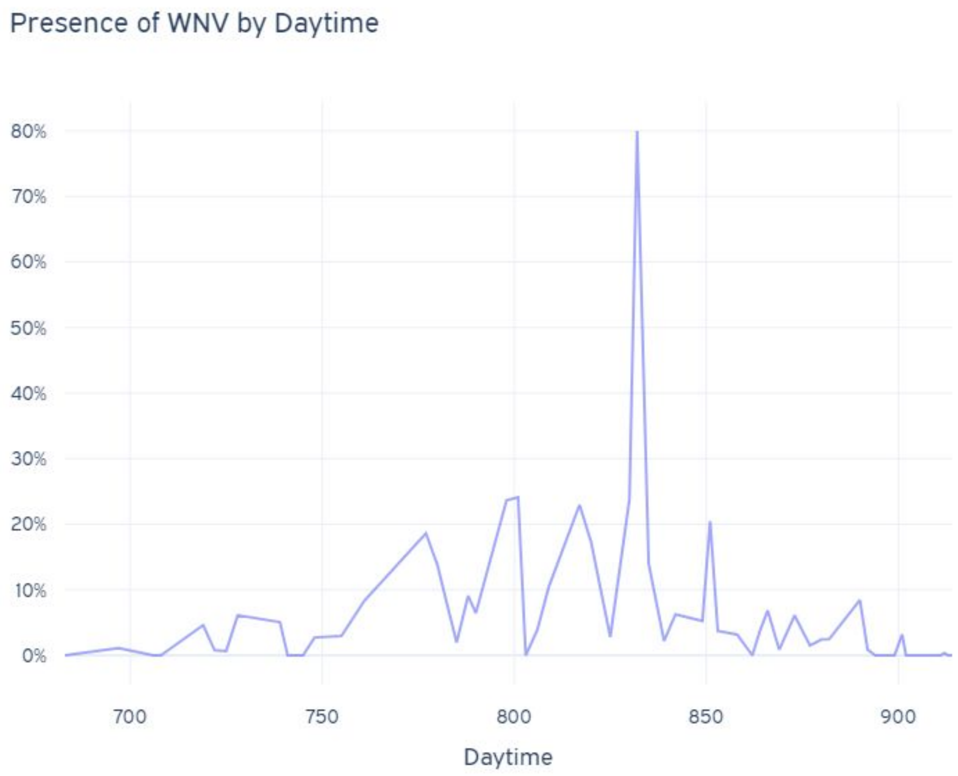


EFFECT OF SPRAYS



- Most sprays are off.
- Hard to tell if effective.
- Numbers of mosquitos go down after sprayed, but so do those not sprayed.
- Effect does not last long. (~14 days)

DAYTIME AS PREDICTOR



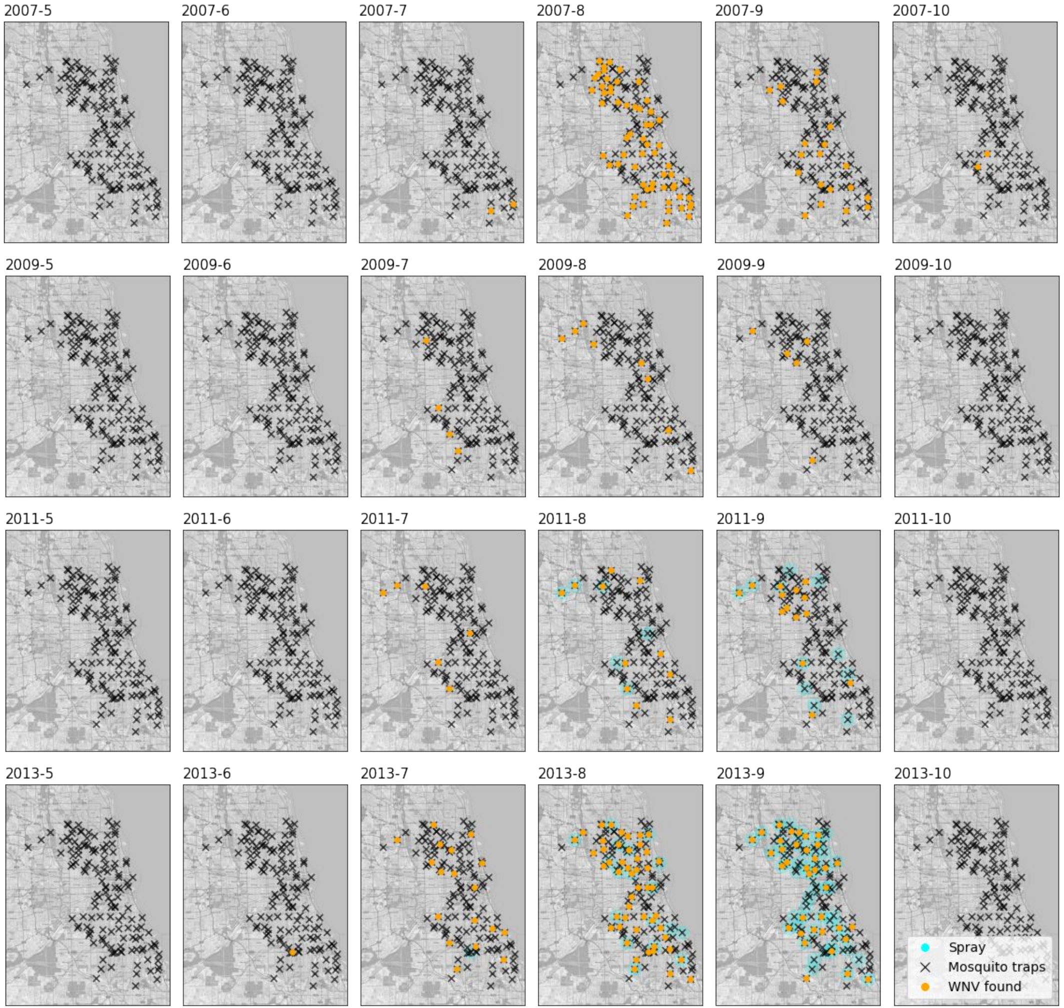
- Daytime = Sunset - Sunrise
- Unexpected peak at ~14 hours
- Correlated but not simply linear

SPECIES AS PREDICTOR

- 3 out of 7 carry WNV.
- Some are dangerous than the others.

WNV and spray location by month/year

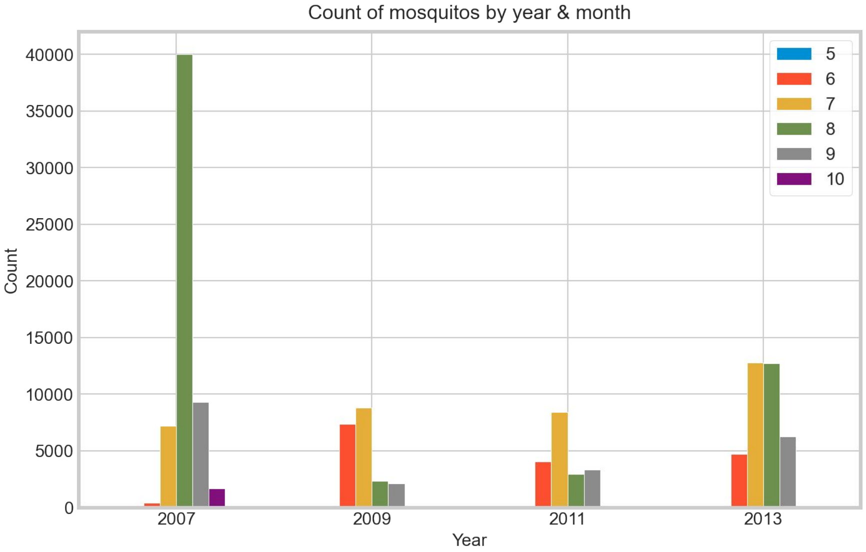
West Nile Virus peak during summer, start to increase from July and peak in the August

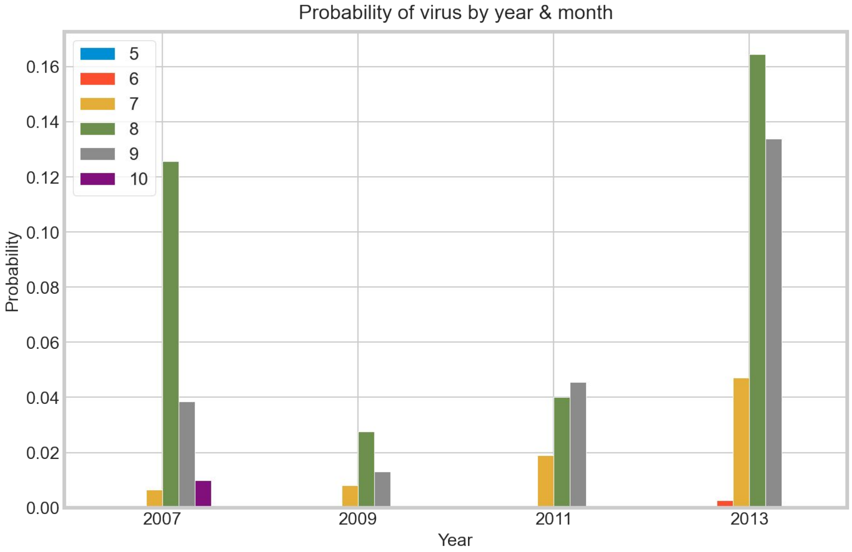


MONTH AS PREDICTOR

- WNV present only in July - October.
- Outbreak peaks in August.

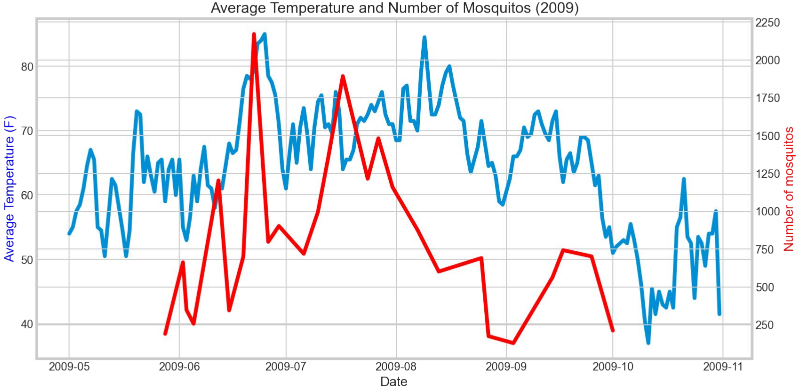
Num Mosquitos
(Number of Mosquitos in a Trap)





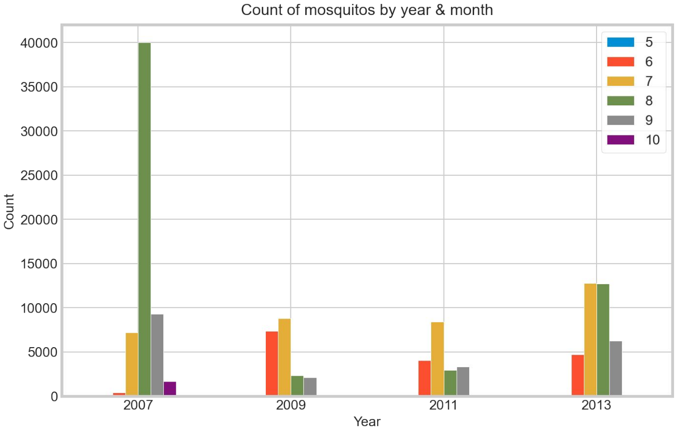
The Number of Mosquitos have the presence of the virus has a positive correlation.
Our training data includes the number of mosquitos in each trap. However, the testing data does not include this feature.
We conclude that it is necessary for our test data to have this feature for predicting the presence of the virus.

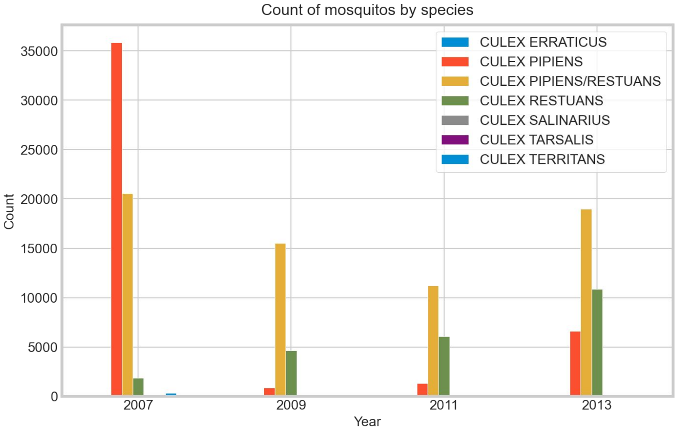
Num Mosquitos
(Number of Mosquitos in a Trap) Correlation



Features that will help us predict the Num Mosquitos:

- Temperature
- Species
- Location
- Month





Predicting Num Mosquitos

```
random forest score: 0.990100755254614

Cross Validation Mean: 0.92198

Cross Validation Std Dev: 0.01681
```

We predicted the Number of mosquitos using a Random Forest Regressor and implement it into the test data to be used for predicting the presence of the Virus

Location Clustering

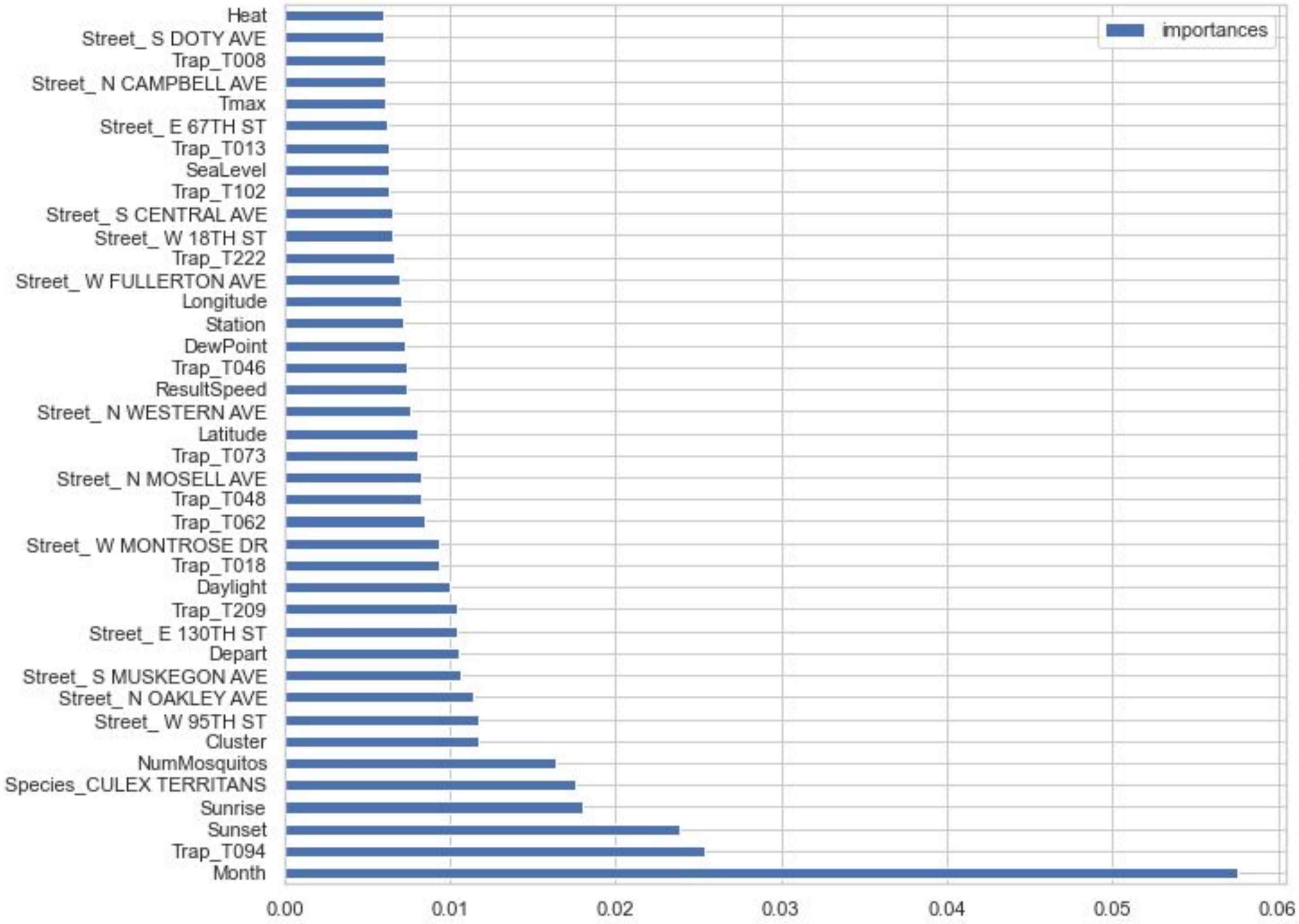
For stronger predictive quality of our models we grouped locations and months into clusters. This will separate months and locations that have more mosquito counts from ones with lower activity.

CLASSIFICATION FEATURES

- Species of Mosquitos
- Number of Mosquitos
- Trap/Street/Cluster/Locations
- Wind/Rainfall
- Daylight
- Temperature

MODELING RESULTS

MODEL	BEST PARAMS	BEST SCORE	ROC/AUC	KAGGLE
Logistic Regression	{'lr__C': 1.0, 'lr__class_weight': 'balanced', 'lr__penalty': 'l2', 'lr__solver': 'liblinear'}	0.8886	0.8899	-
Decision Tree Classifier	{'dt__max_depth': 5, 'dt__min_samples_leaf': 2, 'dt__min_samples_split': 10}	0.9034	0.9008	-
Random Forest Classifier	{'rf__max_depth': 5, 'rf__min_samples_leaf': 1, 'rf__n_estimators': 100}	0.9083	0.9063	-
XGBoost Classifier	{'xgc__colsample_bytree': 0.5, 'xgc__eval_metric': 'auc', 'xgc__gamma': 0.1, 'xgc__learning_rate': 0.1, 'xgc__max_depth': 3, 'xgc__n_estimators': 500, 'xgc__reg_alpha': 1, 'xgc__reg_lambda': 1, 'xgc__scale_pos_weight': 19, 'xgc__subsample': 0.5}	0.9665	0.9720	0.7487

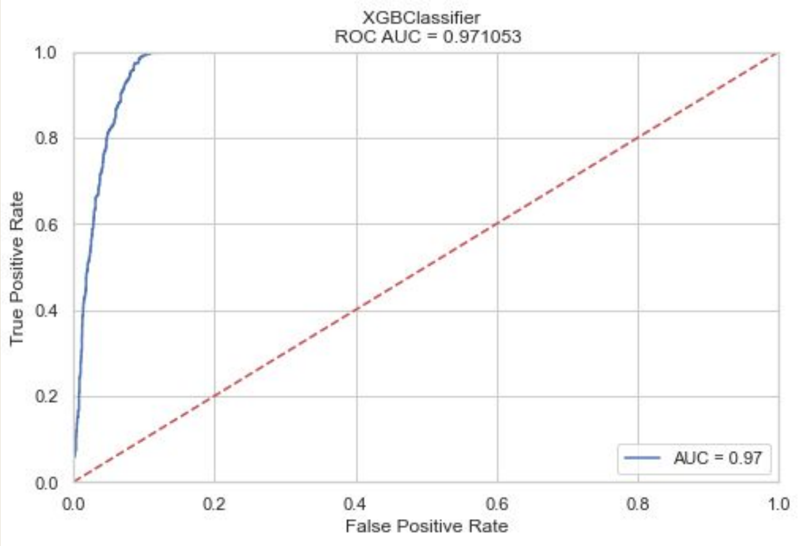


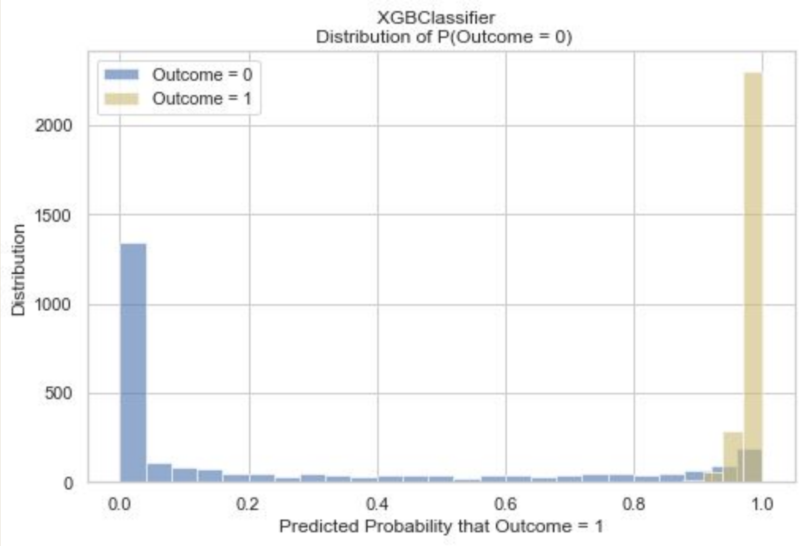
VIRUS INDICATORS

1. Month (7-9)
2. Trap/Street (Location)
3. Cluster
4. DayTime
5. Number of Mosquitos
6. Species

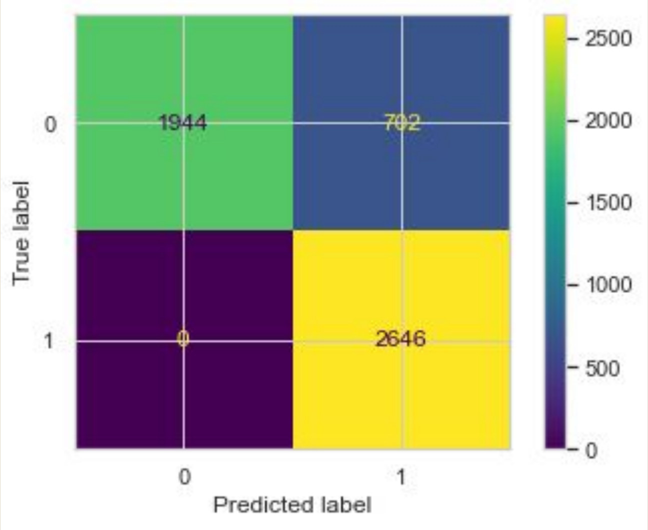
MODEL EVALUATION

The model succeed in distinguishing classes that indicates when virus is presence and when it isn't.





MODEL EVALUATION



Total Predictions: 5292

Correctly Predicting When Virus is Not Presence: 1944

Incorrectly Predicting When Virus is Not Presence: 702

Correctly Predicting When Virus is Presence: 2646

Incorrectly Predicting When Virus is Presence: 0

Accuracy: 86.73%

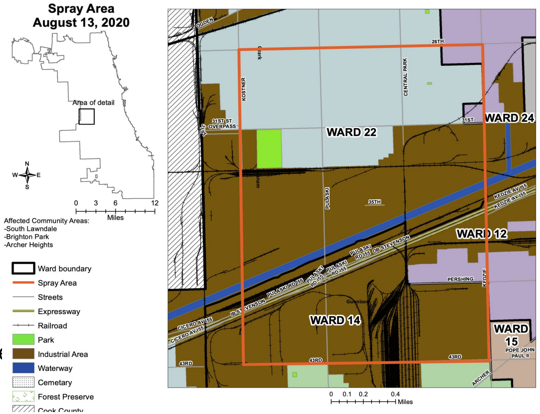
Since we are concern about the prevention of the West Nile virus, it is acceptable that we lean heavily towards predicting that the Virus is presence even in places where there are no viruses. This is because those places have some possibilities of the virus being presence.

Cost-Benefit Analysis

Key Findings

Cost-Benefit Analysis of Spraying Insecticide

1. Cost



- **Vertical control measure - Spraying Insecticide**

  It cost approximately **$852,328** to spray cover an area of **660.1 squared** km by Vector control measure

  - Sprayed at 4.5 ounces per minute
  - Spray duration = 5 hrs conducted from dusk to around midnight from ultra low volume (ULV) sprayer truck which has the overall spray area approximately 0.6 squared km per truck
  - Zenivex E4 cost is about $80 USD per gallon (https://www.cmmcp.org/pesticide-information/pages/zenivex-e4-etofenprox)

- **Medical Fees**
  - Assuming there will be an outbreak, the medical fees are estimated to be
    - $168 million to $250 million per year (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3945683/)

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
Outpatient costs	Cost per case¶				
Outpatient hospital treatment	$333	17	36	$5,668	$15,337
Physician visits	$450	46	100	$20,708	$20,708
Outpatient physical therapy	$909	46	100	$41,810	$41,810
Occupational therapy	$4,037	3	7	$12,111	$185,699
Speech therapy	$588	1	1	$588	$27,032
Total				$80,885	$290,586
Nursing home costs	Cost#				
Nursing home stay**	$190	2	4	$36,195	$36,195
Transportation	$65	46	100	$2,977	$2,977
Home health aides, babysitters, etc.	$1,569	7	14	$10,983	$505,211
Total				$50,154	$544,383
Total for all WNND				$2,140,409	$2,844,339

Table 3. Estimated economic costs of WNND cases due to productivity loss, Sacramento County, California, 2005*

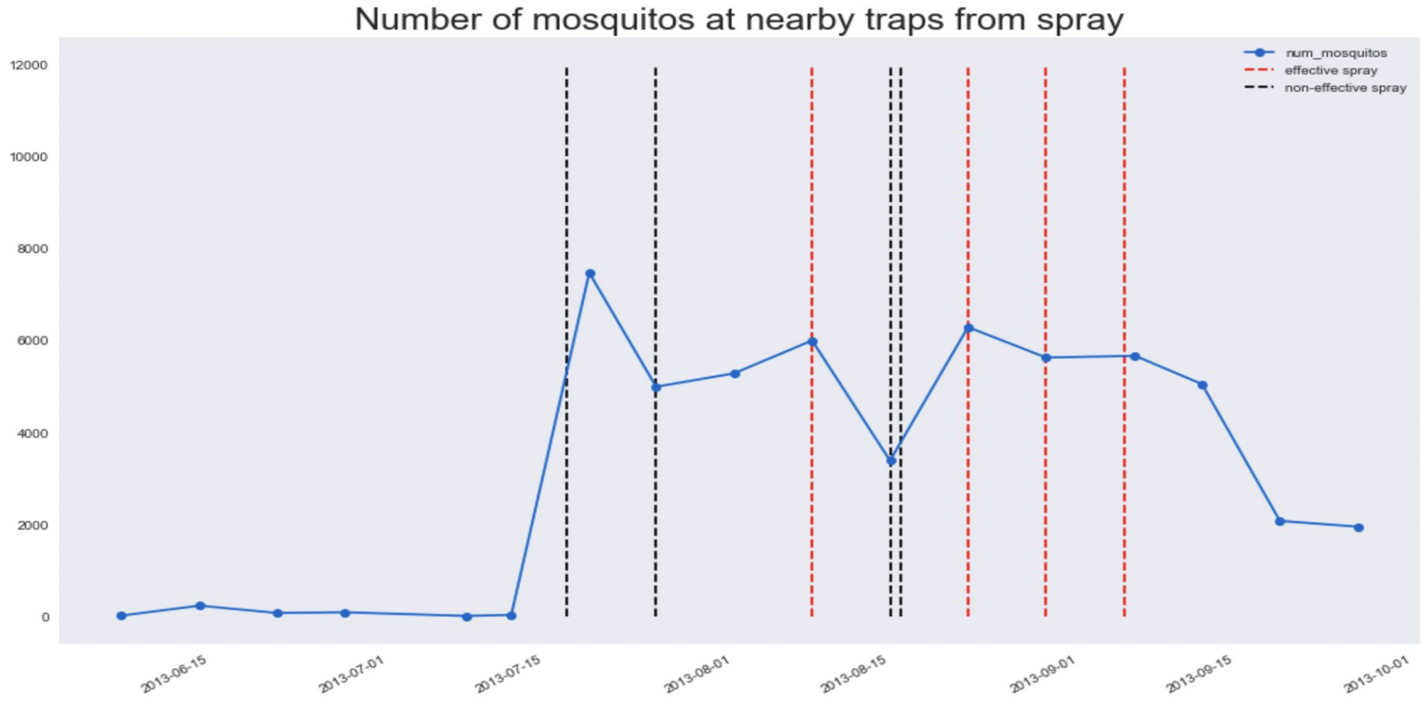
Productivity loss	Value of work day missed†	Value of nonwork day missed‡	No. work days missed	No. nonwork days missed	No. patients <60	No. patients ≥60	% Cases	Total costs for all cases
For patients <60 y	$191	$125	50	10	31		100	$334,800
For patients ≥60 y		$125		60		15	100	$112,500
For caretakers		$125	25		8	4	26	$37,500
Total costs								$484,800

Key Findings

Cost-Benefit Analysis of Spraying Insecticide

2. Benefits

- **Increased quality of life from fewer people falling sick and dying**
- **Increased workplace productivity from fewer people falling ill and going on medical leaves as well as savings in hospital expenses from treating WNV patients**
  - **1 in 5 people** infected with WNV develop West Nile fever
  - and **1 in 150 people** develop more severe symptoms
- **Reduced costs of the state to handle such outbreaks and emergencies**
- In 2017, there were 90 WNV cases, including 8 deaths. Assuming the median household income in Chicago of $55,295 and an average hospital cost of $25,000 per patient, the cost was approximately **$490,000**
  - Assume all were working adults and each took two weeks off work to recover



<u>Effectiveness of spraying efforts thus far</u>

- There was lack of evidence to support the claim that **mosquito spraying had any effect on reducing the number of mosquitos**
  - 50% of effective spray
  - Near the end of summer, breeding conditions were already becoming less favourable for mosquitos
  - Difficult to discern the effects of spraying from the natural decline in mosquito population
- However, we can thus conclude that it still helps to decrease the mosquito population with 50% effective

Conclusion/Recommendation

In conclusion, the main factors driving the spread of the West Nile virus (WNV) are the species of the mosquitos, the number of mosquitos, weather conditions, and the amount of daylight with in a certain period of time. Through our findings, we can conclude that the West Nile Virus is highly seasonal, becoming most prominent in July and August. This is due to the increase in the heat and mosquitos count.

Because our goal is to prevent the spread of the virus, it is recommended to spay in places that have any indications of the virus. As from our model, some places can have the possibility of the virus being presence when there isn't at the time. From a cost standpoint, spraying in these places will be beneficial in the long run due to how the virus can impact a patient.