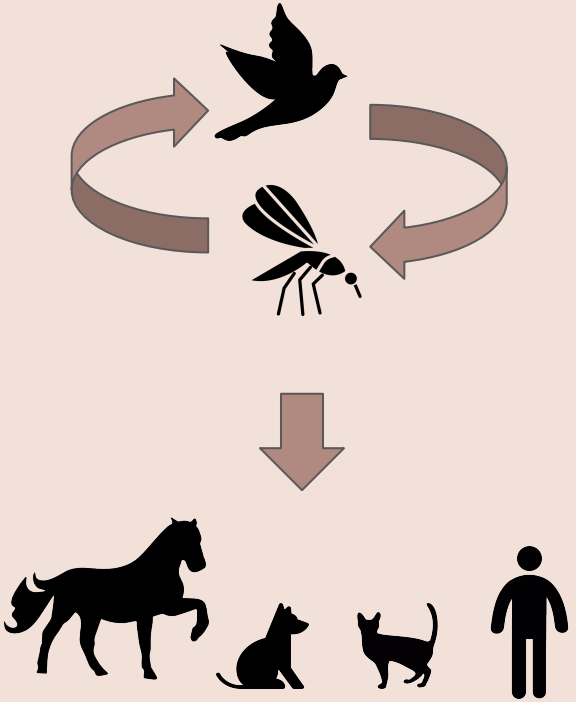
Predicting West Nile Virus In Chicago

Lucus
Dale
Dave

What is West Nile Virus?

- West Nile virus is a mosquito-borne virus of the encephalitis group.
- West Nile Virus occurs mostly in birds.
- Mosquitoes pick the virus up when they feed on infected birds and transmit it when they feed on uninfected birds.



Background

- WNV is the leading cause of mosquito-borne disease.
- Mosquito traps across the city are tested weekly from late spring to fall.
- Results of the tests influences when and where pesticides will be sprayed.

Problem Statement

Although pesticides are known to be effective in dealing with the virus-carrying mosquitoes, it is expensive to deploy pesticides throughout the city of Chicago.

- Understand the factors driving the spread of WNV
- Develop a classification model that could predict the presence of WNV
- Suggest a cost-efficient and effective method of deploying pesticides within the area.

Findings

01

Exploratory Data Analysis

Weather and Timing were the best predictors

02

Cost Benefit Analysis

Efficacy of spraying were inconclusive

03

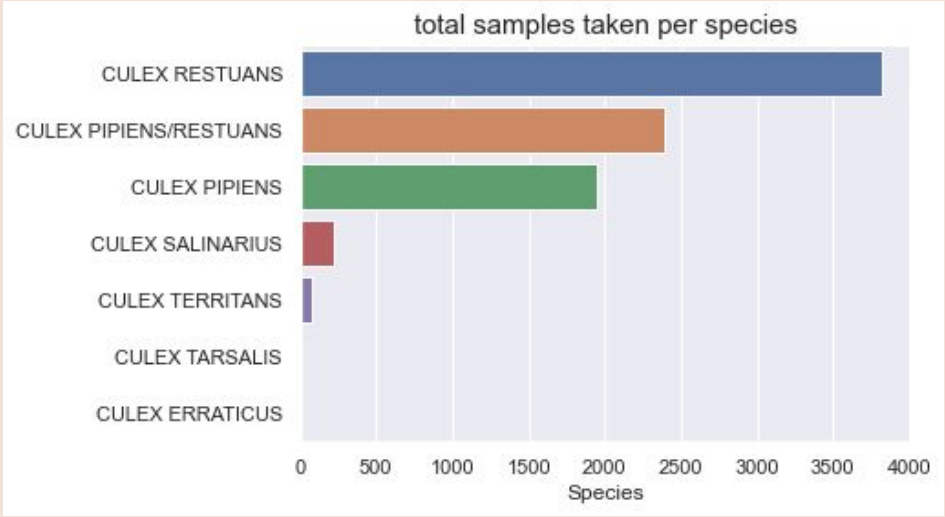
Best Model

Random Forest

01

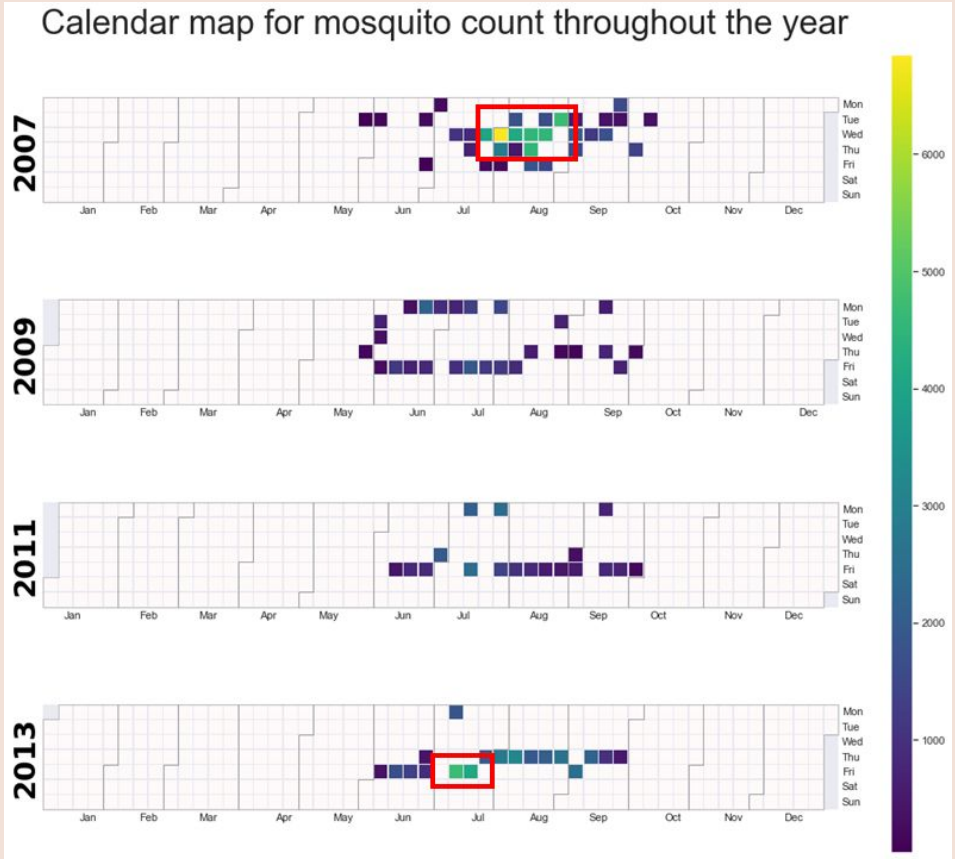
Exploratory Data Analysis

Exploratory Data Analysis



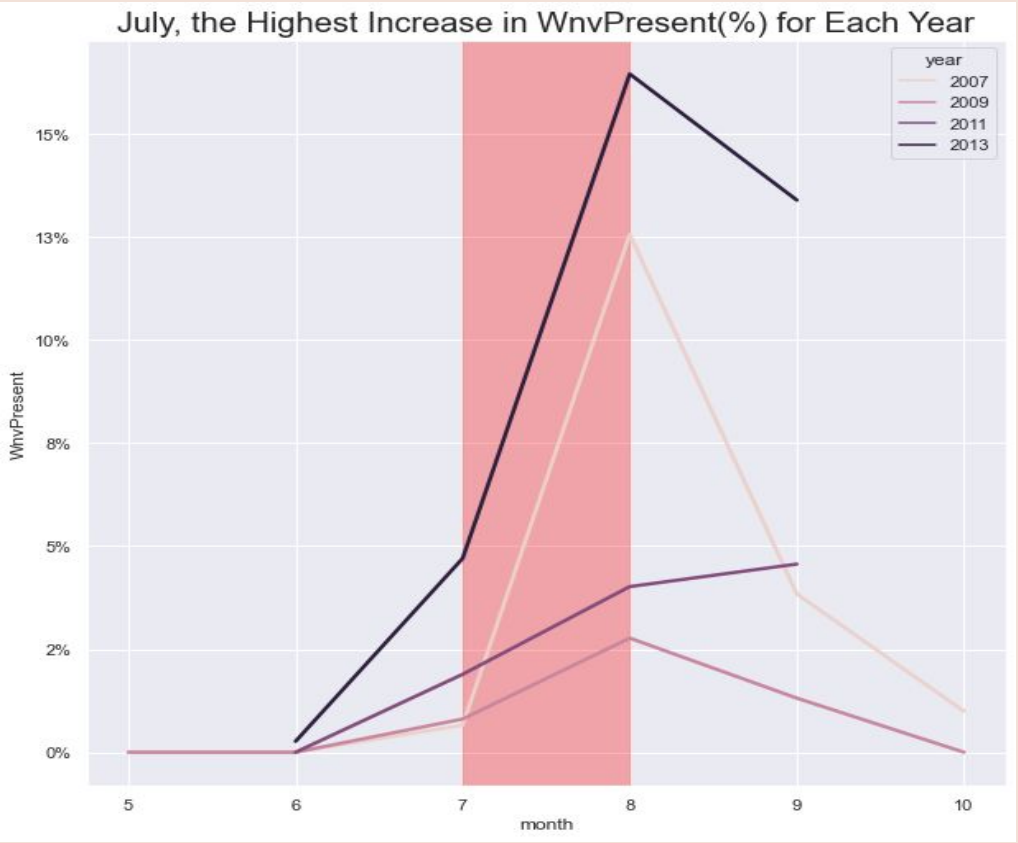
- Culex Restuans and Culex Pipiens made up of ~96% of the species.
- Identified as carriers of West Nile Virus.

Exploratory Data Analysis



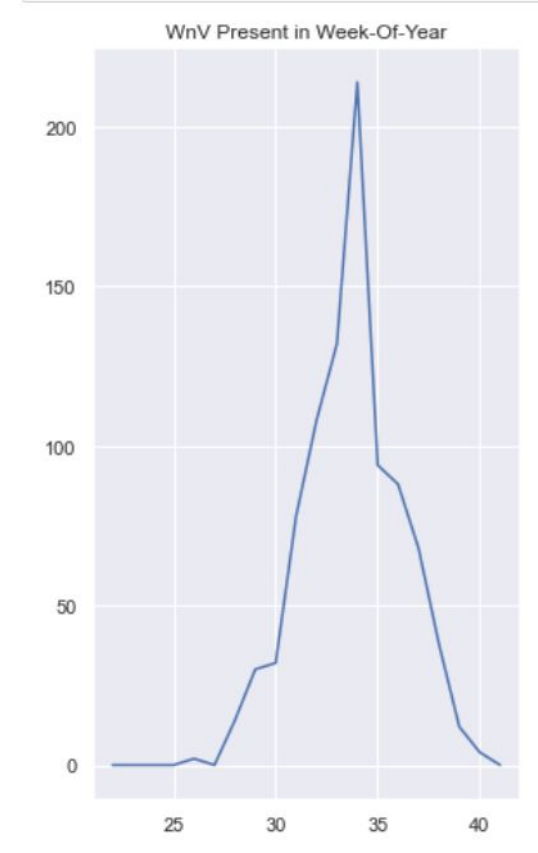
- In 2007, August have days with counts approximate 5000-6000+.
- Mosquito count have generally dropped over the years but an increased has been seen in 2013.

Exploratory Data Analysis



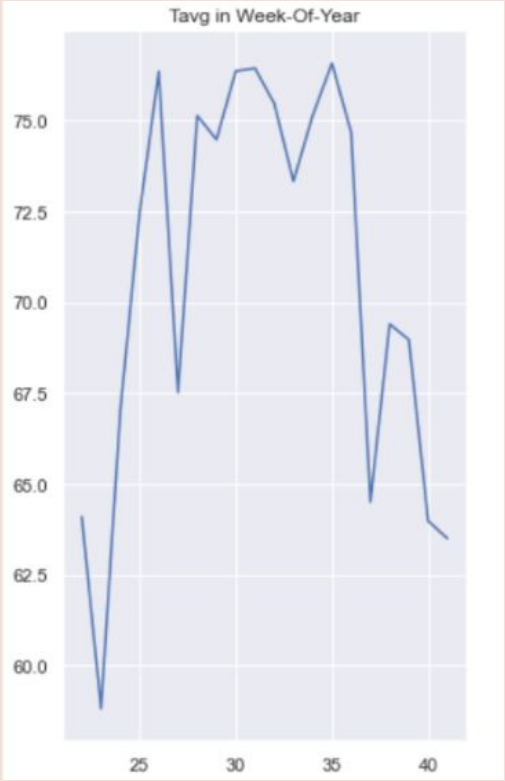
- There is a steep increase of WNV for the month of July in each year.

Exploratory Data Analysis



- WNV occurence peaks around week 33 - 34
- Chicago is having summer season

Exploratory Data Analysis

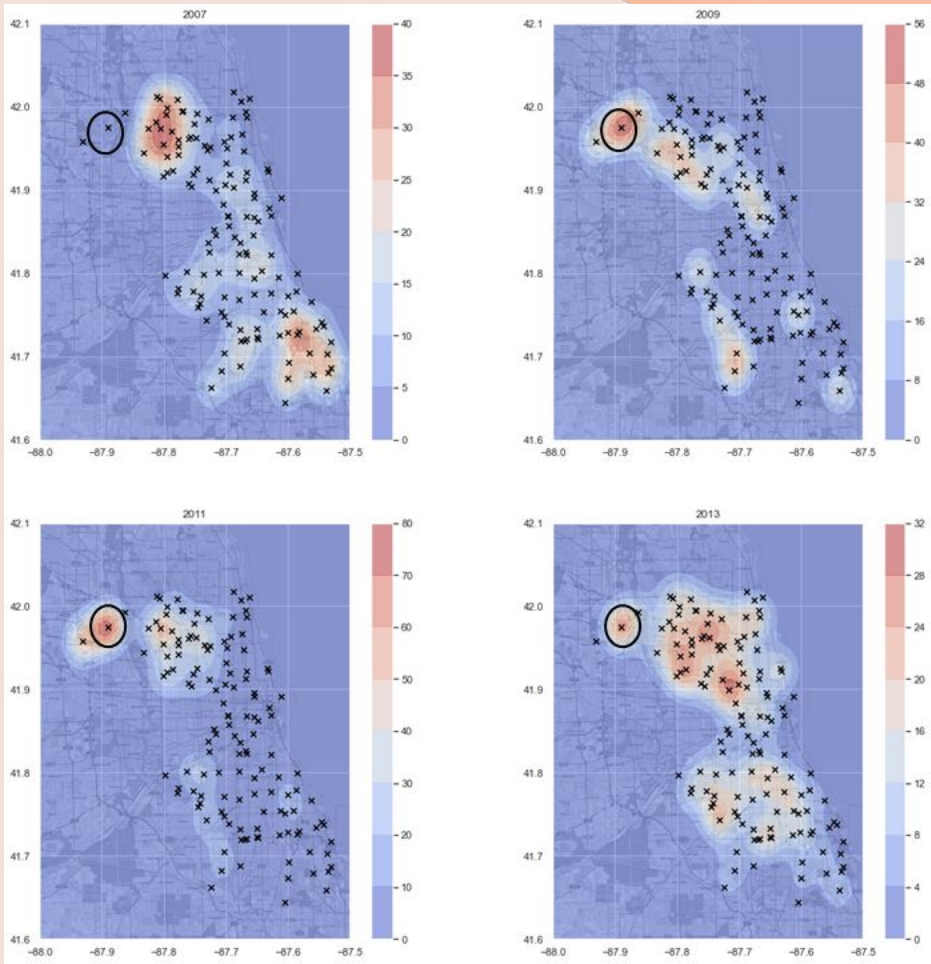


- Average temperature start to pick up at around July.

Exploratory Data Analysis

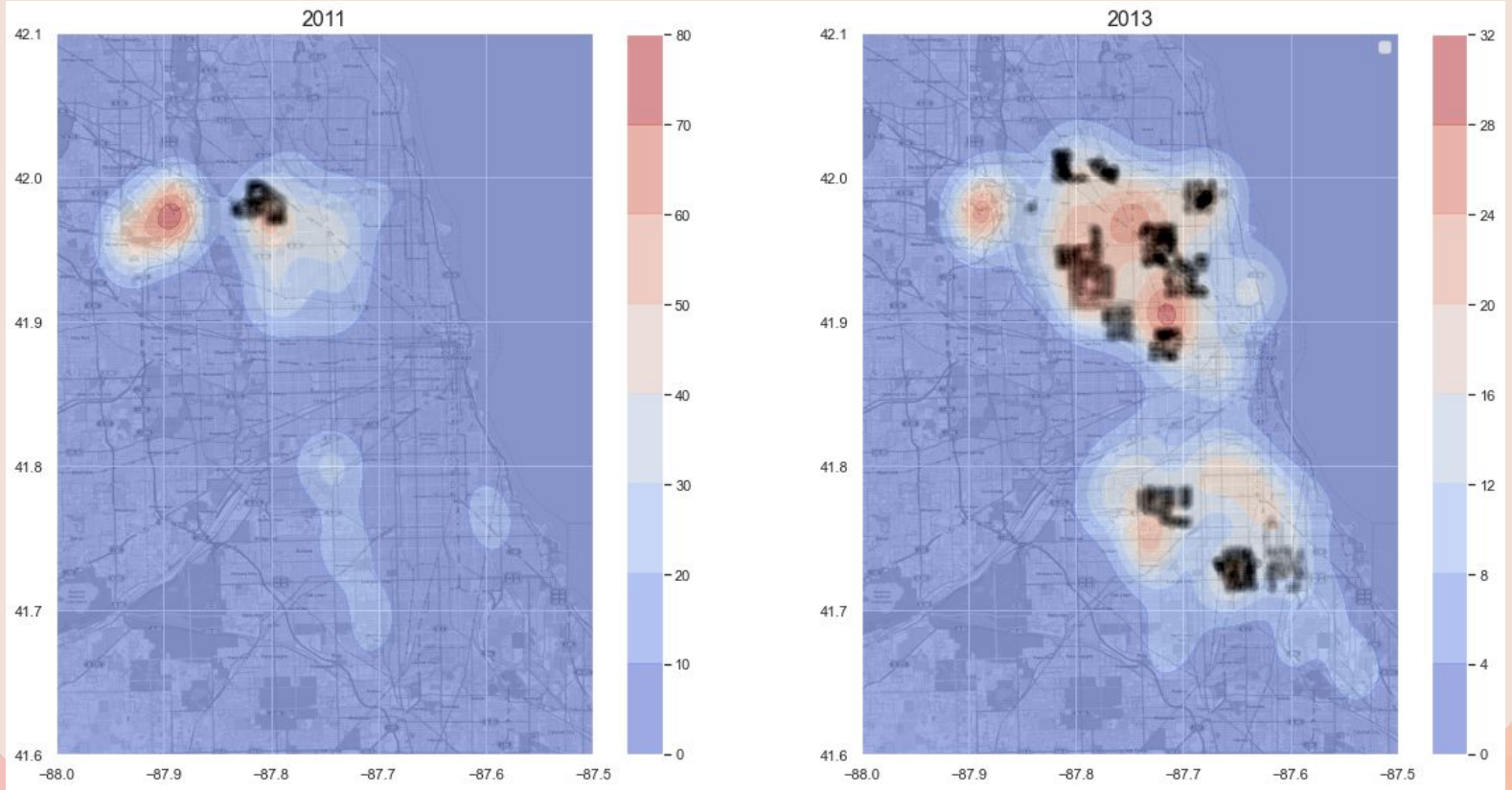
- Trap ‘T900’ has the highest WNV occurrence at 29 throughout the 2007 - 2013.
- Located near O'Hare International Airport

Trap	AddressNumberAndStreet	Latitude	Longitude	WnvPresent
T900	1000 W OHARE AIRPORT, Chicago, IL	41.974689	-87.890615	29



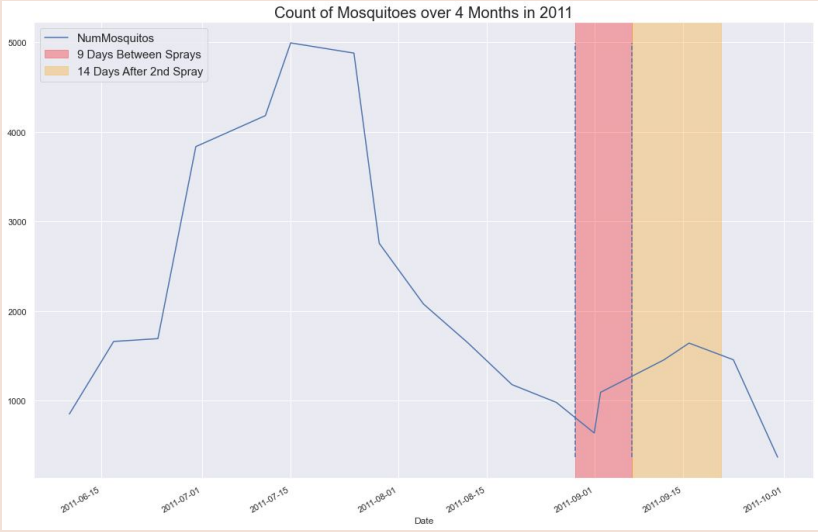
02
Cost Benefit Analysis

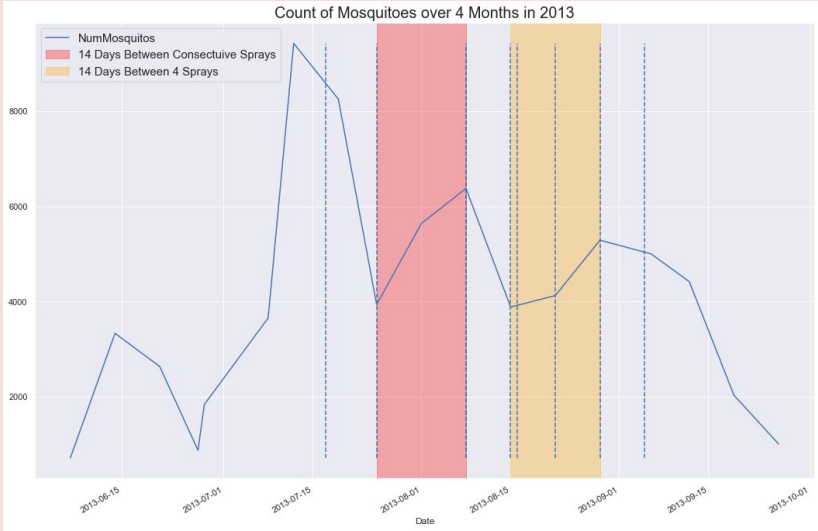
Exploring Spray Data



Unclear efficacy of Spraying

- only two separate dates in 2011
- 10 separate dates in 2013
- Highlighted areas in the graphs shows mosquitoes still had increased
- Common pattern:
  - steady decline in mosquito count occurs towards the end of summer





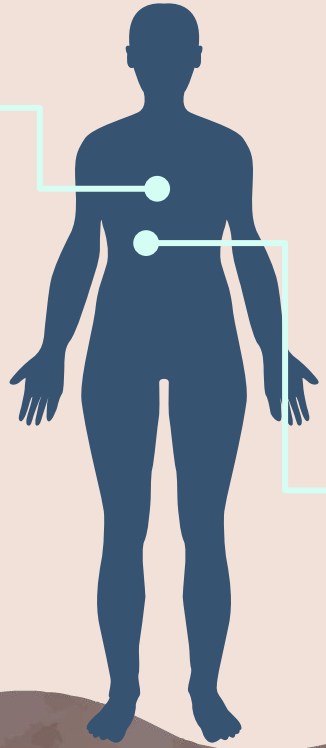
2 POTENTIAL OUTCOMES OF WNV

Fever
Headache
Body aches

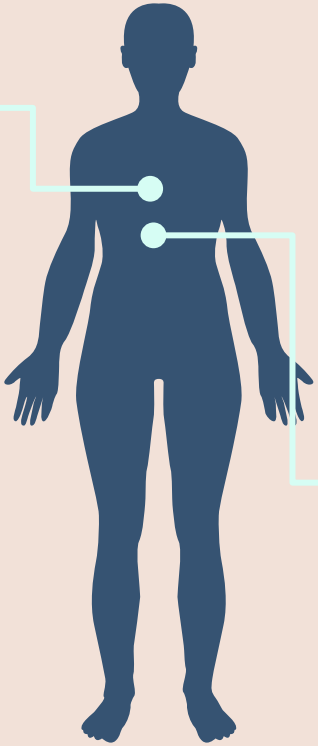
- Affect 1 in 5 people

Ferbrille Illness, Fever (WNF)

2 weeks to recover



2 months to recover



Neuroinvasive disease (WNND)

- Affect 1 in 150 people

Affects CNS

USING SACRAMENTO AS A PROXY

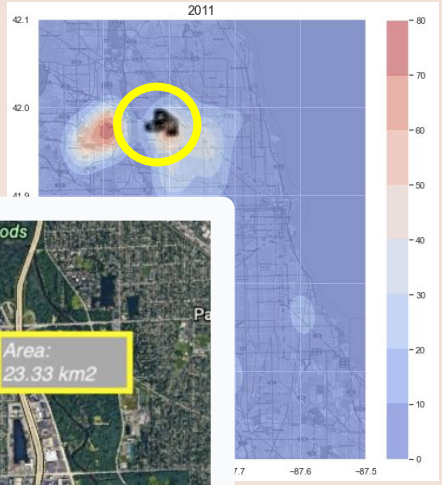
- In 2005, extensive research had already been done
- 1 km2 of spray cost ~ $3016

TYPE OF COST	WNF PATIENT	WNND PATIENT
MEDICAL	$302	$33,143
PRODUCTIVITY LOSS (PER WORK DAY)	$191	$191
PRODUCTIVITY LOSS (PER NON WORK DAY)	$125	$125
TOTAL COST	$2,712 (2 weeks to recover)	$43,547 (2 months to recover)

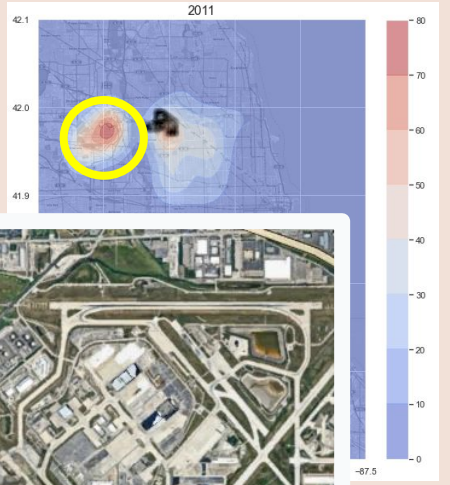
**The numbers above are an approximation of the average cost per patient type in Sacramento, as the extent of treatment per patient received would vary.*

Estimating Total Cost in 2011

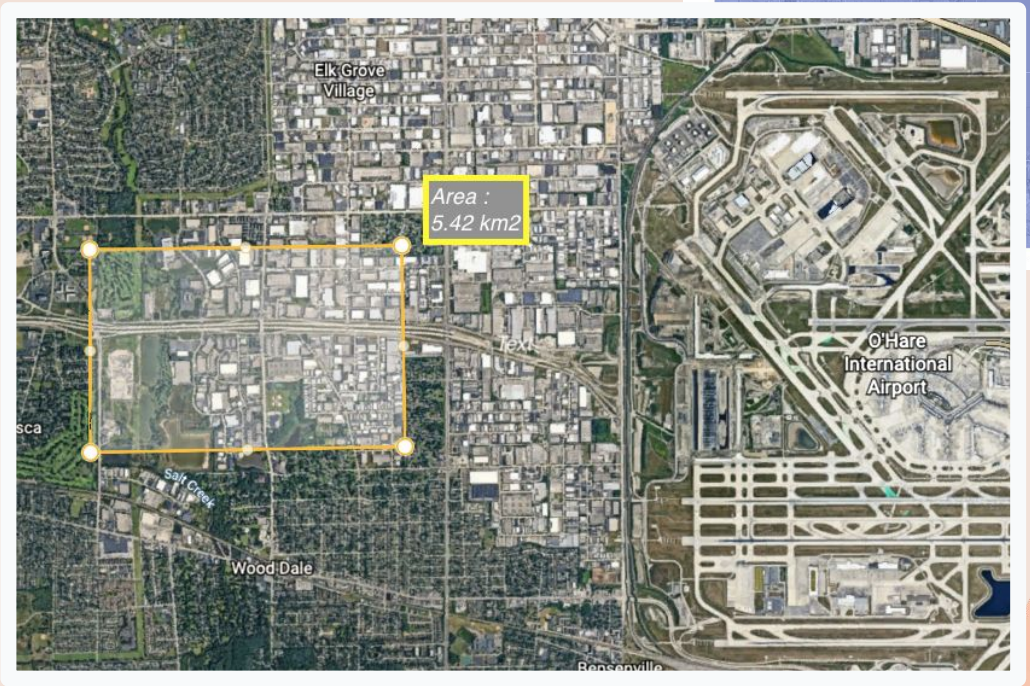
- Cost to spray area: ~ $70,362
- Given 150 people infected (Hypothetical)
  - 30 people with WNF
  - Cost: ~ $81,360
  - 1 person with WNND
  - cost : ~ $43,457
- Total cost: $195,269



Estimating Total Cost Spraying Recommend Location

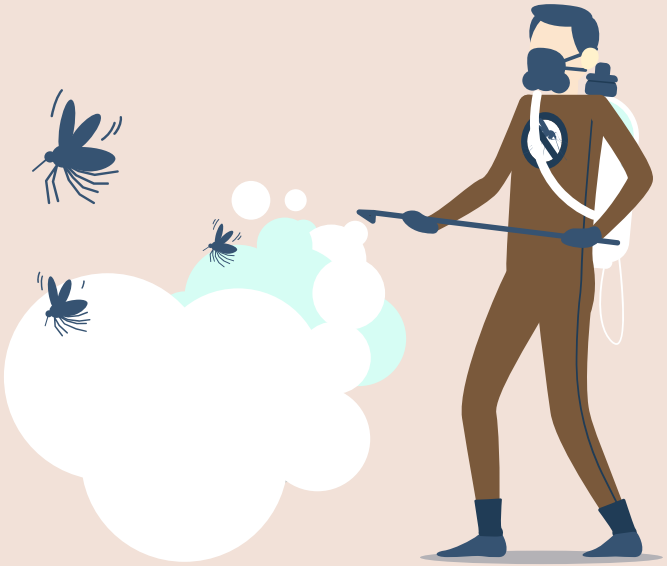


- Focus on area with high WNV
- Cost to spray area: ~ $16,322
- Assuming infected case reduce by 20%
- Given 120 people infected (Hypothetical)
  - 24 people with WNF
  - Cost: ~ $65,088
  - 0 people with WNND
- Total cost: $81,411 ($113,859 savings)



Key Takeaways from Cost-Benefit Analysis

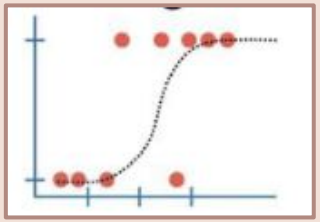
- More data is needed
  - Unable to provide concrete recommendations
- Recommendations that we can offer is where and when spraying efforts could be done.
- When? - Start of July latest August
- Where? - Areas with highest occurrence of WNV
- Potential areas of future studies:
  - relationship between spray effectiveness and wind direction, elevation or landscape.
  - Use of other methods such as larvicide.



03

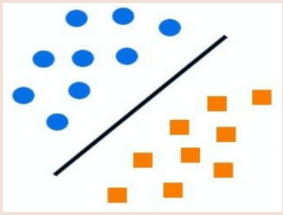
Modeling

Supervised Learning for Classification Model



Logistic Regression

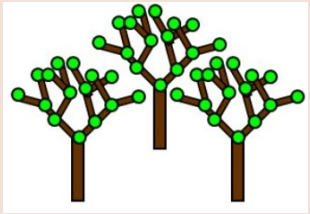
K Nearest Neighbors (KNN)



Support Vector Classifier (SVC)

Decision Tree

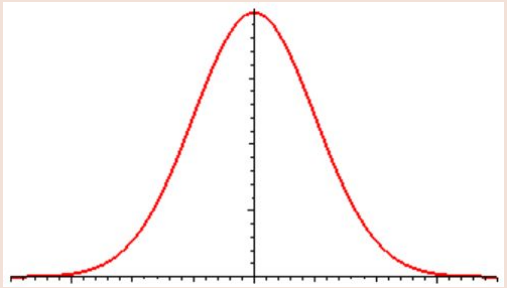
Random Forest



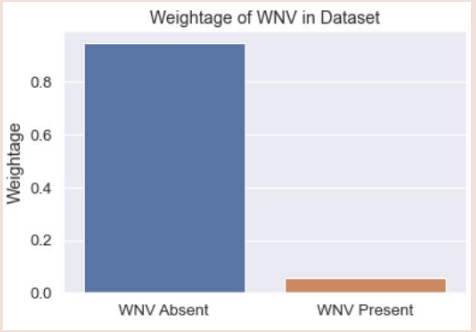
XGBoost

Modeling Process Considerations

1. Handle the different scales of the features by Standard Scaler
   - Logistic Regression
   - KNN
   - SVC



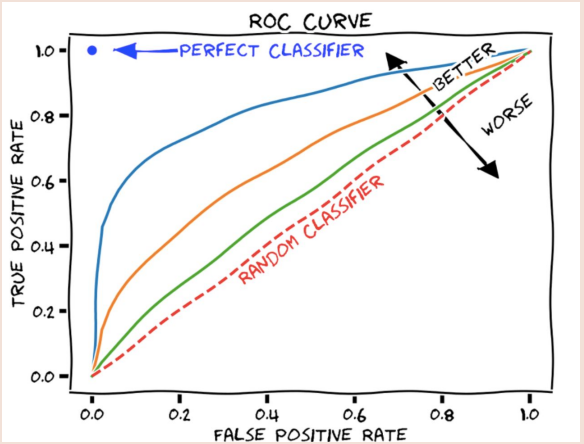
2. Handle class imbalance.
   - Weightage = balanced.



Model Selection

ROC-AUC Score (Baseline Score is 0.5)

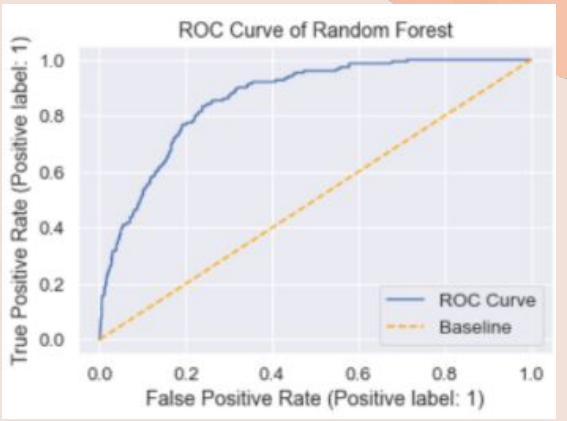
Model	ROC-AUC
Logistic Regression	0.818
K Nearest Neighbor	0.727
Decision Tree	0.610
Random Forest	0.807
SVC	0.805
XGBoost	0.862

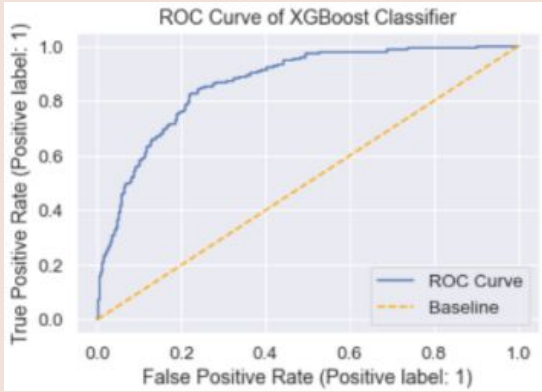


Model Hyperparameter Tuning

Hypertuned

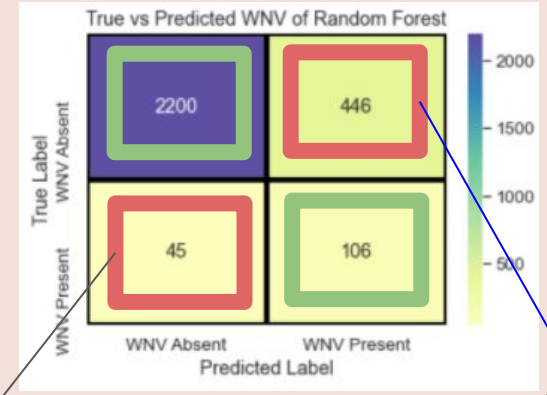
Model	ROC-AUC	ROC-AUC
Logistic Regression	0.818	0.821
Random Forest	0.807	0.863
SVC	0.805	0.855
XGBoost	0.862	0.865





Model - Confusion Matrix

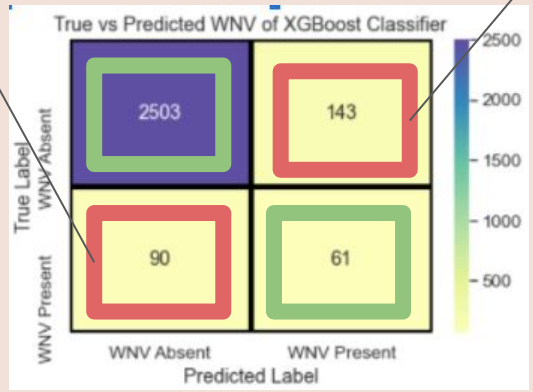
Random Forest



False Negative (FN)

False Positive (FP)

XGBoost

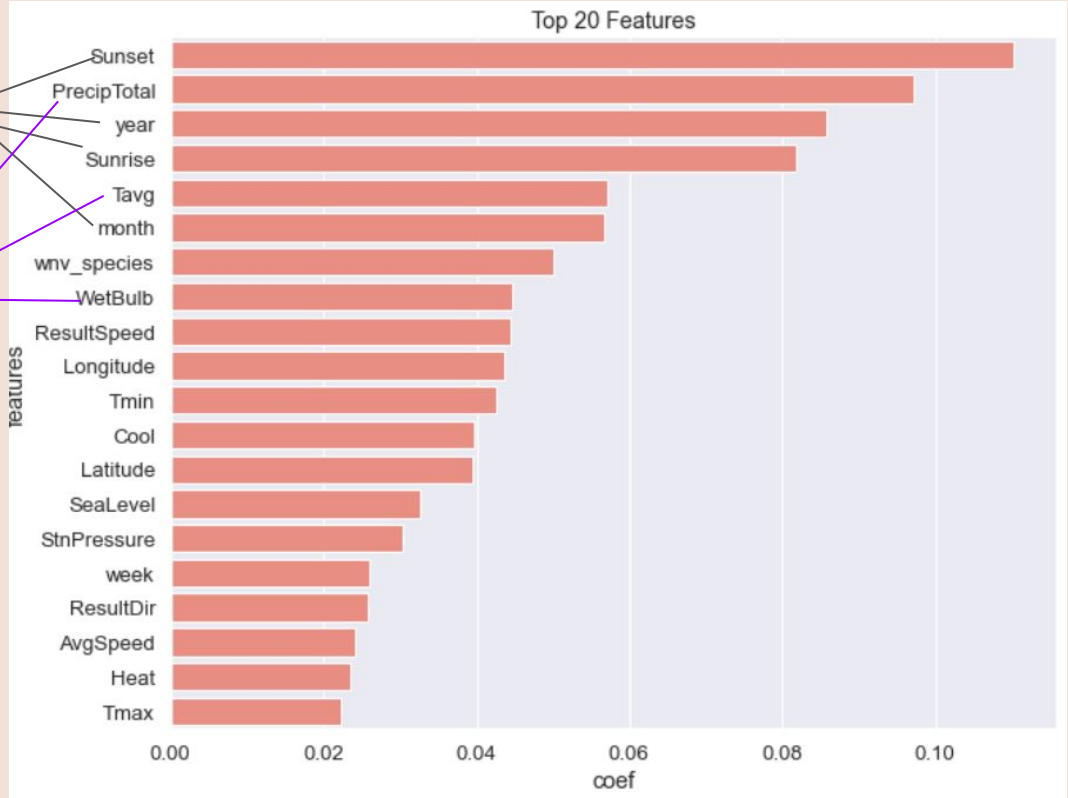


- Assume: When WNV is predicted as present, we will spray pesticide.
- Type I Error (FP)
  - Predict WNV present, but actually there is no WNV present.
  - "Wasted" cost incurred for spraying pesticide.
- Type II Error (FN)
  - Predict WNV not present, but actually WNV present.
  - Potential Health Risk for the people in that area in getting WNV.
- Aim: Reduce FN
  - Lower risk for the people to be exposed to WNV
  - Less chance of getting WNV.

Top 20 Features

Time-related

Weather-related



Conclusion

- Best Classification Model -> XGBoost
  - ROC AUC Score (0.865)
- Preferred Classification Model -> Random Forest
  - Cost benefit analysis
  - Type II Error vs Type I Error
- Top features -> more informed decision to spray insecticide

Recommendation

- Current datasets do not substantially point to a significant impact from spraying.
- Better designed spraying regime, such as spraying at the beginning of August.
- Invest more efforts in prevention awareness such as not leave exposed untreated water.

THANKS

Do you have any questions?
addyouremail@freepik.com
+91 620 421 838
yourcompany.com