



West Nile Virus

Here is where your presentation begins

INTRO 01

DATA CLEANING 03

EDA 05

COST BENEFIT
ANALYSIS 07

02 PROBLEM STATEMENT

04 DATA PRE-PROCESSING

06 MODELS

08 CONCLUSIONS
RECOMMENDATIONS

INTRO

01



Introduction

- The West Nile Virus (WNV) is transmitted to people via infected mosquitos.
- 20% of infected people develop symptoms ranging from persistent fever, up to fatal neurological illnesses
- 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

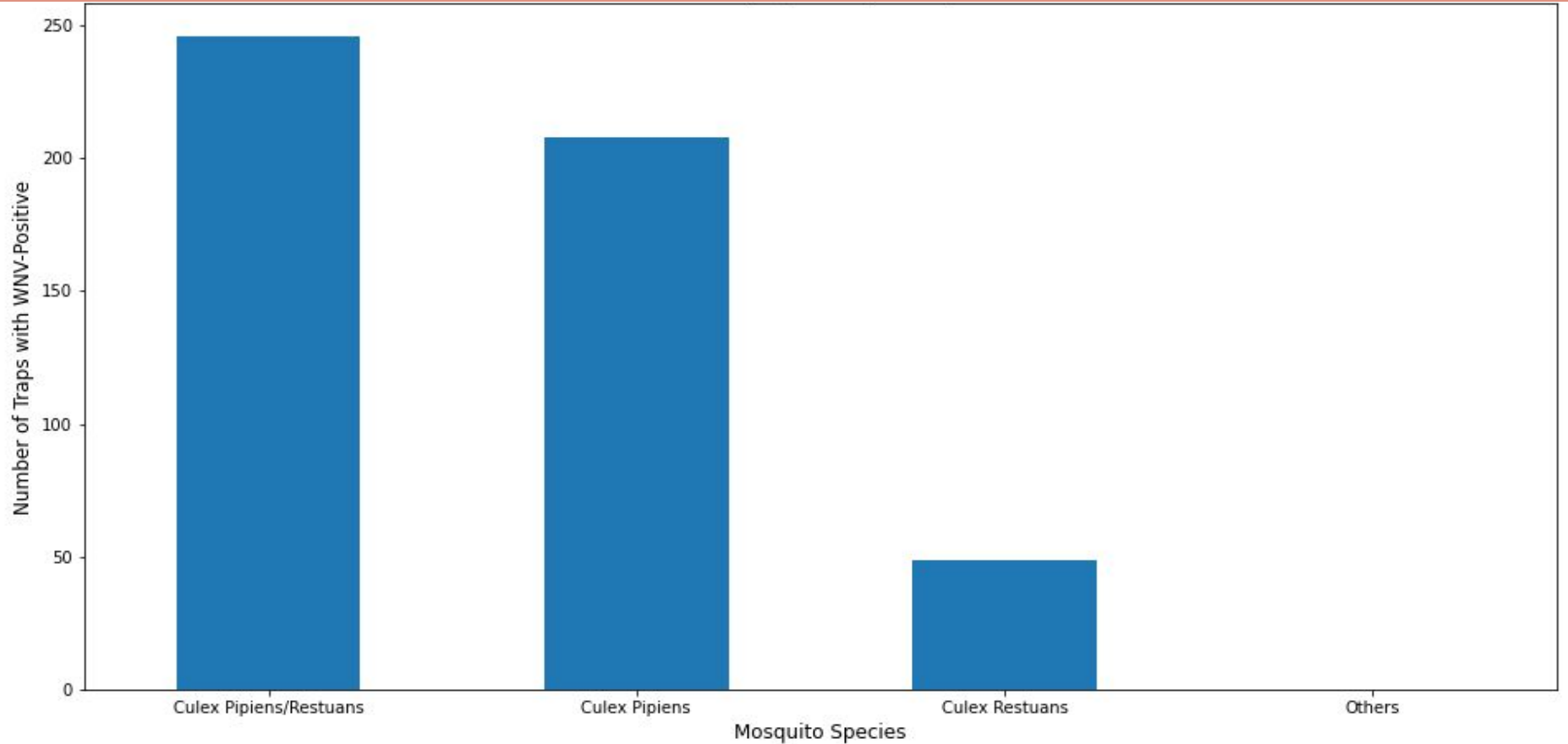
- Replaced missing/ empty rows with mean values or NA

EXPLORATORY DATA ANALYSIS

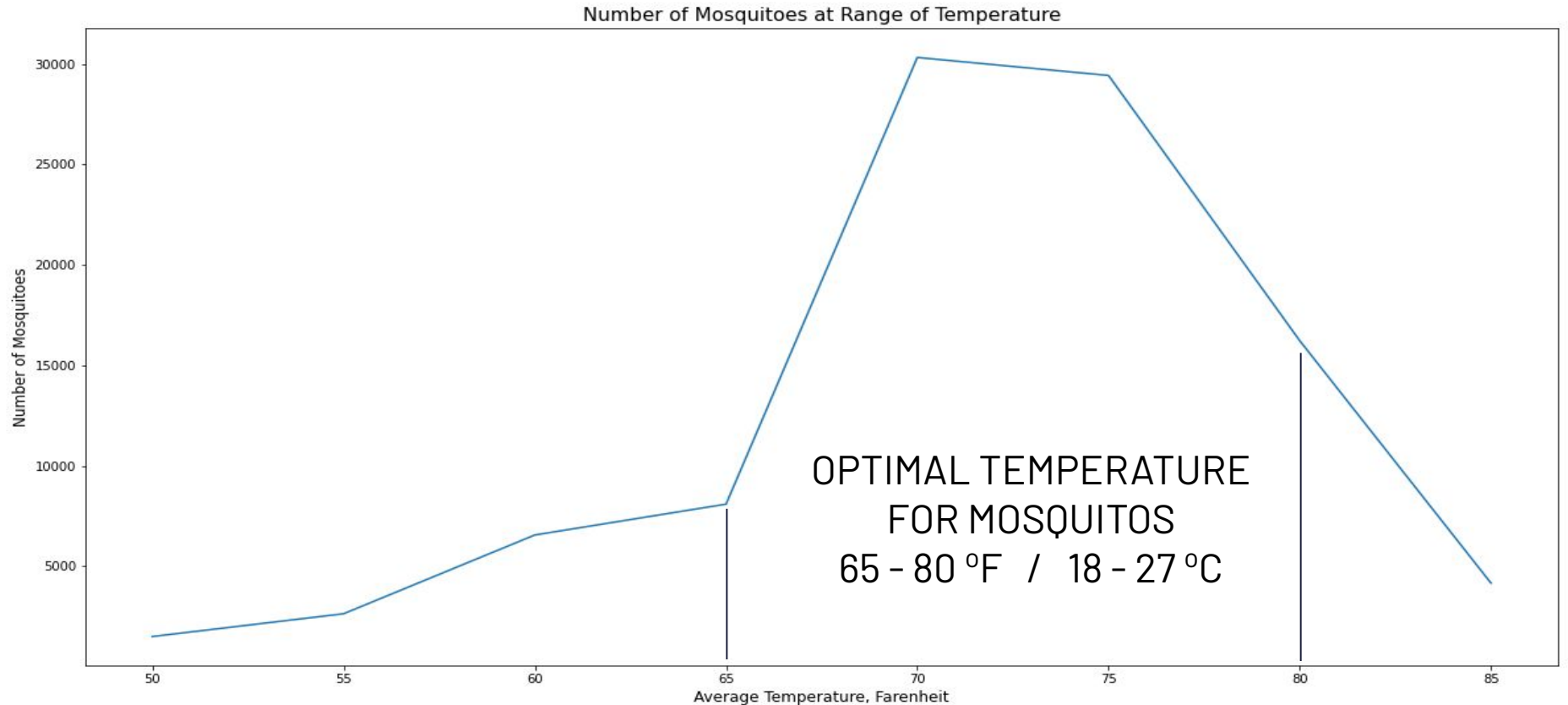


04

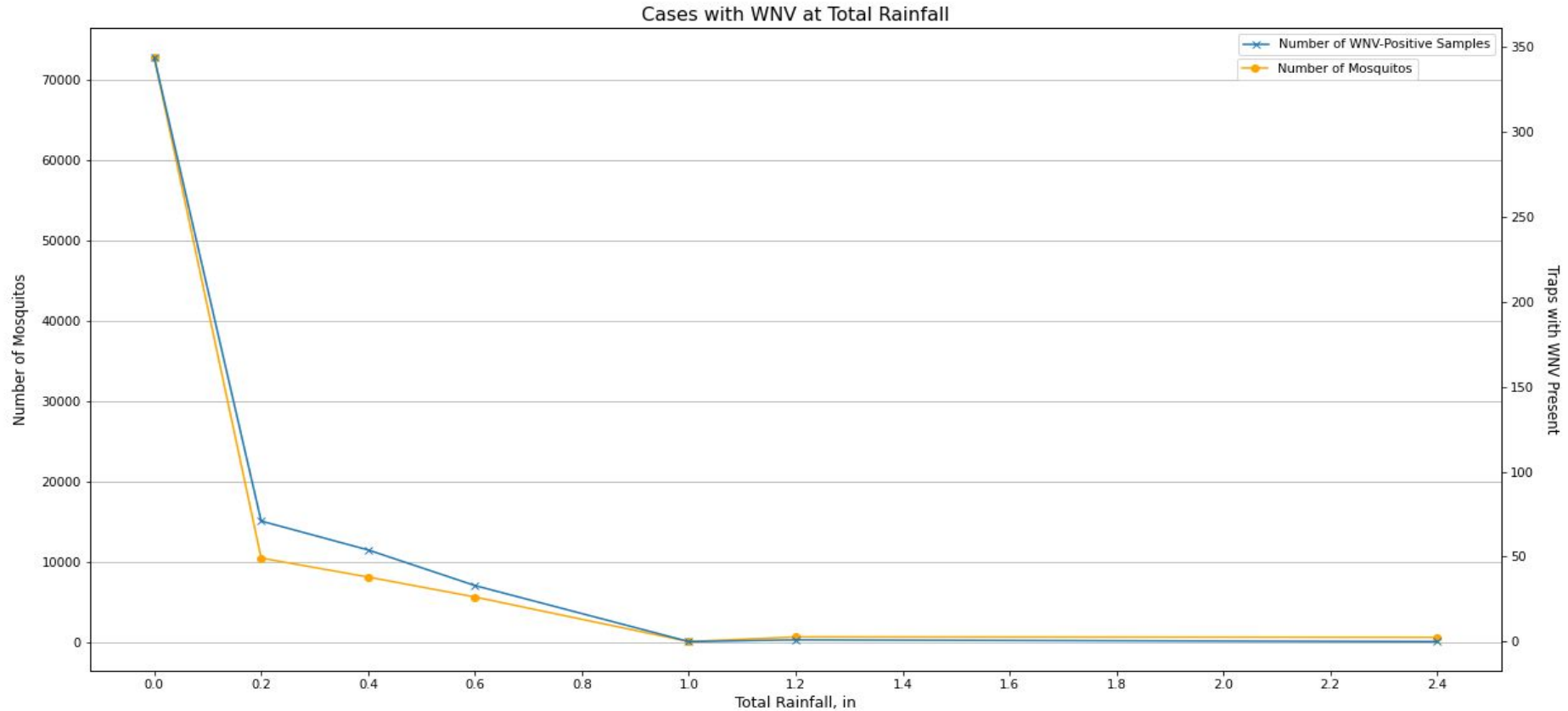
WNV-Carrying Mosquito Species



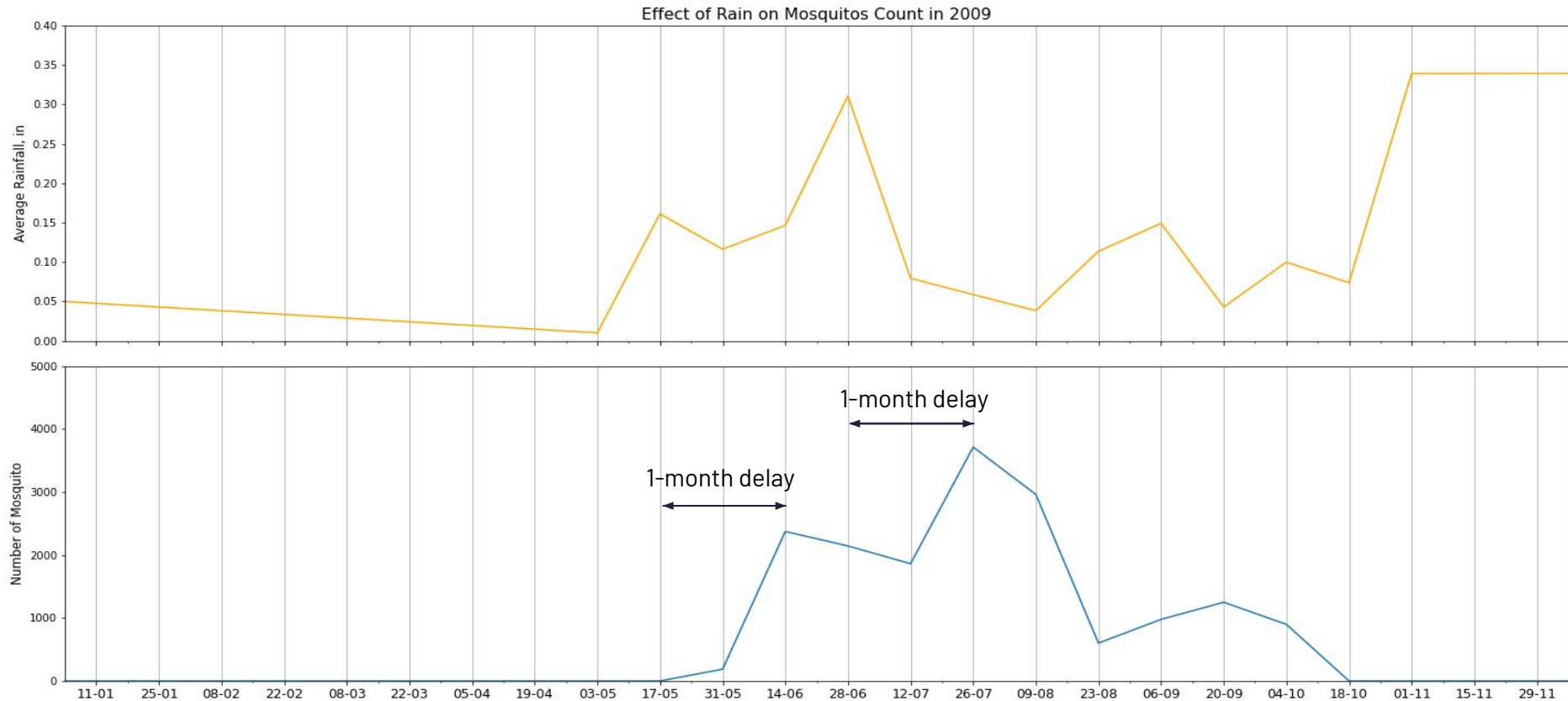
Optimal Temperature Range for Mosquitos



Effect of Precipitation on Mosquitos

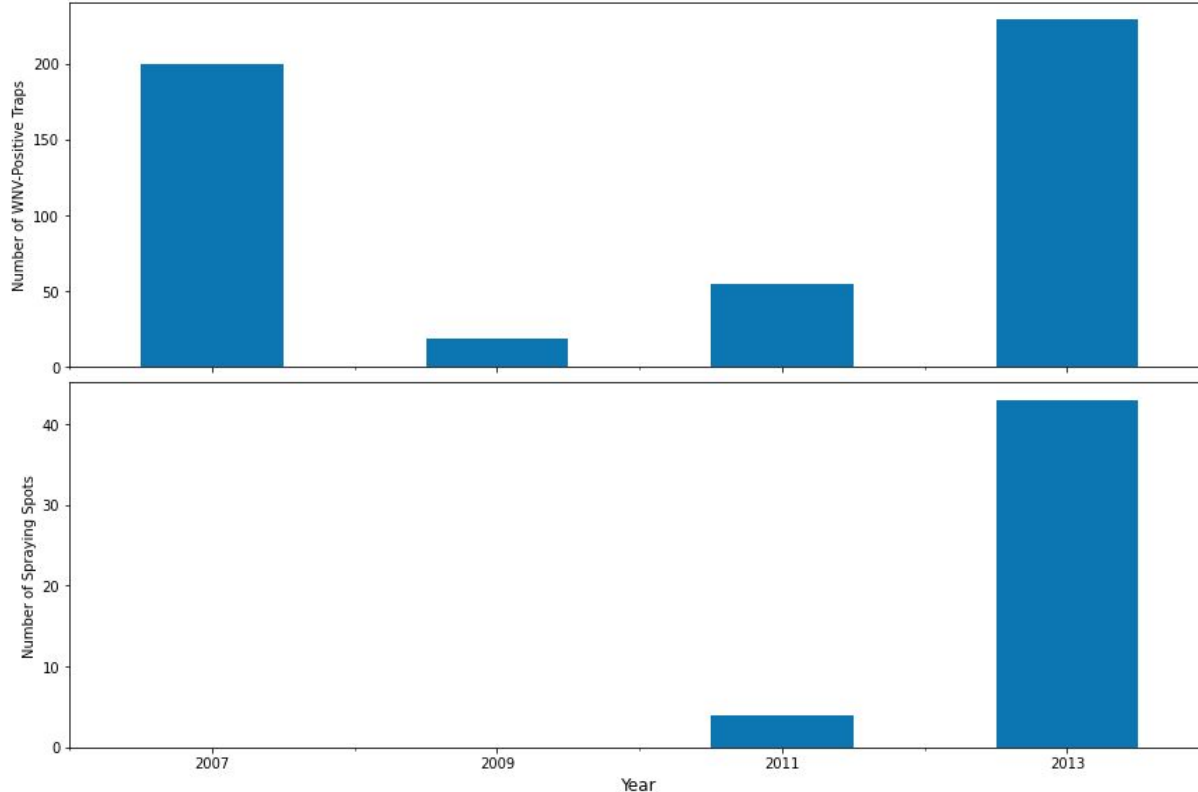


Rain-induced WNV Outbreak



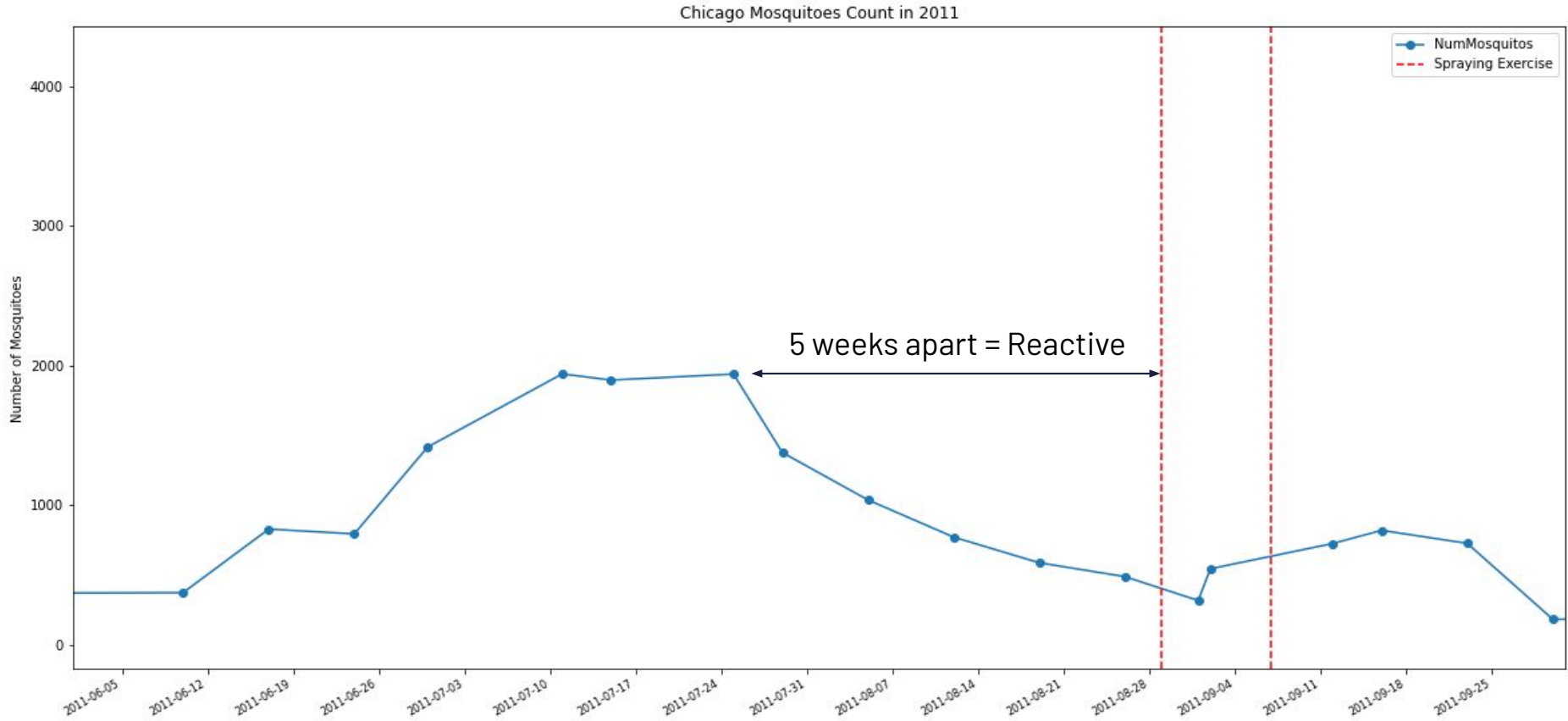
Annual Spraying Activities

Annual Comparison of Spraying Attempt vs. WNV Positive Samples

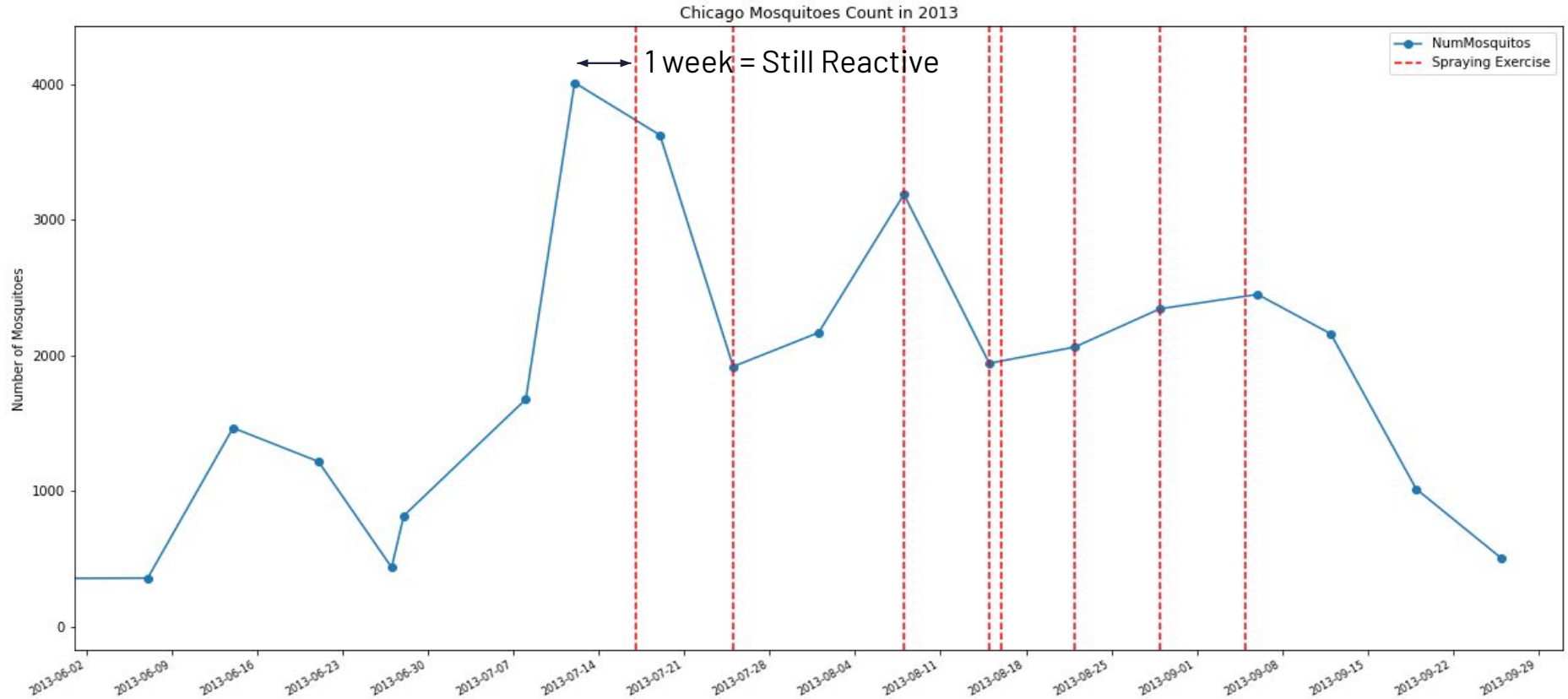


- Data not available for 2007 and 2009
 - Data Missing
 - No Spraying made
- Spraying peaked in 2013
- WNV-positive sample count proportionate to Spraying counts
- Spraying was mostly carried out on reactive basis by authority
- Response only to reports and spikes in cases

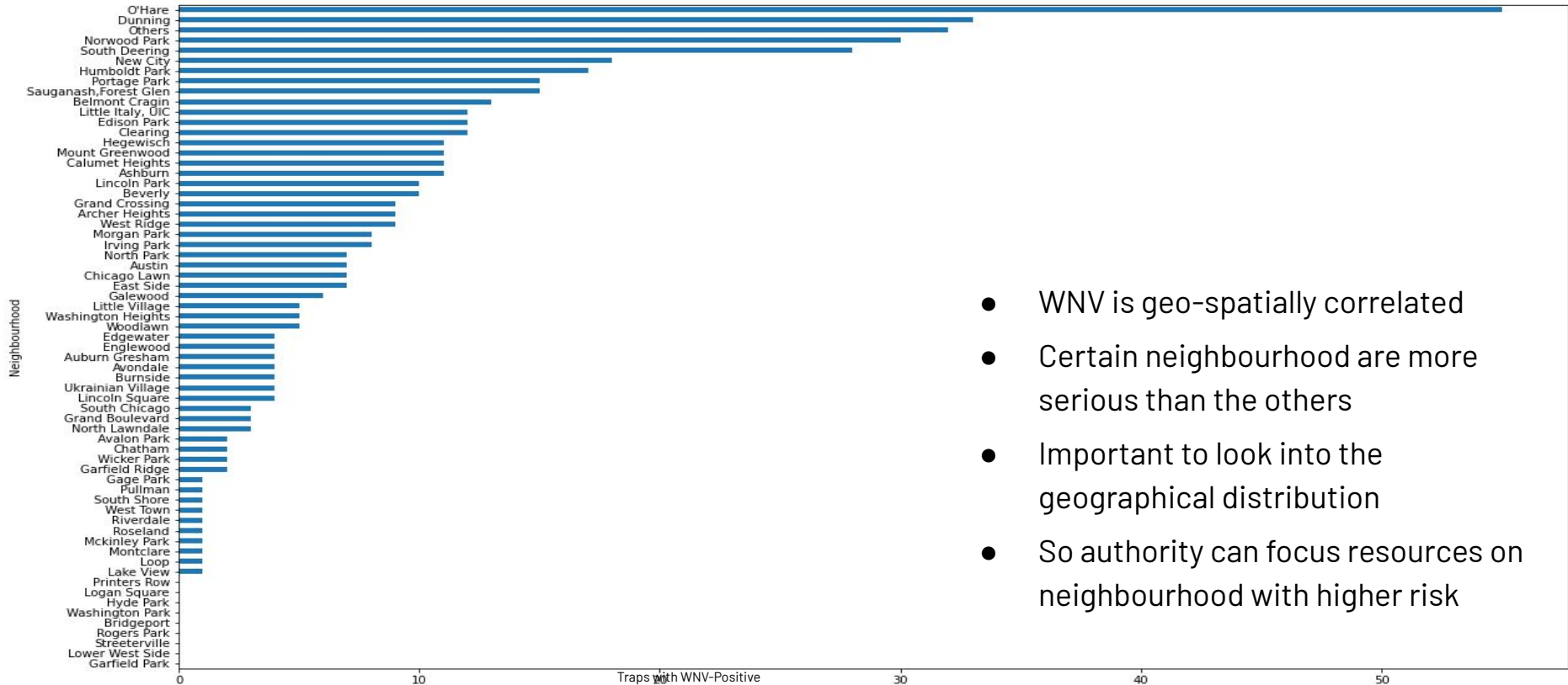
Spraying Timeline in 2011



Spraying Timeline in 2013

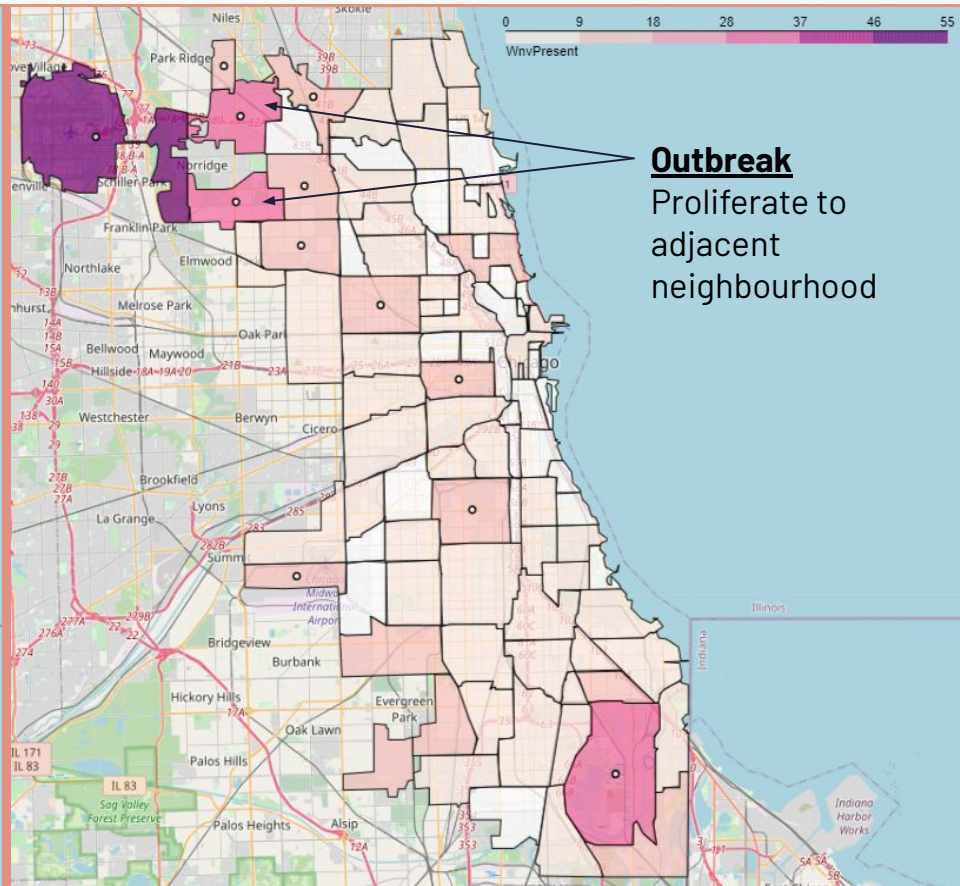
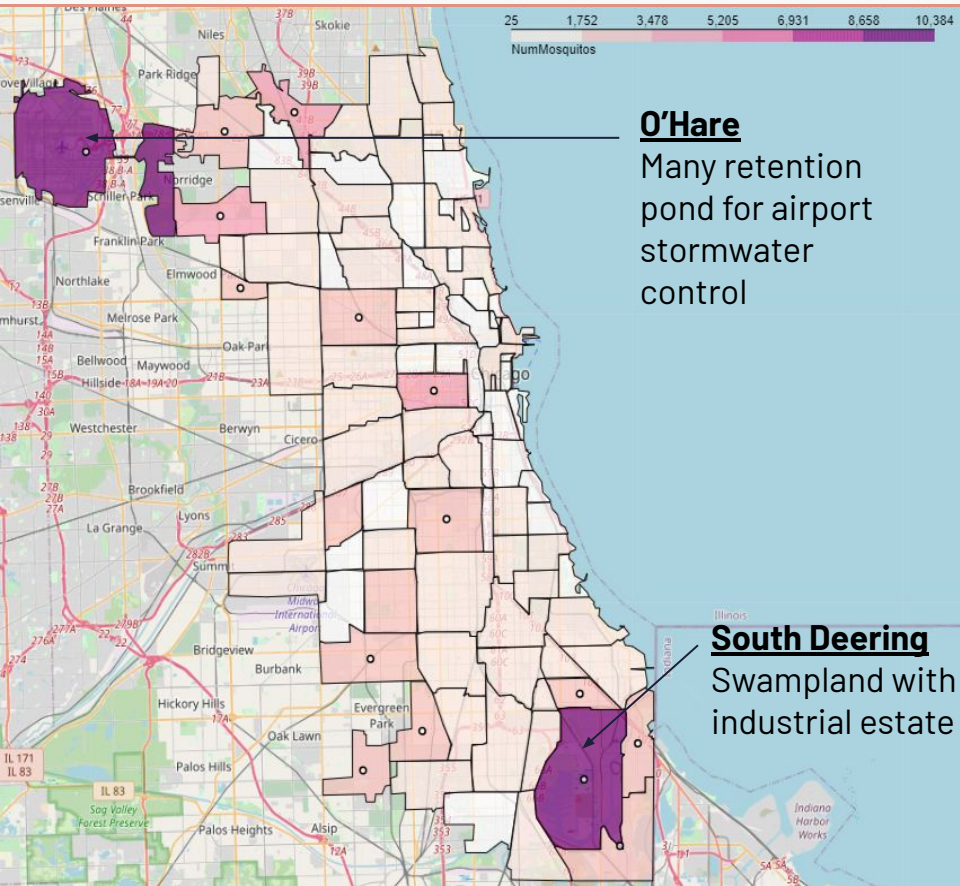


Neighbourhood WNV Outbreak



- WNV is geo-spatially correlated
- Certain neighbourhood are more serious than the others
- Important to look into the geographical distribution
- So authority can focus resources on neighbourhood with higher risk

Geo-spatial Distribution of WNV and Mosquitos



MODELS

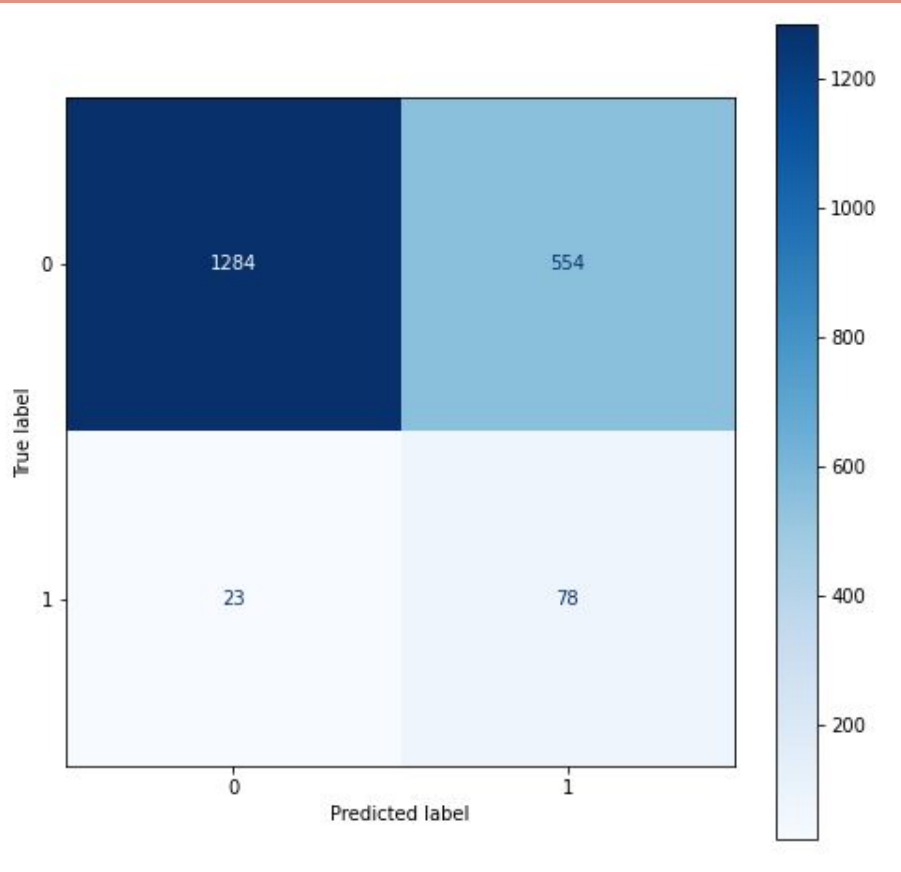
05



Model Accuracy and Recall Scores

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

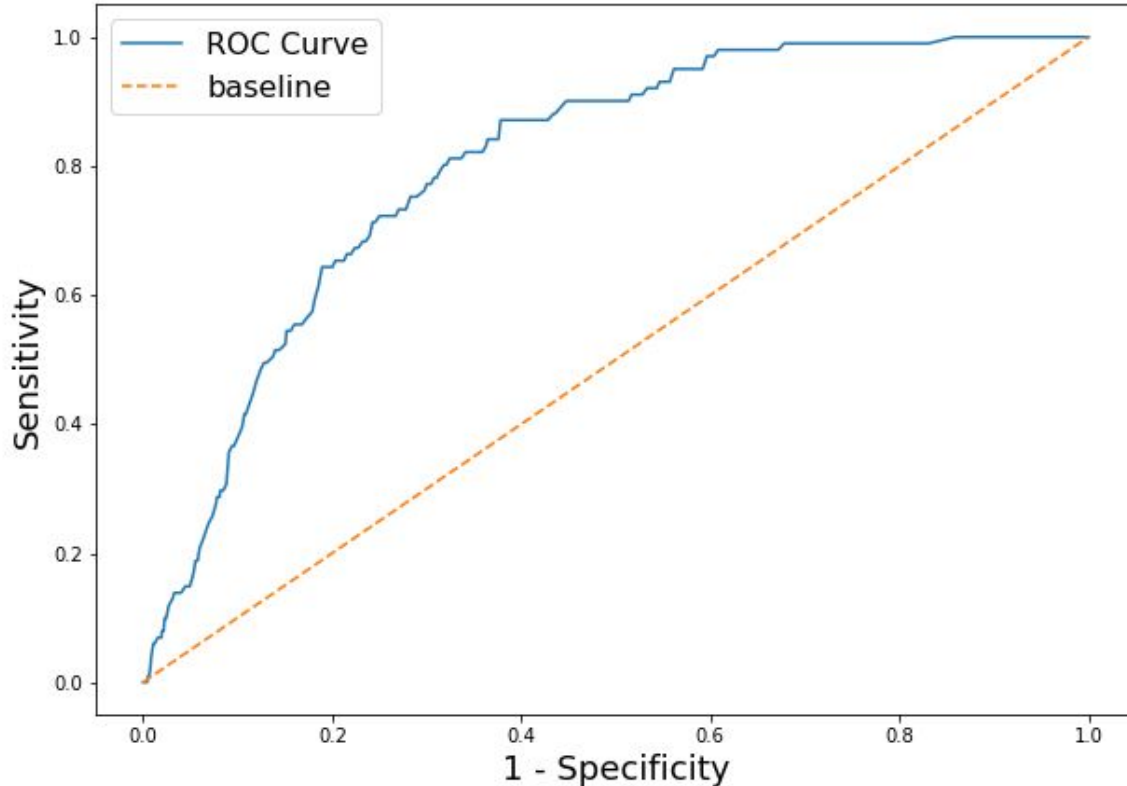
Best Model Performance - Confusion Matrix



- Logistic Regression - the best performing model
- Heavy class-imbalance in the dataset, at 95-to-5 ratio
- SMOTE was used for oversampling so that model is generalized
- Overall Accuracy = 70%
- Recall Score = 77%
- Specificity = 70%

Best Model Performance - ROC-AUC Curve

ROC Curve with AUC = 0.803



- Achieved ROC-AUC score of 80%
- Compared to 50% baseline score
- However, still relatively weak segregation between Class 0 and 1
- No threshold adjustment required at this point
- Balance between Recall and Specificity

COST BENEFIT ANALYSIS



06

Cost Benefit Analysis

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

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

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

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

RECOMMENDATION AND CONCLUSION



10

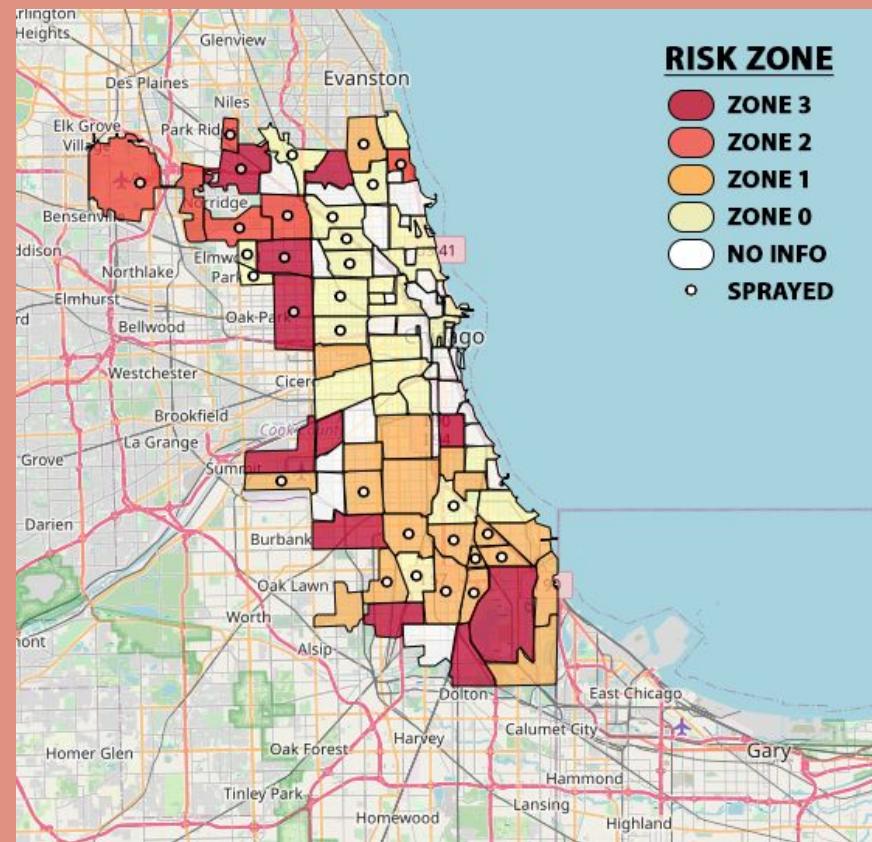
Neighbourhood Outbreak Risk Zoning

- Categorize neighbourhood into zones based on their risk
- Authority can focus their limited resources on specific neighbourhood
- Take into consideration of seniors being more susceptible to WNV

Zone	Risk	WNV-Positive Rate	Dependencies *
3	High	> 75th Percentile	> Median
2	Medium	> 75th Percentile	< Median
1	Low	< 75th Percentile	> Median
0	Very Low	< 75th Percentile	< Median

* Dependent - Population aged less than 18 & more than 64, whom are more vulnerable to WNV

Focus Resource in High Risk Zone



Zone 3 - High Risk Area

Neighbourhood	Area, Square Miles	Wnv Positive Rate, %	Dependent Populations, %
Norwood Park	12.2	12.8	39.5
Morgan Park	9.2	10.1	40.3
Garfield Ridge	11.8	10.0	38.1
Ashburn	13.5	9.0	36.9
North Park	7.0	7.5	39.0
Belmont Cragin	10.9	7.0	37.3
Grand Boulevard	4.8	6.8	39.5
Austin	17.0	6.7	37.9
Archer Heights	5.6	6.3	39.2
Riverdale	9.8	5.9	51.5
South Deering	30.4	5.9	39.5

Original Sprayed Area = 47%
Zone 2 + Zone 3 Area = 36%

Other Considerations

- Spraying effectiveness limited to areas near roads
- Other reasons for WNV decline:
 - Larvicide implementation
 - Varying effectiveness of spraying
 - Increased personal protective measures
- Considerations:
 - Include larvicide data
 - Personal protective measures data

Conclusion

- West Nile Virus outbreak is seasonal, peaking in summertime
- Temperature & Rainfall are main contributing factor, but climate is out of our control
- Curb mosquito breeding by spraying proactively at
 - Beginning of summer before outbreak started
 - Major breeding ground such as O'Hare and South Deering
 - Zone 3 and 2 with the highest risk of WNV outbreak
- Authority can concentrate limited resources to achieve the maximal effect
- By focusing on area with high dependencies, saving healthcare cost and productivity loss in long run

A hand holding a silver and black pen, poised to write on a document. The image is overlaid with a semi-transparent red filter. The word 'THANKS!' is written in large, white, bold, sans-serif capital letters across the center of the image.

THANKS!

Any burning questions?

Shoot!

CREDITS: This presentation template was created by Slidesgo, including icons by FlatIcon, and infographics & images by Freepik.

Please keep this slide for attribution.