In [1]:
```
import pandas as pd
import numpy as np
```

In [4]:
```
df= pd.read_excel('Bird_Strikes_data.xlsx')
df
```

Out[4]:

	Record ID	Aircraft: Type	Airport: Name	Altitude bin	Aircraft: Make/Model	Wildlife: Number struck	Wildlife: Number Struck Actual	Effect Impact to flight
0	202152	Airplane	LAGUARDIA NY	> 1000 ft	B-737-400	Over 100	859	Engine Shut Down
1	208159	Airplane	DALLAS/FORT WORTH INTL ARPT	< 1000 ft	MD-80	Over 100	424	None
2	207601	Airplane	LAKEFRONT AIRPORT	< 1000 ft	C-500	Over 100	261	None
3	215953	Airplane	SEATTLE-TACOMA INTL	< 1000 ft	B-737-400	Over 100	806	Precautionary Landing
4	219878	Airplane	NORFOLK INTL	< 1000 ft	CL-RJ100/200	Over 100	942	None
...	...	...	...	...	...	...	...	...
25553	321151	Airplane	REDDING MUNICIPAL	> 1000 ft	EMB-120	1	1	None
25554	319677	Airplane	ORLANDO INTL	< 1000 ft	A-321	1	1	None
25555	319680	NaN	NaN	NaN	EC-135	NaN	1	NaN
25556	319679	Airplane	DETROIT METRO WAYNE COUNTY ARPT	< 1000 ft	B-757-200	1	1	None
25557	319593	Airplane	ABRAHAM LINCOLN CAPITAL ARPT	< 1000 ft	B-737-400	1	1	None

In [5]:
```
df.shape
```

Out[5]:
```
(25558, 26)
```

In [6]:
```
df.ndim
```

Out[6]:
```
2
```

In [8]:
```
df.describe()
```

Out[8]:

	Record ID	Wildlife: Number Struck Actual	Cost: Total $	Feet above ground	Number of people injured
count	25558.000000	25558.000000	2.555800e+04	25429.000000	25558.000000
mean	253916.085609	2.691525	5.567354e+03	799.028432	0.001056
std	38510.453382	12.793975	1.219713e+05	1740.079843	0.050420
min	1195.000000	1.000000	0.000000e+00	0.000000	0.000000
25%	225783.750000	1.000000	0.000000e+00	0.000000	0.000000
50%	248749.000000	1.000000	0.000000e+00	50.000000	0.000000
75%	269168.750000	1.000000	0.000000e+00	700.000000	0.000000
max	321909.000000	942.000000	1.239775e+07	18000.000000	6.000000

In [9]:
```
df.isnull().sum()
```

Out[9]:
```
Record ID                                0
Aircraft: Type                         129
Airport: Name                          129
Altitude bin                           129
Aircraft: Make/Model                     0
Wildlife: Number struck                129
Wildlife: Number Struck Actual           0
Effect: Impact to flight               129
FlightDate                             129
Effect: Indicated Damage                 0
Aircraft: Number of engines?           267
Aircraft: Airline/Operator             129
Origin State                           449
When: Phase of flight                  129
Conditions: Precipitation                0
Remains of wildlife collected?           0
Remains of wildlife sent to Smithsonian  0
Remarks                               4771
Wildlife: Size                         129
Conditions: Sky                          0
Wildlife: Species                        0
Pilot warned of birds or wildlife?     129
Cost: Total $                            0
Feet above ground                      129
Number of people injured                 0
Is Aircraft Large?                     129
dtype: int64
```

In [10]:
```
df['Remarks'].unique()
```

Out[10]:
```
array(['FLT 753. PILOT REPTD A HUNDRED BIRDS ON UNKN TYPE. #1 ENG WAS SHUT DOWN A
ND DIVERTED TO EWR. SLIGHT VIBRATION. A/C WAS OUT OF SVC FOR REPAIRS TO COWLING,
FAN DUCT ACCOUSTIC PANEL. INGESTION. DENTED FAN BLADE #26 IN #1 ENG. HEAVY BLOOD
STAINS ON L WINGTIP',
       '102 CARCASSES FOUND. 1 LDG LIGHT ON NOSE GEAR WAS DAMAGED AND REPLACED.',
       'FLEW UNDER A VERY LARGE FLOCK OF BIRDS OVER APCH END OF RWY. NO DMG. JUST
A LOT OF BIRD DROPPINGS ON WINDSCREEN.',
       ...,
       'STRUCK BIRD ON RT FRONT DURING T/O. BIRD REPTD AS BROWN/WHITE. TWY.',
       'PILOTS REPORT STRIKING UNKNOWN BIRD ON RWY 21L BTWN TWY F & J. NO REMAINS
FOUND ON RWY OR ON A/C. NO DMG TO A/C.',
       'HIT CENTER OF RADOME, CAVING IN ABOUT 12". RADOME WAS REPLACED. CARCASS F
OUND IN SAFETY ARA ON RT SIDE OF RWY 22 AT INTXN OF RWY 18/36.'],
      dtype=object)
```

In [11]:
```
df['Origin State'].unique()
```

Out[11]:
```
array(['New York', 'Texas', 'Louisiana', 'Washington', 'Virginia', nan,
       'Delaware', 'DC', 'Georgia', 'Florida', 'California', 'Illinois',
       'Connecticut', 'Missouri', 'Rhode Island', 'Hawaii', 'Arizona',
       'Tennessee', 'South Carolina', 'South Dakota', 'New Jersey',
       'Colorado', 'Minnesota', 'Alabama', 'Ohio', 'Wisconsin',
       'Michigan', 'Massachusetts', 'Alaska', 'North Carolina',
       'Kentucky', 'Indiana', 'Oregon', 'Pennsylvania', 'New Hampshire',
       'Arkansas', 'Nevada', 'Mississippi', 'Maryland', 'Maine', 'Quebec',
       'Idaho', 'British Columbia', 'Utah', 'Nebraska', 'Iowa',
       'New Mexico', 'West Virginia', 'Oklahoma', 'North Dakota',
       'Vermont', 'Wyoming', 'Kansas', 'Prince Edward Island', 'Montana',
       'Puerto Rico', 'Ontario', 'Virgin Islands',
       'Newfoundland and Labrador', 'Alberta', 'Saskatchewan'],
      dtype=object)
```

In [12]:
```
df['When: Phase of flight'].unique()
```

Out[12]:
```
array(['Climb', 'Landing Roll', 'Approach', 'Take-off run', 'Descent',
       nan, 'Taxi', 'Parked'], dtype=object)
```

In [13]:
```
df['Aircraft: Type'].unique()
```

Out[13]:
```
array(['Airplane', nan], dtype=object)
```

In [14]:
```
df.columns
```

Out[14]:
```
Index(['Record ID', 'Aircraft: Type', 'Airport: Name', 'Altitude bin',
       'Aircraft: Make/Model', 'Wildlife: Number struck',
       'Wildlife: Number Struck Actual', 'Effect: Impact to flight',
       'FlightDate', 'Effect: Indicated Damage',
       'Aircraft: Number of engines?', 'Aircraft: Airline/Operator',
       'Origin State', 'When: Phase of flight', 'Conditions: Precipitation',
       'Remains of wildlife collected?',
       'Remains of wildlife sent to Smithsonian', 'Remarks', 'Wildlife: Size',
       'Conditions: Sky', 'Wildlife: Species',
       'Pilot warned of birds or wildlife?', 'Cost: Total $',
       'Feet above ground', 'Number of people injured', 'Is Aircraft Large?'],
      dtype='object')
```

In [15]:
```
df.size
```

Out[15]: 664508

In [16]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25558 entries, 0 to 25557
Data columns (total 26 columns):
 #   Column                                  Non-Null Count  Dtype         
---  ------                                  --------------  -----         
 0   Record ID                               25558 non-null  int64         
 1   Aircraft: Type                          25429 non-null  object        
 2   Airport: Name                           25429 non-null  object        
 3   Altitude bin                            25429 non-null  object        
 4   Aircraft: Make/Model                    25558 non-null  object        
 5   Wildlife: Number struck                 25429 non-null  object        
 6   Wildlife: Number Struck Actual          25558 non-null  int64         
 7   Effect: Impact to flight                25429 non-null  object        
 8   FlightDate                              25429 non-null  datetime64[ns]
 9   Effect: Indicated Damage                25558 non-null  object        
 10  Aircraft: Number of engines?            25291 non-null  object        
 11  Aircraft: Airline/Operator              25429 non-null  object        
 12  Origin State                            25109 non-null  object        
 13  When: Phase of flight                   25429 non-null  object        
 14  Conditions: Precipitation               25558 non-null  object        
 15  Remains of wildlife collected?          25558 non-null  bool          
 16  Remains of wildlife sent to Smithsonian 25558 non-null  bool          
 17  Remarks                                 20787 non-null  object        
 18  Wildlife: Size                          25429 non-null  object        
 19  Conditions: Sky                         25558 non-null  object        
 20  Wildlife: Species                       25558 non-null  object        
 21  Pilot warned of birds or wildlife?      25429 non-null  object        
 22  Cost: Total $                           25558 non-null  int64         
 23  Feet above ground                       25429 non-null  float64       
 24  Number of people injured                25558 non-null  int64         
 25  Is Aircraft Large?                      25429 non-null  object        
dtypes: bool(2), datetime64[ns](1), float64(1), int64(4), object(18)
memory usage: 4.7+ MB
```

In [18]:
```
df1 = df.copy()
df1
```

Out[18]:

	Record ID	Aircraft: Type	Airport: Name	Altitude bin	Aircraft: Make/Model	Wildlife: Number struck	Wildlife: Number Struck Actual	Effect Impact to flight
0	202152	Airplane	LAGUARDIA NY	> 1000 ft	B-737-400	Over 100	859	Engine Shut Down
1	208159	Airplane	DALLAS/FORT WORTH INTL ARPT	< 1000 ft	MD-80	Over 100	424	None
2	207601	Airplane	LAKEFRONT AIRPORT	< 1000 ft	C-500	Over 100	261	None
3	215953	Airplane	SEATTLE-TACOMA INTL	< 1000 ft	B-737-400	Over 100	806	Precautionary Landing
4	219878	Airplane	NORFOLK INTL	< 1000 ft	CL-RJ100/200	Over 100	942	None
...	...	...	...	...	...	...	...	...
25553	321151	Airplane	REDDING MUNICIPAL	> 1000 ft	EMB-120	1	1	None
25554	319677	Airplane	ORLANDO INTL	< 1000 ft	A-321	1	1	None
25555	319680	NaN	NaN	NaN	EC-135	NaN	1	NaN
25556	319679	Airplane	DETROIT METRO WAYNE COUNTY ARPT	< 1000 ft	B-757-200	1	1	None
25557	319593	Airplane	ABRAHAM LINCOLN CAPITAL ARPT	< 1000 ft	B-737-400	1	1	None

In [19]:
```
df1.dropna(subset=['Aircraft: Type','Airport: Name', 'Altitude bin','Wildlife: Nur
```

In [20]:
```
df1.isnull().sum()
```

Out[20]:
```
Record ID                                0
Aircraft: Type                           0
Airport: Name                            0
Altitude bin                             0
Aircraft: Make/Model                     0
Wildlife: Number struck                  0
Wildlife: Number Struck Actual           0
Effect: Impact to flight                 0
FlightDate                               0
Effect: Indicated Damage                 0
Aircraft: Number of engines?             0
Aircraft: Airline/Operator               0
Origin State                             0
When: Phase of flight                    0
Conditions: Precipitation                0
Remains of wildlife collected?           0
Remains of wildlife sent to Smithsonian  0
Remarks                                4605
Wildlife: Size                           0
Conditions: Sky                          0
Wildlife: Species                        0
Pilot warned of birds or wildlife?       0
Cost: Total $                            0
Feet above ground                        0
Number of people injured                 0
Is Aircraft Large?                       0
dtype: int64
```

In [21]:
```
df1.describe()
```

Out[21]:

	Record ID	Wildlife: Number Struck Actual	Cost: Total $	Feet above ground	Number of people injured
count	24747.000000	24747.000000	2.474700e+04	24747.000000	24747.000000
mean	254485.775165	2.689255	5.485157e+03	801.538449	0.000849
std	35581.910360	12.506021	1.231439e+05	1736.743268	0.047986
min	200011.000000	1.000000	0.000000e+00	0.000000	0.000000
25%	225827.000000	1.000000	0.000000e+00	0.000000	0.000000
50%	248552.000000	1.000000	0.000000e+00	50.000000	0.000000
75%	268974.500000	1.000000	0.000000e+00	700.000000	0.000000
max	321909.000000	942.000000	1.239775e+07	18000.000000	6.000000

In [22]:
```
df1.drop(labels='Remarks', axis=1,inplace=True)
```

In [23]:
```
df1.isnull().sum()
```

Out[23]:
```
Record ID                                0
Aircraft: Type                           0
Airport: Name                            0
Altitude bin                             0
Aircraft: Make/Model                     0
Wildlife: Number struck                  0
Wildlife: Number Struck Actual           0
Effect: Impact to flight                 0
FlightDate                               0
Effect: Indicated Damage                 0
Aircraft: Number of engines?             0
Aircraft: Airline/Operator               0
Origin State                             0
When: Phase of flight                    0
Conditions: Precipitation                0
Remains of wildlife collected?           0
Remains of wildlife sent to Smithsonian  0
Wildlife: Size                           0
Conditions: Sky                          0
Wildlife: Species                        0
Pilot warned of birds or wildlife?       0
Cost: Total $                            0
Feet above ground                        0
Number of people injured                 0
Is Aircraft Large?                       0
dtype: int64
```

In [24]:
```
df1
```

Out[24]:

	Record ID	Aircraft: Type	Airport: Name	Altitude bin	Aircraft: Make/Model	Wildlife: Number struck	Wildlife: Number Struck Actual	Effect Impact t fligh
0	202152	Airplane	LAGUARDIA NY	> 1000 ft	B-737-400	Over 100	859	Engine Shu Dow
1	208159	Airplane	DALLAS/FORT WORTH INTL ARPT	< 1000 ft	MD-80	Over 100	424	Non
2	207601	Airplane	LAKEFRONT AIRPORT	< 1000 ft	C-500	Over 100	261	Non
3	215953	Airplane	SEATTLE-TACOMA INTL	< 1000 ft	B-737-400	Over 100	806	Precautionar Landin
4	219878	Airplane	NORFOLK INTL	< 1000 ft	CL-RJ100/200	Over 100	942	Non
...	...	...	...	...	...	...	...	.
25552	319672	Airplane	SACRAMENTO INTL	< 1000 ft	B-737-700	1	1	Non
25553	321151	Airplane	REDDING MUNICIPAL	> 1000 ft	EMB-120	1	1	Non
25554	319677	Airplane	ORLANDO INTL	< 1000 ft	A-321	1	1	Non
25556	319679	Airplane	DETROIT METRO WAYNE COUNTY ARPT	< 1000 ft	B-757-200	1	1	Non
25557	319593	Airplane	ABRAHAM LINCOLN CAPITAL ARPT	< 1000 ft	B-737-400	1	1	Non

24747 rows × 25 columns

In [25]:
```
df1.to_excel(r'bird_strike_cleaned.xlsx',index = False)
```

In [ ]:
```
```