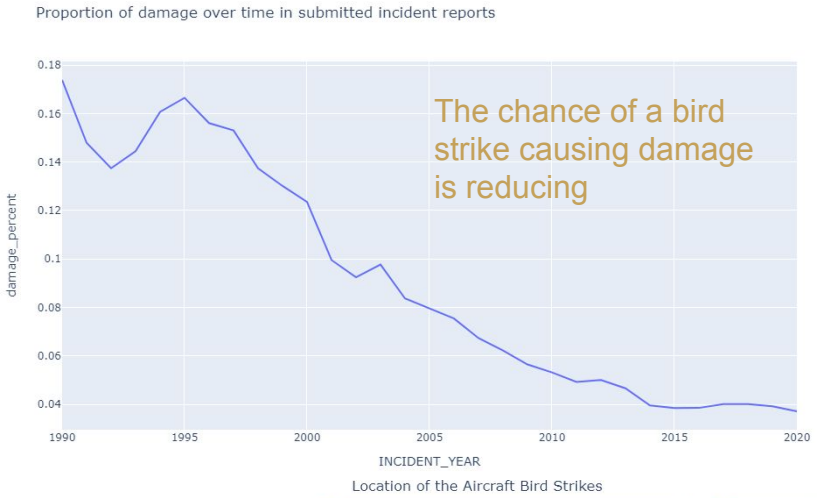
Network Airlines - Bird Strike Model

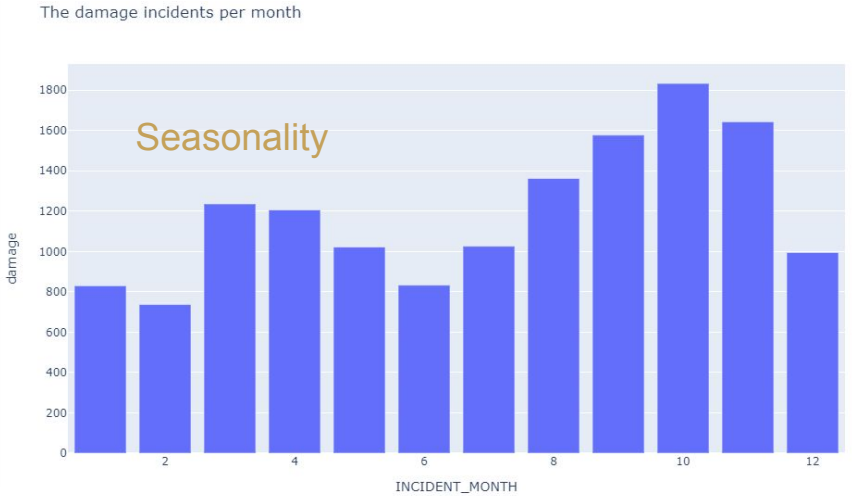
Anthony Wynne 18/01/2022

Network Airlines - Bird Strike Model

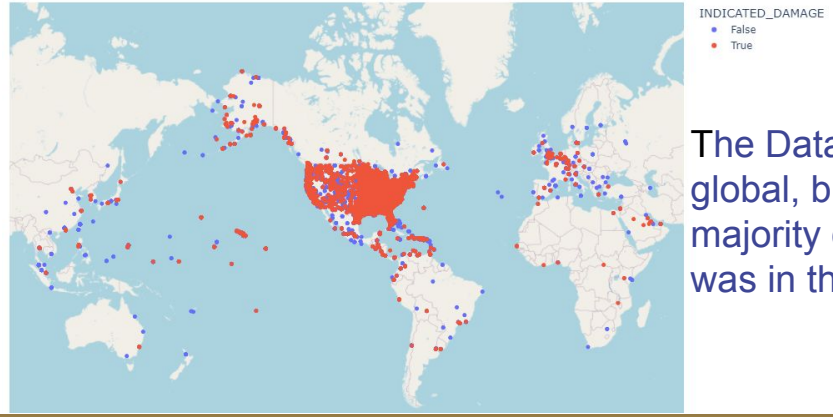
The Data

Anthony Wynne 18/01/2022





Most of the Enroute data was dropped from the model as it lacked clear location data

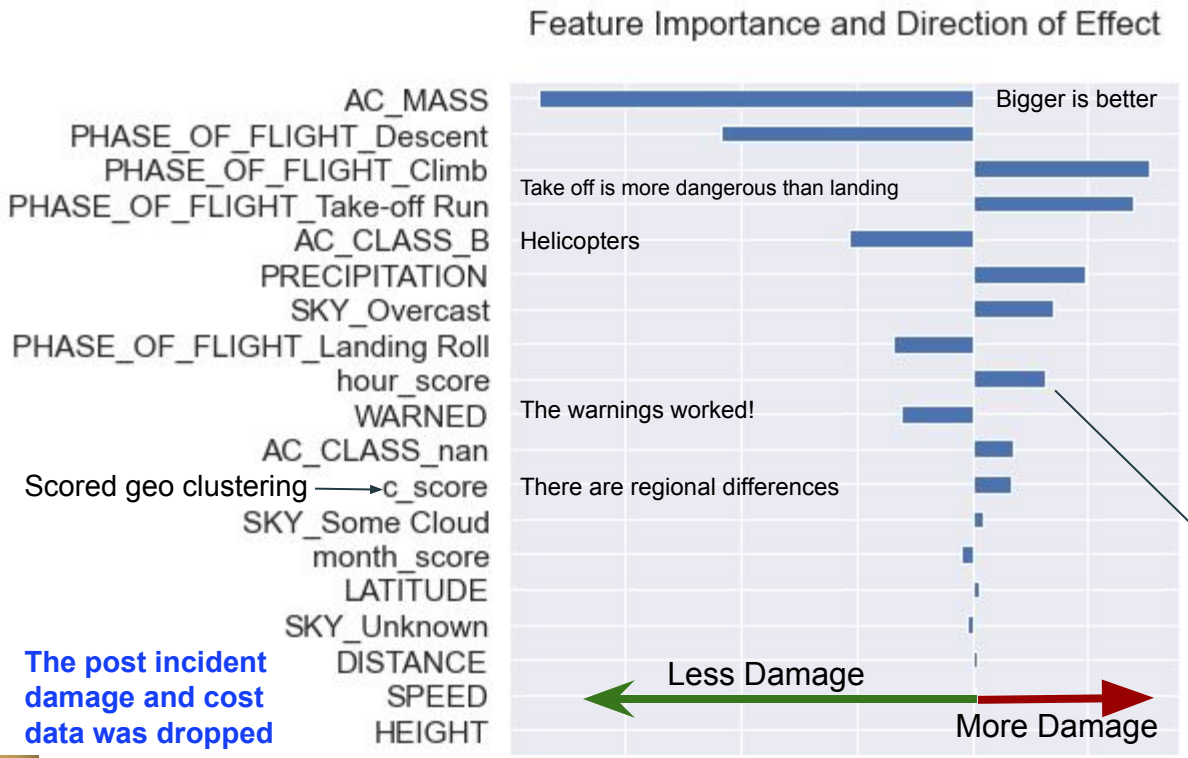


The Data was global, but the majority of it was in the US

The data set was a collection of bird strike reports, so a collision had occurred in all cases. The data was used to predict the chance of damage occurring from a bird strike.

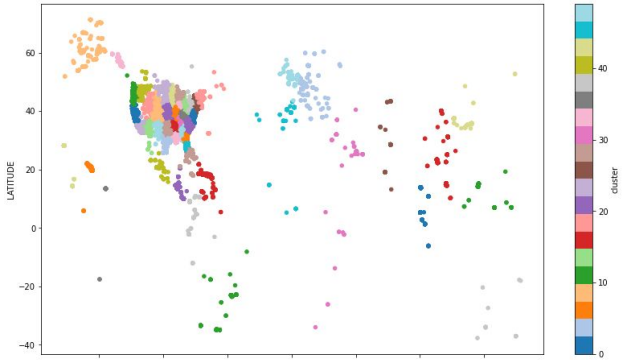
The Features

Anthony Wynne 18/01/2022



The post incident damage and cost data was dropped

Longitude and Latitude were clustered into 50 zones



Time was important but there was not enough quality data to use it in the final model

The rate of damage rate is low at 7% of the incident reports

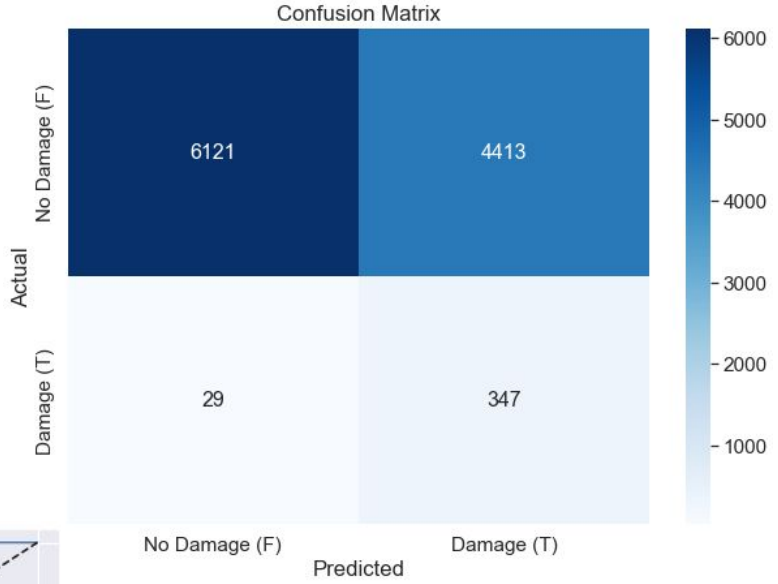
Network Airlines - Bird Strike Model

The Model

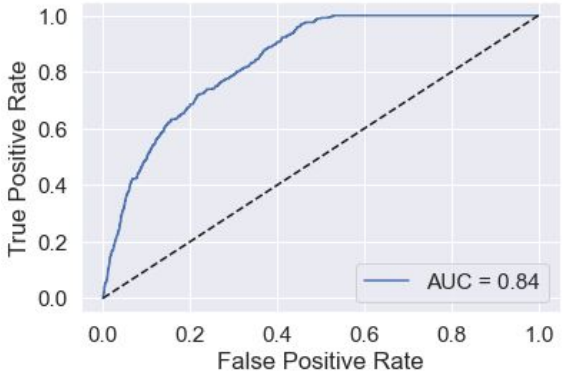
Anthony Wynne 18/01/2022

The final model developed used the following data to predict the indicated damage field using a Logistic Regression model:

- The Aircraft Mass
- The Aircraft Class
- The Phase of flight
- Precipitation
- Sky - Cloud cover
- The Month
- Warned
- And clustered zones derived from the longitude and latitude



2021 data was removed from the data set, and used as a final validation test. The results are shown here. The model was biased towards a low rate of false negative predictions at the cost of more false positives and overall accuracy. I felt this would protect against future damage.



	precision	recall	f1-score	support
False	1.00	0.58	0.73	10534
True	0.07	0.92	0.14	376
accuracy			0.59	10910
macro avg	0.53	0.75	0.43	10910
weighted avg	0.96	0.59	0.71	10910

Next Steps:

- Use a Gradient Boosted Tree or Random Forest
- Try using more Geo Clusters
- Try and improve the time data
- Predict the severity and cost of the damage
- Use imbalance-learn SMOTE sampling