West Nile Virus

Here is where your presentation begins

INTRO

01

Introduction

The West Nile Virus (WNV) is transmitted to people via infected mosquitos. Around 20% of people who become infected with the virus develop symptoms ranging from a persistent fever, to serious neurological illnesses that can result in death.

The spread of WNV is usually seasonal. Patients usually fall ill around July–September, and no. of patients usually peaks in August

In 2002, the first human cases of West Nile virus were reported in Chicago. By 2004 the City of Chicago and the Chicago Department of Public Health (CDPH) had established a comprehensive surveillance and control program that is still in effect today.

Every week from late spring through the fall, mosquitos in traps across the city are tested for the virus. The results of these tests influence when and where the city will spray airborne pesticides to control adult mosquito populations.

PROBLEM STATEMENT

02

PROBLEM STATEMENT

The deadly, mosquito-borne West Nile Virus has long infested the City of Chicago. It is imperative to develop an accurate predictive model that could identify the key contributing factors that leads to proliferation of the virus. The model and the insights drawn could help the authority to predict future outbreaks of West Nile Virus and thus effectively allocate resources to mitigate it.

Logistic Regression, Naive Bayes, Extra Trees and XGboost models will be used to model the data and accuracy and recall score will be used to evaluate and find the best model for production and prediction.

DATA CLEANING

03

Data Cleaning and Preprocessing

Dropped

- Dropped redundant columns
- Dropped duplicate rows

Combined

- Combined Spray, Train and Weather data

Encode

- One hot encoded some categorical features
- Encoded others with relevant values

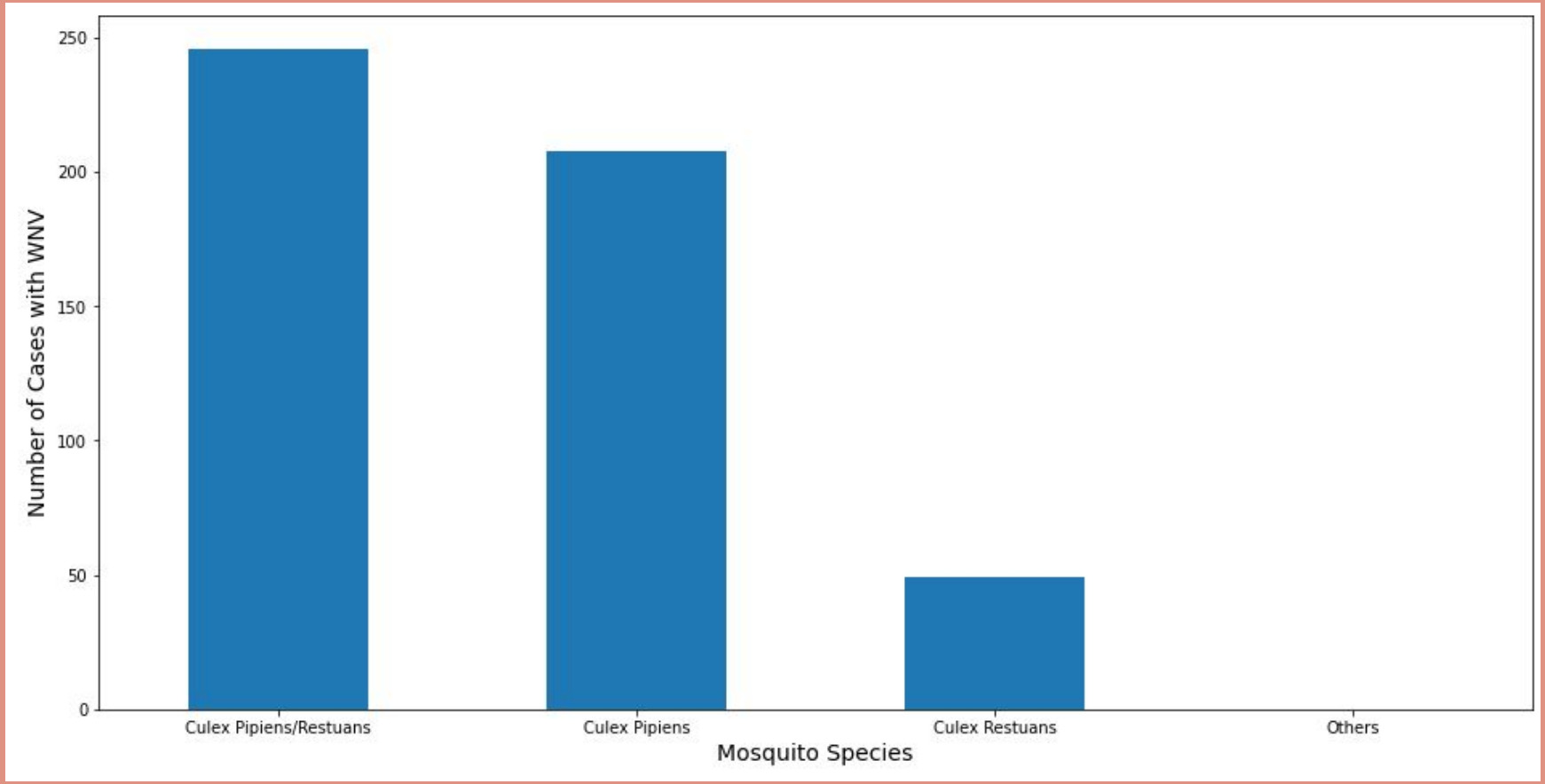
Replaced

- Replaced missing/ empty rows with mean values or NA

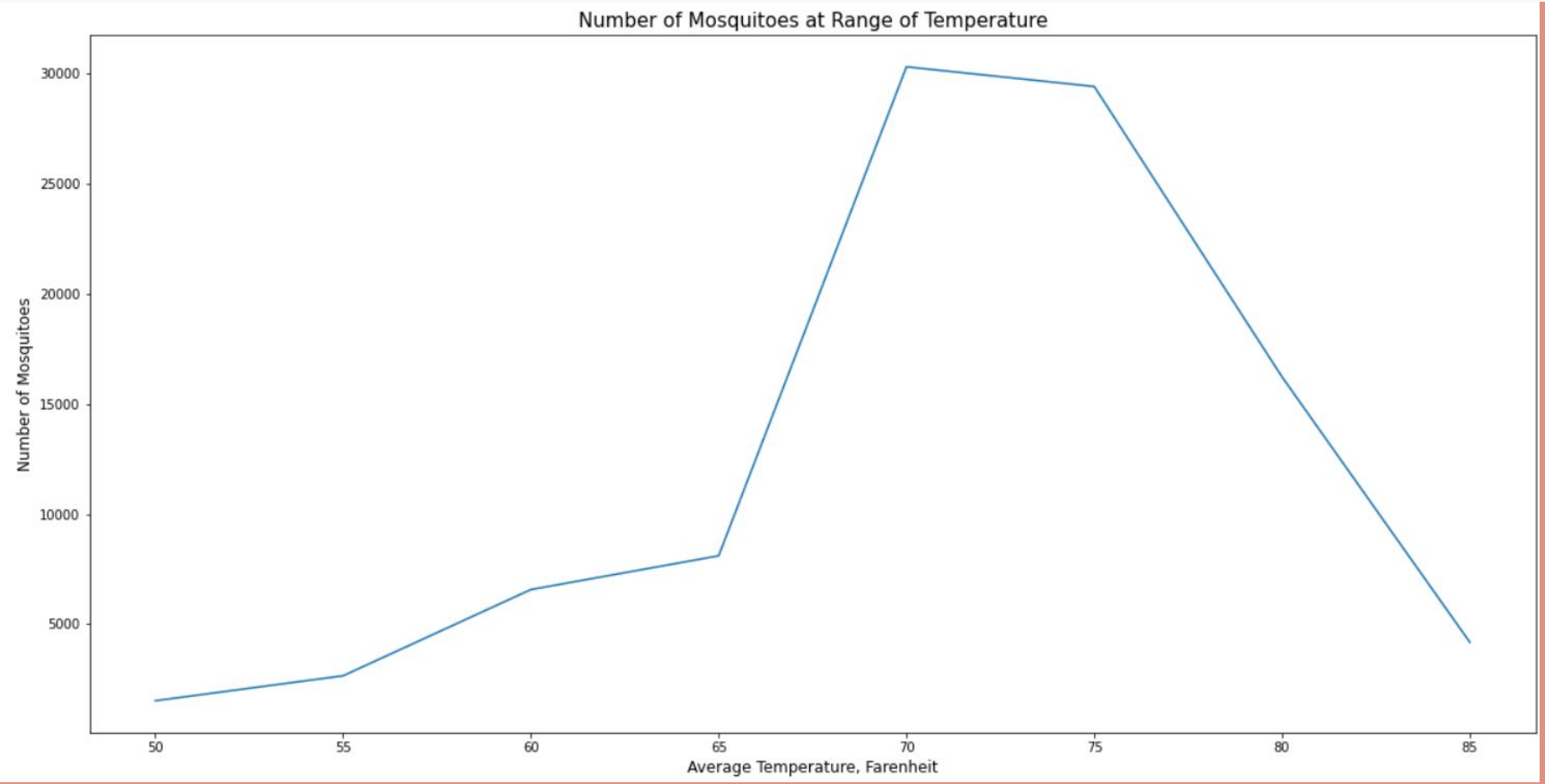
EXPLORATORY DATA ANALYSIS

04

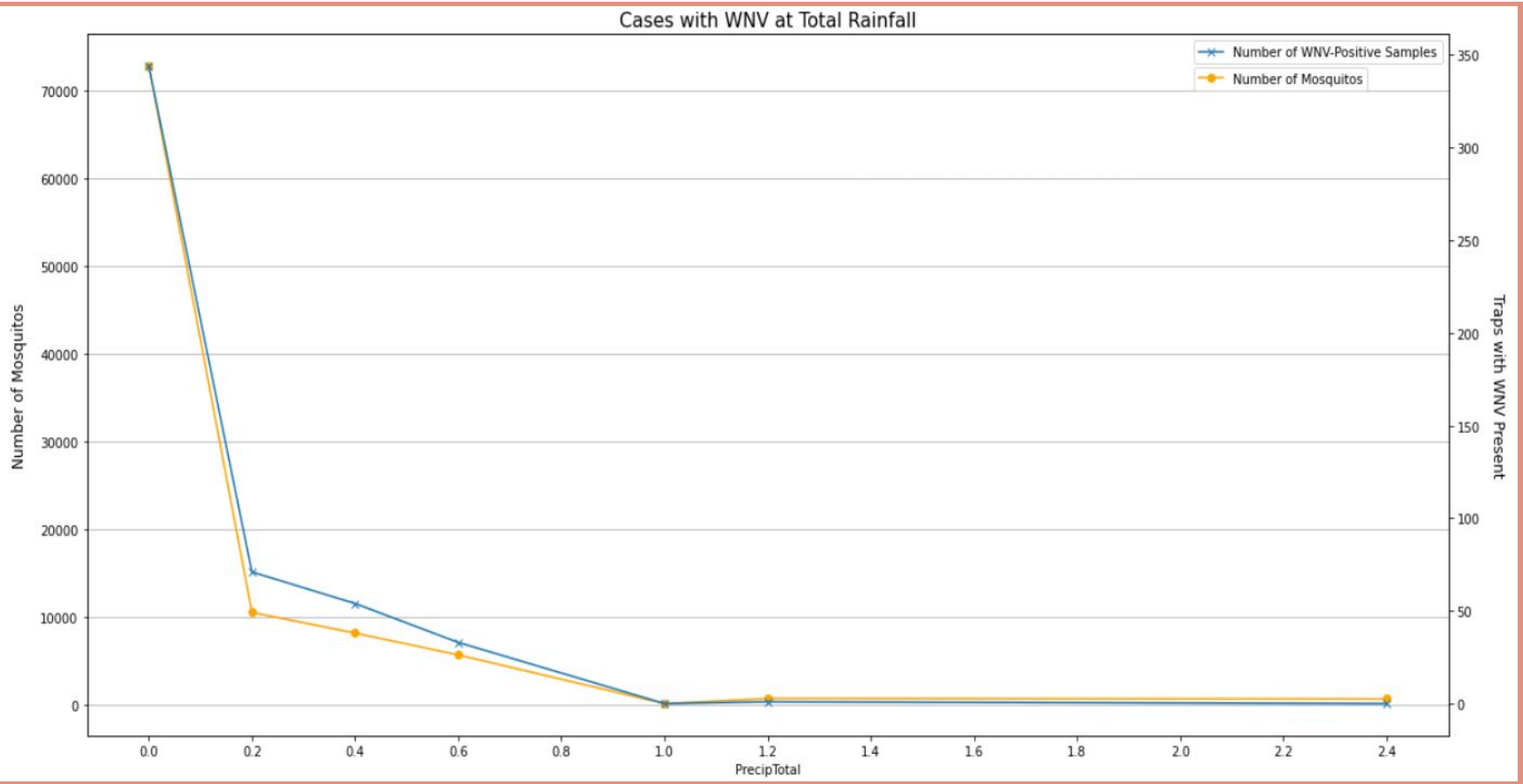
EDA


250
200
150
100
50
0
Number of Cases with WNV
Culex Pipiens/Restuans
Culex Pipiens
Culex Restuans
Others
Mosquito Species

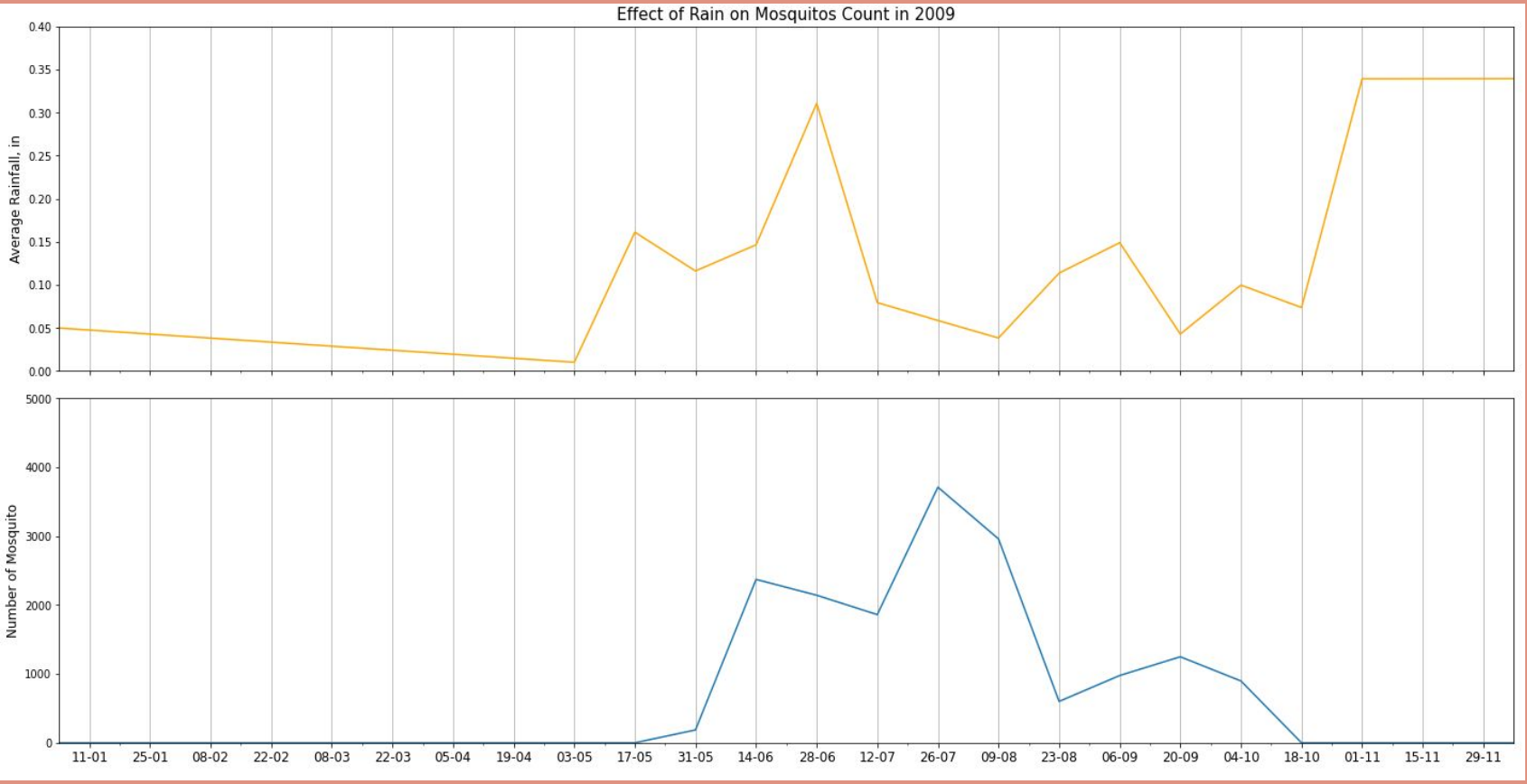
EDA


Number of Mosquitoes at Range of Temperature
Number of Mosquitoes
30000
25000
20000
15000
10000
5000
50
55
60
65
70
75
80
85
Average Temperature, Farenheit

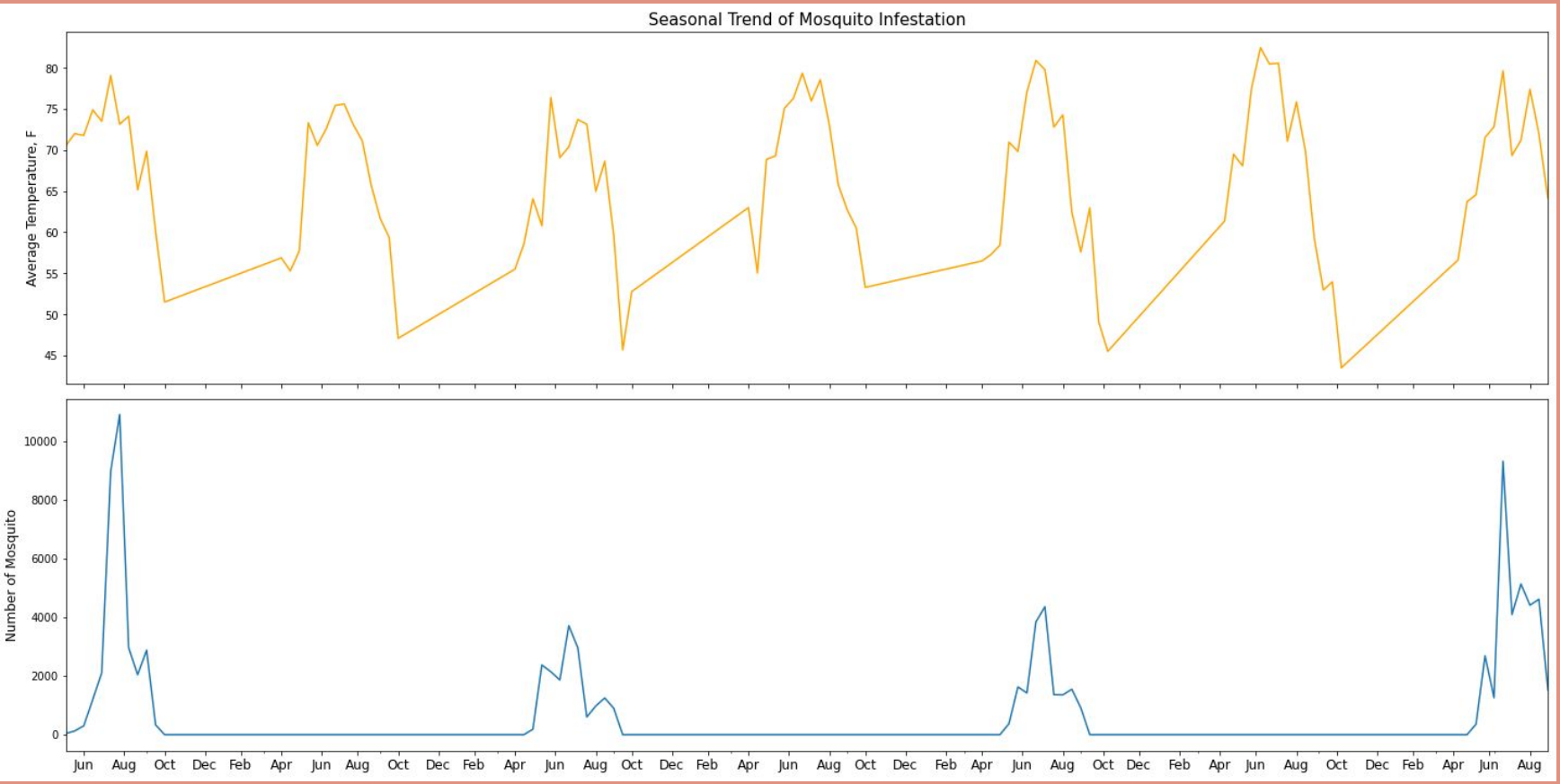
EDA


Cases with WNV at Total Rainfall
Number of WNV-Positive Samples
Number of Mosquitos
Number of Mosquitos
Traps with WNV Present
PrecipTotal
0
10000
20000
30000
40000
50000
60000
70000
0
50
100
150
200
250
300
350
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
2.4

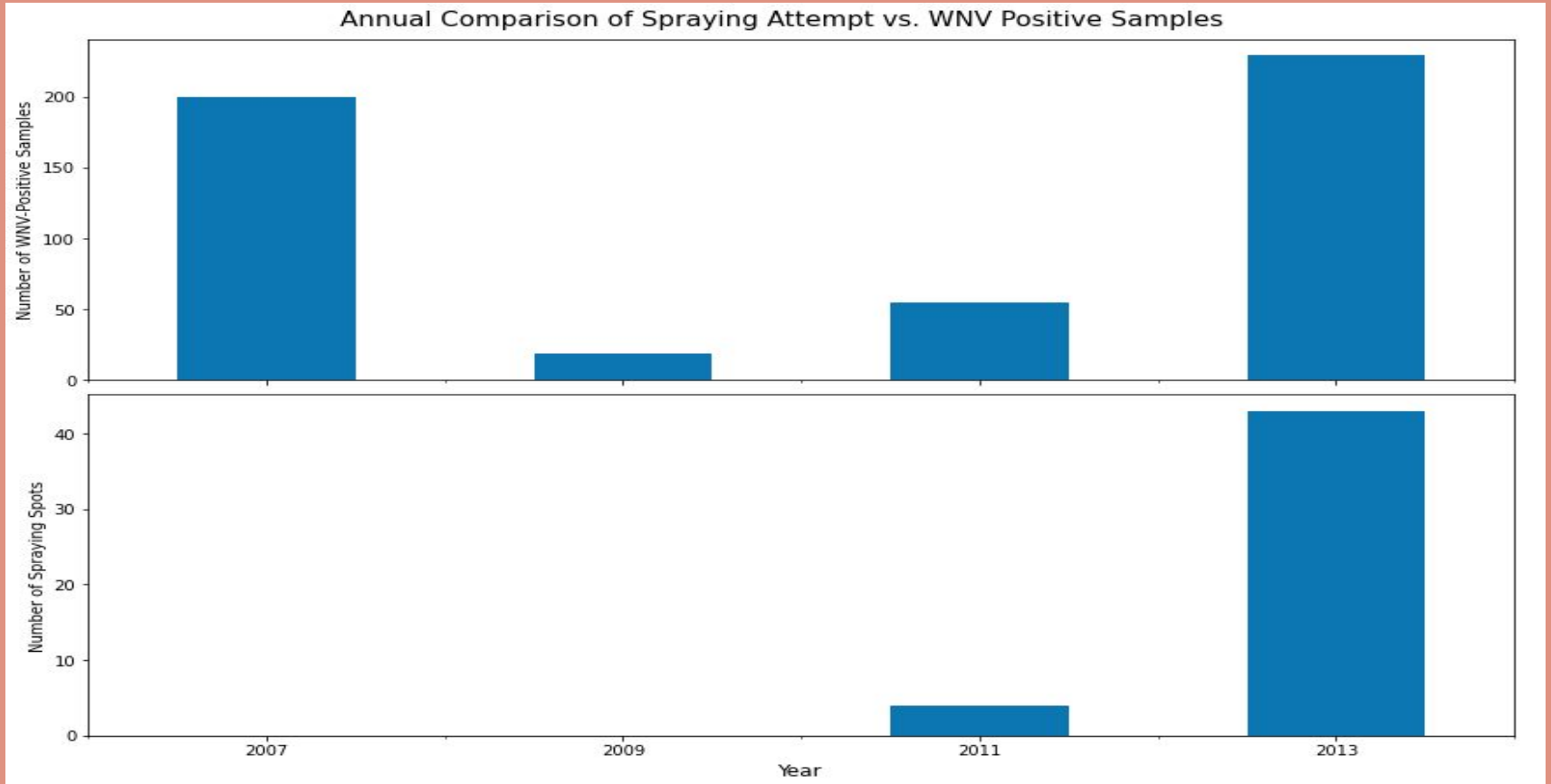
EDA


Effect of Rain on Mosquitos Count in 2009
Average Rainfall, in
0.40
0.35
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Number of Mosquito
5000
4000
3000
2000
1000
0
11-01
25-01
08-02
22-02
08-03
22-03
05-04
19-04
03-05
17-05
31-05
14-06
28-06
12-07
26-07
09-08
23-08
06-09
20-09
04-10
18-10
01-11
15-11
29-11

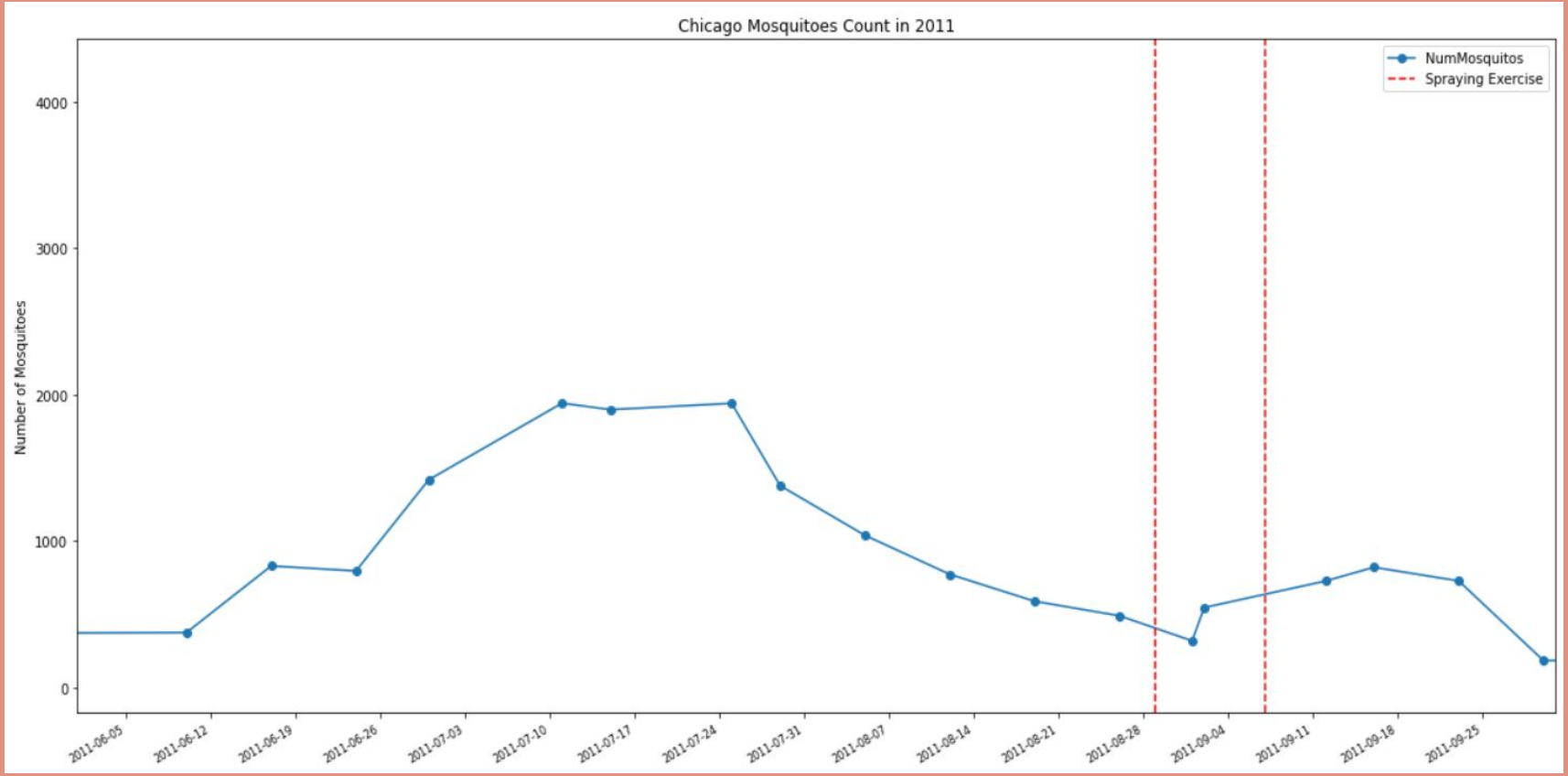
EDA


Seasonal Trend of Mosquito Infestation
Average Temperature, F
80
75
70
65
60
55
50
45
Number of Mosquito
10000
8000
6000
4000
2000
0
Jun Aug Oct Dec Feb Apr Jun Aug Oct Dec Feb Apr Jun Aug Oct Dec Feb Apr Jun Aug Oct Dec Feb Apr Jun Aug Oct Dec Feb Apr Jun Aug Oct Dec Feb Apr Jun Aug

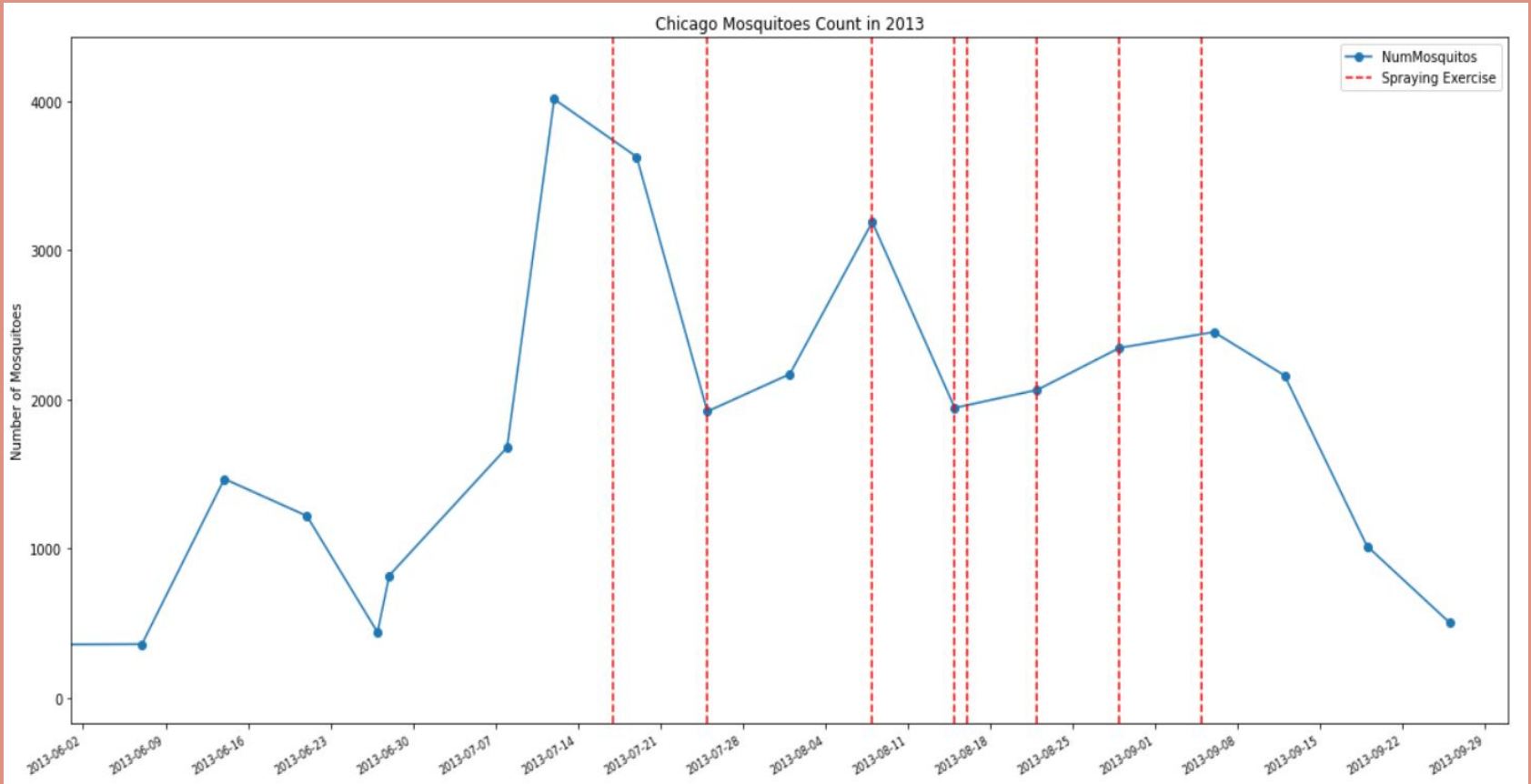
EDA


Annual Comparison of Spraying Attempt vs. WNV Positive Samples
Number of WNV-Positive Samples
0
50
100
150
200
Number of Spraying Spots
0
10
20
30
40
2007
2009
2011
2013
Year

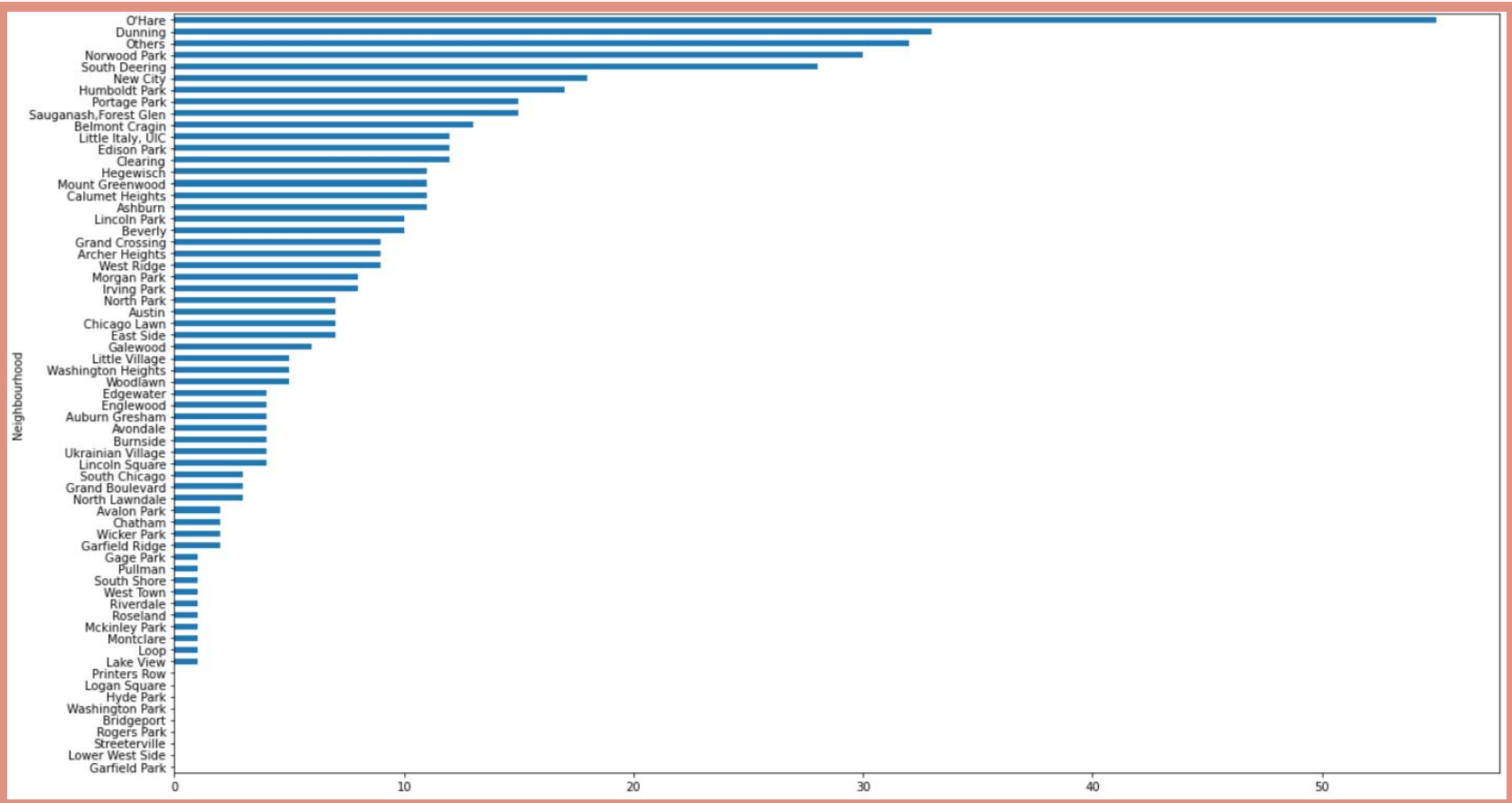
EDA


Chicago Mosquitoes Count in 2011
NumMosquitos
Spraying Exercise
Number of Mosquitoes
4000
3000
2000
1000
0
2011-06-05
2011-06-12
2011-06-19
2011-06-26
2011-07-03
2011-07-10
2011-07-17
2011-07-24
2011-07-31
2011-08-07
2011-08-14
2011-08-21
2011-08-28
2011-09-04
2011-09-11
2011-09-18
2011-09-25

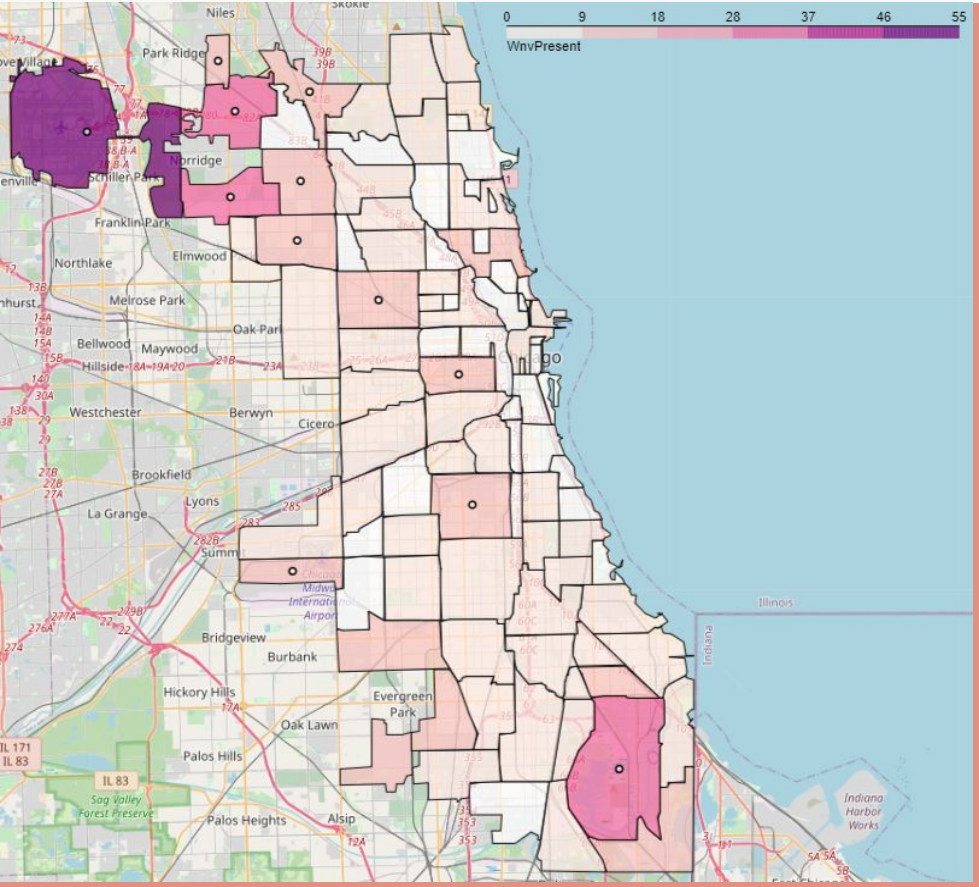
EDA

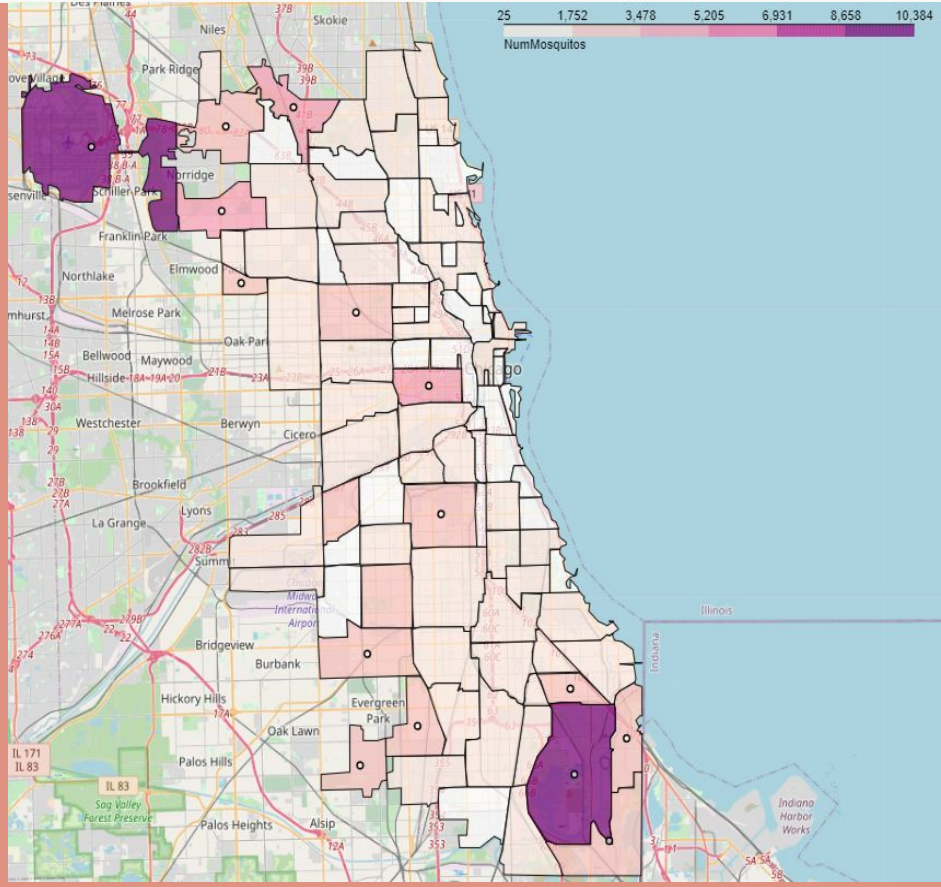

Chicago Mosquitoes Count in 2013
NumMosquitos
Spraying Exercise
Number of Mosquitoes
4000
3000
2000
1000
0
2013-06-02
2013-06-09
2013-06-16
2013-06-23
2013-06-30
2013-07-07
2013-07-14
2013-07-21
2013-07-28
2013-08-04
2013-08-11
2013-08-18
2013-08-25
2013-09-01
2013-09-08
2013-09-15
2013-09-22
2013-09-29

EDA


Neighbourhood
O'Hare
Dunning
Others
Norwood Park
South Deering
New City
Humboldt Park
Portage Park
Sauganash,Forest Glen
Belmont Cragin
Little Italy, UIC
Edison Park
Clearing
Hegewisch
Mount Greenwood
Calumet Heights
Ashburn
Lincoln Park
Beverly
Grand Crossing
Archer Heights
West Ridge
Morgan Park
Irving Park
North Park
Austin
Chicago Lawn
East Side
Galewood
Little Village
Washington Heights
Woodlawn
Edgewater
Englewood
Auburn Gresham
Avondale
Burnside
Ukrainian Village
Lincoln Square
South Chicago
Grand Boulevard
North Lawndale
Avalon Park
Chatham
Wicker Park
Garfield Ridge
Gage Park
Pullman
South Shore
West Town
Riverdale
Roseland
Mckinley Park
Montclare
Loop
Lake View
Printers Row
Logan Square
Hyde Park
Washington Park
Bridgeport
Rogers Park
Streeterville
Lower West Side
Garfield Park
0
10
20
30
40
50

EDA


0 9 18 28 37 46 55
WnvPresent
Niles
Park Ridge
Norridge
Schiller Park
Franklin Park
Northlake
Elmwood
Melrose Park
Oak Park
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Palos Heights
Alsip
Illinois
Indiana
IL 83
Indiana Harbor Works


25 1,752 3,478 5,205 6,931 8,658 10,384
NumMosquitos
Niles
Skokie
Park Ridge
Franklin Park
Northlake
Elmwood
Melrose Park
Oak Park
Bellwood
Maywood
Hillside
Westchester
Berwyn
Cicero
Brookfield
La Grange
Lyons
Bridgeview
Burbank
Hickory Hills
Evergreen Park
Oak Lawn
Palos Hills
Palos Heights
Alsip
Illinois
Indiana
IL 83
Indiana Harbor Works

MODELS

05

Models

Models	Train Accuracy Score	Test Accuracy Score	Train Recall Score	Test Recall Score	Kaggle Score
Logistic Regression (SMOTE)	0.702	0.702	0.786	0.772	0.731
Logistic Regression with PCA (SMOTE)	0.660	0.649	0.704	0.772	0.695
Extra Trees (Balanced)	0.793	0.783	0.973	0.743	0.688
Naive Bayes (SMOTE)	0.613	0.609	0.769	0.802	0.606
Xgboost (SMOTE)	0.843	0.841	0.701	0.653	0.702

COST BENEFIT ANALYSIS

06

Cost Benefit Analysis

Cost Benefit Analysis: Spraying			
Costs			
Category	**Quantity**	**Price**	**Total**
Vector Controls[1]	284.867	$1471 /sqkm	$419,039
		TOTAL COSTS	**$418,754**
Benefits			
Medical & Productivity[2]	114[3]	$13,987 /wnv case	$1,594,518
		TOTAL BENEFITS	**$1,594,518**
		BENEFIT-COST RATIO	**3.8**

[1]Aerial spray by pyrethrins for six nights (sources: Bellini, R., Zeller, H. & Van Bortel, W.)

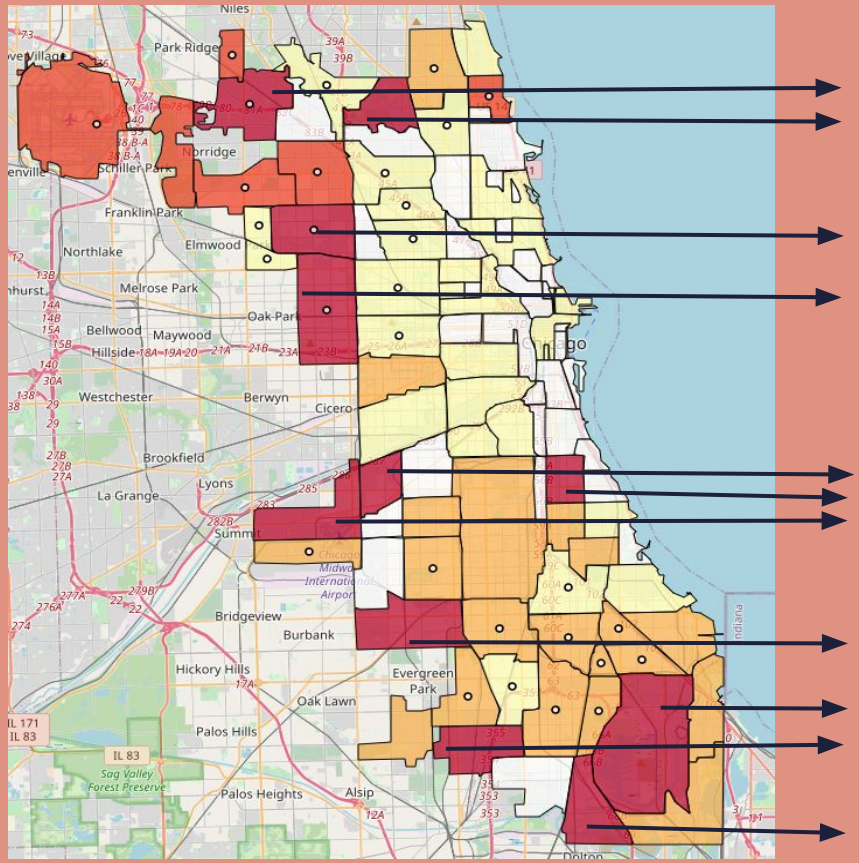
[2]Including medical costs to treat cases and patients' productivity loss (sources: Bellini, R., Zeller, H. & Van Bortel, W.)

[3]Difference in positive cases in Cook county between 2012 and 2013 (sources: Illinois Department of Public Health)

CONCLUSION AND RECOMMENDATION

10

Recommendation



	Neighbourhood	area	WnvPresent	Dependencies
44	Norwood Park	12.195911	0.128205	39.5
39	Morgan Park	9.187734	0.101266	40.3
22	Garfield Ridge	11.789078	0.100000	38.1
1	Ashburn	13.546034	0.090164	36.9
43	North Park	7.028871	0.075269	39.0
6	Belmont Cragin	10.909941	0.070270	37.3
23	Grand Boulevard	4.849250	0.068182	39.5
3	Austin	17.003775	0.066667	37.9
0	Archer Heights	5.592251	0.063380	39.2
50	Riverdale	9.838950	0.058824	51.5
55	South Deering	30.379706	0.058700	39.5

Limitations and other considerations

- Spraying effectiveness limited to areas near roads
- Other reasons for WNV decline:
  - larvicide implementation,
  - varying effectivness of spraying
  - increased personal protective measures
- Considerations:
  - Include larvicide data
  - personal protective measures data

THANKS!

Do you have any questions?

youremail@freepik.com
+91 620 421 838
yourcompany.com

Additional Slides