

# Project 4

## *West Nile Virus Prediction*

## Problem Statement

Build a Classification model:

To predict accurately the Outbreaks of West Nile Virus from Mosquitoes, based on Environmental and Other Variables.

# Exploratory Data Analysis

## **Main dataset:**

Mosquito trap (date, location, species, number of mosquitos, WNV present)

## **Weather data:**

Weather conditions during the months of test

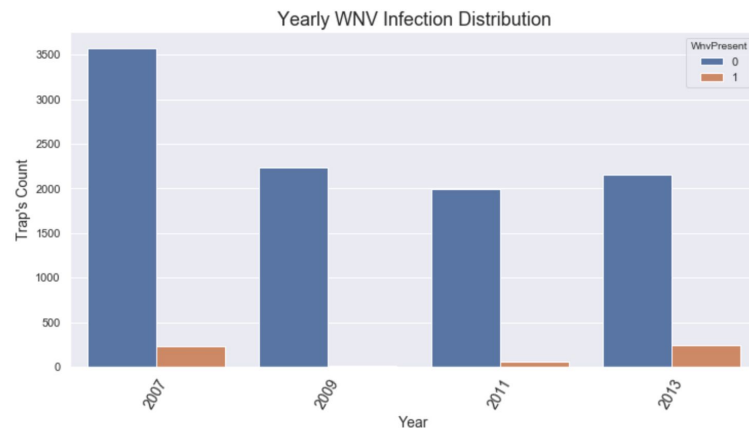
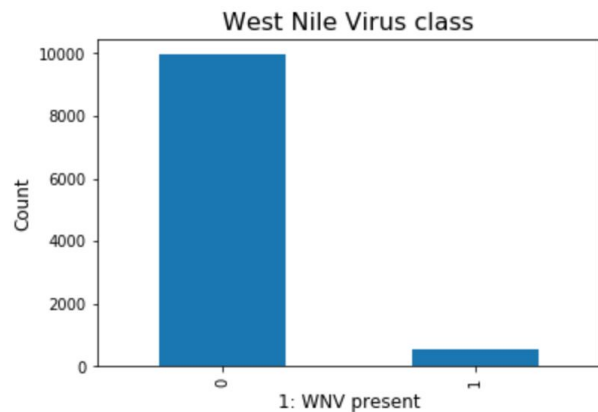
## **Spray dataset:**

GIS data for spray effort

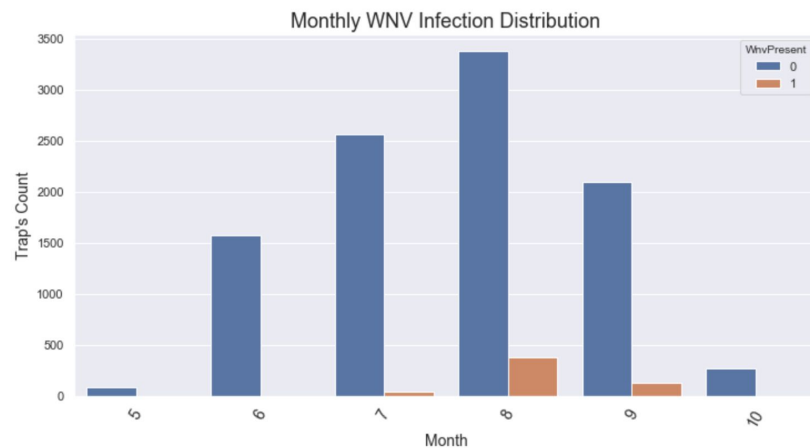
---

# Train data

Wnv Present

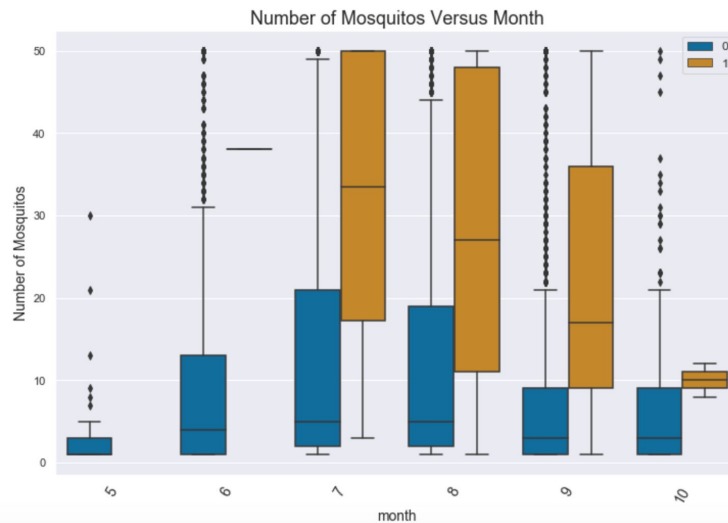
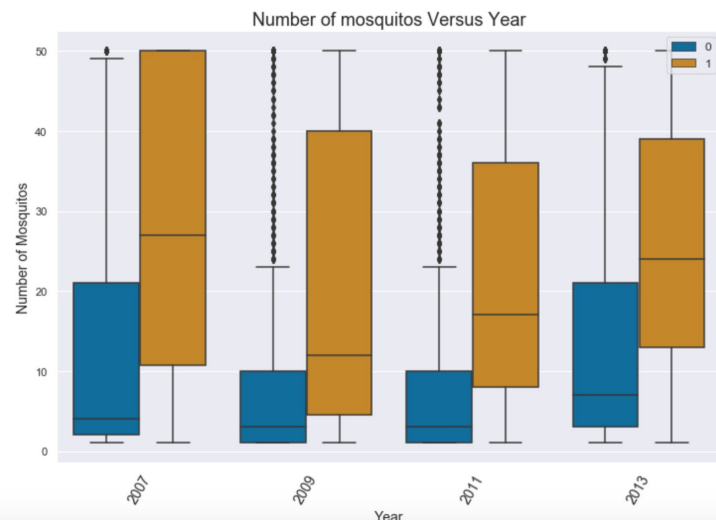
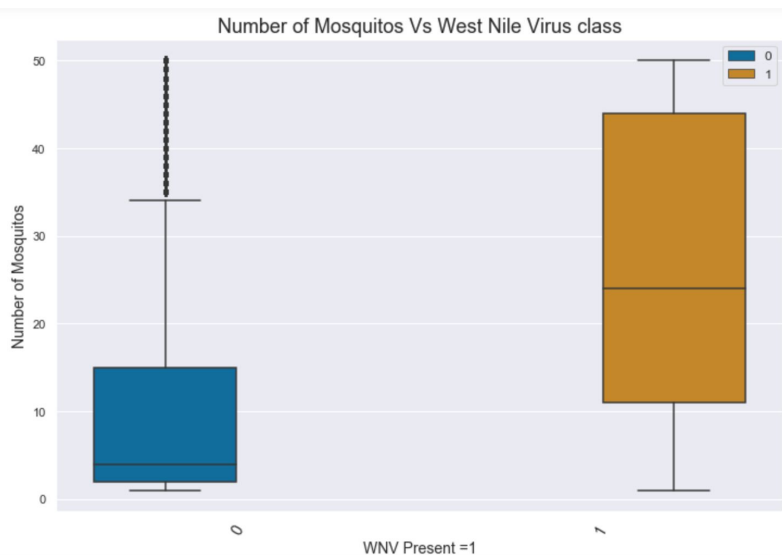


Wnv Infection



# Train data

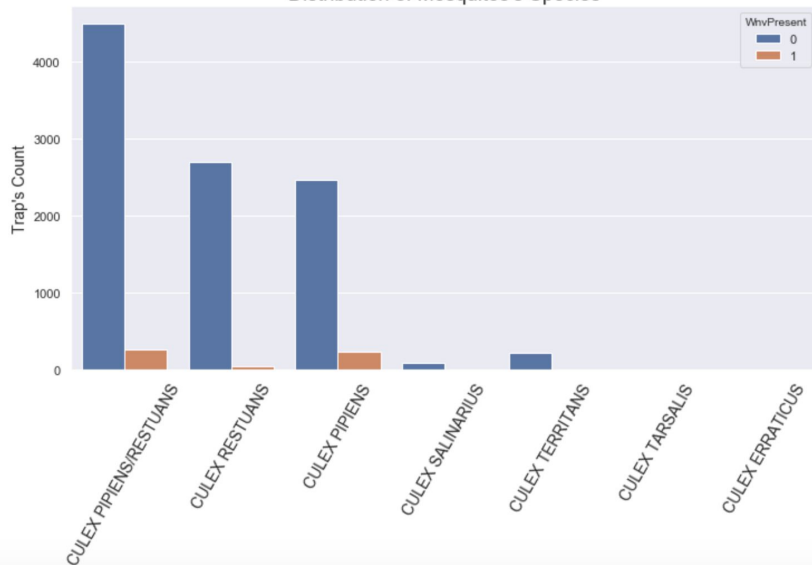
Number of Mosquitos



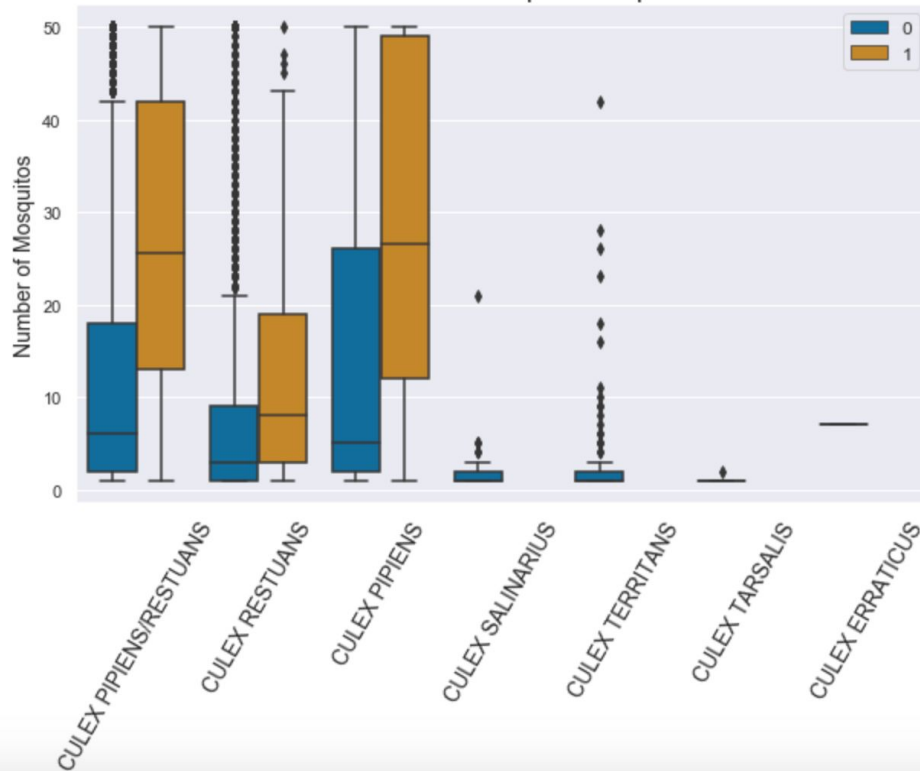
# Train data

Species

Distribution of Mosquitos's Species

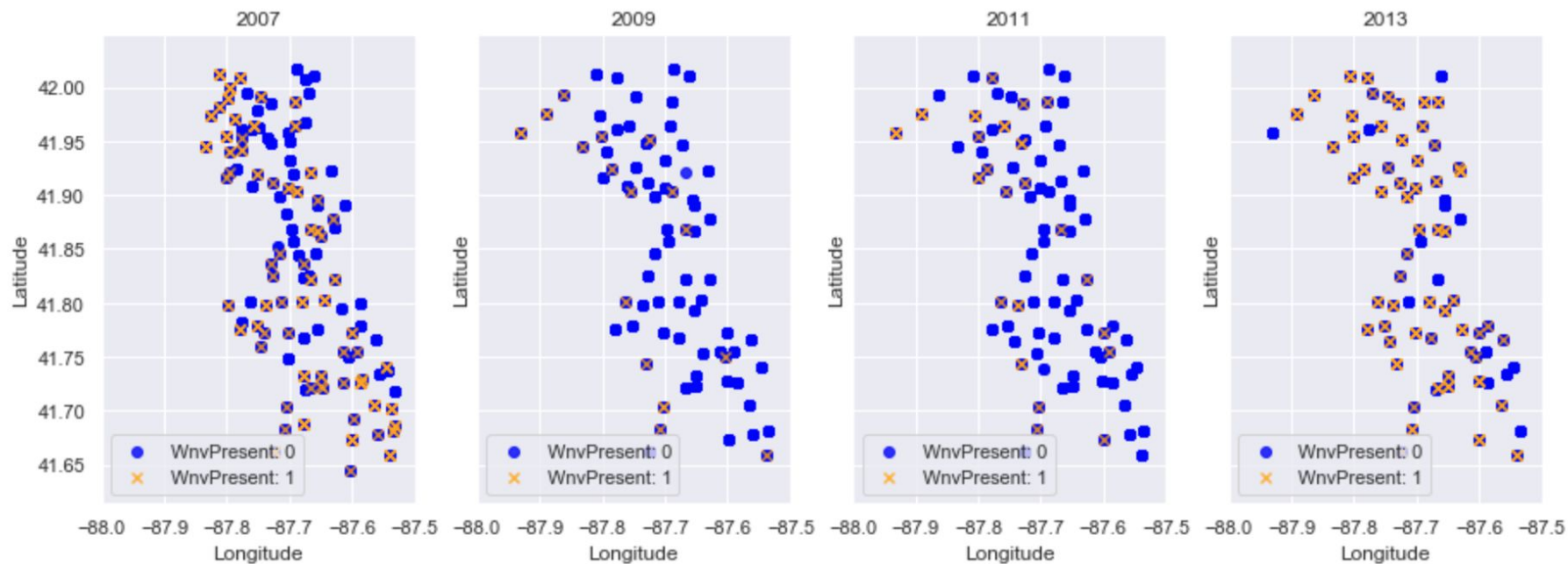


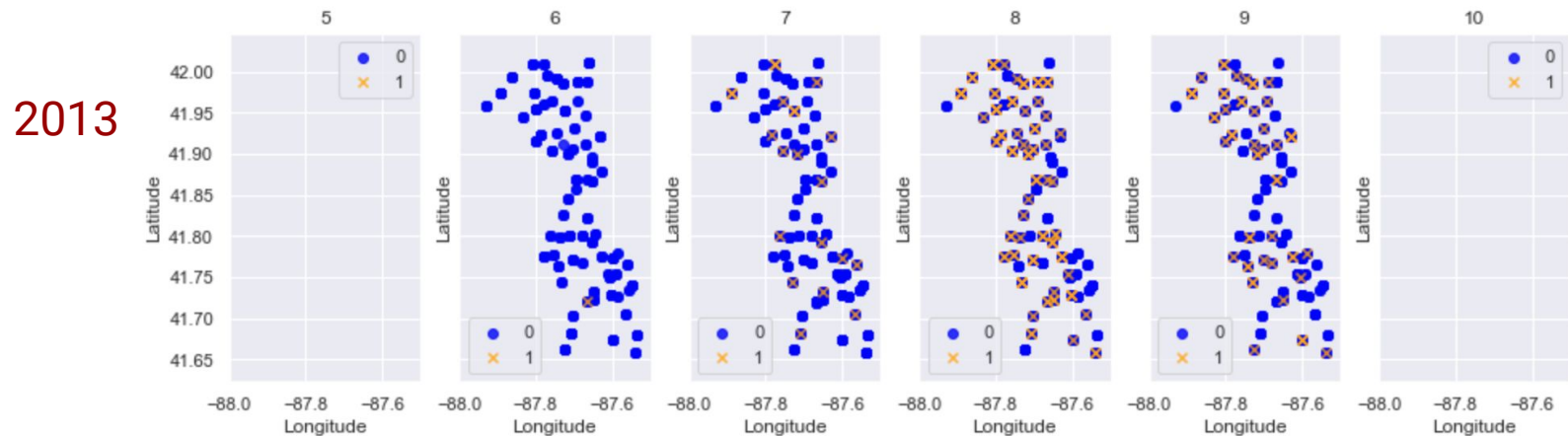
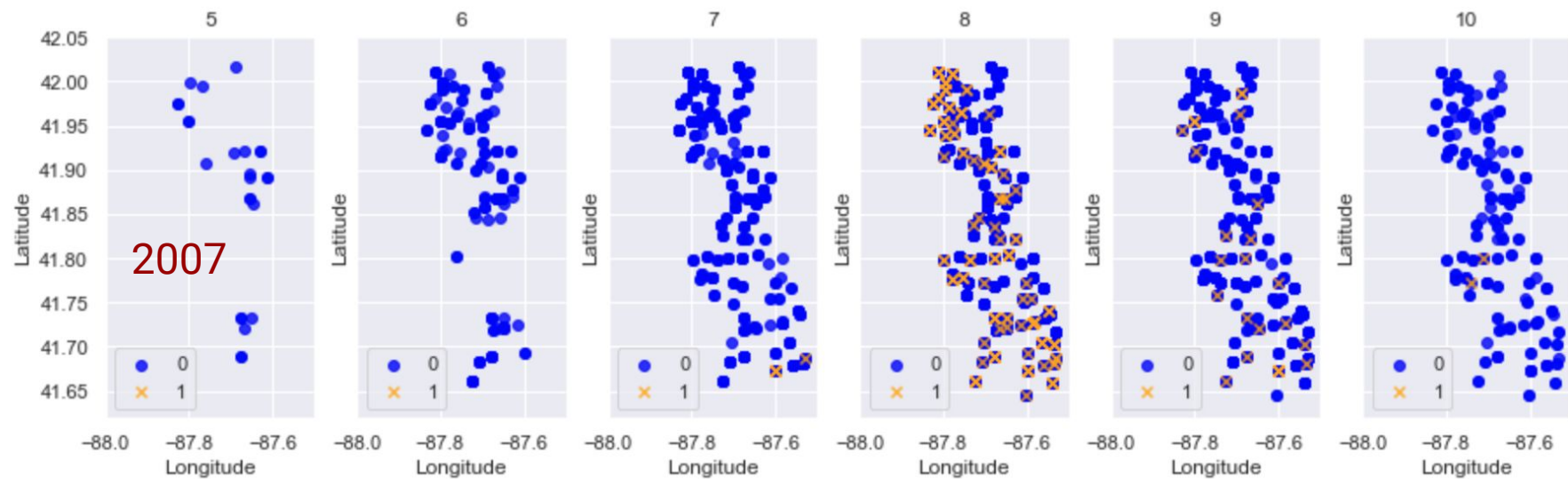
Distribution of Mosquitos's Species



# Train data

Location



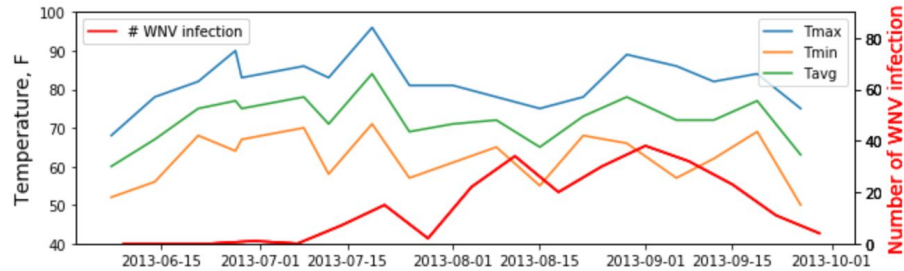
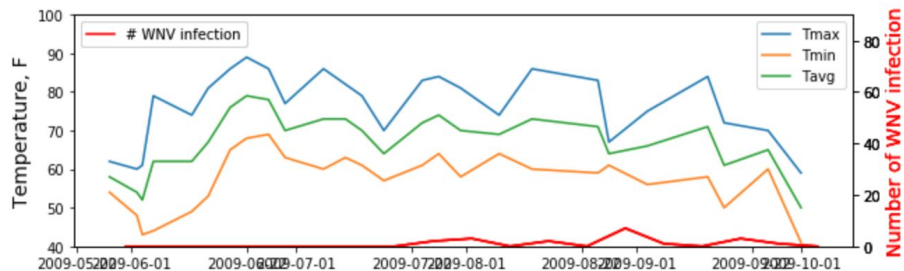
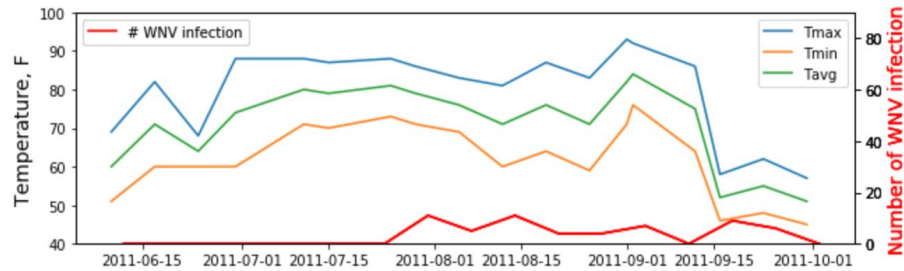
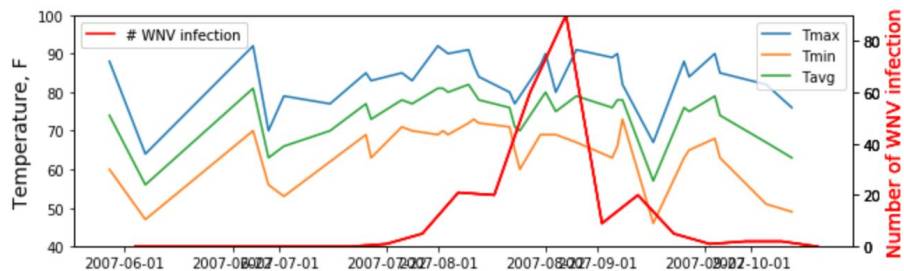




# Weather data

It is believed that hot and dry conditions are more favorable for West Nile virus than cold and wet.

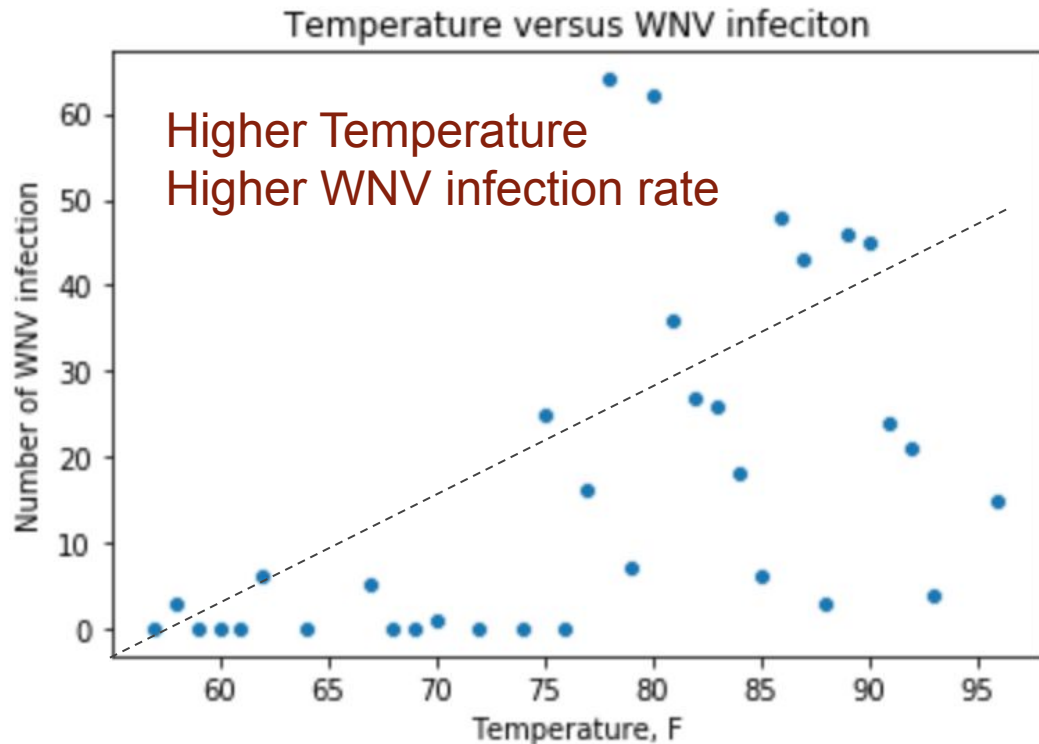
## Temperature



# Weather data

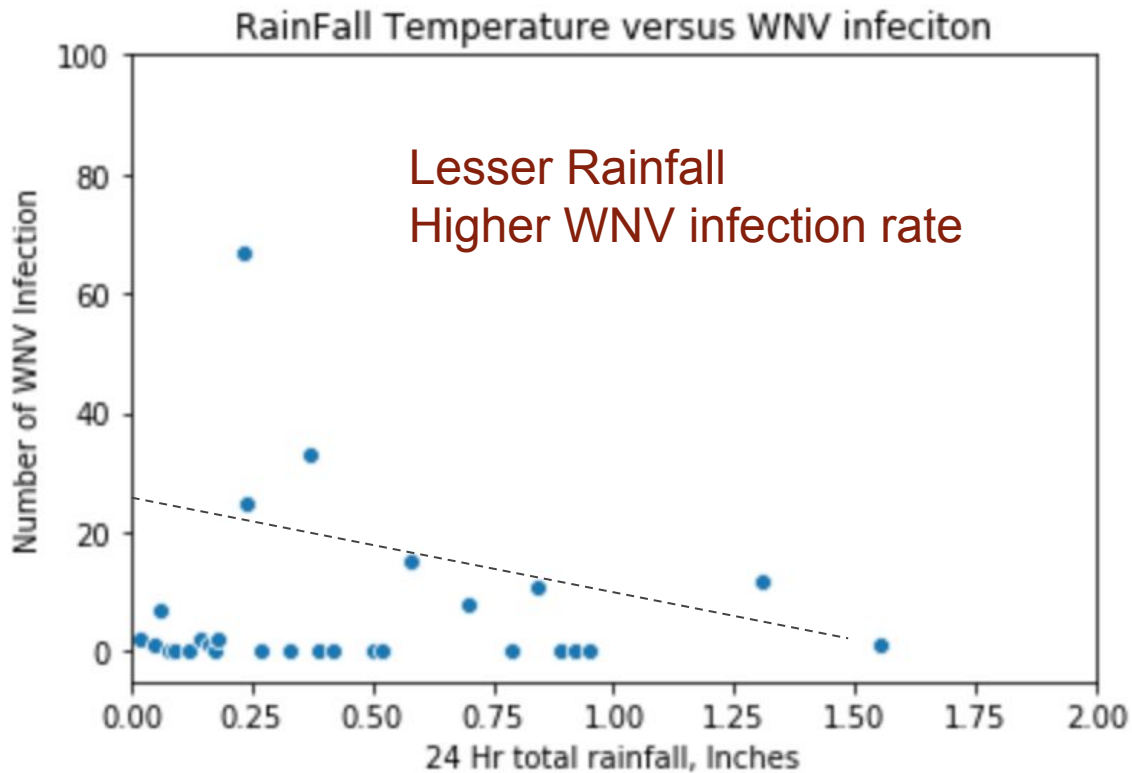
## Temperature

It is believed that hot and dry conditions are more favorable for West Nile virus than cold and wet.



# Weather data

## Rainfall

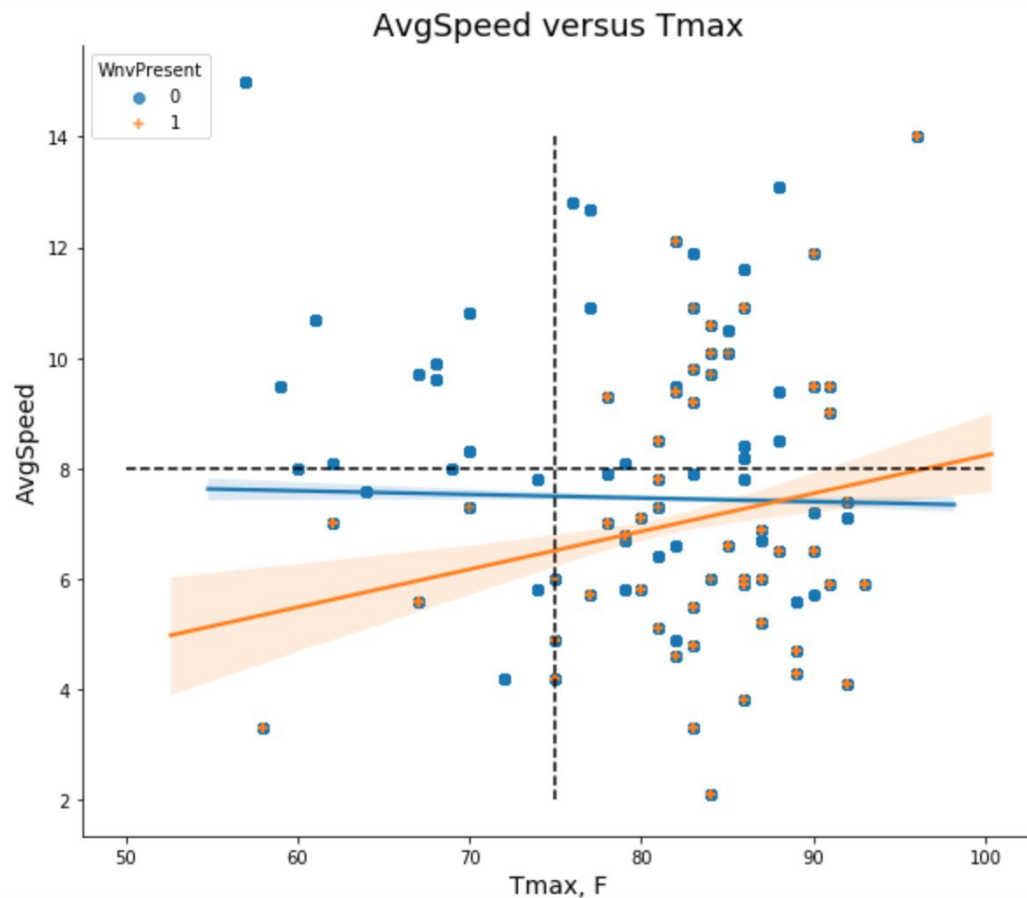


# Weather data

Temperature



Wind Speed

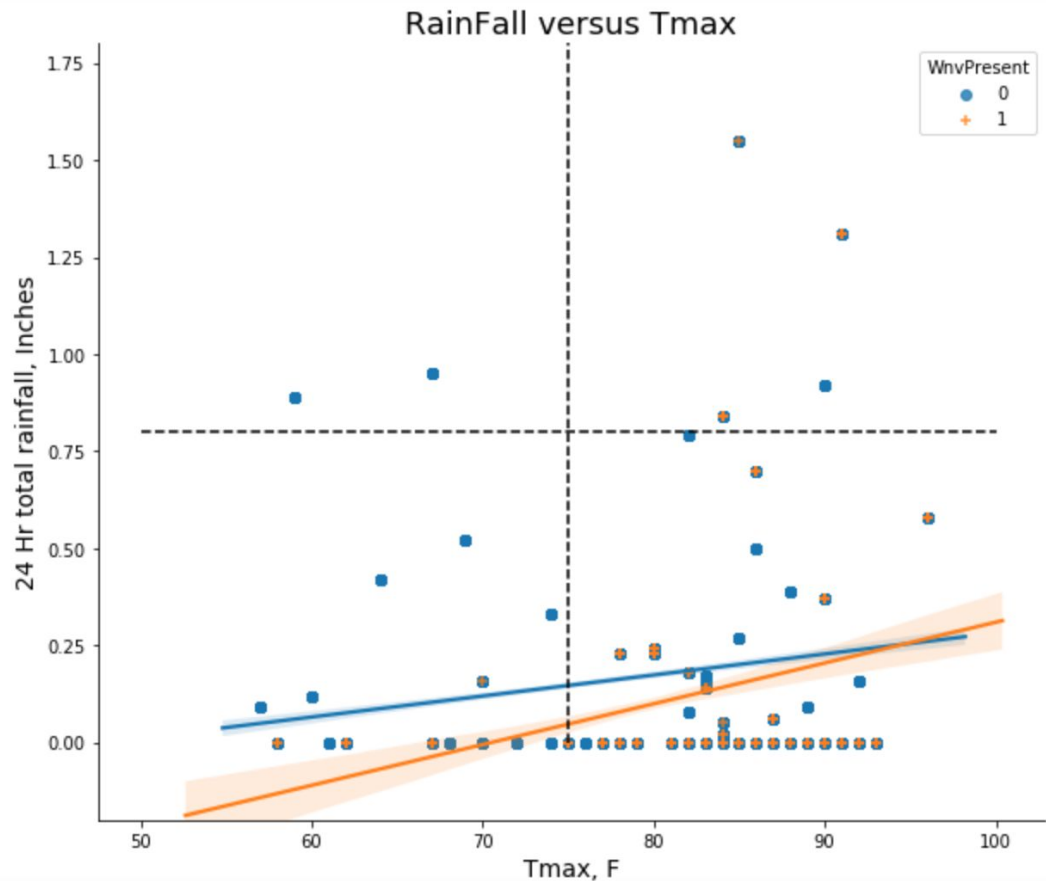


# Weather data

Temperature



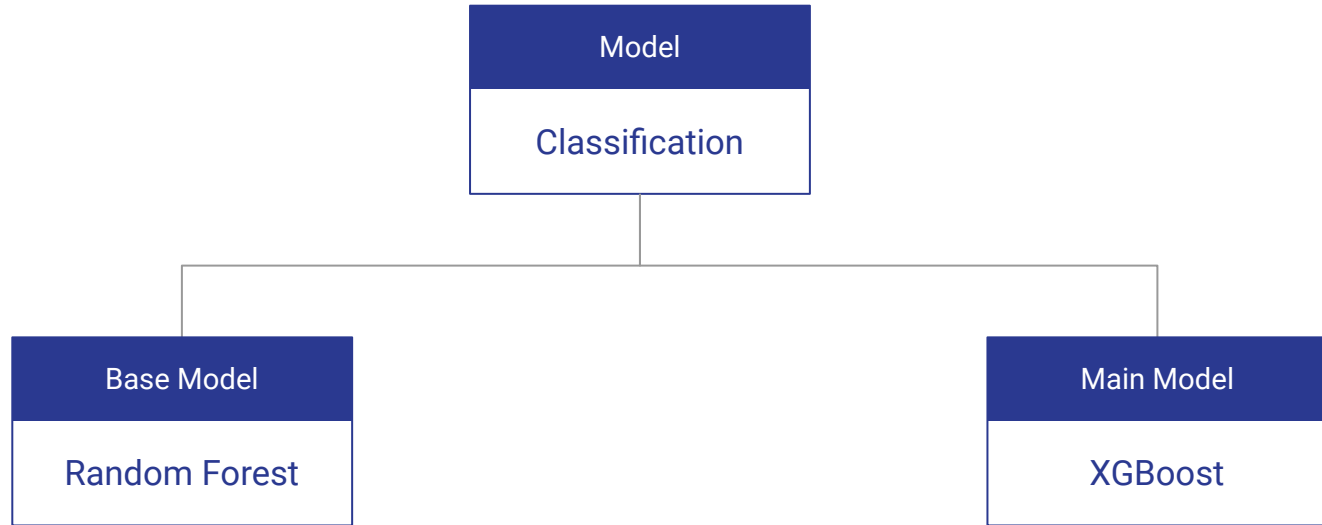
Rainfall



# Predictive Model

# Predictive Model

WNV present (1) or NOT present(0)



# Model Performance

Classifier Model	Accuracy	roc_auc	Recall	Kaggle roc_aucvscore
RandomForest w GridSearchCV	81.5%	81.5%	83.1%	63%
RandomForest w feature engineering	98.3%	98.5%	99.7%	70%
XGBoost	84.1%	91.3%	91.1%	70.7%
<b>XGBoost w GridSearchCV</b>	<b>94.7%</b>	<b>98%</b>	<b>99.4%</b>	<b>75.9%</b>
XGBoost w feature engineering	94.6%	98%	99.4%	75%

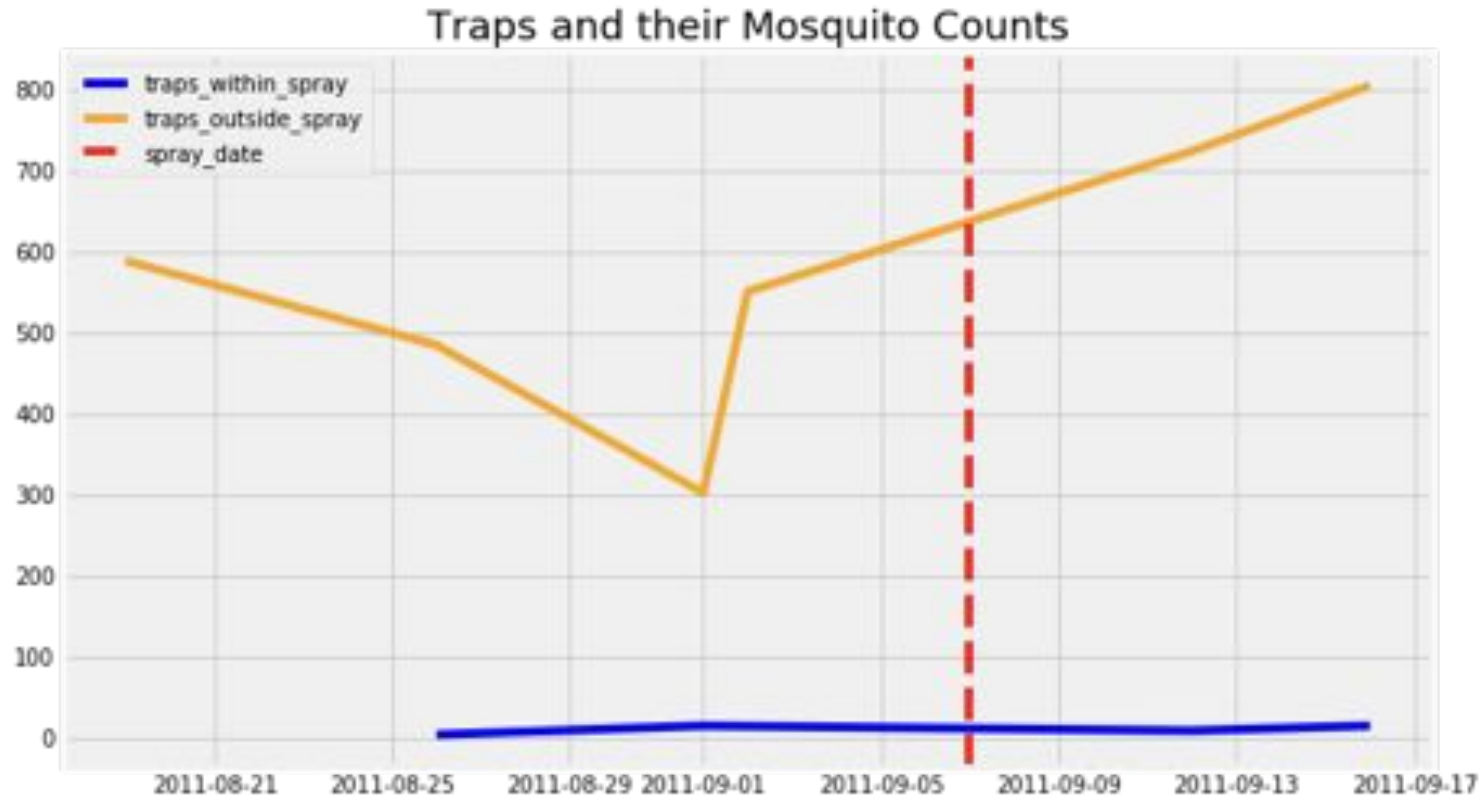


# Confusion Matrix

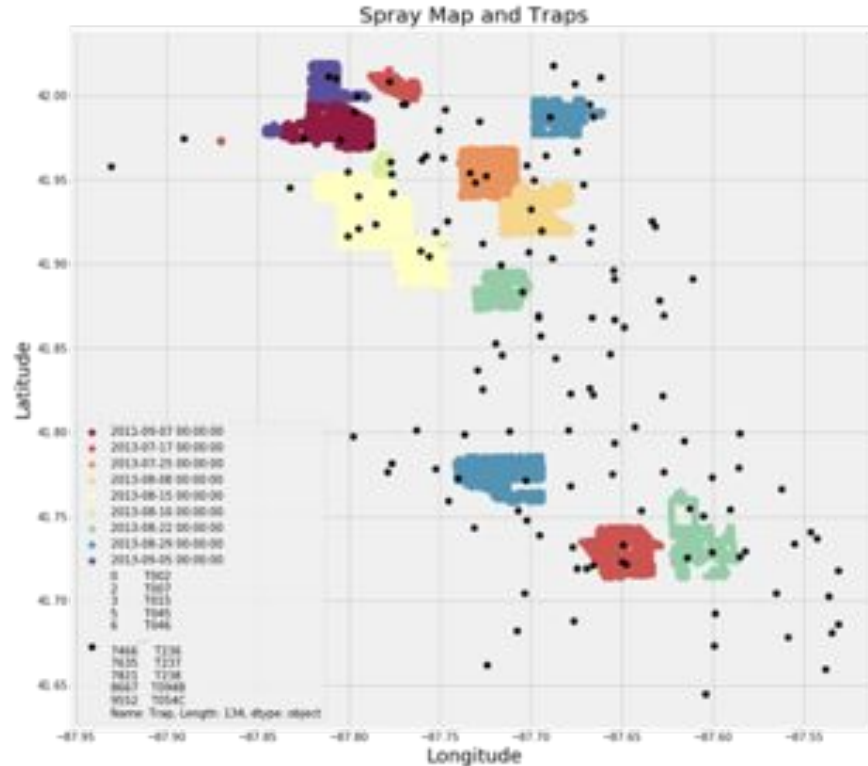
(XGBoost w GridSearchCV)

TP	3316	<ul style="list-style-type: none"><li>• Correctly predicted as West Nile Virus carrier.</li></ul>
TN	2963	<ul style="list-style-type: none"><li>• Correctly predicted as non West Nile Virus Carrier</li></ul>
FP	331	<ul style="list-style-type: none"><li>• Mispredicted someone as a West Nile Virus carrier.</li><li>• TYPE I error.</li></ul>
FN	21	<ul style="list-style-type: none"><li>• Mispredicted someone as a non West Nile Virus carrier.</li><li>• Type II error.</li></ul>

# Conclusion – XG Boost Works! Need more Data!



# Future – Better Geo-Libraries, ‘Dist from Spray Border’, Time-lag Stationarity Tests!



# Other Recommendations – 7 Day Cycle

