



Kaggle Challenge

Team: Beguiling West Niling



Presentation Overview

Part 1 - Problem Statement

Part 2 - EDA

Part 3 - Feature Engineering and Modelling

Part 4 - Cost-Benefit Analysis

Part 5 - Recommendations and Conclusions

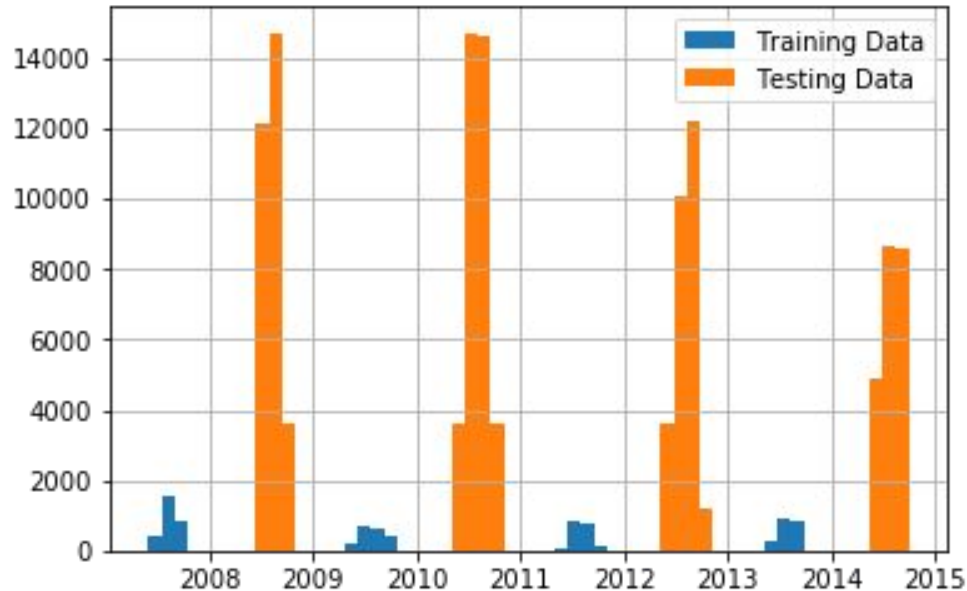




Problem Statement

Can you predict the probability of an outbreak of West Nile Virus provided information of location, time, weather, and mosquitos found?

Exploratory Data Analysis



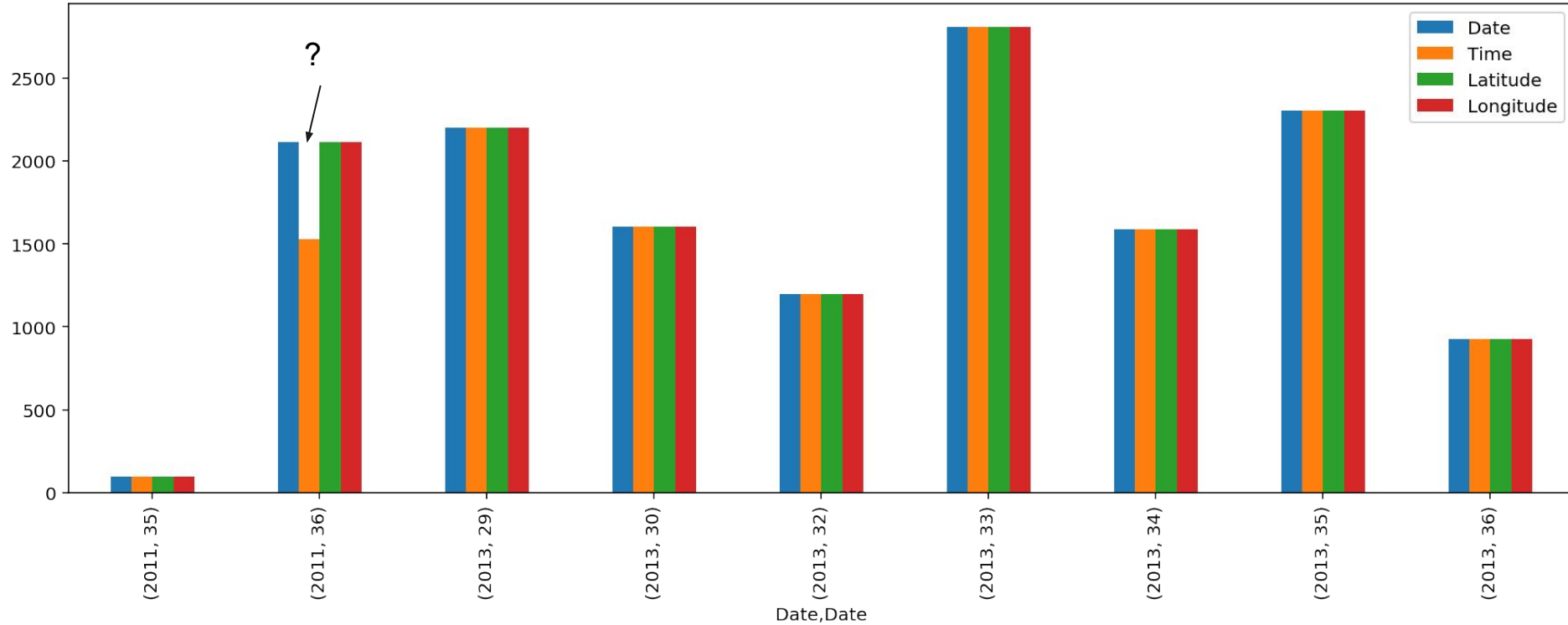
- Training data and testing data were split by alternating years with no overlap
- Training data was far less abundant than the testing data
- Given date, species, location, and trap number



Exploratory Data Analysis

- Weather from multiple stations.
- Various weather data ranging from temperature to precipitation.
- Lots of missing data.

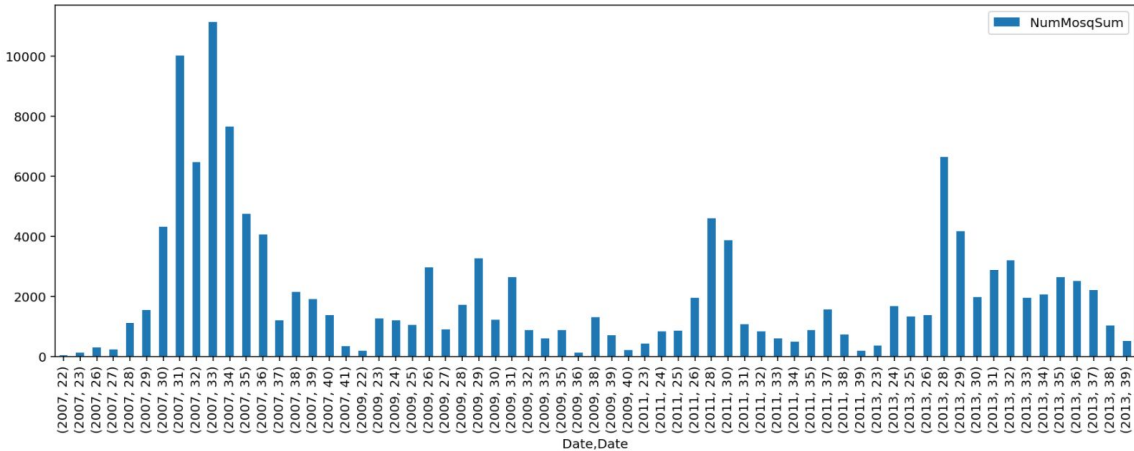
Spray data info per week





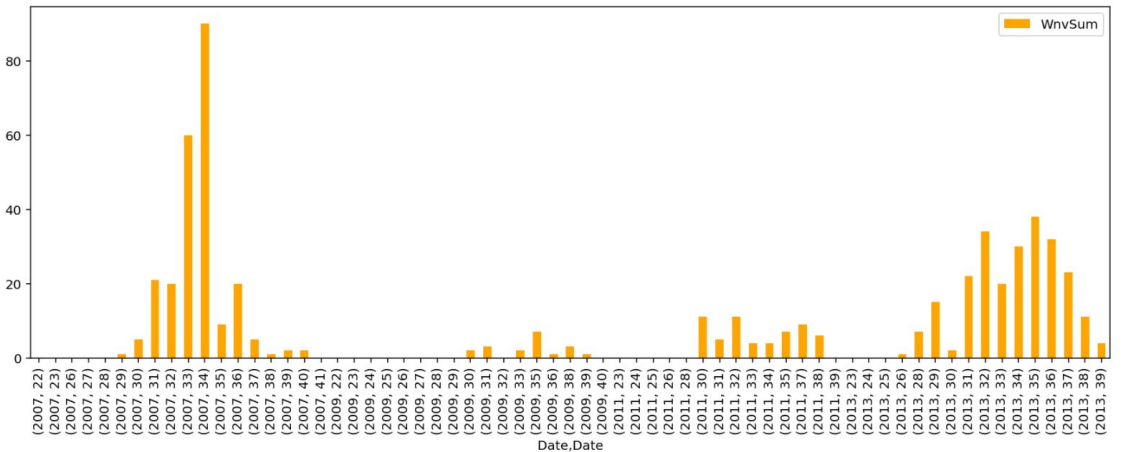
Mosquitoes found

Per week



West Nile Events

Per week





When & Where

Spraying in Time:

https://bricewolfgang.carto.com/viz/4722e417-ac9e-426d-a860-952c86fd1477/public_map

West Nile Events in Time:

https://bricewolfgang.carto.com/viz/ac4a9ac3-568b-4161-a6d5-927266ce491e/public_map



Feature Engineering

- Mosquito Species
- Trap
- Date
- Days Since Last Checked
- Spray Factor
- Weather Codes
- Length of Day
- Aggregated Weather Data

Feature Engineering

- Time Lagged Data
- Weighted Traps
- Leakage



Modeling



- Used an XGBoost model.
- Our best model gave us a public AUC-ROC score of **0.76892** and a private score of 0.76117.
- Our second-best model (which must have been a tad overfit) gave us a public AUC-ROC score of **0.77068** and a private score of 0.75741.



Cost-Benefit Analysis

- Determined medical and productivity costs vs. spraying cost.
- Determined benefits (lives saved).
- Performed modified spray plans.
- Alternatives.
- Analyzed nuances.



Cost of Abatement Program *

- Regular years - ~\$2 Million (maximum \$3 Million per year)
- Outbreak years - with emergency spraying - ~\$6 Million
- 16 Million total for 2008-2013
- **Current Program:** includes treating roughly 40,000 stormwater catch-basins and about 3,500 off-road sites and retention ponds with biological products to prevent mosquito larvae from maturing into adults.

*Spraying is not extremely effective in general. It's also not done very effectively.



Health Care Costs

- Less than 1% of patients develop neuroinvasive WNV. however...
- Median cost for a patient with AFP was **\$20,774** (range \$13k–\$140k)
- Median cost for a patient with WNV non-neuroinvasive was **\$4,467** (range \$3,241–\$8,433)
- **Uninsured patients** → Cook County Health Board reports that **1 of 3 patients** in Cook County health system is uninsured.
- The system spent \$ 400 Million last year on uncompensated health care
- Chicago child homeless and unaccompanied populations: **31%**
 - Adding to immuno compromised population



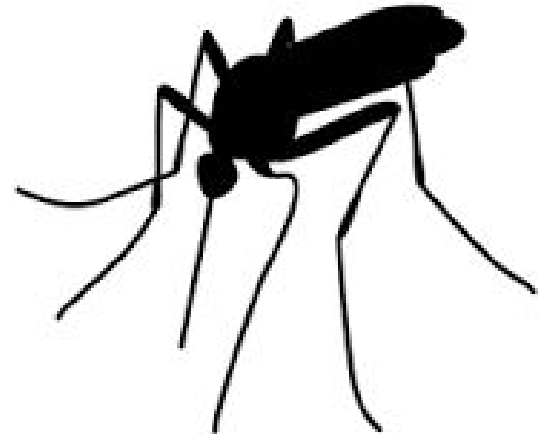
Incidence of West Nile in Cook County, IL

\$2 Million per Year

Year	Cases	Deaths	Death rate	Wnv Found
2007	33	0	0	189
2008	9	0	0	n/a
2009	1	0	0	19
2010	30	1	3.3%	n/a
2011	22	1	4.5%	50
2012	174	5	2.8%	n/a
2013	60	7	11.6%	199
2015	80	9	11.25%	n/a

Productivity Costs

- Roughly \$1,000 per person.
- Not significant relative to other costs.
- Few people develop symptoms





Varied Spray Plans

- Predicted highest-risk locations and times
- Created spray schedules with varying levels of coverage
- \$500k - \$3 million per year
- Used trained model to predict WNV given the various spraying schedules
- Spraying was found to be ineffective



Additional Considerations

- Unforeseen and unquantifiable costs:
 - Example: Overtime costs for emergency spraying
 - Example: cost of a life/ health care
- Community Outreach Programs
- Social Media
- Environmental Ethics



Cost-Benefit Analysis - Conclusions

- ~ \$2 million per year of abatement programs
- higher concerns about WNV due to more high risk populations
- 2017 updates
- Spraying 99.9% ineffective: not recommended to expand

Any Questions?

