



# Project 4

West Nile Virus



Ivan Cho  
Andrea Wong  
Carina Rebellon

# Objectives



**01**

**PROBLEM  
STATEMENT**

**02**

**EXPLORATORY DATA  
ANALYSIS (EDA)**

**03**

**FEATURE  
ENGINEERING**

**04**

**MODELLING  
RESULTS ANALYSIS**

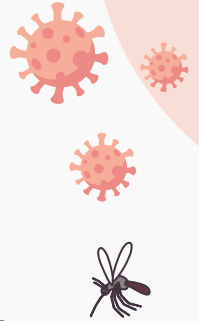
**05**

**CONCLUSIONS &  
RECOMMENDATIONS**



# Problem Statement

Our team aims to build a classifier model to predict the presence of West Nile Virus in Chicago supporting the Chicago Department of Public Health in its prevention efforts and control activities while also educating the general public and health care providers, enabling the Chicago Department of Public Health to effectively plan and manage their resources in preventing West Nile Virus.



# Beginnings of West Nile Virus ...

West Nile Virus was first identified in NYC in the summer and is leading cause of mosquito-borne disease in the US

1999



WNV quickly adapted to the local populations of Culex vector mosquitoes and avian populations, rapidly spreading throughout United States

## Chicago suburbs

Suitability of the environment for mosquito breeding and transmission to key avian species, especially the American robin

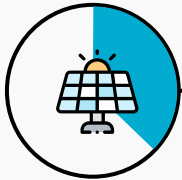


# Datasets



## Weather

Data detailing the weather conditions from 2007 to 2014 during the months of the test  
E.g. temperature, dew point, total precipitation



## Spray

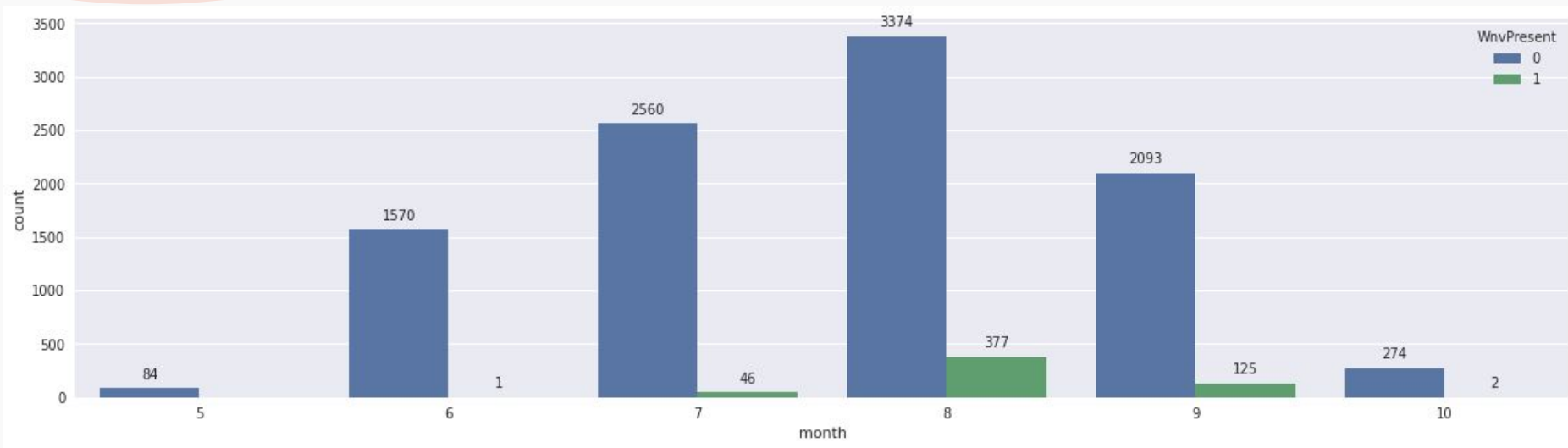
Data on Chicago's spray effort in 2011 & 2013 including the date, time, latitude and longitude of spray locations



## Test

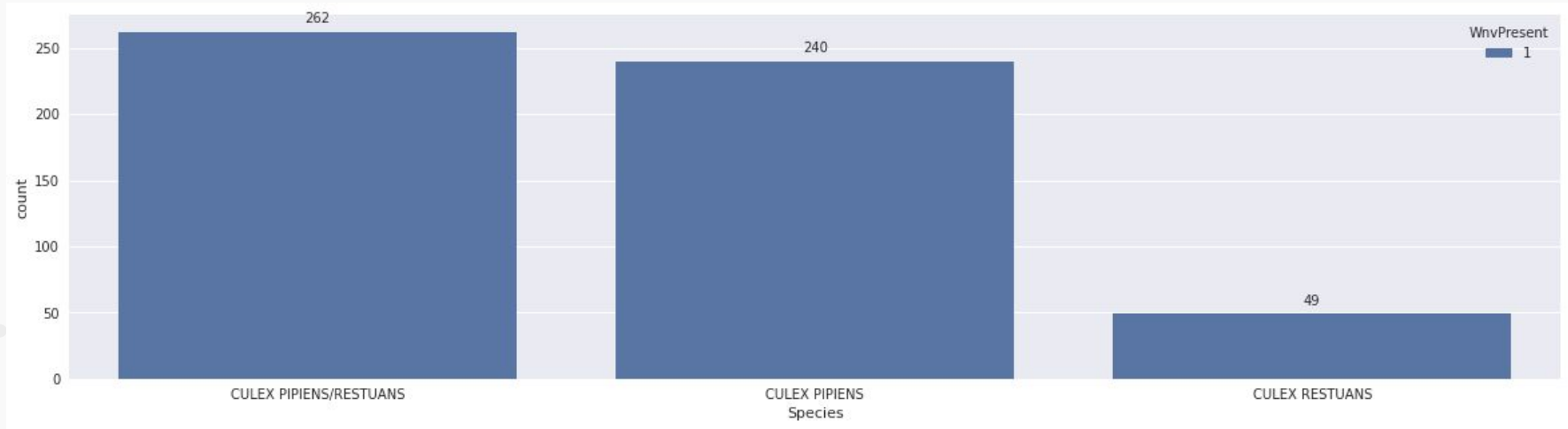
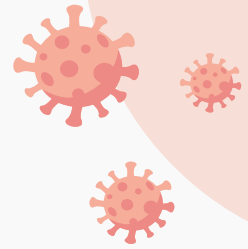
Contains data from 2007, 2009, 2011 and 2011 with details on the traps like number of mosquitoes caught, latitude, longitude, trap id

# Presence of West Nile Virus in traps - Month

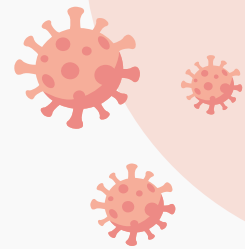


Counts of traps with West Nile Virus present was present in July, August & September

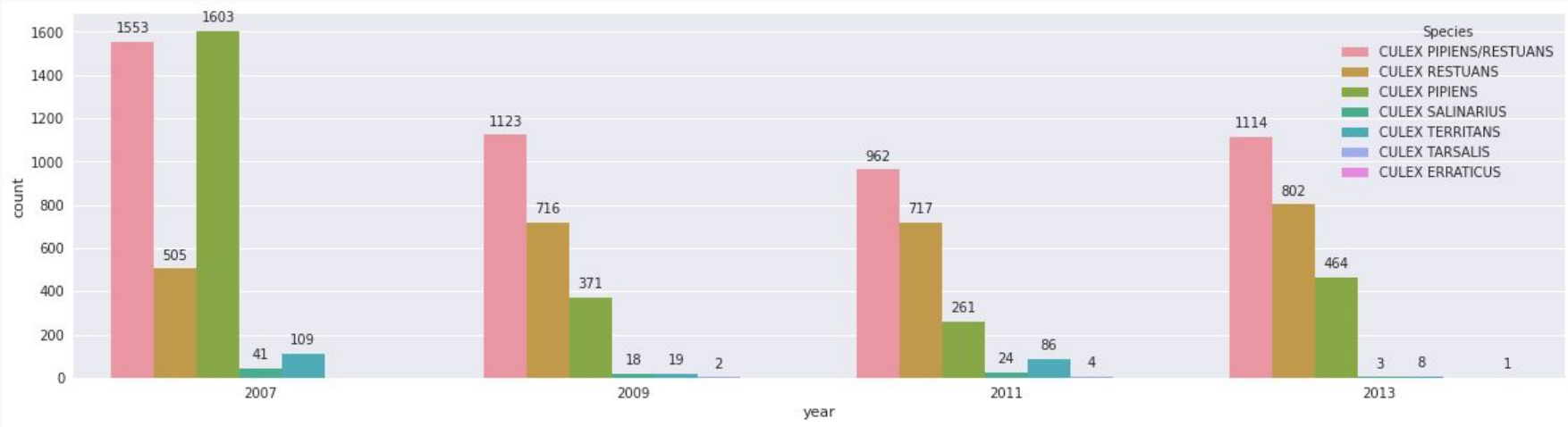
# Counts of traps with mosquitoes carrying West Nile Virus



Out of the 7 species of mosquitoes in our dataset, presence of west nile virus were primarily carrying the west nile virus were primarily in two species - Culex Papiens and Culex Restuans



# Counts of species captured in traps over the years



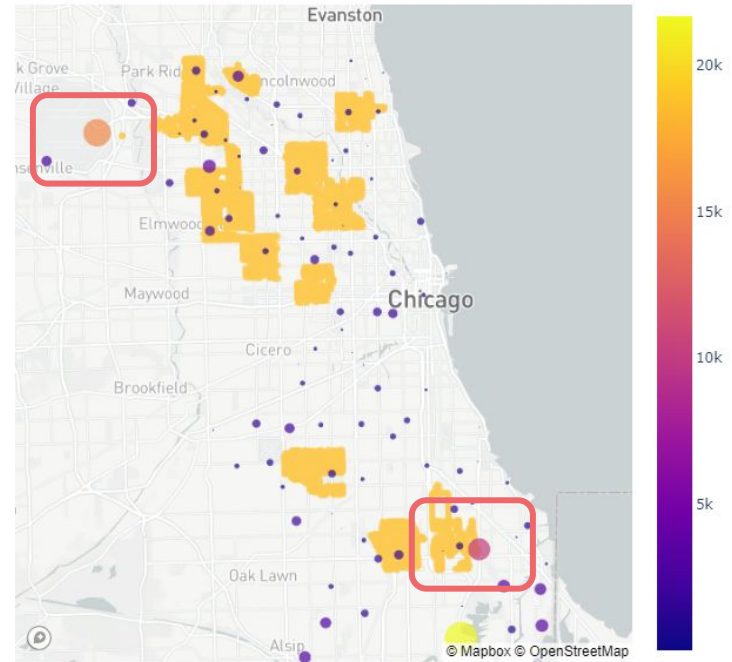
Culex Restuans and Culex Pipiens were the three main species seen in the training dataset. This was seen with these species taking up the bulk of the number of cases of mosquitoes captured in traps in 2007, 2009, 2011 and 2013. 2007 saw Culex Pipiens taking up 42% of traps sampled with mosquitoes caught. However in 2009, 2011 and 2013, Culex Pipiens/Restuans represented the bulk of the traps sampled



# How has spraying affected mosquito counts in the area

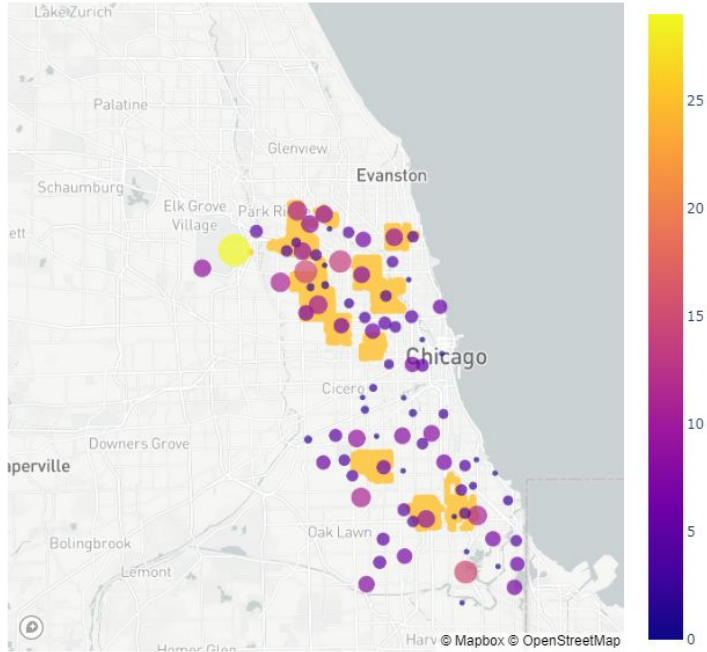
In the areas that have been sprayed, the sizes of the mosquito counts are much smaller. There are two big clusters observed where there is no spray in those area.

Spray vs Number of Mosquito by Cluster



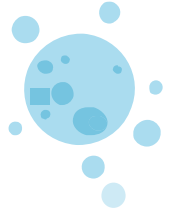
# Does spraying reduce counts of WNV?

Spray vs Count of WNV by Cluster



While spraying may have controlled the number of mosquito, it didn't seem to have significant impact on the number of virus present.

# Feature Engineering Decisions



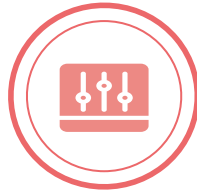
## Cluster

Clusters were created for weather station and breeding clusters



## Weather Elements

Relative Humidity calculations



## Rolling Window

Rolling window for Tavg, dew point, Precip Total and relative humidity



## Lat/ Long Features

Haversine formula to calculate distance between two points

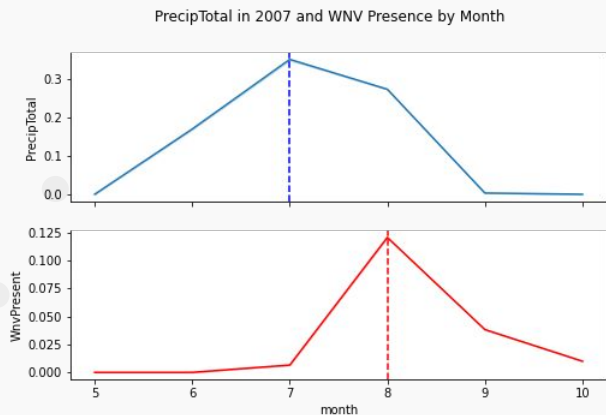


## Date/ Time Features

- Total Sunlight
- Presence of WNV by date of the Year
- # of Days from the Day of Maximum Presence of WNV



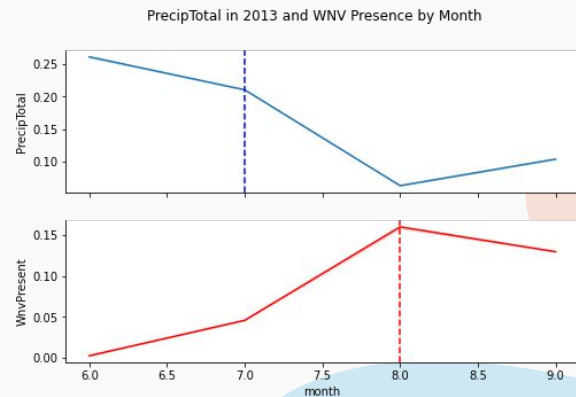
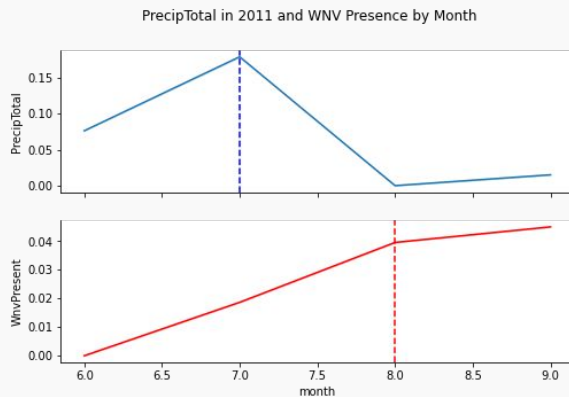
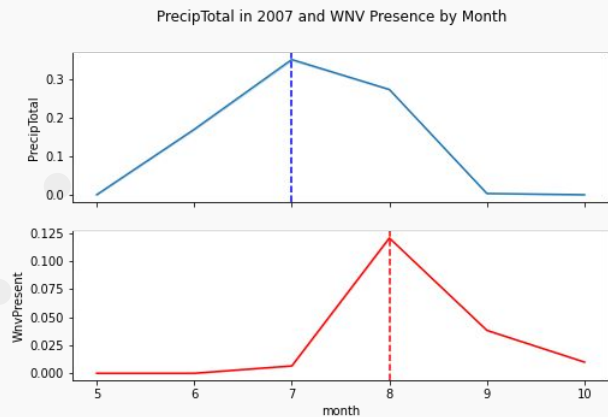
# Precip Total leading to presence of Wnv



When PrecipTotal reaches peak levels for the year, around 30 days after the heavy rain, presence of WNV will increase as well. One of the reasons could be mosquitoes breeding in stagnant water



# Precip Total leading to presence of Wnv



The same was observed in 2011 and 2013 as well.



# Modelling results



Our team will be evaluating our models by optimising these two areas:

- 1) ROC AUC score
- 2) Recall Score

Model	Test ROC AUC	KAGGLE ROC AUC	F1	Precision	Recall	Accuracy
ExtraTreesClassifier	<b>0.7387</b>	0.6357	0.2279	0.1345	<b>0.7456</b>	0.7325
SGDClassifier	0.7198	0.6777	0.2328	0.1410	0.6667	0.7673
RidgeClassifier	0.7175	<b>0.7013</b>	0.2229	0.1331	0.6842	0.7673
RandomForestClassifier	0.7124	0.6330	0.2225	0.1336	0.6667	0.7534
LogisticRegression	0.7953	0.6819	0.2292	0.1396	0.6404	0.7719
AdaBoostClassifier	0.6445	0.6555	0.2989	0.2653	0.3421	0.9150
GradientBoostingClassifier	0.5832	0.5928	0.2316	0.2895	0.1930	0.9322
XGBClassifier	0.6730	0.5846	0.2174	0.2857	0.1754	0.9331



# Modelling results (SelectKBest) ...

Model	Test ROC AUC	KAGGLE ROC AUC	F1	Precision	Recall	Accuracy
ExtraTreesClassifier	<b>0.7358</b>	0.6357	0.2305	0.1370	<b>0.7281</b>	0.7427
SGDClassifier	0.7151	0.6777	0.2229	0.1334	0.6754	0.7506
RidgeClassifier	0.7114	<b>0.7013</b>	0.2181	0.1301	0.6754	0.7436
RandomForestClassifier	0.7251	0.6330	0.2263	0.1349	0.7018	0.7459
LogisticRegression	0.7181	0.6819	0.2303	0.1392	0.6667	0.7641
AdaBoostClassifier	0.6290	0.6555	0.2687	0.2338	0.3158	0.9090
GradientBoostingClassifier	0.6105	0.5928	0.2749	0.2990	0.2544	0.9289
XGBClassifier	0.5650	0.5846	0.1905	0.24	0.1579	0.9289

# Recommendations to Develop and Strengthen Mosquito Control Program in Chicago



Continue spray efforts to reduce counts of mosquitoes as it is scientific proven to kill mosquitoes temporarily.



Remove all potential breeding areas near the most common traps with presence of west Nile virus - T900, T115. Remove, puncture or regularly drain all water-retaining objects



support the mosquito control program by educating the residents on understanding of the mosquitoes and how they are able to prevent certain mosquito-borne diseases.



# Recommendations to Develop and Strengthen Mosquito Control Program in Chicago



Recruit more Mosquito control officials who will be monitoring mosquito traps, investigating breeding sites, educating residents and schools and obtaining feedback from the public.



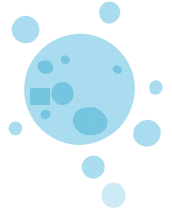
Initiate studies on the specific species that are west nile virus carriers, better understand the behavior of these particular species to predict the mosquito larval occurrences, flight behavior

# References ...

- <https://www.renesas.com/us/en/blogs/understanding-relative-humidity-and-dew-point>
- <https://www.sciencedaily.com/releases/2020/09/200915105932.htm#:~:text=W est%20Nile%20virus%20spreads%20most,published%20today%20in%20eLife%20 shows>
- <https://kestrelmeters.com/blogs/news/the-science-of-mosquito-abatement#:~:text=Wind%20works%20as%20a%20natural,MPH%20wind%20gust%20is%20subst antial>
- [https://en.wikipedia.org/wiki/Rain#:~:text=Light%20rain%20%E2%80%94%20when%20the%20precipitation,50%20mm%20\(2.0%20in\)%20per](https://en.wikipedia.org/wiki/Rain#:~:text=Light%20rain%20%E2%80%94%20when%20the%20precipitation,50%20mm%20(2.0%20in)%20per)



# West Nile Virus Transmission



01

An infected mosquito bites a bird



02

Bird transmit the disease while flying to another location



03

Mosquitoes become infected when they feed on infected birds



04

Infected mosquito feeds on humans who also becomes infected. The virus can also infect other mammals including horses

# METHODOLOGY ...



**DATA CLEANING**



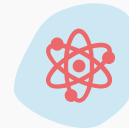
**FEATURE  
ENGINEERING**



**CONCLUSION &  
RECOMMENDATIONS**



**EXPLORATORY  
DATA ANALYSIS  
(EDA)**



**MODELLING**