
WestNile is Coming
Clement, Abirami, Irwin, Keith

Problem Statement

To minimize the presence of the West Nile Virus by suppressing the population of *Culex* mosquitoes, in a cost effective and sustainable manner.

In North America since 1999

Species of dead birds in which West Nile virus has been detected, United States, 1999-2016

Bird species/Common name	Native/Exotic/Captive
Abyssinian Ground-Hornbill	Exotic-Captive
Acorn Woodpecker	Native
African Grey Parrot	Exotic-Captive
African Penguin	Exotic-Captive
American Coot	Native
American Crow	Native
American Dipper	Native
American Goldfinch	Native
American Kestrel	Native
American Robin	
American White Pelic	
Anna's Hummingbird	
Ash-throated Flycatch	
Bald Eagle	
Baltimore Oriole	
Band-tailed Pigeon	
Bank Swallow	
Barn Owl	
Barn Swallow	
Barred Owl	
Belted Kingfisher	
Black Phoebe	
Black Skimmer	

Bird species/Common name	Native/Exotic/Captive
Boat-tailed Grackle	Native
Bobolink	Native
Boreal Owl	Native-Captive
Brewer's Blackbird	Native
Broad-winged Hawk	Native
Bronzed Cowbird	Native
Bronze-winged Duck	Exotic-Captive
Brown Thrasher	Native
Brown-headed Cowbird	Native

Mosquito species in which West Nile virus has been detected, United States, 1999-2016

Mosquito Species	Mosquito Species
Aedes aegypti	Culex erraticus
Aedes albopictus	Culex erythrothorax
Aedes atlanticus/tormentor	Culex nigripalpus
Aedes atropalpus	Culex pipiens
Aedes canadensis	Culex quinquefasciatus
Aedes cantator	Culex restuans
Aedes cinereus	Culex salinarius
Aedes condolescens*	Culex stigmatosoma
Aedes dorsalis	Culex tarsalis
Aedes dupreei	Culex territans
Aedes epactius	Culex thriambus
Aedes fitchii	Culiseta incidens
Aedes fulvus pallens	Culiseta impatiens
Aedes grossbecki	Culiseta inornata
Aedes infirmatus	Culiseta melanura

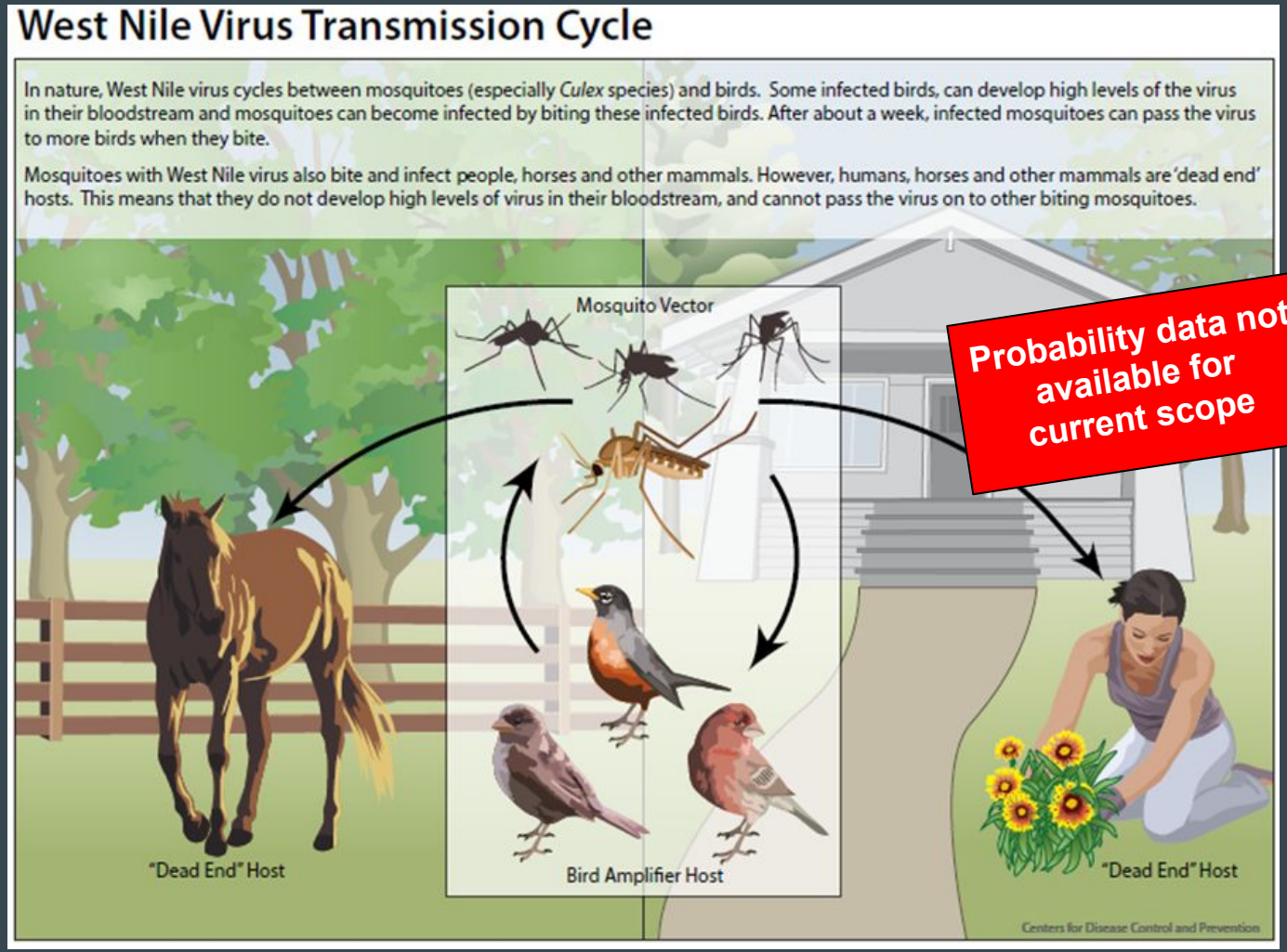
Centers for Disease Control and Prevention
Division of Vector-Borne Diseases
Promote health and quality of life by preventing and controlling vector-borne diseases

West Nile Virus in the United States: Guidelines for Surveillance, Prevention, and Control

U.S. Department of Health and Human Services
Public Health Service
Centers for Disease Control and Prevention
National Center for Emerging and Zoonotic Infectious Diseases
Division of Vector-Borne Diseases
Fort Collins, Colorado

4th Revision
June 14, 2013

Source: Centers for Disease Control
https://www.cdc.gov



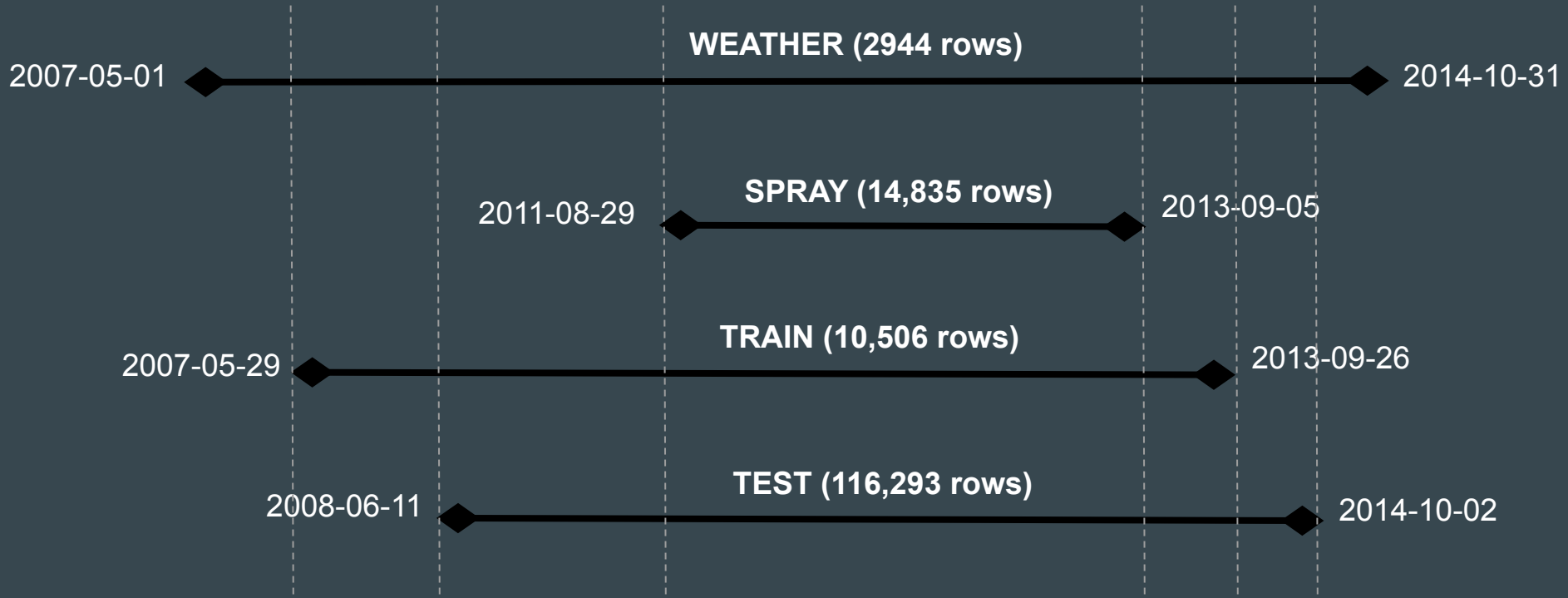
Probability data not available for current scope

Conditions for the **West Nile Virus (WNV)** to be a threat to the public:

1) WNV-Infected birds must be present
2) Mosquitoes (Culex species) have to be infected by WNV upon feeding from the birds
3) Mosquitoes (Culex species) must be present among public

Source: Centers for Disease Control
https://www.cdc.gov/westnile/transmission/

Summary of Data



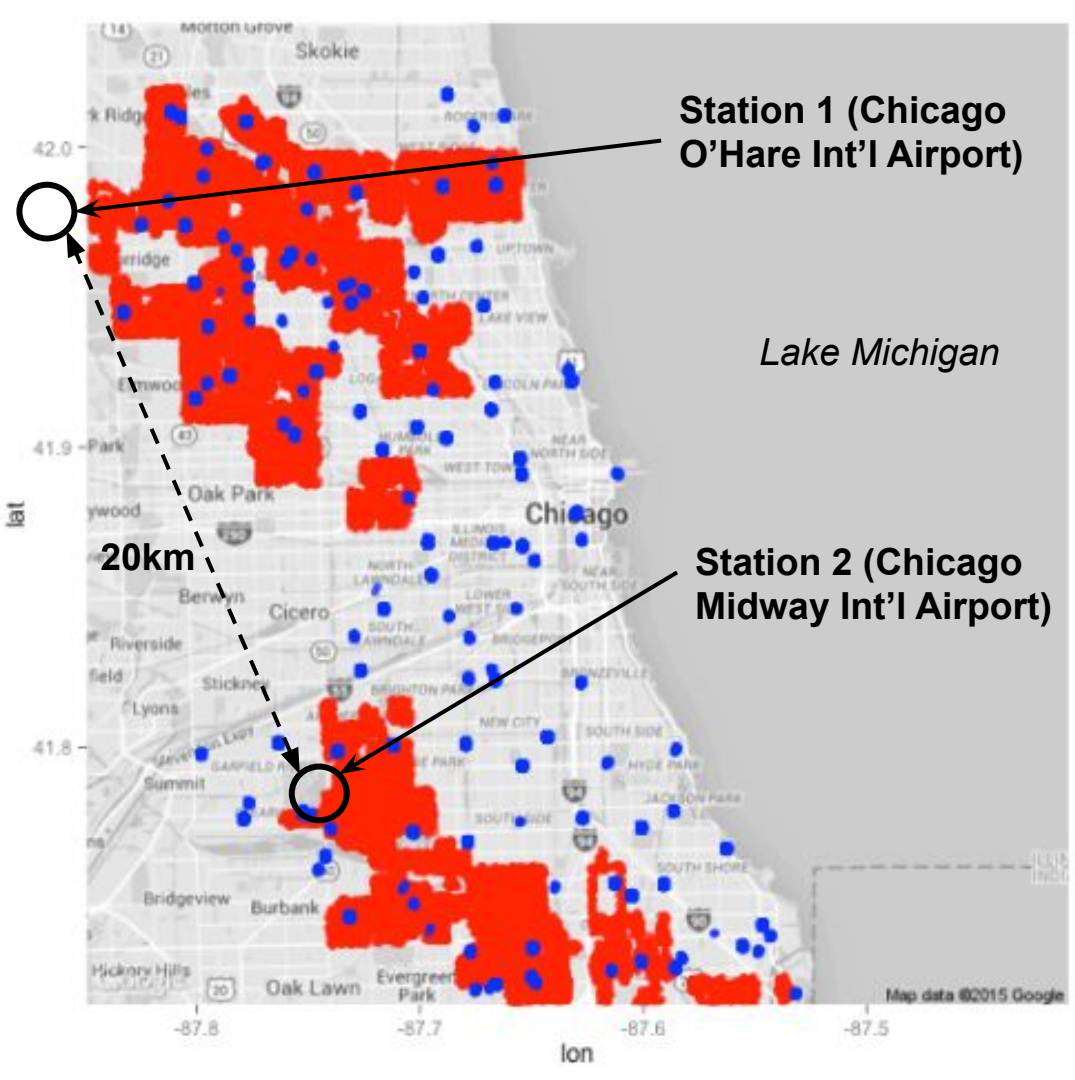
Mosquito Traps

Source: Kaggle (West NIle Virus Prediction
https://www.kaggle.com/c/predict-west-nile-virus/data

These test results are organized in such a way that when the number of mosquitos exceed 50, they are split into another record (another row in the dataset), such that the number of mosquitos are capped at 50.

Please note that not all the locations are tested at all times. Also, records exist only when a particular species of mosquitos is found at a certain trap at a certain time. In the test set, we ask you for all combinations/permutations of possible predictions and are only scoring the observed ones.

1. ***Multiple species (Culex, Aedes, Anopheles) of mosquitoes are found in each trap.***
2. ***Mosquitoes are separated in batches of 50 per batch.***
3. ***Culex mosquitoes are identified from each batch and counted.***
4. **‘*NumMosquitos*’ in Train data refers to the number of *Culex* mosquitoes found in that batch.**



WEATHER DATA

1) Imputed data mostly for weather records for Station 2 e.g. Tavg, Sunrise, Sunset

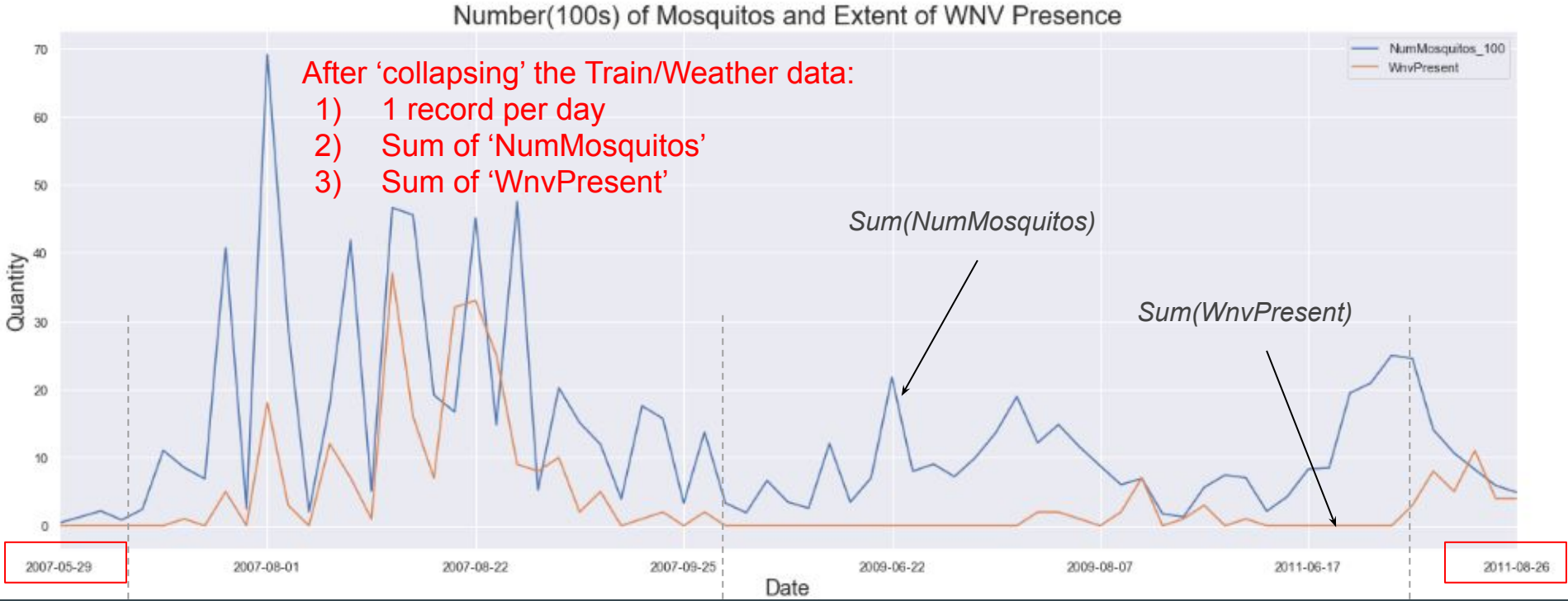
2) Created new features such as daylight time

3) Expanded CodeSum column into dummy columns

Merge TRAIN & TEST with WEATHER data

1) Checked the distance between each observation point (Train/Test) and Station 1 & 2.

2) Merged the weather data to Train/Test data according to the nearest Station.



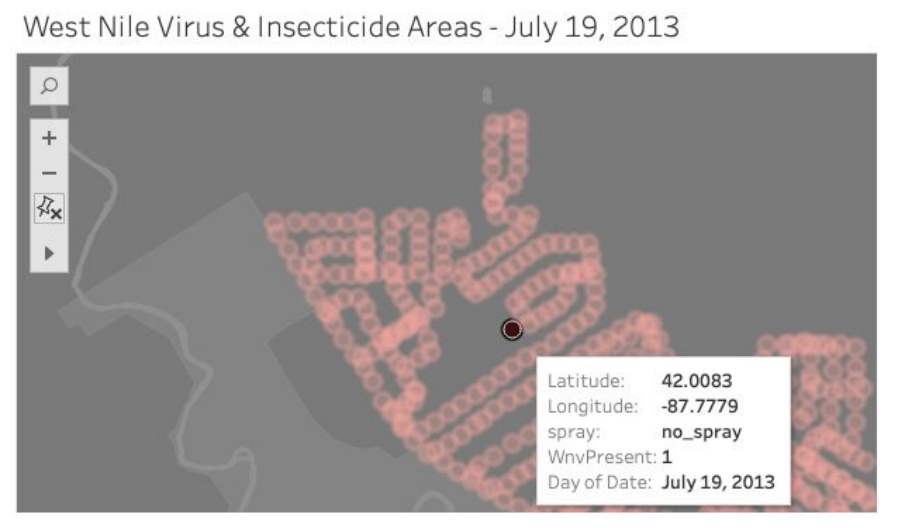
WNV Presence appears to follow quantity of *Culex* mosquitoes found

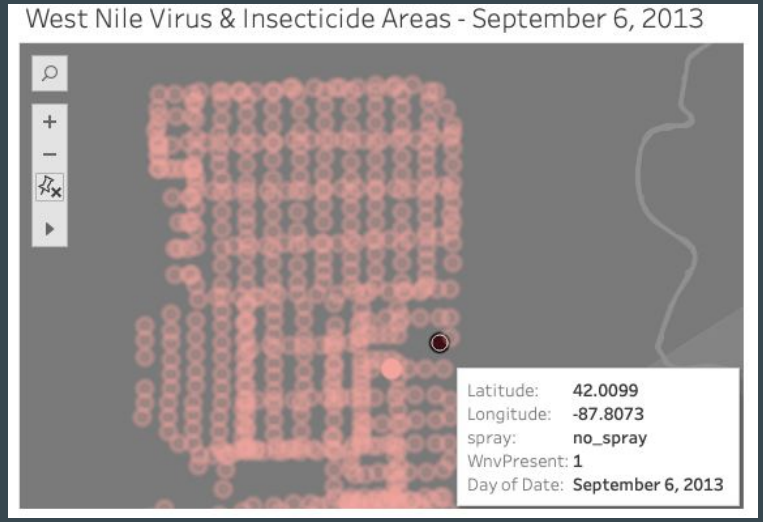
WNV Presence **does not** appear to follow quantity of *Culex* mosquitoes found

As expected, the presence of Culex mosquitoes does not imply WNV presence. WNV presence also depends on the interactions between the Culex mosquitoes and WNV-infected birds.

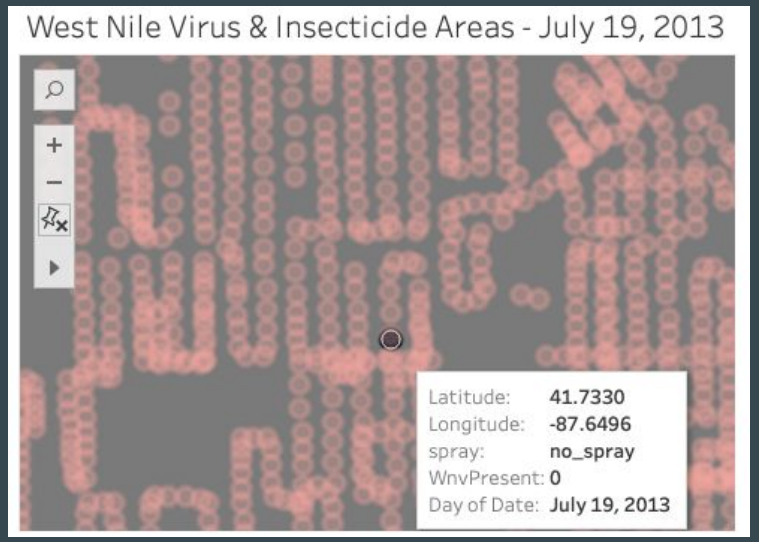
EDA & Time Series Analysis

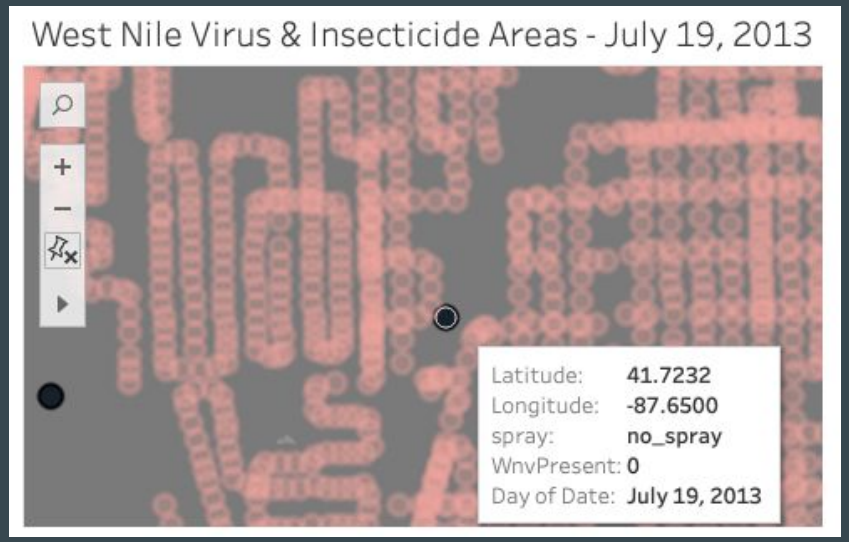
Is spraying effective?

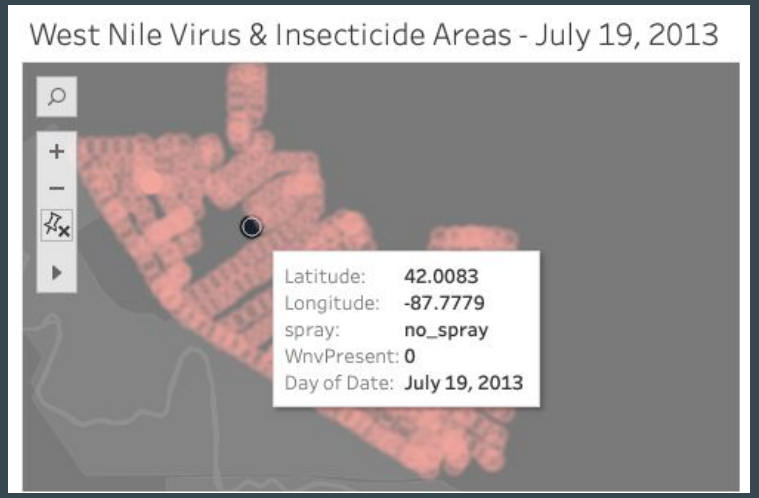




In the above cases, the insecticide was sprayed on July 17 2013 and 2 days after the mosquitos in that trap were tested positive for West Nile Virus.





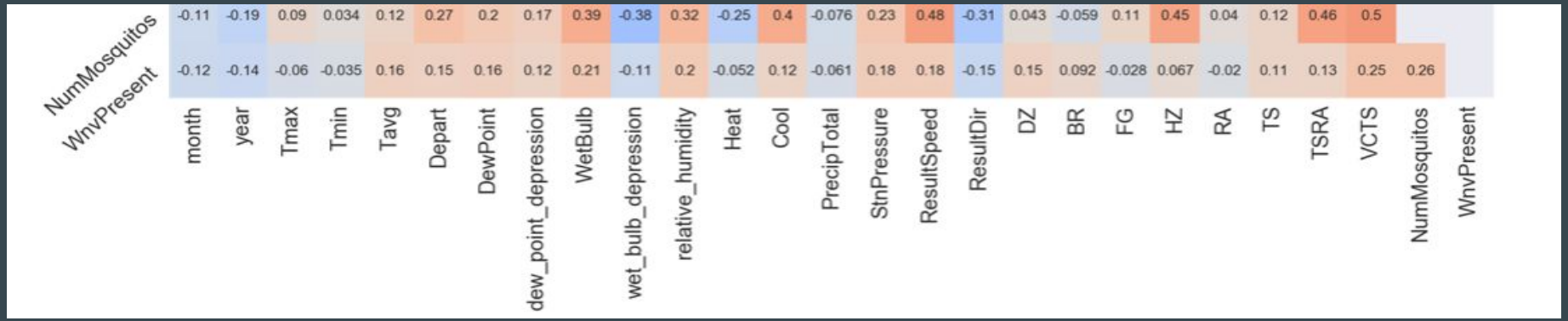


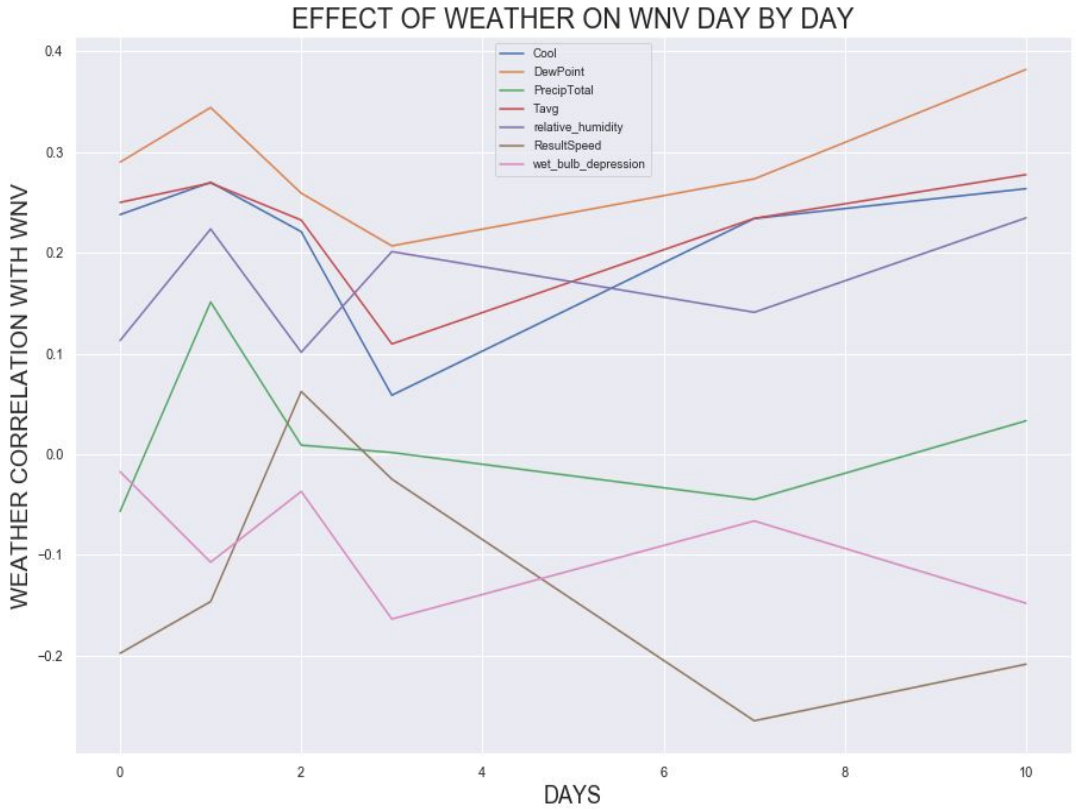
FINDINGS ON SPRAY EFFECTIVENESS

Hence, there is no conclusive evidence that suggests that the spraying at most two days before actually really repels the occurrence of the West Nile Virus. The reason could be because the spraying is not regular enough for it to be effective in subduing the virus.

Heatmap

- The above heatmap shows none of the variable well correlated with both WNV virus and number of mosquitos.
- But weather condition by itself well correlated with another weather condition.
- Tavg, Tmax(To be removed) and Tmin(To be removed)
- Cool and Wetbulb(To be removed)
- Wetbulb Depression and Dewpoint Depression(To be Removed)





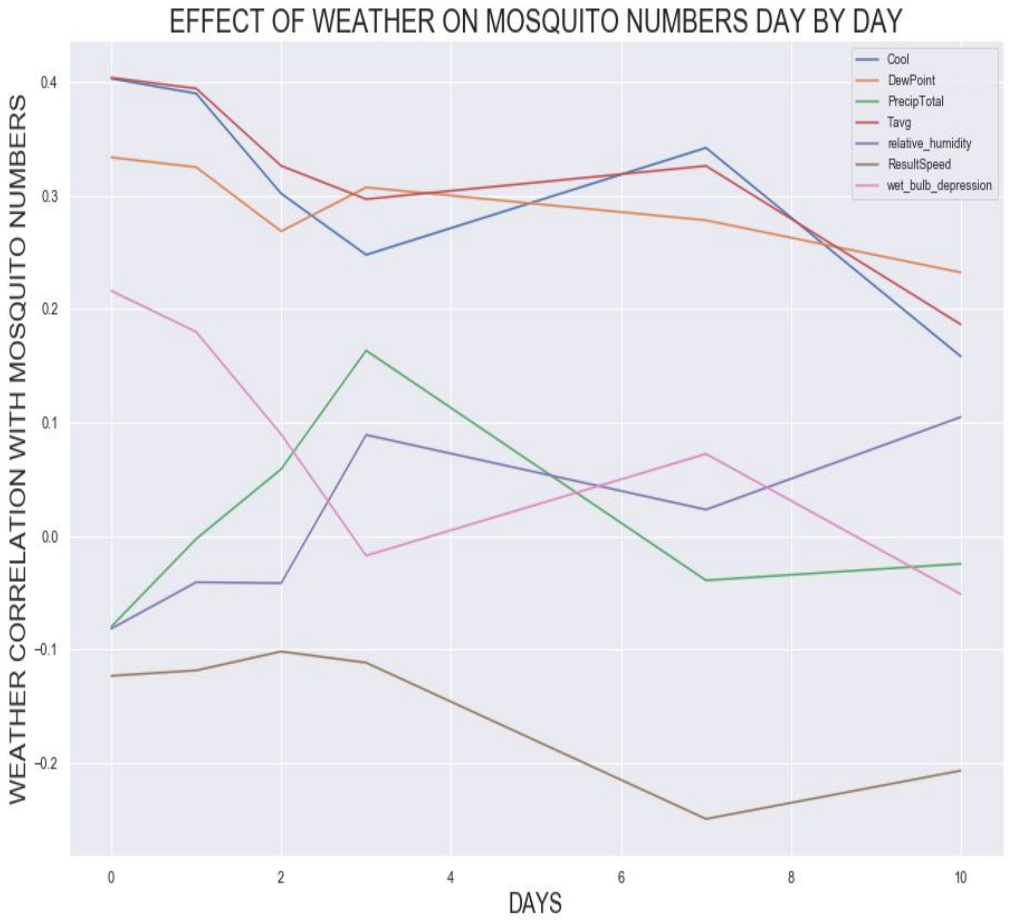
Effect of weather on the presence of WNV on a day-by-day basis

1. On the 1st day, the correlation between the weather condition and WNV is higher compared to same day and the second day except Result speed.

2. Result speed shows effect on the same day and in the 1st and 2nd day the WNV become higher as there maybe a change in wind direction or reduced speed on the days.

3. On the 7th day, in general there is no effect of weather on WNV.

4. On the 10th day, maybe the young mosquitos developed after hatching, so again there is a raise in correlation.



Effect of weather on the presence of mosquitoes on a day-by-day basis

1. Mosquito numbers goes high 0 to 3 days after rain or if the day is humid.

2. If wet bulb depression is higher than the effect on mosquito is almost 0 on the third day. As it is the opposite of humidity.

3. Wind speed has negative correlation with number of mosquito. But after the wind subside the number of mosquito raise from 2nd to 7th day.

4. Cool and Tavg are quiet correlated, so they show same effect on mosquito. Mosquitos reduces with raise in temperature.

5. Dewpoint correlation with number of mosquito is somewhat constant except on 2nd and 7th day.

Modelling

Steps

1. Further Feature Engineering & Selection
    a. Creation of a long-lat column
    b. Dummied
    c. Top 40 correlated feature selected
    d. Train-test-split on train.csv, while test.csv was maintained
    e. Polynomial with degree 2 and standard scalar
    f. Oversampling using Synthetic Minority Over-Sampling Technique (SMOTE)

2. Find best performing model using GridSearch on classification models:
    a. LogisticRegression(),
    b. KNeighborsClassifier(),
    c. DecisionTreeClassifier(),
    d. RandomForestClassifier(n_estimators=100),
    e. ExtraTreesClassifier(n_estimators=100),
    f. GradientBoostingClassifier(n_estimators=100),
    g. AdaBoostClassifier(n_estimators=100)

Model Selection

```
Results from GradientBoostingClassifier:
Train Score: 0.9111
Test Score: 0.8736
Using the following parameters: {'gbc__learning_rat

From Confusion Matrix:
True Negatives: 2237
False Positives: 252
False Negatives: 80
True Positives: 58
```

	precision	recall	f1-score	support
Predict 0	0.965473	0.898755	0.930920	2489.00000
Predict 1	0.187097	0.420290	0.258929	138.00000
accuracy	0.873620	0.873620	0.873620	0.87362

```
Results from AdaBoostClassifier:
Train Score: 0.8629
Test Score: 0.8047
Using the following parameters: {'abc__learning_rate': 0.73}

From Confusion Matrix:
True Negatives: 2029
False Positives: 460
False Negatives: 53
True Positives: 85
```

	precision	recall	f1-score	support
Predict 0	0.974544	0.815187	0.887771	2489.00000
Predict 1	0.155963	0.615942	0.248902	138.00000
accuracy	0.804720	0.804720	0.804720	0.80472

Improvements

1. Limit the observations in the train and test dataset to the 3 species of Culex mosquitoes which carry the virus, to improve the modelling accuracy of the models.

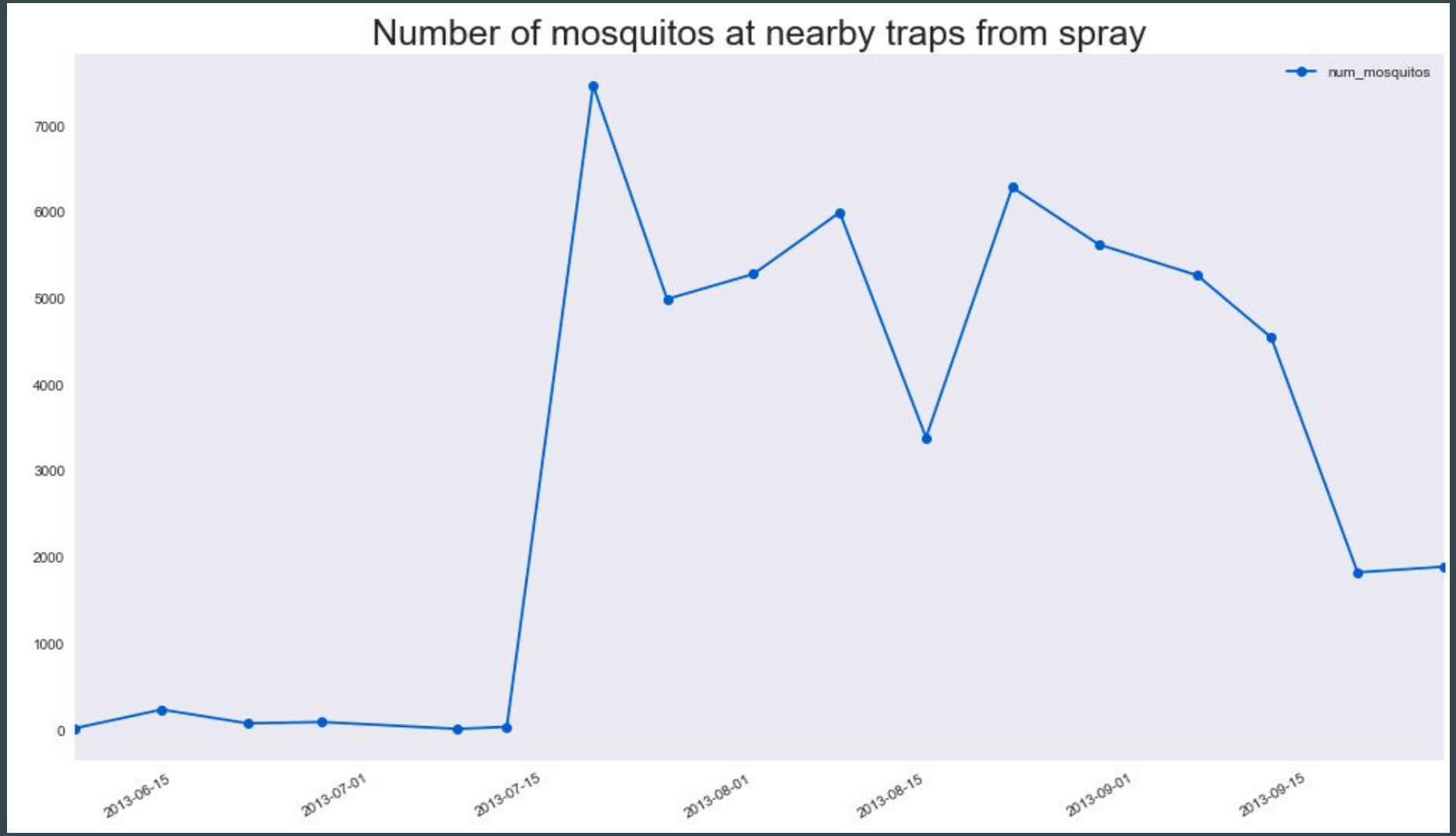
2. Include the number of mosquitoes as a predicted variable, which would then be used to predict the presence of the virus.

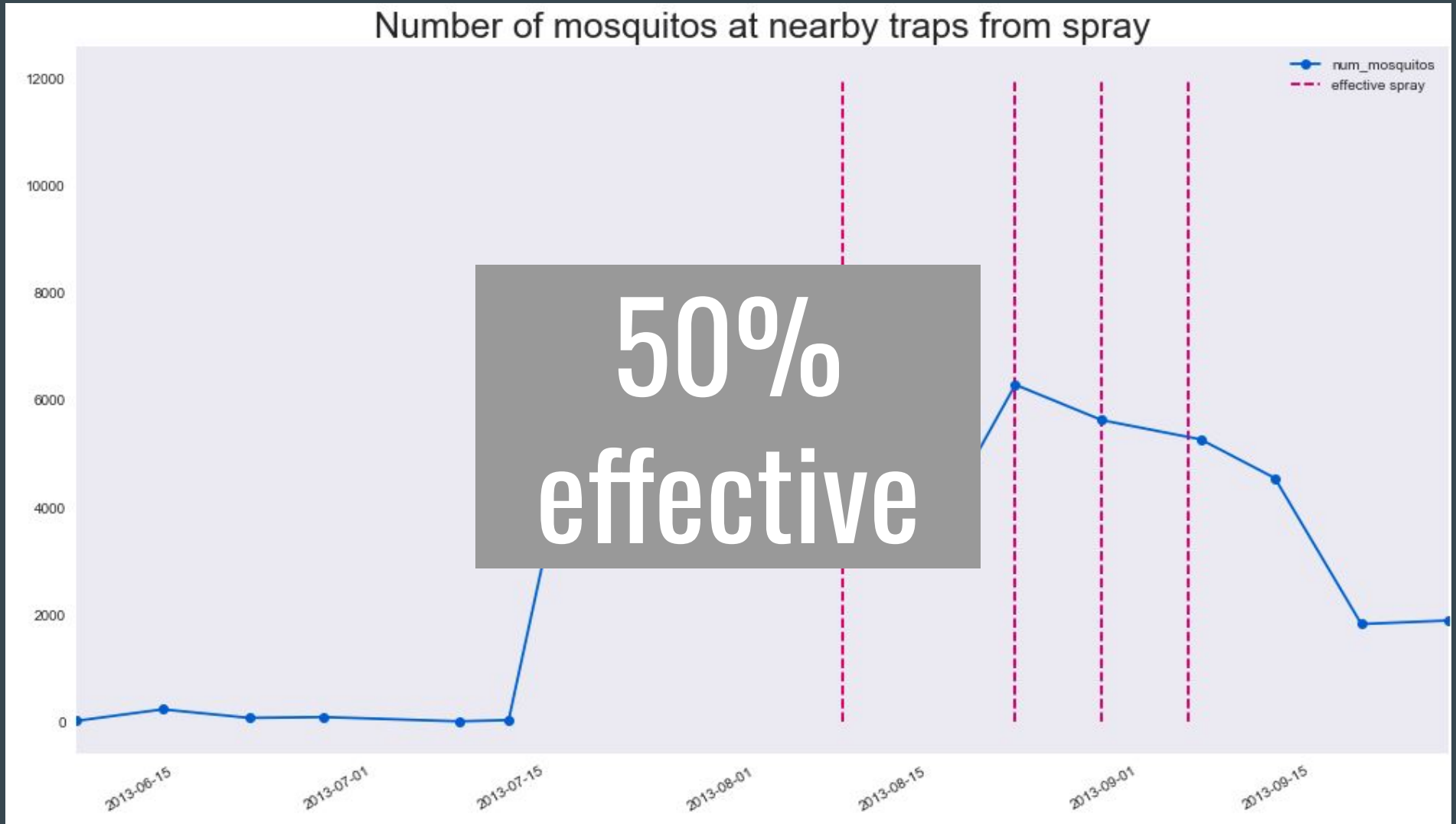
3. Increasing the parameters in the gridsearch with the objective of achieving a model with the highest f-1 score.

Cost and Benefit Analysis

Effectiveness of insecticide spraying

- Previously we looked at insecticide spraying impact on West Nile Virus
- How about its impact on mosquitos, the carriers of the virus?


Number of mosquitos at nearby traps from spray
num_mosquitos
7000
6000
5000
4000
3000
2000
1000
0
2013-06-15
2013-07-01
2013-07-15
2013-08-01
2013-08-15
2013-09-01
2013-09-15


Number of mosquitos at nearby traps from spray
num_mosquitos
effective spray
12000
10000
8000
6000
4000
2000
0
2013-06-15
2013-07-01
2013-07-15
2013-08-01
2013-08-15
2013-09-01
2013-09-15
50%
effective

So is spraying really effective?

- 50% effective on number of mosquitos
- Since it reduces the mosquitos half of the time, we should still do it

Projected cases and costs

Year	Projected Cases
2008	2261
2010	2943
2012	4948
2014	4765

Conclusions and recommendations

- Study migratory patterns of affected birds
- Regression analysis to predict number of mosquitos

To be safe, let's still spray!

And educate the public on reducing at-risk breeding areas!