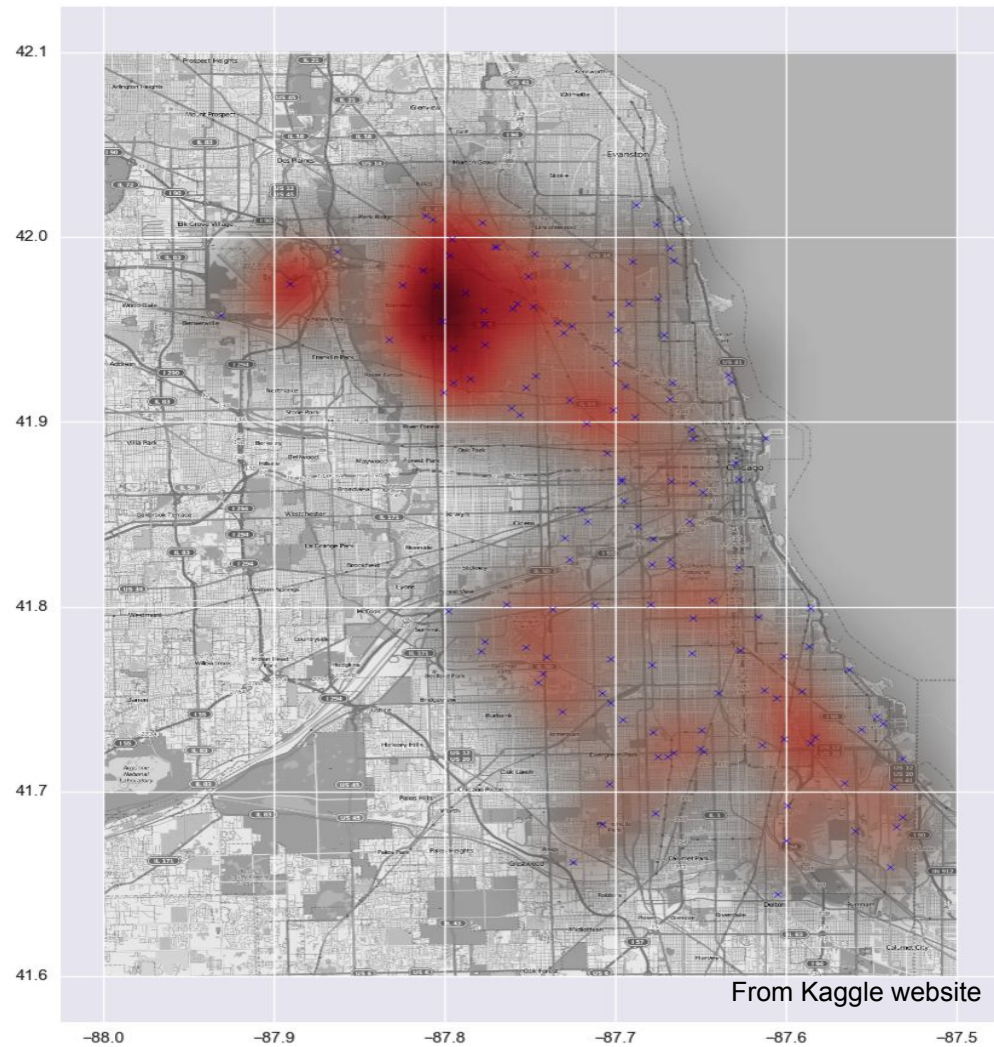
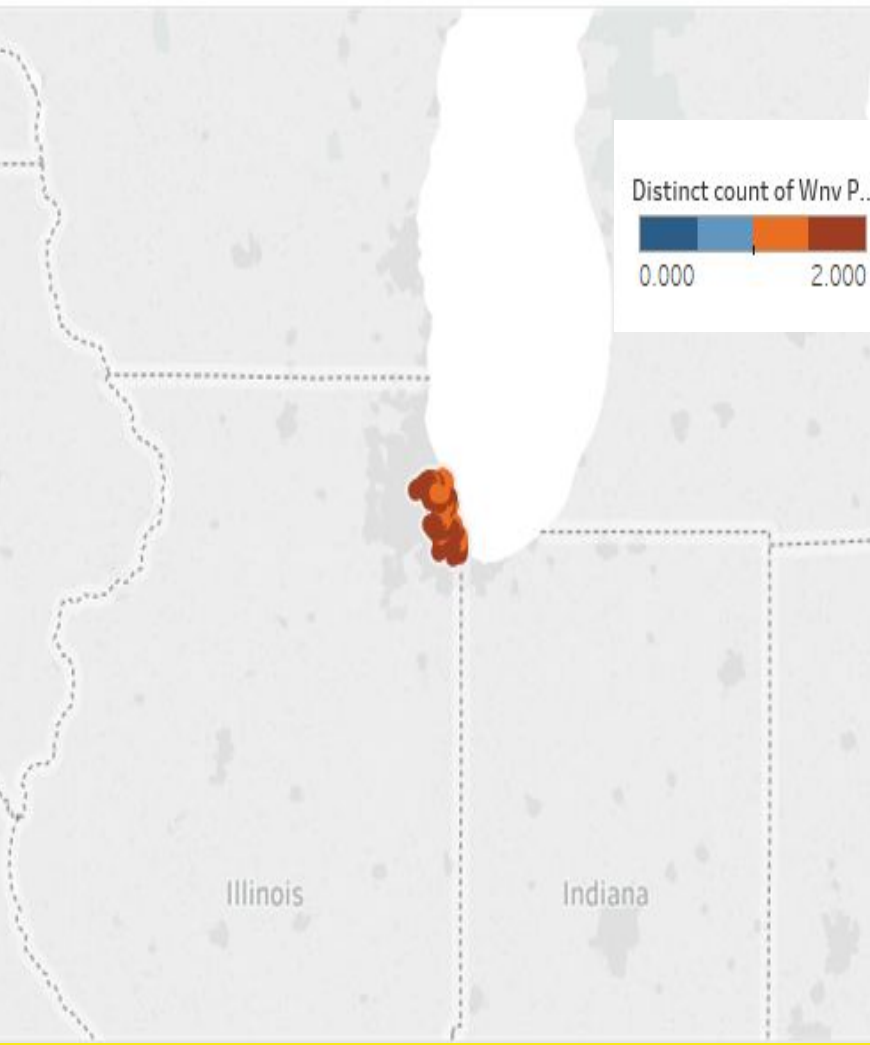


West Nile Virus–Chicago 2007–2014

Khatereh Mohajery, Tim McKinnon, Swara Salih

Background

- Aim of project:
 - Create a model to reliably predict the occurrence of West Nile Virus (WNV) in Chicago
- Data:
 - Weather conditions
 - Mosquito trap contents
 - Spray data
- Limitations:
 - The competition does not permit use of outside data
 - We only considered an aggregate of previous month of weather conditions
 - Data about use of insecticide spray was omitted for test set



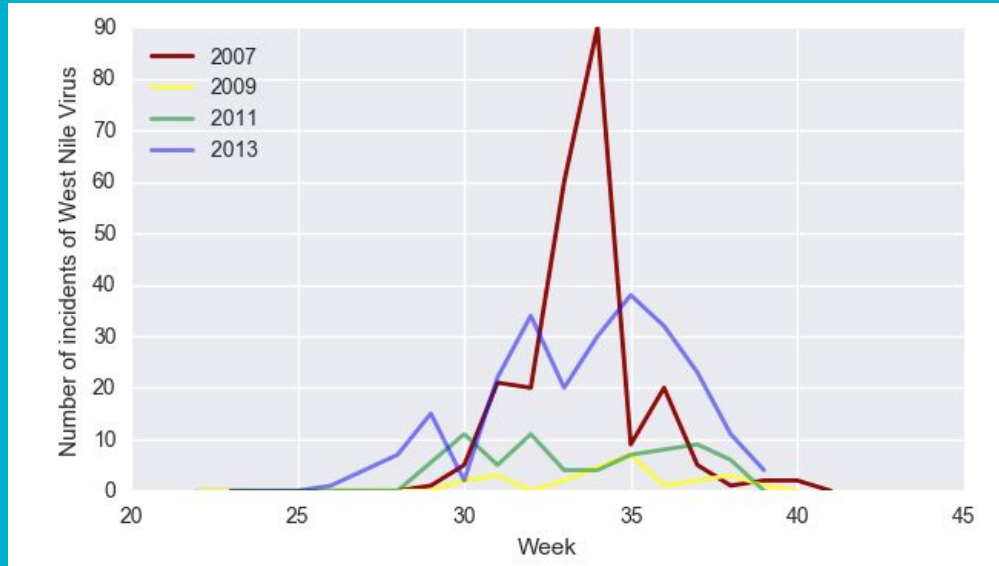
Structure of presentation

- Description of the dataset/EDA
- Models / Features used
- Results/Recommendations

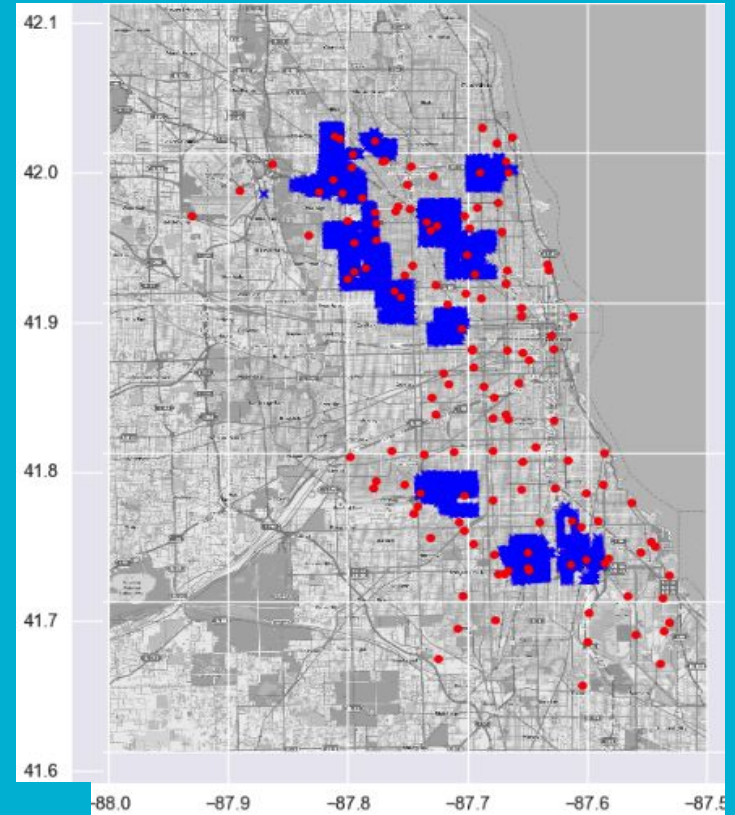
Train Dataset

	Date	Trap	Species	Latitude	Longitude	NumMosquitos	WnvPresent
0	2007-05-29	T002	CULEX PIPIENS/RESTUANS	41.954690	-87.800991	1	0
1	2007-05-29	T002	CULEX RESTUANS	41.954690	-87.800991	1	0
2	2007-05-29	T007	CULEX RESTUANS	41.994991	-87.769279	1	0
3	2007-05-29	T015	CULEX PIPIENS/RESTUANS	41.974089	-87.824812	1	0
4	2007-05-29	T015	CULEX RESTUANS	41.974089	-87.824812	4	0

Traps

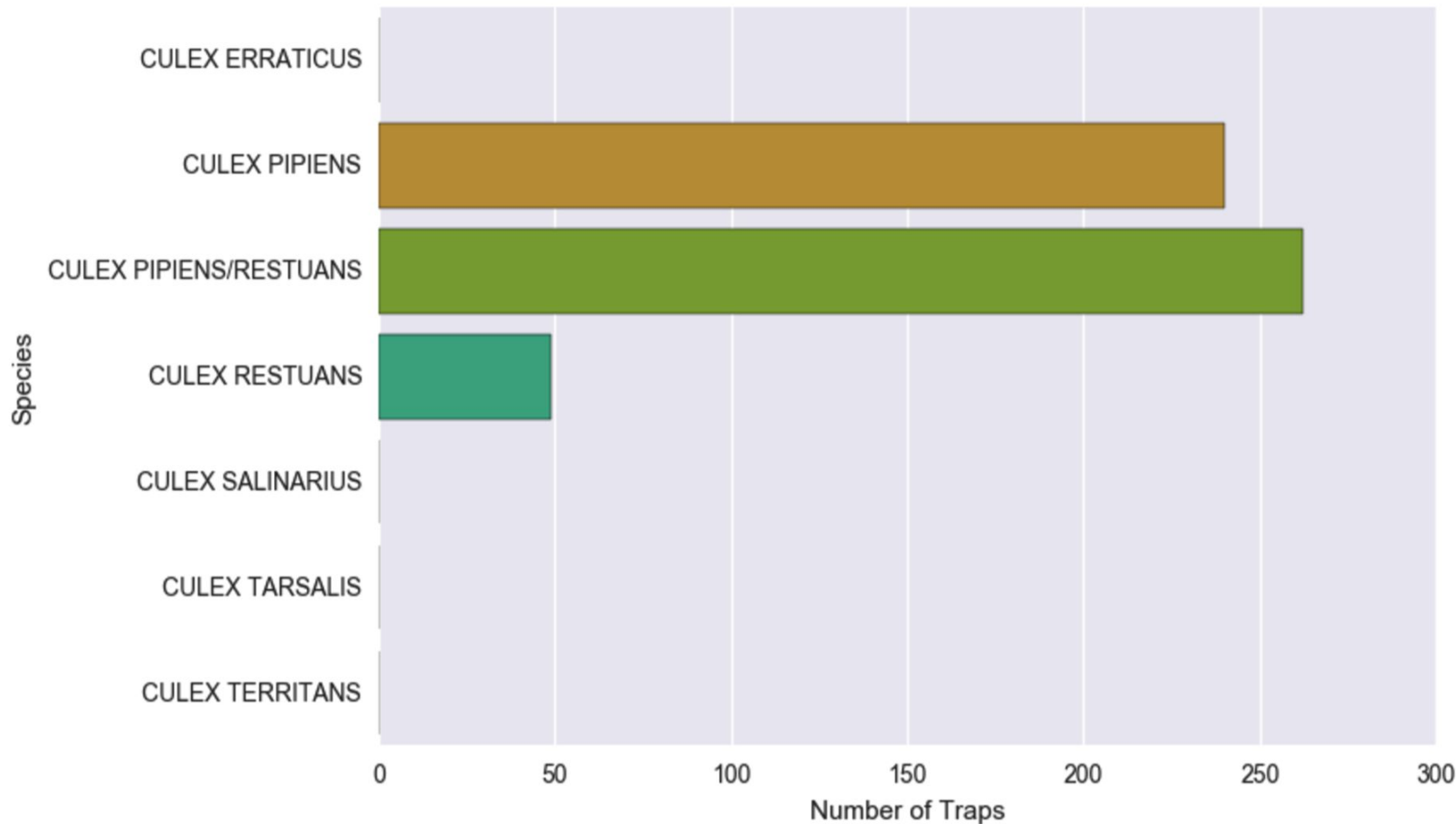


Total number of traps with West Nile Virus



Sprayed Areas in 2011 and 2013

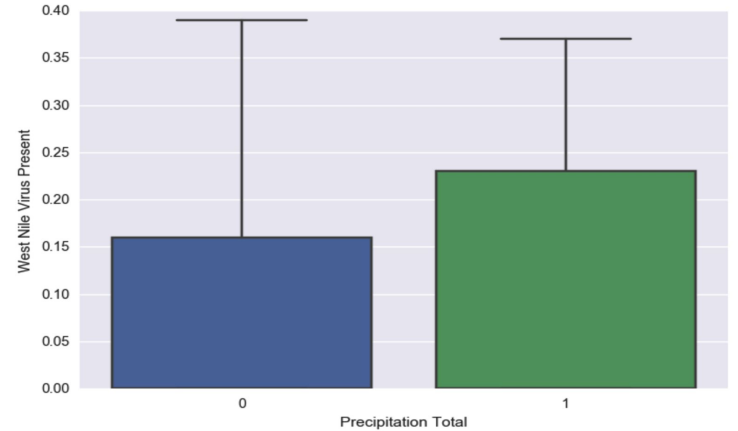
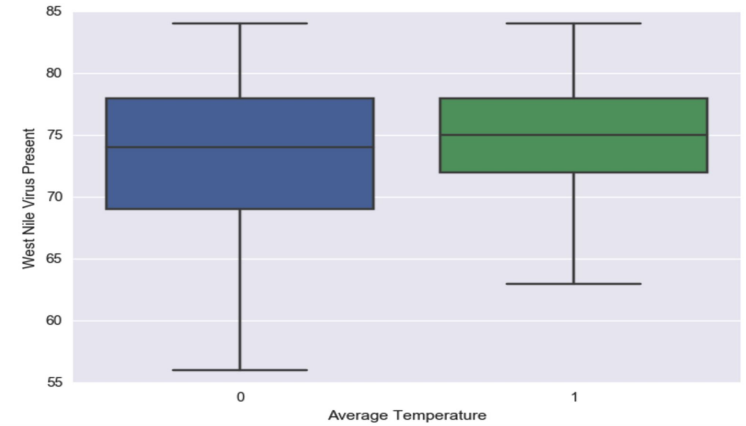
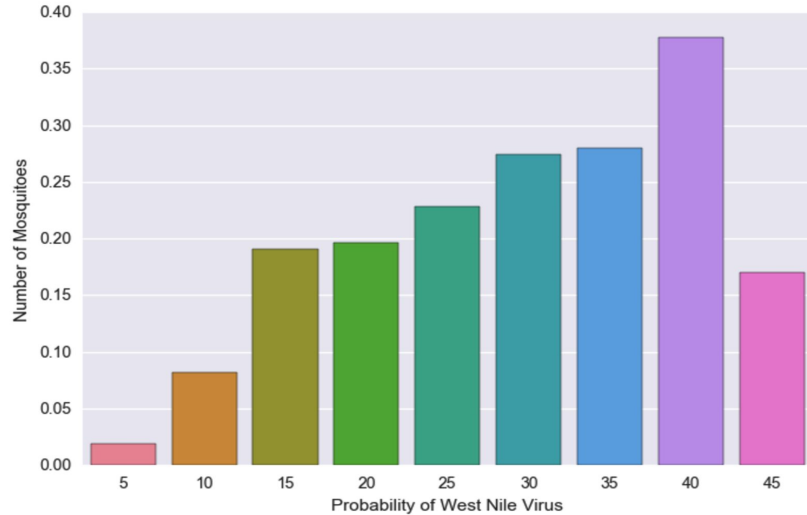
Species caught by Traps



Culex Pipiens



Feature Engineering



Features

● Trap information:

- Locations
- Number of mosquitos and species counts
- Training n=10,000 (500 observations positive for WNV)
- Chronological information

● Weather:

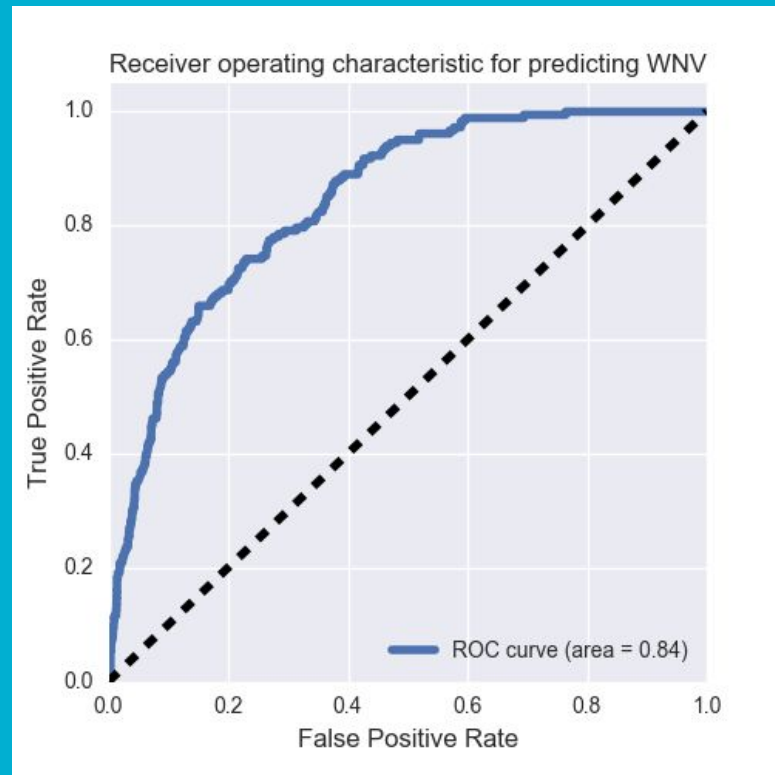
- features aggregated over 30 days previous to observation
 - Dew point
 - Max/Min Temps
 - Avg Temps
 - Weather condition codes
 - etc.

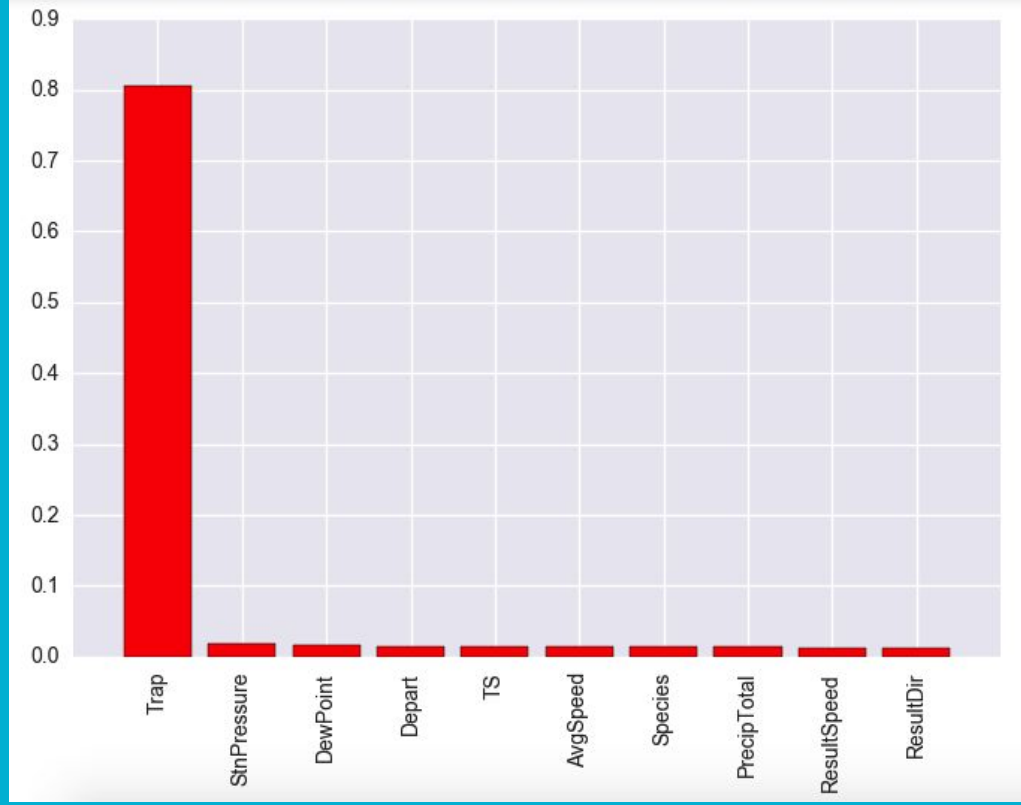
Models:

- KNN
- SVM
- Ada Boosting
- Logistic Regression
- Random Forest
- Gradient Boosting Classification

-*AUC 0.77376 for test set*

-#324 on Kaggle





Recommendations

- Individual trap locations are most predictive of WNV occurrence
 - Weather conditions are also important features
- Although we focused on the previous 30 days before measurements, further research needed to identify the most predictive timeframe
- Total hatch size (as measured by trap counts) highly correlates with WNV detection; however, we failed to include this feature

322	↑19	awilliams	0.77406	12	Wed, 17 Jun 2015 14:18:46
323	↓24	tk	0.77398	24	Mon, 15 Jun 2015 20:52:01 (-3d)
-		Kmohajeri	0.77376	-	Fri, 13 Jan 2017 04:50:59 Post-Deadline

Post-Deadline Entry

If you would have submitted this entry during the competition, you would have been around here on the leaderboard.

324	↓25	Unlabeled	0.77362	25	Mon, 15 Jun 2015 20:52:01 (-3d)
-----	-----	-----------	---------	----	---------------------------------