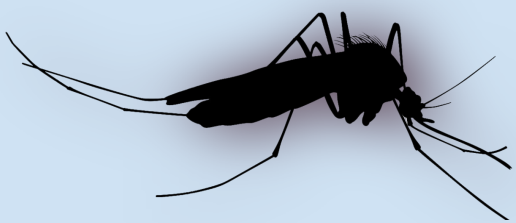
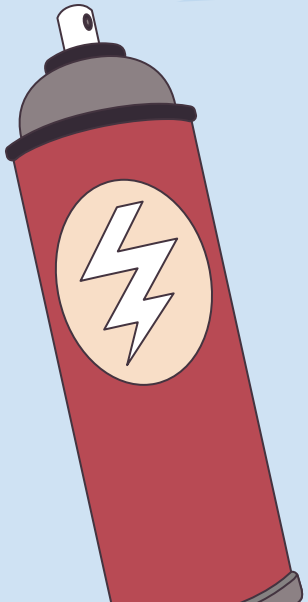


GA Project 4:

West Nile Virus Carrying Mosquitoes in Chicago

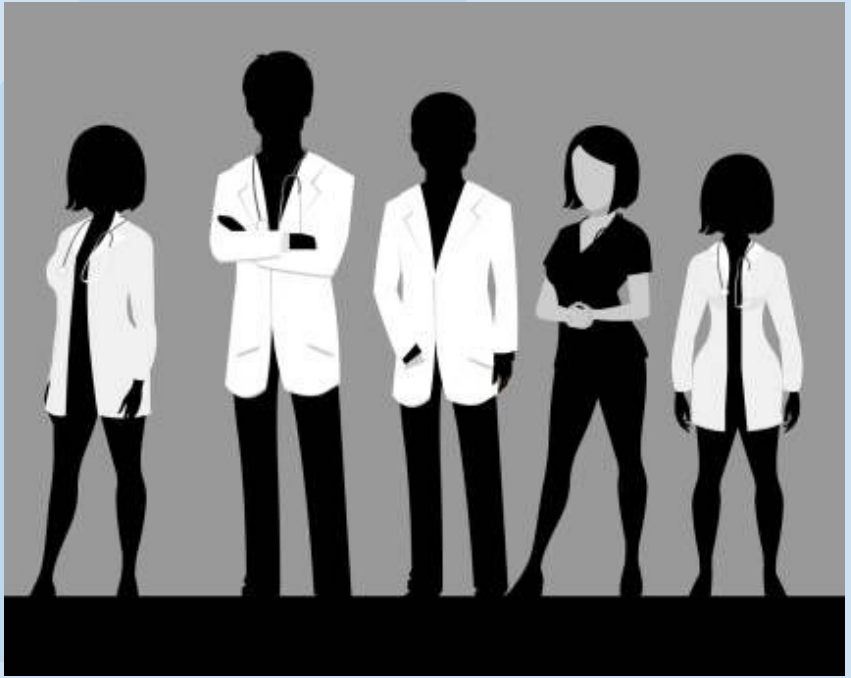
Presented By:
Clara Gan, Luka Chua, Ronnette Chan, Nixon Cheng, Johnny Tseng

Table of contents



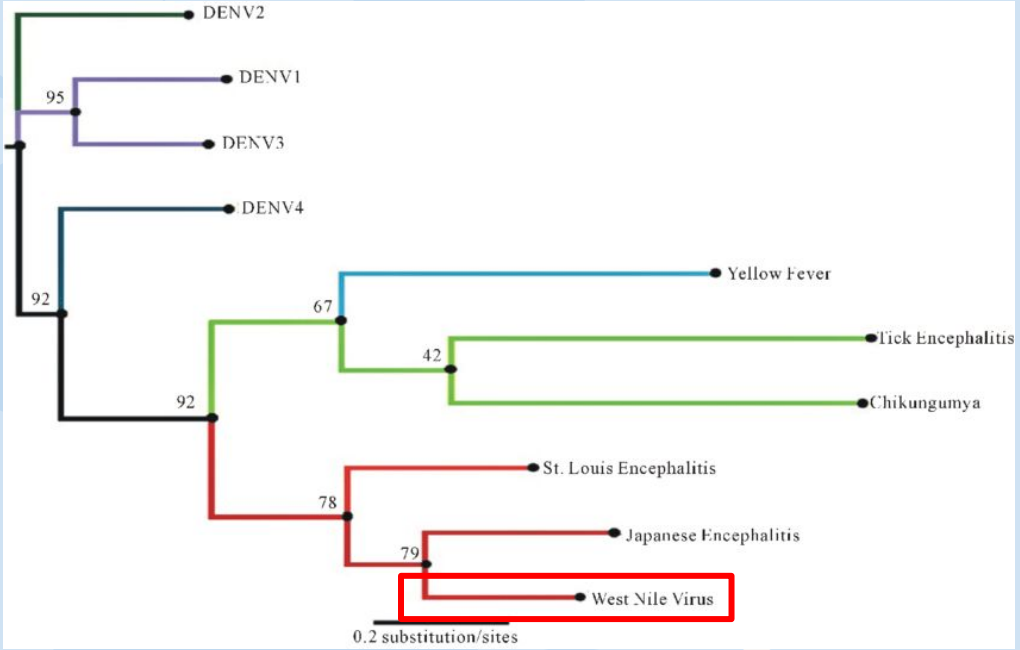
BACKGROUND: Who Are We & Problem Statement

- **Private Consultant to CDC**
  - Environmental Factors affecting WNV transmission
  - Reduce WNV carrying mosquitoes



BACKGROUND: What is West Nile Virus & Resource

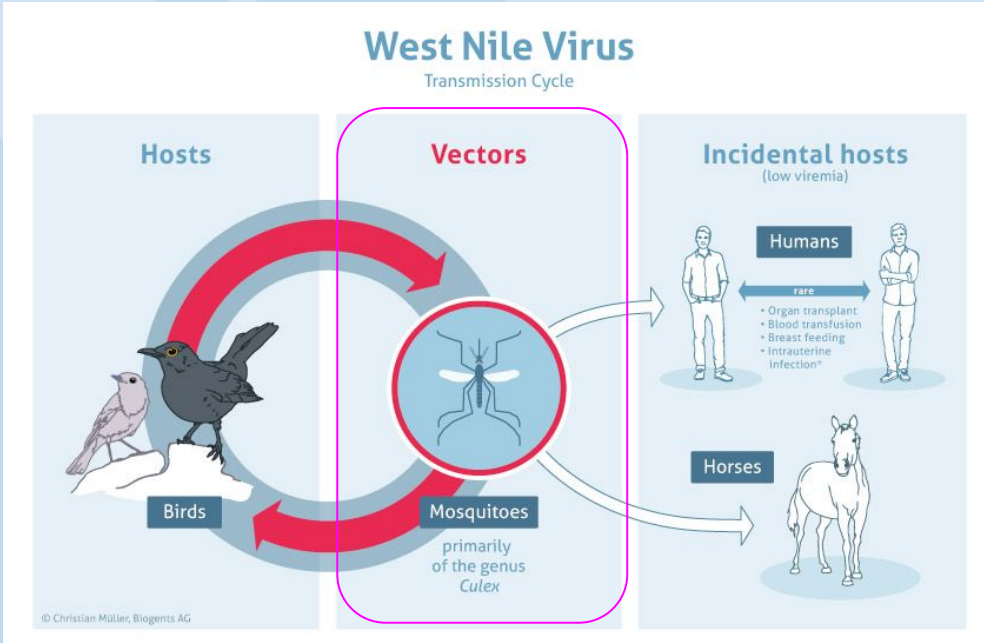
- West Nile Virus
  - **Vector**-borne virus
  - Originated in **Africa**
  - **1 in 5** falls **mildly** ill
  - **1 in 150** falls **severely** ill



BACKGROUND: What is West Nile Virus & Resource

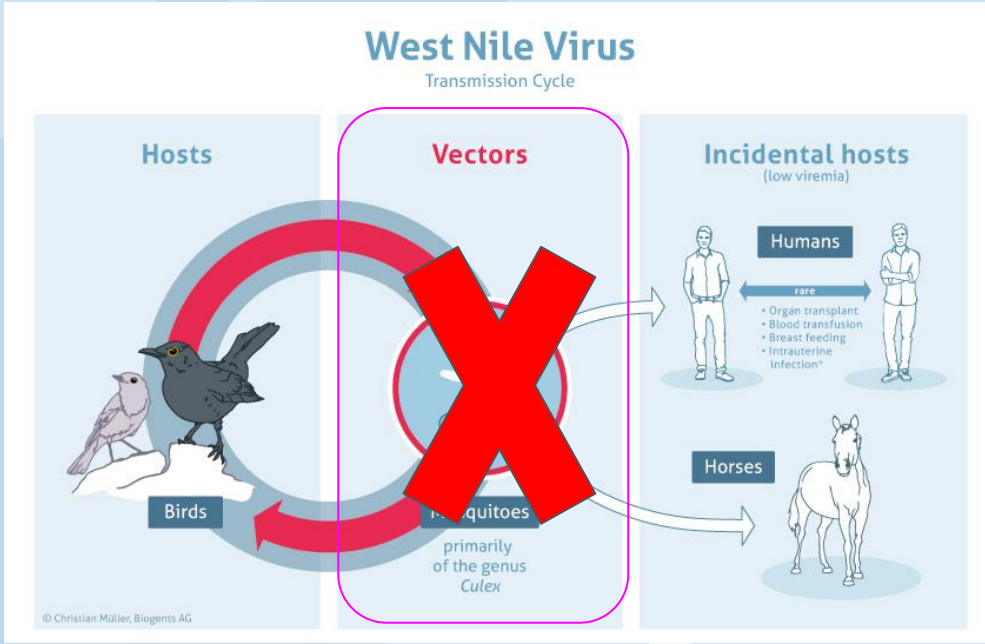
- West Nile Virus
  - Vector-borne virus
  - Originated in Africa
  - 1 in 5 falls mildly ill
  - 1 in 150 falls severely ill

- **Prevention** is the best cure



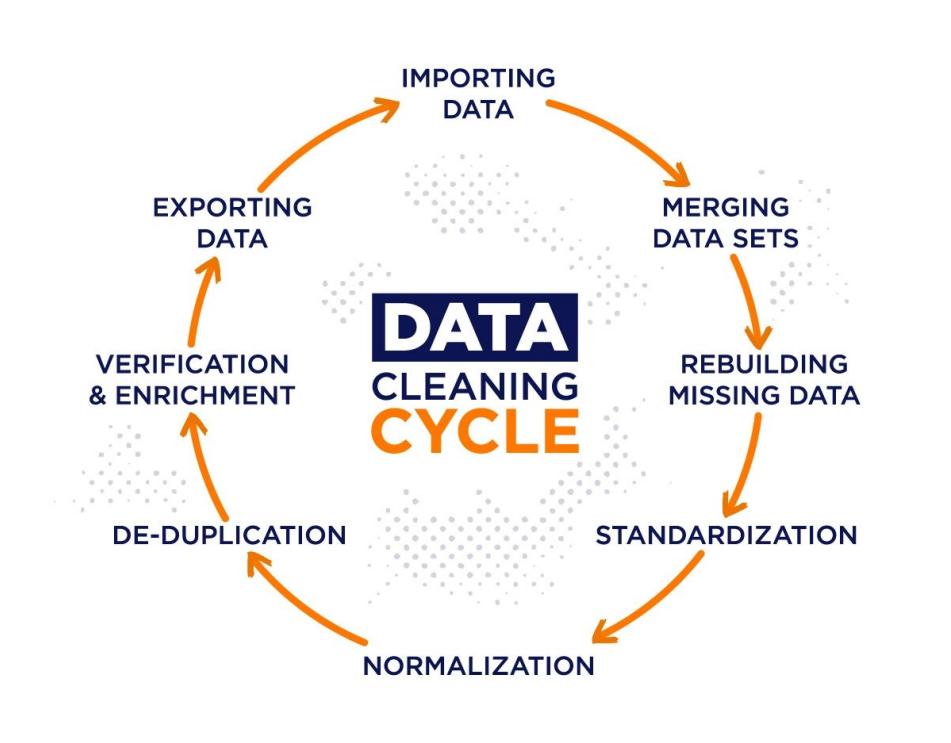
BACKGROUND: What is West Nile Virus & Resource

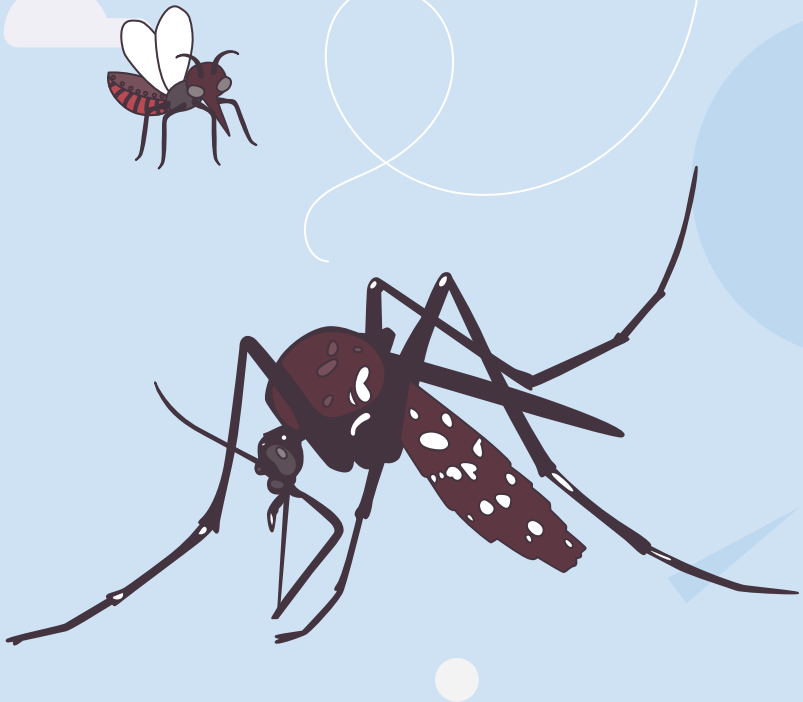
- West Nile Virus
  - Vector-borne virus
  - Originated in Africa
  - 1 in 5 falls mildly ill
  - 1 in 150 falls severely ill
- Prevention is the best cure
- Resource and Data from in-house recordings from **2007 up to 2014**



DATA CLEANING: Filtering / Removing / Replacing/Merge

- Determine Relevant Columns
- Clean nulls, na and drop by:
  - Impute
  - Replace
  - Drop
  - Remove Duplicates
- Merge Train, Spray and Weather

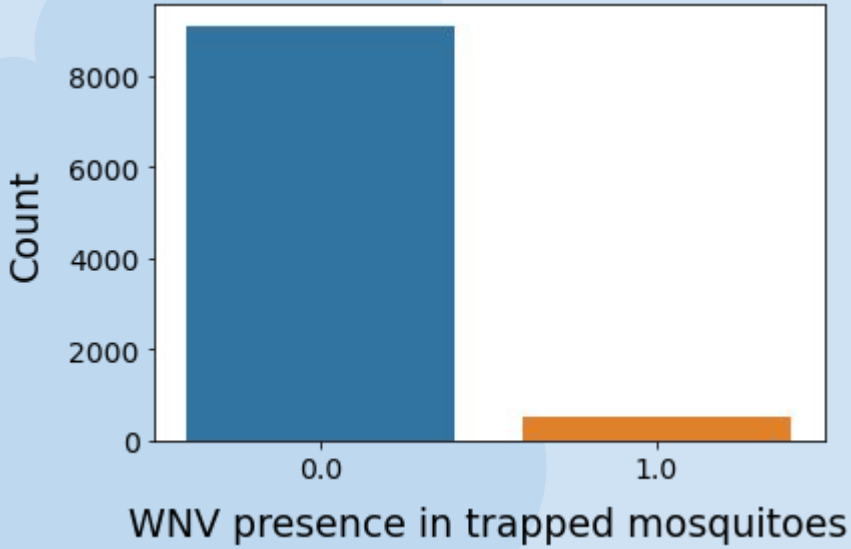




EDA

Unbalanced Class

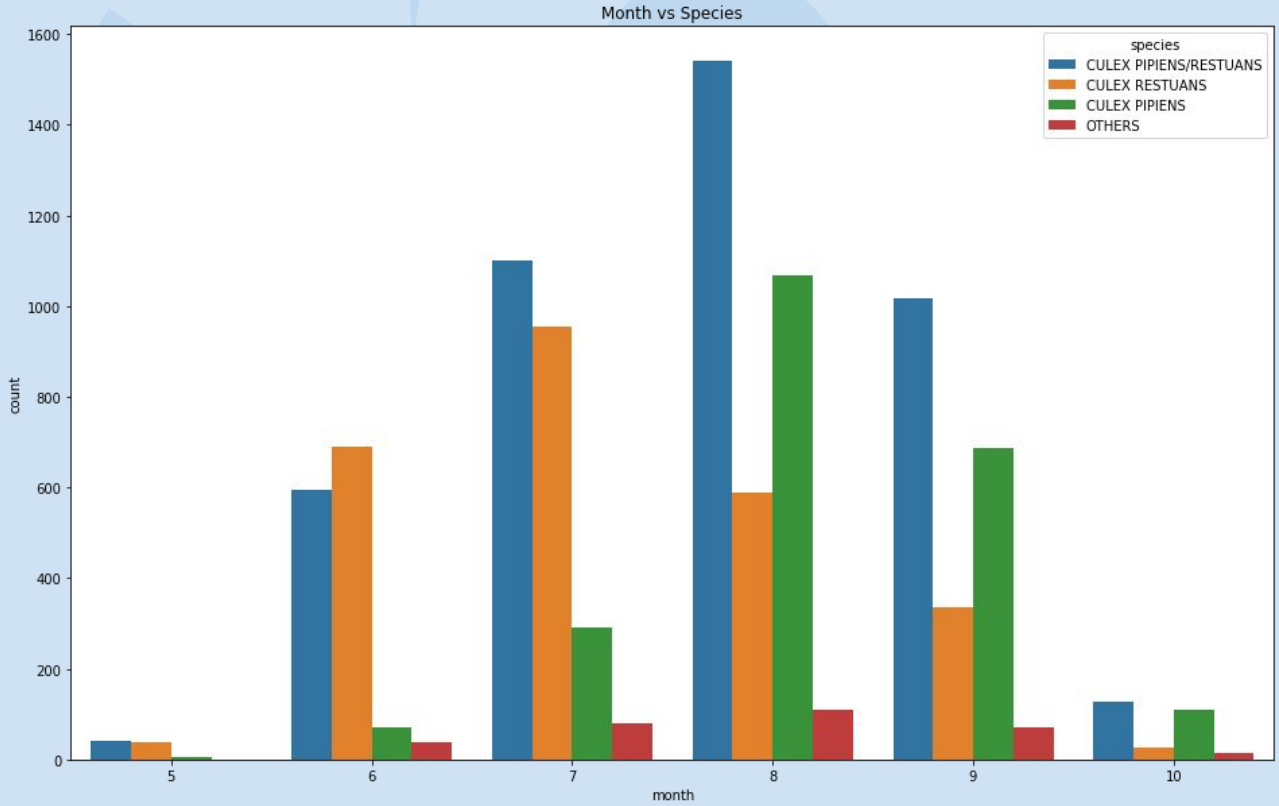
- Imbalanced Data
- Use **SMOTE** and **ADASYN** at modelling stage
- Minimize False Negatives
- Minimize False Positives



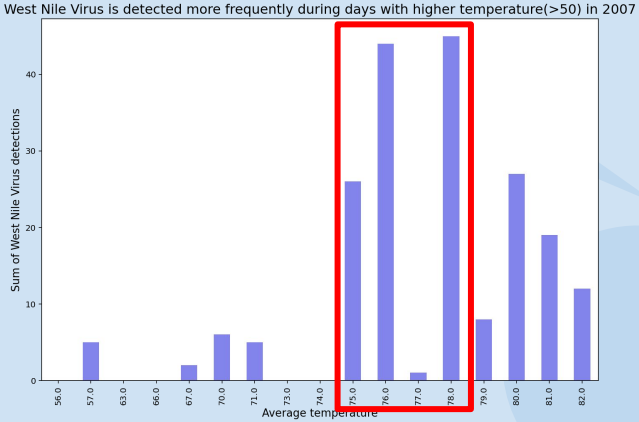
Key:
0 = Not West Nile Virus Carrying Mosquitoes
1 = West Nile Virus Carrying Mosquitoes

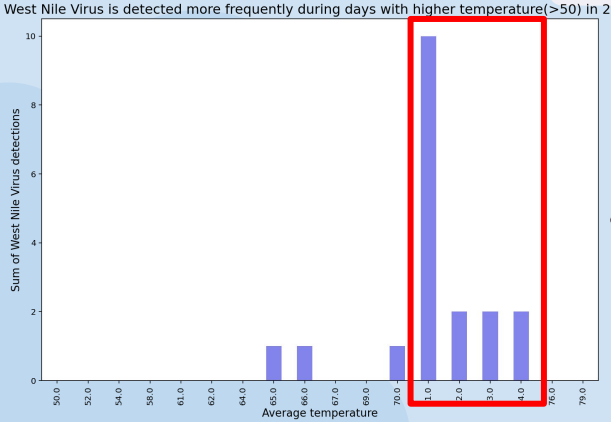
Mosquito Population vs Months

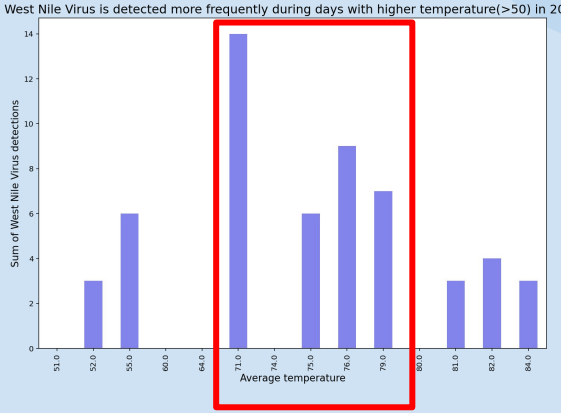
- Peaks around **August** (hottest month)
- Blue, Orange and Green represents the mosquitoes that **carriers** of **WNV**
- Other species were combined together

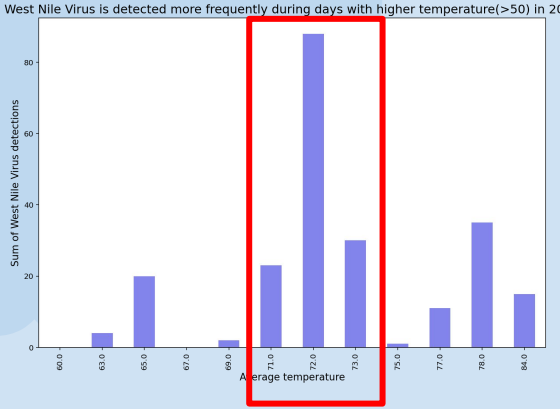


WNV Mosquito Frequency vs Temperature



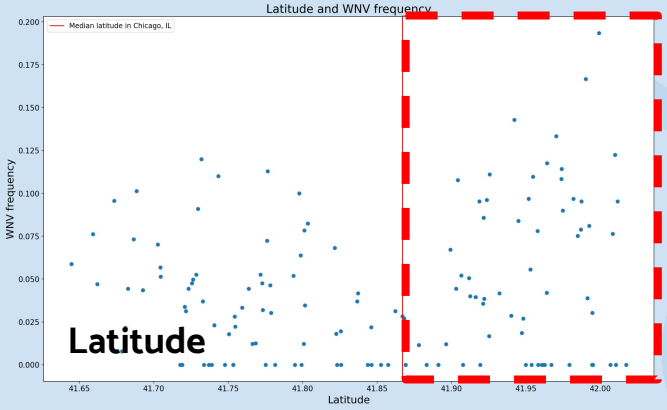




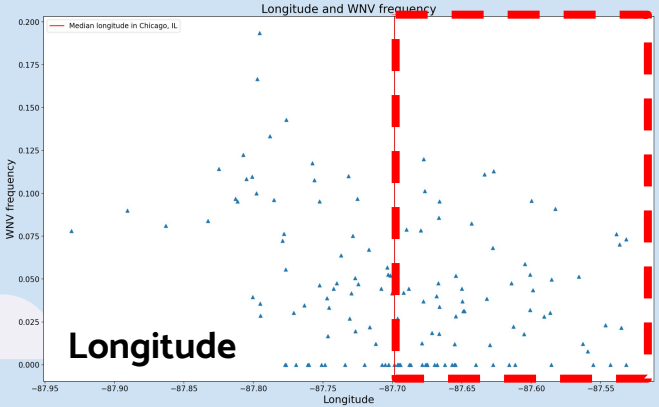


Majority of the Population

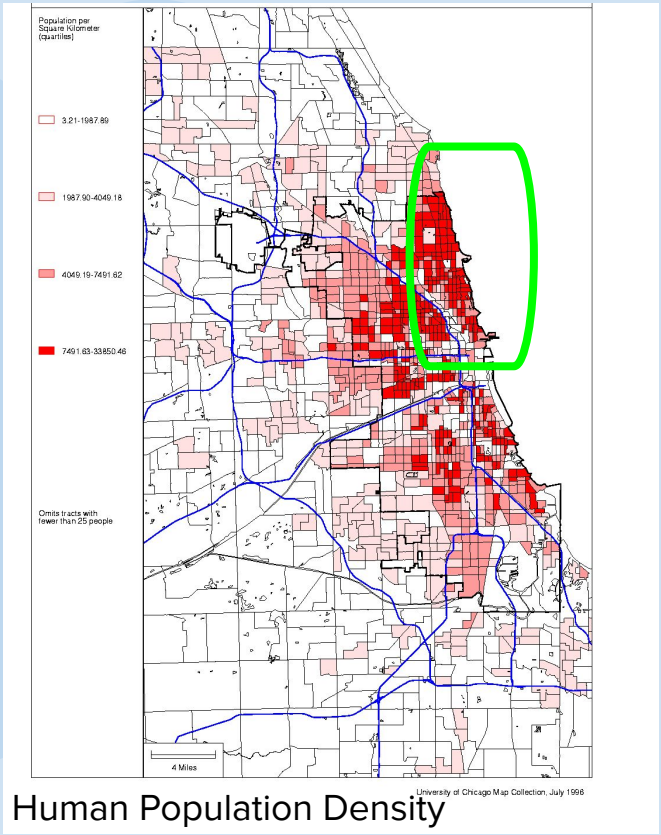
WNV Frequency



Higher density in Northern region

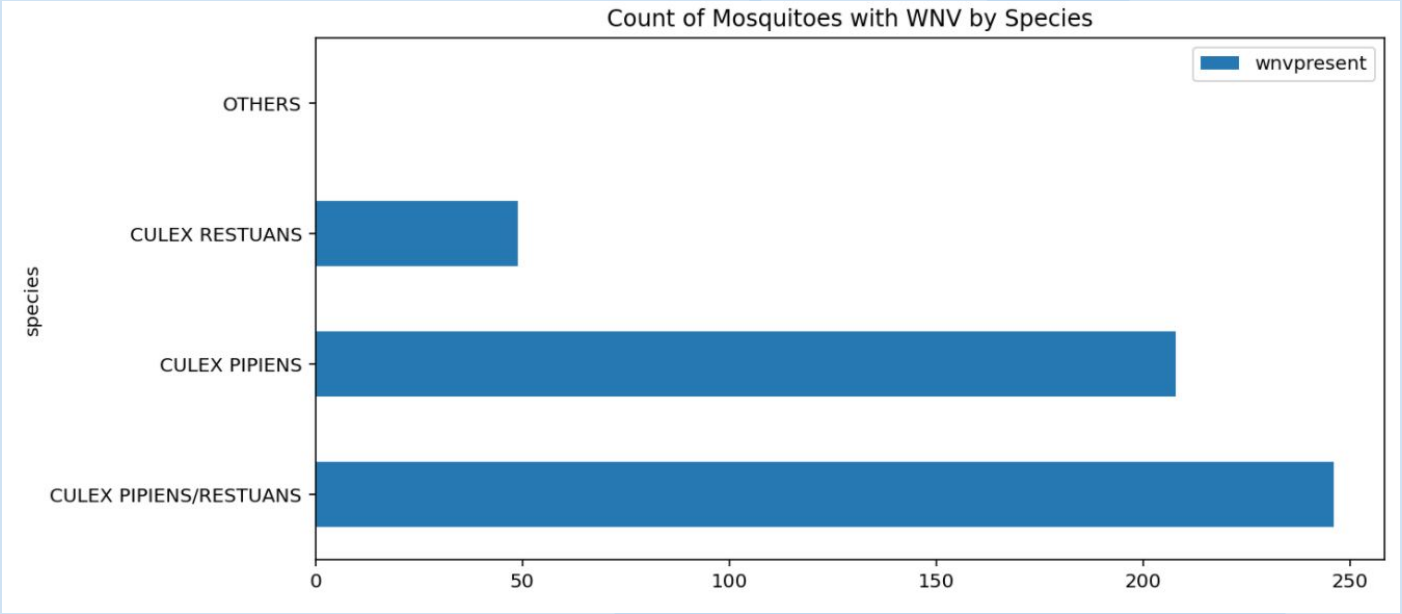


Higher density in Eastern region

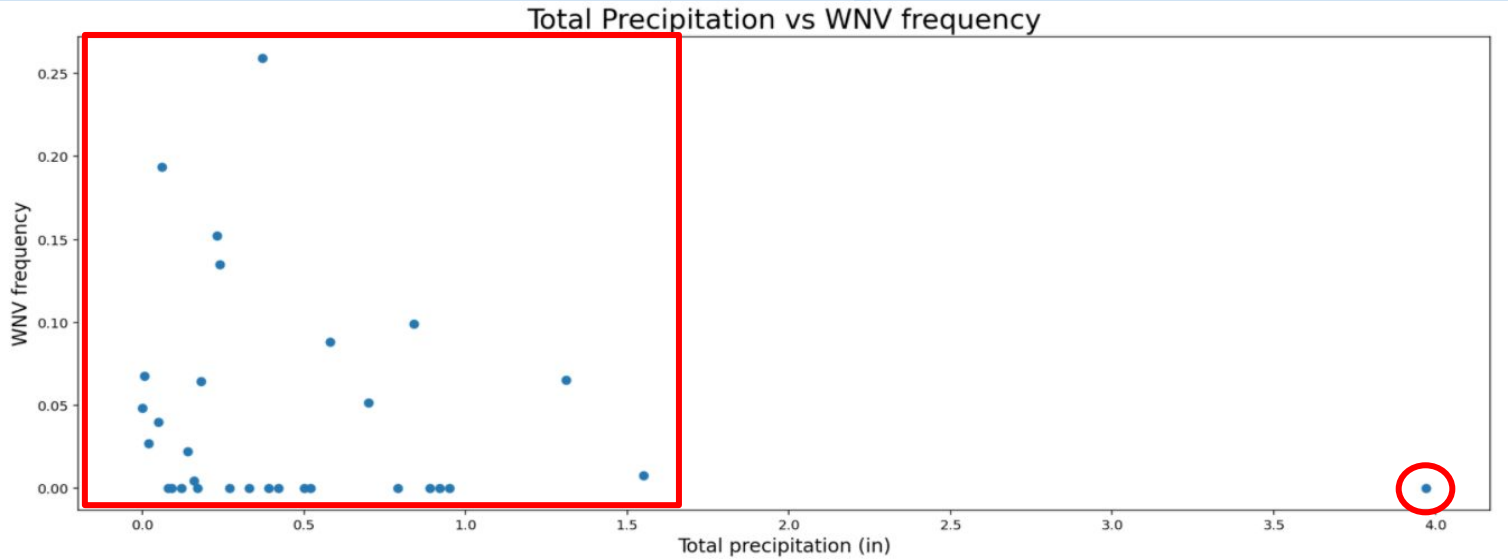


Human Population Density

Mosquito Species with WNV



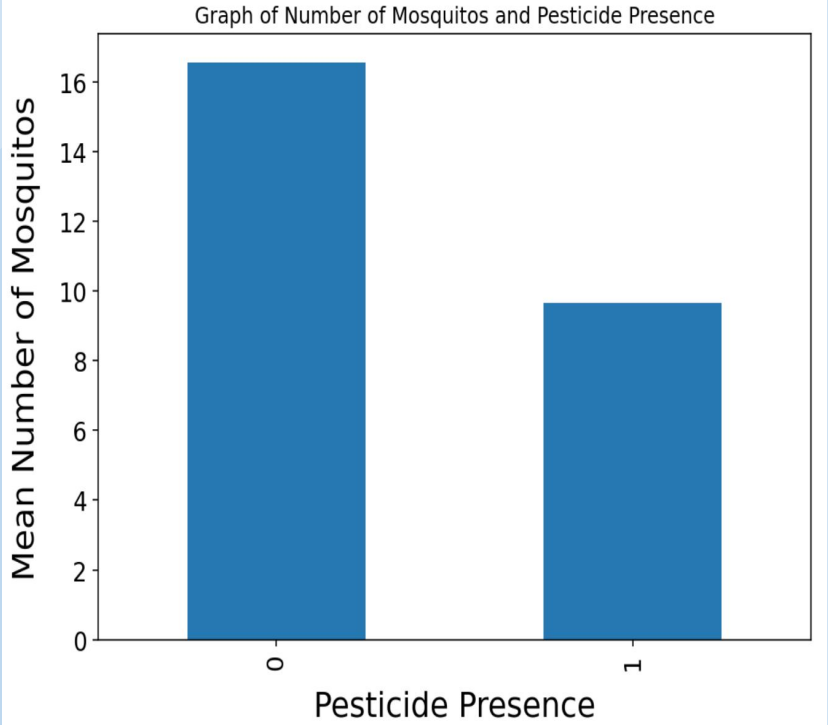
Total Precipitation and WNV



- Very little precipitation required for the mosquitos to breed.
- Excessive rainfall can cause breeding sites to overflow, disrupting mosquito breeding and destroying developing larvae.
- They could also be breeding in other sources of water such as in flower pots.

Effect of Pesticides on Mosquitos

- **Decrease** in number of mosquitos with of pesticides
- But **not as significant** as we would expect from using pesticides

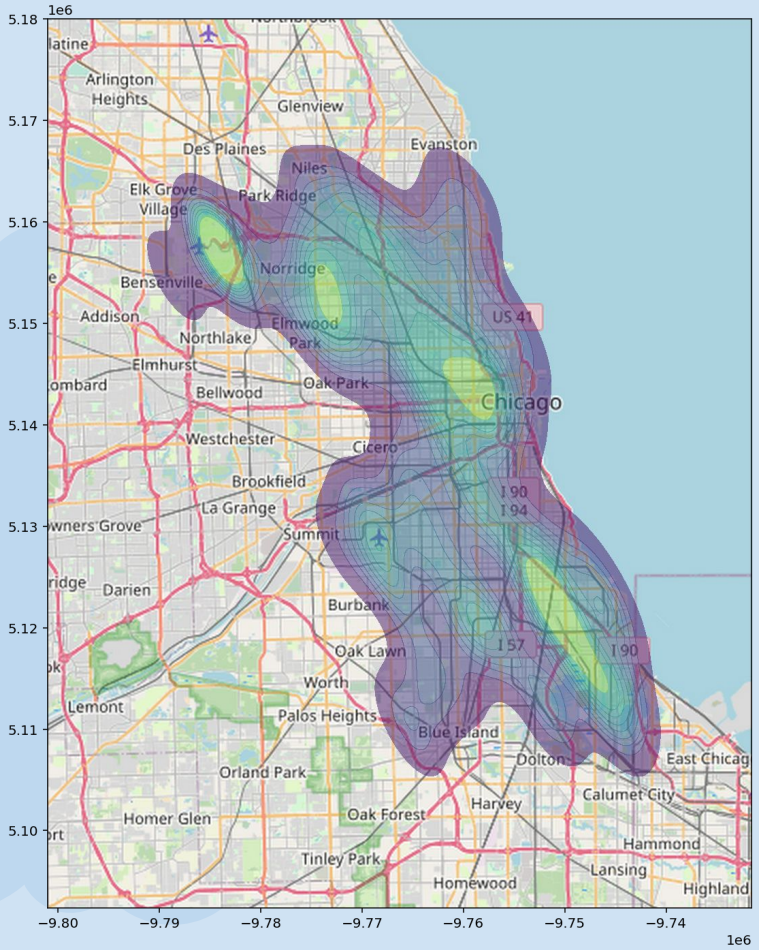


Type of Pesticide Use Matters

- Zenivex is an **Adulticide**
- **Least effective** mosquito control technique
- Programs spray **indiscriminately**

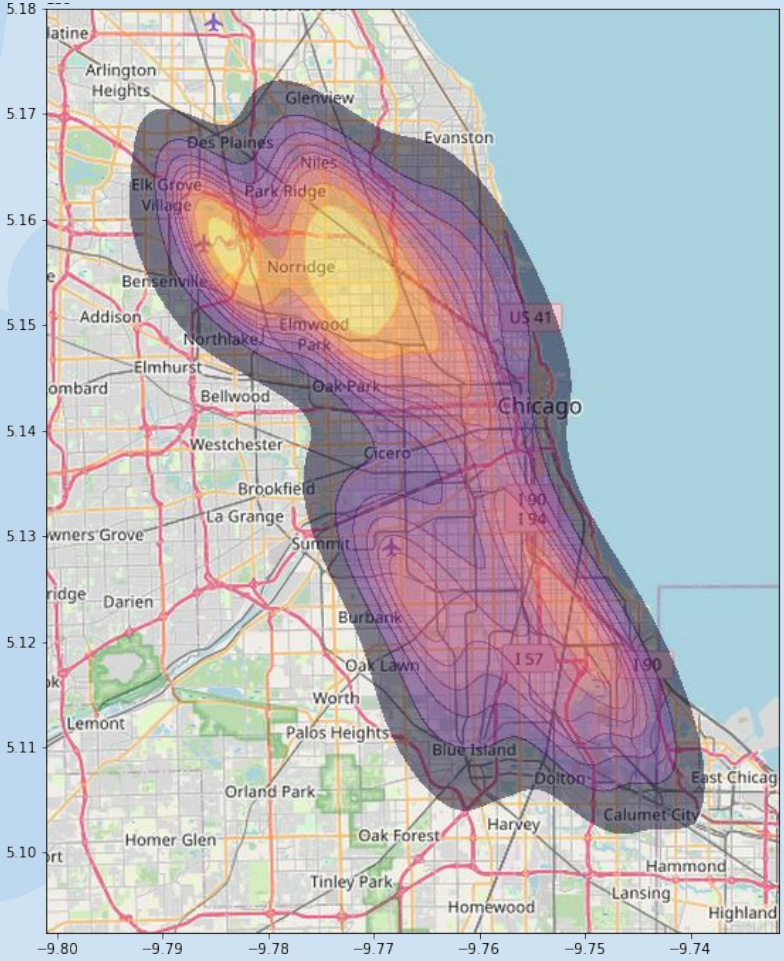
Mosquito Density

- Light green represents high concentration of Mosquitoes
- Majority located near the **coastline**



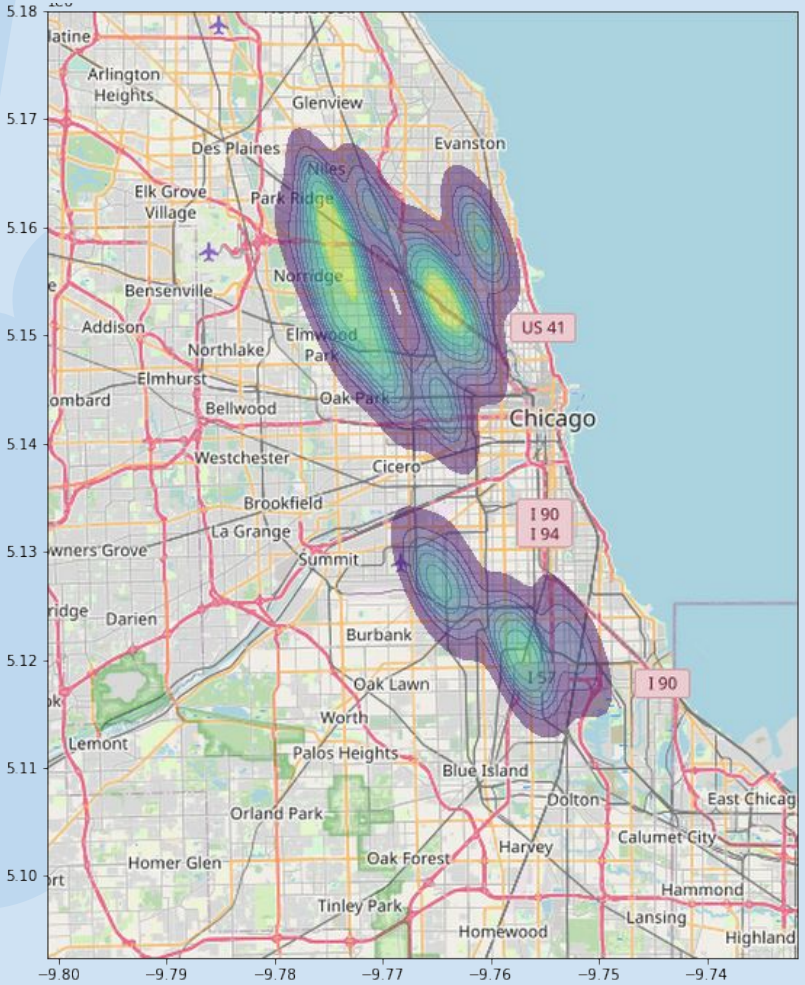
WNV Mosquito Density

- **Two** main hotspots for WNV mosquitoes



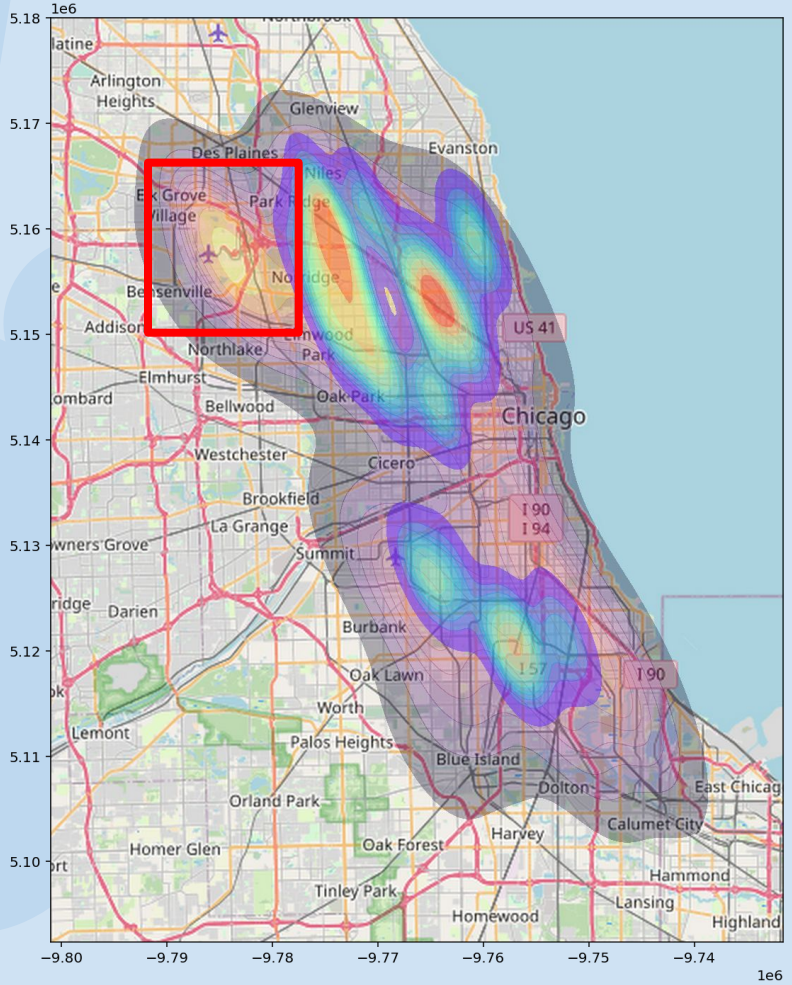
Spray Location

- Concentrated around **2 main areas**
- Density of spray increases in the yellow regions

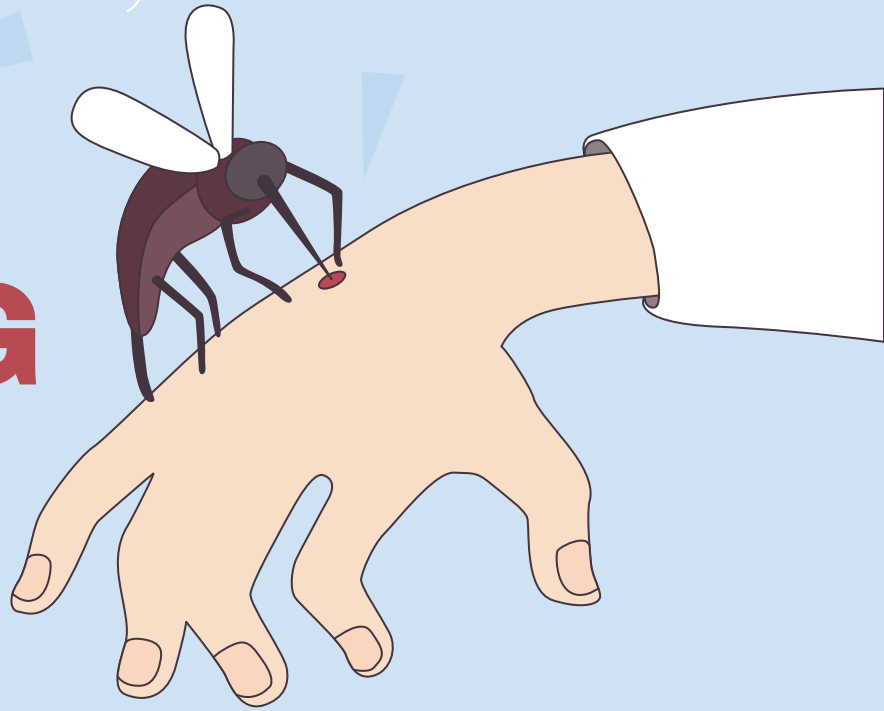


Overlap of WNV Mosquito and Spray Locations

- Most of the areas currently being sprayed with pesticide are also areas with WNV
- There is **one area** (in red) with high WNV density that is **not covered by pesticide**
- Recommended to include that area for pesticide spraying



MODELLING



Models & Resamplers

Models:

- Naive Bayes
- Logistic Regression
- K-Nearest Neighbor
- Random Forest
- Extra Tree
- Decision Tree
- ADA Boost
- Gradient Boost
- XGB
- Light GB
- SVM

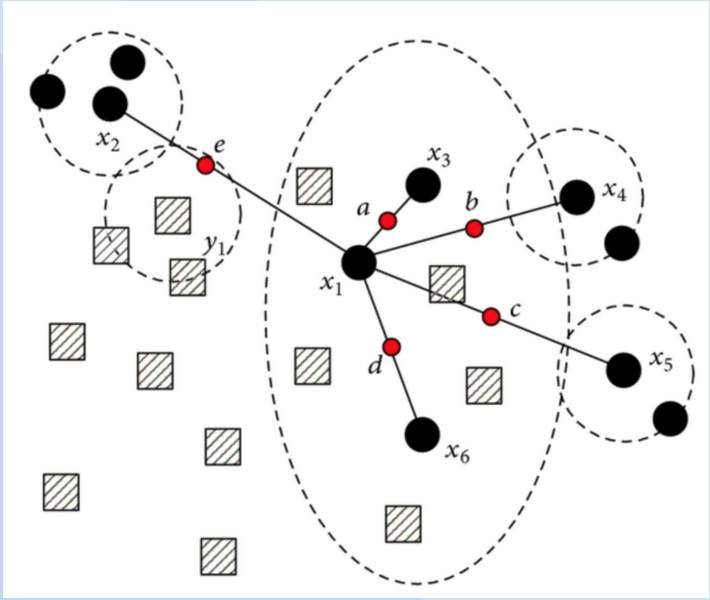
Resamplers:

- SMOTE
- ADASYN

SMOTE

Synthetic Minority Over sampling Technique

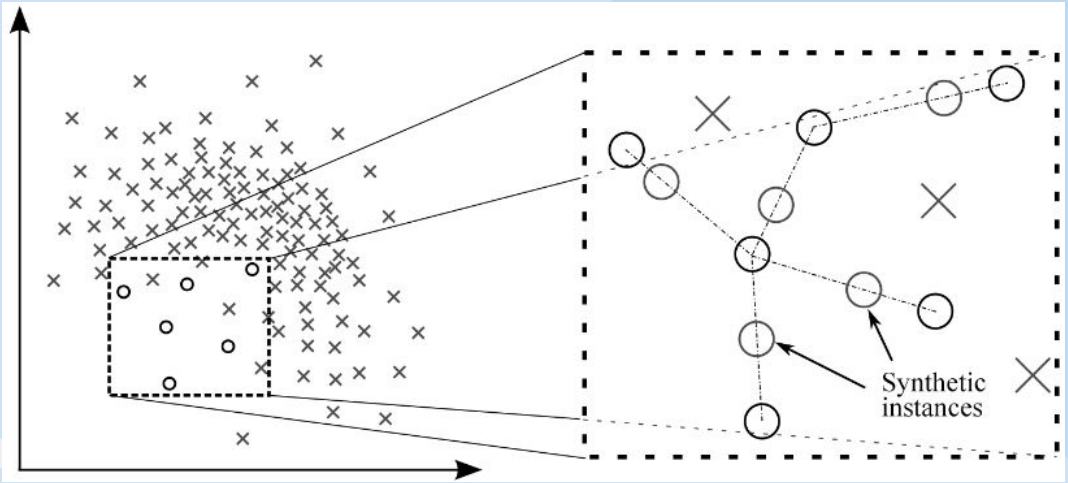
- K-Nearest Neighbours approach
- Draws a line between the neighbours of the minority class
- Generates random points on the lines



ADASYN

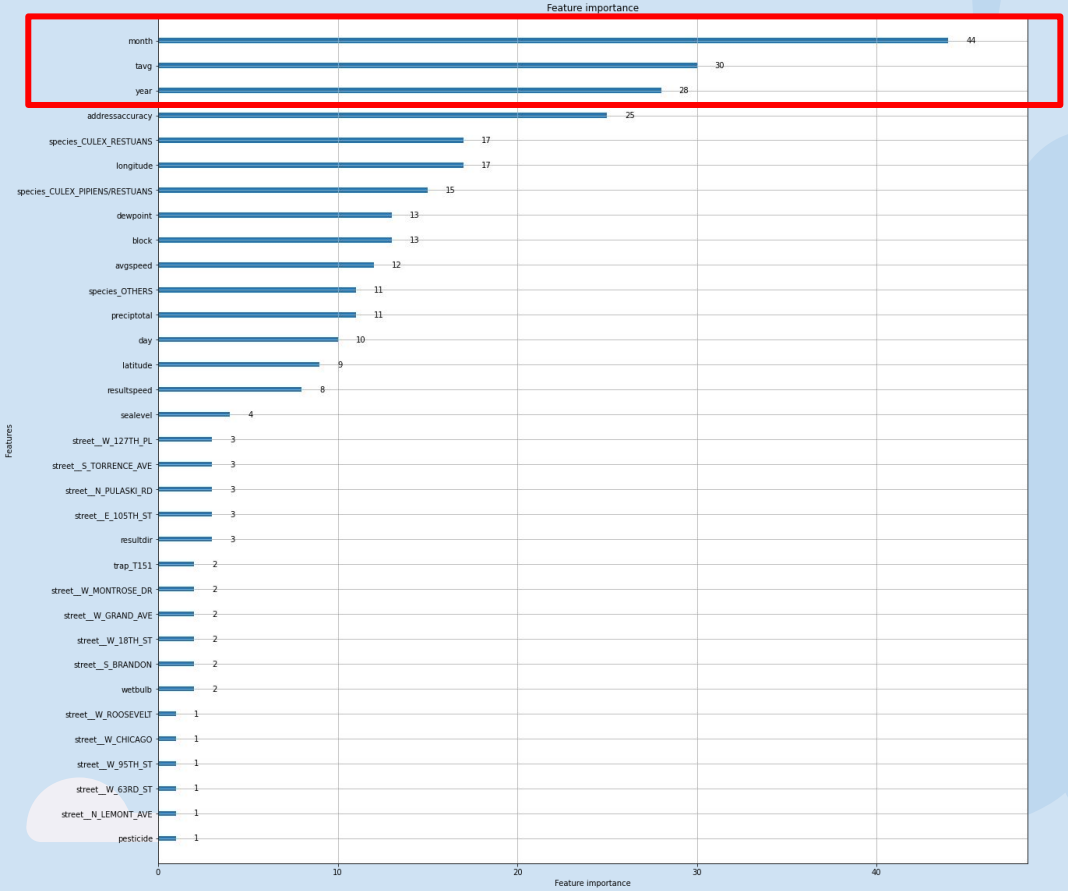
ADAptive SYNthetic

- K-Nearest Neighbours approach
- no assumptions made for the underlying distribution of the data
- Difference: generates the same number of synthetic samples for each original minority sample



Classifier	CV Score	ROC_AUC (train)	ROC_AUC (test)	Accuracy (train)	Accuracy (test)	Sensitivity	Specificity	Precision	Recall	Misclassification	F1 Score
Gaussian Naive Bayes	0.6838	0.77	0.67	0.77	0.67	0.76	0.51	0.08	0.76	0.48	0.1447619
Gaussian Naive Bayes (SMOTE)	0.6753	0.72	0.64	0.72	0.64	0.92	0.23	0.06	0.92	0.73	0.11265306
Gaussian Naive Bayes (ADASYN)	0.6329	0.69	0.6	0.69	0.6	0.86	0.25	0.06	0.86	0.72	0.11217391
Logistic Regression	0.76917	0.83	0.77	0.83	0.77	0	1	0	0	0.05	0
Logistic Regression (SMOTE)	0.7542	0.8	0.75	0.8	0.75	0.49	0.83	0.14	0.49	0.19	0.21777778
Logistic Regression (ADASYN)	0.752	0.79	0.75	0.79	0.75	0.44	0.84	0.13	0.44	0.18	0.20070175
Random Forest	0.8222	0.94	0.83	0.94	0.83	0	1	0	0	0.05	0
Random Forest (SMOTE)	0.8212	0.92	0.84	0.92	0.84	0.61	0.87	0.2	0.61	0.14	0.30123457
Random Forest (ADASYN)	0.8211	0.92	0.83	0.92	0.83	0.62	0.87	0.21	0.62	0.14	0.31373494
Decision Tree	0.6687	0.68	0.65	0.68	0.65	0	1	0	0	0.05	0
Decision Tree (SMOTE)	0.6871	0.64	0.61	0.64	0.61	0.91	0.33	0.07	0.91	0.64	0.13
Decision Tree (ADASYN)	0.7082	0.76	0.72	0.76	0.72	0.75	0.59	0.09	0.75	0.4	0.16071429
Extra Trees	0.8026	0.96	0.81	0.96	0.81	0	1	0	0	0.05	0
Extra Trees (SMOTE)	0.7959	0.92	0.79	0.92	0.79	0.46	0.87	0.17	0.46	0.15	0.24825397
Extra Trees (ADASYN)	0.795	0.91	0.8	0.91	0.8	0.49	0.87	0.17	0.49	0.15	0.25242424
Light Gradient Boost	0.8302	0.87	0.83	0.87	0.83	0	1	0	0	0.05	0
Light Gradient Boost (SMOTE)	0.8195	**0.87**	**0.82**	0.87	0.82	0.62	0.86	**0.2**	**0.62**	0.15	**0.30243902**
Light Gradient Boost (ADASYN)	0.8191	0.87	0.83	0.87	0.83	0.56	0.87	0.19	0.56	0.15	0.28373333
K-Nearest Neighbours	0.7165	0.93	0.73	0.93	0.73	0.05	0.99	0.35	0.05	0.05	0.0875
K-Nearest Neighbours (SMOTE)	0.7498	0.93	0.76	0.93	0.76	0.79	0.66	0.11	0.79	0.34	0.19311111
K-Nearest Neighbours (ADASYN)	0.7443	0.93	0.73	0.93	0.76	0.78	0.66	0.11	0.78	0.33	0.19280899
Gradient Boosting	0.8371	0.92	0.85	0.92	0.85	0.02	1	0.25	0.02	0.05	0.03703704
Gradient Boosting (SMOTE)	0.8239	0.91	0.84	0.91	0.84	0.38	0.93	0.22	0.38	0.1	0.27866667
Gradient Boosting (ADASYN)	0.8233	0.91	0.84	0.91	0.84	0.41	0.92	0.22	0.41	0.11	0.28634921
XG Boost	0.8354	0.9	0.85	0.9	0.85	0.04	1	0.46	0.04	0.05	0.0736
XG Boost (SMOTE)	0.8209	0.87	0.83	0.87	0.83	0.62	0.86	0.2	0.62	0.15	0.30243902
XG Boost (ADASYN)	0.8211	0.88	0.83	0.88	0.83	0.53	0.89	0.21	0.53	0.13	0.30081081
SVM	0.7806	0.97	0.8	0.97	0.8	0	1	0	0	0.05	0
SVM (SMOTE)	0.8136	0.88	0.82	0.88	0.82	0.69	0.84	0.19	0.69	0.17	0.29795455
SVM (ADASYN)	0.8143	0.88	0.82	0.88	0.82	0.67	0.84	0.19	0.67	0.17	0.29604651
ADABoost	0.7765	0.91	0.79	0.91	0.79	0.05	0.99	0.3	0.05	0.06	0.08571429
ADABoost (SMOTE)	0.7813	0.89	0.8	0.89	0.8	0.46	0.88	0.17	0.46	0.15	0.24825397
ADABoost (ADASYN)	0.7831	0.89	0.79	0.89	0.79	0.51	0.88	0.19	0.51	0.14	0.27685714

Best Features



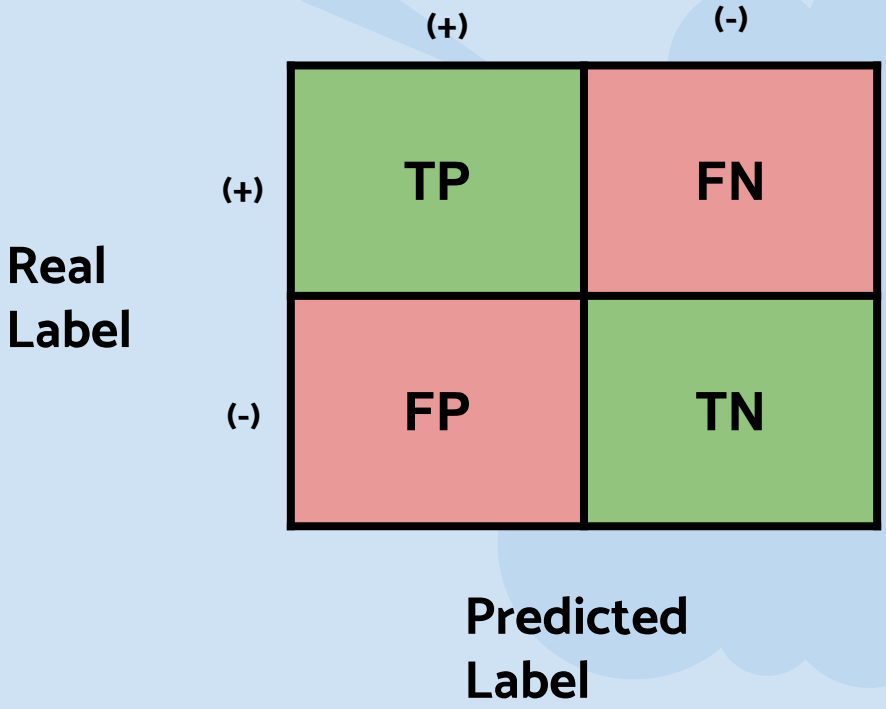
Top 3 features:

Month

Tavg

Year

PARAMETERS: Confusion Matrix



PARAMETERS: Confusion Matrix

	(+)	(-)
(+)	TP Predicted: ✓ WNV Actual: ✓ WNV	FN Predicted: ✖ WNV Actual: ✓ WNV
(-)	FP Predicted: ✓ WNV Actual: ✖ WNV	TN Predicted: ✖ WNV Actual: ✖ WNV

PARAMETERS: Confusion Matrix

	(+)	(-)
(+)	2234	496
(-)	48	103

Predicted: ✓ WNV
Actual: ✓ WNV

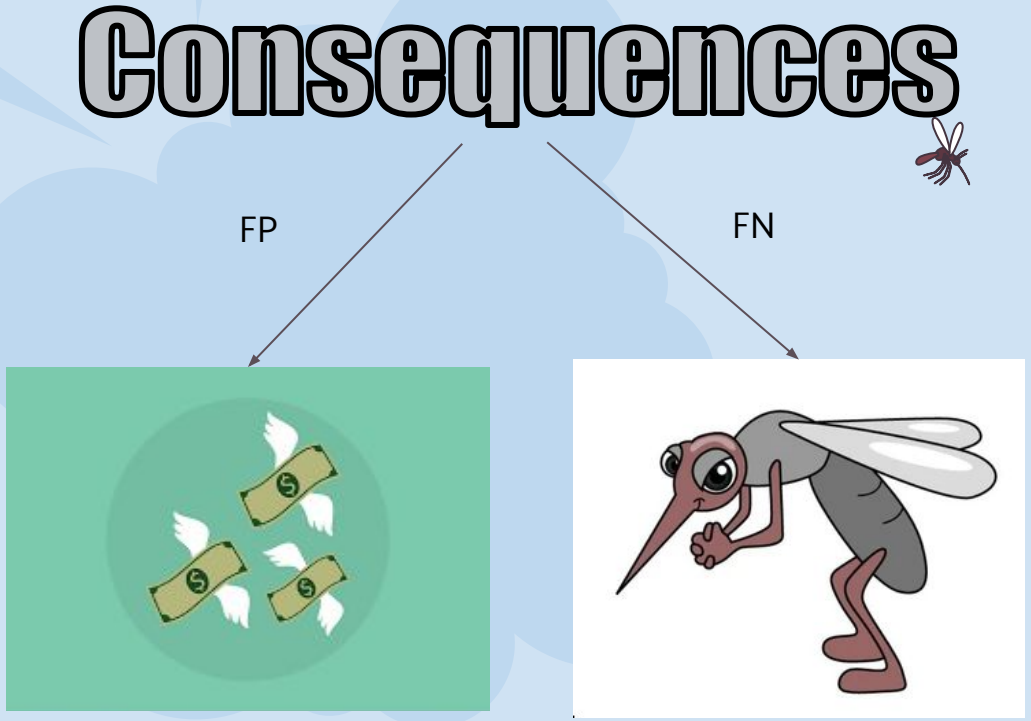
Predicted: ✖ WNV
Actual: ✓ WNV

Predicted: ✓ WNV
Actual: ✖ WNV

Predicted: ✖ WNV
Actual: ✖ WNV

Classifier Evaluation Metric

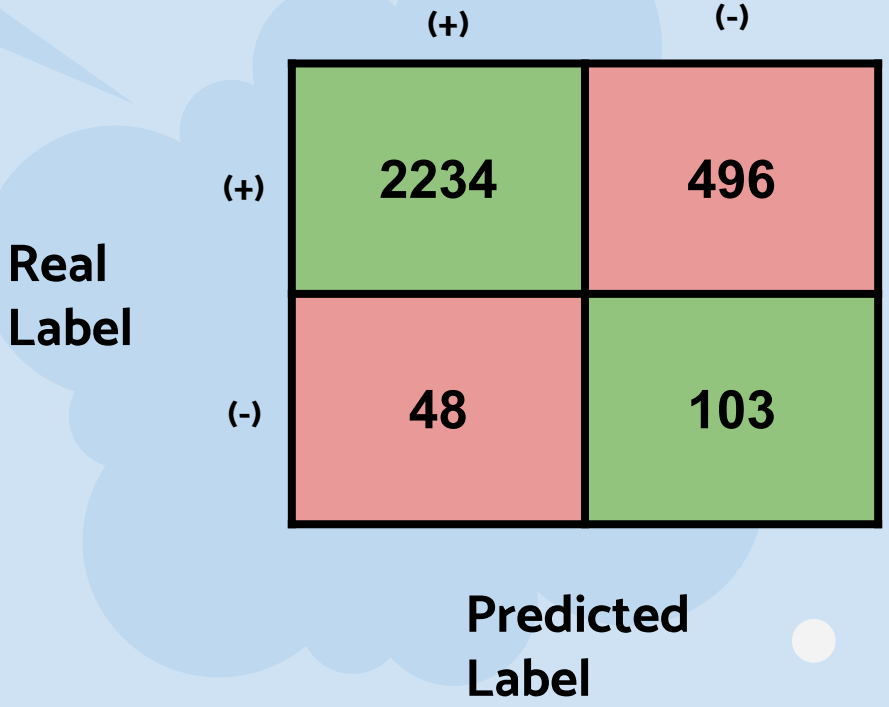
Precision	TP/(TP+TN)
Recall	TP/(TP+FN)
Sensitivity	TP/(TP+FP)
Specificity	TN/(TN+FN)



Best Model

Light GB with SMOTE

- AUC (Train): 0.87
- AUC (Test): 0.82
- Accuracy: 0.82
- Sensitivity: 0.62
- Specificity: 0.86
- Precision: 0.20
- Recall: 0.62
- F1 Score:



Conversion for Cost-Benefit Analysis

Confusion Matrix (Number)

Real \ Predicted	(+)	(-)
(+)	2234	496
(-)	48	103

Confusion Matrix (Proportion)

Real \ Predicted	(+)	(-)
(+)	0.775	0.172
(-)	0.017	0.036

Confusion Matrix (Costs)

Real \ Predicted	(+)	(-)
(+)	$78.85 + (21.7 * 0.281 * 1000)	$21.7 * 7.31 * 1000
(-)	$ 78.85	0

Confusion Matrix (Costs)

Real \ Predicted	(+)	(-)
(+)	$6176.51	0.172
(-)	$6176.51	0.036

Overall Total Cost = $ 32,085.55 per spray

Expected total cost per year = $US 802,138.75

COST BENEFIT ANALYSIS

COST

Indirect Cost

- Higher Tax Rates

Direct Cost

- Cost of Pesticides
- Other Miscellaneous Cost

Intangible Cost

- Lower Quality of Life

DIRECT COST

Direct Cost

- Cost of Pesticides
- Other Miscellaneous Cost

69km²

amount spent on aerial spraying in Chicago in 2020

USD797k

amount spent on spraying in Chicago in 2020

INDIRECT COST

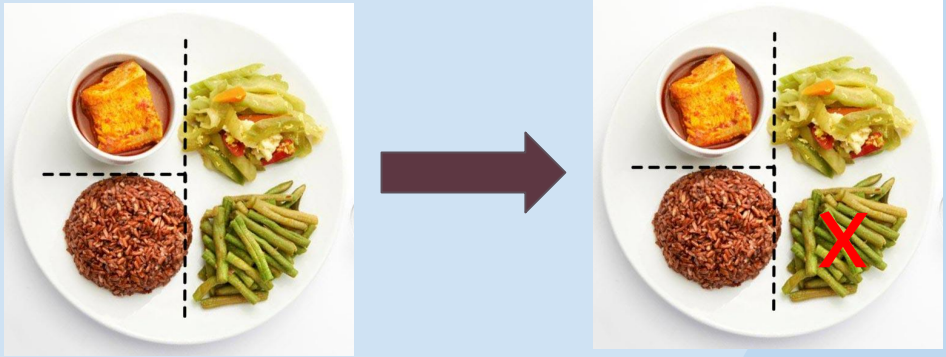
Indirect Cost

- Higher Tax Rates

- Shared cost of pesticides borne by both the state Illinois, Chicago city, and the residents of Chicago
- Eventual increase in higher taxes may be a huge **financial burden**
  - Especially those of the **lower-income group.**

INTANGIBLE COST

- Lower Income Group



Intangible Cost

- Lower Quality of Life

BENEFITS

Direct Benefits

Indirect Benefits

Intangible Benefits

Competitive Benefits

BENEFITS

Direct Benefits

Indirect Benefits

Intangible Benefits

Competitive Benefits

Decrease in probability:

- People contracting the West Nile Virus
- Unemployed residents in Chicago
  - Due to contracting the West Nile Virus or,
  - Being caregivers to the patients

BENEFITS

Direct Benefits

Indirect Benefits

Intangible Benefits

Competitive Benefits

- Decrease in Medical Fees
- Reduce stimulus cheque for patients who cannot work

BENEFITS

Direct Benefits

Indirect Benefits

Intangible Benefits

Competitive Benefits

Increase Living Satisfaction

- Live with **greater assurance** & **lower levels of fears** of contracting the West Nile Virus

BENEFITS

Direct Benefits

Indirect Benefits

Intangible Benefits

Competitive Benefits

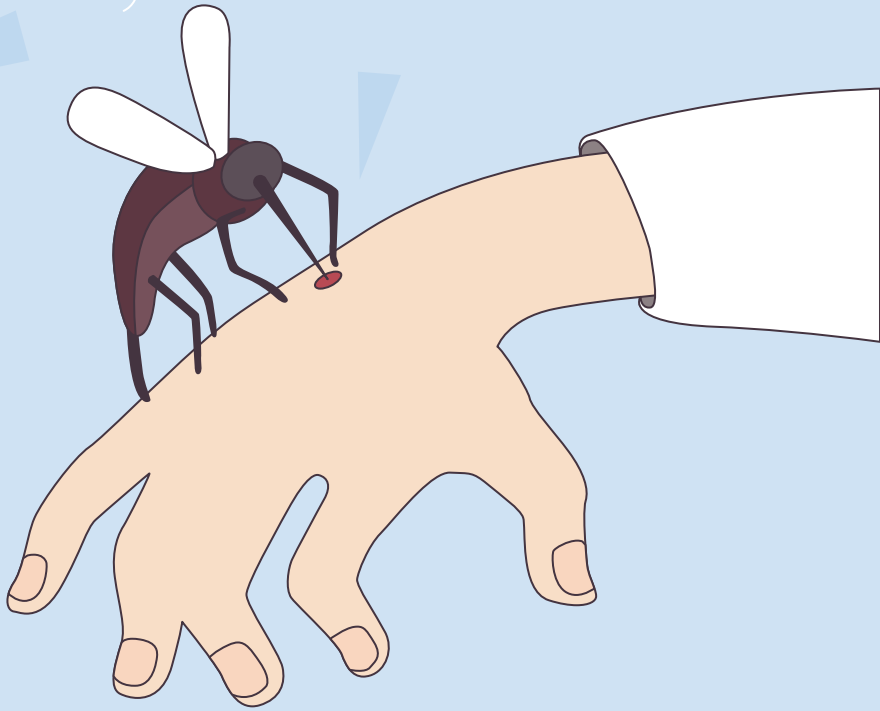
- Stand out from her neighbouring cities
- Promote as a **choice destination** for a safe summer vacation.

COST BENEFIT ANALYSIS

Cannot put a price tag on human lives!

- As compared to the COVID-19 pandemic, the benefits of preventing an add-on factor on top of an ongoing pandemic will definitely outweigh the cost.

RECOMMENDATIONS & FURTHER STEPS



RECOMMENDATIONS

Focus on Education

- Increase awareness on Social Media
- Start education early in schools

Change the Spray Locations

- Blocks 10, 11 & 76 densely populated with the presence of West Nile Virus, but were not sprayed.

RECOMMENDATIONS

Consider alternative pesticides

- Zenivex is an adulticide which is less effective than larvicides.

Utilize effective technologies

- Use drones to spray the pesticides at lower altitude for harder to reach targeted areas.

FURTHER STEPS

Research for less invasive solutions

- Research to optimally use wolbachia or other strains

Conduct decision tree / markov models

- Data on healthcare costs & quality-adjusted life years.
- Economic evaluation to calculate incremental cost-effectiveness ratio.

Account for environmental factors

- Study relationship between climate change and WNV transmission.
- Resistance to pesticides
- Public involvement in curbing WNV.

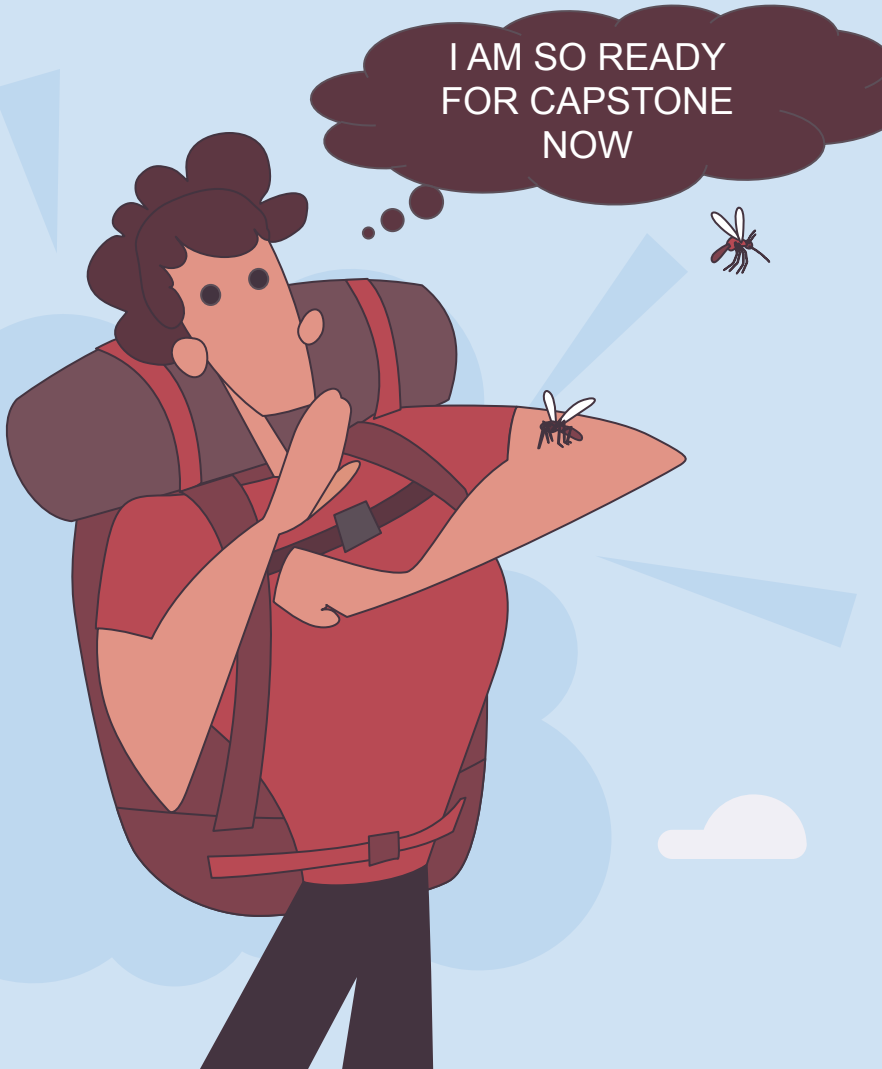
CONCLUSION

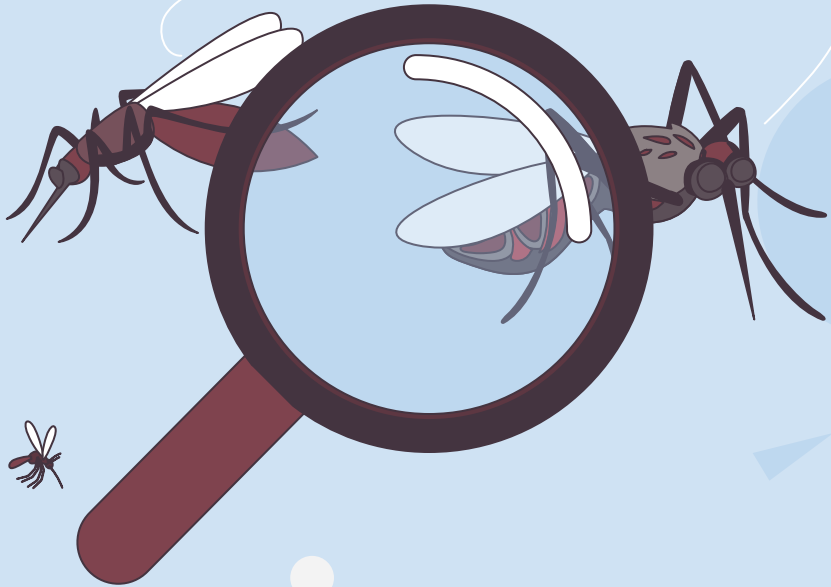
CONCLUSION:

- Forecast of seasonal WNV outbreaks
- Best Model: Light Gradient Boost Model
- Top determinants for the presence of WNV are: Month, Tavg, year
- Conducted a preliminary cost benefit analysis but requires new data to obtain a more detailed analysis
- Future directions with regards to climate change

Thank you!

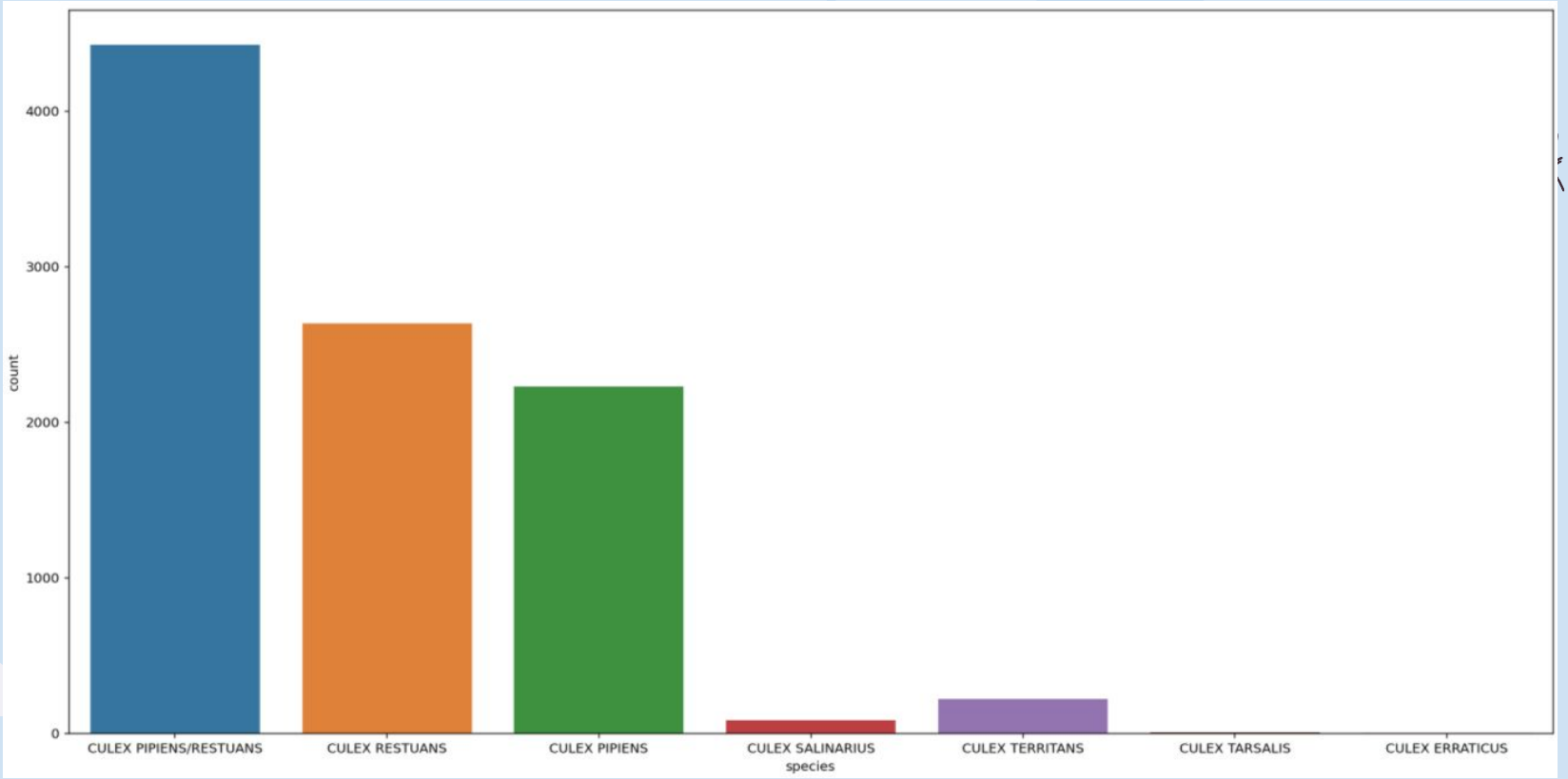
Any Questions?





Appendix

Distribution of all Mosquitos trapped



Pesticide and WNV occurrence