West Nile Virus in Chicago

Preventing outbreaks through predictive analysis

Chin Xia Liew
Thomas Lim
Clarence Thng

About the disease
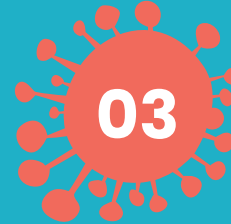
01

West Nile Virus (WNV)

Belongs in the same family as Zika, Dengue, and Yellow Fever
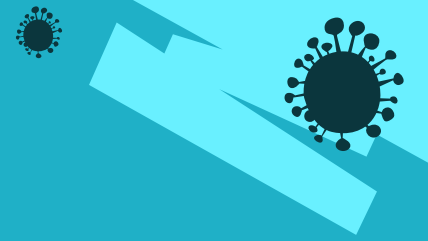
02

United States

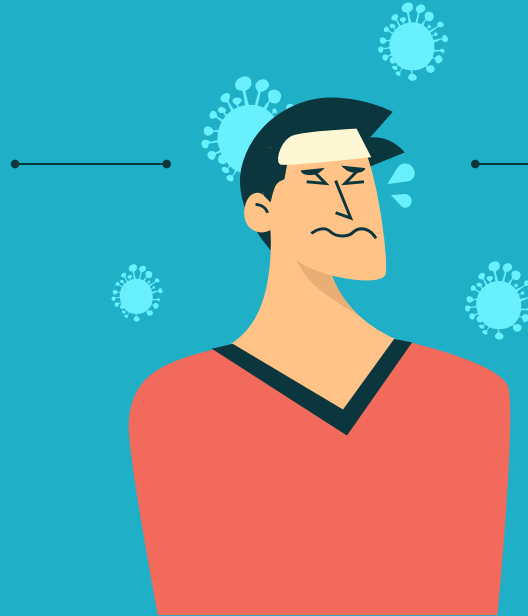Leading mosquito-borne diseases

03

Potentially Fatal

About 1 in 150 people develop severe symptoms, and 1 in 1500 die

Effects of the disease

Central Nervous System

Neck Stiffness, Stupor, Disorientation, Coma, Tremors, Convulsions, Muscle Weakness, Vision Loss, Numbness, Paralysis

Febrile illness

Fever, Headaches, Body Aches, Joint Pains, Vomiting, Diarrhoea, Rash, Fatigue

The City's Efforts

Mosquito Surveillance
Checking mosquito populations for WNV presence

Population Control
Spraying efforts
Larvicide treatment

Effective Resource Allocation
Predict WNV probability for given location, time and species
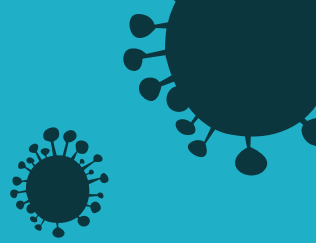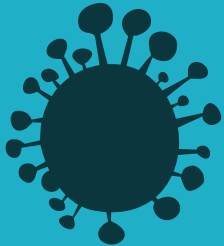
Table of contents

01.
Data Processing

Data Cleaning

Lowercasing

01

Dropping columns

04

Missing / Trace values

02

Consistent data-types

05

Imputing values

03

Removing outliers

06
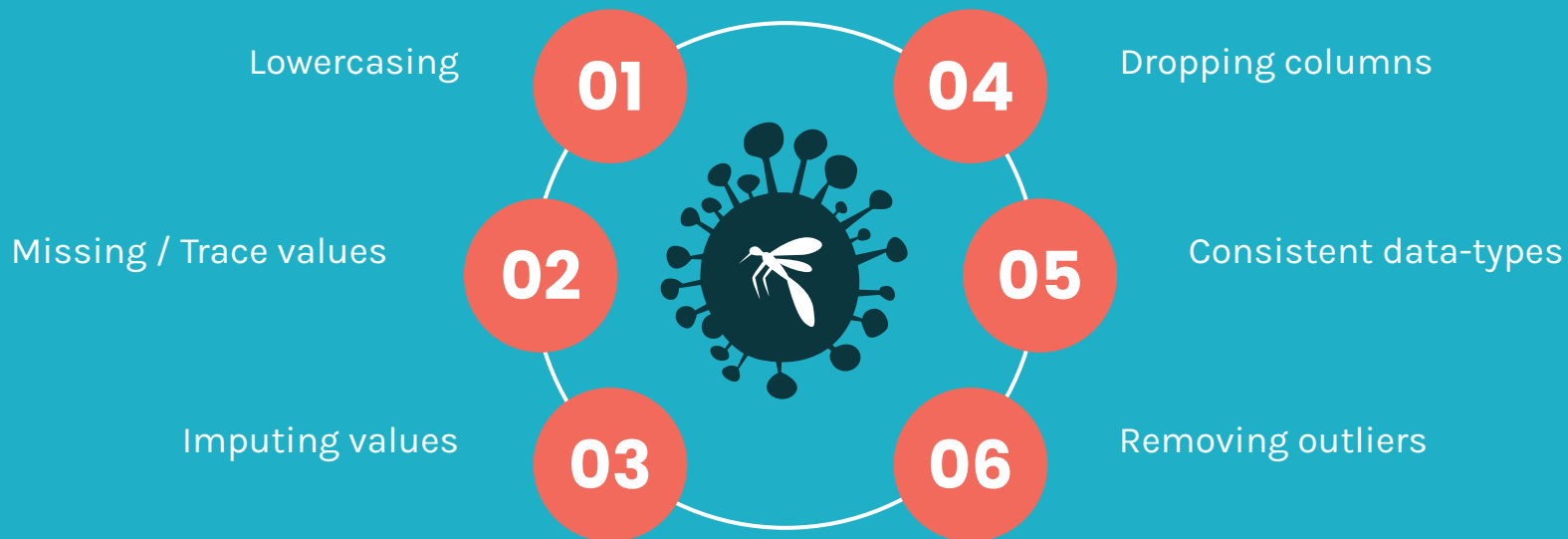
Feature Engineering

01

Relative Humidity

Based on temperature, dewpoint, and pressure

02

Average by days

Averaged data for 2 stations for each date

03

Dummify

Trap, Weather Phenomena, Species

04

(Data) Oversampling

To combat imbalance classes

02.
Exploratory Data Analysis

EDA Summary

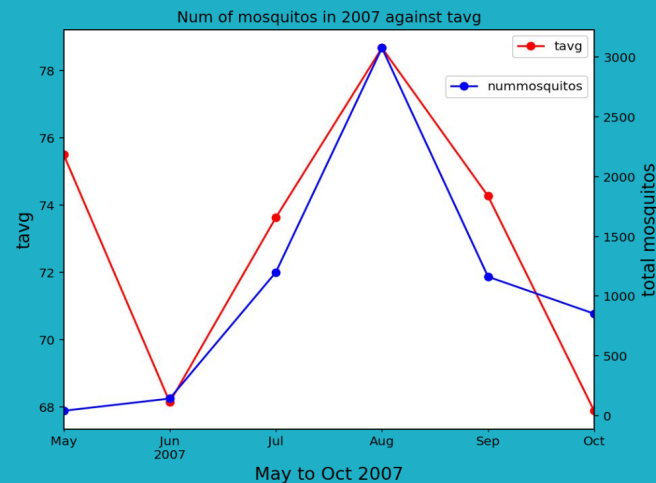High correlation w/ Number of Mosquitos and weather features
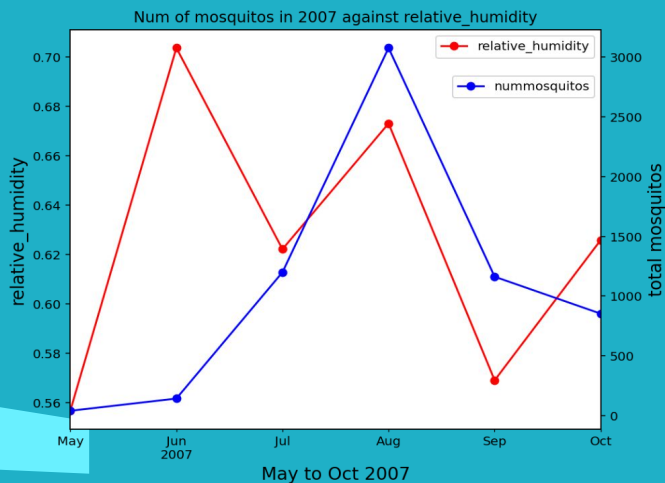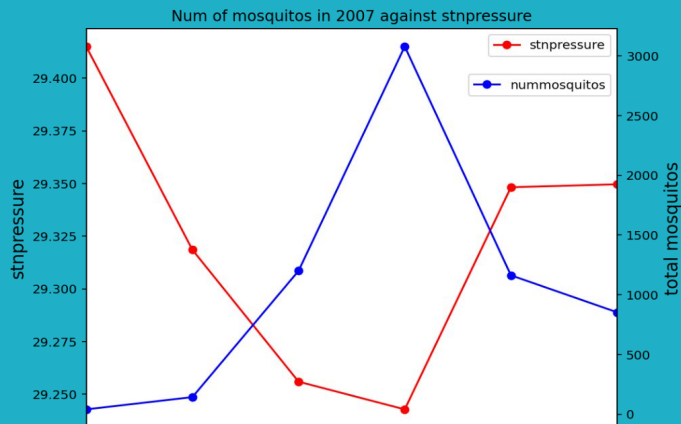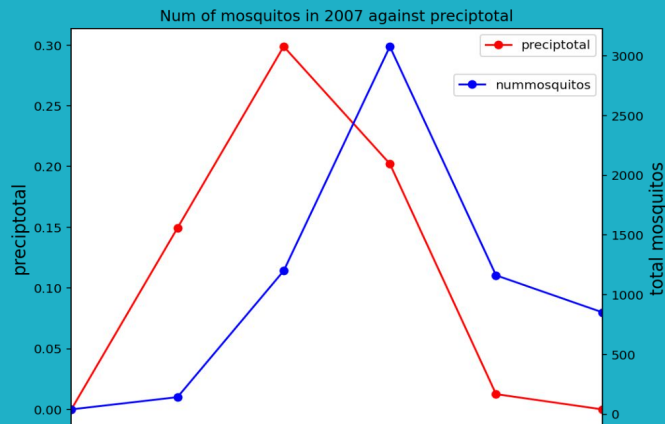
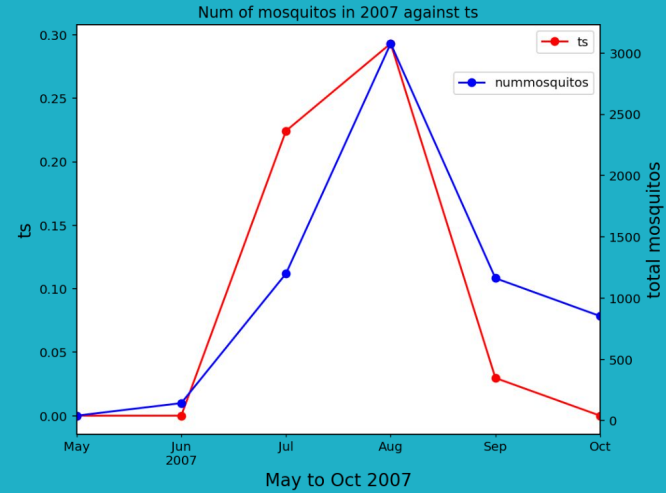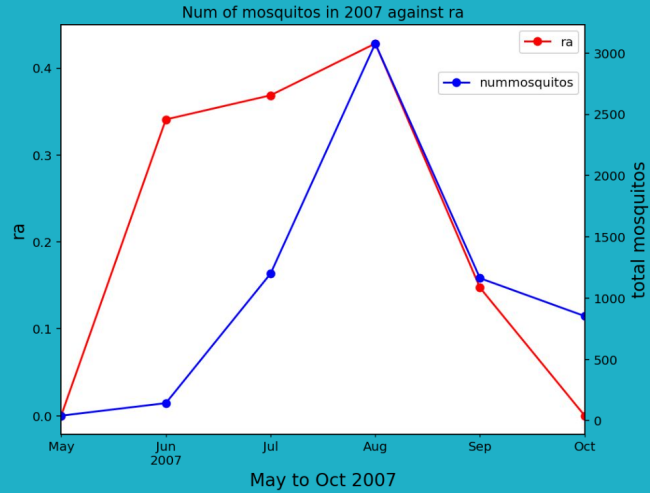Mosquito population peaks between July - Aug

Higher population of two species in particular

Higher population in certain locations

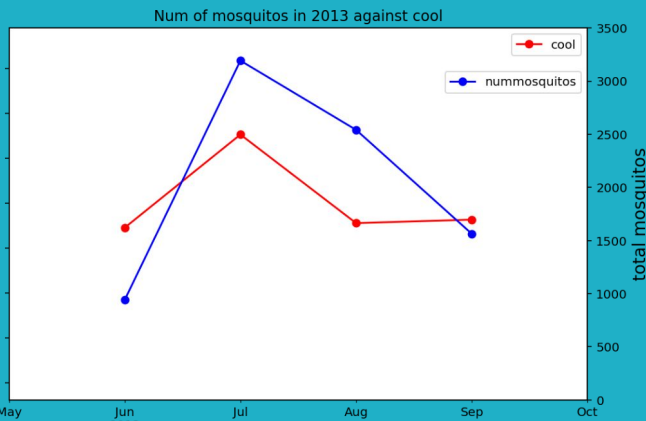# Features highly correlated to number of mosquitos

Other features

Num of mosquitos in 2007 against ra

May to Oct 2007

Num of mosquitos in 2007 against ts

May to Oct 2007

Mosquito numbers peak between Jul to Aug
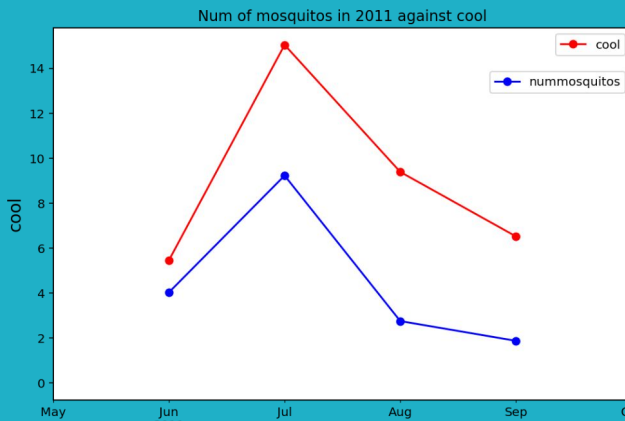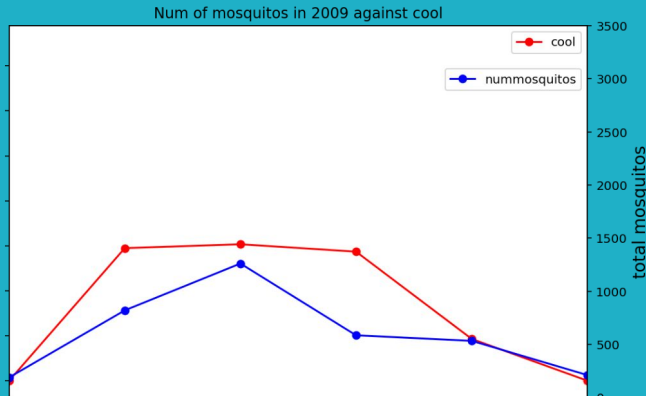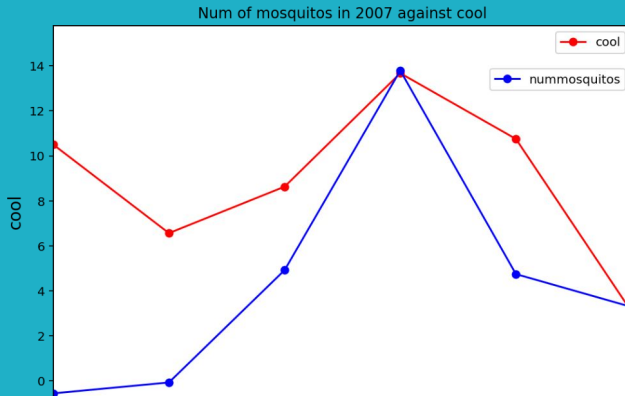
Culex Pipiens and Restuans were more common

Number of mosquitos of different species with or without the presence of west nile virus across 2007 to 2013

Some locations saw higher mosquito populations

Top locations containing more than 200 mosquitos with west nile virus across 2007 to 2013

Spraying can be effective but they were not well targeted
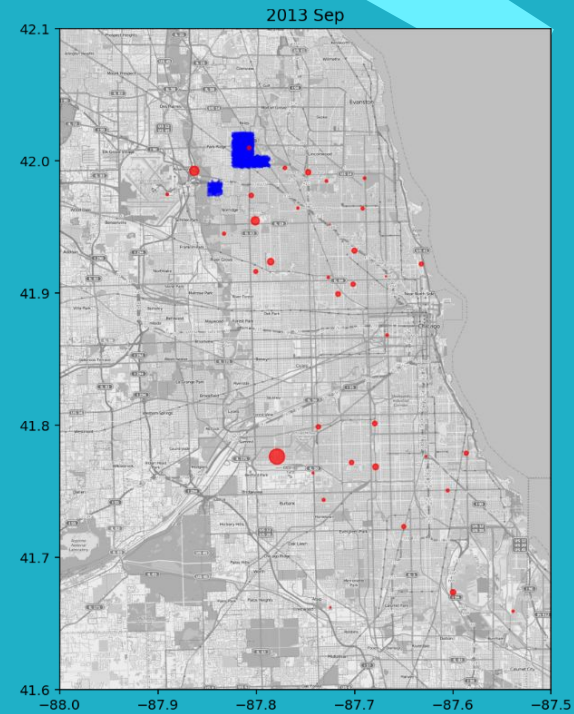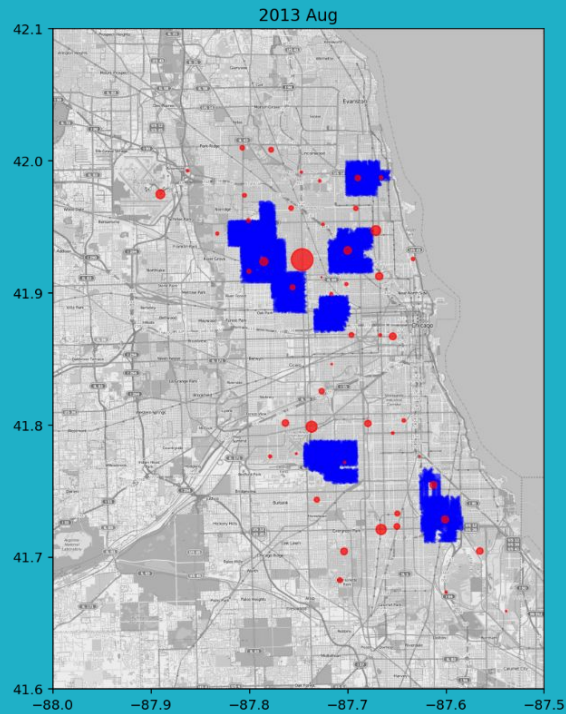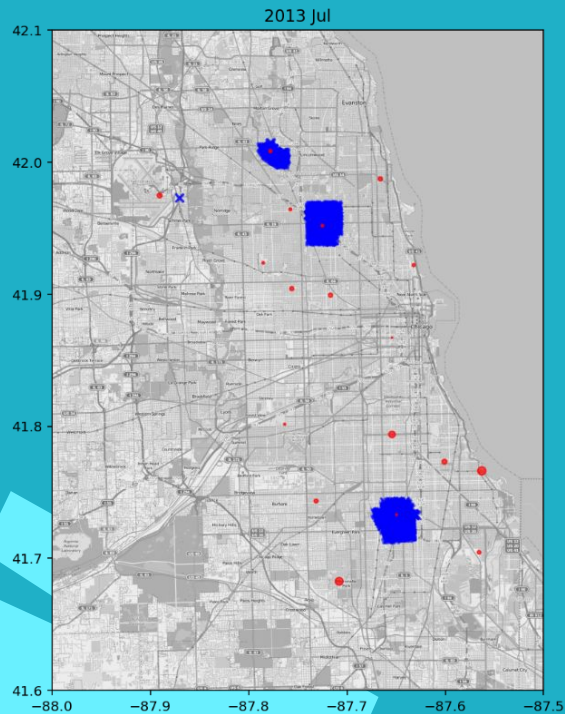
03.

Modelling

LogReg, kNN, RandomForest

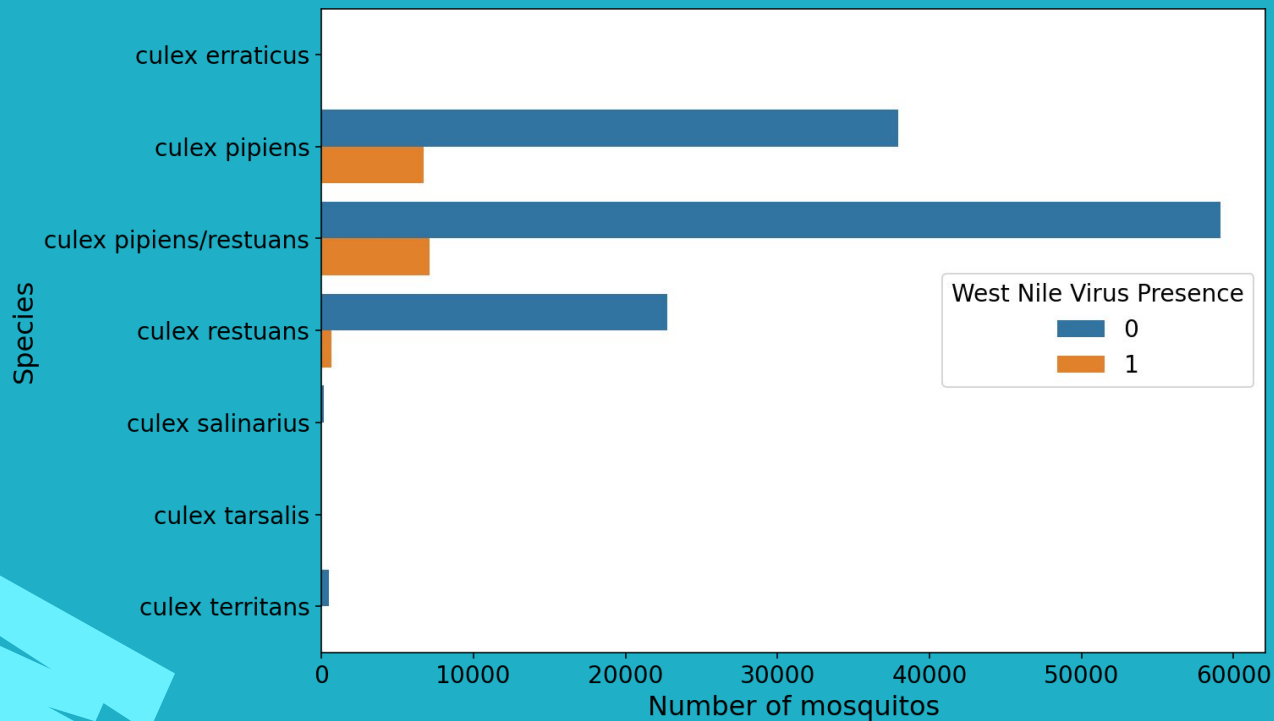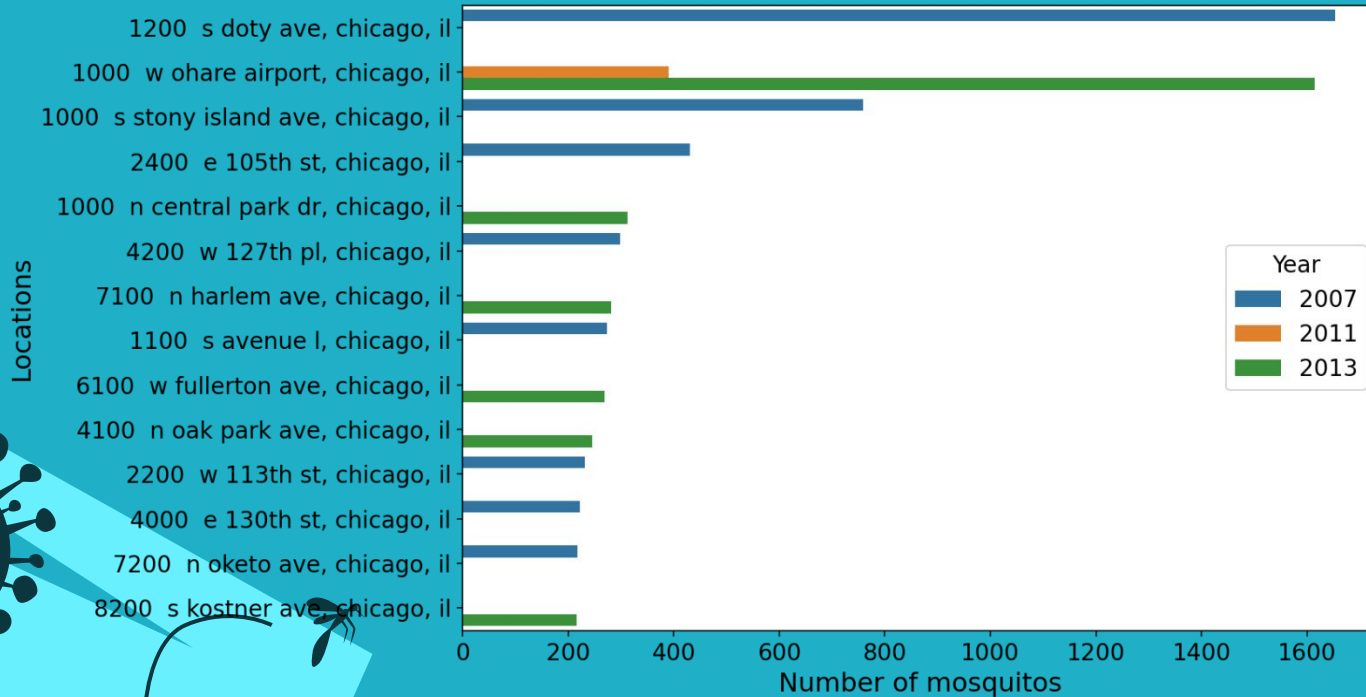| Estimator	Scores
Null Model	Score: 94.8% Score (oversampled): 50%
Logistic Regression	Precision Score: 94% Recall Score: 94% F1 Score: 94% AUC Score: 98%
k-Nearest Neighbours	Precision Score: 95.5% Recall Score: 95% F1 Score: 95% AUC Score: 97%
Random Forest Classifier	Precision Score: 96% Recall Score: 96% F1 Score: 96% AUC Score: 99%

k-Nearest Neighbours

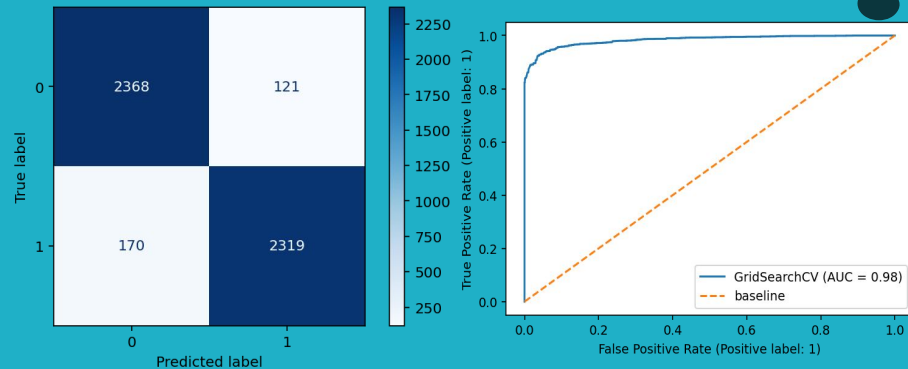|	precision	recall	f1-score	support
0	0.96	0.94	0.95	2489
1	0.94	0.96	0.95	2489
accuracy			0.95	4978
macro avg	0.95	0.95	0.95	4978
weighted avg	0.95	0.95	0.95	4978

Logistic Regression

	precision	recall	f1-score	support
0	0.93	0.95	0.94	2489
1	0.95	0.93	0.94	2489
accuracy			0.94	4978
macro avg	0.94	0.94	0.94	4978
weighted avg	0.94	0.94	0.94	4978

Random Forest
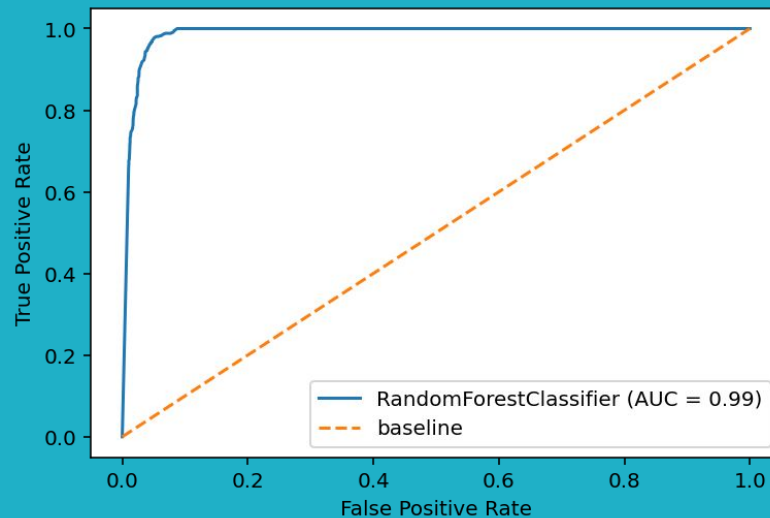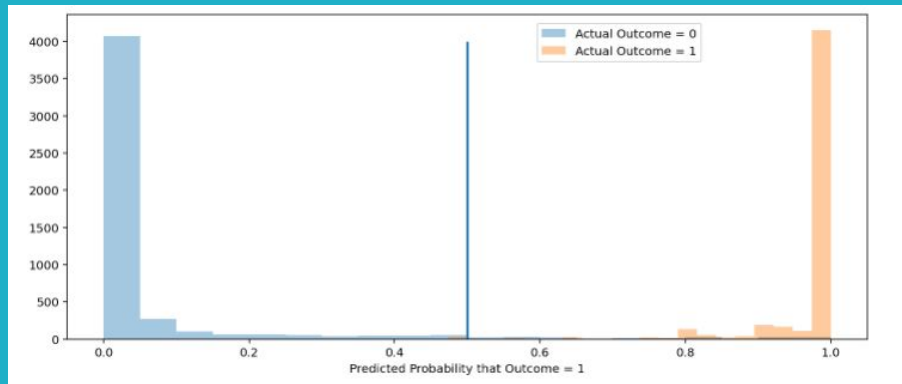
	precision	recall	f1-score	support
0	0.99	0.93	0.96	2489
1	0.93	0.99	0.96	2489
accuracy			0.96	4978
macro avg	0.96	0.96	0.96	4978
weighted avg	0.96	0.96	0.96	4978

Random Forest – important features

Feature importances using Mean Decrease in Impurity
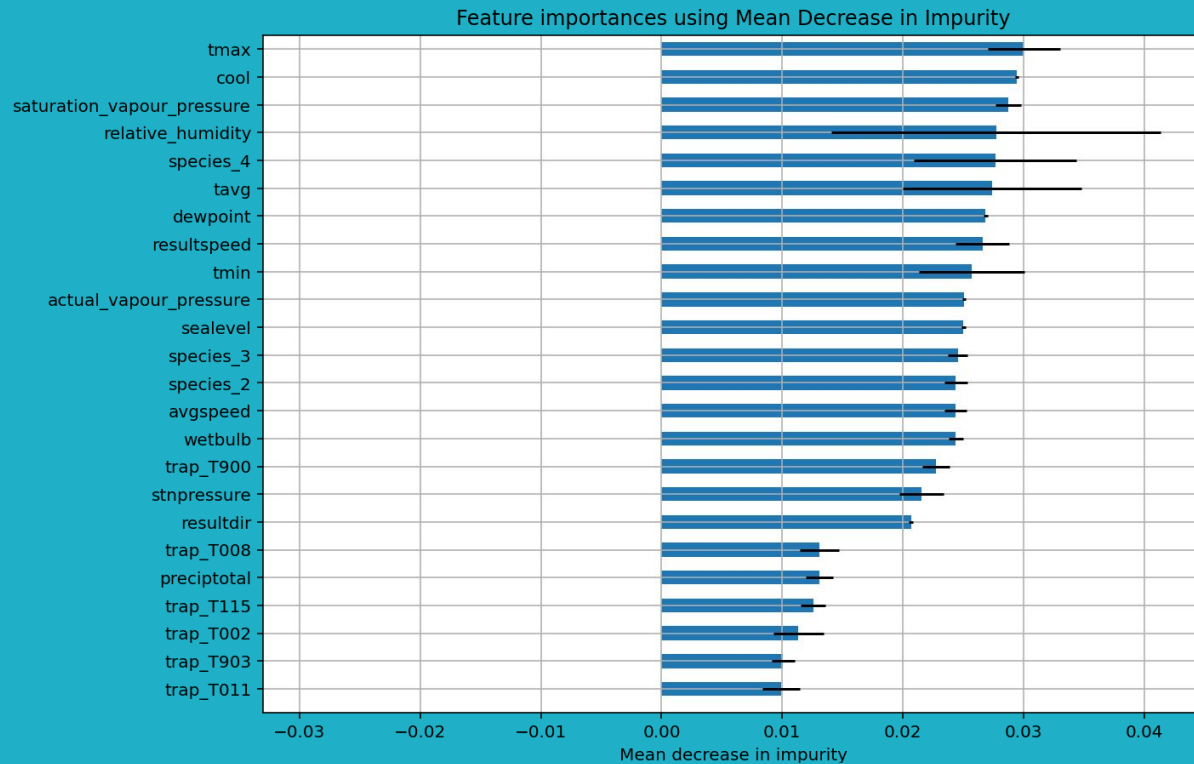
Cost-Benefit Analysis

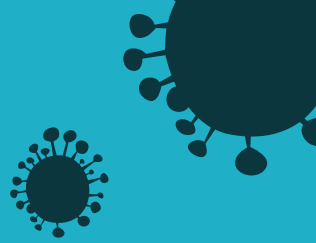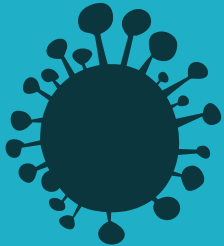| Prediction	No. of sprays	Cost
Without model	16000	$8,000,000
With model	8450	$4,225,000

cost : $500/spray

No. of traps: approx. 130

Spray period: July to September

04.
Conclusions & Recommendations

Conclusions

- Number of mosquitos correlated to weather conditions and WNV presence

- Spraying can be effective, but the locations sprayed previously were not well targeted

- Weather and location are important

- Random Forest worked best and does well for our requirements

- The model has an F1 score of 96%, keeping false positives and negatives at a minimum.

Recommendations

Don't

- Randomly perform spraying

Do

- Continue the spraying efforts
- Target the spraying around certain traps
- Target the spraying during / after specific weather conditions

Thank You!

Do you have any questions?