

West Nile Virus (WNV) Executive Summary:

Guiding Questions:

- 1) Determine where and under what conditions is WNV most prevalent?
- 2) What is the most cost-effective strategy in preventing WNV?
- 3) How can we positively identify WNV carrying mosquitoes?

EDA:

One of the most challenging aspects about this project was the huge class imbalances and merging different datasets together. We also had to feature engineer weather related data with different rolling time frames.

Another interesting aspect related to this project was working with coordinate grids. We used a package called *Vincenty* to calculate distances between spray locations and trap data. We also used *utils* from *sklearn.utils* package for addressing class imbalances. The coordinate grids allowed to plot incidents where WNV was present and see if spraying impacted WNV outbreaks.

Our EDA found that humidity and precipitation correlated with mosquito population and WNV present observations. We also found out which

Modeling:

Since we want to reduce the number of false negative rates, we need a model that is optimized for sensitivity. Our group decided that missing false negative incidents would cause greater harm than false positives. We ran our data through Random Forest, XGBoost, AdaBoost, and Logistic Regression models. Even though our accuracy score using XGBoost was the best, its sensitivity score was not as good as the AdaBoost model.

Policy Recommendations:

Even though our data and external research shows spraying's effectiveness is inconclusive, we still recommend spraying in known 'hot spots' over a duration of 4-5 weeks consecutively. Spraying does seem to decrease the overall population of mosquitos and it will not cost as much as dealing with a WNV epidemic.

The City of Chicago's mosquito abatement budget for 2017 is \$2 million. It is a small price to pay when the average medical cost to treat someone with WNV is around \$25,000. This doesn't take into account the lost productivity costs and loss in tourism.

We also recommend incorporating dead bird surveillance into future datasets for WNV. The City of Chicago has a dead bird surveillance program already in place and a hot line for residents to call. However, it isn't used as a feature that could be used in models.