# West Nile Virus Prediction

By: Solomon Heng, Ng Jun Kang, Preety Bhogal, Vaishnavi, Zoey Cheng

Disease And Treatment Agency (DATA)

Division of Societal Cures In Epidemiology and New Creative Engineering

# Background

- West Nile Virus (WNV)  is most commonly spread to humans through infected mosquitos
- In response to the recent epidemic in  City of Chicago, Chicago Department of Public Health (CDPH) had established a surveillance and control program. The program entails setting mosquito traps, testing for WNV and spraying of airborne pesticides in the city
- Centres for Disease Control and Prevention (CDC) has engaged DATA for the following services:
    - To conduct a review on the surveillance and control program through data analytics
    - To conduct a cost-benefit analysis on the annual cost projections for various levels of pesticide coverage and  the effect of of these various levels of pesticide coverage

# Presentation Scope

1. Aim
2. DATA’s Problem Statement
3. Intro to Datasets
4. EDA
5. Modeling and ROC AUC Curve
6. Cost Benefit Analysis & Possible Inferences
7. Conclusion & Recommendations

# Aim

- To provide CDC and CDPH an update on the data analytics study and cost-benefit analysis
- To review the pesticide deployment plan
- To seek CDC’s and CPDH’s support on DATA’s recommendations

# DATA's Problem Statement

- To build a classification model with high sensitivity & specificity on the presence of WNV at a specific place and time so as to aid CDC in deriving an effective pesticide deployment plan

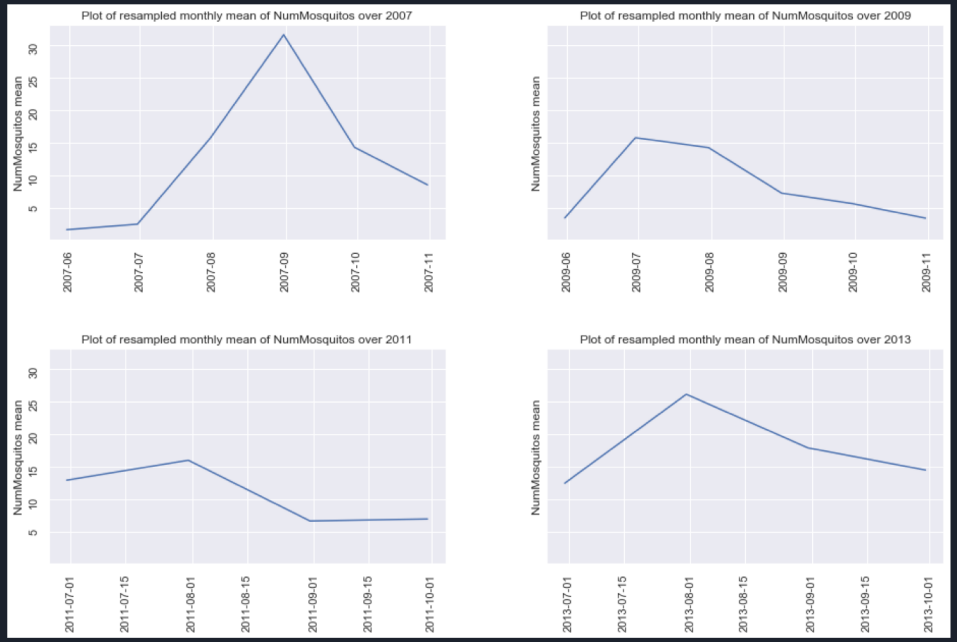
- To conduct a cost-benefit analysis on the annual cost projections for various levels of pesticide coverage and  the effect of of these various levels of pesticide coverage

# Intro to Datasets

- Datasets provided:
  - Data of 2007, 2009, 2011 & 2013
  - Data to predict on: 2008, 2010, 2012, 2014
  - Weather data: 2007 - 2014
  - Spray data: 2011 & 2013

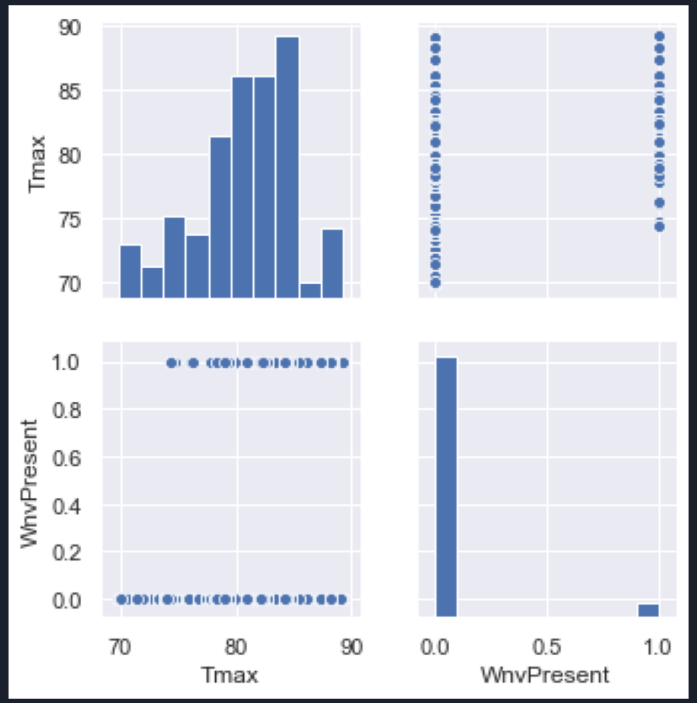
# EDA

- Notice how the number of mosquitoes usually picks up in July and start decreasing after September/October. Perhaps there is some sort of monthly seasonality when it comes to the mosquito counts.

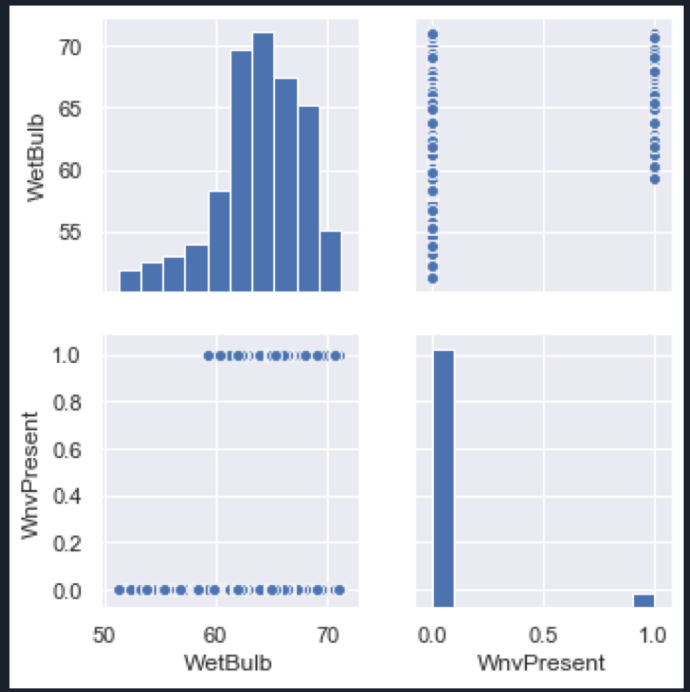


# EDA

- Higher temperatures -> Higher chance of WnvPresence



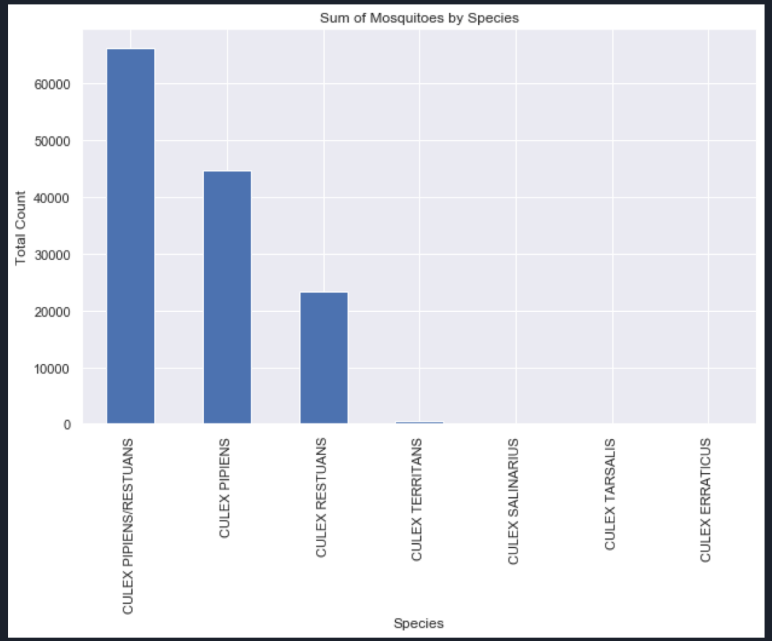

- At low wetbulb temperature, there is no presence of WnV as well.



# EDA

- WNV is only present in 3 species groups of mosquitoes

| Species | WnvPresent |
|---|---|
| CULEX PIPIENS/RESTUANS | 225 |
| CULEX PIPIENS | 184 |
| CULEX RESTUANS | 48 |
| CULEX TERRITANS | 0 |
| CULEX TARSALIS | 0 |
| CULEX SALINARIUS | 0 |
| CULEX ERRATICUS | 0 |



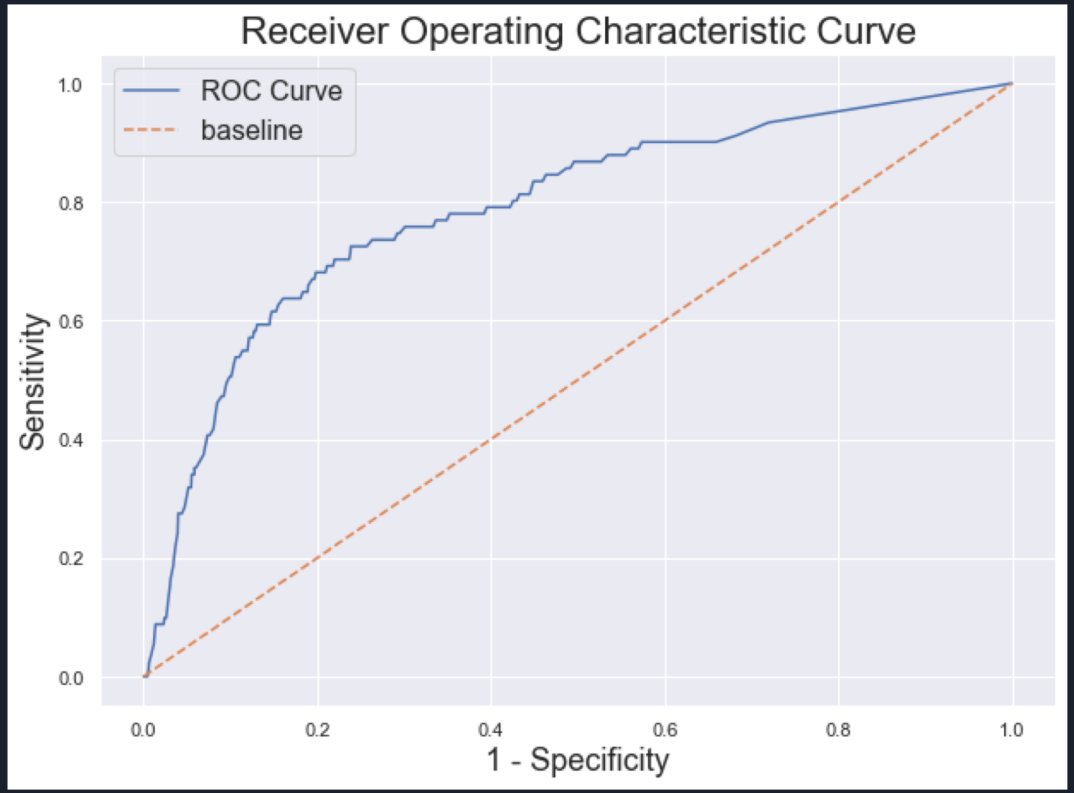
# Modeling

- Modeling Process
  - Train-Test Split
  - Hyperparameter Tuning
  - Cross Validation
  - Test Prediction
- Models and ROC AUC score

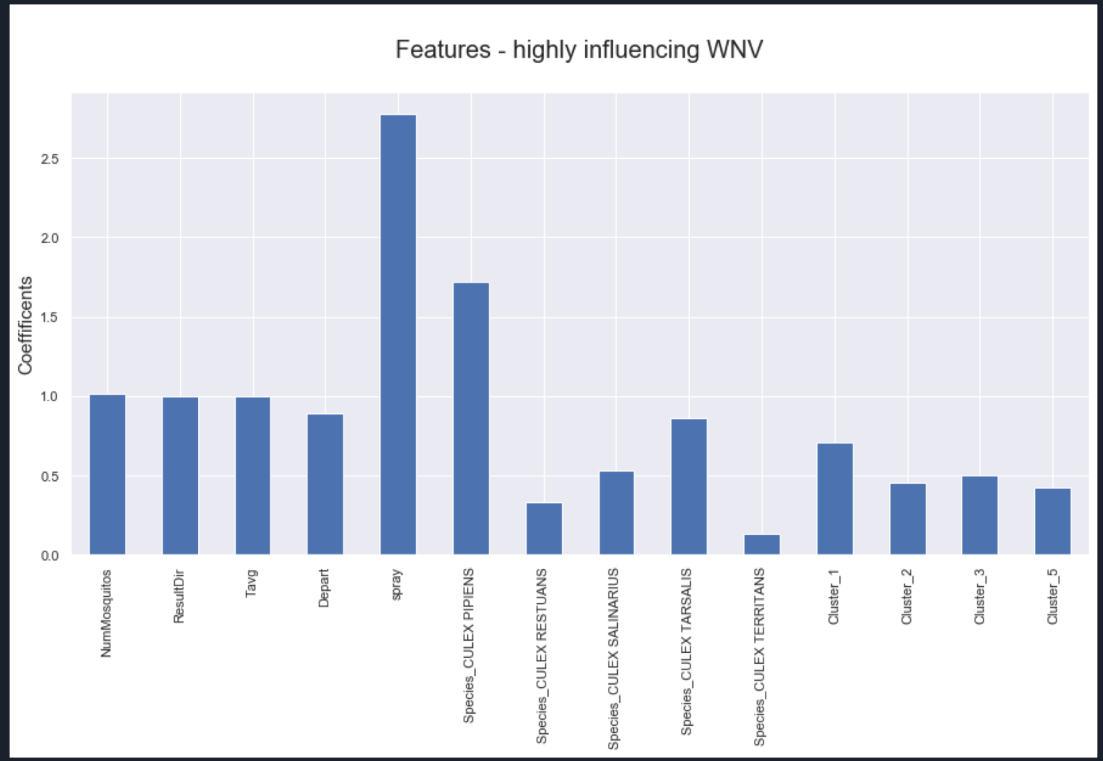
| | Cross Validation | Test |
|---|---|---|
| Logistic Regression | 0.893 | 0.737 |
| Random Forest | 0.972 | 0.719 |
| XGBoost | 0.984 | 0.635 |

- Kaggle score for Logistic Regression model with L1 Regularization: ~ 0.75

# ROC AUC Curve



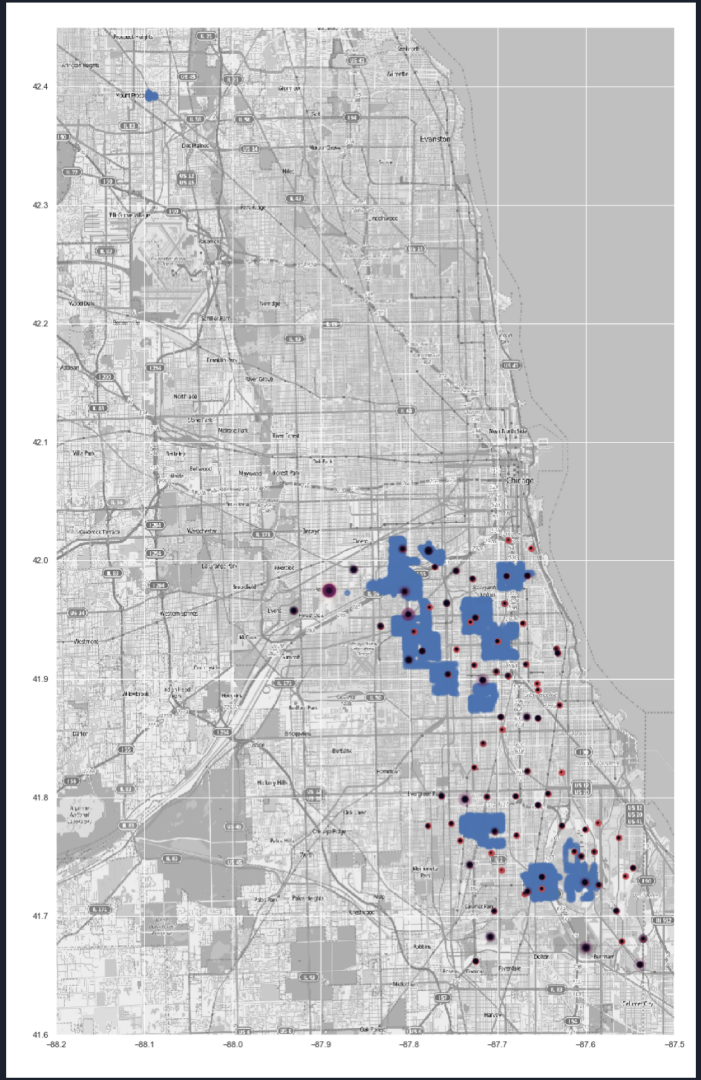
# Cost Benefit Analysis-Possible Inferences




Based on the graph presence of, WNV has an inverse relationship with Spray

# Possible Inferences

- It might be because spraying was done on locations that does not have that many mosquitoes and WNV present.



# Cost Benefit Analysis

- The product used to control the adult mosquitoes in Chicago, Zenivex™
- As per our calculations, the total estimated cost for Mosquito control spray is $6877.7 for an estimated area of 7037.14 acres  and spray cost at unwanted locations is $ 4011.99(SPRAY COSTS)
- A typical household with employer health coverage spends about USD 800 a year in out-of-pocket costs (MEDICAL COSTS)
- BENEFIT - decreases medical expenses for residents in Chicago, fewer people being infected with mosquitos, people taking few or no medical leaves.

# Conclusion

DATA's EDA, Modeling & CBA arrives at the following conclusion:

- EDA shows that while it is true that dry condition (low wet bulb temperature) and hot temperature is more favourable for presence of WNV,
- High mosquito counts (approx 1500) shows that the presence of WNV is almost guaranteed
- 2 species (CULEX PIPIENS & CULEX PIPIENS/RESTUANS) of mosquitos exceeded the count of 1000
- WNV is only seen in 3 mosquito species groups: CULEX PIPIENS/RESTUANS, CULEX PIPIENS and CULEX RESTUANS
- High likelihood that not all species of mosquitoes found in the trap will carry WNV
- Effectiveness of current pesticide deployment plan might not be optimal as the emphasis is not on areas with high mosquito counts or presence of WNV

# Recommendations

CDC and CDPH to support the following recommendations:

- New pesticide deployment plan which places more emphasis on areas with high mosquito counts or presence of WNV
- DATA to further narrow the study to the control of CULEX PIPIENS & RESTUANS mosquitoes for a period of 12 months as the EDA shows these are the main carriers of WNV
- Attachment of 2 x DATA data warehouse engineers to aid in the data collection process for spray datasets for a period of 12 months