# Identifying Trends of Laser Attacks on Aircraft

Applied Data Science Capstone Project

James Brambley

June 2020

james.brambley@gmail.com

Table of Contents

**Abstract**

There is a growing trend in the number of laser attacks on aircraft in the USA. In the 9 months to October 2016 there were 5,352 attacks [1], up from 2,837 for all of 2010. These attacks pose a direct threat to flight safety and can lead to large fines or imprisonment for the offender.

This report aims to analyse laser attack data from one year in the USA to identify trends in the data and to use machine learning models to predict future attacks. This could help law enforcement agencies to target resources on the ground and to design safety policy for airspace users in identifying the locations and time of highest risk.

**Introduction**

Laser attacks on aircraft have been steadily increasing since the early 1990s[2]. These pose a direct threat to flight safety as pilots can be blinded or distracted during critical phases of flight such as takeoff or landing. As well as being a safety issue there is also an economic one – flights have been forced to divert as pilots are left with potentially lasting damage making it unsafe to continue the flight to destination.[3]

There are some measures that can be used to mitigate the safety threat, including goggles or protective film on the aircraft windshield but these are generally impractical for long term use[4]. It would therefore be beneficial to raise awareness of the areas of highest risk of a laser attack so that pilots can be aware and plan accordingly. Commercial pilots tend to perform take off and approach briefings where they identify potential threats and how to avoid them. An occasional refresher if you were to operate in these higher risk environments as to the best strategies to deal with a laser attack could lead to increased overall safety margins.

This report analyses laser attack data provided by the Federal Aviation Administration over 2014[5] to identify the major variables that increase the risk of an attack. These include time, location, day of the week and altitude.

[1] https://eu.usatoday.com/story/news/2015/11/12/laser-strikes-aircraft-record-pace-another-flurry-overnight/75649430/

[2] A sampling of NASA ASRS laser incident reports can be done by searching for the term "laser".

[3] "Virgin Atlantic laser incident: pilots' union demands action"

[4] Sample page provided by Google Book Search, from *Weapons of Mass Casualties and Terrorism Response* by Charles Edward Stewart, Jones & Bartlett Publishers, 2006.

[5] https://www.faa.gov/about/initiatives/lasers/laws/

## Data

The primary source of data for this investigation is the history of laser attacks provided by the Federal Aviation Administration (FAA) in the USA.[6]

This lists all attacks that are registered with the agency over each year, saved into an Excel spreadsheet. Fields are Date, Time, Aircraft Registration, Aircraft Type, Altitude, Laser Colour, Injury Reported, City and State.

This data was reviewed and imported into a Jupyter Notebook using Python. There were fewer than ten reports that had missing data and these were removed from the dataframe as being immaterial for the analysis of nearly 4,000 reports.

| DATE | TIME | Aircraft ID | No. A/C | TYPE A/C | ALT | MAJOR CITY | COLOR | Injury Reported | CITY | STATE |
|---|---|---|---|---|---|---|---|---|---|---|
| 2014-01-01 | 0:08 | CHQ6045 | 1 | E145 | 8000 | CAE | Green | NO | Columbia | South Carolina |
| 2014-01-01 | 0:35 | PSA | 1 | CRJ2 | 2100 | PHL | Green | NO | Philadelphia | Pennsylvania |
| 2014-01-01 | 2:02 | SWA563 | 1 | B737 | 4200 | PHX | Green | NO | Phoenix | Arizona |
| 2014-01-01 | 2:35 | RPA3242 | 1 | E175 | FL240 | EWR | Green | NO | Newark | New Jersey |
| 2014-01-01 | 2:59 | AAL1153 | 1 | B738 | 12000 | POM | Green | NO | Pomona | California |
| 2014-01-01 | 3:00 | AWE1954 | 1 | A321 | 2000 | LGA | UNKN | NO | New York | New York |
| 2014-01-01 | 3:02 | SWA549 | 1 | B737 | FL260 | SBA | Green | NO | Santa Barbara | California |
| 2014-01-01 | 3:07 | SKW2617 | 1 | CRJ2 | 4000 | SMO | Green | NO | Santa Monica | California |

## Methodology

The data was extracted and counts of attacks were made by comparison to Date, Hour (UTC), Altitude and State. The latitude and longitude of each City were calculated and plotted on a map of the USA to identify hotspot locations using k-means clustering.

After reviewing the data it was decided to use UTC Hour and State as the dependent variables and the Count of Attacks was the target variable. Using this method a pilot could see where they may be taking off or landing and the time and this could output a risk factor for a laser attack.

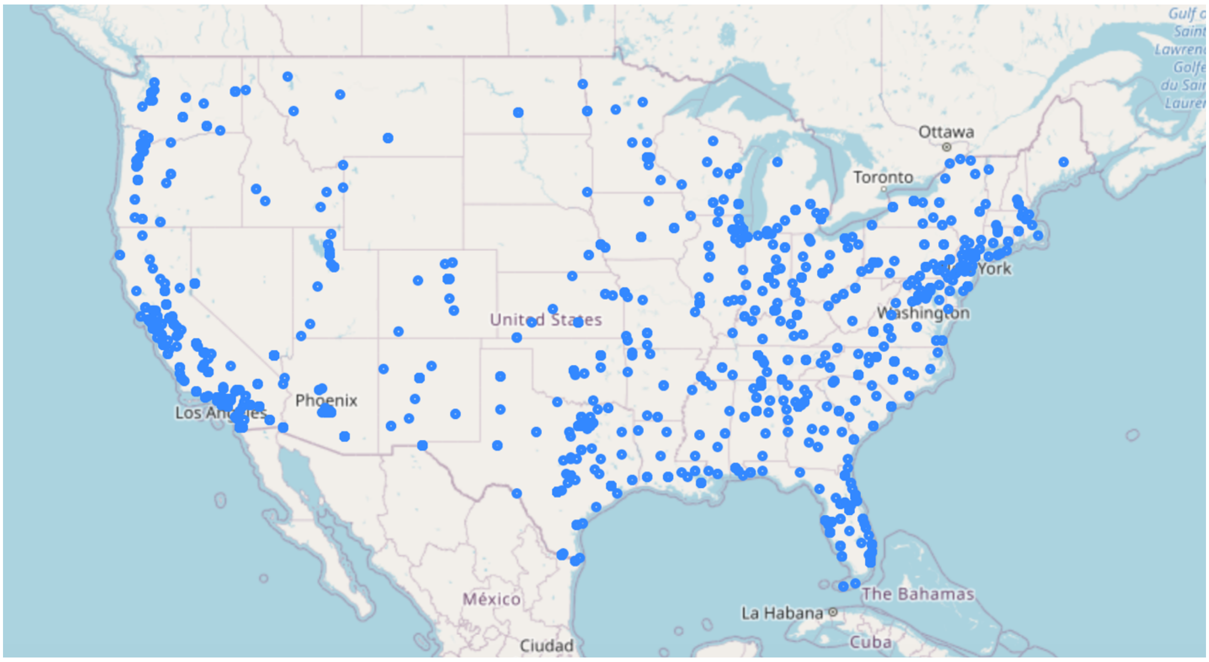
The sklearn package for K Neighbor Regressors was used to model the training data and compared with the test data. The intent is for this regressor model to be continually updated as fresh reports are added to the database to train the model further over time.

[6] https://www.faa.gov/about/initiatives/lasers/laws/

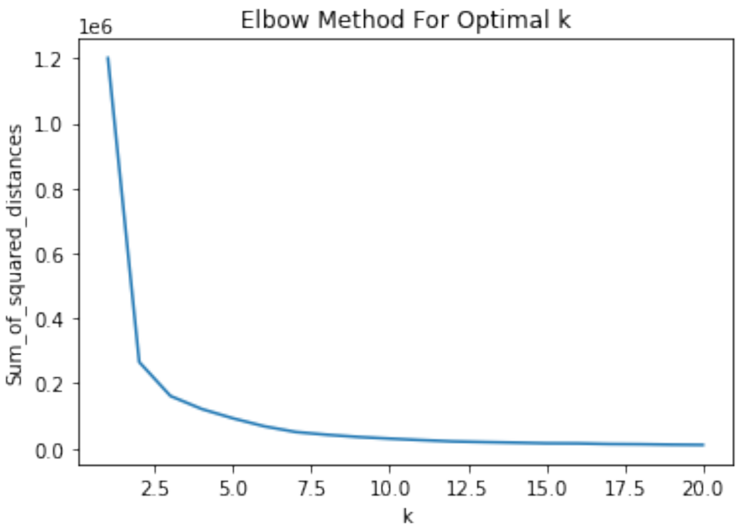
**Analysis**

The initial data had some reports from outside the continental 48 states of the USA which skewed the clusters away from sensible positions so the data was cleaned to remove these.

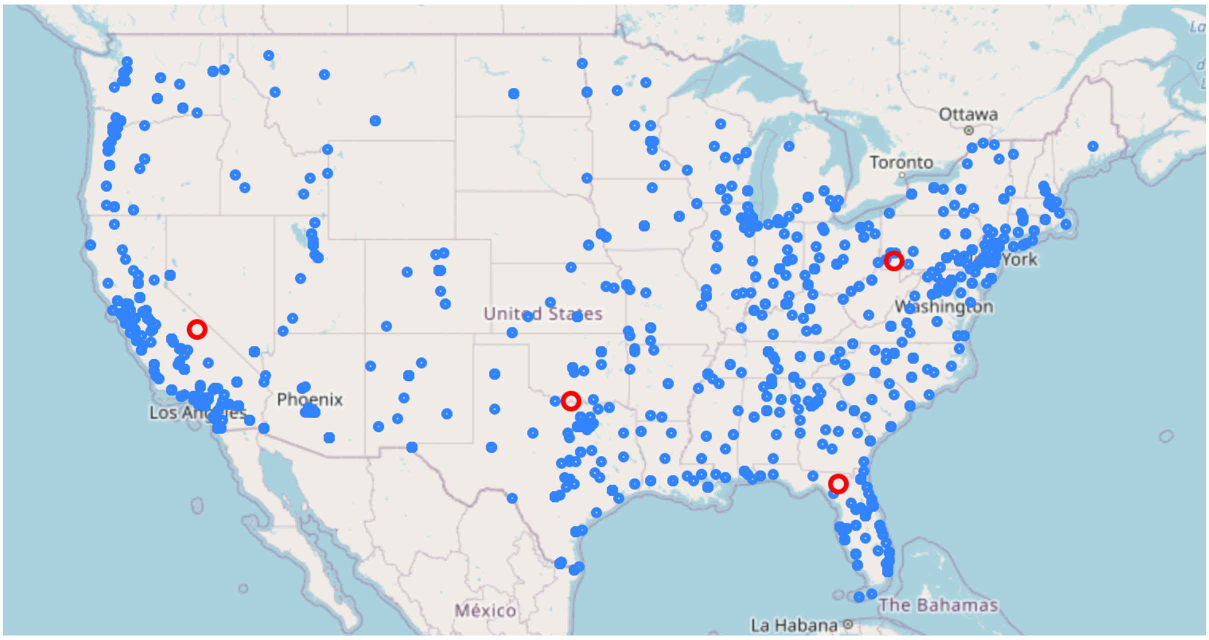
This is the visualization of 2014 attacks:




The elbow method of k-means clustering was used to find the optimal k:

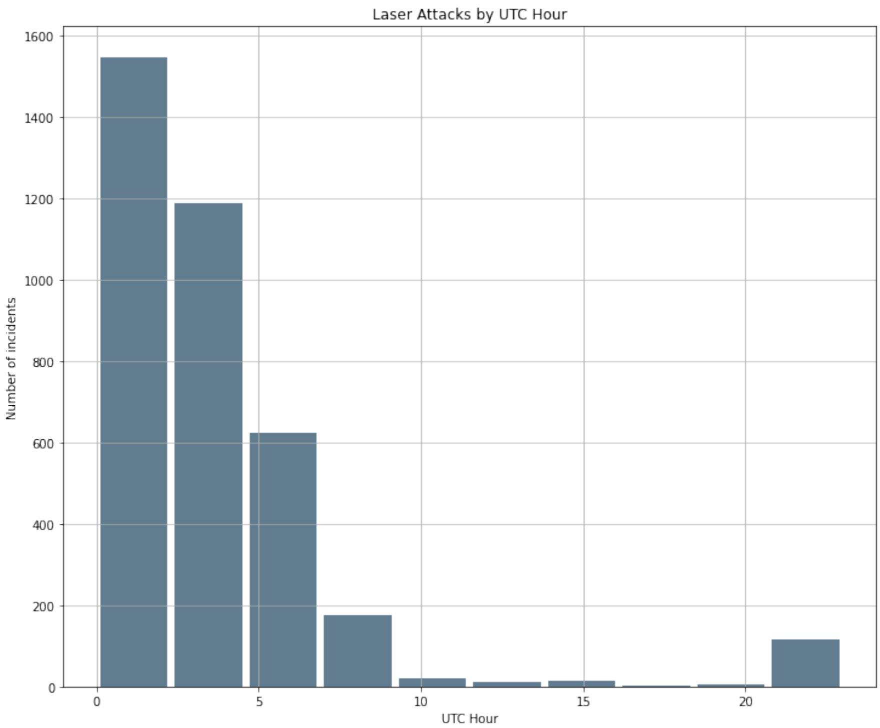


A k of 4 was selected and the clusters were replotted onto the map:

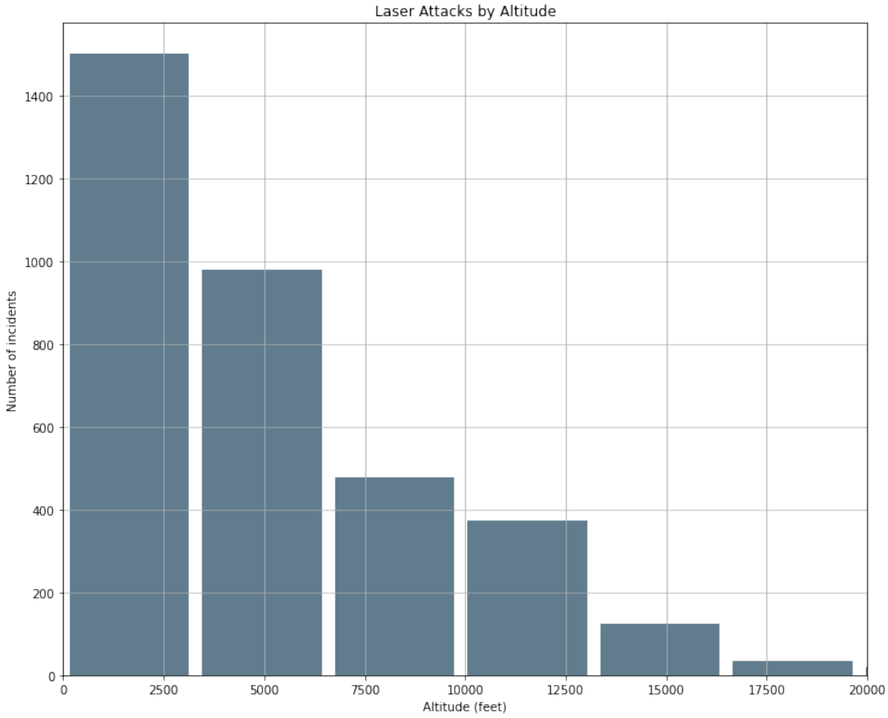



It is clear that there are several hotspot areas around California, New York, Florida and Texas.

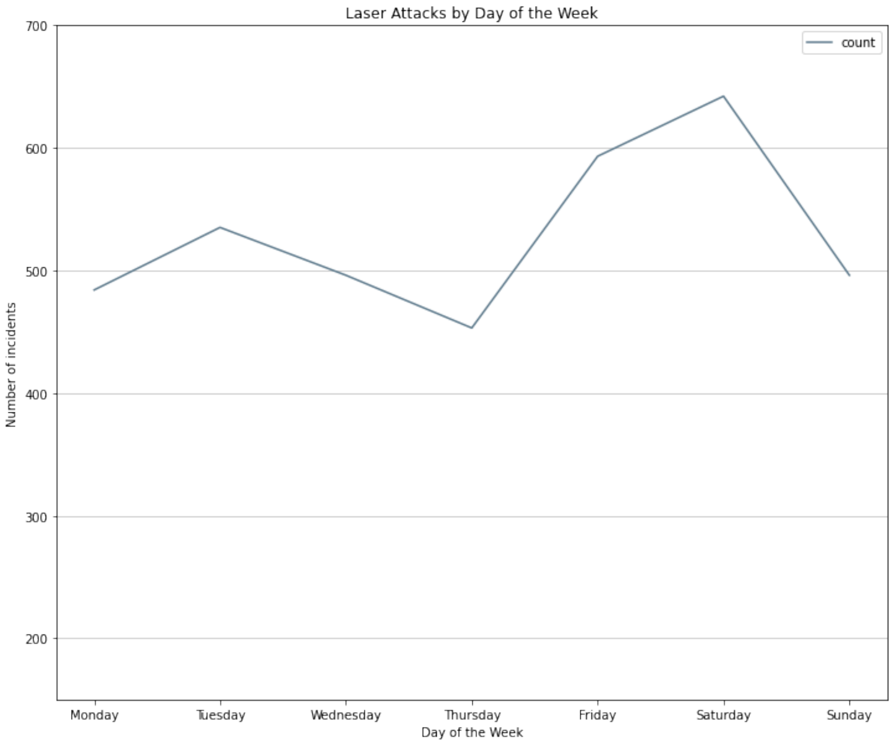
Some analysis of other factors were then performed to see if they had a material impact on the number of attacks.



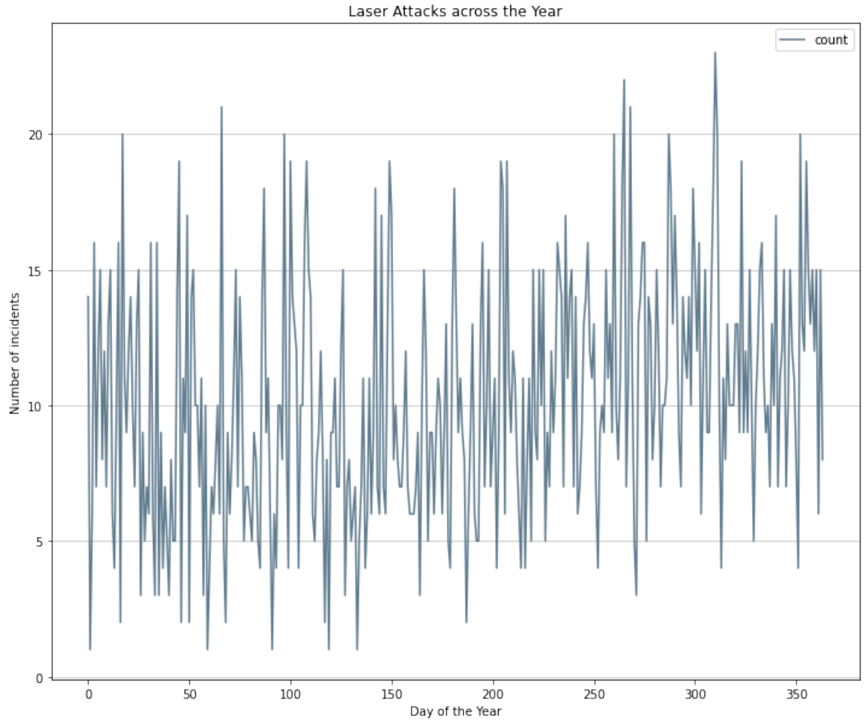

There is definitely a peak in the range of 00:00 – 06:00 UTC which is entirely logical as this translates to evening local time in the USA.



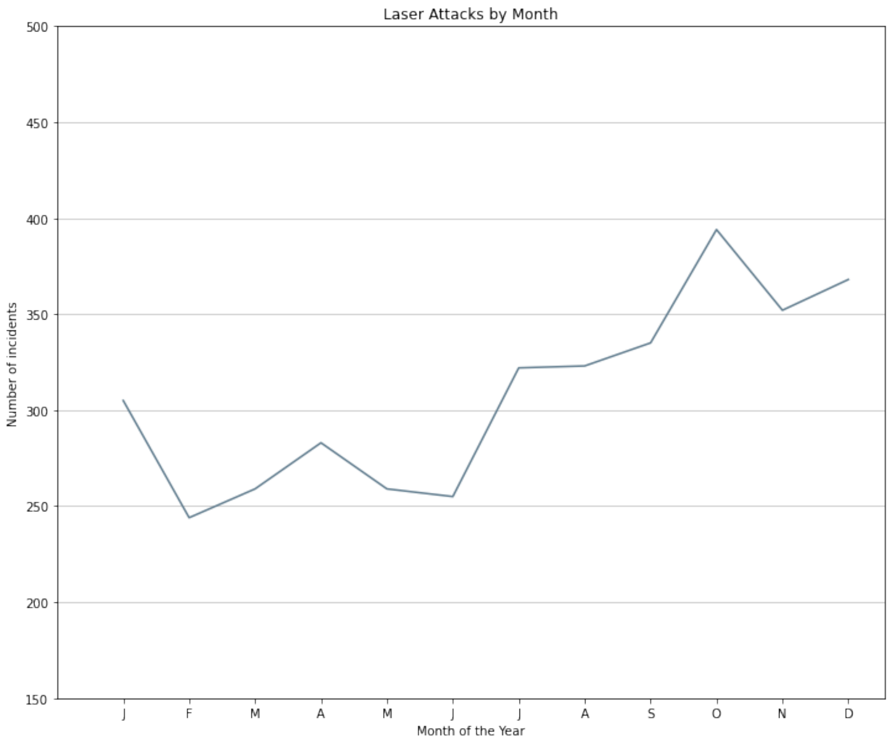

This chart shows that the risk is highest at low altitude – indicating that the most relevant part of the risk is during take off and landing. The enroute phase sees virtually no attacks at all.



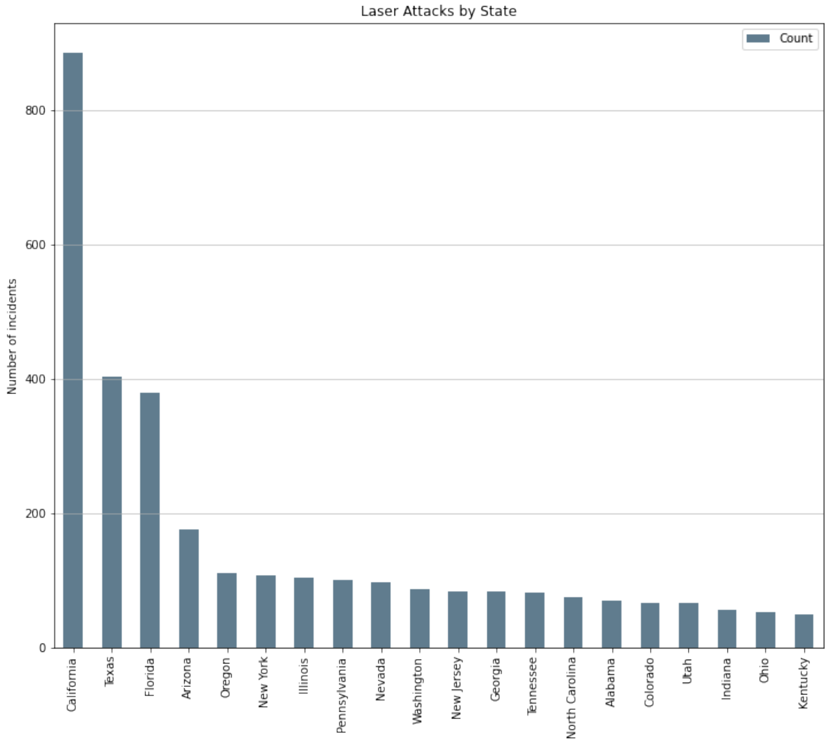

There is a reasonable spread of attacks across the week, with a very small peak on Fridays and Saturdays.




There is perhaps a very gentle increase across the year, it is perhaps more easily viewed when the data is grouped by month.




There is a gentle increase across the year and this ties in with the knowledge that attacks have been increasing year on year.




It is clear that California has by far the largest number of laser attacks, which links back to the clustering map we produced earlier.

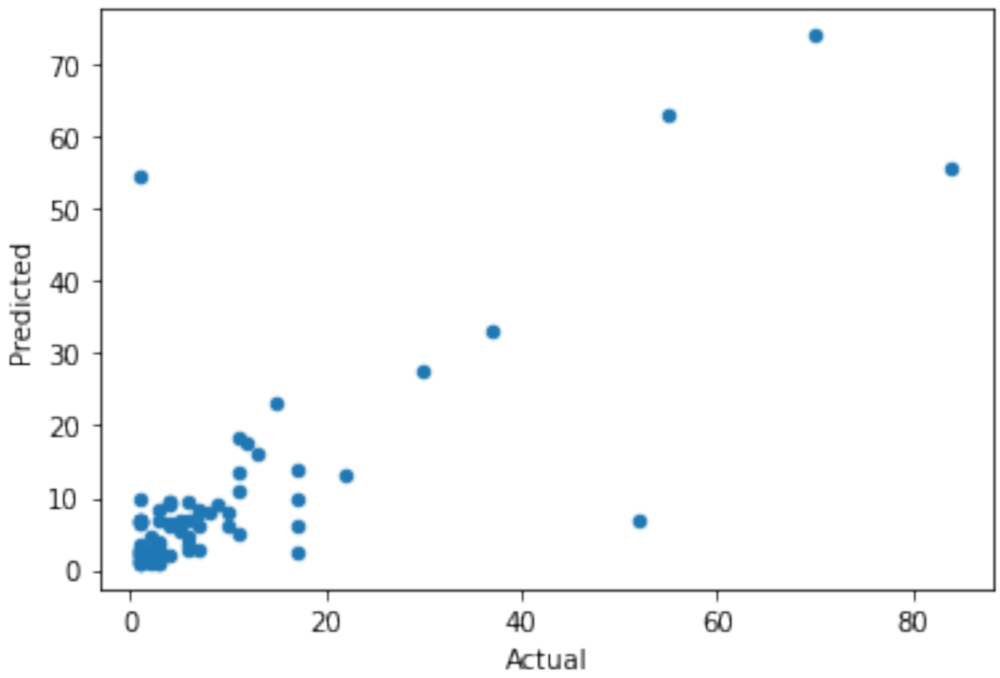
If the altitude variable is considered by only including take off and landing location instead of the cruise routing, the variables that have the largest influence on the number of attacks are Time and State. These were combined together to produce a list of higher risk combinations.

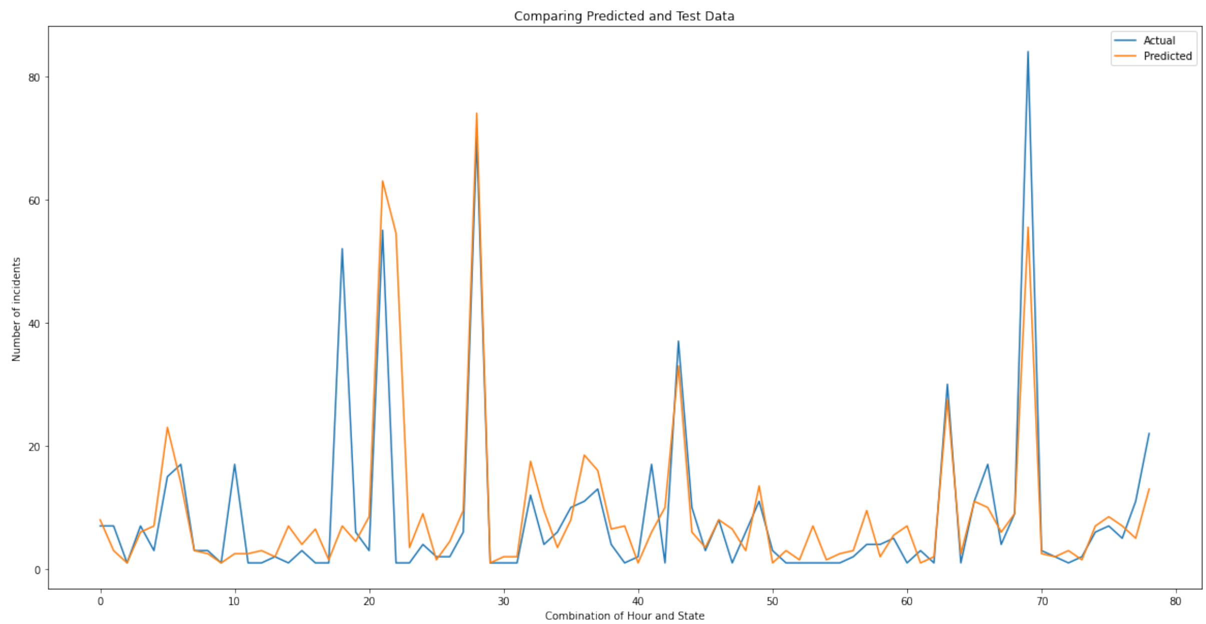
| hour | STATE | Count |
|---|---|---|
| 4 | California | 212 |
| 3 | California | 162 |
| 5 | California | 161 |
| 6 | California | 128 |
| 2 | Texas | 108 |
| 2 | Florida | 97 |
| 1 | Texas | 89 |
| 2 | California | 87 |
| 1 | Florida | 84 |
| 3 | Texas | 70 |
| 3 | Florida | 55 |
| 7 | California | 53 |
| 0 | Florida | 52 |
| 4 | Texas | 40 |
| 1 | California | 39 |

This data could then be usefully categorised into risk factors to enable awareness during the takeoff and approach briefings.

RED – 03:00 <–> 06:00 UTC in California
AMBER – 01:00 <–> 03:00 UTC in Texas and Florida, 02:00 and 07:00 in California.
GREEN – Any other time and location.

To assist with future predictions it was decided to create a KNN Regressor using the 2014 data. A k of 1 gave the highest correlation between the training and test data using a 30% test split.



The model seems to cope quite well with the time and state combinations provided in predicting number of attacks. This could be continually updated each month as further data arrives.

```
Correlation is: 0.78433804448034
R^2 value: 0.7843380444803401
P Value: 1.2346883252204964e-17
Standard Error: 0.0643753361598501
```

The predicted and actual values have an R^2 value of 0.78 which indicates a strong relationship.

**Results**

We investigated the effect of all the reported variables on the likelihood of a laser attack. Most attacks were found to be between 00:00 and 05:00 UTC in California with other areas in Florida and Texas showing a medium risk.

This information could prove genuinely useful to pilots in assessing the risks of a laser attack and allowing them to prebrief their response.

A Nearest Neighbor model was used to help predict future occurrences and this could prove useful in directing resources to prevention and deterrence.

**Conclusion and Further Study**

There are certain pockets of higher risk combinations of time and location for laser attack.

Given that California has the highest incidence of attack across the country it may prove useful to concentrate analysis in that area in future and assess if any prevention measures are currently being used and their effectiveness.

It would also be useful to measure the rate per takeoff and landing and combine this into the risk model.