Business Understanding

Our company is expanding into the shipping industry, necessitating purchasing and operating airplanes for shipping cargo purposes. Since we are unfamiliar with the potential risks of different types of aircraft, here we are looking to determine which aircraft present the lowest risk in order to begin the new shipping endeavor.

Data Understanding

Here we are working with a dataset provided by the National Transportation Safety Board which includes aviation accident data from 1962 through 2003 about civil aviation accidents and selected incidents in the United States and international waters. Every incident has a unique event ID, and the data files provide the dates and types of each event, as well as other pertinent safety information (e.g. aircraft make and model, number of injuries).

In [22]:
```
import pandas as pd
import numpy as np
```

In [23]:
```
data = pd.read_csv('./data/Aviation_Data.csv', low_memory=False)
```

In [52]:
```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 90348 entries, 0 to 90347
Data columns (total 31 columns):
 #   Column                  Non-Null Count  Dtype  
---  ------                  --------------  -----  
 0   Event.Id                88889 non-null  object 
 1   Investigation.Type      90348 non-null  object 
 2   Accident.Number         88889 non-null  object 
 3   Event.Date              88889 non-null  object 
 4   Location                88837 non-null  object 
 5   Country                 88663 non-null  object 
 6   Latitude                34382 non-null  object 
 7   Longitude               34373 non-null  object 
 8   Airport.Code            50249 non-null  object 
 9   Airport.Name            52790 non-null  object 
 10  Injury.Severity         87889 non-null  object 
 11  Aircraft.damage         85695 non-null  object 
 12  Aircraft.Category       32287 non-null  object 
 13  Registration.Number     87572 non-null  object 
 14  Make                    88826 non-null  object 
 15  Model                   88797 non-null  object 
 16  Amateur.Built           88787 non-null  object 
 17  Number.of.Engines       82805 non-null  float64
 18  Engine.Type             81812 non-null  object 
 19  FAR.Description         32023 non-null  object 
 20  Schedule                12582 non-null  object 
 21  Purpose.of.flight       82697 non-null  object 
 22  Air.carrier             16648 non-null  object 
 23  Total.Fatal.Injuries    77488 non-null  float64
 24  Total.Serious.Injuries  76379 non-null  float64
 25  Total.Minor.Injuries    76956 non-null  float64
 26  Total.Uninjured         82977 non-null  float64
 27  Weather.Condition       84397 non-null  object 
 28  Broad.phase.of.flight   61724 non-null  object 
 29  Report.Status           82508 non-null  object 
 30  Publication.Date        73659 non-null  object 
dtypes: float64(5), object(26)
memory usage: 21.4+ MB
```

In [4]:
```
data.head()
```

Out[4]:

	Event.Id	Investigation.Type	Accident.Number	Event.Date	Location	Country	
0	20001218X45444	Accident	SEA87LA080	1948-10-24	MOOSE CREEK, ID	United States	
1	20001218X45447	Accident	LAX94LA336	1962-07-19	BRIDGEPORT, CA	United States	
2	20061025X01555	Accident	NYC07LA005	1974-08-30	Saltville, VA	United States	3(
3	20001218X45448	Accident	LAX96LA321	1977-06-19	EUREKA, CA	United States	
4	20041105X01764	Accident	CHI79FA064	1979-08-02	Canton, OH	United States	

5 rows × 31 columns

In [24]:
```
data['Event Date'] = pd.to_datetime(data['Event.Date'])
data['Event Date'].describe()
```

```
<ipython-input-24-ab274c447368>:2: FutureWarning: Treating datetime d
ata as categorical rather than numeric in `.describe` is deprecated a
nd will be removed in a future version of pandas. Specify `datetime_i
s_numeric=True` to silence this warning and adopt the future behavior
now.
  data['Event Date'].describe()
```

Out[24]:
```
count                   88889
unique                  14782
top       1982-05-16 00:00:00
freq                       25
first     1948-10-24 00:00:00
last      2022-12-29 00:00:00
Name: Event Date, dtype: object
```

In [117]:
```
data['Make'].value_counts()
```

Out[117]:
```
Cessna                          9106
Piper                           4078
CESSNA                          2117
Beech                           1689
Bell                            1564
                                ...
Goodyear Aerospace                 1
Gulfstream American Corp           1
Sorrel                             1
Zlin                               1
CIRRUS DESIGN CORPORATION          1
Name: Make, Length: 1055, dtype: int64
```

In [7]:
```
1 data['Injury.Severity'].value_counts()
```

Out[7]:
```
Non-Fatal     67357
Fatal(1)       6167
Fatal          5262
Fatal(2)       3711
Incident       2219
               ...
Fatal(121)        1
Fatal(123)        1
Fatal(96)         1
Fatal(111)        1
Fatal(49)         1
Name: Injury.Severity, Length: 109, dtype: int64
```

In [8]:
```
1 data['Number.of.Engines'].value_counts()
```

Out[8]:
```
1.0    69582
2.0    11079
0.0     1226
3.0      483
4.0      431
8.0        3
6.0        1
Name: Number.of.Engines, dtype: int64
```

In [9]:
```
1 data['Investigation.Type'].value_counts()
```

Out[9]:
```
Accident      85015
Incident       3874
25-09-2020      702
26-09-2020       60
02-02-2021       39
               ...
03-11-2020        1
31-03-2021        1
05-01-2021        1
04-03-2021        1
05-08-2022        1
Name: Investigation.Type, Length: 71, dtype: int64
```

In [10]:
```
1 data['Aircraft.Category'].value_counts()
```

Out[10]:
```
Airplane             27617
Helicopter            3440
Glider                 508
Balloon                231
Gyrocraft              173
Weight-Shift           161
Powered Parachute       91
Ultralight              30
Unknown                 14
WSFT                     9
Powered-Lift             5
Blimp                    4
UNK                      2
ULTR                     1
Rocket                   1
Name: Aircraft.Category, dtype: int64
```

In [11]:
```
1 data['Amateur.Built'].value_counts()
```

Out[11]:
```
No     80312
Yes     8475
Name: Amateur.Built, dtype: int64
```

In [12]:
```
data['FAR.Description'].value_counts()
```

Out[12]:
```
091                                  18221
Part 91: General Aviation             6486
NUSN                                  1584
NUSC                                  1013
137                                   1010
135                                    746
121                                    679
Part 137: Agricultural                 437
UNK                                    371
Part 135: Air Taxi & Commuter          298
PUBU                                   253
129                                    246
Part 121: Air Carrier                  165
133                                    107
Part 129: Foreign                      100
Non-U.S., Non-Commercial                97
Non-U.S., Commercial                    93
Part 133: Rotorcraft Ext. Load          32
Unknown                                 22
Public Use                              19
091K                                    14
ARMF                                     8
125                                      5
Part 125: 20+ Pax,6000+ lbs              5
107                                      4
103                                      2
Public Aircraft                          2
Armed Forces                             1
Part 91F: Special Flt Ops.               1
437                                      1
Part 91 Subpart K: Fractional            1
Name: FAR.Description, dtype: int64
```

In [13]:
```
data['Schedule'].value_counts()
```

Out[13]:
```
NSCH    4474
UNK     4099
SCHD    4009
Name: Schedule, dtype: int64
```

In [14]:
```
1 data['Report.Status'].value_counts()
```

Out[14]:
```
Probable Cause
61754
Foreign
1999
<br /><br />
167
Factual
145
The pilot's failure to maintain directional control during the landin
g roll.
58

...
A partial loss of engine power during takeoff due to failure of the l
eft magneto. Contributing to the accident was unsuitable terrain to c
onduct a forced landing.
1
The pilot's failure to maintain directional control during the takeof
f roll.  Contributing to the accident was the pilot's inadvertent use
of the toe brake.
1
The pilot misjudged his altitude which resulted in the airplane conta
cting a wire during the low altitude aerial application maneuver.
1
The pilot's failure to maintain adequate airspeed while turning onto
the final approach leg of the traffic pattern, which led to the airpl
ane exceeding its critical angle-of-attack and experiencing an aerody
namic stall.        1
The aircraft's encounter with deep snow during the landing roll out.
Contributing to the accident was the airport manager's failure to upd
ate the AWOS recording to reflect the closure of the runway due to sn
ow.                 1
Name: Report.Status, Length: 17007, dtype: int64
```

In [15]:
```
1 data['Aircraft.damage'].value_counts()
```

Out[15]:
```
Substantial    64148
Destroyed      18623
Minor           2805
Unknown          119
Name: Aircraft.damage, dtype: int64
```

In [16]:
```
data['Engine.Type'].value_counts()
```

Out[16]:
```
Reciprocating      69530
Turbo Shaft         3609
Turbo Prop          3391
Turbo Fan           2481
Unknown             2051
Turbo Jet            703
None                  19
Geared Turbofan       12
Electric              10
LR                     2
NONE                   2
UNK                    1
Hybrid Rocket          1
Name: Engine.Type, dtype: int64
```

In [17]:
```
data['Weather.Condition'].value_counts()
```

Out[17]:
```
VMC    77303
IMC     5976
UNK      856
Unk      262
Name: Weather.Condition, dtype: int64
```

In [18]:
```
data['Broad.phase.of.flight'].value_counts()
```

Out[18]:
```
Landing        15428
Takeoff        12493
Cruise         10269
Maneuvering     8144
Approach        6546
Climb           2034
Taxi            1958
Descent         1887
Go-around       1353
Standing         945
Unknown          548
Other            119
Name: Broad.phase.of.flight, dtype: int64
```

In [19]:
```
data['Number.of.Engines'].value_counts()
```

Out[19]:
```
1.0    69582
2.0    11079
0.0     1226
3.0      483
4.0      431
8.0        3
6.0        1
Name: Number.of.Engines, dtype: int64
```

The dataset includes records from 1982 through 2002. It includes a wide variety of aircraft types, makes and models. Most injuries are non-fatal. While some aircraft have 2+ engines, a significant majority have only one engine. Exploring all data to help determine its relevance to the business problem.

Data Preparation

I make the data easier to work with by dropping unnecessary columns which contain irrelevant information, as well as records/rows relating to obviously irrelevant incidents (ie. since we are interested in the safest airplanes only, we are removing ameteur built aircraft as well as things like helicopters, parachutes, etc.)

In [25]:
```
, 'Air.carrier', 'Report.Status', 'Publication.Date', 'Broad.phase.of.
```

In [21]:
```
data.head()
```

Out[21]:

	Event.Date	Injury.Severity	Aircraft.damage	Aircraft.Category	Make	Model	Amateur.Built
0	1948-10-24	Fatal(2)	Destroyed	NaN	Stinson	108-3	No
1	1962-07-19	Fatal(4)	Destroyed	NaN	Piper	PA24-180	No
2	1974-08-30	Fatal(3)	Destroyed	NaN	Cessna	172M	No
3	1977-06-19	Fatal(2)	Destroyed	NaN	Rockwell	112	No
4	1979-08-02	Fatal(1)	Destroyed	NaN	Cessna	501	No

In [26]:
```
data.drop(data[data['Amateur.Built'] == 'Yes'].index, inplace=True
```

In [27]:
```
data.drop(data[data['Purpose.of.flight'] == 'Personal'].index, inp
```

In [24]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 39866 entries, 5 to 90344
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              38407 non-null  object
 1   Injury.Severity         37424 non-null  object
 2   Aircraft.damage         35580 non-null  object
 3   Aircraft.Category       14122 non-null  object
 4   Make                    38361 non-null  object
 5   Model                   38343 non-null  object
 6   Amateur.Built           38325 non-null  object
 7   Number.of.Engines       33935 non-null  float64
 8   Engine.Type             33589 non-null  object
 9   Purpose.of.flight       32271 non-null  object
 10  Total.Fatal.Injuries    33365 non-null  float64
 11  Total.Serious.Injuries  32970 non-null  float64
 12  Total.Minor.Injuries    33142 non-null  float64
 13  Total.Uninjured         36070 non-null  float64
 14  Weather.Condition       34484 non-null  object
 15  Event Date              38407 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(10)
memory usage: 5.2+ MB
```

In [25]:
```
data.head()
```

Out[25]:

	Event.Date	Injury.Severity	Aircraft.damage	Aircraft.Category	Make	Model	Amateur.B
5	1979-09-17	Non-Fatal	Substantial	Airplane	Mcdonnell Douglas	DC9	
8	1982-01-01	Non-Fatal	Substantial	Airplane	Cessna	401B	
20	1982-01-02	Non-Fatal	Substantial	Airplane	Cessna	152	
22	1982-01-02	Non-Fatal	Substantial	Helicopter	Bell	206L-1	
25	1982-01-03	Fatal(8)	Destroyed	Airplane	Cessna	414A	

In [28]:
```
non_airplanes = ['Helicopter', 'Glider', 'Balloon', 'Gyrocraft', '
data = data[~data['Aircraft.Category'].isin(non_airplanes)]
```

In [29]:
```
data['Make'] = data['Make'].str.lower()
```

In [217]:
```
1 data['Make'].value_counts()
```

Out[217]:
```
cessna                    11247
piper                      5101
boeing                     2439
beech                      2067
bell                       1599
                          ...
firefly balloon, inc.         1
c a tecnam srl                1
moyes                         1
eiriavion oy                  1
bell-campbell                 1
Name: Make, Length: 869, dtype: int64
```

In [12]:
```
1 data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 36836 entries, 5 to 90344
Data columns (total 16 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              35377 non-null  object
 1   Injury.Severity         34464 non-null  object
 2   Aircraft.damage         32679 non-null  object
 3   Aircraft.Category       11092 non-null  object
 4   Make                    35332 non-null  object
 5   Model                   35313 non-null  object
 6   Amateur.Built           35297 non-null  object
 7   Number.of.Engines       31481 non-null  float64
 8   Engine.Type             31404 non-null  object
 9   Purpose.of.flight       29941 non-null  object
 10  Total.Fatal.Injuries    30659 non-null  float64
 11  Total.Serious.Injuries  30248 non-null  float64
 12  Total.Minor.Injuries    30390 non-null  float64
 13  Total.Uninjured         33156 non-null  float64
 14  Weather.Condition       32007 non-null  object
 15  Event Date              35377 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(10)
memory usage: 4.8+ MB
```

In [30]:
```
1 data.drop(columns=['Aircraft.Category', 'Amateur.Built'], inplace=
```

In [33]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 36836 entries, 5 to 90344
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              35377 non-null  object
 1   Injury.Severity         34464 non-null  object
 2   Aircraft.damage         32679 non-null  object
 3   Make                    35332 non-null  object
 4   Model                   35313 non-null  object
 5   Number.of.Engines       31481 non-null  float64
 6   Engine.Type             31404 non-null  object
 7   Purpose.of.flight       29941 non-null  object
 8   Total.Fatal.Injuries    30659 non-null  float64
 9   Total.Serious.Injuries  30248 non-null  float64
 10  Total.Minor.Injuries    30390 non-null  float64
 11  Total.Uninjured         33156 non-null  float64
 12  Weather.Condition       32007 non-null  object
 13  Event Date              35377 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(8)
memory usage: 4.2+ MB
```

In [34]:
```
data.head()
```

Out[34]:

	Event.Date	Injury.Severity	Aircraft.damage	Make	Model	Number.of.Engines	Engine.T
5	1979-09-17	Non-Fatal	Substantial	Mcdonnell Douglas	DC9	2.0	Turbo
8	1982-01-01	Non-Fatal	Substantial	Cessna	401B	2.0	Reciproca
20	1982-01-02	Non-Fatal	Substantial	Cessna	152	1.0	Reciproca
25	1982-01-03	Fatal(8)	Destroyed	Cessna	414A	2.0	Reciproca
31	1982-01-03	Non-Fatal	Substantial	Air Tractor	AT-301	1.0	Reciproca

Exploratory Data Analysis

In [31]:
```
import matplotlib
import matplotlib.pyplot as plt

%matplotlib inline
```

In [36]:
```
data['Number.of.Engines'].value_counts()
```

Out[36]:
```
1.0    22723
2.0     7570
3.0      477
4.0      416
0.0      293
6.0        1
8.0        1
Name: Number.of.Engines, dtype: int64
```

In [32]:
```
data = data[~data['Number.of.Engines'].isin([0.0, 6.0, 8.0])]
```

In [222]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 36541 entries, 5 to 90344
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              35082 non-null  object
 1   Injury.Severity         34169 non-null  object
 2   Aircraft.damage         32427 non-null  object
 3   Make                    35037 non-null  object
 4   Model                   35018 non-null  object
 5   Number.of.Engines       31186 non-null  float64
 6   Engine.Type             31159 non-null  object
 7   Purpose.of.flight       29647 non-null  object
 8   Total.Fatal.Injuries    30441 non-null  float64
 9   Total.Serious.Injuries  30013 non-null  float64
 10  Total.Minor.Injuries    30165 non-null  float64
 11  Total.Uninjured         32907 non-null  float64
 12  Weather.Condition       31713 non-null  object
 13  Event Date              35082 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(8)
memory usage: 4.2+ MB
```

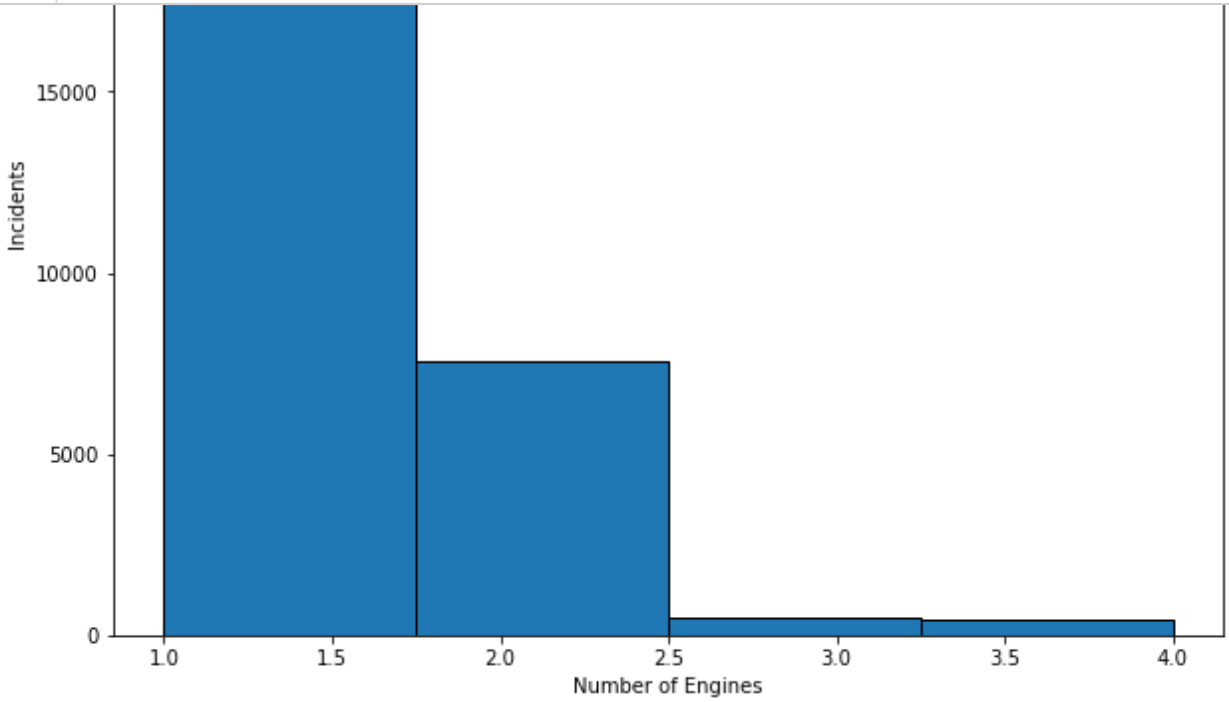
In [223]:
```
data.shape
```

Out[223]:
```
(36541, 14)
```

In [158]:
```
data['Number.of.Engines'].value_counts()
```

Out[158]:
```
1.0    22723
2.0     7570
3.0      477
4.0      416
Name: Number.of.Engines, dtype: int64
```

In [50]:
```
fig, ax = plt.subplots(figsize=(10, 8))
plt.hist(data['Number.of.Engines'], bins=4, edgecolor='black')
ax.set_xlabel("Number of Engines")
ax.set_ylabel("Incidents")
ax.set_title("Incidents by Number of Engines")
plt.savefig('plot1.jpg')
plt.show()
```



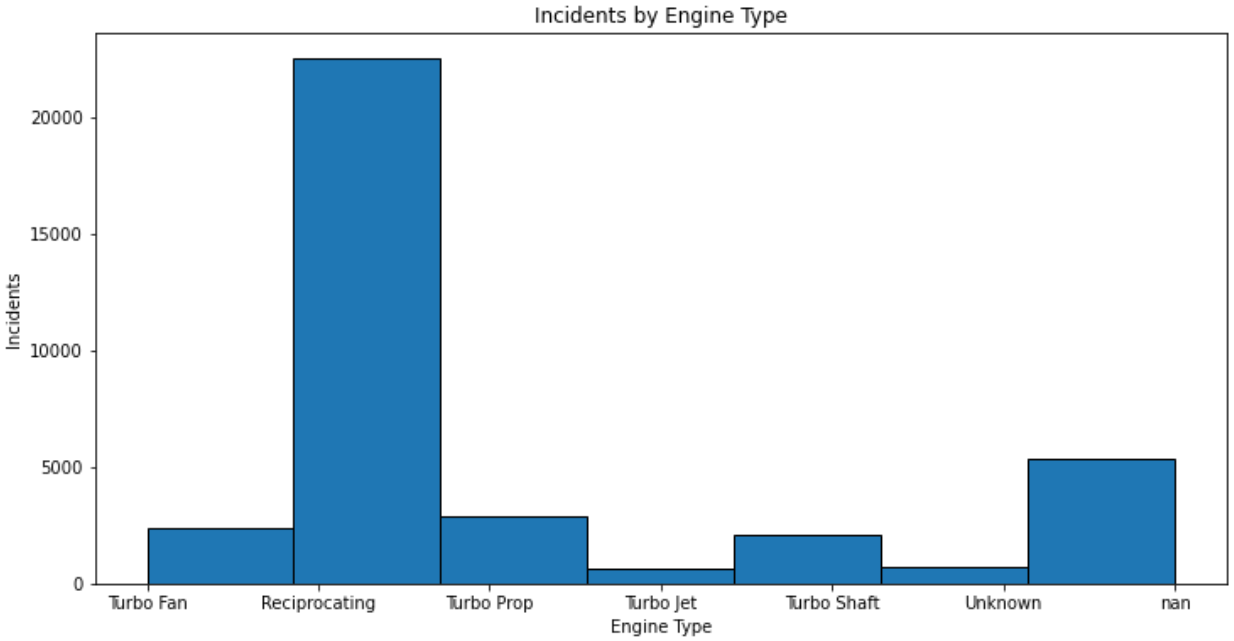
In [161]:
```
data['Engine.Type'].value_counts()
```

Out[161]:
```
Reciprocating      22517
Turbo Prop          2879
Turbo Fan           2378
Turbo Shaft         2076
Unknown              686
Turbo Jet            607
Geared Turbofan       12
Electric               3
UNK                    1
Name: Engine.Type, dtype: int64
```

Type *Markdown* and LaTeX: α^2

In [35]:
```
p(data[data['Engine.Type'].isin(['Geared Turbofan', 'Electric', 'UNK'])
```

In [49]:
```
fig, ax = plt.subplots(figsize=(12, 6))
plt.hist(data['Engine.Type'], bins=7, edgecolor='black')
ax.set_xlabel("Engine Type")
ax.set_ylabel("Incidents")
ax.set_title("Incidents by Engine Type")
plt.savefig('plot2.jpg')
plt.show()
```



In [37]:
```
safest_aircraft = data[data['Number.of.Engines'] != 1.0]
```

In [119]:
```
safest_aircraft.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 13818 entries, 5 to 90344
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              12359 non-null  object
 1   Injury.Severity         11475 non-null  object
 2   Aircraft.damage         9886 non-null   object
 3   Make                    12314 non-null  object
 4   Model                   12302 non-null  object
 5   Number.of.Engines       8463 non-null   float64
 6   Engine.Type             8938 non-null   object
 7   Purpose.of.flight       7748 non-null   object
 8   Total.Fatal.Injuries    10627 non-null  float64
 9   Total.Serious.Injuries  10385 non-null  float64
 10  Total.Minor.Injuries    10282 non-null  float64
 11  Total.Uninjured         11409 non-null  float64
 12  Weather.Condition       9354 non-null   object
 13  Event Date              12359 non-null  datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(8)
memory usage: 1.6+ MB
```

In [164]:
```
safest_aircraft['Make'].value_counts()
```

Out[164]:
```
boeing                          2341
cessna                          1988
piper                           1460
beech                           1374
mcdonnell douglas                486
                                ...
aficionado                         1
md helicopters                     1
stearman                           1
indonesian aerospace               1
aerostar international inc.        1
Name: Make, Length: 435, dtype: int64
```

In [38]:
```
safest_aircraft = safest_aircraft[safest_aircraft['Engine.Type'] !
```

In [166]:
```
safest_aircraft.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10451 entries, 5 to 90344
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              8992 non-null   object
 1   Injury.Severity         8113 non-null   object
 2   Aircraft.damage         6568 non-null   object
 3   Make                    8947 non-null   object
 4   Model                   8935 non-null   object
 5   Number.of.Engines       5248 non-null   float64
 6   Engine.Type             5571 non-null   object
 7   Purpose.of.flight       4676 non-null   object
 8   Total.Fatal.Injuries    7567 non-null   float64
 9   Total.Serious.Injuries  7392 non-null   float64
 10  Total.Minor.Injuries    7297 non-null   float64
 11  Total.Uninjured         8218 non-null   float64
 12  Weather.Condition       6084 non-null   object
 13  Event Date              8992 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(8)
memory usage: 1.2+ MB
```

In [167]:
```
safest_aircraft['Make'].value_counts()
```

Out[167]:
```
boeing                          2334
cessna                           952
beech                            676
mcdonnell douglas                484
piper                            389
                                ...
aero vodochody                     1
zenair                             1
rockwell commander                 1
hawker siddely                     1
aerostar international inc.        1
Name: Make, Length: 389, dtype: int64
```

In [169]:
```
safest_aircraft['Aircraft.damage'].value_counts()
```

Out[169]:
```
Substantial    3118
Destroyed      1740
Minor          1659
Unknown          51
Name: Aircraft.damage, dtype: int64
```

In [39]:
```
t[safest_aircraft['Aircraft.damage'].isin(['Substantial', 'Destroyed'])
```

In [231]:
```
safest_aircraft.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 5581 entries, 79 to 90344
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype
---  ------                  --------------  -----
 0   Event.Date              4122 non-null   object
 1   Injury.Severity         3412 non-null   object
 2   Aircraft.damage         1709 non-null   object
 3   Make                    4094 non-null   object
 4   Model                   4090 non-null   object
 5   Number.of.Engines       2585 non-null   float64
 6   Engine.Type             2556 non-null   object
 7   Purpose.of.flight       1747 non-null   object
 8   Total.Fatal.Injuries    3440 non-null   float64
 9   Total.Serious.Injuries  3564 non-null   float64
 10  Total.Minor.Injuries    3483 non-null   float64
 11  Total.Uninjured         3985 non-null   float64
 12  Weather.Condition       2509 non-null   object
 13  Event Date              4122 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(8)
memory usage: 654.0+ KB
```

In [40]:
```
top_3_makes = safest_aircraft['Make'].value_counts().nlargest(3).i
```

In [41]:

In [41]:

```
safest_aircraft = safest_aircraft[safest_aircraft['Make'].isin(top
print (safest_aircraft)
```

```
      Event.Date Injury.Severity Aircraft.damage    Make      Model
\
320   1982-02-15        Incident           Minor  boeing   B737-2H4
351   1982-02-19        Incident             NaN  boeing  B-727-200
506   1982-03-08        Incident             NaN  boeing   707-131B
796   1982-04-09        Incident           Minor  boeing    727-200
1834  1982-07-05        Incident           Minor  boeing    727-233
...          ...             ...             ...     ...        ...
90285 2022-11-26       Non-Fatal             NaN  boeing    737-800
90308 2022-12-05             NaN             NaN  boeing        737
90310 2022-12-05       Non-Fatal             NaN  boeing    737-800
90314 2022-12-08         Serious             NaN  boeing    767-322
90338 2022-12-18             NaN             NaN  airbus   A330-243

      Number.of.Engines Engine.Type Purpose.of.flight  Total.Fatal.I
njuries  \
320                 2.0   Turbo Jet           Unknown
0.0
351                 3.0   Turbo Jet               NaN
NaN
506                 4.0   Turbo Fan           Unknown
1.0
796                 3.0   Turbo Fan           Unknown
0.0
1834                3.0   Turbo Fan               NaN
NaN
...                 ...         ...             ...
...
90285               NaN         NaN             NaN
0.0
90308               NaN         NaN             NaN
0.0
90310               NaN         NaN             NaN
0.0
90314               NaN         NaN             NaN
0.0
90338               NaN         NaN             NaN
0.0

      Total.Serious.Injuries  Total.Minor.Injuries  Total.Uninjured
\
320                      0.0                   0.0            119.0
351                      NaN                   NaN             83.0
506                      0.0                   0.0              0.0
796                      0.0                   0.0            108.0
1834                     NaN                   NaN             74.0
...                      ...                   ...              ...
90285                    0.0                   0.0            186.0
90308                    0.0                   0.0              0.0
90310                    0.0                   0.0            102.0
90314                    3.0                   1.0            175.0
```

```
90338                    0.0                    0.0              0.0

      Weather.Condition Event Date

320                 IMC 1982-02-15
351                 IMC 1982-02-19
506                 VMC 1982-03-08
796                 IMC 1982-04-09
1834                VMC 1982-07-05
...                 ...        ...
90285               NaN 2022-11-26
90308               NaN 2022-12-05
90310               NaN 2022-12-05
90314               NaN 2022-12-08
90338               NaN 2022-12-18

[2330 rows x 14 columns]
```

In [234]:
```
1 safest_aircraft.info()
```

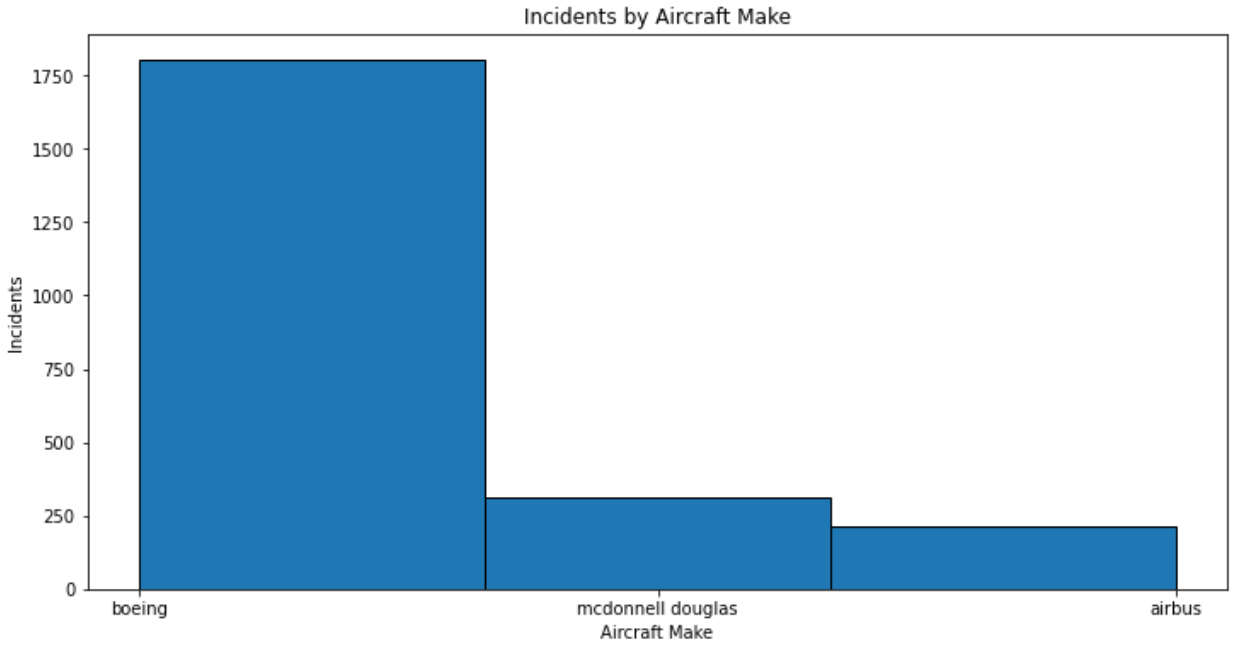
```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 2330 entries, 320 to 90338
Data columns (total 14 columns):
 #   Column                  Non-Null Count  Dtype         
---  ------                  --------------  -----         
 0   Event.Date              2330 non-null   object        
 1   Injury.Severity         1799 non-null   object        
 2   Aircraft.damage         915 non-null    object        
 3   Make                    2330 non-null   object        
 4   Model                   2326 non-null   object        
 5   Number.of.Engines       1441 non-null   float64       
 6   Engine.Type             1348 non-null   object        
 7   Purpose.of.flight       814 non-null    object        
 8   Total.Fatal.Injuries    1960 non-null   float64       
 9   Total.Serious.Injuries  2037 non-null   float64       
 10  Total.Minor.Injuries    1996 non-null   float64       
 11  Total.Uninjured         2237 non-null   float64       
 12  Weather.Condition       1275 non-null   object        
 13  Event Date              2330 non-null   datetime64[ns]
dtypes: datetime64[ns](1), float64(5), object(8)
memory usage: 273.0+ KB
```

In [235]:
```
safest_aircraft.head()
```

Out[235]:

	Event.Date	Injury.Severity	Aircraft.damage	Make	Model	Number.of.Engines	Engine.Ty
320	1982-02-15	Incident	Minor	boeing	B737-2H4	2.0	Turbo
351	1982-02-19	Incident	NaN	boeing	B-727-200	3.0	Turbo
506	1982-03-08	Incident	NaN	boeing	707-131B	4.0	Turbo F
796	1982-04-09	Incident	Minor	boeing	727-200	3.0	Turbo F
1834	1982-07-05	Incident	Minor	boeing	727-233	3.0	Turbo F

In [48]:
```
fig, ax = plt.subplots(figsize=(12, 6))
plt.hist(safest_aircraft['Make'], bins=3, edgecolor='black')
ax.set_xlabel("Aircraft Make")
ax.set_ylabel("Incidents")
ax.set_title("Incidents by Aircraft Make")
plt.savefig('plot3.jpg')
plt.show()
```



In [ ]:
```

```

Conclusions

1) Aircraft with multiple engines are more reliable. Overwhelmingly, aircraft involved in incidents are single engine aircraft. 2) Turbo fan engines appear most reliable, though avoiding reciprocating engines is most crucial when it comes to engine type. 3) Boeing and Airbus made aircraft are safest due to their overall reliability.

Limitations

The dataset outlines "selected" incidents only and does not include information on the total number of aircraft or total flights, leaving room for more precision using a more complete dataset.

Recommendations

1) Aircraft with multiple engines are recommended, as single engine aircraft are overwhelmingly involved in more incidents. 2) Non-reciprocating engines are recommended, as reciprocating engines are involved in more incidents. 3) Boeing and Airbus made aircraft are specifically recommmended for safety.

Next Steps

Evaluating cost and other variables will be crucial, such as ease of service and repair, as well as ability to modify for business purposes, since these considerations will factor into choosing aircraft to purchase.

In [ ]:
```
1
```