In [3]:
```
import csv
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

In [4]:
```
data = pd.read_csv('AviationData.csv', encoding = "latin-1")
data['Location'].value_counts()
data
```

```
/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/IPython/core/in
teractiveshell.py:3145: DtypeWarning: Columns (6,7,28) have mixed types.S
pecify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

Out[4]:

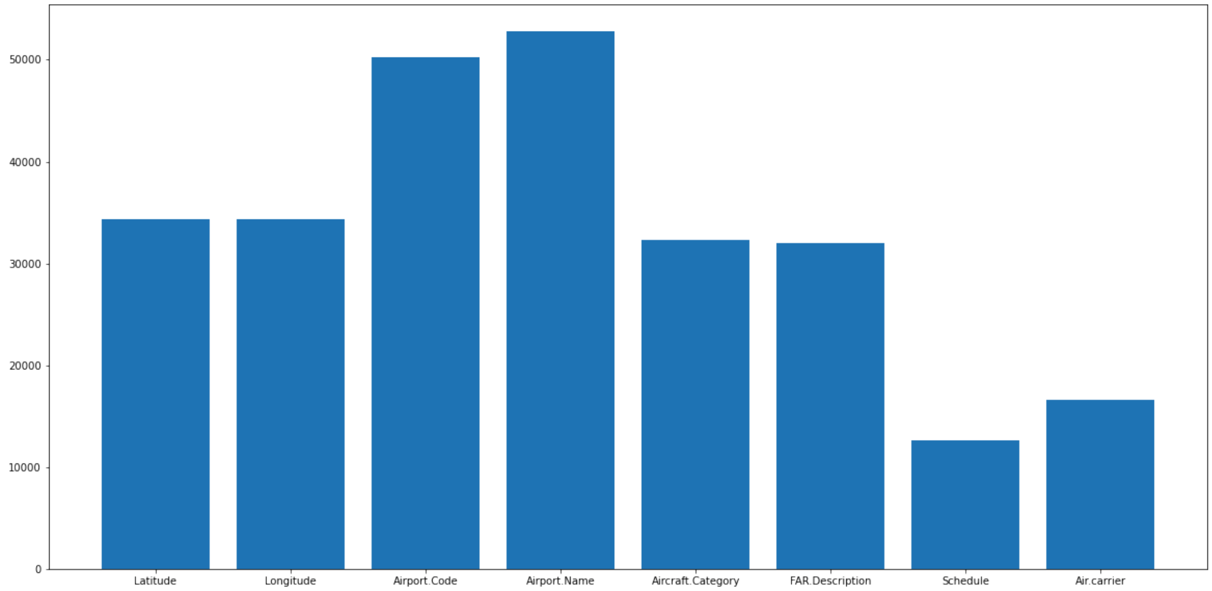
| | Event.Id | Investigation.Type | Accident.Number | Event.Date | Location | Country | Lat |
|---|---|---|---|---|---|---|---|
| **0** | 20001218X45444 | Accident | SEA87LA080 | 1948-10-24 | MOOSE CREEK, ID | United States | |
| **1** | 20001218X45447 | Accident | LAX94LA336 | 1962-07-19 | BRIDGEPORT, CA | United States | |
| **2** | 20061025X01555 | Accident | NYC07LA005 | 1974-08-30 | Saltville, VA | United States | 36 |
| **3** | 20001218X45448 | Accident | LAX96LA321 | 1977-06-19 | EUREKA, CA | United States | |
| **4** | 20041105X01764 | Accident | CHI79FA064 | 1979-08-02 | Canton, OH | United States | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **88884** | 20221227106491 | Accident | ERA23LA093 | 2022-12-26 | Annapolis, MD | United States | |
| **88885** | 20221227106494 | Accident | ERA23LA095 | 2022-12-26 | Hampton, NH | United States | |
| **88886** | 20221227106497 | Accident | WPR23LA075 | 2022-12-26 | Payson, AZ | United States | 341 |
| **88887** | 20221227106498 | Accident | WPR23LA076 | 2022-12-26 | Morgan, UT | United States | |
| **88888** | 20221230106513 | Accident | ERA23LA097 | 2022-12-29 | Athens, GA | United States | |

88889 rows × 31 columns

Plotting the sum of all the non NaN values, where the sum is less than 60,000.

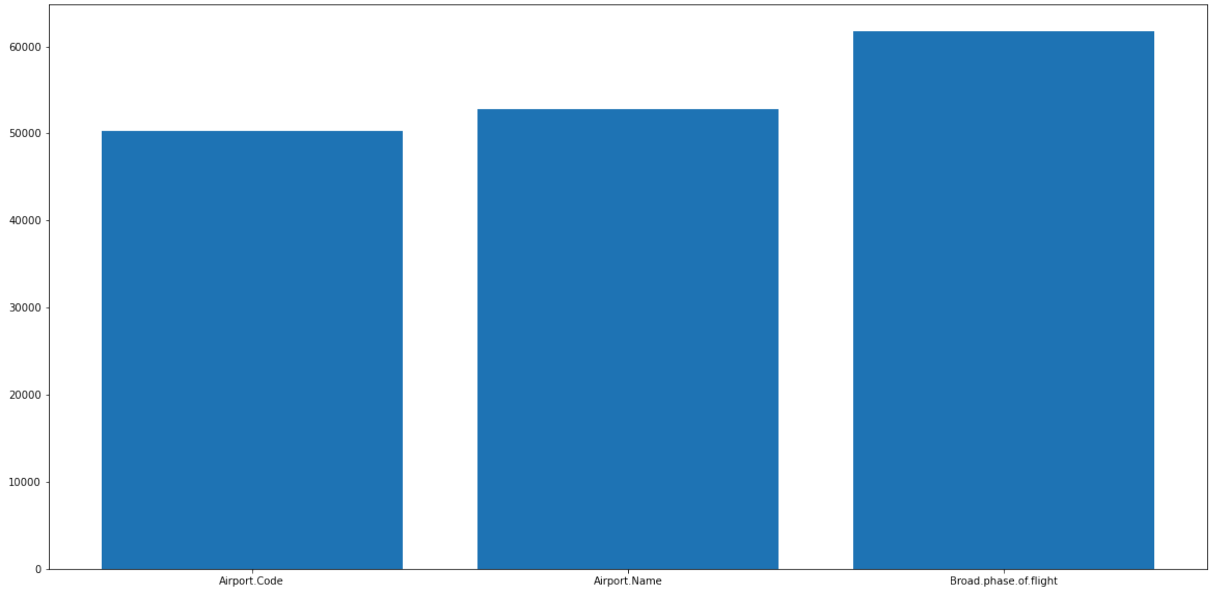
In [5]:

```
x_list= []
y_list = []

for column in data:
    y_list.append((data[column].notna().sum()))
    x_list.append(column)

final_x_list =[]
final_y_list = []

for x, y in zip(x_list, y_list):
    if y < 60000:
        final_x_list.append(x)
        final_y_list.append(y)



fig, ax = plt.subplots(1,1, figsize=(20,10))

ax.bar(final_x_list, final_y_list)
plt.show()
```



In [6]:

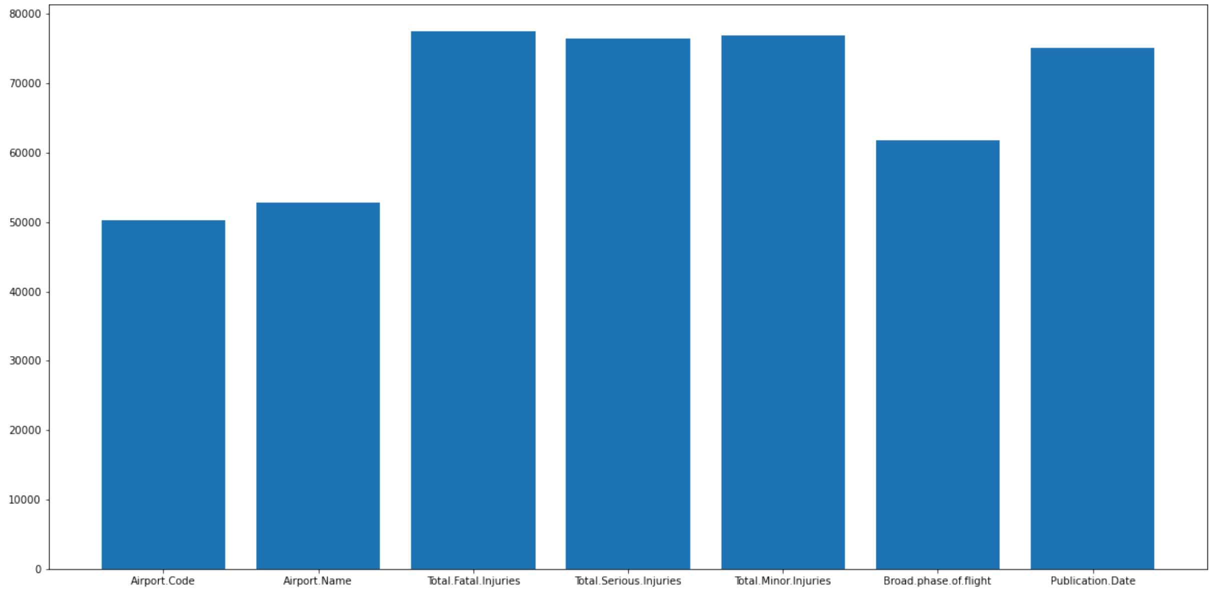
```
x_list= []
y_list = []

for column in data:
    y_list.append((data[column].notna().sum()))
    x_list.append(column)

final_x_list =[]
final_y_list = []

for x, y in zip(x_list, y_list):
    if 50000 < y < 70000:
        final_x_list.append(x)
        final_y_list.append(y)



fig, ax = plt.subplots(1,1, figsize=(20,10))

ax.bar(final_x_list, final_y_list)
plt.show()
```



Plotting the sum of all the non NaN values, where the sum is greater than 50,000 and less than 80,000.

In [7]:

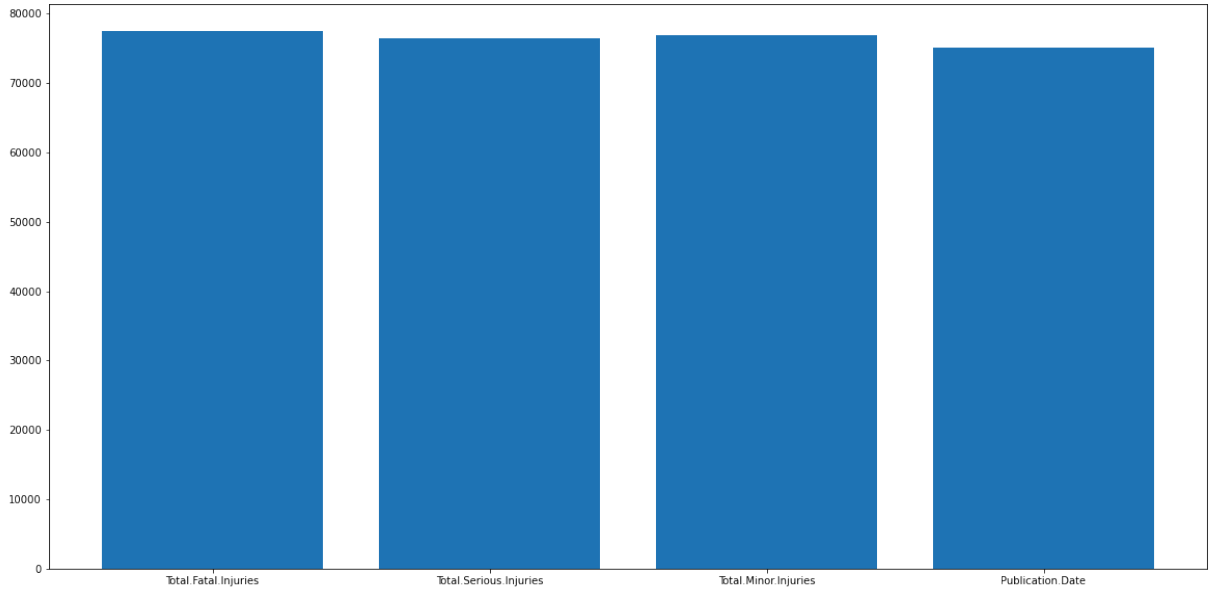
```
x_list= []
y_list = []

for column in data:
    y_list.append((data[column].notna().sum()))
    x_list.append(column)

final_x_list =[]
final_y_list = []

for x, y in zip(x_list, y_list):
    if 50000 < y < 80000:
        final_x_list.append(x)
        final_y_list.append(y)



fig, ax = plt.subplots(1,1, figsize=(20,10))

ax.bar(final_x_list, final_y_list)
plt.show()
```



Plotting the sum of all the non NaN values, where the sum is greater than 70,000 and less than 80,000.

In [163]:

```
x_list= []
y_list = []

for column in data:
    y_list.append((data[column].notna().sum()))
    x_list.append(column)

final_x_list =[]
final_y_list = []

for x, y in zip(x_list, y_list):
    if 70000 < y <= 80000:
        final_x_list.append(x)
        final_y_list.append(y)



fig, ax = plt.subplots(1,1, figsize=(20,10))

ax.bar(final_x_list, final_y_list)
plt.show()
```



In [8]:

```
data['Broad.phase.of.flight'].value_counts(ascending = False)
```

Out[8]:

```
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 [9]:
```
print(data['Weather.Condition'].unique())
```

```
['UNK' 'IMC' 'VMC' nan 'Unk']
```

In [10]:
```
print(data['Air.carrier'].unique())
```

```
[nan 'Air Canada' 'Rocky Mountain Helicopters, In' ...
 'SKY WEST AVIATION INC TRUSTEE' 'GERBER RICHARD E' 'MC CESSNA 210N LLC']
```

In [11]:
```
print(len(data['Air.carrier'].unique()))
```

```
13591
```

In [126]:
```
print(data['Air.carrier'].value_counts())
```

```
Pilot                                                 258
American Airlines                                      90
United Airlines                                        89
Delta Air Lines                                        53
SOUTHWEST AIRLINES CO                                  42
                                                     ... 
Chemtec Aviation LLC                                    1
High Adventur Air Charter, Guides & Outfitters Inc      1
FAA FLYING CLUB INC                                     1
KEITER KENNETH A                                        1
ADAPTIVE AERO INC                                       1
Name: Air.carrier, Length: 13590, dtype: int64
```

In [160]:

```
personal_df = data[data['Purpose.of.flight']=='Personal']
personal_df[:25]
```

Out[160]:

| vent.Date | Location | Country | Latitude | Longitude | Airport.Code | Airport.Name | ... | Purpose.of. |
|---|---|---|---|---|---|---|---|---|
| 1948-10-24 | MOOSE CREEK, ID | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1962-07-19 | BRIDGEPORT, CA | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1974-08-30 | Saltville, VA | United States | 36.9222 | -81.8781 | NaN | NaN | ... | Per |
| 1977-06-19 | EUREKA, CA | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1979-08-02 | Canton, OH | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1981-08-01 | COTTON, MN | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-01 | PULLMAN, WA | United States | NaN | NaN | NaN | BLACKBURN AG STRIP | ... | Per |
| 1982-01-01 | JACKSONVILLE, FL | United States | NaN | NaN | JAX | JACKSONVILLE INTL | ... | Per |
| 1982-01-01 | HOBBS, NM | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-01 | TUSKEGEE, AL | United States | NaN | NaN | NaN | TUSKEGEE | ... | Per |
| 1982-01-02 | HOMER, LA | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-02 | HEARNE, TX | United States | NaN | NaN | T72 | HEARNE MUNICIPAL | ... | Per |
| 1982-01-02 | CHICKASHA, OK | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-02 | LITTLE ROCK, AR | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-02 | MIDWAY, UT | United States | NaN | NaN | NaN | FIELD RANCH | ... | Per |
| 1982-01-02 | SKWENTA, AK | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-02 | GALETON, PA | United States | NaN | NaN | 5G6 | CHERRY SPRINGS | ... | Per |
| 1982-01-02 | MIAMI, FL | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-02 | CHARLOTTE, MI | United States | NaN | NaN | 49G | TINKERBELL | ... | Per |
| 1982-01-03 | VAN NUYS, CA | United States | NaN | NaN | VNY | VAN NUYS | ... | Per |
| 1982-01-03 | 10 NM W LEE VIN, CA | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-03 | PINEHURST, NC | United States | NaN | NaN | SOP | MOORE COUNTY | ... | Per |
| 1982-01-03 | WHITE PLAINS, NY | United States | NaN | NaN | HPN | WESTCHESTER COUNTY | ... | Per |

| vent.Date | Location | Country | Latitude | Longitude | Airport.Code | Airport.Name | ... | Purpose.of. |
|---|---|---|---|---|---|---|---|---|
| 1982-01-03 | COCOA, FL | United States | NaN | NaN | NaN | NaN | ... | Per |
| 1982-01-03 | SAN CLEMENTINE, CA | United States | NaN | NaN | NaN | NaN | ... | Per |

In [127]:
```
new_df = data[data['Air.carrier'] =='Pilot']
new_df
```

Out[127]:

| Investigation.Type | Accident.Number | Event.Date | Location | Country | Latitude | Longitude | Airport.Cod |
|---|---|---|---|---|---|---|---|
| Accident | CHI08CA090 | 2008-03-01 | Apple River, IL | United States | 042303N | 0090521W | Na |
| Accident | CHI08CA095 | 2008-03-25 | North Canton, OH | United States | 405535N | 0812652W | Na |
| Accident | CHI08CA094 | 2008-03-26 | Traverse City, MI | United States | 444429N | 0853556W | TV |
| Accident | CHI08CA116 | 2008-04-27 | Howell, MI | United States | 423817N | 0835932W | 3H |
| Accident | CHI08CA117 | 2008-04-27 | Madison, IN | United States | 394532N | 0852756W | IM |
| ... | ... | ... | ... | ... | ... | ... | .. |
| Accident | CEN23LA012 | 2022-10-12 | Paola, KS | United States | 383226N | 0945513W | K8 |
| Accident | CEN23LA027 | 2022-11-05 | Aguilar, CO | United States | 037246N | 1043920W | Na |
| Accident | CEN23LA040 | 2022-11-20 | Cypress, TX | United States | 295824N | 0954114W | Na |
| Accident | CEN23LA052 | 2022-11-22 | Denton, TX | United States | 292548N | 1005924W | Na |
| Accident | CEN23LA067 | 2022-12-21 | Auburn Hills, MI | United States | NaN | NaN | Na |

In [128]:

```
df = new_df[new_df['Purpose.of.flight']=='Personal']
df
```

Out[128]:

| | Event.Id | Investigation.Type | Accident.Number | Event.Date | Location | Country | Latitude |
|---|---|---|---|---|---|---|---|
| **64136** | 20080401X00406 | Accident | CHI08CA090 | 2008-03-01 | Apple River, IL | United States | 042303N |
| **64239** | 20080425X00551 | Accident | CHI08CA095 | 2008-03-25 | North Canton, OH | United States | 405535N |
| **64418** | 20080513X00657 | Accident | CHI08CA116 | 2008-04-27 | Howell, MI | United States | 423817N |
| **64422** | 20080604X00785 | Accident | CHI08CA117 | 2008-04-27 | Madison, IN | United States | 394532N |
| **64990** | 20080822X01288 | Accident | CHI08CA206 | 2008-07-23 | Lowell, MI | United States | 425713N |
| **...** | ... | ... | ... | ... | ... | ... | ... |
| **88658** | 20221013106117 | Accident | CEN23LA012 | 2022-10-12 | Paola, KS | United States | 383226N |
| **88758** | 20221107106255 | Accident | CEN23LA027 | 2022-11-05 | Aguilar, CO | United States | 037246N |
| **88803** | 20221121106332 | Accident | CEN23LA040 | 2022-11-20 | Cypress, TX | United States | 295824N |
| **88816** | 20221128106371 | Accident | CEN23LA052 | 2022-11-22 | Denton, TX | United States | 292548N |
| **88881** | 20221221106483 | Accident | CEN23LA067 | 2022-12-21 | Auburn Hills, MI | United States | NaN |

214 rows × 31 columns

In [136]:

```
print(data.columns)
```

```
Index(['Event.Id', 'Investigation.Type', 'Accident.Number', 'Event.Date',
       'Location', 'Country', 'Latitude', 'Longitude', 'Airport.Code',
       'Airport.Name', 'Injury.Severity', 'Aircraft.damage',
       'Aircraft.Category', 'Registration.Number', 'Make', 'Model',
       'Amateur.Built', 'Number.of.Engines', 'Engine.Type', 'FAR.Descript
ion',
       'Schedule', 'Purpose.of.flight', 'Air.carrier', 'Total.Fatal.Injur
ies',
       'Total.Serious.Injuries', 'Total.Minor.Injuries', 'Total.Uninjure
d',
       'Weather.Condition', 'Broad.phase.of.flight', 'Report.Status',
       'Publication.Date'],
      dtype='object')
```

```
In [138]: data['Make'].unique()
```

```
Out[138]: array(['Stinson', 'Piper', 'Cessna', ..., 'JAMES R DERNOVSEK',
                 'ORLICAN S R O', 'ROYSE RALPH L'], dtype=object)
```

```
In [140]: data['Make'].value_counts(ascending = False)
```

```
Out[140]: Cessna                22227
          Piper                 12029
          CESSNA                 4922
          Beech                  4330
          PIPER                  2841
                                ...  
          Intl Ultralight           1
          Morris George             1
          AAA AIRCRAFT LLC          1
          SUD AVIATION              1
          KLEMP GREGORY J           1
          Name: Make, Length: 8237, dtype: int64
```

```
In [141]: data['Model'].value_counts(ascending = False)
```

```
Out[141]: 152               2367
          172               1756
          172N              1164
          PA-28-140          932
          150                829
                            ... 
          GLOBE GC-1A          1
          Avid Aerobat         1
          ASW-22               1
          CL-600-2D15          1
          PEGAZAIR 100         1
          Name: Model, Length: 12318, dtype: int64
```

```
In [142]: data['Injury.Severity'].value_counts(ascending = False)
```

```
Out[142]: Non-Fatal     67357
          Fatal(1)       6167
          Fatal          5262
          Fatal(2)       3711
          Incident       2219
                        ...  
          Fatal(72)         1
          Fatal(49)         1
          Fatal(114)        1
          Fatal(228)        1
          Fatal(47)         1
          Name: Injury.Severity, Length: 109, dtype: int64
```

In [143]:
```
data['Aircraft.damage'].value_counts(ascending = False)
```

Out[143]:
```
Substantial    64148
Destroyed      18623
Minor           2805
Unknown          119
Name: Aircraft.damage, dtype: int64
```

In [144]:
```
data['Total.Fatal.Injuries'].value_counts(ascending = False)
```

Out[144]:
```
0.0      59675
1.0       8883
2.0       5173
3.0       1589
4.0       1103
         ...  
31.0         1
169.0        1
150.0        1
117.0        1
156.0        1
Name: Total.Fatal.Injuries, Length: 125, dtype: int64
```

In [146]:
```
data['Purpose.of.flight'].value_counts(ascending = False)
```

Out[146]:
```
Personal                     49448
Instructional                10601
Unknown                       6802
Aerial Application            4712
Business                      4018
Positioning                   1646
Other Work Use                1264
Ferry                          812
Aerial Observation             794
Public Aircraft                720
Executive/corporate            553
Flight Test                    405
Skydiving                      182
External Load                  123
Public Aircraft - Federal      105
Banner Tow                     101
Air Race show                   99
Public Aircraft - Local         74
Public Aircraft - State         64
Air Race/show                   59
Glider Tow                      53
Firefighting                    40
Air Drop                        11
ASHO                             6
PUBS                             4
PUBL                             1
Name: Purpose.of.flight, dtype: int64
```

In [147]:

```
data['Weather.Condition'].value_counts(ascending = False)
```

Out[147]:

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

In [148]:

```
data['Broad.phase.of.flight'].value_counts(ascending = False)
```

Out[148]:

```
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 [156]:

```
data['Schedule'].dropna()
```

Out[156]:

```
5        SCHD
22       NSCH
33       NSCH
39       SCHD
43       NSCH
         ...
88866    NSCH
88867    NSCH
88876    SCHD
88879    SCHD
88881    NSCH
Name: Schedule, Length: 12582, dtype: object
```

In [157]:

```
data['Air.carrier'].dropna()
```

Out[157]:

```
5                              Air Canada
22       Rocky Mountain Helicopters, In
33                       Lang Air Service
39                        Empire Airlines
43          Joel Frederick's Monarch Air
                            ...
88877                    GERBER RICHARD E
88879               HAWAIIAN AIRLINES INC
88880                Chandler Air Service
88881                               Pilot
88887                  MC CESSNA 210N LLC
Name: Air.carrier, Length: 16648, dtype: object
```

In [158]:
```
data['Air.carrier'].value_counts()
```

Out[158]:
```
Pilot                                                  258
American Airlines                                       90
United Airlines                                         89
Delta Air Lines                                         53
SOUTHWEST AIRLINES CO                                   42
                                                      ...
Chemtec Aviation LLC                                     1
High Adventur Air Charter, Guides & Outfitters Inc       1
FAA FLYING CLUB INC                                      1
KEITER KENNETH A                                         1
ADAPTIVE AERO INC                                        1
Name: Air.carrier, Length: 13590, dtype: int64
```

In [184]:
```
subset = data[['Airport.Code', 'Airport.Name', 'Air.carrier']].dropna()
subset
```

Out[184]:

| | Airport.Code | Airport.Name | Air.carrier |
|---|---|---|---|
| **43** | Q35 | SPRINGERVILLE | Joel Frederick's Monarch Air |
| **44** | BET | BETHEL | Executive Charter Service |
| **79** | ORD | CHICAGO O'HARE INTER'L | Trans World Airlines |
| **80** | CKB | BENEDUM | Aeromech Incorporated |
| **93** | JFK | JOHN F. KENNEDY INT'L | Pocono Airlines |
| **...** | ... | ... | ... |
| **88839** | KVNC | Venice Municipal Airport | ST PETE AVIATION SERVICES LLC |
| **88845** | CRZ | Corning Municipal Airport | PORTER STEVEN B |
| **88865** | DKX | KNOXVILLE DOWNTOWN ISLAND | Knoxville Flight Training Academy |
| **88873** | SIG | FERNANDO LUIS RIBAS DOMINICCI | SKY WEST AVIATION INC TRUSTEE |
| **88877** | BKV | BROOKSVILLE-TAMPA BAY RGNL | GERBER RICHARD E |

10137 rows × 3 columns

In [185]:

```
data.iloc[15]
```

Out[185]:

```
Event.Id                             20020917X02117
Investigation.Type                         Accident
Accident.Number                          FTW82FPG08
Event.Date                               1982-01-02
Location                            LITTLE ROCK, AR
Country                               United States
Latitude                                        NaN
Longitude                                       NaN
Airport.Code                                    NaN
Airport.Name                                    NaN
Injury.Severity                            Fatal(2)
Aircraft.damage                           Destroyed
Aircraft.Category                          Airplane
Registration.Number                          N9779L
Make                                          Beech
Model                                            19
Amateur.Built                                    No
Number.of.Engines                                 1
Engine.Type                           Reciprocating
FAR.Description           Part 91: General Aviation
Schedule                                        NaN
Purpose.of.flight                          Personal
Air.carrier                                     NaN
Total.Fatal.Injuries                              2
Total.Serious.Injuries                            0
Total.Minor.Injuries                              0
Total.Uninjured                                   0
Weather.Condition                               IMC
Broad.phase.of.flight                        Cruise
Report.Status                        Probable Cause
Publication.Date                         02-01-1983
Name: 15, dtype: object
```

In [213]:
```
data.iloc[114]
```

Out[213]:
```
Event.Id                             20020917X02476
Investigation.Type                         Accident
Accident.Number                          MKC82FCQ08
Event.Date                               1982-01-19
Location                             CEDAR VALE, KS
Country                               United States
Latitude                                        NaN
Longitude                                       NaN
Airport.Code                                   KS30
Airport.Name                            MILLS RANCH
Injury.Severity                            Fatal(1)
Aircraft.damage                           Destroyed
Aircraft.Category                          Airplane
Registration.Number                          N7761M
Make                                         Mooney
Model                                          M20F
Amateur.Built                                    No
Number.of.Engines                                 1
Engine.Type                           Reciprocating
FAR.Description           Part 91: General Aviation
Schedule                                        NaN
Purpose.of.flight                          Business
Air.carrier                                     NaN
Total.Fatal.Injuries                              1
Total.Serious.Injuries                            0
Total.Minor.Injuries                              0
Total.Uninjured                                   0
Weather.Condition                               VMC
Broad.phase.of.flight                       Takeoff
Report.Status                        Probable Cause
Publication.Date                         19-01-1983
Name: 114, dtype: object
```

```
Keep:

Qualitative:
Investigation.Type
Location
Country
Aircraft.damage
Make
Model
Amateur.Built
Engine.Type
Purpose.of.flight
Weather.Condition
Broad.phase.of.flight

Quantiative:
Number.of.Engines
Total.Fatal.Injuries
Total.Serious.Injuries
Total.Minor.Injuries
Total.Uninjured
Injury.Severity
Event.Date
```

```
Drop:
Event.Id
Accident.Number
Latitude
Longitude
Airport.Code
Airport.Name
Registration.Number
FAR.Description
Schedule
Air.carrier
Publication.Date
```

In [250]:
```
data['Publication.Date'].value_counts()
```

Out[250]:
```
25-09-2020    17019
26-09-2020     1769
03-11-2020     1155
31-03-1993      452
25-11-2003      396
              ...  
10-06-1997        1
02-06-2021        1
04-12-2009        1
06-06-2019        1
24-02-2009        1
Name: Publication.Date, Length: 2924, dtype: int64
```

In [251]:
```
data['Publication.Date'].isna().value_counts()
```

Out[251]:
```
False    75118
True     13771
Name: Publication.Date, dtype: int64
```

In [253]:

```
columns_to_drop = ['Event.Id',
'Accident.Number',
'Latitude',
'Longitude',
'Airport.Code',
'Airport.Name',
'Registration.Number',
'FAR.Description',
'Schedule',
'Air.carrier',
'Publication.Date']

final_data_set = data.drop(columns = columns_to_drop)

print(final_data_set)
```

```
      Investigation.Type  Event.Date         Location        Country  \
0               Accident  1948-10-24  MOOSE CREEK, ID  United States   
1               Accident  1962-07-19   BRIDGEPORT, CA  United States   
2               Accident  1974-08-30    Saltville, VA  United States   
3               Accident  1977-06-19       EUREKA, CA  United States   
4               Accident  1979-08-02       Canton, OH  United States   
...                  ...         ...              ...            ...   
88884           Accident  2022-12-26    Annapolis, MD  United States   
88885           Accident  2022-12-26      Hampton, NH  United States   
88886           Accident  2022-12-26       Payson, AZ  United States   
88887           Accident  2022-12-26       Morgan, UT  United States   
88888           Accident  2022-12-29       Athens, GA  United States   

      Injury.Severity Aircraft.damage Aircraft.Category  \
0            Fatal(2)       Destroyed               NaN   
1            Fatal(4)       Destroyed               NaN   
2            Fatal(3)       Destroyed               NaN   
3            Fatal(2)       Destroyed               NaN   
4            Fatal(1)       Destroyed               NaN   
...               ...             ...               ...   
88884           Minor             NaN               NaN   
88885             NaN             NaN               NaN   
88886       Non-Fatal     Substantial          Airplane   
88887             NaN             NaN               NaN   
88888           Minor             NaN               NaN   

                             Make      Model Amateur.Built  Number.of.Eng
ines  \
0                         Stinson      108-3            No              
1.0   
1                           Piper   PA24-180            No              
1.0   
2                          Cessna       172M            No              
1.0   
3                        Rockwell        112            No              
1.0   
4                          Cessna        501            No              
NaN   
...                           ...        ...           ...              
...   
88884                       PIPER  PA-28-151            No              
NaN   
88885                    BELLANCA       7ECA            No              
NaN   
88886  AMERICAN CHAMPION AIRCRAFT      8GCBC            No              
1.0   
88887                      CESSNA       210N            No              
NaN   
88888                       PIPER  PA-24-260            No              
NaN   

         Engine.Type Purpose.of.flight  Total.Fatal.Injuries  \
0      Reciprocating          Personal                   2.0   
1      Reciprocating          Personal                   4.0   
2      Reciprocating          Personal                   3.0   
3      Reciprocating          Personal                   2.0   
4                NaN          Personal                   1.0   
...              ...               ...                   ...   
```

```
88884            NaN           Personal                   0.0
88885            NaN                NaN                   0.0
88886            NaN           Personal                   0.0
88887            NaN           Personal                   0.0
88888            NaN           Personal                   0.0

       Total.Serious.Injuries  Total.Minor.Injuries  Total.Uninjured  \
0                         0.0                   0.0              0.0   
1                         0.0                   0.0              0.0   
2                         NaN                   NaN              NaN   
3                         0.0                   0.0              0.0   
4                         2.0                   NaN              0.0   
...                       ...                   ...              ...   
88884                     1.0                   0.0              0.0   
88885                     0.0                   0.0              0.0   
88886                     0.0                   0.0              1.0   
88887                     0.0                   0.0              0.0   
88888                     1.0                   0.0              1.0   

      Weather.Condition Broad.phase.of.flight   Report.Status
0                   UNK                Cruise  Probable Cause
1                   UNK               Unknown  Probable Cause
2                   IMC                Cruise  Probable Cause
3                   IMC                Cruise  Probable Cause
4                   VMC              Approach  Probable Cause
...                 ...                   ...             ...
88884               NaN                   NaN             NaN
88885               NaN                   NaN             NaN
88886               VMC                   NaN             NaN
88887               NaN                   NaN             NaN
88888               NaN                   NaN             NaN

[88889 rows x 20 columns]
```

```
In [254]: final_data_set.to_csv('final_dataset.csv')
```

In [1]:
```
import csv
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from datetime import datetime
```

In [3]:
```
data = pd.read_csv('final_dataset_1.csv')
```

```
/opt/anaconda3/envs/learn-env/lib/python3.8/site-packages/IPython/cor
e/interactiveshell.py:3145: DtypeWarning: Columns (19) have mixed typ
es.Specify dtype option on import or set low_memory=False.
  has_raised = await self.run_ast_nodes(code_ast.body, cell_name,
```

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

Out[4]:

| Make | Model | ... | Number.of.Engines | Engine.Type | Purpose.of.flight | Total.Fatal.Injuries | Total.S |
|---|---|---|---|---|---|---|---|
| Stinson | 108-3 | ... | 1.0 | Reciprocating | Personal | 2.0 | |
| Piper | PA24-180 | ... | 1.0 | Reciprocating | Personal | 4.0 | |
| Cessna | 172M | ... | 1.0 | Reciprocating | Personal | 3.0 | |
| Rockwell | 112 | ... | 1.0 | Reciprocating | Personal | 2.0 | |
| Cessna | 501 | ... | NaN | NaN | Personal | 1.0 | |

In [11]:
```
final_data_set = data.drop(columns = 'Unnamed: 0')
final_data_set
```

Out[11]:

| | Investigation.Type | Event.Date | Location | Country | Injury.Severity | Aircraft.damage | A |
|---|---|---|---|---|---|---|---|
| **0** | Accident | 1948-10-24 | MOOSE CREEK, ID | United States | Fatal(2) | Destroyed | |
| **1** | Accident | 1962-07-19 | BRIDGEPORT, CA | United States | Fatal(4) | Destroyed | |
| **2** | Accident | 1974-08-30 | Saltville, VA | United States | Fatal(3) | Destroyed | |
| **3** | Accident | 1977-06-19 | EUREKA, CA | United States | Fatal(2) | Destroyed | |
| **4** | Accident | 1979-08-02 | Canton, OH | United States | Fatal(1) | Destroyed | |
| **...** | ... | ... | ... | ... | ... | ... | |
| **88884** | Accident | 2022-12-26 | Annapolis, MD | United States | Minor | NaN | |
| **88885** | Accident | 2022-12-26 | Hampton, NH | United States | NaN | NaN | |
| **88886** | Accident | 2022-12-26 | Payson, AZ | United States | Non-Fatal | Substantial | |
| **88887** | Accident | 2022-12-26 | Morgan, UT | United States | NaN | NaN | |
| **88888** | Accident | 2022-12-29 | Athens, GA | United States | Minor | NaN | |

88889 rows × 20 columns

```
Filtered the data frame to only incldue quantitative values
(Injury.Severtiy is incldued here because of the numeric values
associated with the Fatal catagory)
```

In [6]:
```
columns = ['Number.of.Engines', 'Total.Fatal.Injuries', 'Total.Serious.
quant_columns = data[columns]
quant_columns[89:100]
```

Out[6]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Uninjur |
|---|---|---|---|---|---|
| **89** | 2.0 | NaN | NaN | NaN | 2 |
| **90** | 1.0 | 0.0 | 0.0 | 0.0 | 2 |
| **91** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |
| **92** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |
| **93** | 2.0 | 0.0 | 1.0 | 2.0 | 12 |
| **94** | 2.0 | 0.0 | 0.0 | 0.0 | 2 |
| **95** | 1.0 | 0.0 | 0.0 | 0.0 | 4 |
| **96** | 1.0 | 1.0 | 2.0 | 0.0 | 0 |
| **97** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |
| **98** | 1.0 | 1.0 | 0.0 | 0.0 | 0 |
| **99** | 1.0 | 1.0 | 0.0 | 0.0 | 0 |

In [7]:
```
quant_columns['Number.of.Engines'].unique()
```

Out[7]:
```
array([ 1., nan,  2.,  0.,  3.,  4.,  8.,  6.])
```

```
In [25]: quant_columns[['Number.of.Engines', 'Total.Fatal.Injuries']]
```

Out[25]:

| | Number.of.Engines | Total.Fatal.Injuries |
|---|---|---|
| **0** | 1.0 | 2.0 |
| **1** | 1.0 | 4.0 |
| **2** | 1.0 | 3.0 |
| **3** | 1.0 | 2.0 |
| **4** | NaN | 1.0 |
| **...** | ... | ... |
| **88884** | NaN | 0.0 |
| **88885** | NaN | 0.0 |
| **88886** | 1.0 | 0.0 |
| **88887** | NaN | 0.0 |
| **88888** | NaN | 0.0 |

88889 rows × 2 columns

```
In [41]: quant_columns['Number.of.Engines'].value_counts()
```

```
Out[41]: 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
```

Randomly assing 1 and 2s weighed based off data

```
In [113]: quant_columns['Number.of.Engines'].isna().value_counts()
```

```
Out[113]: False    82805
          True      6084
          Name: Number.of.Engines, dtype: int64
```

```
In [116]: quant_columns['Total.Fatal.Injuries'].value_counts()
```

```
Out[116]: 0.0      59675
          1.0       8883
          2.0       5173
          3.0       1589
          4.0       1103
                   ...
          31.0         1
          169.0        1
          150.0        1
          117.0        1
          156.0        1
          Name: Total.Fatal.Injuries, Length: 125, dtype: int64
```

```
In [45]: quant_columns['Total.Serious.Injuries'].unique()
```

```
Out[45]: array([  0.,  nan,   2.,   1.,   6.,   4.,   5.,  10.,   3.,   8.,
         9.,
                 7.,  15.,  17.,  28.,  26.,  47.,  14.,  81.,  13., 106.,   6
         0.,
                16.,  21.,  50.,  44.,  18.,  12.,  45.,  39.,  43.,  11.,   2
         5.,
                59.,  23.,  55.,  63.,  88.,  41.,  34.,  53.,  33.,  67.,   3
         5.,
                20., 137.,  19.,  27., 125., 161.,  22.])
```

```
In [61]: quant_columns['Total.Uninjured'].unique()
```

```
Out[61]: array([  0.,  nan,  44.,   2.,   1.,   3.,   6.,   4., 149.,  12., 18
         2.,
                154.,   5.,  10.,   7., 119.,  36.,  51.,  16.,  83.,   9.,   6
         8.,
                 30.,  20.,  18.,   8., 108.,  11., 152.,  21.,  48.,  56., 11
         3.,
                129., 109.,  29.,  13.,  84.,  74., 142., 102., 393., 128., 11
         2.,
                 17.,  65.,  67., 136.,  23., 116.,  22.,  57.,  58.,  73., 20
         3.,
                 31., 201., 412., 159.,  39., 186., 588.,  82.,  95., 146., 19
         0.,
                245., 172.,  52.,  25.,  59., 131., 151., 180., 150.,  86.,   1
         9.,
                133., 240.,  15., 145., 125., 440.,  77., 122., 205., 289., 11
         0.,
                 79.,  66.,  87.,  78.,  49., 104., 250.,  33., 138., 100.,   5
         3.,
                158., 127., 160., 260.,  47.,  38., 165., 495.,  81.,  41.,   1
         4.,
                 72.,  98., 263., 188., 239.,  27., 105., 111., 212., 157.,   4
         6.,
                121.,  75.,  71.,  45.,  91.,  99.,  85.,  96.,  50.,  93., 27
         6.,
                365., 371., 200., 103., 189.,  37., 107.,  61.,  26., 271., 13
         0.,
                 89., 439., 132., 219.,  43., 238., 195., 118., 175.,  32., 50
         7.,
                421.,  90., 225., 269., 169., 236., 224., 134., 106., 331., 14
         0.,
                 94., 192., 161., 270.,  69., 436., 213., 233., 115.,  42., 16
         7.,
                137., 114., 148., 222.,  92., 375.,  76., 171., 173., 246., 23
         4.,
                123., 220., 202., 408., 279., 363., 135., 528., 334., 178., 14
         7.,
                126.,  62.,  70.,  97., 228., 226.,  64., 290., 206., 297., 34
         9.,
                208., 144.,  54.,  24., 258., 304., 274., 286.,  55., 199., 22
         1.,
                 80., 272., 211., 262., 441., 194., 309., 185., 261., 241., 38
         3.,
                177., 259., 244., 254., 156.,  40.,  34., 247., 176.,  63.,   2
         8.,
                218., 282., 320., 204., 124., 215., 298., 120., 280., 179., 31
         5.,
                461., 153.,  60., 308.,  88., 361., 277., 191., 235., 187., 10
         1.,
                162.,  35., 197., 193., 164., 370., 387., 163., 139., 267., 35
         7.,
                339., 288., 231., 300., 255., 306., 443., 385., 248., 459., 14
         1.,
                414., 229., 166., 209., 184., 168., 170., 198., 299., 573., 22
         3.,
                265., 322., 196., 117., 253., 399., 360., 252., 217., 155., 18
         3.,
                227., 249., 329., 340., 699., 325., 287., 143., 243., 230., 38
         6.,
                181., 257., 283., 404., 319., 450., 356., 216., 174., 558., 21
         4.,
                448., 324., 338., 273., 232., 401., 312., 368., 501., 237., 30
         7.,
                296., 291., 403., 314., 285., 311., 293., 352., 332., 384., 27
         5.,
                210., 268., 326., 454., 278., 576., 380., 394., 362., 397., 35
         9.,
                264., 333., 367., 302., 348., 351., 358., 295., 321., 521., 30
         1.,
                294., 378., 207., 406., 251., 455.])
```

In [54]:
```
of.Engines', 'Total.Fatal.Injuries', 'Total.Serious.Injuries', 'Total.M
```

In [57]:
```
cleaned_nans[21:41]
```

Out[57]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Uninjur |
|---|---|---|---|---|---|
| **21** | 1.0 | 0.0 | 0.0 | 0.0 | 2 |
| **22** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |
| **23** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |
| **24** | 1.0 | 2.0 | 1.0 | 0.0 | 0 |
| **25** | 2.0 | 8.0 | 0.0 | 0.0 | 0 |
| **26** | 2.0 | 1.0 | 0.0 | 0.0 | 0 |
| **27** | 1.0 | 0.0 | 0.0 | 0.0 | 2 |
| **28** | 1.0 | 0.0 | 0.0 | 1.0 | 1 |
| **29** | 1.0 | 0.0 | 0.0 | 2.0 | 0 |
| **30** | 1.0 | 0.0 | 0.0 | 2.0 | 0 |
| **31** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |
| **32** | 1.0 | 4.0 | 0.0 | 0.0 | 0 |
| **33** | 2.0 | 2.0 | 0.0 | 0.0 | 0 |
| **34** | 2.0 | 2.0 | 2.0 | 0.0 | 0 |
| **35** | 1.0 | 0.0 | 0.0 | 2.0 | 0 |
| **36** | 1.0 | 1.0 | 0.0 | 1.0 | 0 |
| **37** | 1.0 | 3.0 | 0.0 | 0.0 | 0 |
| **38** | 1.0 | 1.0 | 0.0 | 0.0 | 0 |
| **39** | 2.0 | 2.0 | 0.0 | 0.0 | 0 |
| **40** | 1.0 | 0.0 | 0.0 | 0.0 | 1 |

In [64]:
```
cleaned_quant = cleaned_nans[(((cleaned_nans['Number.of.Engines']) < 7
                              (cleaned_nans['Total.Fatal.Injuries']) > 0
                              (cleaned_nans['Total.Serious.Injuries']) >
                              (cleaned_nans['Total.Minor.Injuries']) > 0
                              (cleaned_nans['Total.Uninjured']) > 0 or (
```

```
  File "<ipython-input-64-88d25a9d1392>", line 5
    (cleaned_nans['Total.Uninjured']) > 0 or (cleaned_nans['Total.Uni
njured'] != 'Nan' or 'nan' )]

^
SyntaxError: closing parenthesis ']' does not match opening parenthes
is '(' on line 4
```

In [59]:
```
print(quant_columns.dtypes)
```

```
Number.of.Engines         float64
Total.Fatal.Injuries      float64
Total.Serious.Injuries    float64
Total.Minor.Injuries      float64
Total.Uninjured           float64
Injury.Severity            object
Event.Date                 object
dtype: object
```

In [78]:

```
test_set = quant_columns[quant_columns['Injury.Severity'].isna()]
test_set
```

Out[78]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **63918** | NaN | 0.0 | 0.0 | 0.0 | |
| **63962** | NaN | 0.0 | 0.0 | 0.0 | |
| **63987** | NaN | 0.0 | 0.0 | 0.0 | |
| **64026** | 2.0 | 0.0 | 0.0 | 0.0 | |
| **64128** | NaN | 0.0 | 0.0 | 0.0 | |
| **...** | ... | ... | ... | ... | |
| **88863** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88874** | NaN | 0.0 | 0.0 | 0.0 | |
| **88879** | NaN | 0.0 | 0.0 | 0.0 | |
| **88885** | NaN | 0.0 | 0.0 | 0.0 | |
| **88887** | NaN | 0.0 | 0.0 | 0.0 | |

1000 rows × 7 columns

In [79]:
```
IS_updated_set = quant_columns.dropna(subset=['Injury.Severity'])
IS_updated_set
```

Out[79]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **1** | 1.0 | 4.0 | 0.0 | 0.0 | |
| **2** | 1.0 | 3.0 | NaN | NaN | |
| **3** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **4** | NaN | 1.0 | 2.0 | NaN | |
| **...** | ... | ... | ... | ... | |
| **88882** | NaN | 0.0 | 1.0 | 0.0 | |
| **88883** | NaN | 1.0 | 0.0 | 0.0 | |
| **88884** | NaN | 0.0 | 1.0 | 0.0 | |
| **88886** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88888** | NaN | 0.0 | 1.0 | 0.0 | |

87889 rows × 7 columns

In [68]:
```
test_set['Number.of.Engines'].value_counts()
```

Out[68]:
```
2.0    213
1.0     46
4.0     19
3.0      1
Name: Number.of.Engines, dtype: int64
```

In [76]:
```
quant_columns['Number.of.Engines'].value_counts()
```

Out[76]:
```
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 [70]: len(test_set['Number.of.Engines'])

Out[70]: 1000
```

```
In [71]: test_set['Total.Fatal.Injuries'].value_counts()

Out[71]: 0.0    1000
         Name: Total.Fatal.Injuries, dtype: int64
```

In [74]:
```
quant_columns['Total.Serious.Injuries'].value_counts()
```

Out[74]:
```
0.0      63289
1.0       9125
2.0       2815
3.0        629
4.0        258
5.0         78
6.0         41
7.0         27
9.0         16
8.0         13
10.0        13
13.0         9
11.0         6
26.0         5
14.0         5
12.0         5
25.0         3
20.0         3
28.0         3
17.0         2
50.0         2
59.0         2
21.0         2
47.0         2
55.0         1
88.0         1
41.0         1
67.0         1
33.0         1
18.0         1
161.0        1
81.0         1
39.0         1
137.0        1
27.0         1
15.0         1
45.0         1
125.0        1
23.0         1
44.0         1
106.0        1
22.0         1
34.0         1
16.0         1
35.0         1
53.0         1
43.0         1
63.0         1
19.0         1
60.0         1
Name: Total.Serious.Injuries, dtype: int64
```

In [75]:
```
quant_columns
```

Out[75]:
```
       Number.of.Engines  Total.Fatal.Injuries  Total.Serious.Injuries  Total.Minor.Injuries  Total.Unir
     0               1.0                   2.0                     0.0                   0.0
     1               1.0                   4.0                     0.0                   0.0
     2               1.0                   3.0                     NaN                   NaN
     3               1.0                   2.0                     0.0                   0.0
     4               NaN                   1.0                     2.0                   NaN
   ...               ...                   ...                     ...                   ...
 88884               NaN                   0.0                     1.0                   0.0
 88885               NaN                   0.0                     0.0                   0.0
 88886               1.0                   0.0                     0.0                   0.0
 88887               NaN                   0.0                     0.0                   0.0
 88888               NaN                   0.0                     1.0                   0.0

88889 rows × 7 columns
```

In [80]:
```
IS_updated_set
```

Out[80]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **1** | 1.0 | 4.0 | 0.0 | 0.0 | |
| **2** | 1.0 | 3.0 | NaN | NaN | |
| **3** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **4** | NaN | 1.0 | 2.0 | NaN | |
| **...** | ... | ... | ... | ... | |
| **88882** | NaN | 0.0 | 1.0 | 0.0 | |
| **88883** | NaN | 1.0 | 0.0 | 0.0 | |
| **88884** | NaN | 0.0 | 1.0 | 0.0 | |
| **88886** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88888** | NaN | 0.0 | 1.0 | 0.0 | |

87889 rows × 7 columns

In [84]:
```
IS_updated_set['Injury.Severity'].unique()
```

Out[84]:
```
array(['Fatal(2)', 'Fatal(4)', 'Fatal(3)', 'Fatal(1)', 'Non-Fatal',
       'Incident', 'Fatal(8)', 'Fatal(78)', 'Fatal(7)', 'Fatal(6)',
       'Fatal(5)', 'Fatal(153)', 'Fatal(12)', 'Fatal(14)', 'Fatal(2
3)',
       'Fatal(10)', 'Fatal(11)', 'Fatal(9)', 'Fatal(17)', 'Fatal(1
3)',
       'Fatal(29)', 'Fatal(70)', 'Unavailable', 'Fatal(135)', 'Fatal
(31)',
       'Fatal(256)', 'Fatal(25)', 'Fatal(82)', 'Fatal(156)', 'Fatal(2
8)',
       'Fatal(18)', 'Fatal(43)', 'Fatal(15)', 'Fatal(270)', 'Fatal(14
4)',
       'Fatal(174)', 'Fatal(111)', 'Fatal(131)', 'Fatal(20)', 'Fatal
(73)',
       'Fatal(27)', 'Fatal(34)', 'Fatal(87)', 'Fatal(30)', 'Fatal(1
6)',
       'Fatal(47)', 'Fatal(56)', 'Fatal(37)', 'Fatal(132)', 'Fatal(6
8)',
       'Fatal(54)', 'Fatal(52)', 'Fatal(65)', 'Fatal(72)', 'Fatal(16
0)',
       'Fatal(189)', 'Fatal(123)', 'Fatal(33)', 'Fatal(110)',
       'Fatal(230)', 'Fatal(97)', 'Fatal(349)', 'Fatal(125)', 'Fatal
(35)',
       'Fatal(228)', 'Fatal(75)', 'Fatal(104)', 'Fatal(229)', 'Fatal
(80)',
       'Fatal(217)', 'Fatal(169)', 'Fatal(88)', 'Fatal(19)', 'Fatal(6
0)',
       'Fatal(113)', 'Fatal(143)', 'Fatal(83)', 'Fatal(24)', 'Fatal(4
4)',
       'Fatal(64)', 'Fatal(92)', 'Fatal(118)', 'Fatal(265)', 'Fatal(2
6)',
       'Fatal(138)', 'Fatal(206)', 'Fatal(71)', 'Fatal(21)', 'Fatal(4
6)',
       'Fatal(102)', 'Fatal(115)', 'Fatal(141)', 'Fatal(55)',
       'Fatal(121)', 'Fatal(45)', 'Fatal(145)', 'Fatal(117)',
       'Fatal(107)', 'Fatal(124)', 'Fatal(49)', 'Fatal(154)', 'Fatal
(96)',
       'Fatal(114)', 'Fatal(199)', 'Fatal(89)', 'Fatal(57)', 'Fatal',
       'Minor', 'Serious'], dtype=object)
```

Dropping the numbers connected to the Fatal values because they are already represented as the number of fatal injuries

In [117]:
```
IS_updated_set = IS_updated_set.copy()
for index, word in IS_updated_set['Injury.Severity'].items():
    if word not in ['Incident', 'Unavailable', 'Non-Fatal', 'Fatal']:
        holder = word.split("(", 1)
        IS_updated_set.loc[index, 'Injury.Severity'] = holder[0]
IS_updated_set
```

Out[117]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **1** | 1.0 | 4.0 | 0.0 | 0.0 | |
| **2** | 1.0 | 3.0 | NaN | NaN | |
| **3** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **4** | NaN | 1.0 | 2.0 | NaN | |
| **...** | ... | ... | ... | ... | |
| **88882** | NaN | 0.0 | 1.0 | 0.0 | |
| **88883** | NaN | 1.0 | 0.0 | 0.0 | |
| **88884** | NaN | 0.0 | 1.0 | 0.0 | |
| **88886** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88888** | NaN | 0.0 | 1.0 | 0.0 | |

87889 rows × 7 columns

Filling the NaN values with the assumption that if one of the values in the row was reported, then the NaN most likely means they are zeros.

In [146]:
```
IS_updated_set[['Total.Fatal.Injuries', 'Total.Serious.Injuries', 'Tota
IS_updated_set
```

Out[146]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **1** | 1.0 | 4.0 | 0.0 | 0.0 | |
| **2** | 1.0 | 3.0 | 0.0 | 0.0 | |
| **3** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **4** | NaN | 1.0 | 2.0 | 0.0 | |
| **...** | ... | ... | ... | ... | |
| **88882** | NaN | 0.0 | 1.0 | 0.0 | |
| **88883** | NaN | 1.0 | 0.0 | 0.0 | |
| **88884** | NaN | 0.0 | 1.0 | 0.0 | |
| **88886** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88888** | NaN | 0.0 | 1.0 | 0.0 | |

87889 rows × 7 columns

In [128]:
```
IS_updated_set['Total.Serious.Injuries'].fillna(0)
```

Out[128]:
```
0        0.0
1        0.0
2        0.0
3        0.0
4        2.0
        ... 
88882    1.0
88883    0.0
88884    1.0
88886    0.0
88888    1.0
Name: Total.Serious.Injuries, Length: 87889, dtype: float64
```

```
In [129]: IS_updated_set['Total.Minor.Injuries'].fillna(0)
```

```
Out[129]: 0        0.0
          1        0.0
          2        0.0
          3        0.0
          4        0.0
                  ...
          88882    0.0
          88883    0.0
          88884    0.0
          88886    0.0
          88888    0.0
          Name: Total.Minor.Injuries, Length: 87889, dtype: float64
```

```
In [130]: IS_updated_set['Total.Uninjured'].fillna(0)
```

```
Out[130]: 0        0.0
          1        0.0
          2        0.0
          3        0.0
          4        0.0
                  ...
          88882    1.0
          88883    0.0
          88884    0.0
          88886    1.0
          88888    1.0
          Name: Total.Uninjured, Length: 87889, dtype: float64
```

In [131]:
```
IS_updated_set
```

Out[131]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **1** | 1.0 | 4.0 | 0.0 | 0.0 | |
| **2** | 1.0 | 3.0 | NaN | NaN | |
| **3** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **4** | NaN | 1.0 | 2.0 | NaN | |
| **...** | ... | ... | ... | ... | |
| **88882** | NaN | 0.0 | 1.0 | 0.0 | |
| **88883** | NaN | 1.0 | 0.0 | 0.0 | |
| **88884** | NaN | 0.0 | 1.0 | 0.0 | |
| **88886** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88888** | NaN | 0.0 | 1.0 | 0.0 | |

87889 rows × 7 columns

In [147]:
```
IS_updated_set['Number.of.Engines'].unique()
```

Out[147]:
```
array([ 1., nan,  2.,  0.,  3.,  4.,  8.,  6.])
```

In [171]:
```
IS_updated_set['Number.of.Engines'].value_counts()
```

Out[171]:
```
1.0    69536
2.0    10866
0.0     1226
3.0      482
4.0      412
8.0        3
6.0        1
Name: Number.of.Engines, dtype: int64
```

In [149]:
```
IS_updated_set['Number.of.Engines'].value_counts()[0]
```

Out[149]:
```
1226
```

In [150]:
```
IS_updated_set['Number.of.Engines'].value_counts()[1]
```
Out[150]:
```
69536
```

In [155]:
```
total = 69536+10866+1226+482+412+3+1
total
```
Out[155]:
```
82526
```

In [160]:
```
IS_updated_set['Number.of.Engines'].value_counts()[1]/ total
```
Out[160]:
```
0.8425950609504883
```

In [166]:
```
IS_updated_set['Number.of.Engines'].value_counts(1)[1]
```
Out[166]:
```
0.8425950609504883
```

In [169]:
```
sum(IS_updated_set['Number.of.Engines'].value_counts())
```
Out[169]:
```
82526
```

In [172]:
```
updated_total = sum(IS_updated_set['Number.of.Engines'].value_counts()
```

In [181]:
```
P_1 = 69536/updated_total
P_1
```
Out[181]:
```
0.8553013530135302
```

In [180]:
```
P_2 = 10866/updated_total
P_2
```
Out[180]:
```
0.13365313653136532
```

In [179]:
```
P_3 = 482/updated_total
P_3
```
Out[179]:
```
0.005928659286592866
```

In [178]:
```
P_4 = 412/updated_total
P_4
```
Out[178]:
```
0.005067650676506765
```

In [185]:

```
IS_updated_set['Number.of.Engines'] = IS_updated_set['Number.of.Engines
IS_updated_set
```

Out[185]:

```
       Number.of.Engines  Total.Fatal.Injuries  Total.Serious.Injuries  Total.Minor.Injuries  Total.Unin
    0                  1                   2.0                     0.0                   0.0
    1                  1                   4.0                     0.0                   0.0
    2                  1                   3.0                     0.0                   0.0
    3                  1                   2.0                     0.0                   0.0
    4                1.0                   1.0                     2.0                   0.0
  ...                ...                   ...                     ...                   ...
88882                1.0                   0.0                     1.0                   0.0
88883                1.0                   1.0                     0.0                   0.0
88884                1.0                   0.0                     1.0                   0.0
88886                  1                   0.0                     0.0                   0.0
88888                1.0                   0.0                     1.0                   0.0

87889 rows × 7 columns
```

In [191]:

```
IS_updated_set['Total.Uninjured'].unique()
```

Out[191]:

```
array([  0.,  44.,   2.,   1.,   3.,   6.,   4., 149.,  12., 182., 15
4.,
         5.,  10.,   7., 119.,  36.,  51.,  16.,  83.,   9.,  68.,  3
0.,
        20.,  18.,   8., 108.,  11., 152.,  21.,  48.,  56., 113., 12
9.,
       109.,  29.,  13.,  84.,  74., 142., 102., 393., 128., 112.,   1
7.,
        65.,  67., 136.,  23., 116.,  22.,  57.,  58.,  73., 203.,   3
1.,
       201., 412., 159.,  39., 186., 588.,  82.,  95., 146., 190., 24
5.,
       172.,  52.,  25.,  59., 131., 151., 180., 150.,  86.,  19., 13
3.,
       240.,  15., 145., 125., 440.,  77., 122., 205., 289., 110.,   7
9.,
        66.,  87.,  78.,  49., 104., 250.,  33., 138., 100.,  53., 15
8.,
       127., 160., 260.,  47.,  38., 165., 495.,  81.,  41.,  14.,   7
2.,
        98., 263., 188., 239.,  27., 105., 111., 212., 157.,  46., 12
1.,
        75.,  71.,  45.,  91.,  99.,  85.,  96.,  50.,  93., 276., 36
5.,
       371., 200., 103., 189.,  37., 107.,  61.,  26., 271., 130.,   8
9.,
       439., 132., 219.,  43., 238., 195., 118., 175.,  32., 507., 42
1.,
        90., 225., 269., 169., 236., 224., 134., 106., 331., 140.,   9
4.,
       192., 161., 270.,  69., 436., 213., 233., 115.,  42., 167., 13
7.,
       114., 148., 222.,  92., 375.,  76., 171., 173., 246., 234., 12
3.,
       220., 202., 408., 279., 363., 135., 528., 334., 178., 147., 12
6.,
        62.,  70.,  97., 228., 226.,  64., 290., 206., 297., 349., 20
8.,
       144.,  54.,  24., 258., 304., 274., 286.,  55., 199., 221.,   8
0.,
       272., 211., 262., 441., 194., 309., 185., 261., 241., 383., 17
7.,
       259., 244., 254., 156.,  40.,  34., 247., 176.,  63.,  28., 21
8.,
       282., 320., 204., 124., 215., 298., 120., 280., 179., 315., 46
1.,
       153.,  60., 308.,  88., 361., 277., 191., 235., 187., 101., 16
2.,
        35., 197., 193., 164., 370., 387., 163., 139., 267., 357., 33
9.,
       288., 231., 300., 255., 306., 443., 385., 248., 459., 141., 41
4.,
       229., 166., 209., 184., 168., 170., 198., 299., 573., 223., 26
5.,
       322., 196., 117., 253., 399., 360., 252., 217., 155., 183., 22
7.,
       249., 329., 340., 699., 325., 287., 143., 243., 230., 386., 18
1.,
       257., 283., 404., 319., 450., 356., 216., 174., 558., 214., 44
8.,
       324., 338., 273., 232., 401., 312., 368., 501., 237., 307., 29
6.,
       291., 403., 314., 285., 311., 293., 352., 332., 384., 275., 21
0.,
       268., 326., 454., 278., 576., 380., 394., 362., 397., 359., 26
4.,
       333., 367., 302., 348., 351., 358., 295., 321., 521., 301., 29
4.,
       378., 207., 406., 251., 455.])
```

In [193]:

```
Engines','Total.Fatal.Injuries', 'Total.Serious.Injuries', 'Total.Minor
```

Out[193]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **1** | 1.0 | 4.0 | 0.0 | 0.0 | |
| **2** | 1.0 | 3.0 | 0.0 | 0.0 | |
| **3** | 1.0 | 2.0 | 0.0 | 0.0 | |
| **4** | 1.0 | 1.0 | 2.0 | 0.0 | |
| **...** | ... | ... | ... | ... | |
| **88882** | 1.0 | 0.0 | 1.0 | 0.0 | |
| **88883** | 1.0 | 1.0 | 0.0 | 0.0 | |
| **88884** | 1.0 | 0.0 | 1.0 | 0.0 | |
| **88886** | 1.0 | 0.0 | 0.0 | 0.0 | |
| **88888** | 1.0 | 0.0 | 1.0 | 0.0 | |

87889 rows × 7 columns

In [194]:
```
f.Engines','Total.Fatal.Injuries', 'Total.Serious.Injuries', 'Total.Mir
```

Out[194]:

| | Number.of.Engines | Total.Fatal.Injuries | Total.Serious.Injuries | Total.Minor.Injuries | Total.Unir |
|---|---|---|---|---|---|
| **0** | 1 | 2 | 0 | 0 | |
| **1** | 1 | 4 | 0 | 0 | |
| **2** | 1 | 3 | 0 | 0 | |
| **3** | 1 | 2 | 0 | 0 | |
| **4** | 1 | 1 | 2 | 0 | |
| **...** | ... | ... | ... | ... | |
| **88882** | 1 | 0 | 1 | 0 | |
| **88883** | 1 | 1 | 0 | 0 | |
| **88884** | 1 | 0 | 1 | 0 | |
| **88886** | 1 | 0 | 0 | 0 | |
| **88888** | 1 | 0 | 1 | 0 | |

87889 rows × 7 columns

In [196]:
```
final_quant_columns = IS_updated_set
```

In [197]:
```
final_quant_columns
```

Out[197]:
```
       Number.of.Engines  Total.Fatal.Injuries  Total.Serious.Injuries  Total.Minor.Injuries  Total.Unir

     0                 1                     2                       0                     0
     1                 1                     4                       0                     0
     2                 1                     3                       0                     0
     3                 1                     2                       0                     0
     4                 1                     1                       2                     0
   ...               ...                   ...                     ...                   ...
 88882                 1                     0                       1                     0
 88883                 1                     1                       0                     0
 88884                 1                     0                       1                     0
 88886                 1                     0                       0                     0
 88888                 1                     0                       1                     0

87889 rows × 7 columns
```

In [ ]:
```
updated_df.to_csv("clean_data_final_1.22.csv", index=False)
```