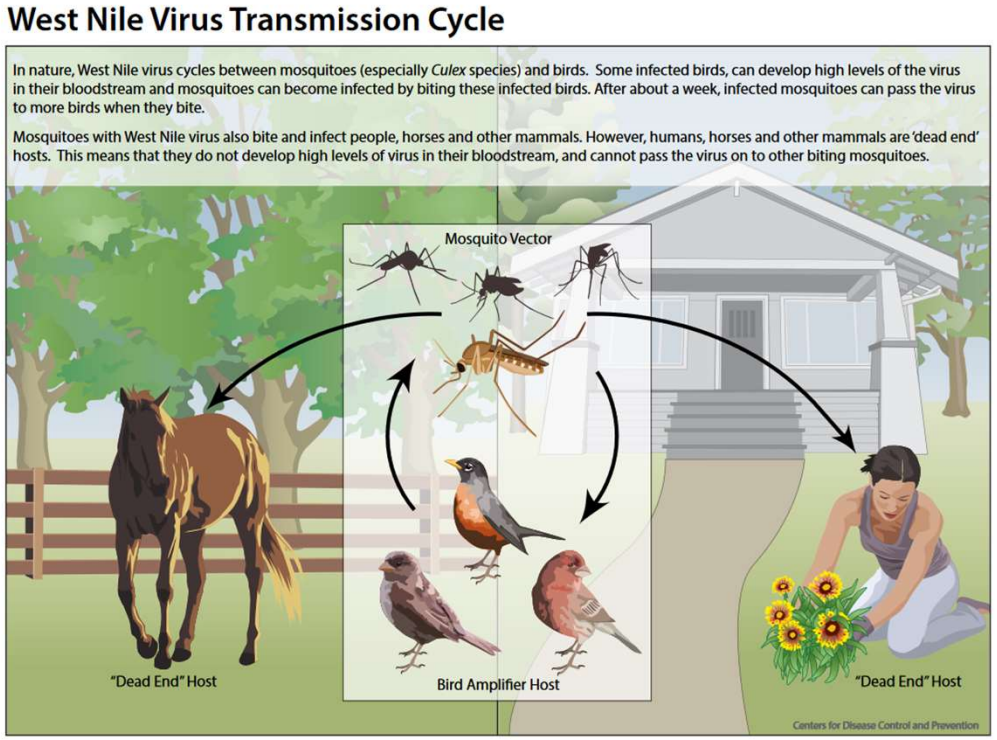
Keeping our city safe from West Nile virus

Mosquito Breeding Team

West Nile virus (WNV) - the facts

- Most commonly spread to people by the bite of an infected mosquito.
- Cases of WNV occur during mosquito season - July and August.
- There are no vaccines to prevent or medications to treat WNV in people.
- Fortunately, most people infected with WNV do not feel sick.
    - About 1 in 5 people who are infected develop a fever and other symptoms.
    - About 1 out of 150 infected people develop a serious, sometimes fatal, illness.



Problem statement

To develop a predictive model that will forecast the probability of WNV presence in 138 mosquito traps around Chicago over the course of a season. WNV is a communicable disease that is spread through its most common vector, mosquitoes

Information provided:

- Mosquito trap data: location of traps, how many mosquitos on which date, and whether they have the West Nile Virus
- Weather data: 2 Chicago weather station reports with many indicators (temperature, dew point, sunrise/sunset)
- Spray data: gis data for insecticide spraying by the city.

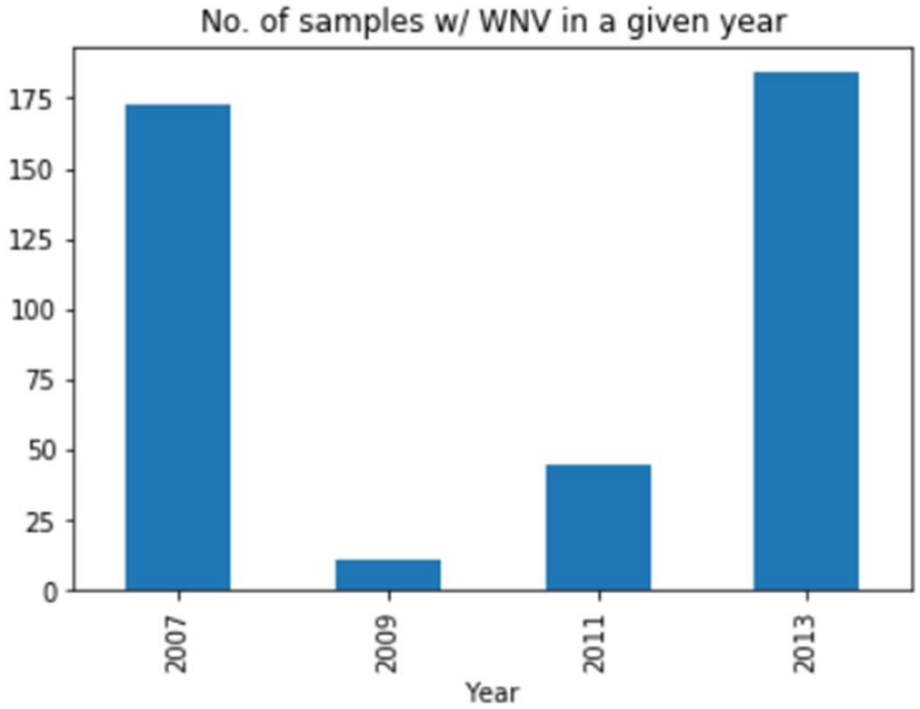
EDA

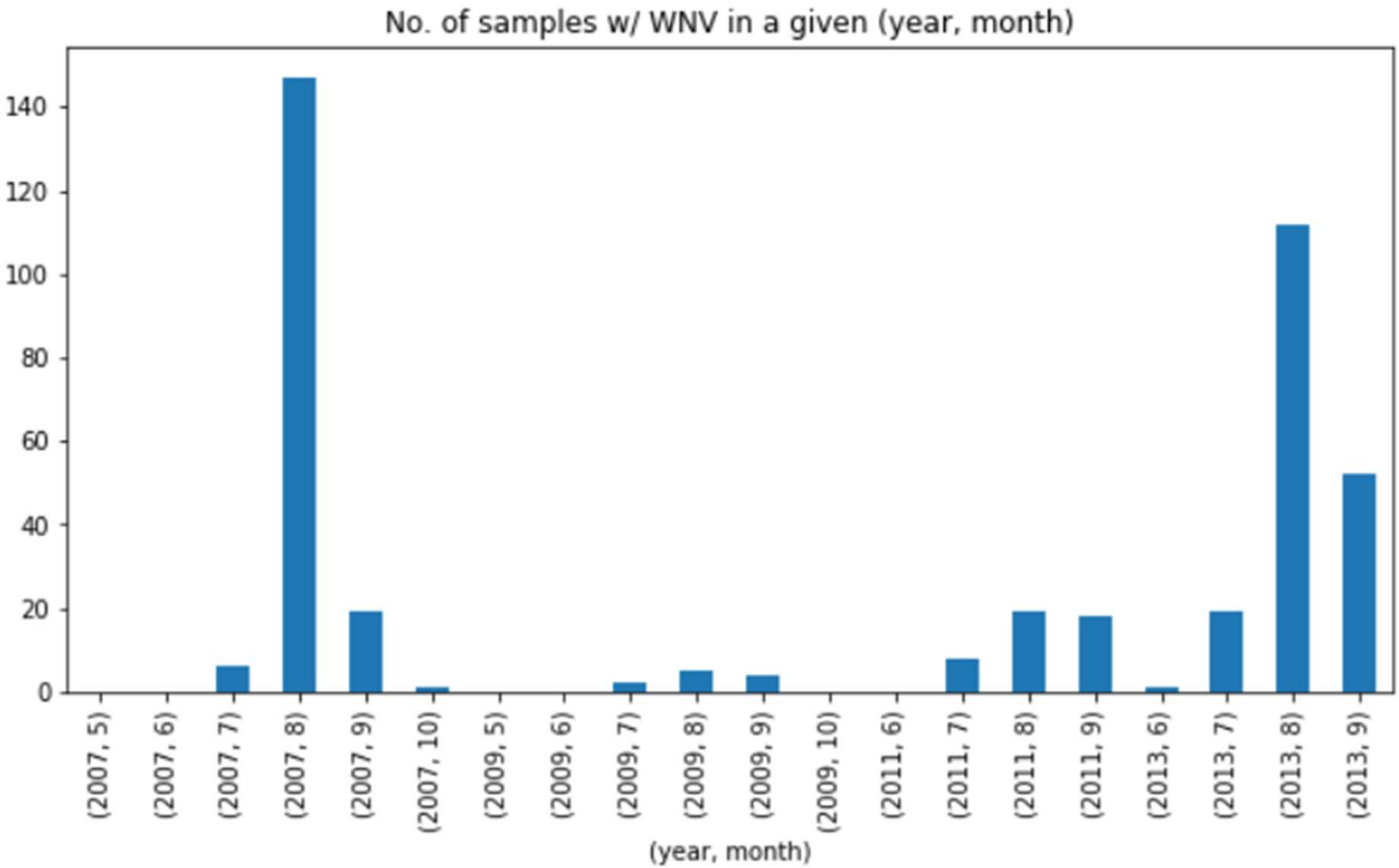
Mapped weather data across to training set

Final list of features selected -

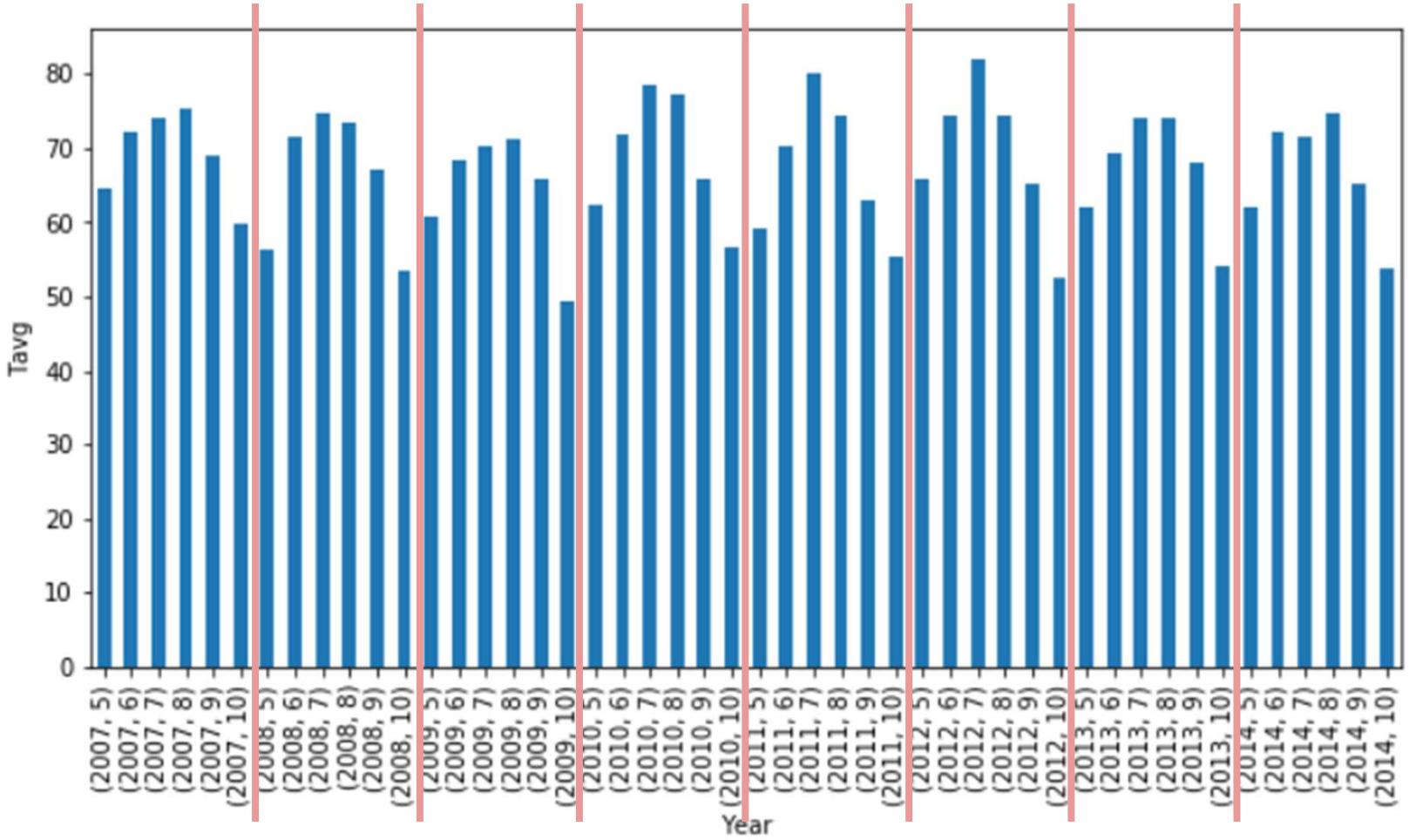
- Latitude
- Longitude
- Tavg
- DewPoint
- Days since rain
- Tavg_rolling3day, Tavg_rolling7day, Tavg_rolling14day
- DewPoint_rolling3day, DewPoint_rolling7day, DewPoint_rolling14day

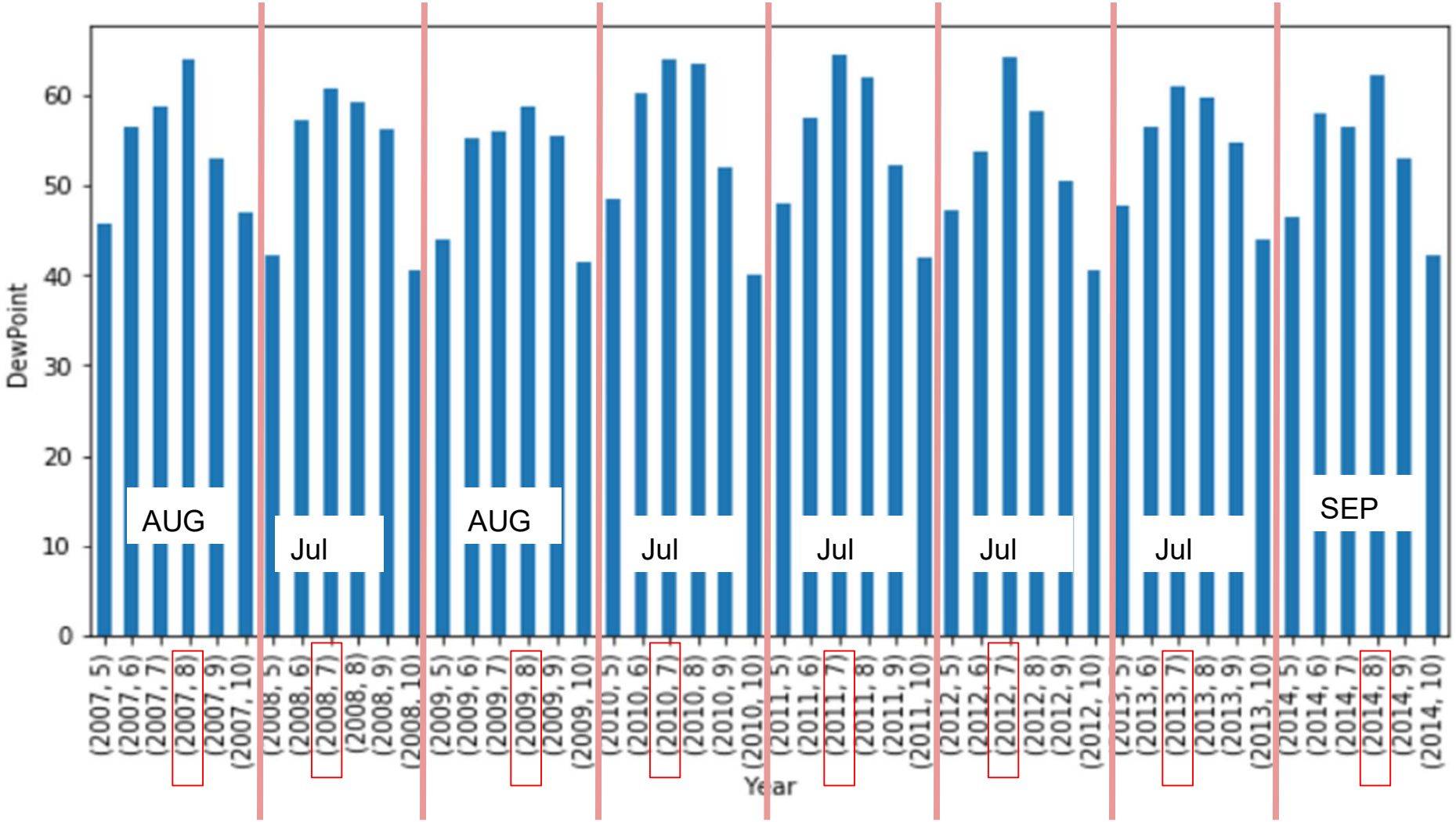
Visualising training data

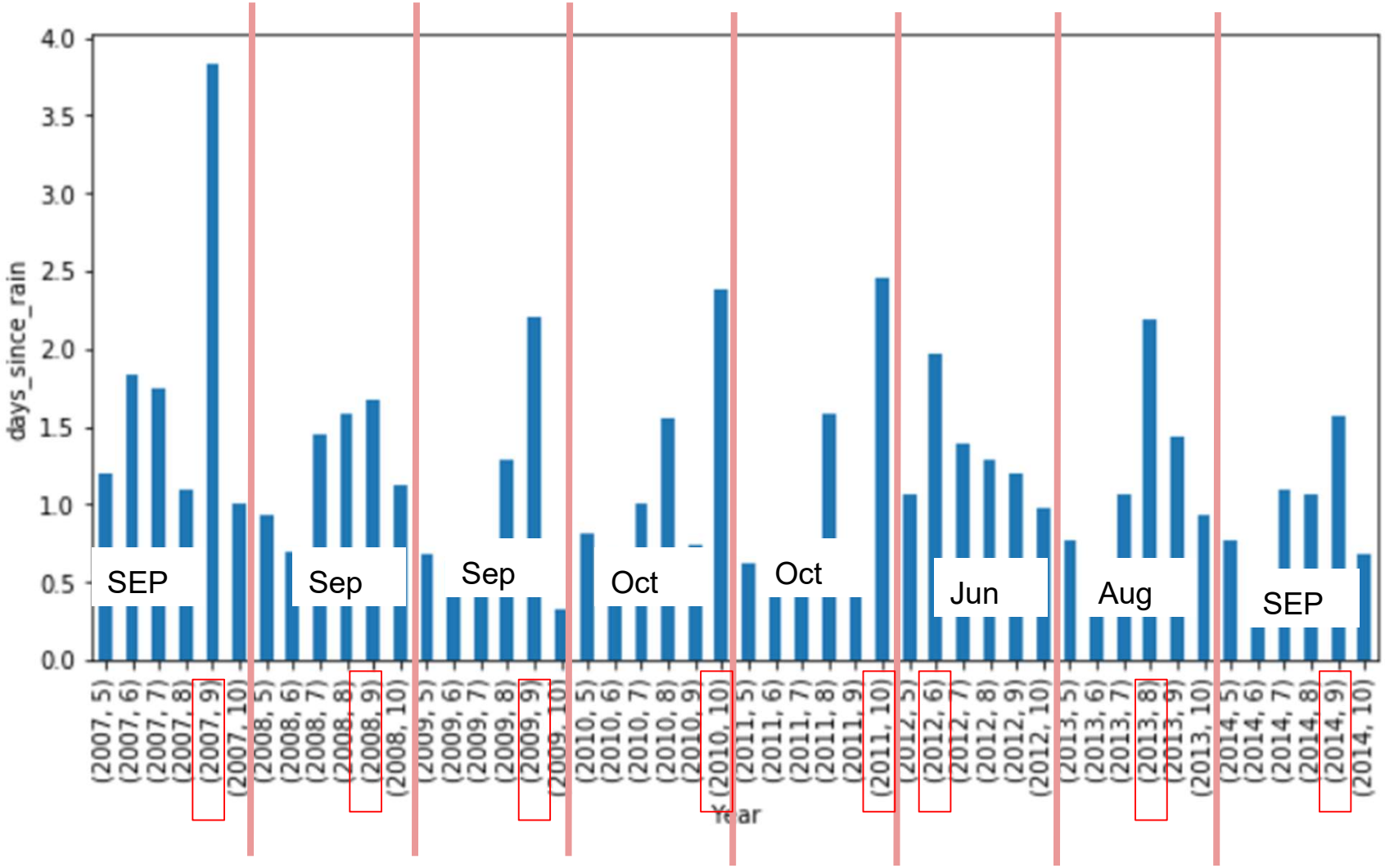


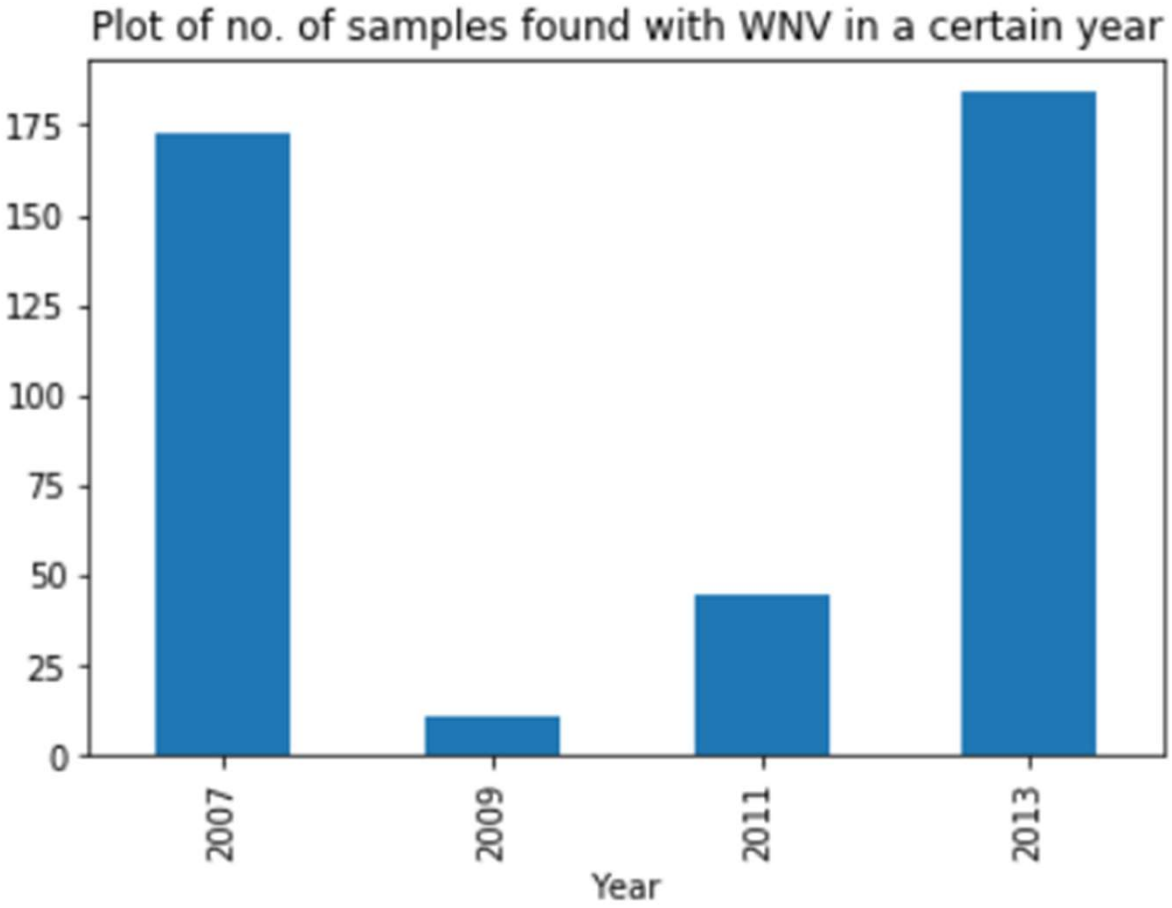

No. of samples w/ WNV in a given (year, month)
140
120
100
80
60
40
20
0
(2007, 5)
(2007, 6)
(2007, 7)
(2007, 8)
(2007, 9)
(2007, 10)
(2009, 5)
(2009, 6)
(2009, 7)
(2009, 8)
(2009, 9)
(2009, 10)
(2011, 6)
(2011, 7)
(2011, 8)
(2011, 9)
(2013, 6)
(2013, 7)
(2013, 8)
(2013, 9)
(year, month)

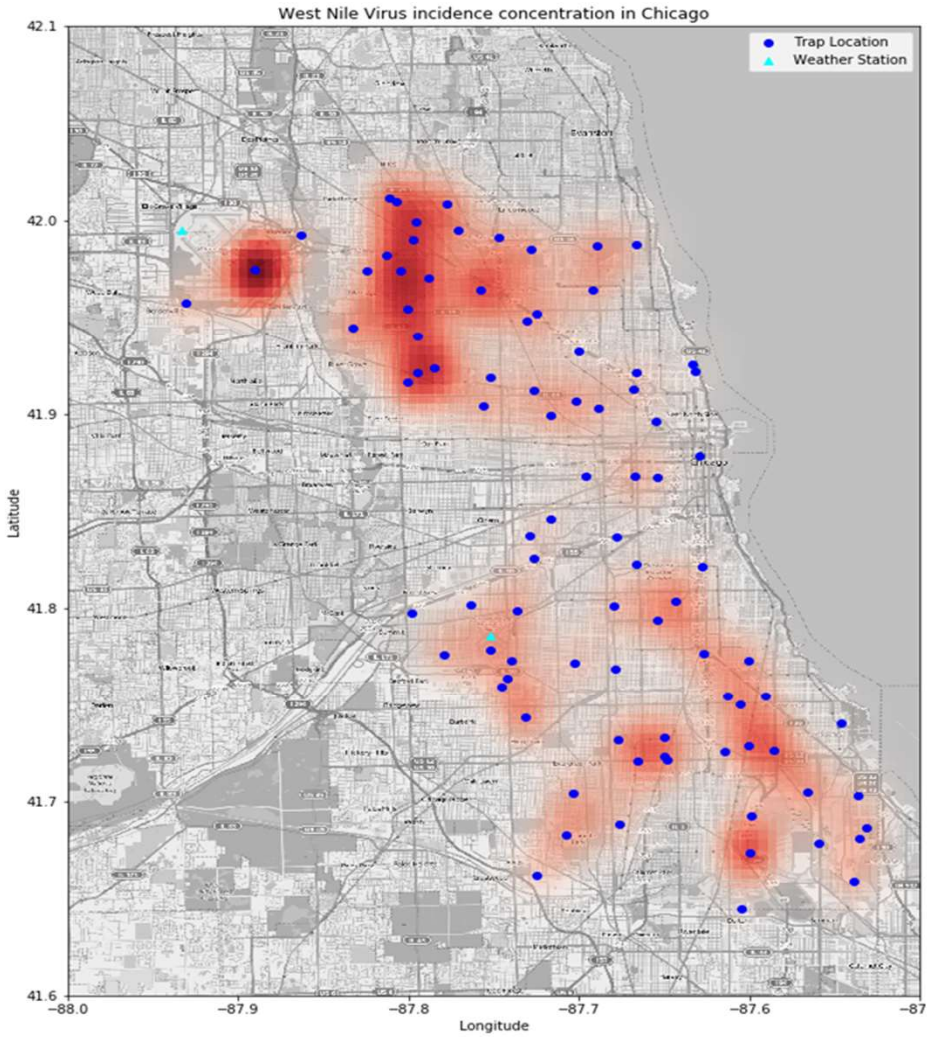
Visualising weather data

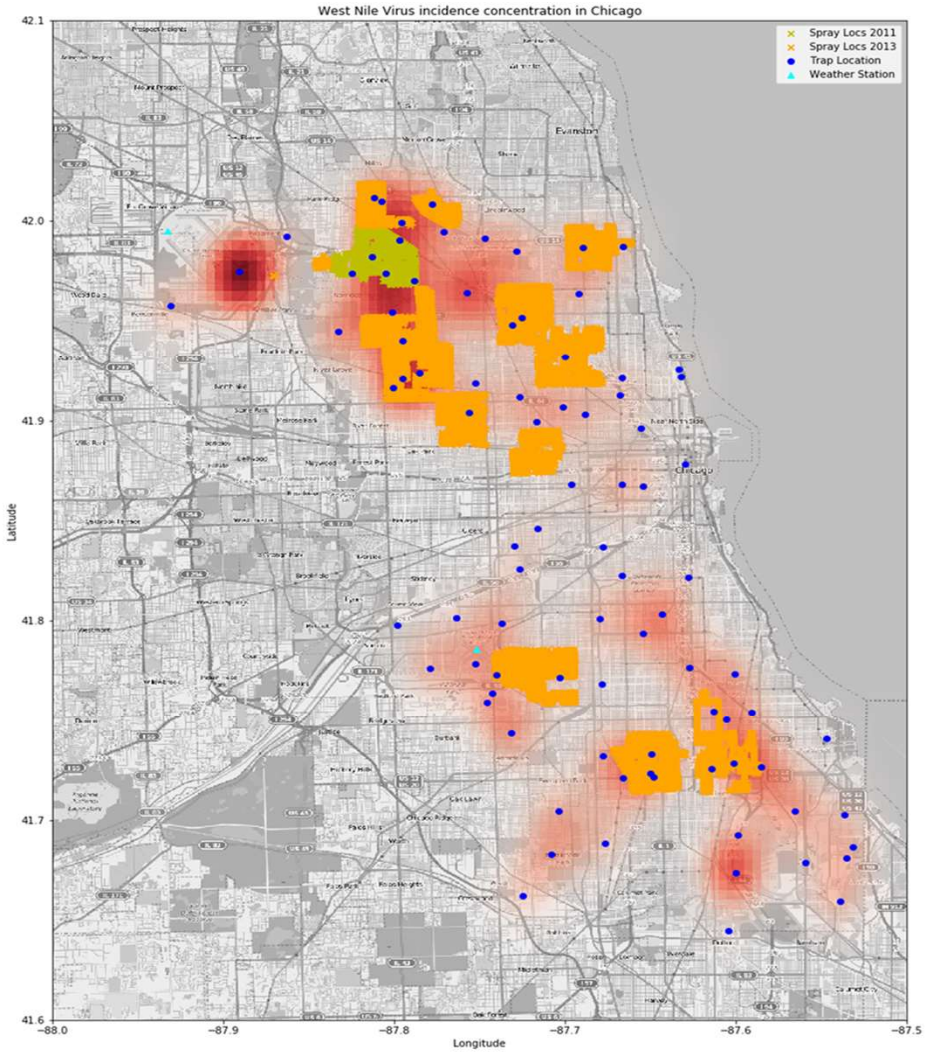


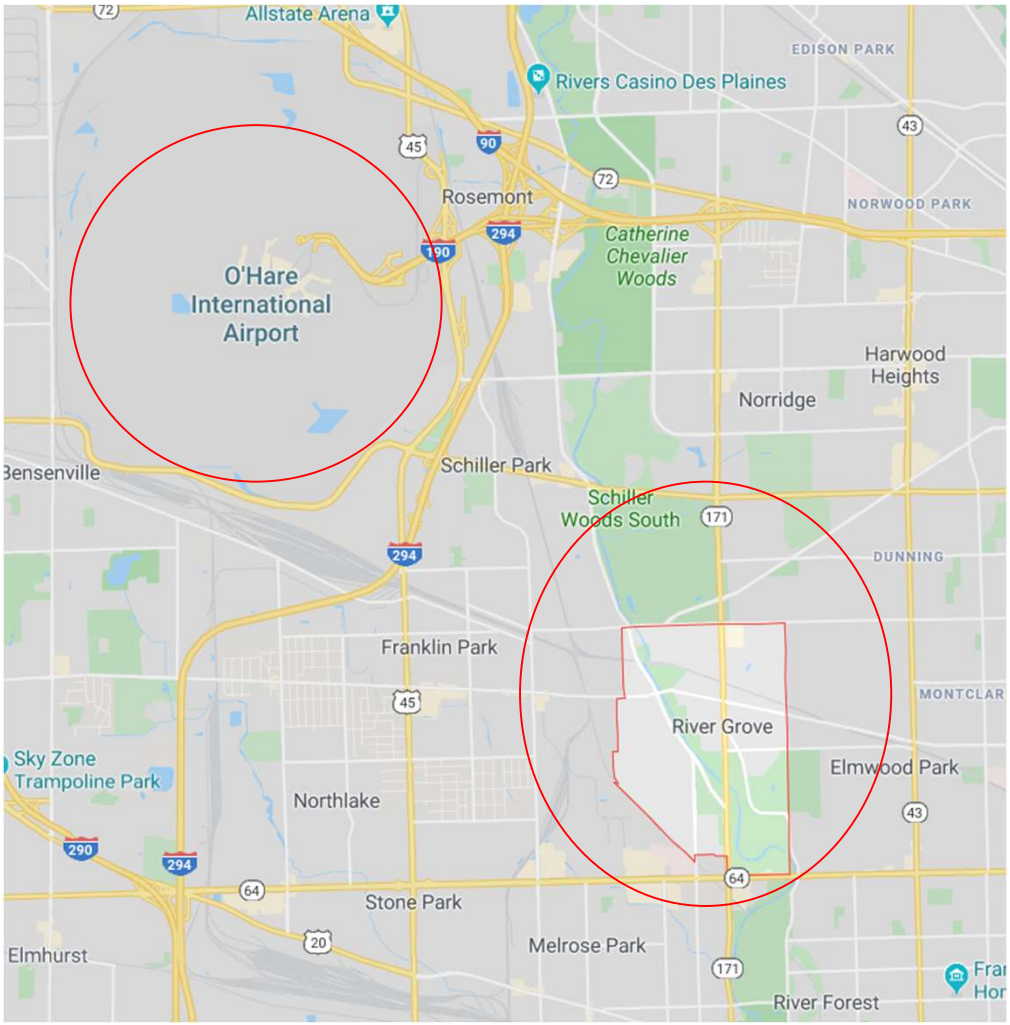

DewPoint
0
10
20
30
40
50
60
AUG
Jul
AUG
Jul
Jul
Jul
Jul
SEP
(2007, 5)
(2007, 6)
(2007, 7)
(2007, 8)
(2007, 9)
(2007, 10)
(2008, 5)
(2008, 6)
(2008, 7)
(2008, 8)
(2008, 9)
(2008, 10)
(2009, 5)
(2009, 6)
(2009, 7)
(2009, 8)
(2009, 9)
(2009, 10)
(2010, 5)
(2010, 6)
(2010, 7)
(2010, 8)
(2010, 9)
(2010, 10)
(2011, 5)
(2011, 6)
(2011, 7)
(2011, 8)
(2011, 9)
(2011, 10)
(2012, 5)
(2012, 6)
(2012, 7)
(2012, 8)
(2012, 9)
(2012, 10)
(2013, 5)
(2013, 6)
(2013, 7)
(2013, 8)
(2013, 9)
(2013, 10)
(2014, 5)
(2014, 6)
(2014, 7)
(2014, 8)
(2014, 9)
(2014, 10)
Year


days_since_rain
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
SEP
Sep
Sep
Oct
Oct
Jun
Aug
SEP
(2007, 5)
(2007, 6)
(2007, 7)
(2007, 8)
(2007, 9)
(2007, 10)
(2008, 5)
(2008, 6)
(2008, 7)
(2008, 8)
(2008, 9)
(2008, 10)
(2009, 5)
(2009, 6)
(2009, 7)
(2009, 8)
(2009, 9)
(2009, 10)
(2010, 5)
(2010, 6)
(2010, 7)
(2010, 8)
(2010, 9)
(2010, 10)
(2011, 5)
(2011, 6)
(2011, 7)
(2011, 8)
(2011, 9)
(2011, 10)
(2012, 5)
(2012, 6)
(2012, 7)
(2012, 8)
(2012, 9)
(2012, 10)
(2013, 5)
(2013, 6)
(2013, 7)
(2013, 8)
(2013, 9)
(2013, 10)
(2014, 5)
(2014, 6)
(2014, 7)
(2014, 8)
(2014, 9)
(2014, 10)
Year


Plot of no. of samples found with WNV in a certain year
175
150
125
100
75
50
25
0
2007
2009
2011
2013
Year


West Nile Virus incidence concentration in Chicago
Trap Location
Weather Station
Latitude
Longitude
42.1
42.0
41.9
41.8
41.7
41.6
−88.0
−87.9
−87.8
−87.7
−87.6
−87


West Nile Virus incidence concentration in Chicago
Spray Locs 2011
Spray Locs 2013
Trap Location
Weather Station
Evanston
Latitude
Longitude
42.1
42.0
41.9
41.8
41.7
41.6
−88.0
−87.9
−87.8
−87.7
−87.6
−87.5


Allstate Arena
EDISON PARK
Rivers Casino Des Plaines
Rosemont
NORWOOD PARK
Catherine Chevalier Woods
O'Hare International Airport
Harwood Heights
Norridge
Schiller Park
Bensenville
Schiller Woods South
DUNNING
Franklin Park
MONTCLAR
River Grove
Sky Zone Trampoline Park
Elmwood Park
Northlake
Stone Park
Melrose Park
Elmhurst
River Forest

Cost-Benefit Analysis

- Spraying not really worth it
- USD ~500 to spray 2000 m^2
- O’hare Airport would cost USD ~6,525,000
- Killing adult mosquitoes is not really effective in reducing WNV incidences
    - Instead, larvae should be targetted

Searching searching searching

model	train_score	test_score	accuracy	misclassification	precision	recall	specificity
Logistic Regression	0.643319	0.649525	63.42	36.58	9.14	0.6667	0.6324
Decision Tree Classifier	0.729349	0.730546	72.93	27.07	13.03	0.7319	0.7292
Random Forest Classifier	0.867910	0.603166	86.41	13.59	14.10	0.3116	0.8947
AdaBoost Classifier	0.746098	0.732541	74.61	25.39	13.62	0.7174	0.7477
Gradient Boosting Classifier	0.825276	0.699034	82.53	17.47	16.21	0.5580	0.8401
XG Boost Classifier	0.768177	0.737347	76.82	23.18	14.59	0.7029	0.7718

Modelling

Model	AUC value for ROC curve	Kaggle private score	Kaggle public score
Logistic Regression	0.724	0.6095	0.6360
Decision Tree	0.788	0.6212	0.6544
Random Forest	0.770	0.6150	0.6432
ADA Boost	0.797	0.6586	0.6776
Gradient Boost	0.801	0.6540	0.6809
XG Boost	0.807	0.6737	0.7045

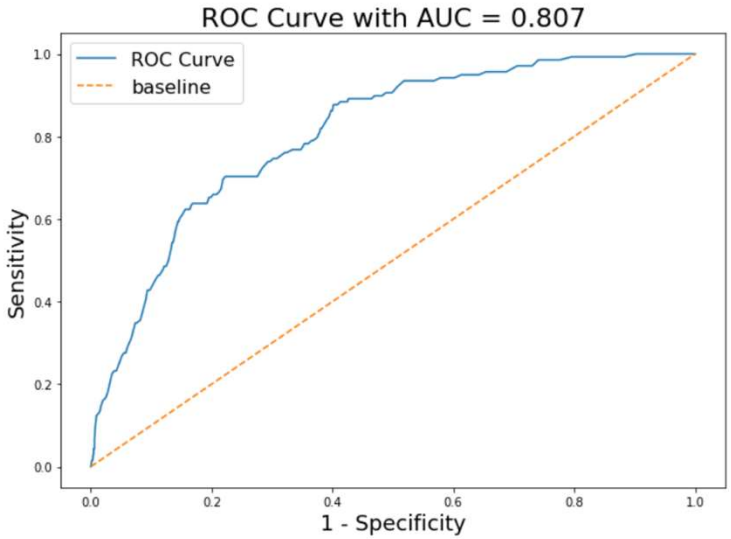
Selected model - XG Boost Classifier

```
    ##########################
    XG Boost Classifier model
    ##########################

Best params: {'learning_rate': 0.1, 'max_depth': 3, 'n_estimators': 100}
Train score: 0.7681766273315569, Test score 0.73734722634665
```

	actual WnV Not Present	actual WnV Present
Predict WnV Not Present	1921	568
Predict WnV Present	41	97

```
** Accuracy %: 76.82% **
Misclassification %: 23.18&
Precision %: 0.1459
Recall: 0.7029
Specificity: 0.7718
```



Conclusion

- We would recommend against spraying on a large scale
  - Maybe areas with really high density of mosquitoes should be sprayed though
  - Instead, educate population on how to prevent mosquito breeding
- O'hare Airport and River Grove areas should be focused on
- Maybe usage of weather forecasts might be useful

Limitations

- No bird data (birds are WNV carriers as well)

Further work

- Try out neural network (possible increase in performance)