Welcome to the RESG Safety Data Analysis Workshop

Do the right thing

Never stop learning

Build collaborative relationships

Respect everyone

Why are we here?

Workshop Plan

Monday 1400 – 1700

1400 – 1415 – Introduction and Workshop Plan
1415 – 1515 – Data Sources - *Open and closed data sources, how to combine the data? (Interactive)*
1515 – 1545 – *Coffee Break*
1545 – 1700 – Analysis techniques - *Web scraping examples*

Tuesday 0900 – 1700

0900 – 0915 – Recap and Questions - *Data sources and Web scraping*
0915 – 1015 – Analysis techniques continued - *Our Tools and utility - What tools do you use? (Interactive)*
1015 – 1045 – Coffee Break
1045 – 1200 – Analysis techniques continued - *Statistical based approaches*
1200 – 1300 – Lunch
1300 – 1400 – Barrier Modelling - *Risk evaluation*
1400 – 1430 – Coffee Break
1430 – 1515 – Safety Performance Indicators - *What are they?*
1515 – 1600 – Dashboard creation - *Combining Barrier Modelling and SPIs (Interactive)*
1600 – 1700 – Actionable Insights *(Action* and Prioritisation)
1700 – 1730 – Round up and Questions

Data Sources

Do the
right thing

Never stop
learning

Build collaborative
relationships

Respect
everyone

Which data sources do we use?

Internal Sources

Search the G-INFO aircraft register
Search for an aircraft's details

Aircraft Register

Dynamics 365 | EPT

Entity Performance

AvStats

Aircraft Utilisation

Q-Pulse® an ideagen product

Audit Data

Big Data

MOR Data

ECCAIRS AVIATION

RAIL

MARITIME

ECCAIRS2 - SRIS2

Airprox Data

UK AIRPROX BOARD

AIB Data

AAIB Air Accidents Investigation Branch

External Sources

Global Accident Data

CIRIUM aviation analytics

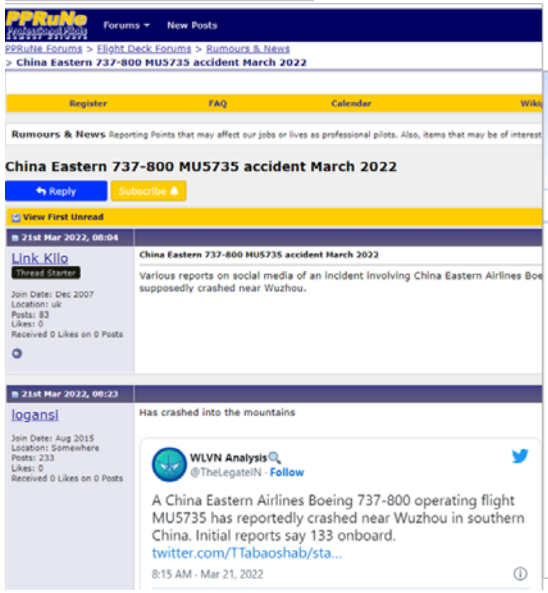
CH!RP Confidential Human Factors Incident Reporting Programme

Confidential Incident Reporting

What other data can we use?

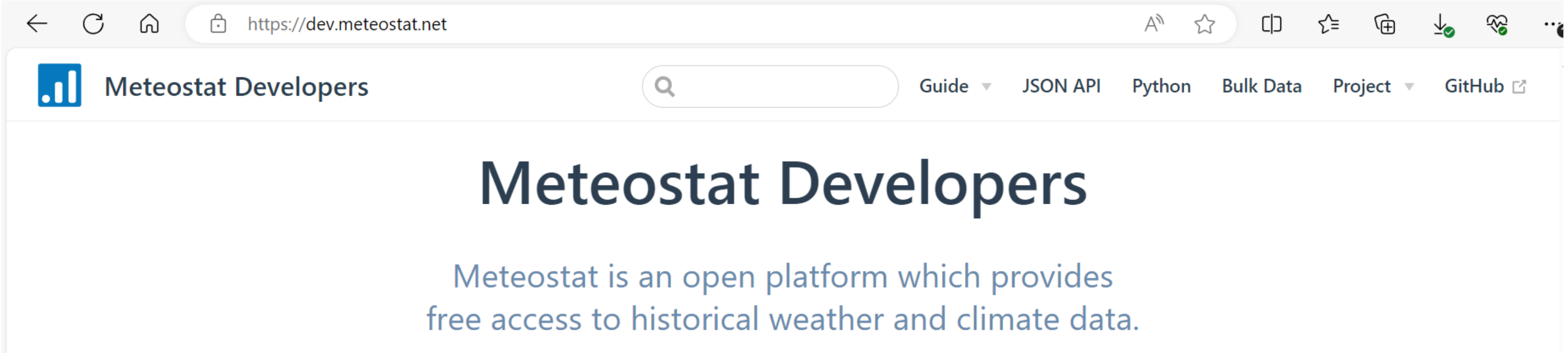
PPRuNe Professional Pilots Rumour Network

AVIATION SAFETY NETWORK



Web Scraping (Python or VBA)

Python Library or JSON API



Web Scraping Example

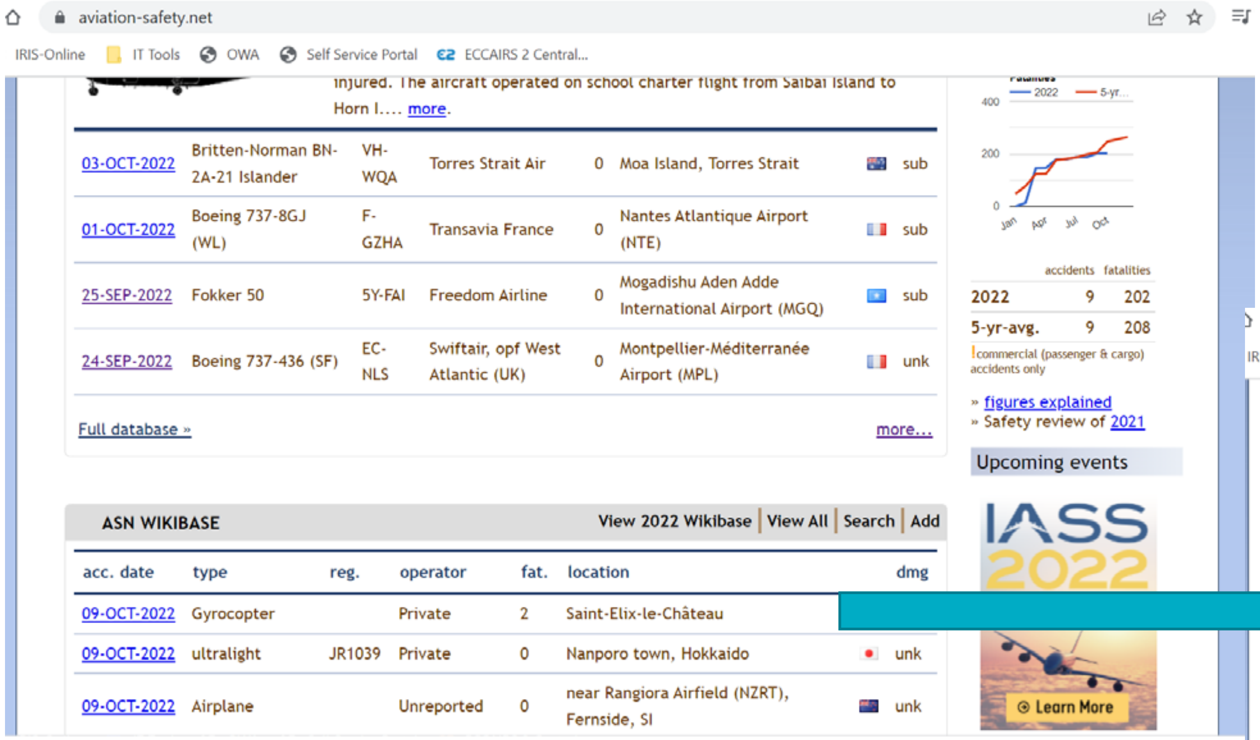
Do the right thing

Never stop learning

Build collaborative relationships

Respect everyone

Aviation Safety Network (ASN) Web Scraping



aviation-safety.net

IRIS-Online | IT Tools | OWA | Self Service Portal | ECCAIRS 2 Central...

injured. The aircraft operated on school charter flight from Saibai Island to Horn I.... more.

03-OCT-2022	Britten-Norman BN-2A-21 Islander	VH-WQA	Torres Strait Air	0	Moa Island, Torres Strait		sub
01-OCT-2022	Boeing 737-8GJ (WL)	F-GZHA	Transavia France	0	Nantes Atlantique Airport (NTE)		sub
25-SEP-2022	Fokker 50	5Y-FAI	Freedom Airline	0	Mogadishu Aden Adde International Airport (MGQ)		sub
24-SEP-2022	Boeing 737-436 (SF)	EC-NLS	Swiftair, opf West Atlantic (UK)	0	Montpellier-Méditerranée Airport (MPL)		unk

Full database » more...

ASN WIKIBASE — View 2022 Wikibase | View All | Search | Add

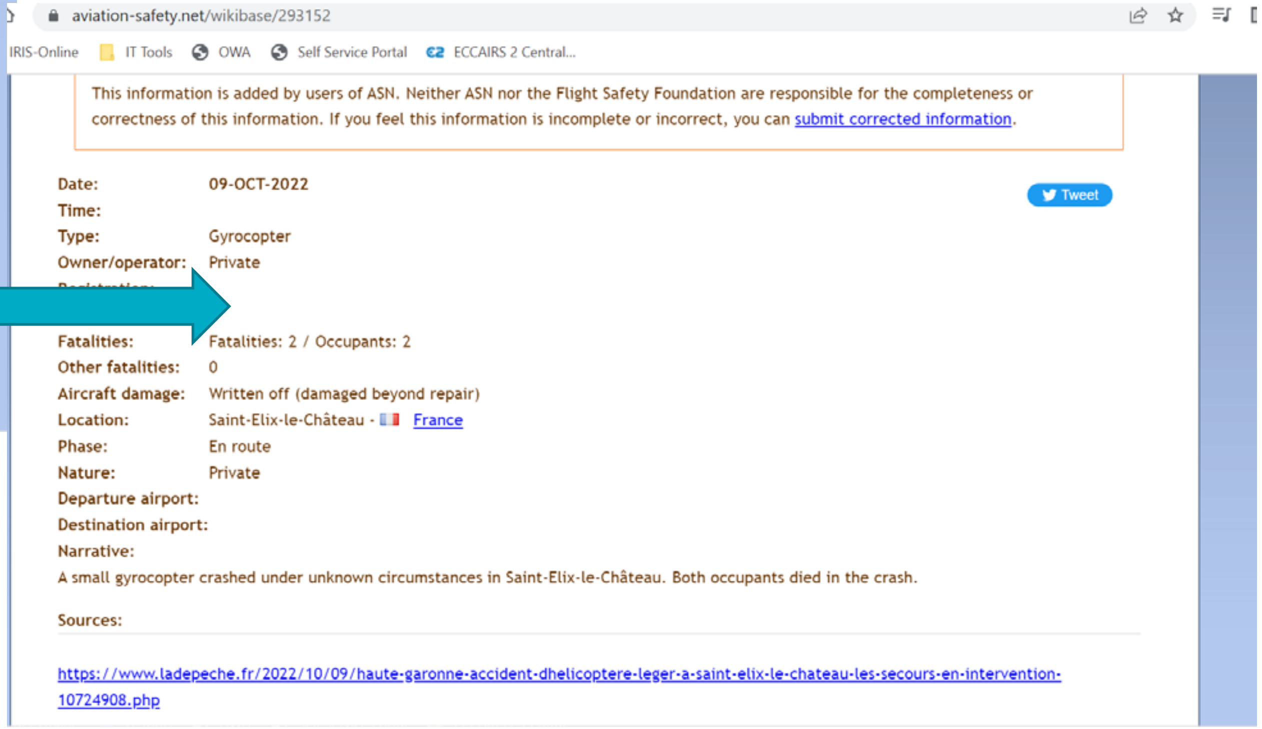
acc. date	type	reg.	operator	fat.	location		dmg
09-OCT-2022	Gyrocopter		Private	2	Saint-Elix-le-Château		
09-OCT-2022	ultralight	JR1039	Private	0	Nanporo town, Hokkaido		unk
09-OCT-2022	Airplane		Unreported	0	near Rangiora Airfield (NZRT), Fernside, SI		unk

	accidents	fatalities
2022	9	202
5-yr-avg.	9	208

commercial (passenger & cargo) accidents only

» figures explained
» Safety review of 2021

Upcoming events



aviation-safety.net/wikibase/293152

IRIS-Online | IT Tools | OWA | Self Service Portal | ECCAIRS 2 Central...

This information is added by users of ASN. Neither ASN nor the Flight Safety Foundation are responsible for the completeness or correctness of this information. If you feel this information is incomplete or incorrect, you can submit corrected information.

Tweet

Date:	09-OCT-2022
Time:	
Type:	Gyrocopter
Owner/operator:	Private
Fatalities:	Fatalities: 2 / Occupants: 2
Other fatalities:	0
Aircraft damage:	Written off (damaged beyond repair)
Location:	Saint-Elix-le-Château - France
Phase:	En route
Nature:	Private
Departure airport:	
Destination airport:	

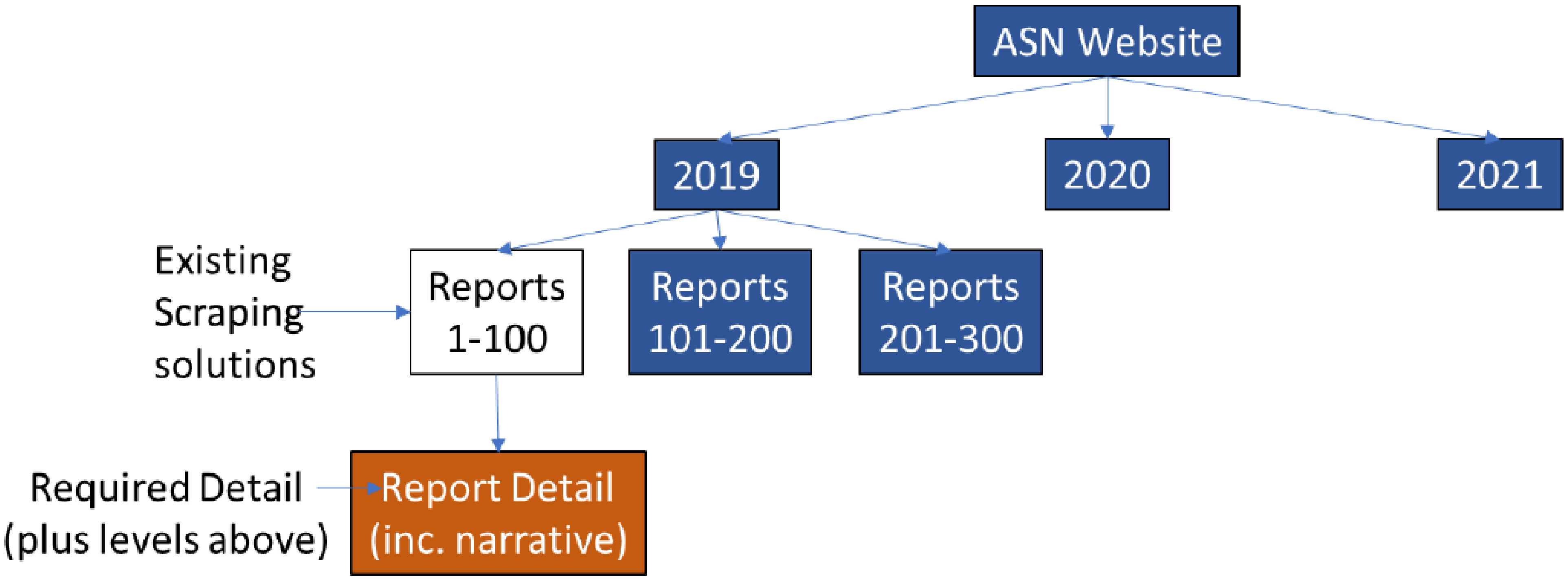
Narrative:

A small gyrocopter crashed under unknown circumstances in Saint-Elix-le-Château. Both occupants died in the crash.

Sources:

https://www.ladepeche.fr/2022/10/09/haute-garonne-accident-dhelicoptere-leger-a-saint-elix-le-chateau-les-secours-en-intervention-10724908.php

Aviation Safety Network (ASN) Web Scraping



Aviation Safety Network (ASN) Web Scraping

Project - weekly brief

- weekly brief (ASN
  - Microsoft Access C
    - Form_frmASN
    - Form_frmASN
  - Modules

Properties - Command5

Command5 CommandBu

Alphabetic | Categorized

BackStyle	1
BackThemeC	4
BackTint	60
Bevel	0
BorderColor	14461583
BorderShade	100
BorderStyle	1
BorderTheme	4
BorderTint	60
BorderWidth	0
BottomPaddi	30
Cancel	False
Caption	Process Dat
ControlTipTe	
ControlType	104
CursorOnHo	0 - acCursoi
Default	False
DisplayWhen	0
Enabled	True
EventProcPre	Command5
FontBold	0

Command5 | Click

```
        rs.AddNew

        For Each delement In delements
            With delement
                If .className = "list" Or .className = "listmain" Then
                    numoccs = 100
                    numpages = 1

                    str1 = .textContent
                    If ptr = 10 Then
                        rs.Update
                        ptr = 0
                        rs.AddNew
                    End If
                    If Strings.Len(str1) > 0 Then
                        rs.Fields(ptr + 1).Value = Trim(str1)
                    End If
                    ptr = ptr + 1
                End If
            End With
        Next delement
        rs.Update

        'close and reopen data set

        rs.Close
        Set rs = Nothing

        sqls = "SELECT [tblASN WikiBase].ID, [tblASN WikiBase].URLID"
        sqls = sqls & " FROM [tblASN WikiBase]"
        sqls = sqls & " WHERE [tblASN WikiBase].URLID = '' or [tblASN WikiBase].URLID is null"
        sqls = sqls & " ORDER BY [tblASN WikiBase].ID"

        Set rs = db.OpenRecordset(sqls, dbOpenDynaset)

        If Not rs.EOF Then
            rs.MoveFirst
```

OccDate	OccType	Registration	Operator	Fat	Location	Damage	URLID
02-JAN-2021	Learjet 31A	PP-BBV	Brasil Vida Táxi	0	Diamantina Air	sub	https://aviation-safety.net/database/record.p
02-JAN-2021	Eurocopter AS 350B3 Ecureuil	ZS-OXK	SANParks	0	Cape Town Inte	sub	https://aviation-safety.net/wikibase/246429
02-JAN-2021	Beech 200 Super King Air	N831WP	Vagus Group	0	White Plains-W	unk	https://aviation-safety.net/database/record.p
02-JAN-2021	Piper PA-24-250 Comanche	N8347P	Aircom LLC	3	N of Oakland S	w/o	https://aviation-safety.net/wikibase/246440
02-JAN-2021	Progressive Aerodyne Searey LSX	N17TS	Private	0	Burnet County,	unk	https://aviation-safety.net/wikibase/246469
03-JAN-2021	Mooney M20F	C-GYGN	Private	0	near pper Kana	non	https://aviation-safety.net/wikibase/265402
03-JAN-2021	Beechcraft 200 Super King Air	C-GFSG	Transwest Air	0	60 nm N of La F	min	https://aviation-safety.net/wikibase/246729
03-JAN-2021	Comco Ikarus C42B	EI-ERM	Private	0	Bartragh Island	sub	https://aviation-safety.net/wikibase/276350
03-JAN-2021	Airbus A321-251NX	G-UZMI	easyJet	0	Bristol Airport,	non	https://aviation-safety.net/wikibase/246979
03-JAN-2021	Boeing 737-8EH (WL)	PR-GGP	Gol	0	São Paulo-Guai	min	https://aviation-safety.net/wikibase/247967
03-JAN-2021	Boeing 787-8 Dreamliner	N782AM	Aeroméxico	0	Cancún Airport	min	https://aviation-safety.net/wikibase/246456
03-JAN-2021	Boeing 777-FF2	TC-LJN	Turkish Airlines	0	near Istanbul-A	min	https://aviation-safety.net/wikibase/246443
04-JAN-2021	Beechcraft [illegible]58 Baron	N325GC	Stratus Sales LL	2	Credit, AR	w/o	https://aviation-safety.net/wikibase/246459
[illegible]	[illegible]usky	N271TM	Private	0	Benjamín River	sub	https://aviation-safety.net/wikibase/246475
04-JAN-2021	Murphy Re[illegible]el	N616PM	Private	0	Suwannee Bell	sub	https://aviation-safety.net/wikibase/246533
04-JAN-2021	Beechcraft B300 King Air 350i	N856UP	Wheels Up	0	Coral Creek Air	min	https://aviation-safety.net/wikibase/246492
04-JAN-2021	Pilatus PC-12/45	N238VM	Guardian Flight	0	Phoenix Sky Ha	unk	https://aviation-safety.net/wikibase/246493
04-JAN-2021	Aviat A-1B Husky	D-EOGY	Private	0	Vöraner Alm	non	https://aviation-safety.net/wikibase/246497
05-JAN-2021	Piper PA-30 Twin Comanche	N7305Y	Arkadelphia Ai	0	Florence Memo	min	https://aviation-safety.net/wikibase/246534
05-JAN-2021	Piper PA-25 Pawnee 235	PR-TCE	Aeroagrícola Cl	0	Guará, SP	sub	https://aviation-safety.net/wikibase/246752
05-JAN-2021	Ayres S-2R-T34 Turbo Thrush 510P	HC-CTQ	Grupo Manoba	0	Cenepa	min	https://aviation-safety.net/wikibase/246483
05-JAN-2021	Beechcraft 1900C	N31702	Ameriflight	0	San Antonio Int	non	https://aviation-safety.net/wikibase/246518
05-JAN-2021	Piper PA-28-236 Dakota	XB-CFP	Escuela de Vue	0	Las Antenas, Co	sub	https://aviation-safety.net/wikibase/246494
05-JAN-2021	Hindustan MiG-21 Bison		Indian Air Forc	0	Suratgarh Air F	w/o	https://aviation-safety.net/wikibase/246477
05-JAN-2021	Boeing Chinook HC6A (CH-47F)	ZA679	RAF, 28 Sqn	0	Wantage, Oxfo	non	https://aviation-safety.net/wikibase/246543

Aviation Safety Network (ASN) Dashboard

Year, Month

- 2021
- 2022

Manufacturer

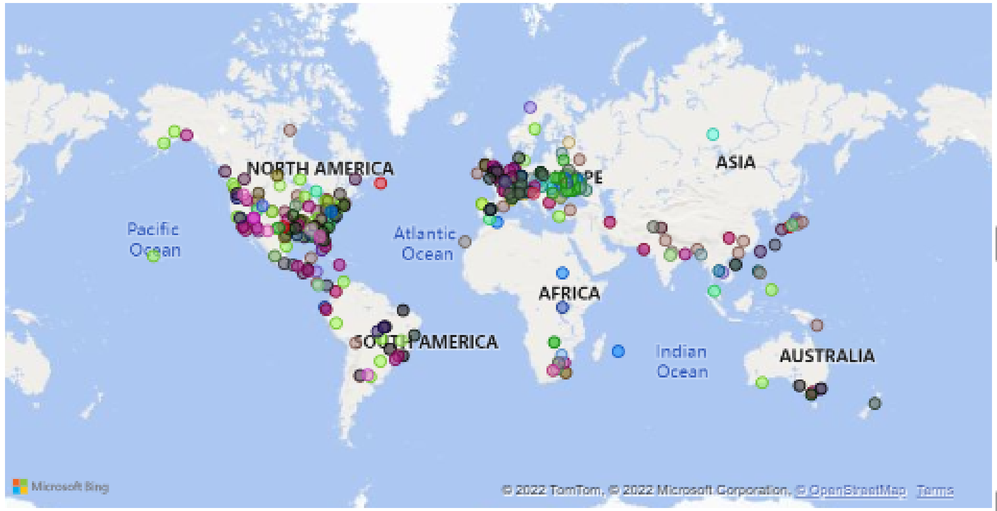
All

AVIATION SAFETY NETWORK

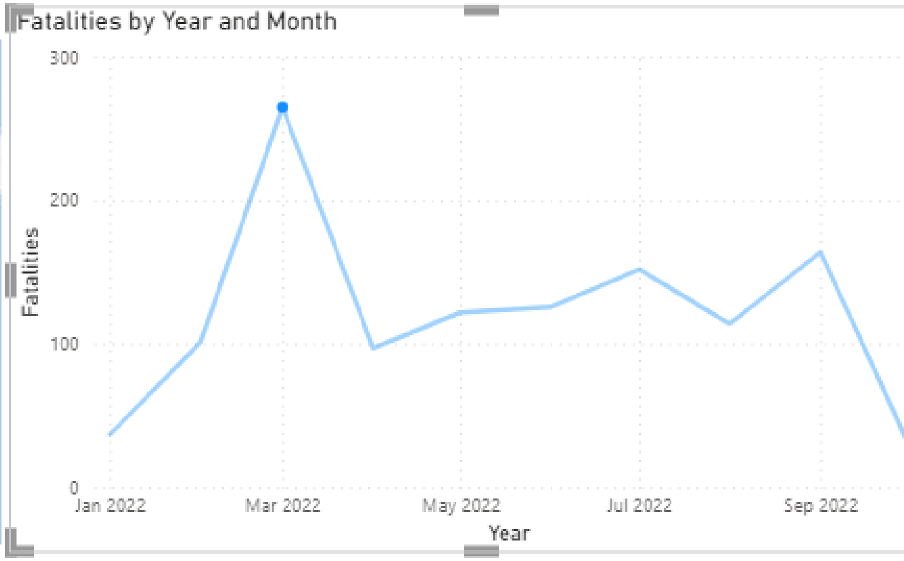
AN EXCLUSIVE SERVICE OF FLIGHT SAFETY

Location: and Manufacturer

Manufacturer (Blank) A1-SM Aero Aeronca Aeroprakt Aeropro Aerospati... Aérospati...



Fatalities by Year and Month



Occ Date	Location:	Fatalities	OccType	Narrative:
21/03/2022 00:00:00	20 km (12.5 mls) SW of Wuzhou, Tengxian County (China)	132	Boeing 737-89P (WL)	China Eastern Airlines flight 5735, a Boeing 737-800, crashed in a mountainous area at 13:15 (05:15 UTC) on a domestic passenger service to Guangzhou Baiyun Internati flying along Airway 599 towards Guangzhou. ADS-B data released by Flightradar24 s 06:22:00 around 7400 feet and the aircraft climbed to about 8600 feet before the air dashcam appears to show the aircraft in an extremely steep descent before it crashe point where flight MU5735 began its descent (also from 29100 feet), the day before t
31/03/2022 00:00:00	near Mariupol - Ukraine	15	Mil Mi-8MTV	The China Global Television Network showed footage on March 31, 2022, of a helico onboard
Total		265		

Meteostat Python Code

- **Meteostat weather data can obtained using a Python Library.**

- **A number of variables are available for global weather stations, daily and hourly.**

```
#set variables for data retrieval
lat = 51.1536621
lng = -0.1842516
start = datetime(2018, 1, 1)
end = datetime(2021, 12, 31, 23, 59)

#import required code libraries
import pandas as pd
from meteostat import Stations
from itertools import product
from datetime import datetime
from meteostat import Daily
from functools import reduce

#get station list
stations = Stations()
stations = stations.nearby(lat, lng)
station = stations.fetch()
#filter list for GB stations and write to csv
station_GB = station[station.country.eq('GB')]
station_GB.to_csv(r'C:\Users\paul_\Desktop\station_GB.csv', sep='\t', encoding='utf-8')

i=0
#loop through GB weather stations and get weather data from 2018
for index, row in station_GB.iterrows():
    data = Daily(index, start, end)
    data = data.fetch()
    if i==0:
        df = pd.DataFrame()
        df = data
        df['station'] = index
        i = i + 1
    else:
        df2 = pd.DataFrame()
        df2 = data
        df2['station'] = index
        df = pd.concat([df, df2])
        i = i + 1

#write data to csv
df.to_csv(r'C:\Users\paul_\Desktop\Merged_Weather_GB.csv', sep='\t', encoding='utf-8')
```

TEMP	Air Temperature
TAVG	Average Temperature
TMIN	Minimum Temperature
TMAX	Maximum Temperature
DWPT	Dew Point
PRCP	Total Precipitation
WDIR	Wind (From) Direction
WSPD	Average Wind Speed
WPGT	Wind Peak Gust
RHUM	Relative Humidity
PRES	Sea-Level Air Pressure
SNOW	Snow Depth
TSUN	Total Sunshine Duration
COCO	Weather Condition Code

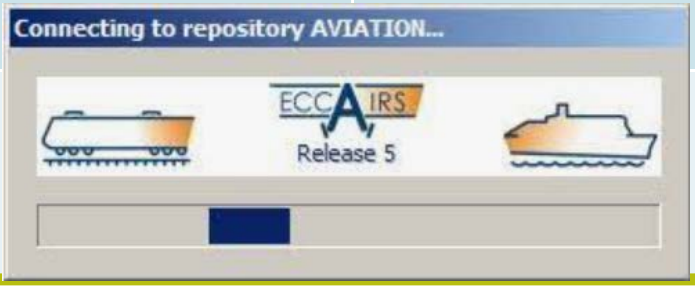
Tools

Microsoft

INGESTION

Azure

Microsoft SQL Server



Microsoft SQL Server Management

STORAGE

INTEGRATION

MANAGEMENT

PROCESSING

Excel

BowTie XP
IncidentXP
CGE Risk Management Solutions BV
www.cgerisk.com

alteryx
The Thrill of Solving

REPORTING

Power BI

Natural Language Analysis with Python NLTK

RStudio

DATA SCIENCE CAPABILITY

Snipping Tool
New Mode Delay Cancel

Google

how to make aviation safer using data

Safety Intelligence team

Group Exercise

- **How can the following be used together to provide more insight than just MOR data?**
- **What additional data would we need to make this work?**

Internal Sources

Search the G-INFO aircraft register
Search for an aircraft's details

Aircraft Register

Dynamics 365 | EPT

Entity Performance

AvStats

Aircraft Utilisation

Q-Pulse®
an ideagen product

Audit Data

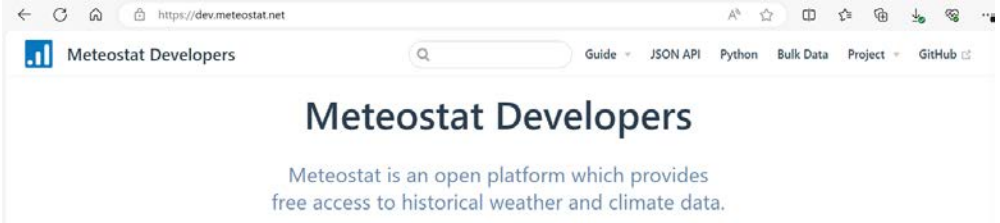
Big Data

MOR Data

ECCAIRS2 - SRIS2

External Sources

Weather Data



Global Incidents and Accidents

AVIATION SAFETY NETWORK
AN EXCLUSIVE SERVICE OF FLIGHT SAFETY

Analysis Techniques

Do the right thing

Never stop learning

Build collaborative relationships

Respect everyone

The Excel One-Line Macro – ECCAIRS 1

A one-line macro was written using VBA in Excel for ECCAIRS 1 Data Manager outputs to transform multiple lines of data to a single line with multiple columns for records with the same file number.

File Number	Event Type 1	Event Type 2	Event Type 3	Event Type 4	Event Type 5
123456789	123	456	789	234	567

File Number	Event Type
123456789	123
123456789	456
123456789	789
123456789	234
123456789	567

One-Line Macro Equivalent in SQL

The same conversion of rows to columns can be achieved using SQL.

1. In tables with an instance ID (e.g. Event Type)

```
SELECT File_Number, [1] as [Event Type 1], [2] as [Event Type 2], [3] as [Event Type 3], [4] as [Event Type
4], [5] as [Event Type 5], [6] as [Event Type 6], [7] as [Event Type 7], [8] as [Event Type 8], [9] as [Event Type
9], [10] as [Event Type 10], [11] as [Event Type 11], [12] as [Event Type 12], [13] as [Event Type 13], [14] as [Event
Type 14]

FROM

   ( SELECT File_Number, Event_Instance_ID, Event_Type_Description

     FROM Eccairs.Eccairs_Events_View

       where [Event_Type_Description] <> '' and Is_Active <> 2

   ) PS

PIVOT

   ( MAX(Event_Type_Description)

     FOR Event_Instance_ID IN ( [1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14])

   ) AS PVT
```

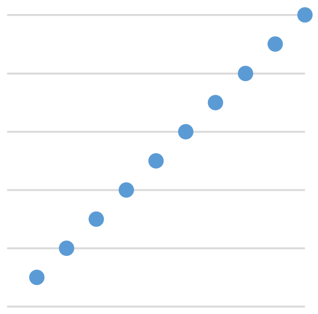
One-Line Macro Equivalent in SQL

2. In MVA tables without an instance ID (e.g. Occurrence Code)

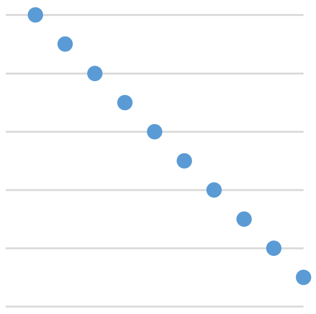
```
SELECT File_Number, [1] AS [Occ Cat 1], [2] AS [Occ Cat 2], [3] AS [Occ Cat 3], [4] AS [Occ Cat 4]
FROM
   ( SELECT Occ1.File_Number, Occ1.Attribute_Desc, 1+ ([rowid] - minrowid) AS OCC_CAT_Instance_ID FROM
(SELECT [File_Number], [Attribute_Name], [Attribute_Desc], ROW_NUMBER() OVER(ORDER BY file_Number, attribute_value) AS rowid
  FROM [Eccairs].[Eccairs_Occurrence_MVA_View]
  WHERE Attribute_Name = 'Occurrence_Category' and Is_Active <> 2) Occ1
inner join
(  SELECT OM.[File_Number], MIN(OM.rowid) AS minrowid FROM
  (SELECT [File_Number], [Attribute_Name], [Attribute_Desc], ROW_NUMBER() OVER(ORDER BY file_Number, attribute_value) AS rowid
  FROM [Eccairs].[Eccairs_Occurrence_MVA_View]
  WHERE Attribute_Name = 'Occurrence_Category' and Is_Active <> 2) OM
  GROUP BY OM.File_Number) Occ2
  ON Occ1.file_number = Occ2.file_number
   ) PS
PIVOT
   ( MAX(Attribute_Desc)
     FOR OCC_CAT_Instance_ID IN ( [1], [2], [3], [4])
   ) AS PVT
```

Correlation

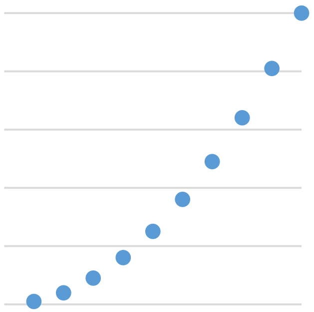
Correlation means that there is a relationship between two variables



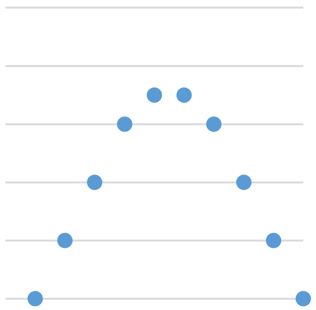
Positive Correlation



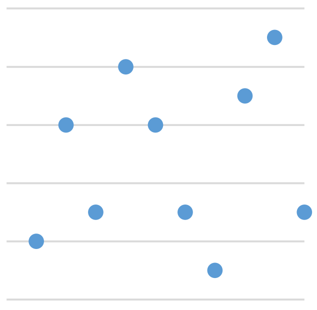
Negative Correlation



Polynomial Curve



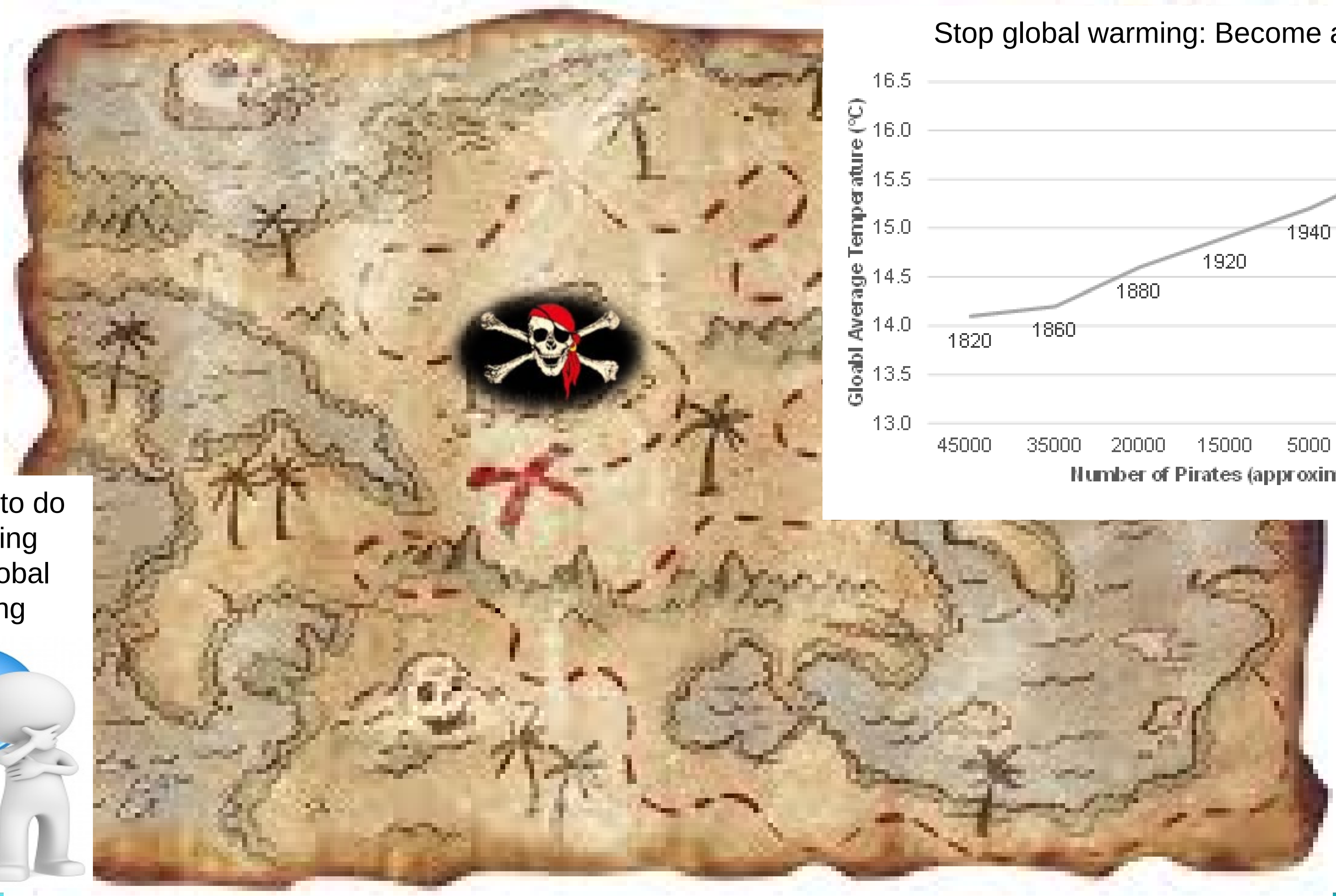
Bell Curve



No Correlation

Correlation and Causation

Beware! Not all correlations are meaningful.



We need to do something about global warming

Regression Lines

REMEMBER: Statistical Tests Should only be used when it is appropriate

Regression lines define the relationship between selected values of X and observed values of y

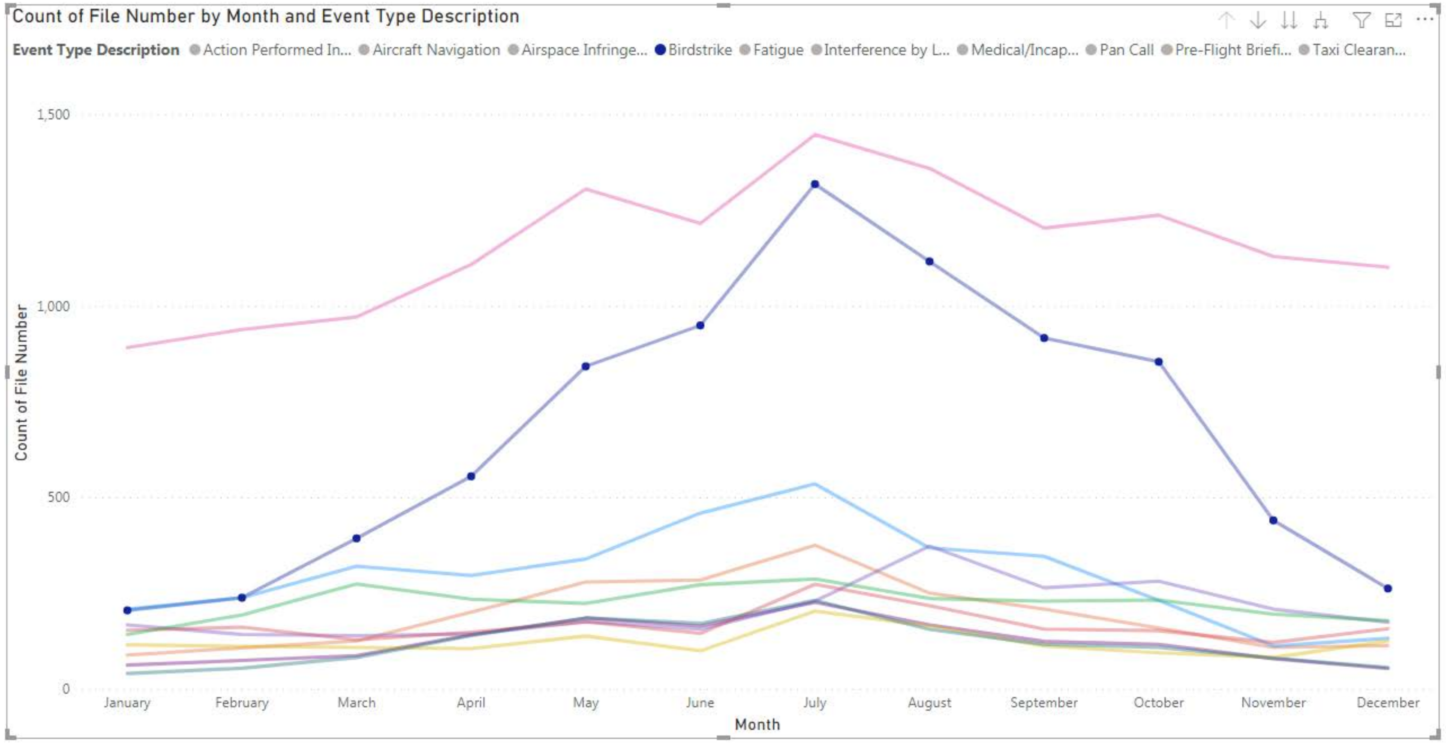
In reality data rarely fits exactly on this line, hence the name line of best fit

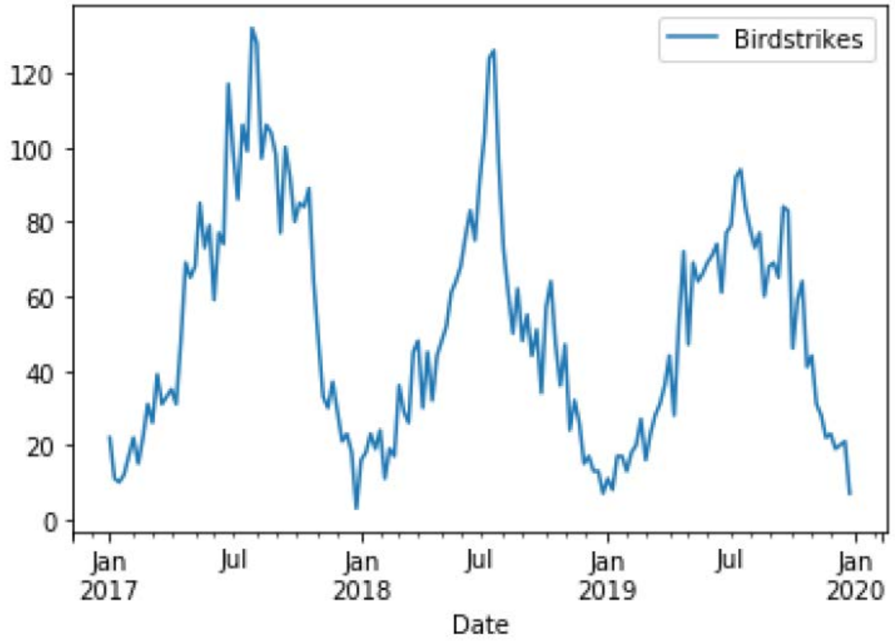
Regression lines are also referred to as trend lines

Upon this line data can be plotted to visualise any interdependencies that may exist

Forecasting Birdstrikes

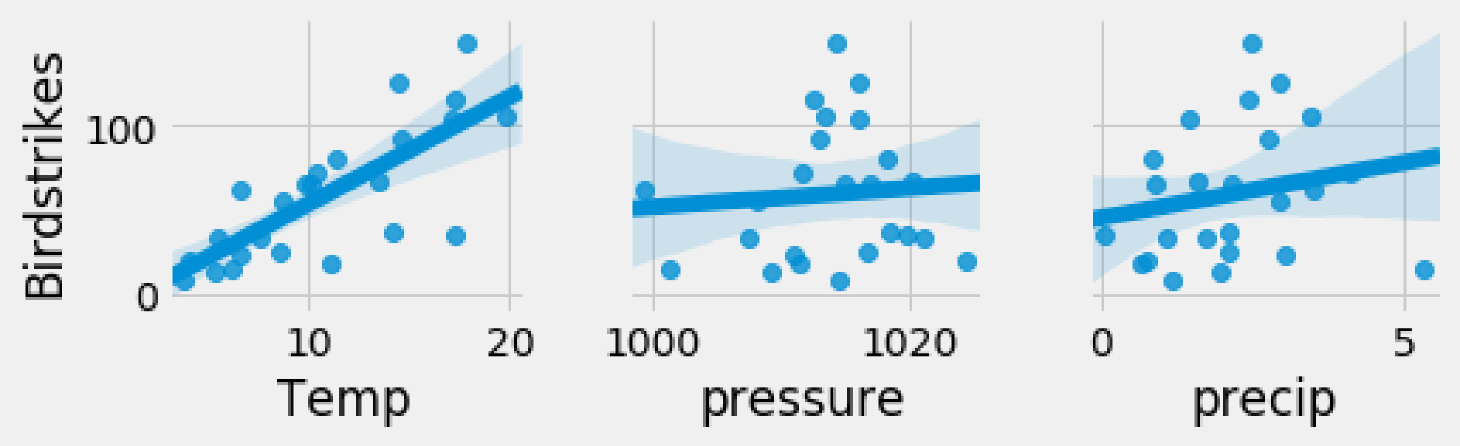
- Of the most common event types, birdstrikes appear to show the greatest degree of seasonality.
- The number of birdstrikes will be forecast using 2 approaches:
  - Linear Regression using weather data.
  - Time Series using historical birdstrike data.
- Both techniques are demonstrated using Python.





Forecasting Birdstrikes - Regression

- We can use weather data to estimate the number of birdstrikes using linear regression.
- The approach relies on accurate weather forecasts, so can only be used reliably for a few days ahead.
- The number of birdstrikes is correlated with Temperature, Pressure and Precipitation.
- For temperature alone, **weekly number of birdstrikes = (6.3*temperature) – 9.5 (can anyone see a limitation here?).** The R^2 statistic is 0.585, so most of the variance is accounted for by temperature alone.
- Adding Precipitation improves the R^2 statistic to 0.631. Adding pressure does little to improve the R^2 value.



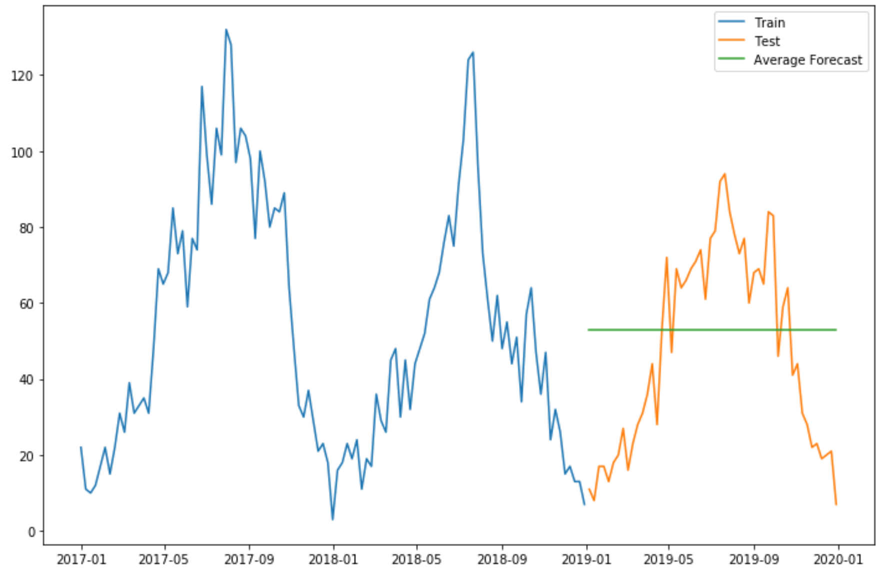

```
feature_cols = ['Temp']

# Create X and y.
X = weather[feature_cols]
y = weather.Birdstrikes

# Instantiate and fit.
linreg = LinearRegression()
linreg.fit(X, y)

# Print the coefficients.
print(linreg.intercept_)
print(linreg.coef_)

# Single Feature Model
print(linreg.score(X,y))

# Single Feature Model
predict_single=linreg.predict(X)

# Single Feature Model
print('MAE:', metrics.mean_absolute_error(y, predict_single))
print('MSE:', metrics.mean_squared_error(y, predict_single))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y, predict_single)))

feature_cols = ['Temp', 'precip']

# Create X and y.
X = weather[feature_cols]
y = weather.Birdstrikes

# Instantiate and fit.
linreg = LinearRegression()
linreg.fit(X, y)

# Print the coefficients.
print(linreg.intercept_)
print(linreg.coef_)

# Multiple Feature Model
print(linreg.score(X,y))

# Multiple Feature Model
predict_multiple=linreg.predict(X)

# Multiple Feature Model
print('MAE:', metrics.mean_absolute_error(y, predict_multiple))
print('MSE:', metrics.mean_squared_error(y, predict_multiple))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y, predict_multiple)))
```

Forecasting Birdstrikes – Time Series

- We can also estimate the number of birdstrikes using a Time Series approach.
- There are 2 components in a time series forecast: trend and seasonality.
- The approach relies on historical data, so may not be accurate for any given future year.

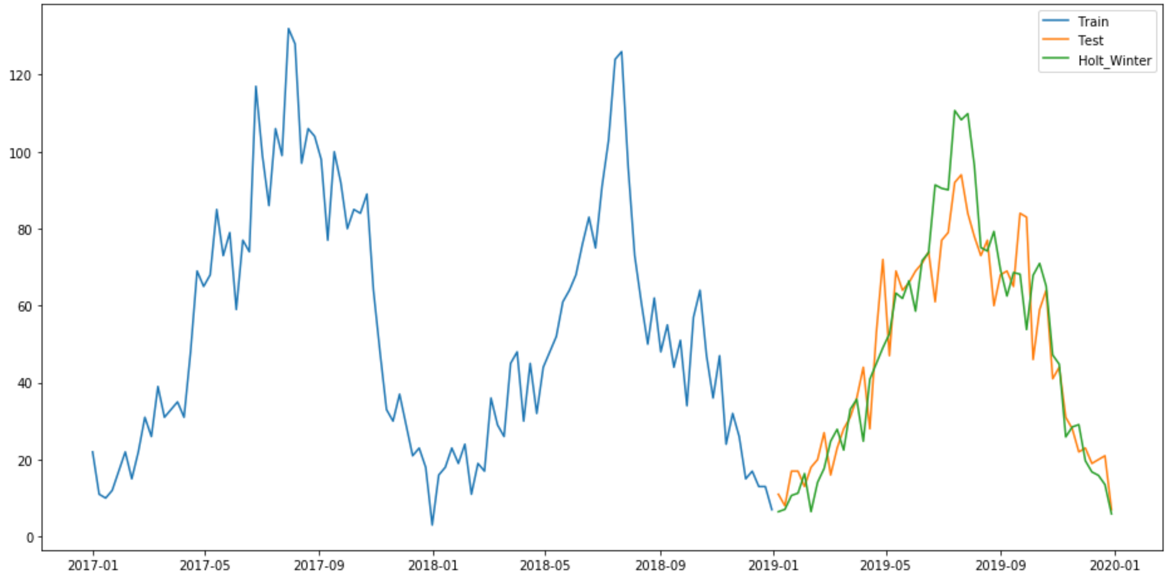
Average Forecast Method

```
y_hat_avg = test.copy()
y_hat_avg['avg_forecast'] = train['Birdstrikes'].mean()
plt.figure(figsize=(12,8))
plt.plot(train['Birdstrikes'], label='Train')
plt.plot(test['Birdstrikes'], label='Test')
plt.plot(y_hat_avg['avg_forecast'], label='Average Forecast')
plt.legend(loc='best')
plt.show();
```

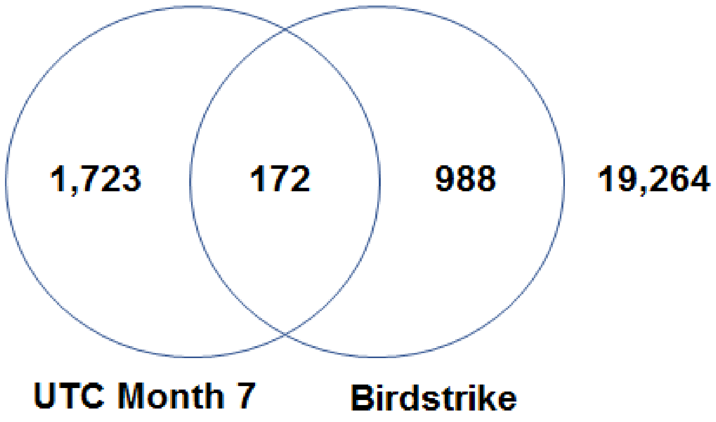


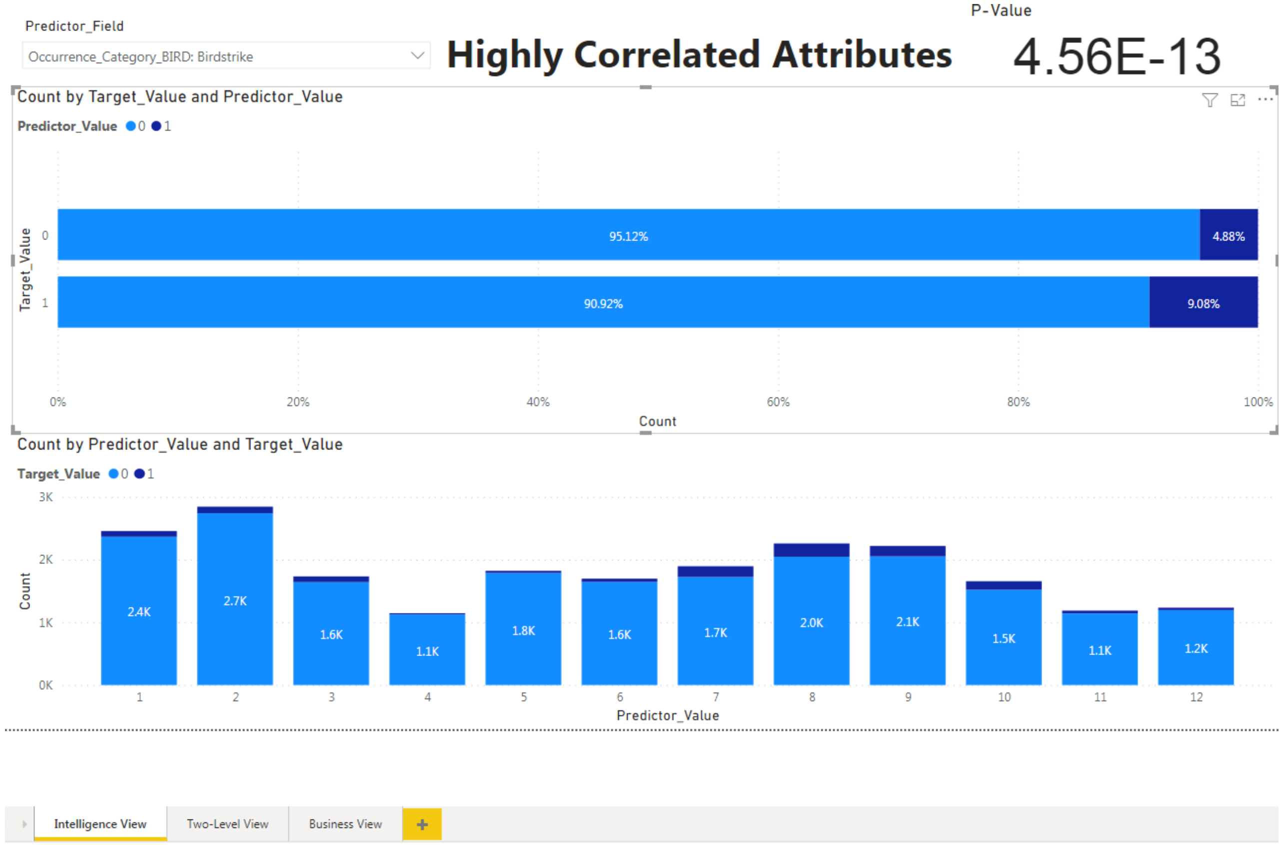
Holt-Winters Forecasting Method

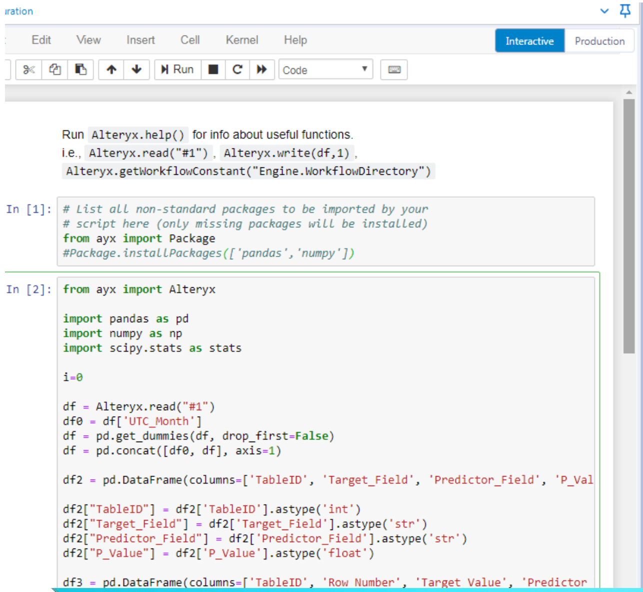
```
y_hat_avg = test.copy()
fit1 = ExponentialSmoothing(np.asarray(train['Birdstrikes'])
,seasonal_periods=52, seasonal='add', trend='add').fit()
y_hat_avg['Holt_Winter'] = fit1.forecast(len(test))
plt.figure(figsize=(16,8))
plt.plot( train['Birdstrikes'], label='Train')
plt.plot(test['Birdstrikes'], label='Test')
plt.plot(y_hat_avg['Holt_Winter'], label='Holt_Winter')
plt.legend(loc='best')
plt.show()
```

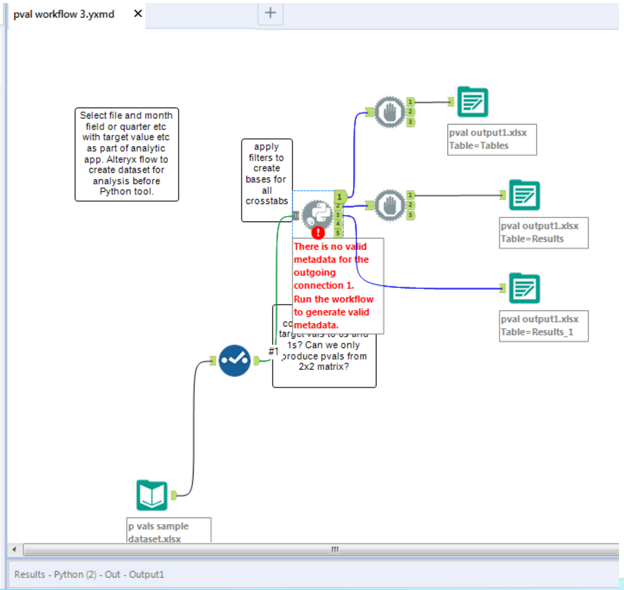


Automatically Identifying Correlations in Data – Python in Alteryx









Automatically Identifying Correlations in Data – Python Code in Alteryx

```python
from ayx import Alteryx

import pandas as pd
import numpy as np
import scipy.stats as stats

i=0

df = Alteryx.read("#1")
df0 = df['UTC_Month']
df = pd.get_dummies(df, drop_first=False)
df = pd.concat([df0, df], axis=1)

df2 = pd.DataFrame(columns=['TableID', 'Target_Field', 'Predictor_Field', 'P_Value'])

df2["TableID"] = df2['TableID'].astype('int')
df2["Target_Field"] = df2['Target_Field'].astype('str')
df2["Predictor_Field"] = df2['Predictor_Field'].astype('str')
df2["P_Value"] = df2['P_Value'].astype('float')

df3 = pd.DataFrame(columns=['TableID', 'Row_Number', 'Target_Value', 'Predictor_Value', 'Count'])
df3["TableID"] = df3['TableID'].astype('int')
df3["Row_Number"] = df3['Row_Number'].astype('int')
df3["Target_Value"] = df3['Target_Value'].astype('int')
df3["Predictor_Value"] = df3['Predictor_Value'].astype('int')
df3["Count"] = df3['Count'].astype('int')

df4 = pd.DataFrame(columns=['TableID', 'Row_Number', 'Target_Value', 'Predictor_Value', 'Count'])
df4["TableID"] = df4['TableID'].astype('int')
df4["Row_Number"] = df4['Row_Number'].astype('int')
df4["Target_Value"] = df4['Target_Value'].astype('int')
df4["Predictor_Value"] = df4['Predictor_Value'].astype('str')
df4["Count"] = df4['Count'].astype('int')

ptr = 0
f1 = 'UTC_Month_7'
f2 = 'UTC_Month'
```

```python
for col in df.columns:
    if ptr > 12:
        df_tb = pd.crosstab(df[f1],df[col])
        oddsratio, pvalue = stats.fisher_exact(df_tb)
        res = pvalue
        if res <= 0.05:
            df_tb = pd.crosstab(df[f1],df[col])
            df_tb2 = pd.crosstab(df[f2],df[col])
            df2 = df2.append({'TableID': ptr, 'Target_Field': f1, 'Predictor_Field': col, 'P_Value': res},ignore_index=
            df3 = df3.append({'TableID': ptr, 'Row_Number': 1, 'Target_Value': 0, 'Predictor_Value': 0, 'Count': df_tb.
            df3 = df3.append({'TableID': ptr, 'Row_Number': 2, 'Target_Value': 0, 'Predictor_Value': 1, 'Count': df_tb.
            df3 = df3.append({'TableID': ptr, 'Row_Number': 3, 'Target_Value': 1, 'Predictor_Value': 0, 'Count': df_tb.
            df3 = df3.append({'TableID': ptr, 'Row_Number': 4, 'Target_Value': 1, 'Predictor_Value': 1, 'Count': df_tb.
            #if strong correlation and filtered sample sufficient then produce level 2 stats
            if (df_tb.iloc[0,1] + df_tb.iloc[1,1]) >= 300 and res <= 0.00001:
                df_filt = df[df[col] == 1]
                print(df_filt.head(5))
                ptr2 = 2
                for col2 in df_filt.columns:
                    if ptr2 > 12:
                        df_tb_filt = pd.crosstab(df_filt[f1],df[col2])
                        #check left of field name up to last underscore. Skip ctab if the same
                        print(df_tb_filt + ' shape' + df_tb_filt.shape[1])
                        print(f1 + ' ' + col2 + ' ' + col)
                        oddsratio, pvalue = stats.fisher_exact(df_tb_filt)
                        res2 = pvalue
                        print (str(res) + ' ' + str(res2))
                        if res2 < res:
                            print(df_tb_filt)
                    ptr2 = ptr2 + 1
                #do loop for 2nd level
            i=0
            for index, row in df_tb2.iterrows():
                df4 = df4.append({'TableID': ptr, 'Row_Number': 2*i + 1, 'Target_Value': 0, 'Predictor_Value': index, '
                df4 = df4.append({'TableID': ptr, 'Row_Number': 2*(i + 1), 'Target_Value': 1, 'Predictor_Value': index,
                i = i + 1
    ptr = ptr + 1
```

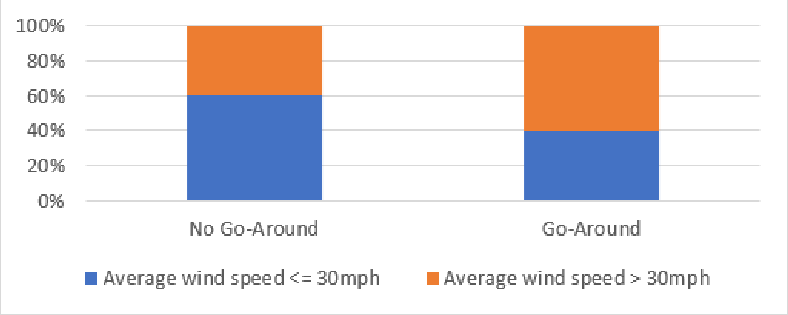
Effect of Sample Size in Hypothesis Testing

- Sample size is important in hypothesis testing.
- An increase or decrease in sample size can make a difference to a hypothesis being accepted or rejected.

Hypothesis: We see more go-arounds on days where the average wind speed is more than 30mph

Actual Values

	No Go-Around	Go-Around	
Average wind speed <= 30mph	**30**	**20**	50%
Average wind speed > 30mph	**20**	**30**	50%
	50%	50%	100%



Expected Values

	No Go-Around	Go-Around	
Average wind speed <= 30mph	**25**	**25**	50%
Average wind speed > 30mph	**25**	**25**	50%
	50%	50%	100%

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(A_{ij}-E_{ij})^2}{E_{ij}}$$

Chi Square **4**

Degrees of Freedom df = (2-1)*(2-1) = 1

CHISQ.TEST(actual_range,expected_range)

0.045500

TABLE 9 Critical Values of the Chi-Square Distribution

Note: Column headings are non-directional (omni-directional) *P*-values. If H_A is directional (which is only possible when df = 1), the directional *P*-values are found by dividing the column headings in half.

	TAIL PROBABILITY						
df	0.20	0.10	0.05	0.02	0.01	0.001	0.0001
1	1.64	2.71	3.84	5.41	6.63	10.83	15.14
2	3.22	4.61	5.99	7.82	9.21	13.82	18.42
3	4.64	6.25	7.81	9.84	11.34	16.27	21.11

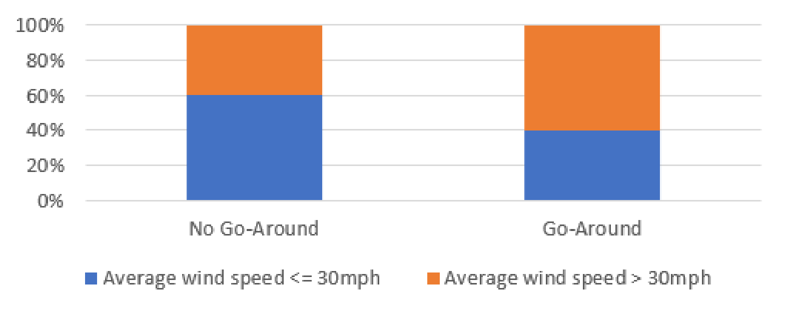
Effect of Sample Size in Hypothesis Testing - Smaller

- Sample size is important in hypothesis testing.
- A smaller sample size leads to the hypothesis being **rejected**.

Hypothesis: We see more go-arounds on days where the average wind speed is more than 30mph

Actual Values

	No Go-Around	Go-Around	
Average wind speed <= 30mph	**3**	**2**	50%
Average wind speed > 30mph	**2**	**3**	50%
	50%	50%	100%



Expected Values

	No Go-Around	Go-Around	
Average wind speed <= 30mph	**2.5**	**2.5**	50%
Average wind speed > 30mph	**2.5**	**2.5**	50%
	50%	50%	100%

Chi Square **0.04**

Degrees of Freedom **df = (2-1)*(2-1) = 1**

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(A_{ij}-E_{ij}\right)^2}{E_{ij}}$$

CHISQ.TEST(actual_range,expected_range)

0.527089

TABLE 9 Critical Values of the Chi-Square Distribution

Note: Column headings are non-directional (omni-directional) *P*-values. If H_A is directional (which is only possible when df = 1), the directional *P*-values are found by dividing the column headings in half.

	TAIL PROBABILITY						
df	0.20	0.10	0.05	0.02	0.01	0.001	0.0001
1	1.64	2.71	3.84	5.41	6.63	10.83	15.14
2	3.22	4.61	5.99	7.82	9.21	13.82	18.42
3	4.64	6.25	7.81	9.84	11.34	16.27	21.11

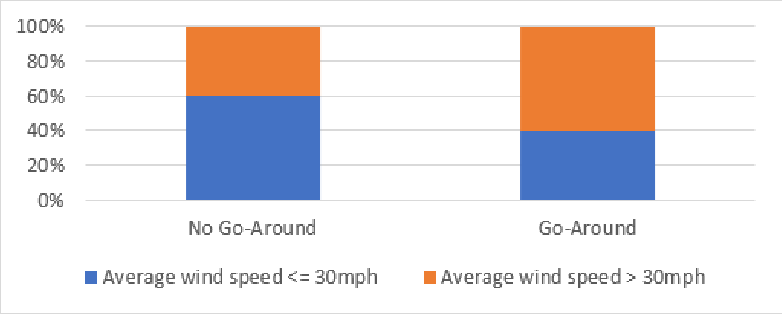
Effect of Sample Size in Hypothesis Testing - Larger

- Sample size is important in hypothesis testing.
- A larger sample size leads to the hypothesis being **accepted**.

Hypothesis: We see more go-arounds on days where the average wind speed is more than 30mph

Actual Values

	No Go-Around	Go-Around	
Average wind speed <= 30mph	**300**	**200**	50%
Average wind speed > 30mph	**200**	**300**	50%
	50%	50%	100%



Expected Values

	No Go-Around	Go-Around	
Average wind speed <= 30mph	**250**	**250**	50%
Average wind speed > 30mph	**250**	**250**	50%
	50%	50%	100%

Chi Square 400

Degrees of Freedom df = (2-1)*(2-1) = 1

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(A_{ij}-E_{ij}\right)^2}{E_{ij}}$$

CHISQ.TEST(actual_range,expected_range)

2.53963E-10

TABLE 9 Critical Values of the Chi-Square Distribution

Note: Column headings are non-directional (omni-directional) *P*-values. If H_A is directional (which is only possible when df = 1), the directional *P*-values are found by dividing the column headings in half.

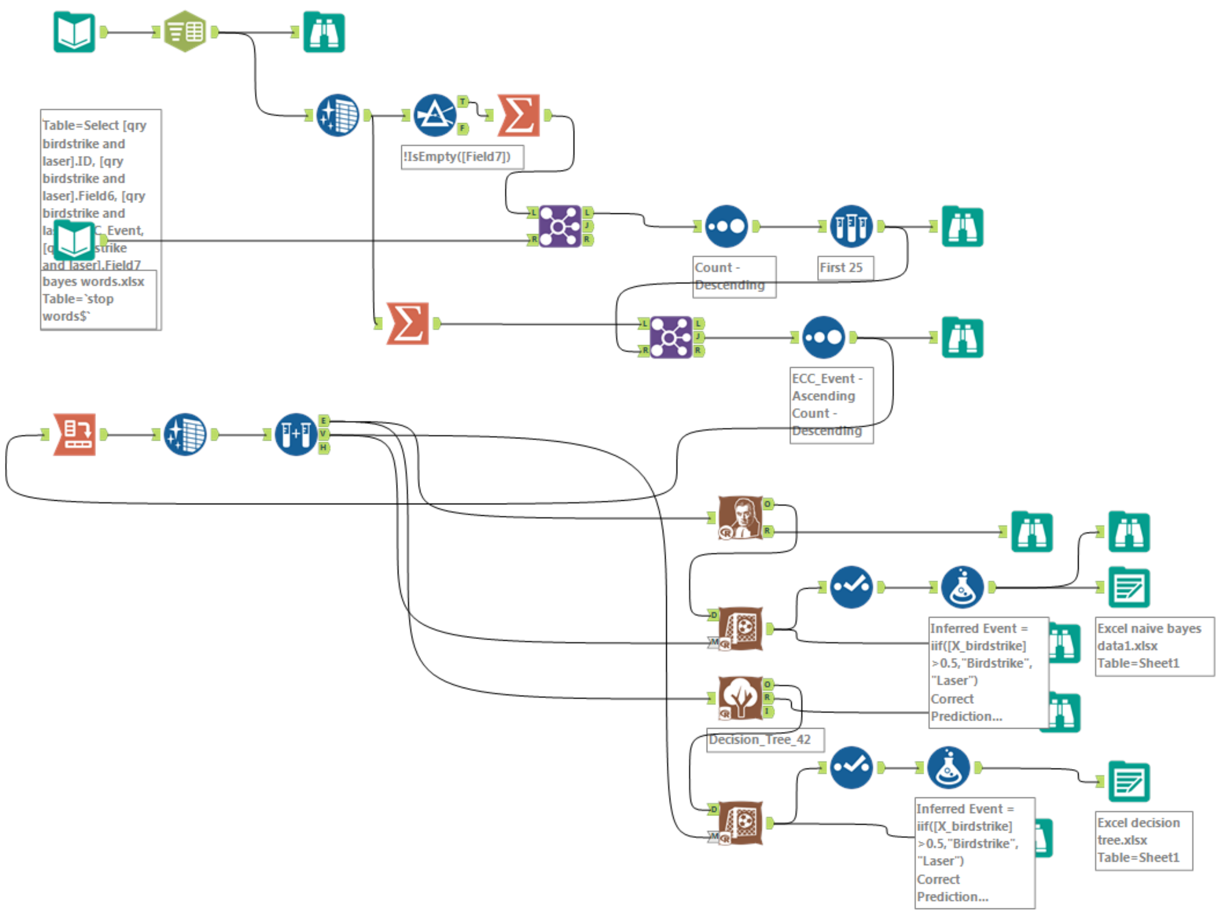
	TAIL PROBABILITY						
df	0.20	0.10	0.05	0.02	0.01	0.001	0.0001
1	1.64	2.71	3.84	5.41	6.63	10.83	15.14
2	3.22	4.61	5.99	7.82	9.21	13.82	18.42
3	4.64	6.25	7.81	9.84	11.34	16.27	21.11

Text Classification

- A Python text classifier was built to predict ECCAIRS Event Codes from the Narrative.
- The classifier was then tested to evaluate its accuracy vs. human event code tagging.
- Some event types were more successfully identified by the classifier than others.

Text Classification – Alteryx workflow

- The same text classification approach can be achieved with a workflow in the Alteryx ETL tool.
- This workflow attempts to classify occurrences as birdstrikes or laser strikes using 2 approaches.



Text Classification – Narrative Words

- Certain words in the narrative allow a successful classification.
- Some event types are not so easy to identify in this way due to the variation of the narrative.

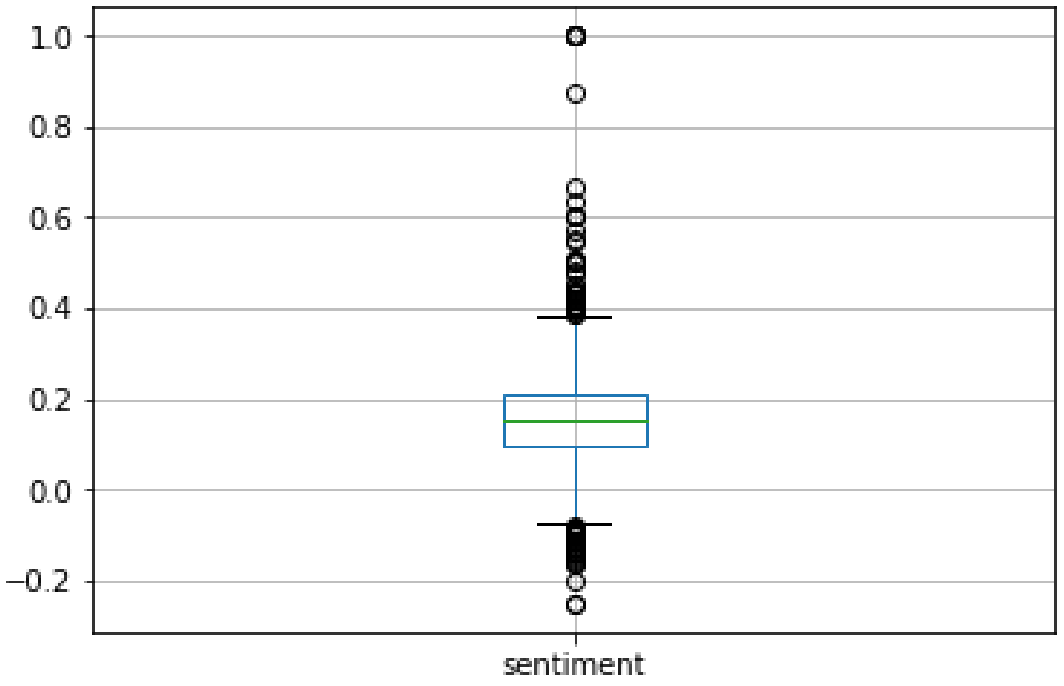
Word	Count	Order	Naïve Bayes - Correct Predictions	Decision Tree - Correct Predictions
bird	3,147	1	-	-
aircraft	2,352	2	90.0%	90.0%
reported	1,935	3	87.0%	87.1%
laser	1,880	4	98.6%	98.7%
strike	1,843	5	90.5%	98.9%
runway	1,635	6	89.6%	98.6%
found	1,356	7	92.5%	98.8%
green	1,161	8	99.2%	99.4%
landing	1,102	9	99.2%	99.1%
inspection	965	10	95.2%	99.3%
approach	837	11		
flight	772	12		
side	728	13		
remains	699	14		
atc	694	15	92.0%	99.3%
pilot	674	16		
informed	655	17		
damage	620	18		
small	618	19		
left	601	20	81.6%	99.2%
birds	553	21		
final	537	22		
engine	531	23		
right	514	24		
birdstrike	494	25	75.5%	99.0%

Field6	ECC_Event	X_birdstrike	X_interference by lasersbeamer	Inferred Event	Correct Prediction
201800199	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800303	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800316	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800390	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800448	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800601	interference by lasersbeamer	0	1	Laser	1
201800608	interference by lasersbeamer	0	1	Laser	1
201800629	interference by lasersbeamer	0	1	Laser	1
201800635	interference by lasersbeamer	0	1	Laser	1
201800642	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800645	interference by lasersbeamer	0	1	Laser	1
201800658	birdstrike	0.991836735	0.008163265	Birdstrike	1
201800666	birdstrike	0.991836735	0.008163265	Birdstrike	1

Sentiment Analysis with Python

- We have used Sentiment Analysis with Python to attempt to triage emails based on the tone (positive or negative) of the email.
- We found that an untrained Sentiment Analysis approach was unable to do this task successfully.

```
import pandas as pd
from textblob import TextBlob, Word

%matplotlib inline
path = r'./data1/[redacted].csv'
emails = pd.read_csv(path, sep='}',encoding='latin1')

# Define a function that accepts text and returns the polarity.
def detect_sentiment(text):
    return TextBlob(text).sentiment.polarity
# Create a new DataFrame column for email body length.
emails['length'] = emails.Text.apply(len)
# Create a new DataFrame column for sentiment.
emails['sentiment'] = emails.Text.apply(detect_sentiment)
# Box plot of sentiment
emails.boxplot(column='sentiment')
# Reviews with most positive sentiment
print(emails[emails.sentiment > 0.75].head())
# Reviews with least positive sentiment
print(emails[emails.sentiment < -0.15].head())
```



Analysis Summary

- We have looked at a number of approaches to analysis using the following tools:
  - SQL
  - VBA (Access, Excel)
  - Power BI
  - Python
  - Alteryx
- Python is a powerful and flexible tool which can help us automate processes:
  - Text classification (e.g. event types and occurrence categories).
  - Identifying correlations in data (e.g. commonly-occurring event types for a certain engine type, event types frequently seen at a certain airport).
  - Forecasting certain types of occurrences based on historical data or weather data.
  - Email triaging.

Barrier Modelling

Do the
right thing

Never stop
learning

Build collaborative
relationships

Respect
everyone

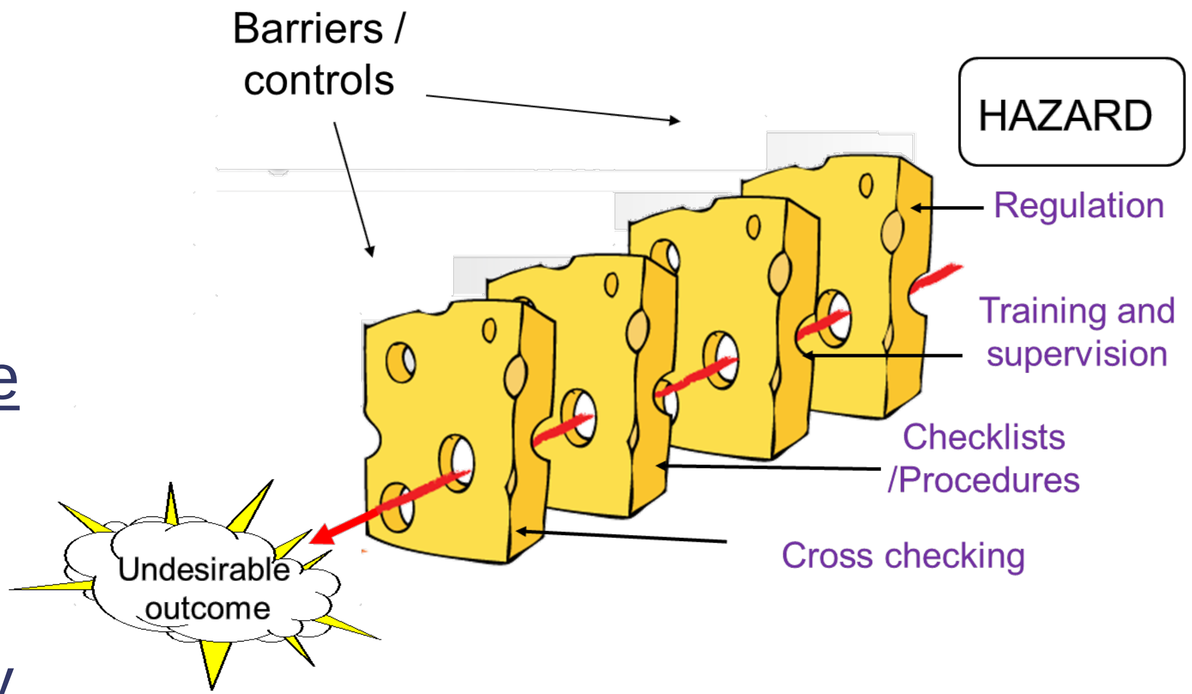
Module Overview

- Gain an appreciation of what a barrier model is
- How a barrier model fits into the overall regulatory and SMS framework
- Understand barrier modelling methodologies
- Identify the different elements of barrier modelling and how they fit together
- Understand the differences between critical and non-critical controls
- Visualise the concept of barrier efficacy
- Identify the importance of a barrier model as a live document that requires constant review

What is a Barrier Model

A barrier model provides a framework through which you can identify defences in your safety system and quantify their respective effectiveness

- The Barrier model provides an established method to assess the probability of progression to an accident
- Barriers <u>are not always 100% effective</u> - holes in the system
- Focus on improving performance of barriers by initiating appropriate safety actions

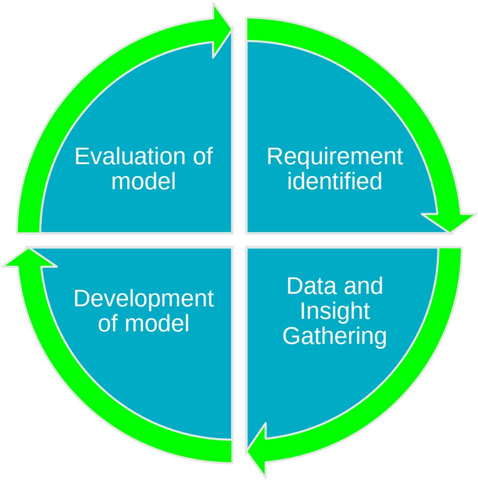


The Lifecycle of a Barrier Model

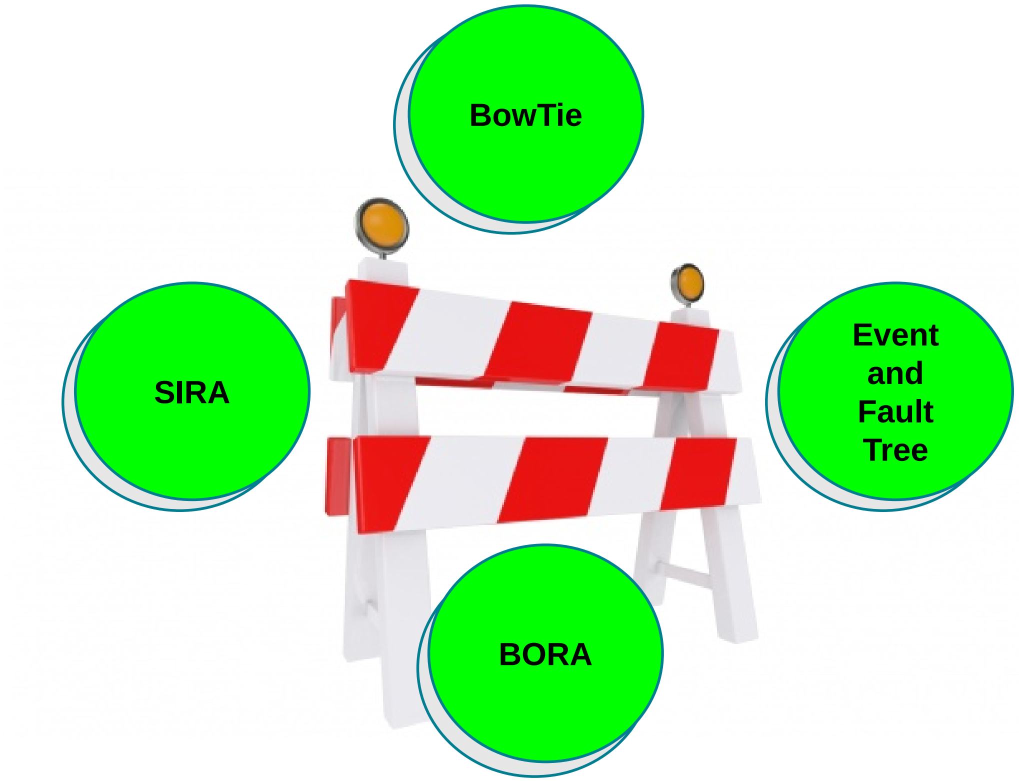
Similar in many ways to an effective SMS, your barrier models should be created and maintained as part of a never-ending improvement cycle

A simple SMS cycle is presented above and enable you to identify hazards/risks and take/measure actions to manage them to an acceptable level

The lifecycle of a barrier model is very similar and should be maintained as part of a cycle that keep the model updated and current



Examples of Barrier Models



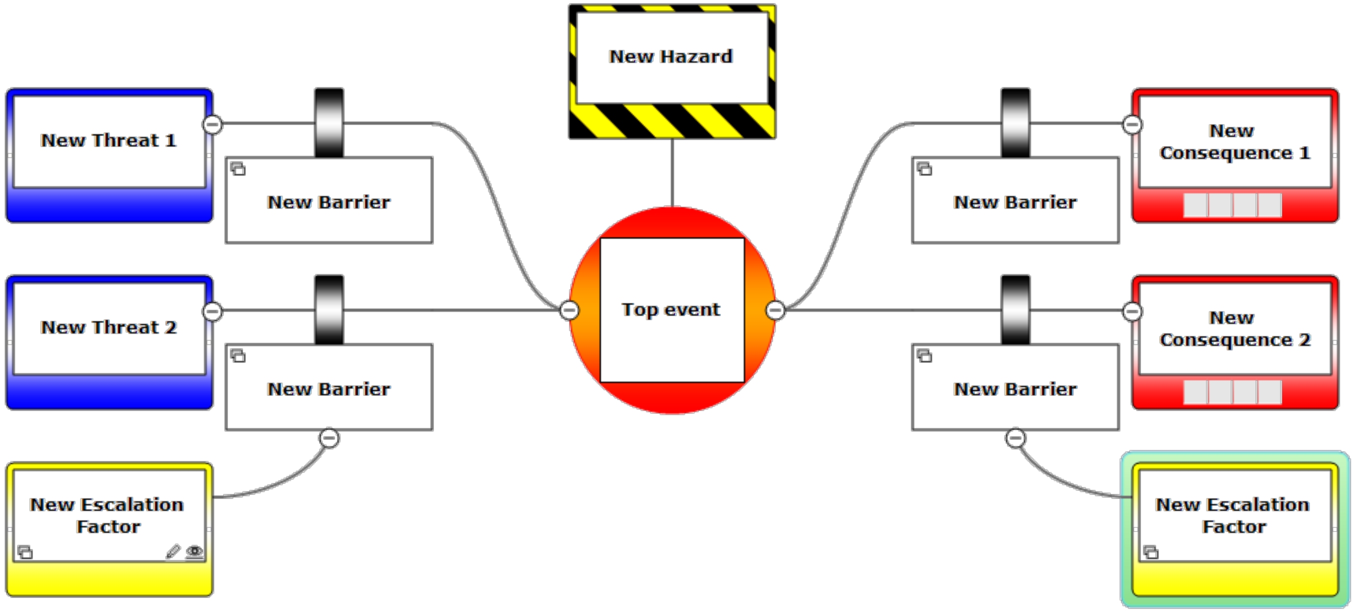
There are many different types of barrier modelling that you can use, and it really comes down to which methodology you think will work most effectively for your business

Although visually these methodologies can be very different, they all share similar components

At the UK CAA we use Bowtie modelling extensively and more information can be found in CAP1329

BowTie Model

Similar to SIRA, the BowTie model visualises risk along a timeline starting from the cause and ending in the effect. Each line on the BowTie is a threat line upon which barriers can be placed.



Key Components of a Barrier Model

Regardless of how you want to visualise your barrier model they all share the same 6 basic components that we will now explain in more detail

Hazard

Consequences

Event

Barriers

Threats

Escalation Factors

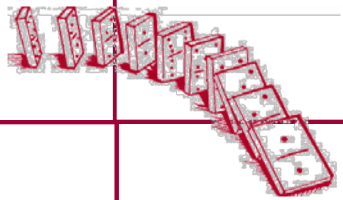
Key Components of a Barrier Model

Hazard

The ever-present object or condition that is involved in most if not all activities we participate in everyday

Top Event

What the hazard could lead to



Threat

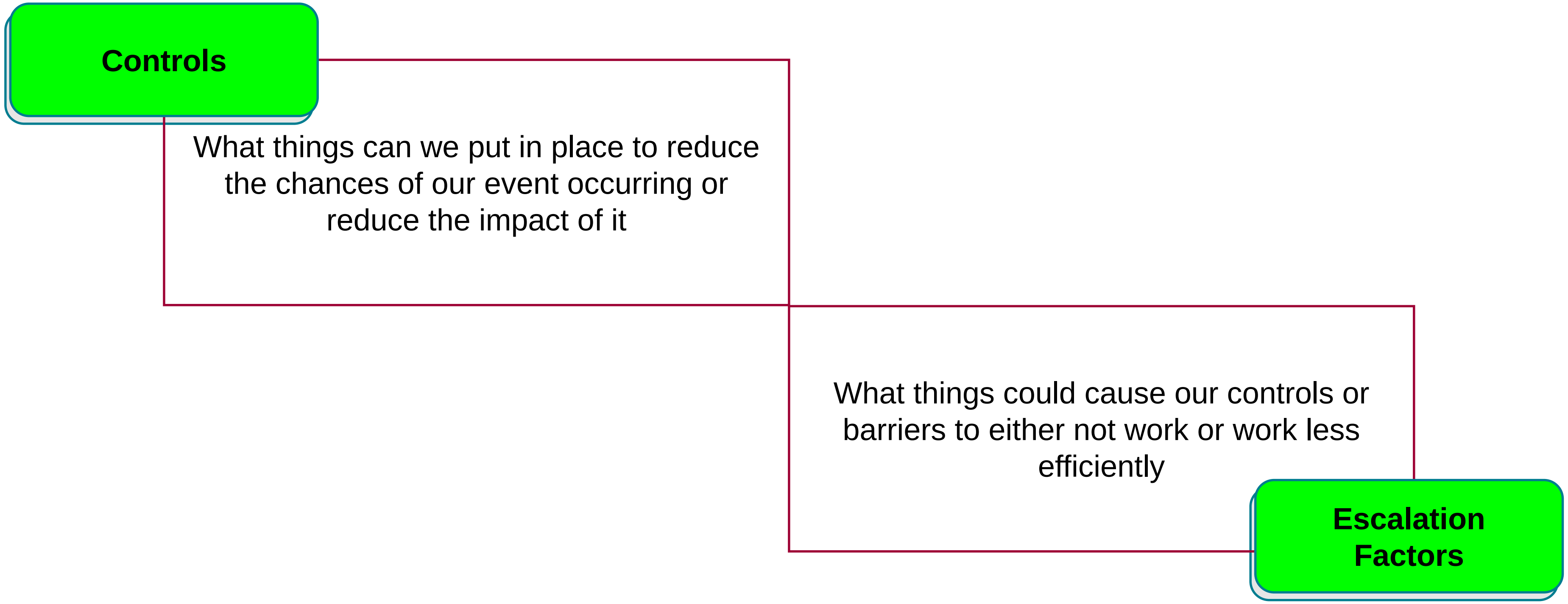
The thing that could lead to the event becoming a reality

Consequence

What could happen if the top event were to happen

Can you spot what is missing?

Key Components of a Barrier Model (2)



When to Create a Barrier Model

When assessing or reassessing risk areas

When trying to understand existing risks in your SMS

When trying to identify potential weaknesses in your system

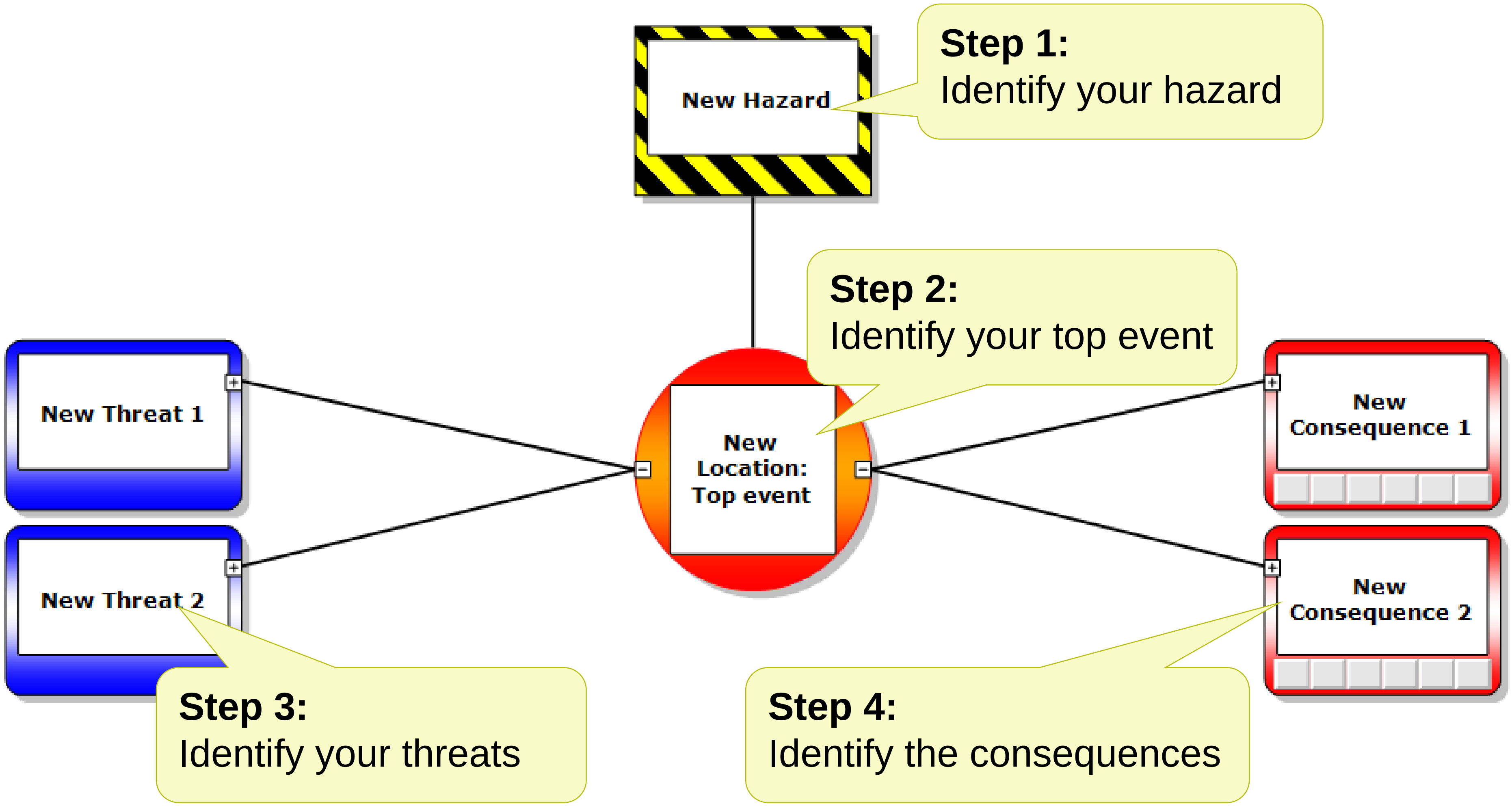
To define and monitor associated safety actions

Building a Bowtie Model – Key Steps

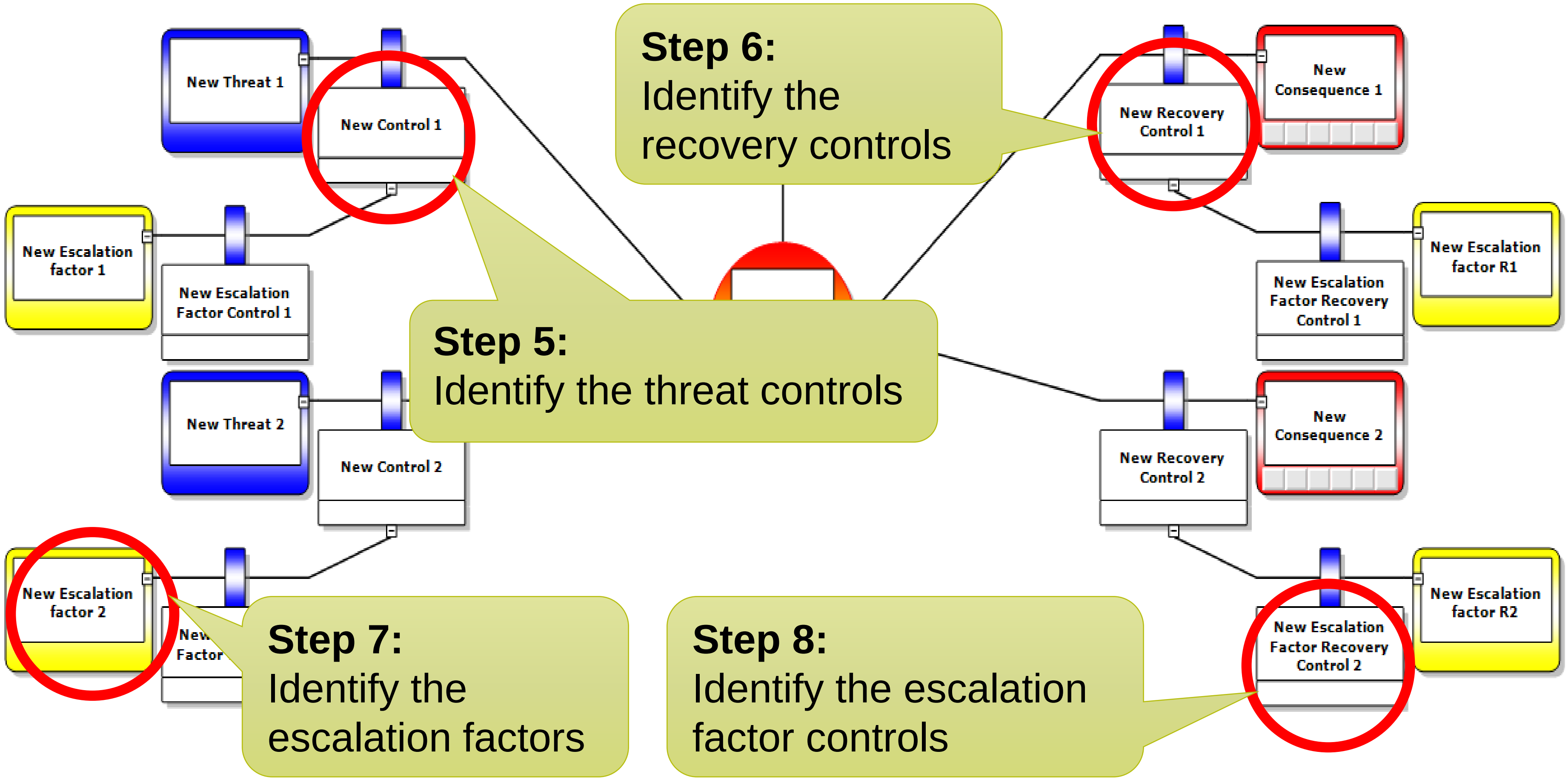
Building a Bowtie model can be done in 8 key steps:

1. Identify Hazard
2. Identify Top Event
3. Identify Threats
4. Assess Consequences
5. Identify Controls
6. Identify Recovery Controls
7. Identify Escalation Factors
8. Identify Escalation Factors Controls

Bowtie Modelling Guidelines



Bowtie Modelling Guidelines



Hints and Tips on Bowties

- The perfect Bowtie does not exist, there is more than one approach
- The level of detail depends on the size of the organisation and the intended use of it (communication or risk management) and the targeted audience
- It is recommended that representatives of all disciplines within an organisation participate in the construction
- It is normal to start with lots of threats and consequences and reduce them as you develop the model
- Once you identify the hazard, the top events and the threats the rest is easier

Safety Performance Indicators

Go to

www.menti.com

Enter the code

3872 4589

Or use QR code

What is an Insight?

An insight is a piece of information uncovered or identified through our day to day activities, activities such as analysis or stakeholder engagement that further develops or progresses our understanding of a given topic.

If you think about any of your analysis you will start to see that insights have a depth to them, a depth we need to factor into our thinking. These are:

Visual

- Level 1 – Visual, observation (increasing/decreasing numbers, trends, it's doing this, it's happening there, …)

Interpretation

- Level 2 – Interpretation, judgement, suggestion, indication (using knowledge and experience, it could be, looks to be…)

Confirmation

- Level 3 – Confirmation, understanding, validation (Comprehension, this is why it is or isn't, its known and controlled)

Actionable Intelligence

Actionable Intelligence; something becomes ‘actionable’ when it can be used as a reason for doing something.

Supporting Aviation Data Driven Decision making (A3DM)

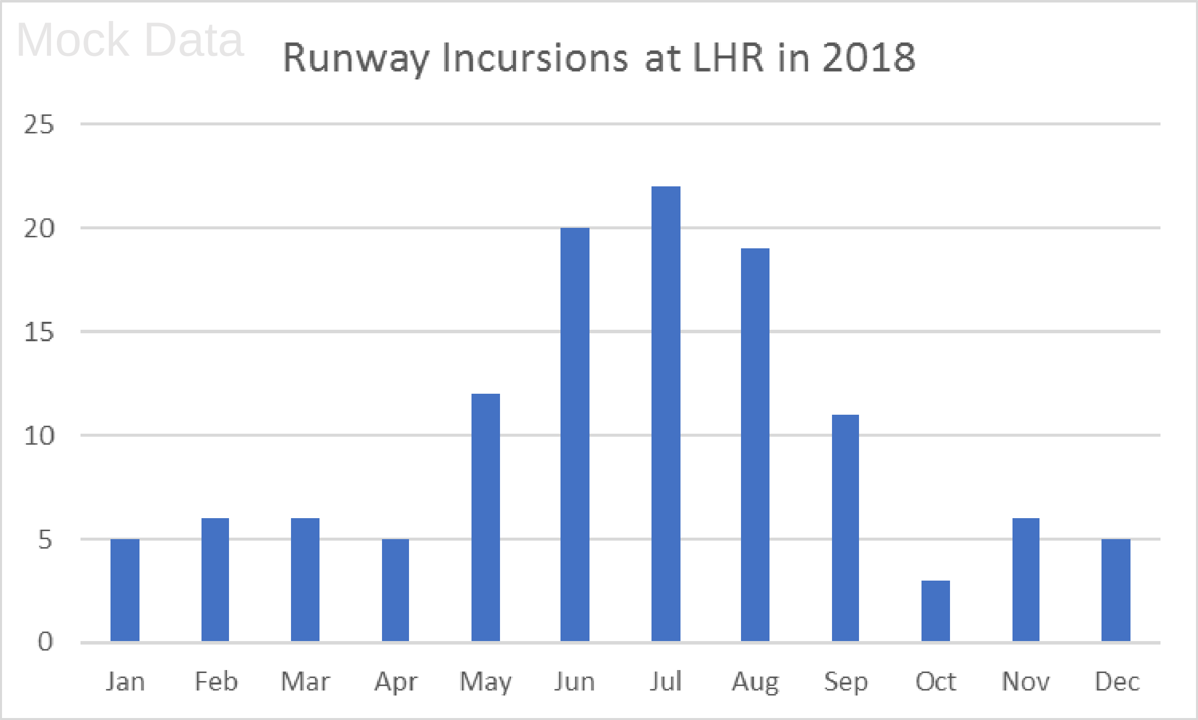
The best Insights lead to the best decisions

We should always aim for confirmatory Insights and attempt to answer the “So What?” question.

Insight example

Not much time - Level 1 Insight, A visual observation

The Analysis Task: ‘*How many Runway Incursions were there at LHR last year?*’



Insight

There is an increase in runway incursions at LHR in 2018 over previous years events.

Suggested Action

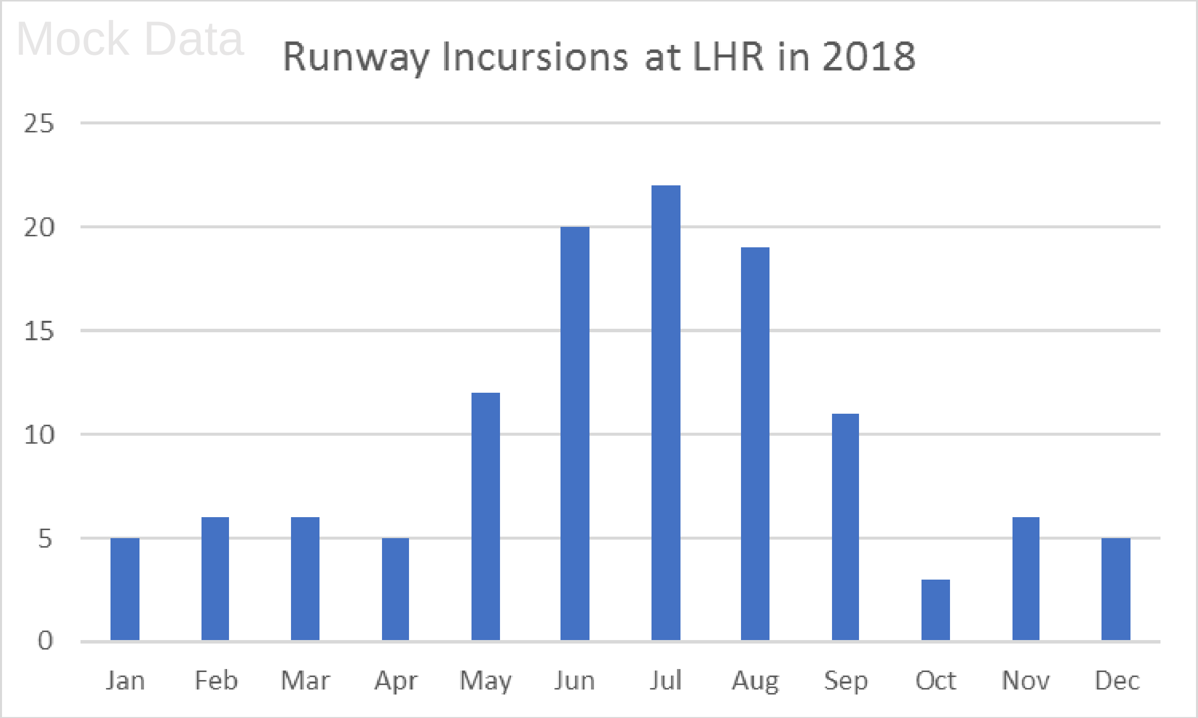
Review the reports for causes

The Reported Result: *‘There were 120 Runway Incursions at LHR in 2018’*

The Response: ‘What’s this telling me, do I need to be concerned?’

Insight example

Moderate amount of time - Level 2 Insight, An interpretation of available information



Insight

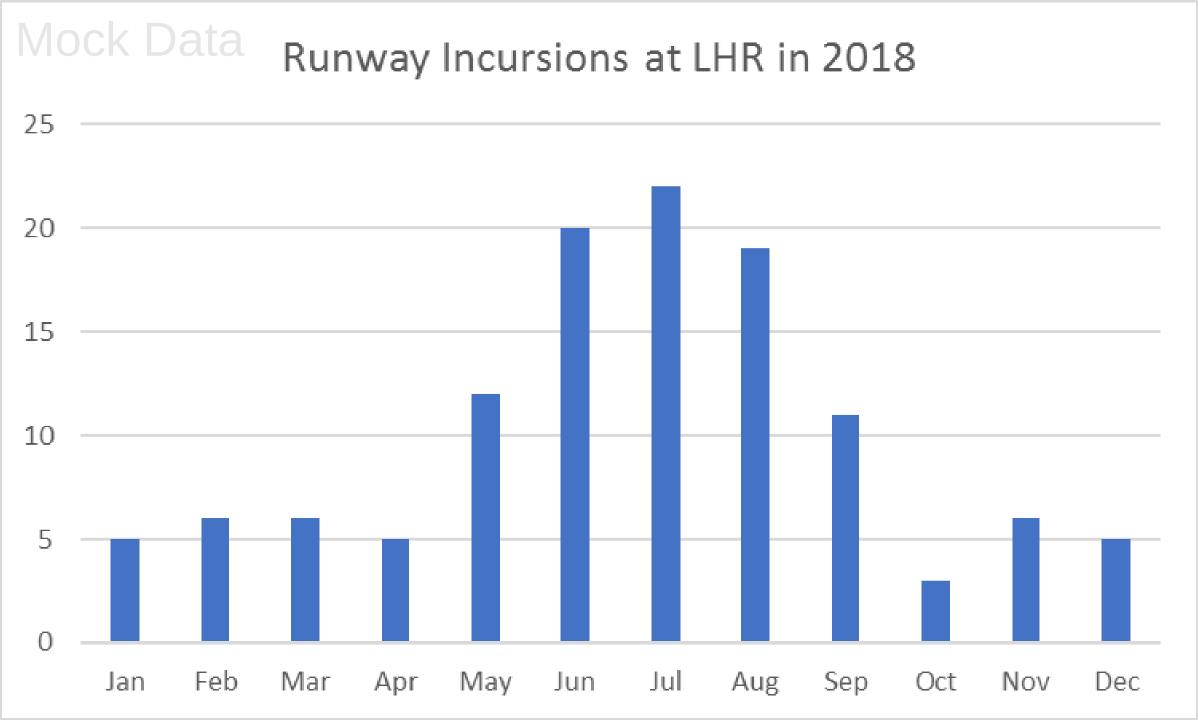
The increase of runway incursions at LHR in 2018 over previous years may be linked to a recent change in ground operating procedures due to construction on the airfield.

Suggested Action

Seek further information to confirm interpretation

Insight example

Plenty of time - Level 3 Insight – A validated and confirmed Actionable Insight.



Insight

The increase of runway incursions at LHR in 2018 over previous years, has been confirmed from feedback via Service Providers as being attributable to the construction taking place near terminal 3 causing aircraft to take different routes to and from the terminal.

Suggested Action

Review change management, WIP procedures and communication

Insight Evolution

Visual

Insight

There is an increase in runway incursions at LHR in 2018 over previous years events.

Suggested Action

Review the reports for causes

Interpretive

Insight

The increase of runway incursions at LHR in 2018 over previous years may be linked to a recent change in ground operating procedures due to construction on the airfield.

Suggested Action

Seek further information to confirm interpretation

Confirmatory

Insight

The increase of runway incursions at LHR in 2018 over previous years, has been confirmed from feedback via Service Providers as being attributable to the construction taking place near terminal 3 causing aircraft to take different routes to and from the terminal.

Suggested Action

Review change management, WIP procedures and communication