Predicting Survival Outcomes of Helicopter Accidents

Bronwyn Middleton May 2019

Background

Helicopter Accidents usually make the news due to their severity

New Zealand helicopter crash: two Australians and four Britons among seven killed

Leicester City owner among five dead in helicopter crash

Britons killed in Grand Canyon helicopter crash named

Success Metrics

- To explore the dataset and find interesting facts
- To identify trends
- To discover whether predictions can be made about the severity of an accident given circumstances of the accident

Data Source

National Transportation Safety Board

Government agency

https://www.ntsb.gov/_layouts/ntsb.aviation/index.aspx

Exploratory Data Analysis

gitude	Airport Code	Airport Name	Injury Severity	Aircraft Damage	Aircraft Category	Registration Number	Mak
911944		N/A	Non-Fatal	Substantial	Helicopter	N593C	M HELICOPTE
979723	FUL	Fullerton Muni	Unavailable	Substantia		N9964L	Cessn
831389	MMK	Meriden Markham Muni	Non-Fatal	Substantial	Airplane	N38658	Pipe
765000		N/A	Fatal(1)	Destroyed	Airplane	N7314D	Air Tracto
411667	T67	HICKS AIRFIELD	Non-Fatal	Substantial	Airplane	N801R	Zena

BUT df.isna.sum() = result 0

Exploratory Data Analysis cont.

df.replace({' ': np.nan}, inplace = True)

event_id	0
investigation_type	3
accident_number	0
event_date	0
location	77
country	507
latitude	53929
longitude	53938
airport_code	35861
airport_name	30462
injury_severity	0
aircraft_damage	2622
aircraft_category	56737
registration_number	3633
make	71
model	99
amateur_built	635
number_of_engines	4670
engine_type	4004
far_description	57070
schedule	71221
purpose_of_flight	4613
air_carrier	78865
total_fatal_injuries	26214
total_serious_injuries	28762
total_minor_injuries	27568
total_uninjured	13968
weather_condition	2791
broad_phase_of_flight	6550
report_status	0
publication_date	14073

Exploratory Data Analysis cont.

Approach

- Drop columns that don't add value

- Subset dataset
  - Aircraft_category = Helicopters (2781 rows)
  - Investigation_type = Accidents (2701 rows)

https://en.wikipedia.org/wiki/Aviation_accidents_and_incidents

~~event_id~~	~~0~~
~~investigation_type~~	~~3~~
~~accident_number~~	~~0~~
event_date	0
location	77
country	507
latitude	53929
longitude	53938
~~airport_code~~	~~35861~~
~~airport_name~~	~~30462~~
injury_severity	0
~~aircraft_damage~~	~~2622~~
~~aircraft_category~~	~~56737~~
~~registration_number~~	~~3633~~
make	71
model	99
amateur_built	635
number_of_engines	4670
engine_type	4004
~~far_description~~	~~57070~~
~~schedule~~	~~71221~~
purpose_of_flight	4613
~~air_carrier~~	~~78865~~
total_fatal_injuries	26214
total_serious_injuries	28762
total_minor_injuries	27568
total_uninjured	13968
weather_condition	2791
broad_phase_of_flight	6550
~~report_status~~	~~0~~
~~publication_date~~	~~14073~~

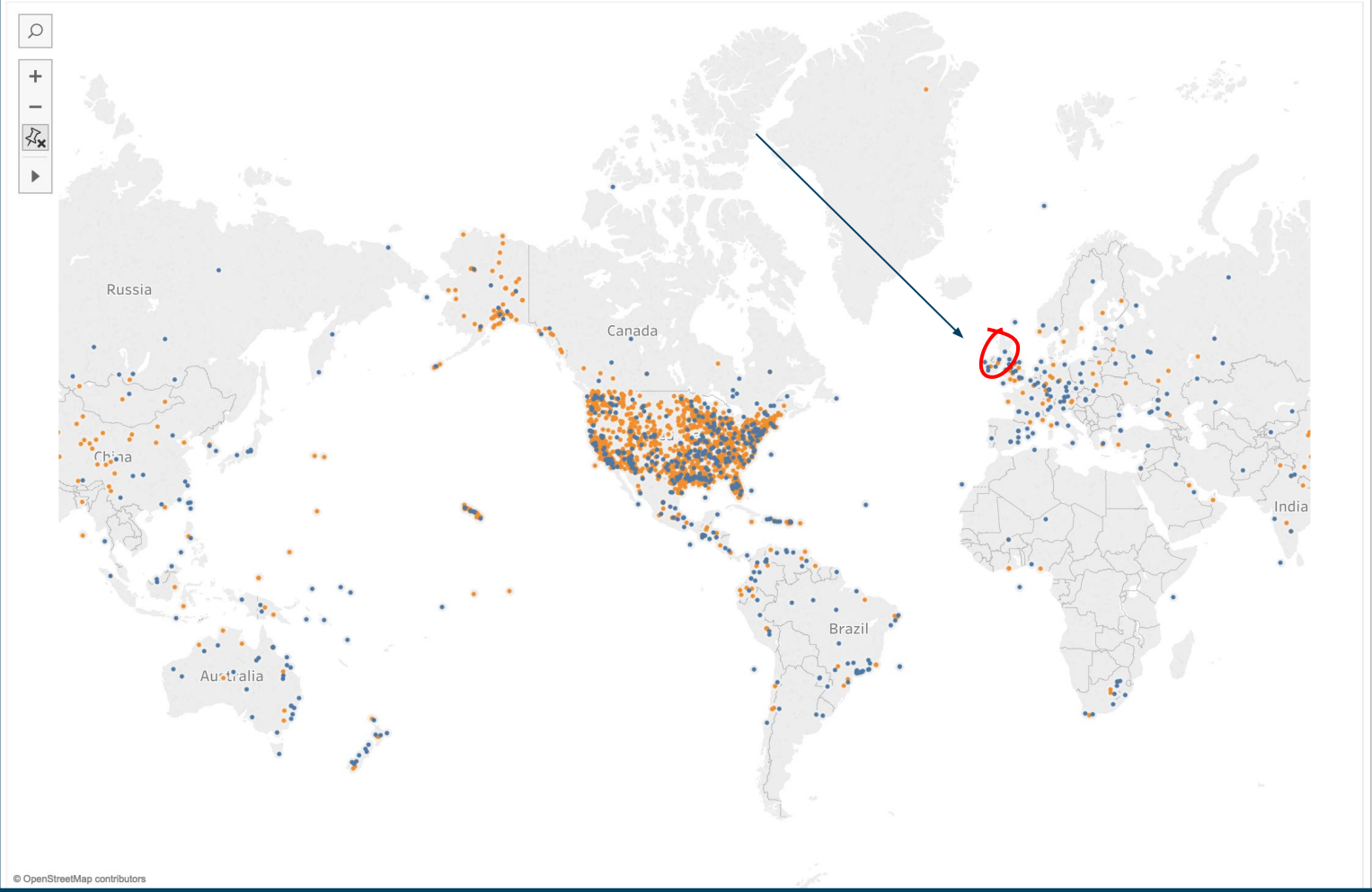
Exploratory Data Analysis cont.

location	6
country	13
latitude	583
longitude	584

```
geolocator = Nominatim()
city ="McCool Junction, Nebraska"
country ="united states"
loc = geolocator.geocode(city+','+ country)
print("latitude is :" ,loc.latitude,"\nlongtitude is:" ,loc.longitude)
```

```
latitude is : 40.7442155
longtitude is: -97.5935135
```

Accidents using Longitude, Latitude and Fatalities columns



Fatalities
- fatal
- non_fatal

Exploratory Data Analysis cont.

2/1/1982 - 16/4/2019

NTSB - 22 reports for UK helicopter accidents

AAIB - 1002 reports for UK helicopter accidents

https://www.gov.uk/aaib-reports?parent=&keywords=&aircraft_category%5B%5D=commercial-rotorcraft&aircraft_category%5B%5D=general-aviation-rotorcraft&date_of_occurrence%5Bfrom%5D=1%2F1%2F1982&date_of_occurrence%5Bto%5D=

Rest of the World vs US

Rest of the world (2:1)

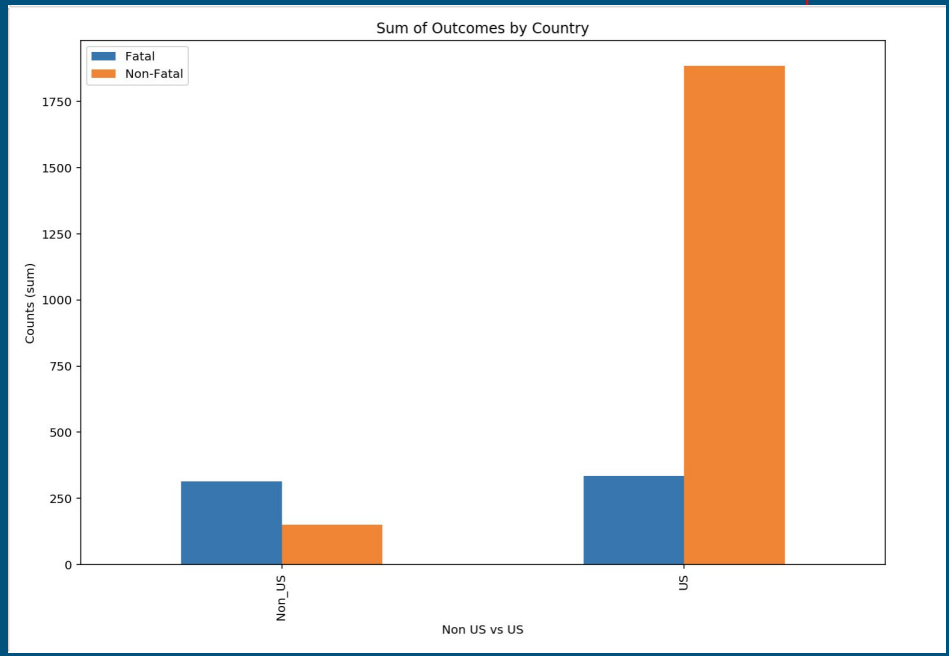
Fatal = 314

Non Fatal = 150

US (1:6)

Fatal = 334

Non Fatal =1886



Cleaning

- White spaces between quote marks
  - ' robinson '
- Inconsistency in naming convention of makes and models

r22_beta r-22_hp r_22_beta r-22hp

r-22 r-22_beta_2 r22_-_alpha

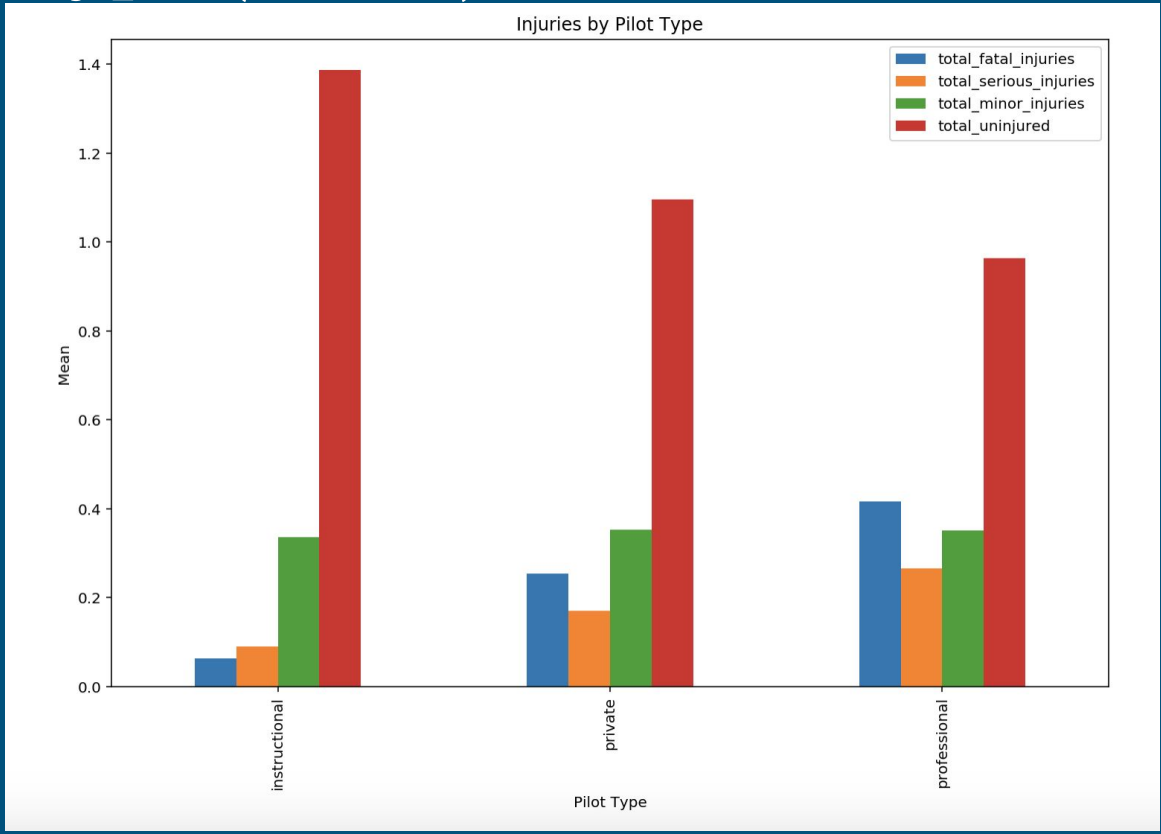
r22_beta_ii r-22_beta_ii r-22a

Feature Engineering

purpose_of_flight	pilot_type
personal or private	private
instructional	instructional
unknown	unknown
All others	professional

Injuries by pilot type (mean)

- Professional pilots fatality rate



Instructional flights

```
len(us_df[(us_df['pilot_type'] == 'instructional') & (us_df['fatalities'] != 'fatal')&(us_df['people_on_board']==2)])
```

```
318
```

```
len(us_df[(us_df['pilot_type'] == 'instructional') & (us_df['fatalities'] != 'fatal')&(us_df['people_on_board']==1)])
```

```
65
```

```
len(us_df[(us_df['pilot_type'] == 'instructional') & (us_df['fatalities'] == 'fatal')&(us_df['people_on_board']==2)])
```

```
10
```

```
len(us_df[(us_df['pilot_type'] == 'instructional') & (us_df['fatalities'] == 'fatal')&(us_df['people_on_board']==1)])
```

```
7
```

Weather

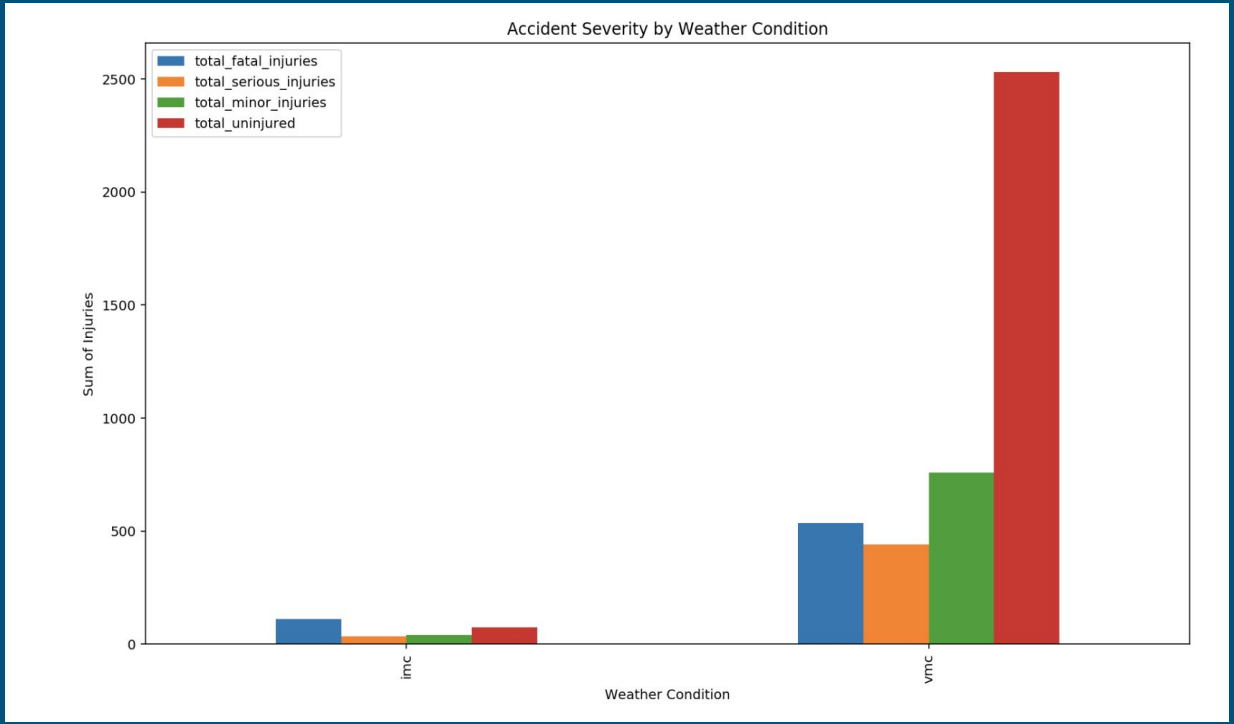
Visual Meteorological Conditions:

VMC minima (day)

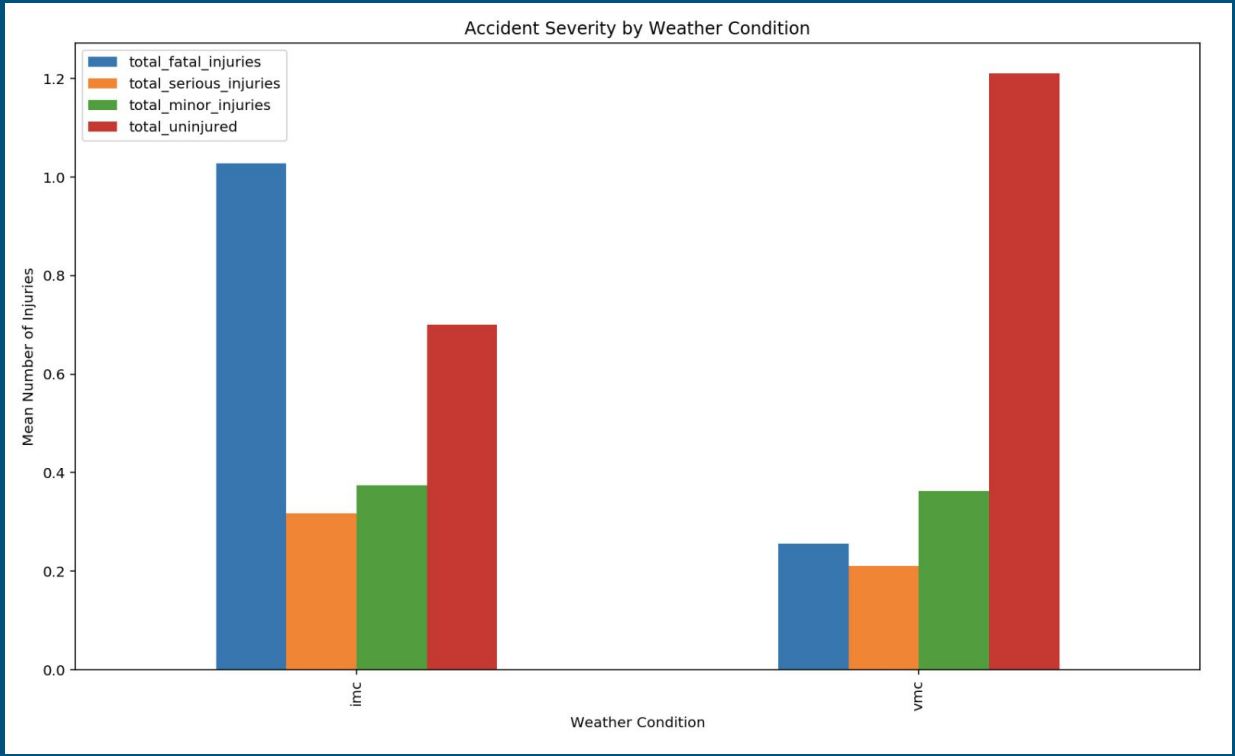
- One mile horizontal visibility
- Clear of clouds
- In sight of the ground

https://en.wikipedia.org/wiki/Visual_meteorological_conditions

Weather Condition results

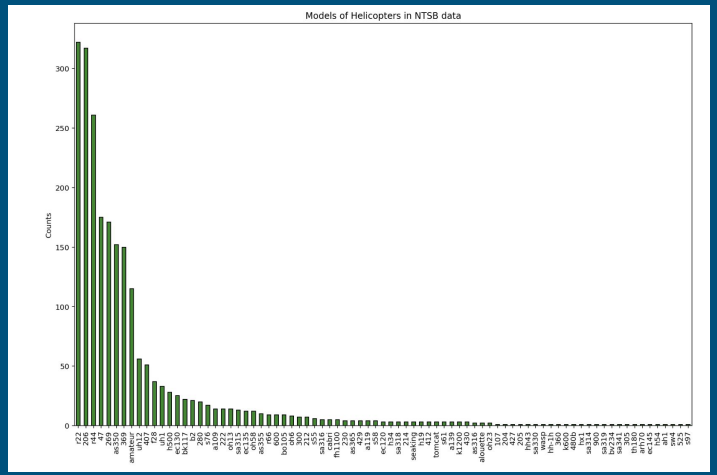


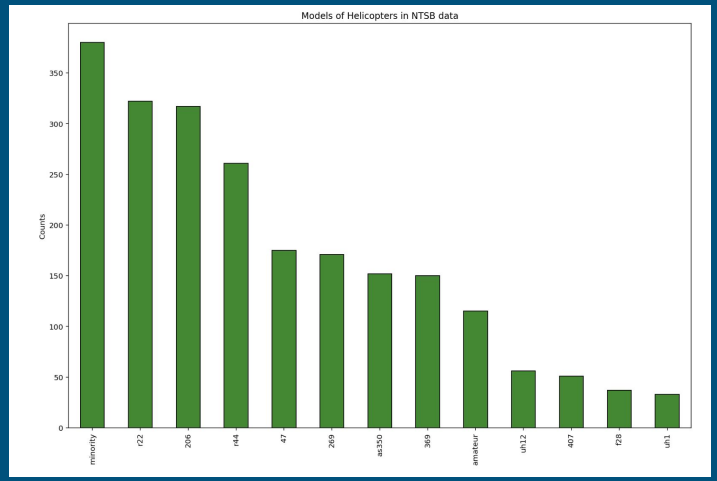
Weather Condition results



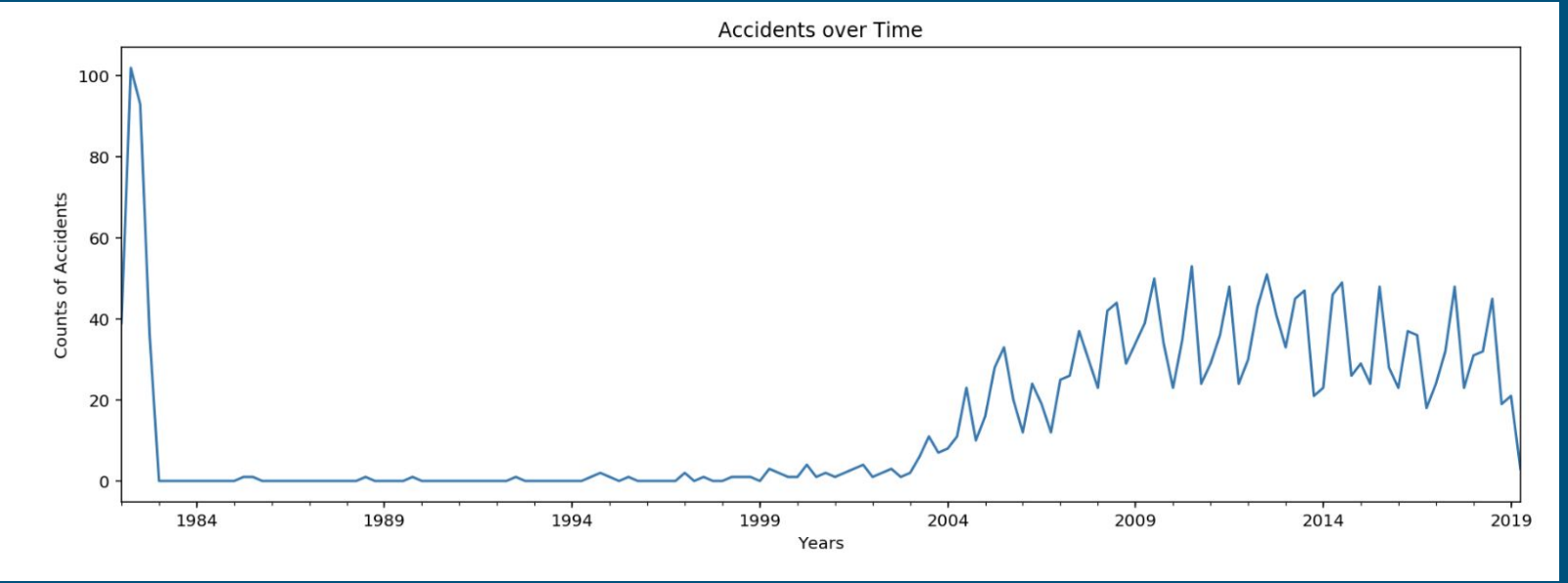
Exploratory Data Analysis (models)

- Several helicopter models
  - <30





EDA - Helicopter Accidents over Time



Modelling

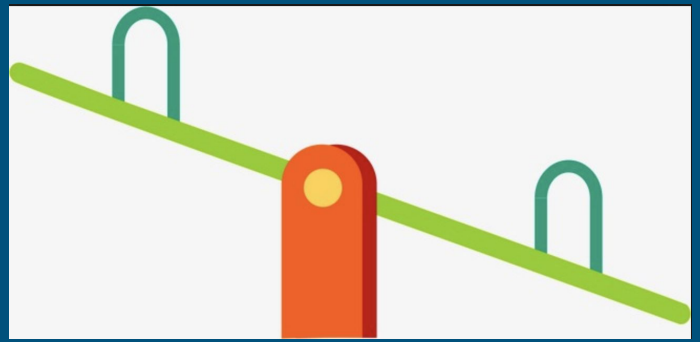
Classification problem:

Logistic Regression

KNN

Decision Tree

Random Forest



Consideration: Class imbalance 272 : 1631

Baseline: 0.8570677877036259

Modelling - Tree Models

Decision Trees

	precision	recall	f1-score
0	0.86	1.00	0.92
1	0.00	0.00	0.00
micro avg	0.86	0.86	0.86
macro avg	0.43	0.50	0.46
weighted avg	0.73	0.86	0.79

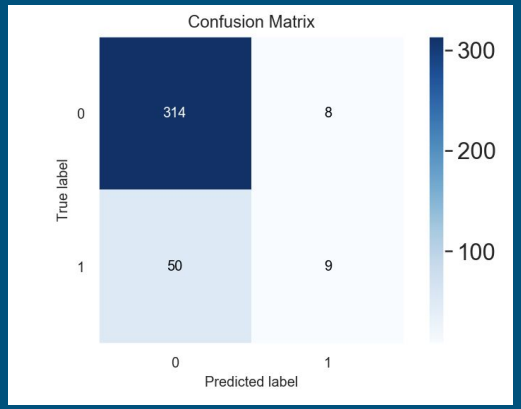
Random Forest

	precision	recall	f1-score
0	0.87	1.00	0.93
1	0.00	0.00	0.00
micro avg	0.87	0.87	0.87
macro avg	0.43	0.50	0.46
weighted avg	0.75	0.87	0.81

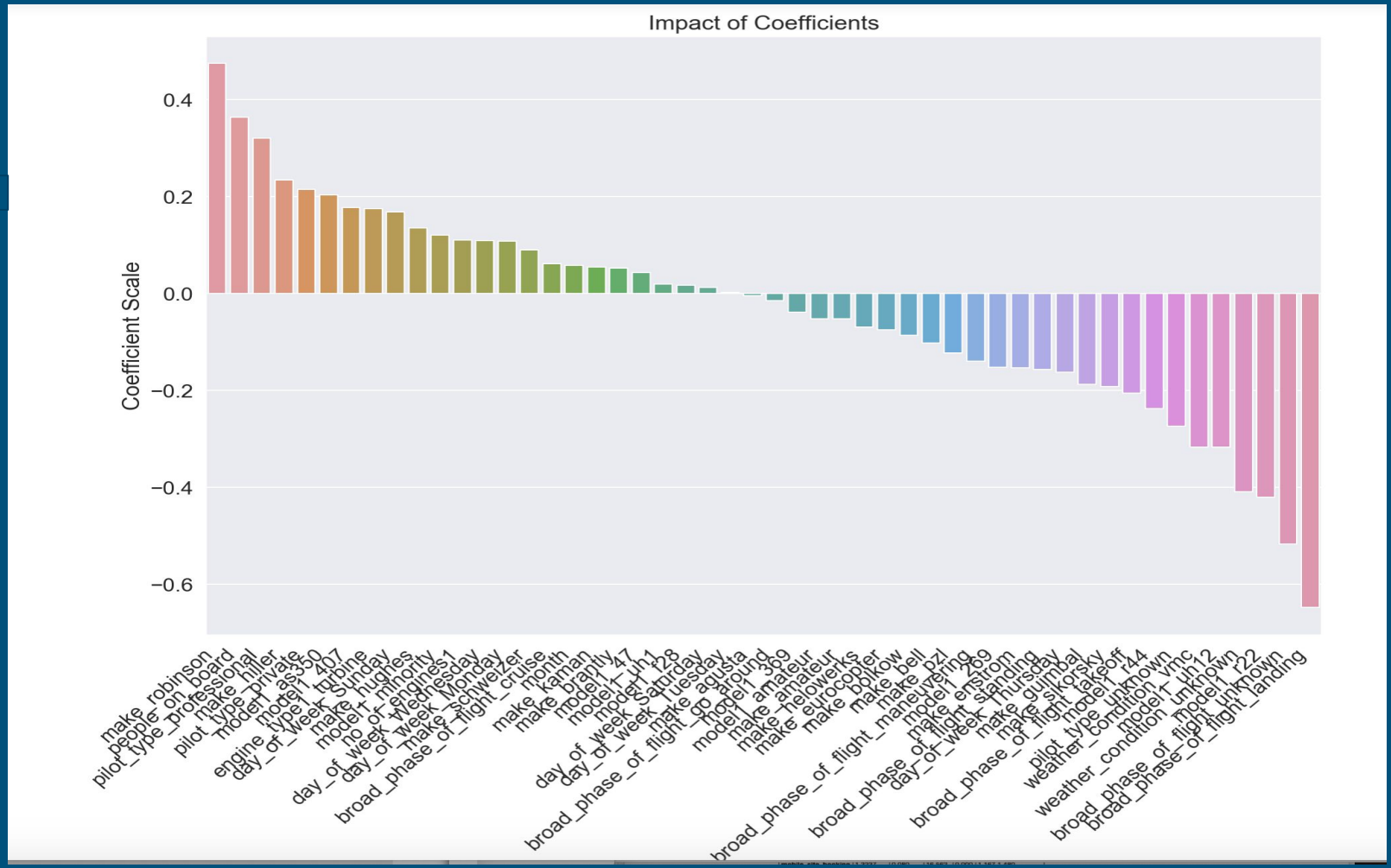
Modelling Logistic Regression (GridSearch)

```
LogisticRegression(C=0.19306977288832497, class_weight=None, dual=False,
          fit_intercept=True, intercept_scaling=1, max_iter=1000,
          multi_class='ovr', n_jobs=None, penalty='l2', random_state=None,
          solver='newton-cg', tol=0.0001, verbose=0, warm_start=False)
```

Gives a test score of 0.850

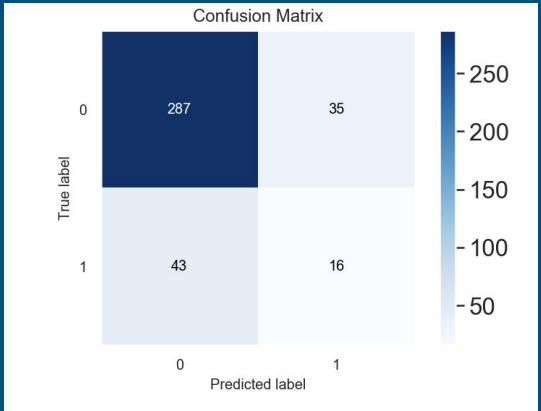


	precision	recall	f1-score
0	0.86	0.98	0.92
1	0.53	0.15	0.24
micro avg	0.85	0.85	0.85
macro avg	0.70	0.56	0.58
weighted avg	0.81	0.85	0.81


Impact of Coefficients
Coefficient Scale
0.4
0.2
0.0
−0.2
−0.4
−0.6
make_robinson
people_on_board
pilot_type_professional
make_hiller
pilot_type_private
model_as350
engine_type_turbine
day_of_week_sunday
make_hughes
no_of_engines_1
day_of_week_wednesday
day_of_week_monday
broad_phase_of_flight_cruise
month
make_kaman
make_brantly
model_47
model_uh1
model_r28
day_of_week_saturday
day_of_week_tuesday
broad_phase_of_flight_go_around
make_agusta
model_369
make_amateur
make_helowerks
make_eurocopter
make_bolkow
make_bell
make_pzl
broad_phase_of_flight_maneuvering
model_269
make_enstrom
day_of_week_thursday
make_standing
make_gulmbal
make_sikorsky
broad_phase_of_flight_takeoff
pilot_type_unknown
model_r44
weather_condition_vmc
model_uh12
weather_condition_unknown
broad_phase_of_flight_unknown
model_r22
broad_phase_of_flight_landing

Modelling KNN

Gives a test score of 0.884



	precision	recall	f1-score
0	0.87	0.89	0.88
1	0.31	0.27	0.29
micro avg	0.80	0.80	0.80
macro avg	0.59	0.58	0.59
weighted avg	0.78	0.80	0.79

Risks and Limitations

- Data Provenance
- Aircraft Failure vs Pilot Error
- Currency of the Pilot
  - Total Number of Flying Hours
- Other Licences
- Accidents only but not non accident flights

Next Steps

To collect data from the Air Accident Investigation Branch

Look into areas that were not possible to be covered in this project

Summary

Several interesting insights during EDA

I do not believe these models are accurate enough to show me areas where predictions can be made

Thank you