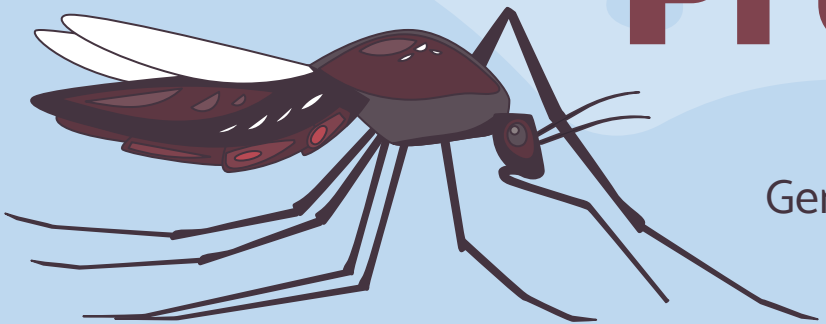
West Nile Virus Prediction

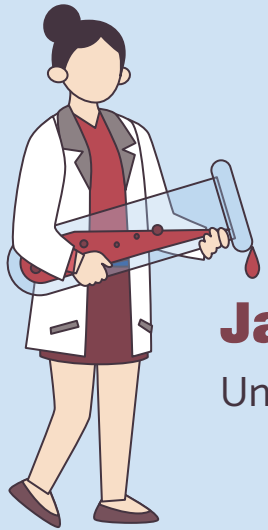


Group 3 - 'Pizzzzz Inc'
Gerald - Janet - Jun Siong - Xue Feng

Our team

Gerald

Finance Guy



Janet

Unpaid Intern

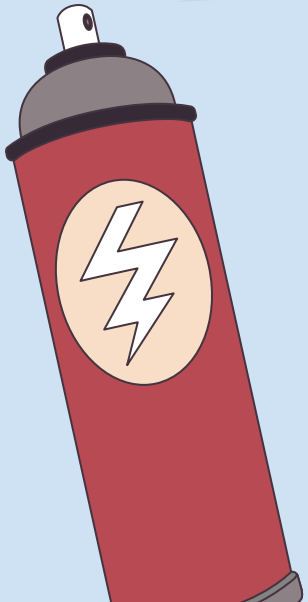
Jun Siong

Big Boss/ ~~Mad Scientist~~

Xue Feng

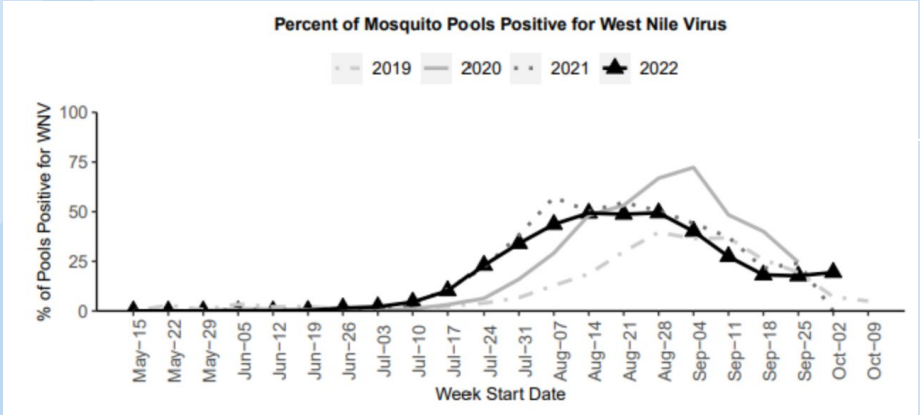
Code Debugger

Table of contents



Background

- WNV is the leading cause of mosquito-borne disease in the United States
- Jun (Summer) till after October (Fall)
- About 1 in 5 people who are infected, develop symptoms ranging from mild, flu-like symptoms, to neurological illnesses that might result in death
- By the end of 2002, Illinois had counted more human cases (884) and deaths (64) than any other state in the United States.



Year to Date Summary	3 Year Average	2022
Number of mosquitoes tested	273190	297250
Number of mosquito pools tested	7051	8088
Number of positive pools	1509	1830
Percent of pools testing positive	21.3%	22.6%
Number of communities with positive pools	67	72

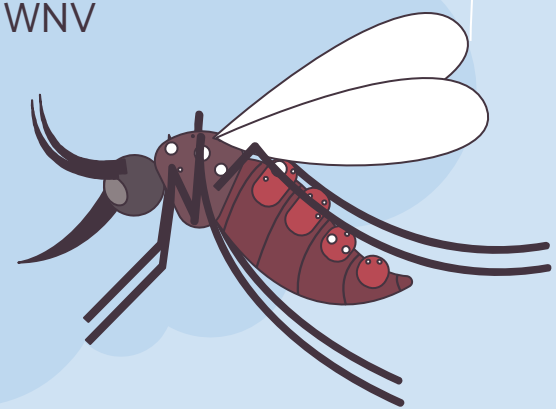
Problem Statement

Who?

We are Data Scientists working for Chicago & CPHD

How?

To combine weather, time and location features to develop a model predicting the presence of WNV



Why?

To combat the outbreak of WNV

What?

Effective usage of public funds to finance the spraying in order to reduce the number of WNV cases

Mosquito Life Cycle

7 to 10 days for an egg to develop into an adult mosquito

Phase 1

Eggs

Most hatch within 48 hours

Phase 2

Larvae

Live in water. Develop to pupae < 5 days

Phase 3

Pupae

Develop into adults within 2 to 3 days

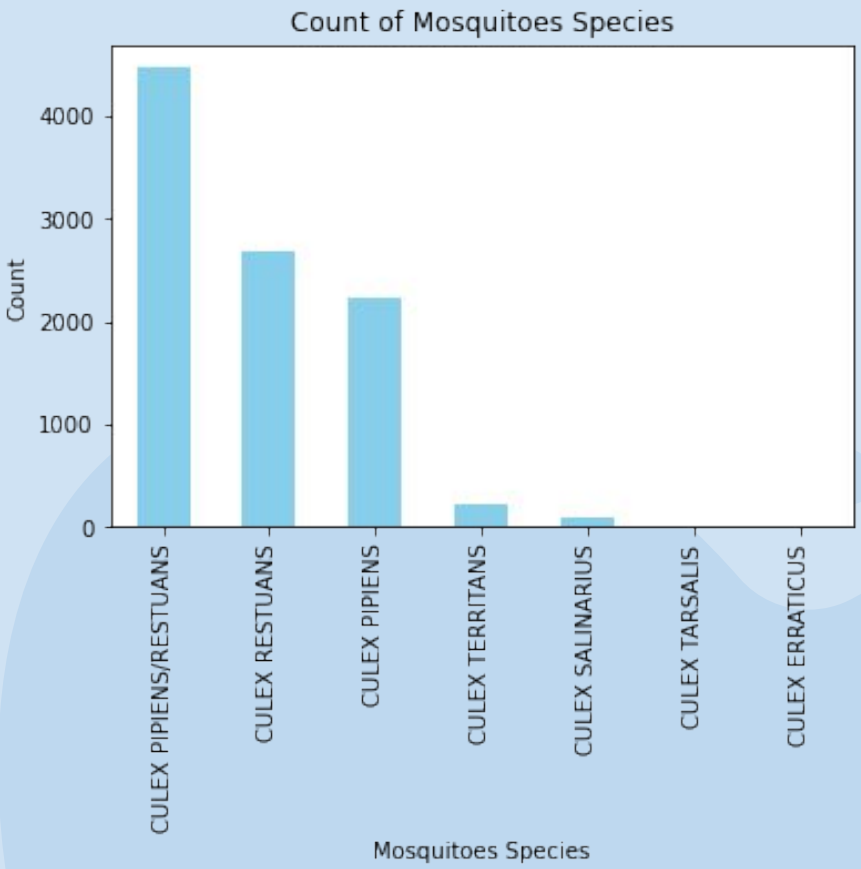
Phase 4

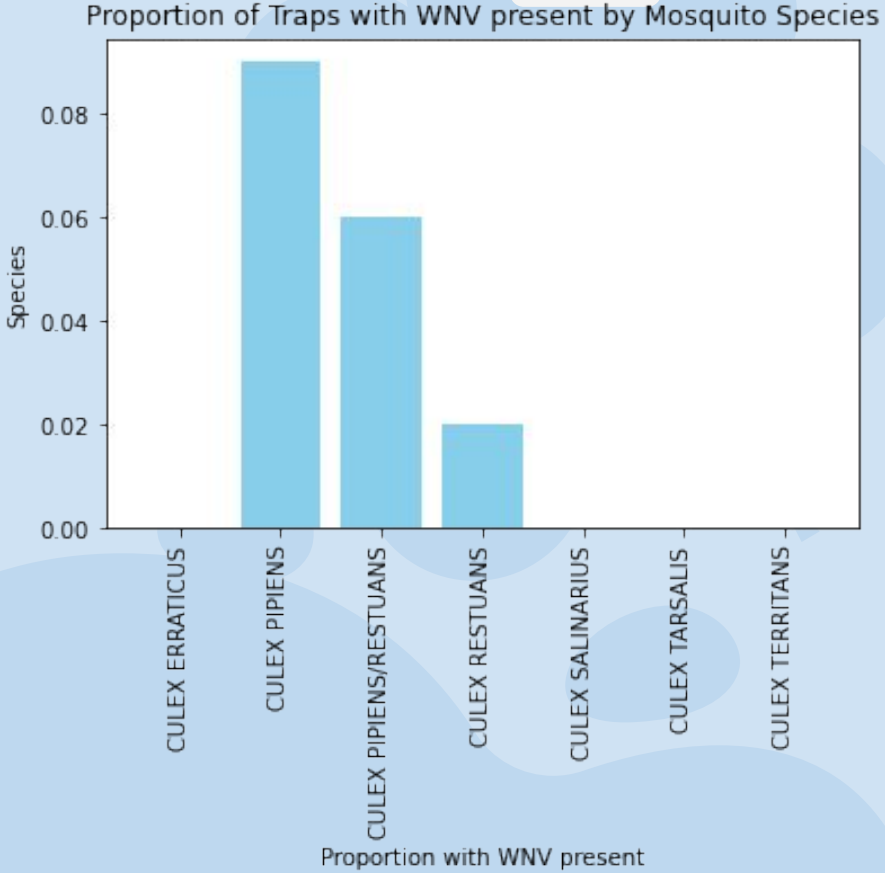
Adult

02a Exploratory Data Analysis

Train Dataset

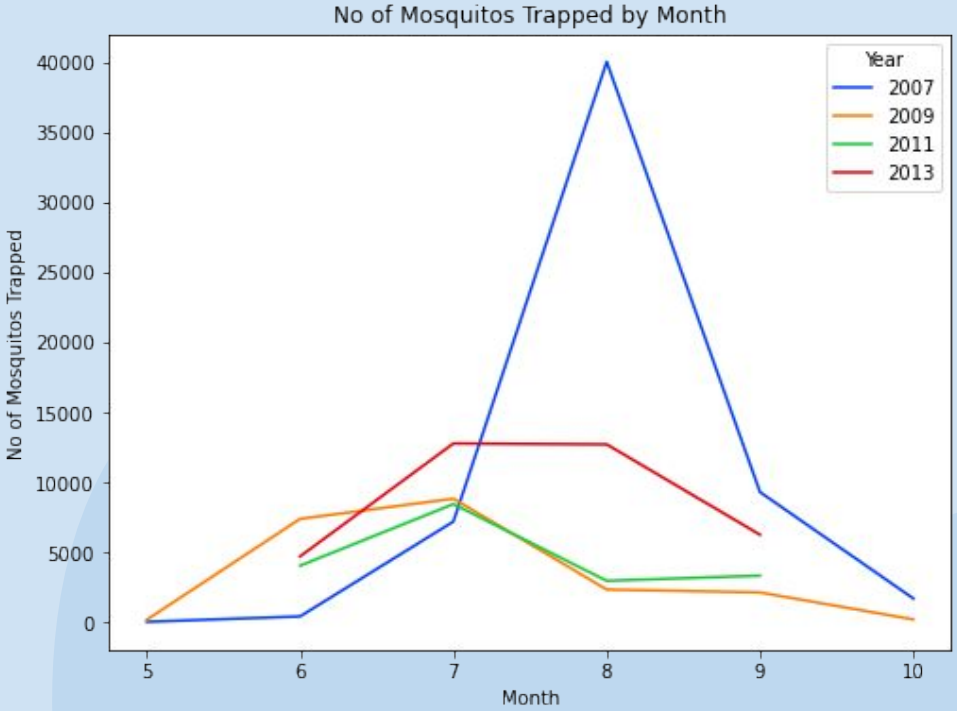
Mosquito Species in Traps & Species with WNV

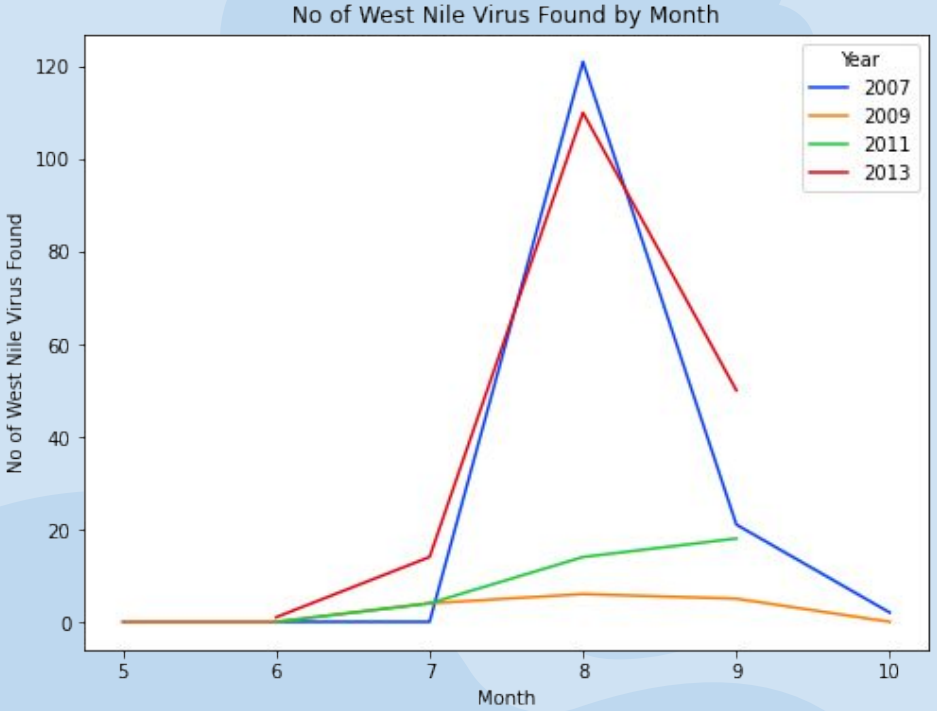




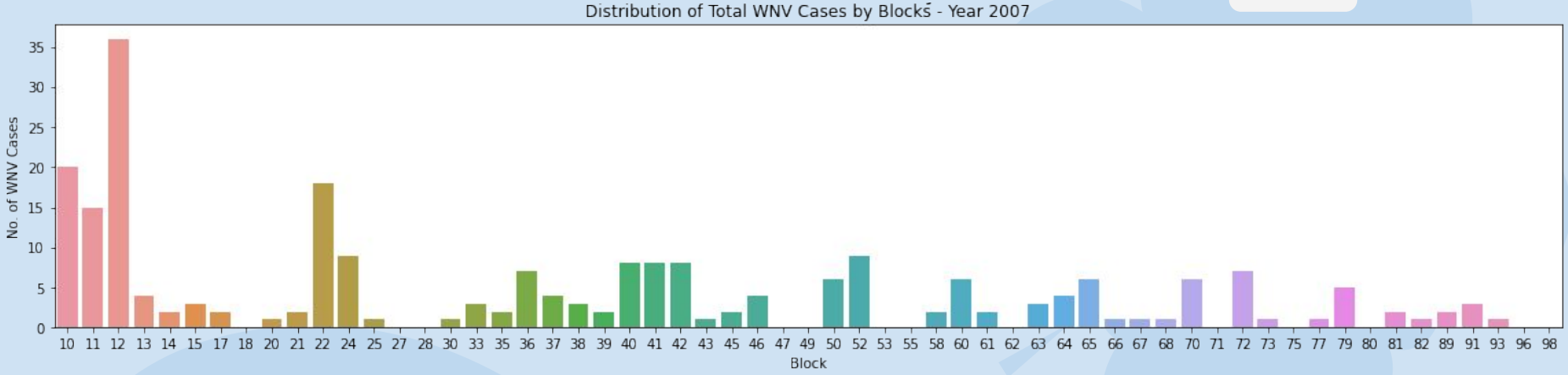
Train Dataset

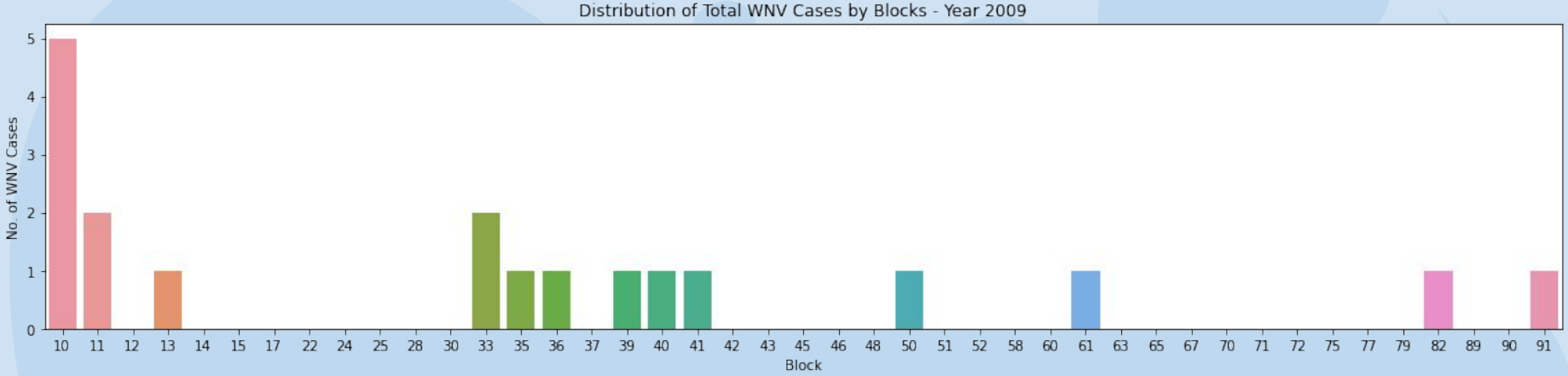
Mosquito trapped & WNV detected per month

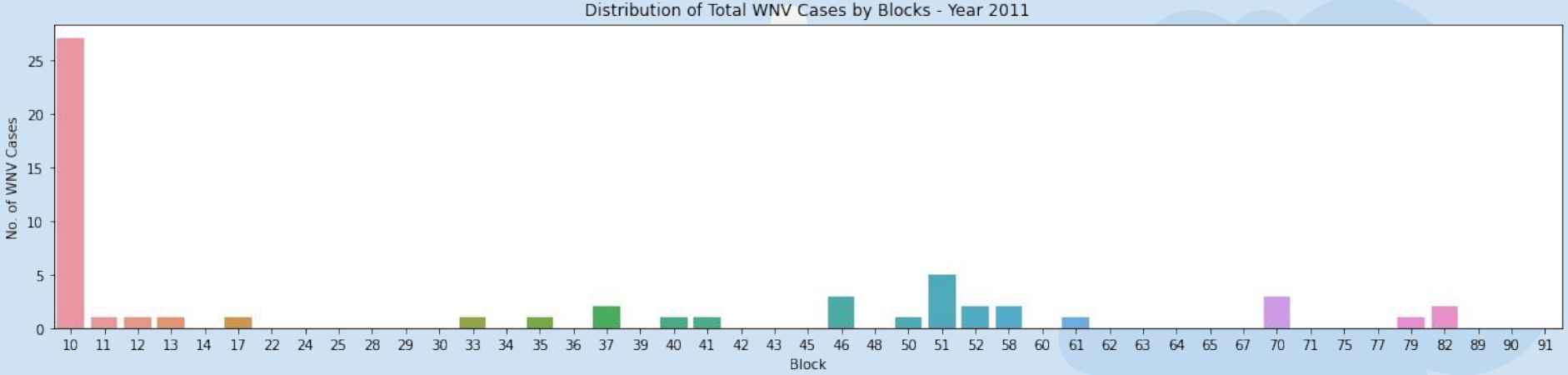


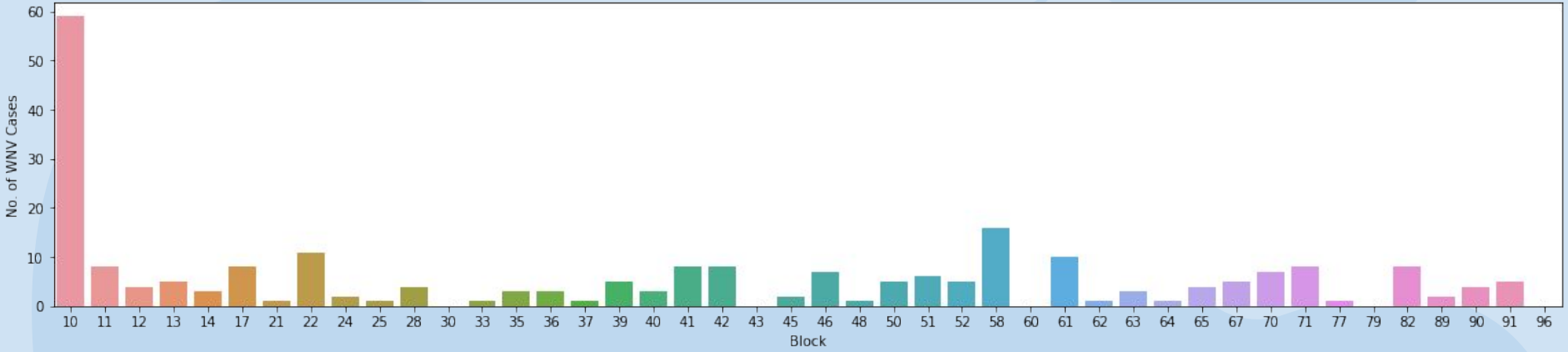


Train Dataset
WNV cases by blocks

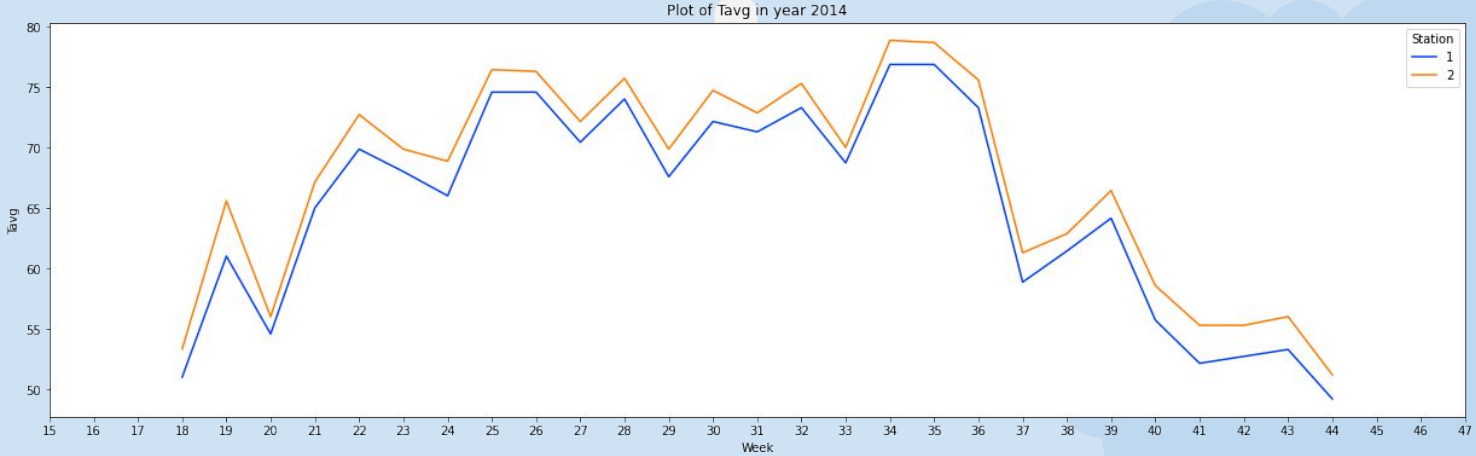




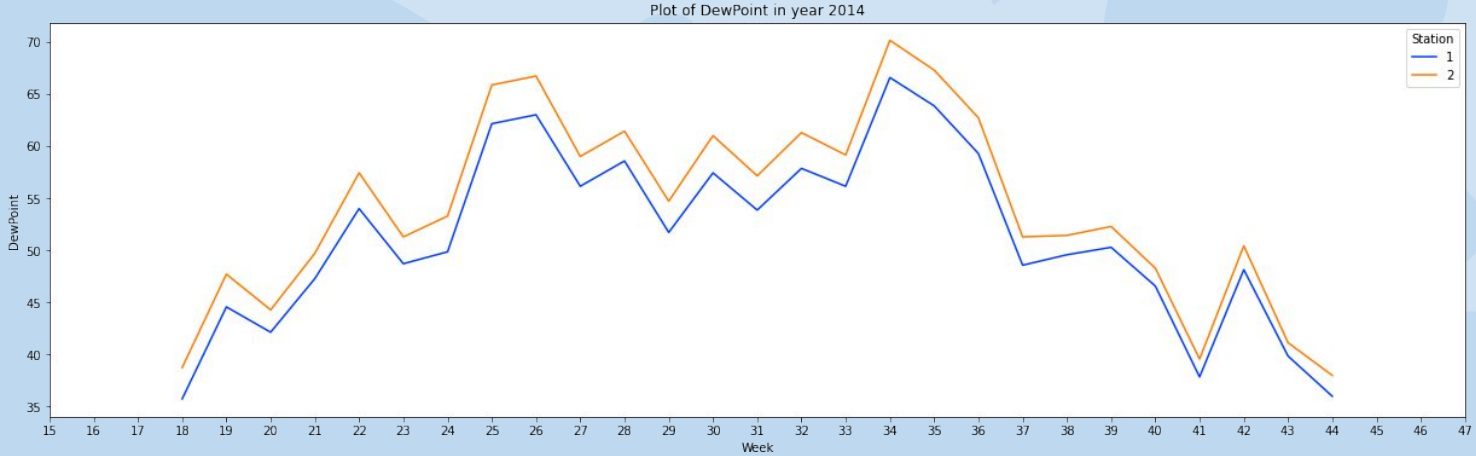

Distribution of Total WNV Cases by Blocks - Year 2011
No. of WNV Cases
0
5
10
15
20
25
10 11 12 13 14 17 22 24 25 28 29 30 33 34 35 36 37 39 40 41 42 43 45 46 48 50 51 52 58 60 61 62 63 64 65 67 70 71 75 77 79 82 89 90 91
Block


Distribution of Total WNV Cases by Blocks - Year 2013
No. of WNV Cases
0
10
20
30
40
50
60
10 11 12 13 14 17 21 22 24 25 28 30 33 35 36 37 39 40 41 42 43 45 46 48 50 51 52 58 60 61 62 63 64 65 67 70 71 77 79 82 89 90 91 96
Block

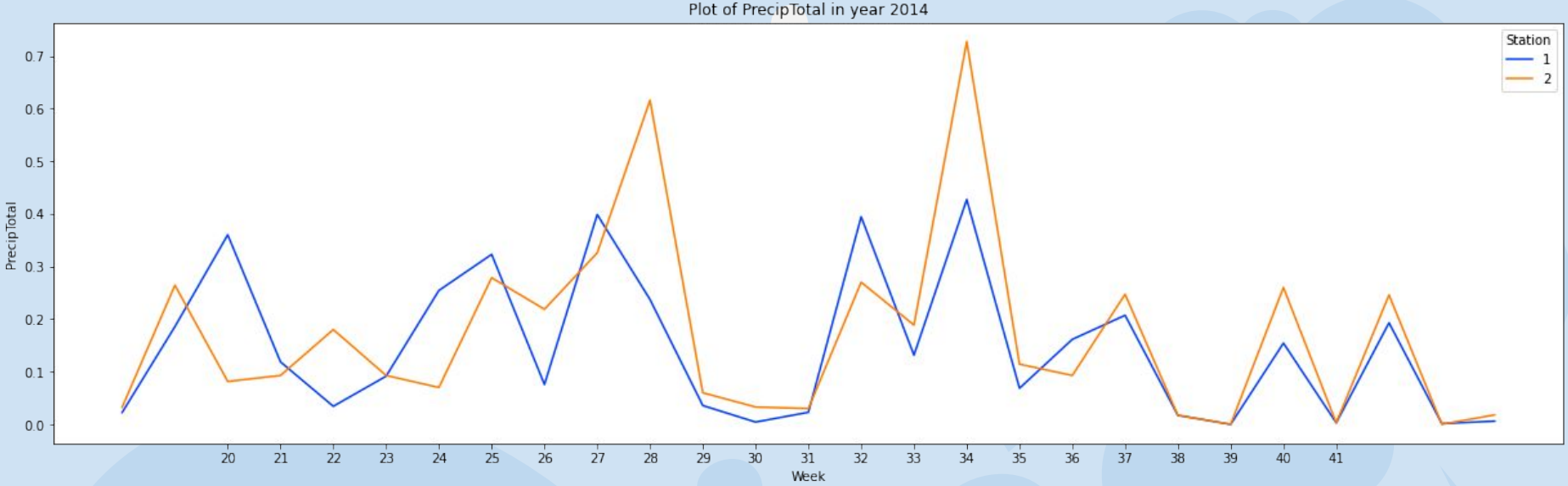
Weather Dataset - Average Temperature



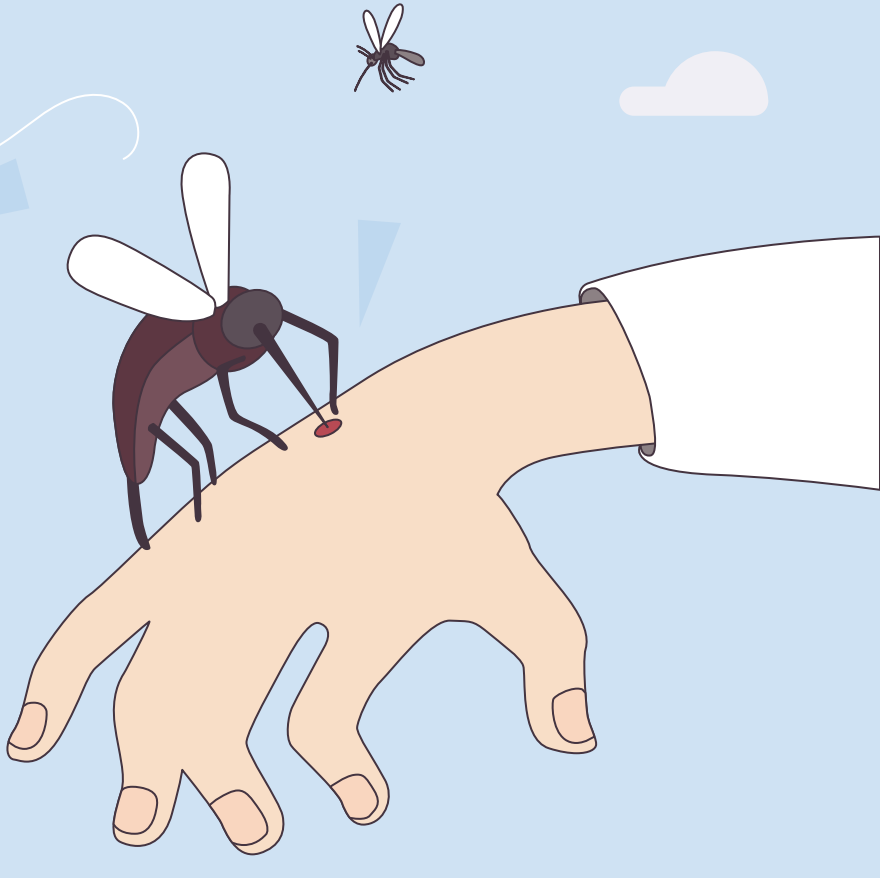
Weather Dataset - DewPoint



Weather Dataset - Precipitation

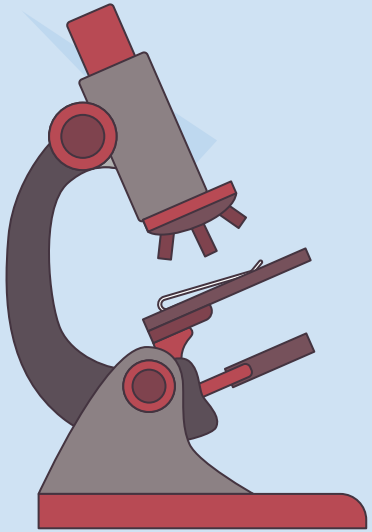


02b Feature Selection



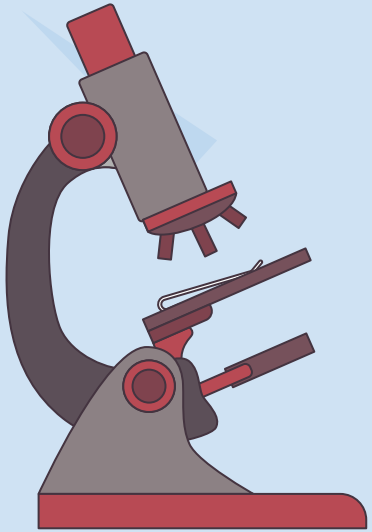
Test & Train Dataset

No. of Mosquitos (Test & Train Dataset)	To be removed as it is not found in the test set
Traps (Test & Train Dataset)	Satellite traps will be defaulted to their parent traps
Unspecified Culex (Test Dataset)	Change to CULEX PIPIENS/RESTUANS (99% of mosquito species)



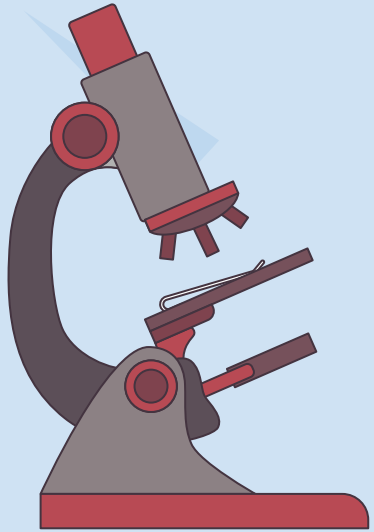
Weather Dataset (Missing Data)

Tavg	To get the average of Tmax & Tmin
Wetbulb, PrecipTotal, StnPressure, SeaLevel, AvgSpeed	To get data from alternate Weather Station



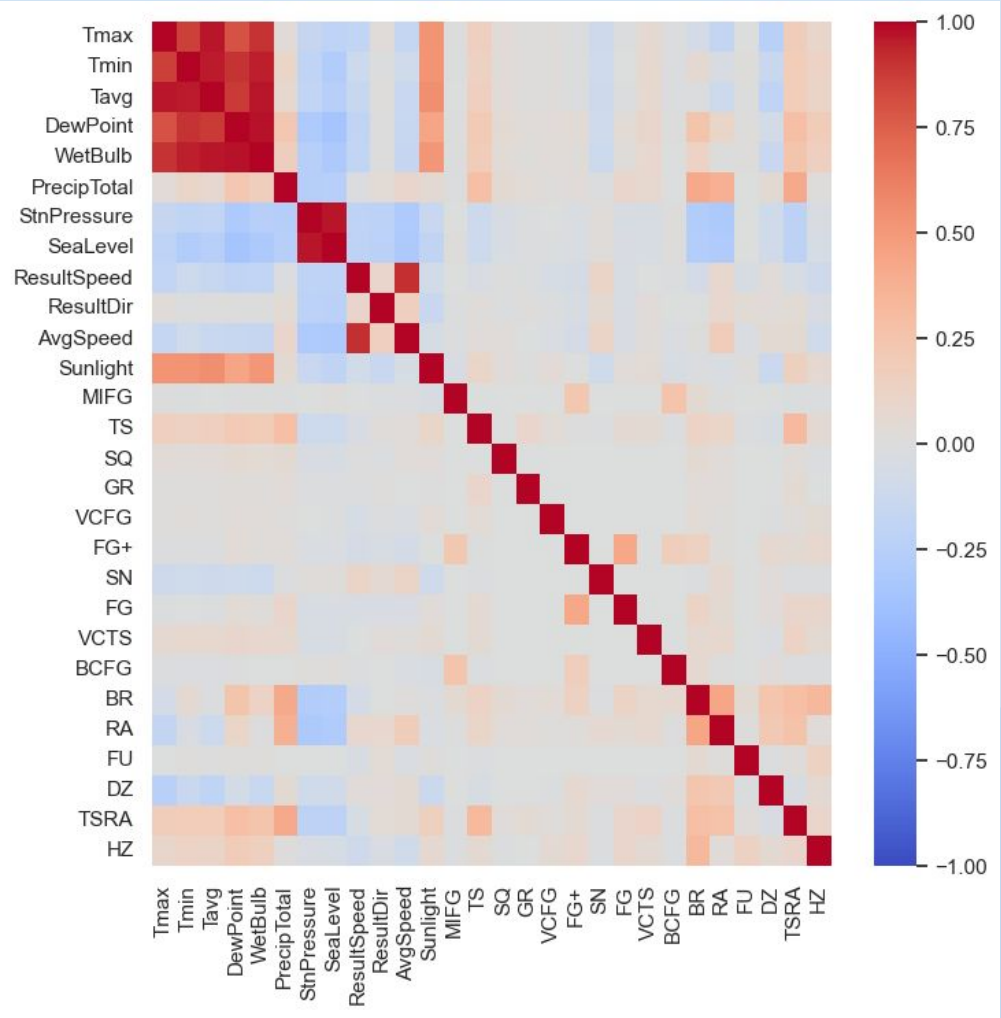
Weather Dataset (Feature Selection & Engineering)

Heat, Cool, Depart	To be removed as they are irrelevant
Depth, Snowfall & Water1	To be removed as most of data are 0 or missing
Sunrise and SunSet	To convert to Daylight hours



Weather Dataset (CodeSum)

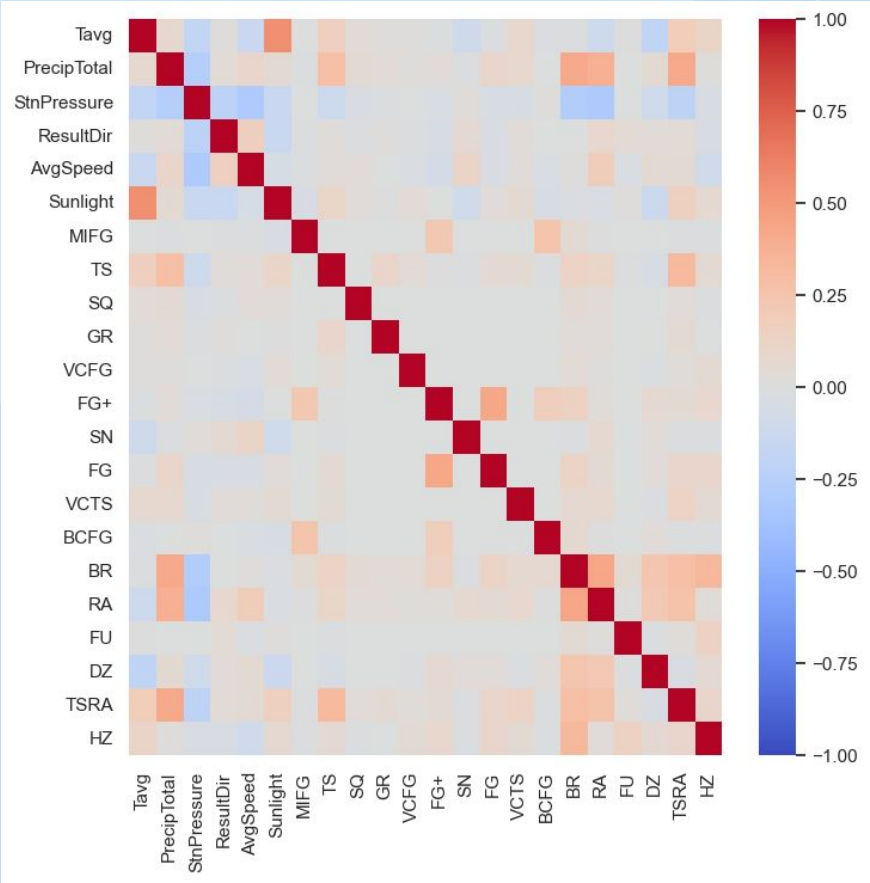
	HA	BR	DZ	FG	TS	FU	GR	HZ	.....	RA	SN
BR HZ	0	1	0	0	0	0	0	1		0	0
RA SN BR	0	1	0	0	0	0	0	0		1	1
...											
DZ HA FG	0	0	1	1	0	0	0	0		0	0



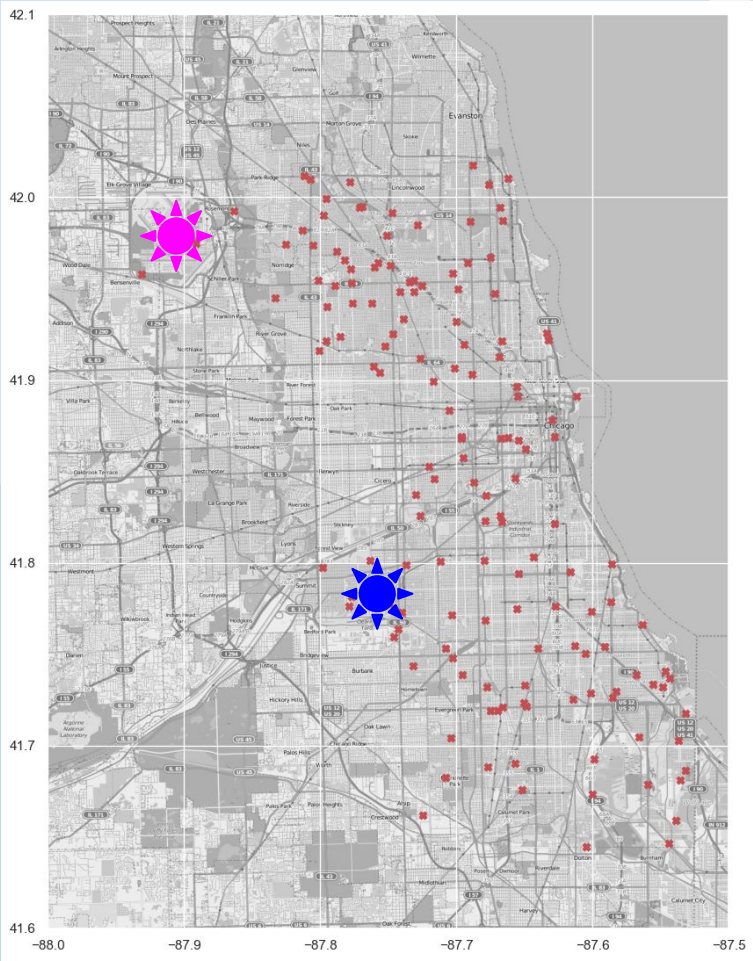
Weather Dataset (Heatmap)

- **Tmax, Tmin, Tavg, DewPoint, and WetBulb**
  - Keep Tavg
  - Drop: DewPoint & WetBulb
- **StnPressure and SeaLevel**
  - Keep: StnPressure
  - Drop: SeaLevel
- **AvgSpeed and ResultSpeed**
  - Keep: AvgSpeed
  - Drop: ResultSpeed

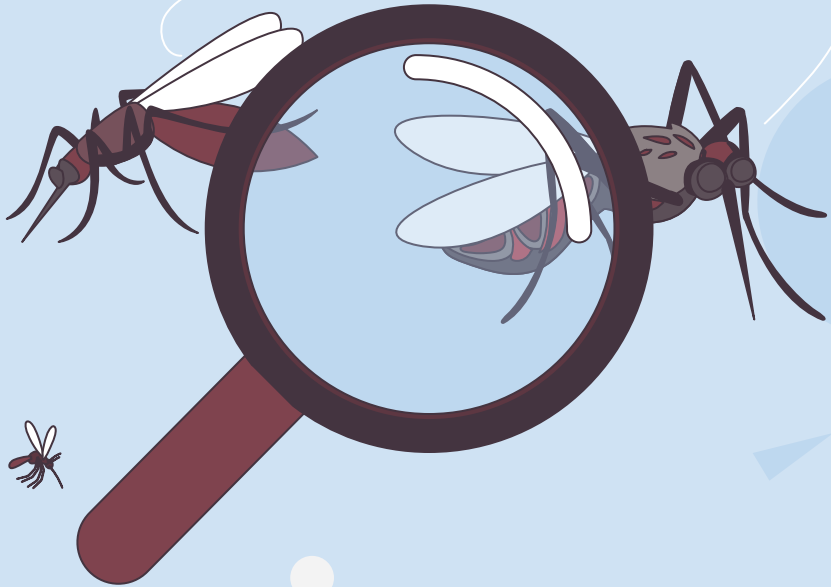
Final Heatmap on Weather Features



Combining Weather & Train Dataset



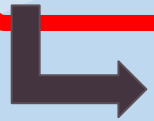
- Station 1 & Station 2 has different weather conditions
- Assign weather dataset to train dataset based on the weighted average



03 Modelling

Quick Recap - Our Data Thus Far

Category	Variables Available
Trap	Trap ID, Mosquito Species Captured
“CodeSum”	Exceptional Weather Events
Weather	Temperature, Precipitation, Pressure, Sunlight Hours, Wind



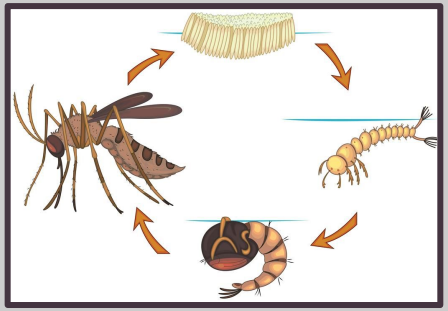
Feature Engineering Alert!

Mosquito: Life and Birth

MOSQUITO STAGES: COMING OUT FROM WINTER

Waking Up from Hibernation

+



Egg to Larvae to an Adult

+

Getting Into the Mood for Love

Bioclock / Weather Triggered

Egg to Adult: 8 to 10 days

Female Life Span: 30 Days

1 day of Good Weather ≠ Peak Summer Mosquito Activity

Rolling Averages of Weather-Related Variables Used!

Rolling Averages of Weather-Related Data

Gets Better and Better....

Dataset	Logistics Regression Performance	
	Train AUC	Test AUC
w/ Single Day Weather	0.805	0.763
w/ 10 -Days Rolling Average	0.853	0.797
w/ 20-Days Rolling Average	0.860	0.818
w/ 30-Days Rolling Average	**0.867**	**0.827**

Best in Class!!

Modelling and Analysis

(1) Candidates from Model Scan:

- Logistics Regression
- Gradient Boosting Decision Tree
- *Light* Gradient Boosting Trees

(2) Hyperparameters Tuning:

- Logistics Regression Chosen *(with Weights Balancing)*
- Not Best AUC, *But* Most Explainable

Logistics Regression

(w/ 30 Days Rolling Average Weather Related Variables)

0.867	0.827	0.754
Train AUC	Test AUC	Kaggle Score

Most Important Variables:

\+ Precipitation (Rain / Drizzle)

\- Fog

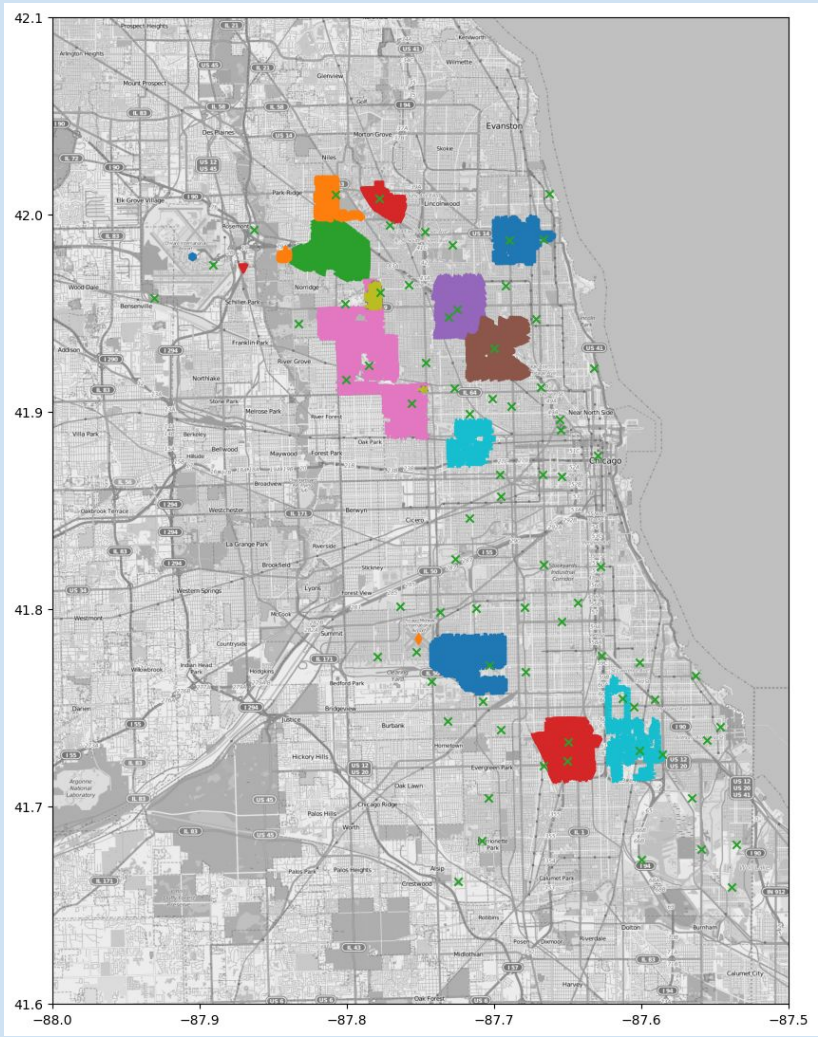
04

Cost-Benefit Analysis

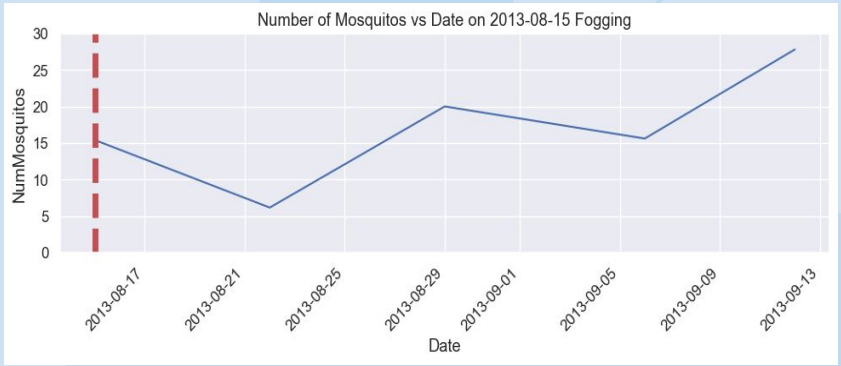
Spray Locations

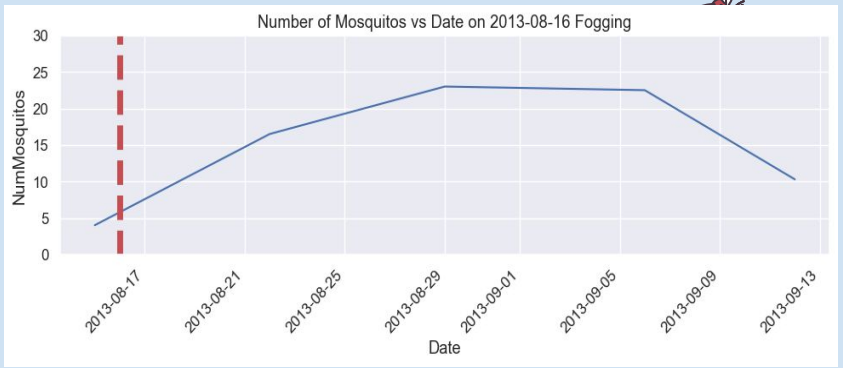
Locations:

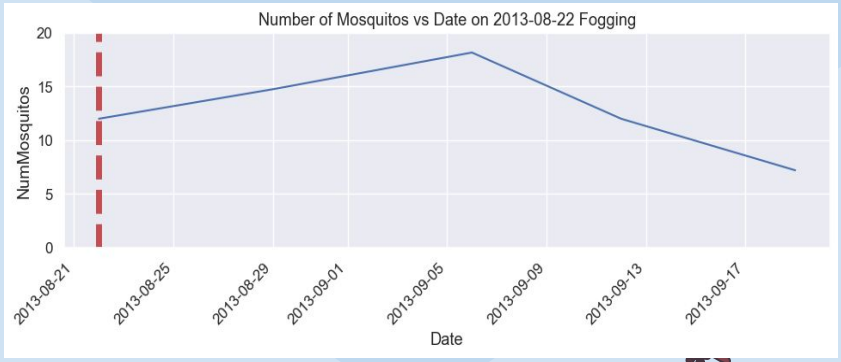
- Each spray data is almost independent to each other.
- Culex mosquitoes don't fly long distances, but have been known to fly up to 2 miles (3.2 km).
- Thus, we will only consider traps that is within the mosquito flight range from the spray

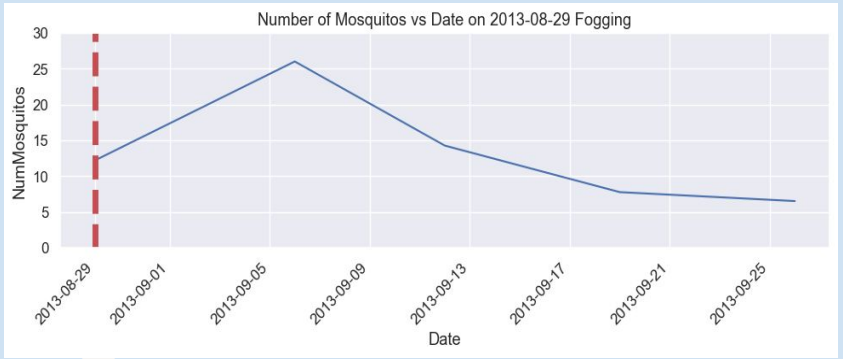


Efficacy of Spray









Cost of Zenivex E4

	Price (gallon)	Pounds AI/gallon	Price per Pound	Application Rate/Acre	Cost/Acre	Annual Acres Treated	Annual Cost
275 gal Zenivex® E20	$282.00*	1.48	$190.54	.0035	$0.67	20,000	$13,338
275 gal Zenivex® E4	$78.85*	.3	$262.83	.0035	$0.92	20,000	$18,398
2.5 gal Zenivex® E20	$296.00*	1.48	$200.00	.0035	$0.70	20,000	$14,000
2.5 gal Zenivex® E4	$80.75*	.3	$269.17	.0035	$0.94	20,000	$18,842

TREATING 20,000 ACRES

Annual Cost

Economy & Politics › Economy

GDP of the Chicago metro area from 2001 to 2020

(in billion chained 2012 U.S. dollars)

Additional Information

Show source

Science News *from research organizations*

West Nile virus hospitalizations cost nearly $800 million in U.S. since 1999, study shows

Date: February 10, 2014

Source: American Society of Tropical Medicine and Hygiene

Summary: In a study of the economic impact of West Nile virus in the United States, a research team reports that in the 14 years since the virus was first detected in New York, hospitalized cases of WNV disease have cost a cumulative $778 million in health care expenditures and lost productivity.

Estimated inpatient and outpatient economic costs of WNND cases, Sacramento County, California, 2005*

Item	Cost per case†	No. cases to which cost applies‡	% Cases to which cost applies§	Total cost for all cases	Total cost if treatment/service were used in all cases
Inpatient treatment costs	**$33,143**	**46**	**100**	**$1,524,570**	**$1,524,570**
Outpatient costs	Cost per case¶				
Outpatient hospital treatment	$333	17	36	$5,668	$15,337
Physician visits	$450	46	100	$20,708	$20,708
Outpatient physical therapy	$909	46	100	$41,810	$41,810
Occupational therapy	$4,037	3	7	$12,111	$185,699
Speech therapy	$588	1	1	$588	$27,032
Total				$80,885	$290,586
Nursing home costs	Cost#				
Nursing home stay**	$190	2	4	$36,195	$36,195
Transportation	$65	46	100	$2,977	$2,977
Home health aides, babysitters, etc.	$1,569	7	14	$10,983	$505,211
Total				$50,154	$544,383
Total for WNND				$2,140,409	$2,844,339

COST-BENEFIT ANALYSIS

Spray

$ 0.6 MILLION

Based on Chicago Land Area
Spray Twice a Year

Hospitalization

$ 1.2 MILLION

Based on 26 cases
reported in Illinois

Prevention

05

Conclusion & Recommendation

Model	Training AUC	Testing AUC	Kaggle AUC
Logistic Regression Model (30-day Rolling Average)	86.7%	82.7%	75.4%

Recommendation	
Direct Research Effort into Understanding Mosquito Hibernation Behaviour	Recommend further studies into the behaviour of mosquito coming out of hibernation. This can potentially lead to the development of a model with far stronger predictive capability.
Examination of Measures Beyond Spraying	Public education, improvement to drainage systems, and strict enforcement of fines for public violaters are effective in curbing WNV rates.

Thank You

Don’t Say, Just Spray

and educate