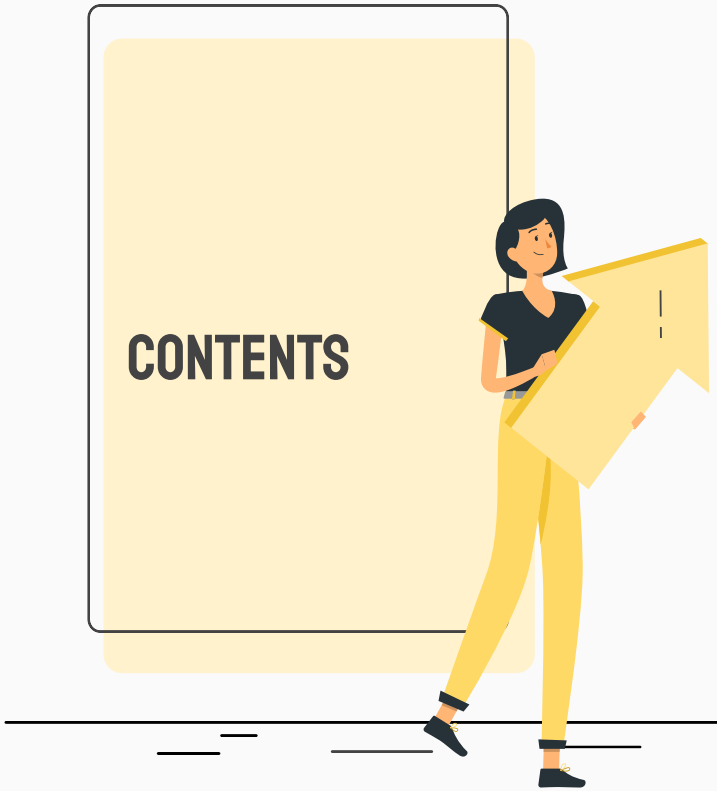
WEST NILE VIRUS (WNV) DETECTION PLAN

Allen
Jean
Mark

CONTENTS



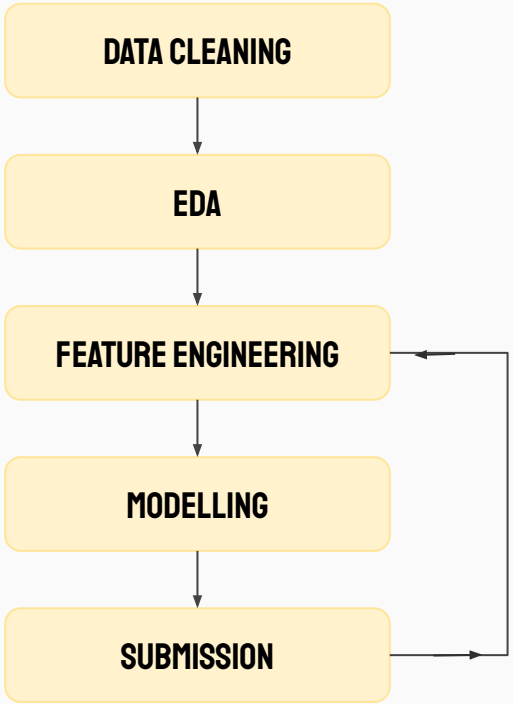
01. PROBLEM STATEMENT

OUR TEAM WILL UTILISE THE DATA TO ANSWER THE FOLLOWING PROBLEM STATEMENTS:

1. PREDICT THE PRESENCE OF THE WNV IN 2008, 2010, 2012 AND 2014
2. IDENTIFY THE KEY FEATURES IN THE MODEL THAT AFFECT WNV
3. PROVIDE RECOMMENDATIONS ON HOW TO USE TRANSMISSION.

EXECUTIVE SUMMARY

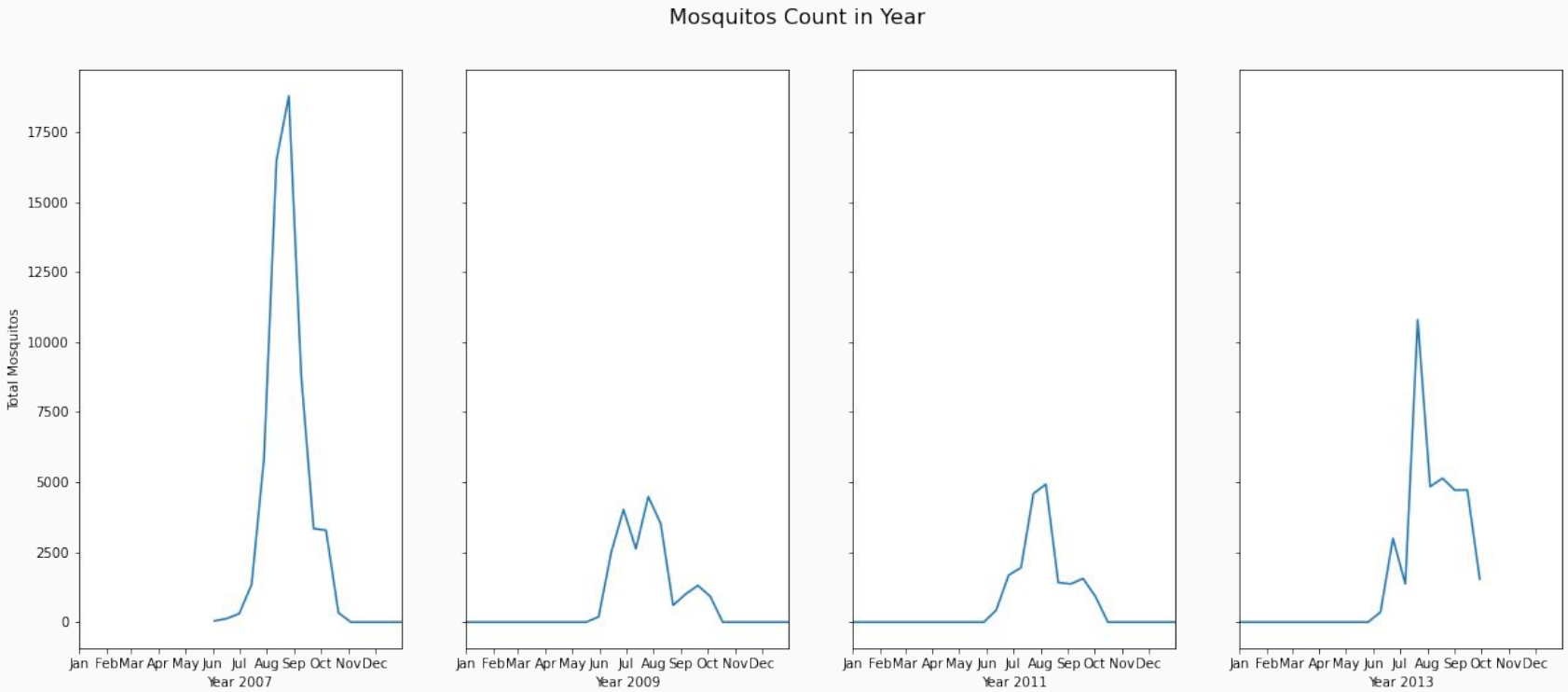
- **WNC**
- **1 in 5**
  - Fever Headaches, body aches, joint pains
- **1 in 150**
  - Severe nervous system breakdown, brain inflammation, meningitis
- Data showed rate of infection for mosquitoes but not rate of transmission to humans
- Location and SPecies are the most important factors for the presence Of wnv



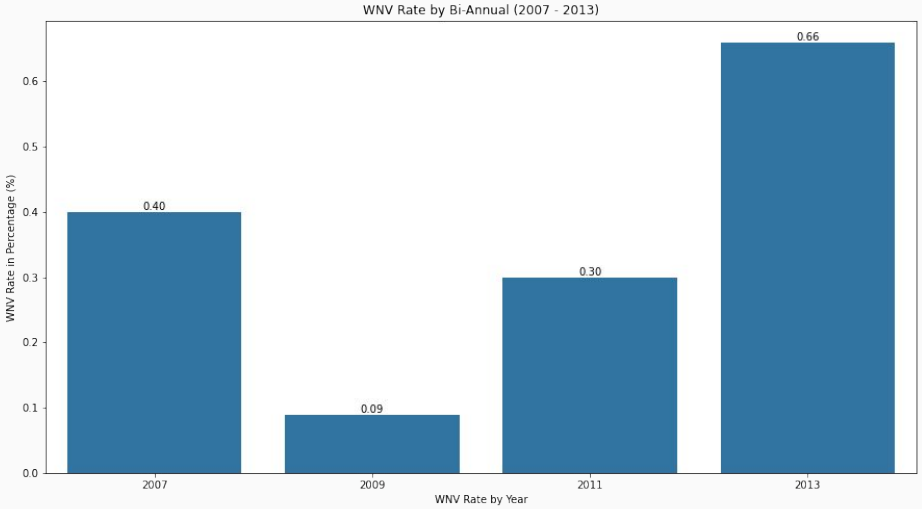
03. EXPLORATORY DATA ANALYSIS

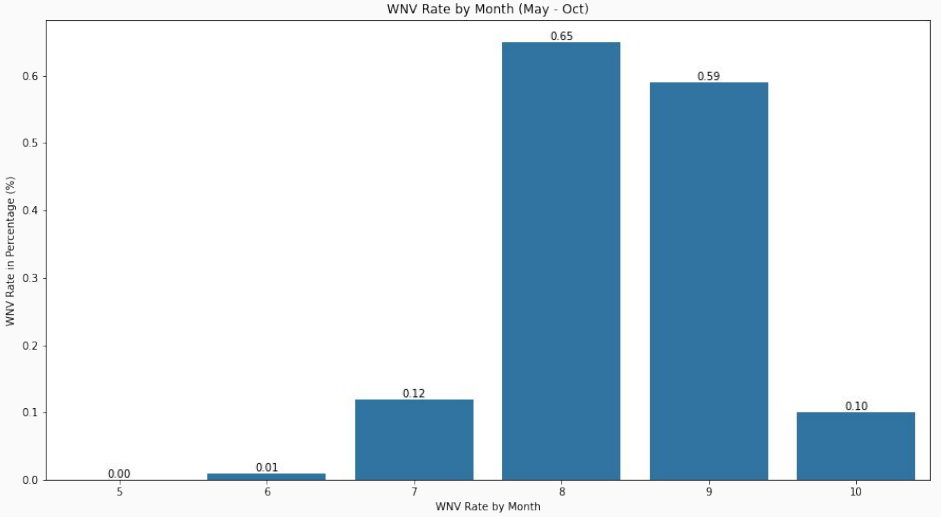
1. EDA ON TRAIN DATASET
2. EDA ON WEATHER DATASET
3. EDA ON SPRAY DATASET

EXPLORATORY DATA ANALYSIS - MOSQUITO COUNT ANNUALLY

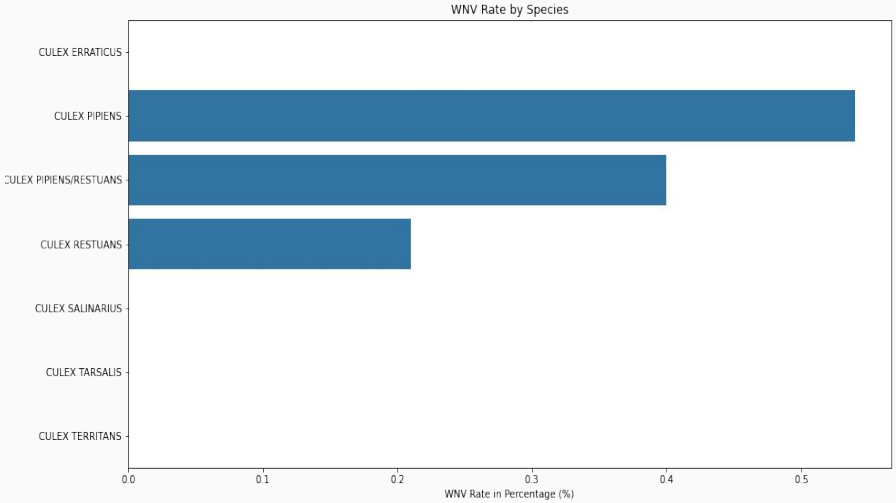


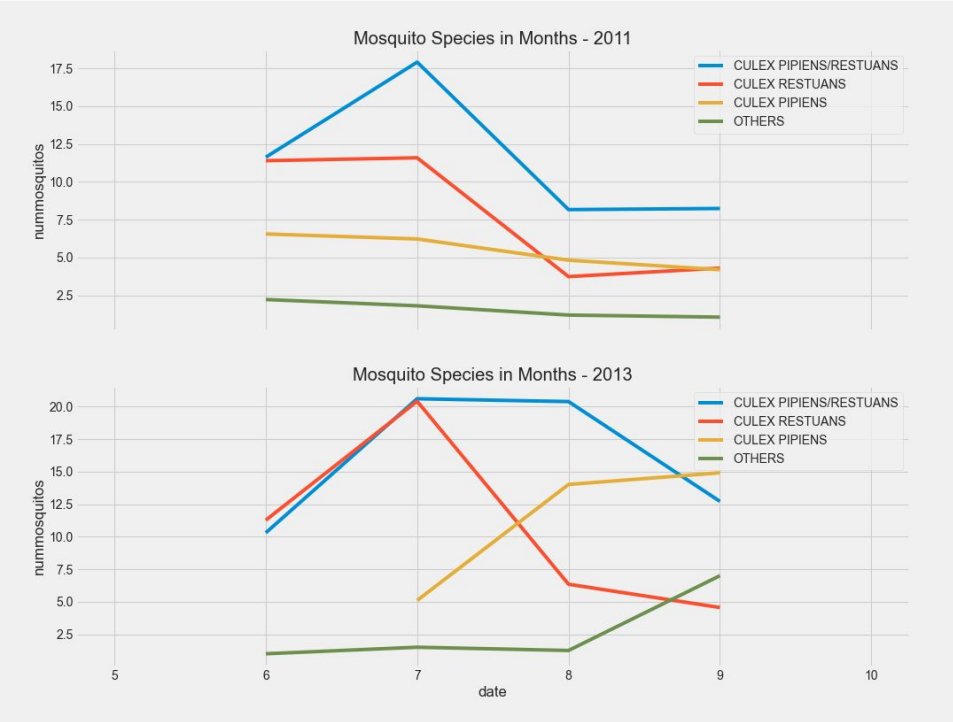
EXPLORATORY DATA ANALYSIS
WNV RATE BY MONTHS AND BI-YEAR



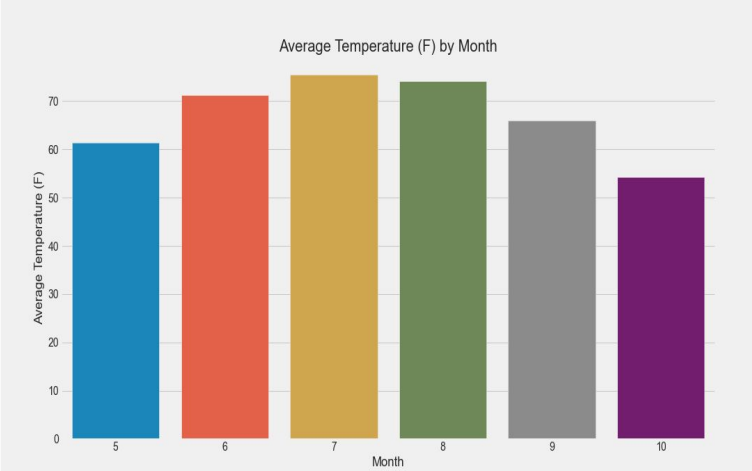


EXPLORATORY DATA ANALYSIS - MOSQUITO SPECIES

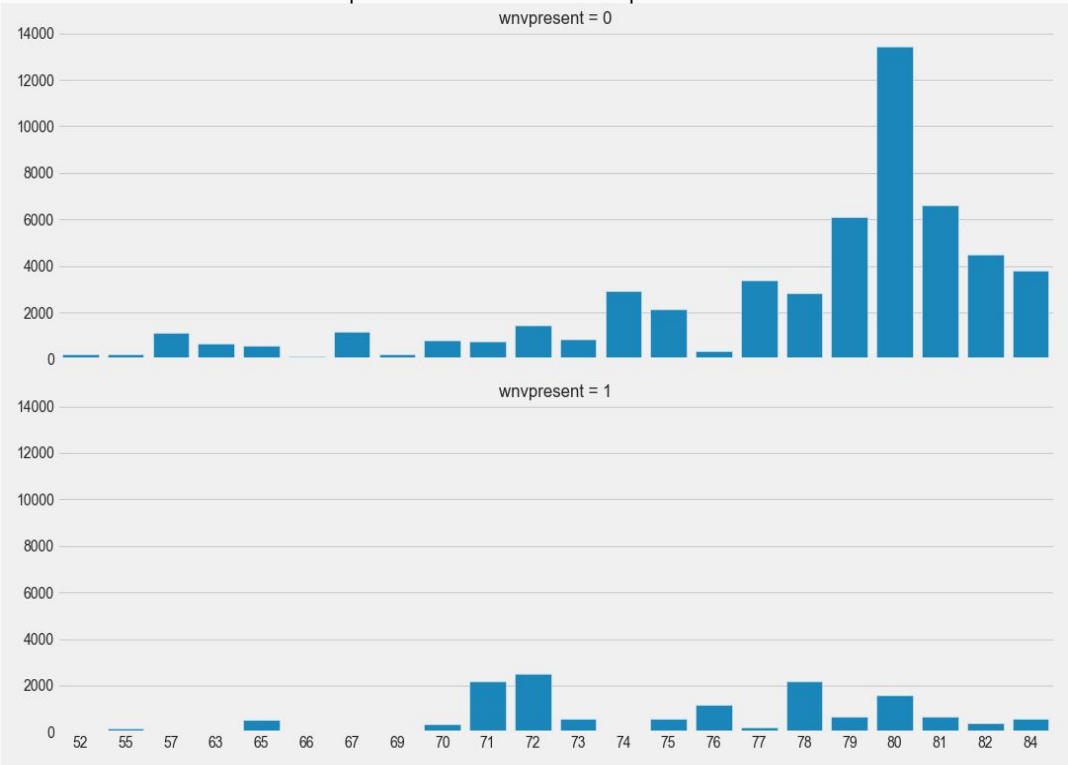




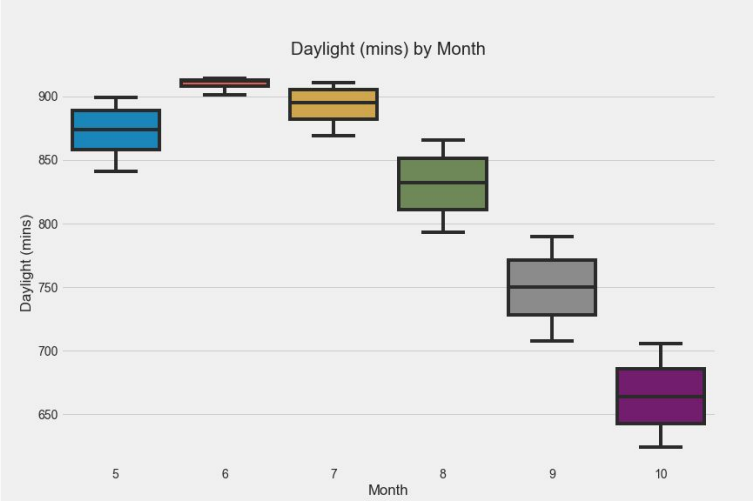
EXPLORATORY DATA ANALYSIS - AVERAGE TEMPERATURE

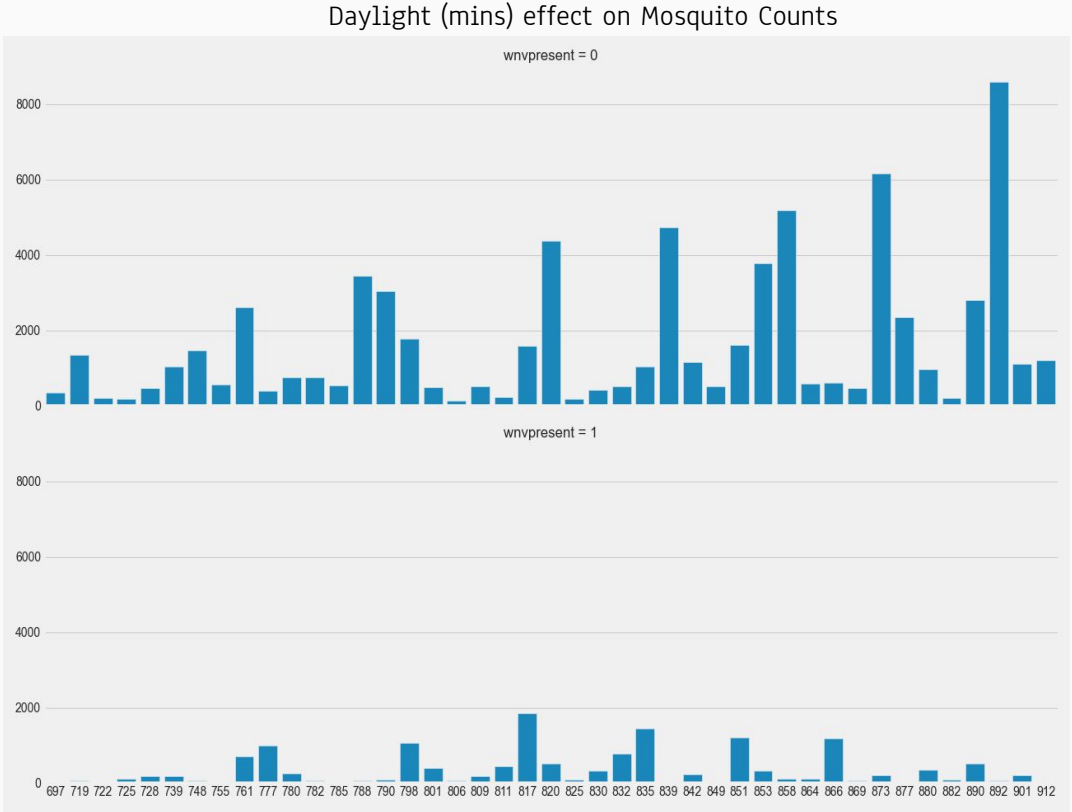


Temperature effect on Mosquito Counts

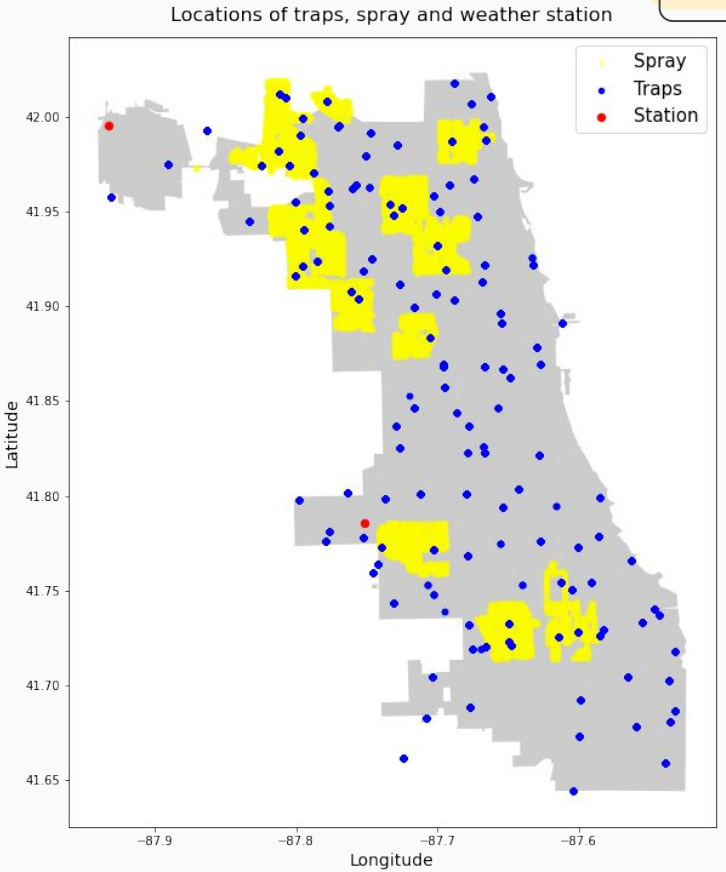


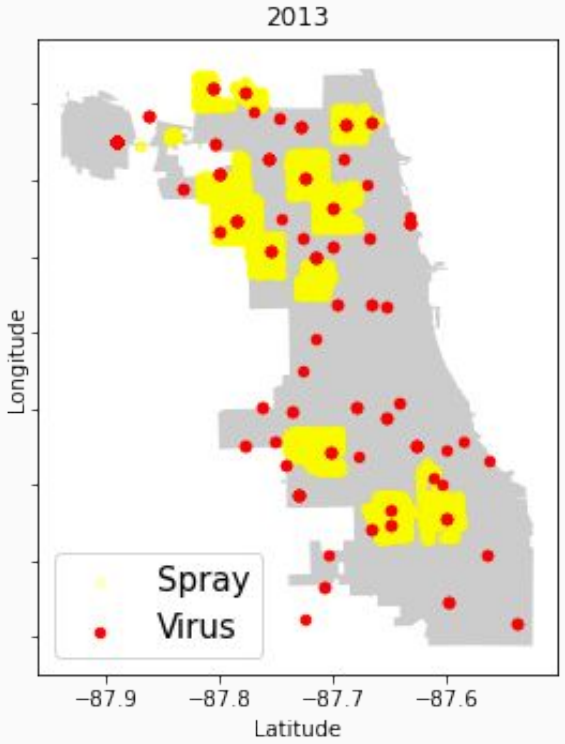
EXPLORATORY DATA ANALYSIS - DAYLIGHT (MINS)





EXPLORATORY DATA ANALYSIS - LOCATION OVERVIEW





04. MODELLING

MODELS USED WITH SMOTE:

- LOGISTIC REGRESSION
- K-NEAREST NEIGHBORS
- RANDOM FOREST CLASSIFIER

BEST MODEL:

- RANDOM FOREST CLASSIFIER

KAGGLE SUBMISSION ITERATION

SUBMISSION	ROC AUC SCORE
1	0.564
2	0.569
3	0.610
4	0.631
5	0.672

01 GENERAL FEATURE ENGINEERING DONE
- MISSING VALUES IMPUTED
- SPECIES ONLY CATEGORICAL FEATURE KEPT
- FEATURES CREATED

02 WEATHER NUMERICAL VARIABLES SQUARED

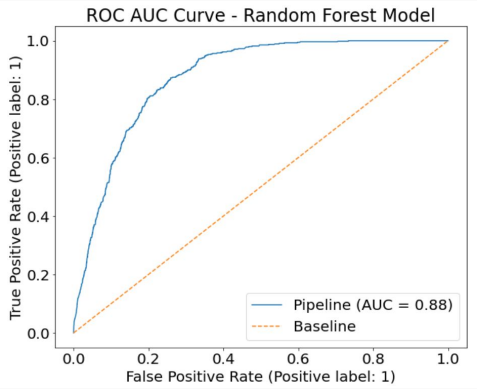
03 STREET ADDED AS FEATURE

04 MOVING AVERAGES FOR WEATHER VARIABLES ADDED AS FEATURES
- 3 AND 7 DAY CYCLES

05 CERTAIN FEATURES REMOVED DUE TO LOW FEATURE IMPORTANCE

05. CONCLUSION RECOMMENDATIONS

1. CONCLUSION
2. RECOMMENDATIONS



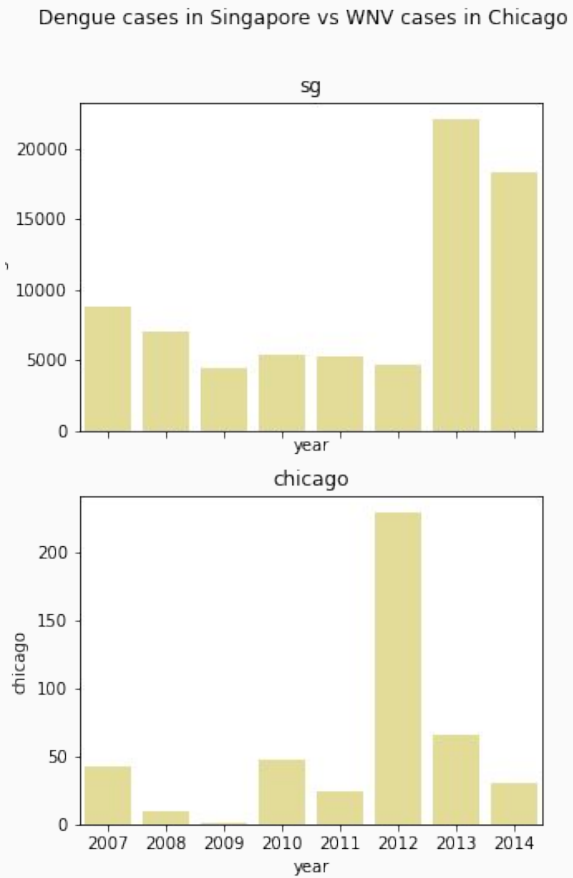
	0	1
1	longitude	0.109392
0	latitude	0.096560
178	species_CULEX PIPIENS/RESTUANS	0.054656
177	species_CULEX PIPIENS	0.048690
179	species_CULEX RESTUANS	0.029478
164	daylightmin_sma_7	0.019180
163	daylightmin_sma_3	0.018983
144	daylightmins	0.018868
123	street_ W OHARE AIRPORT	0.017050
34	street_ N OAK PARK AVE	0.016469

CONCLUSION

1. Model achieved 38% improvement on baseline with an ROCAUC score of 88% on the test set.

2. Feature engineering did improve accuracy of model

3. Relative importance of features
   a. Location data (longitude, latitude, certain streets)
   b. Species of mosquito
   c. Length of day
   d. Other features such as precipitation and temperatures were less important.



Source: Singapore statistics, Chicago statistics

FINAL THOUGHTS: WNV MAY NOT BE A PRESSING PROBLEM FOR CHICAGO

1. The model only measures whether mosquitos in the trap carry WNV.

2. Weather in Chicago is much less conducive for transmission.

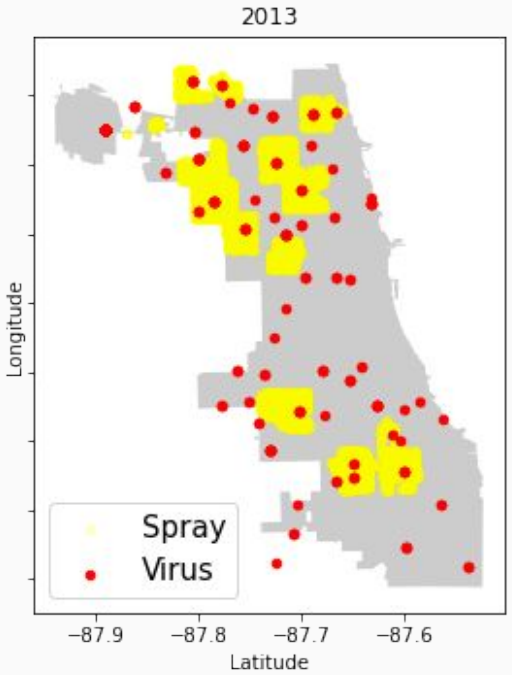
3. It is still possible to reduce the breeding of mosquitoes based on the key features.

RECOMMENDATION #1: OUTREACH AND MONITORING

New mosquito control toolkit for the community, launched by Singapore's National Environment Agency in 2020

Source: NEA

1. Public health education on source control
2. Include weekly checking and eliminating of potential mosquito breeding grounds in asset management operating procedures.
3. Active community involvement for personal protection.
4. Monitoring trap data, especially in key clusters, for early detection.



RECOMMENDATION #2 SPRAY MORE (AND IN THE RIGHT PLACES)

1. Only started spraying in 2011 and even in 2013, was not widespread.
2. Recommendation is only to do fogging when mosquito population is high.
3. Targeted spraying in key clusters to reduce cost.
4. More drastic measures include introduction of larvicides near water-bodies in compliance with EPA guidelines

US$105M TO SPRAY THE WHOLE OF CHICAGO

Land area of Chicago (149,894 acres) * Cost per acre per season ($700) = **$104,925,800**

COST- BENEFIT ANALYSIS OF SPRAYING (US$/YEAR)

Cost to cover the entire Chicago metropolitan area:

US$105M

Land area of Chicago (149,894 acres) * Cost per acre per season ($700)

Avg yearly value of human lives saved from 2007 to 2014:

US$9.1 M

Avg number of deaths from WNV / year (1) * Value of human life ($9.1m)

CHICAGO SHOULD LIMIT EXPENDITURE ON MOSQUITO CONTROL POLICIES

COST- BENEFIT ANALYSIS OF SPRAYING (S$/YEAR)

NEA Budget for mosquito control (2013):

S$85M

Source: Today Online

Avg yearly value of human lives saved from 2007 to 2014:

S$63M

Avg number of deaths from dengue / year (8.75) * Value of human life (S$7.2m)

THE VALUE OF LIFE IN SINGAPORE IS LOW

Entomologist checks on mosquitoes being raised in a laboratory.

Source: CDC

RECOMMENDATION #3 BIOENGINEERING

1. Culling bird populations especially crows
2. Sterile insect techniques
3. Introduction of mosquito predators
4. GM mosquitoes do not pose a risk to people, animals or the environment.
5. Should be used in conjunction with other mosquito control methods during mosquito season.

THANKS

Does anyone have any questions?