* Case Study On Airplane Crashes *

Import Libraries

In [1]:
```
import pandas as pd
import warnings
warnings.filterwarnings('ignore')
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
```

Extract Dataset

In [2]:
```
df=pd.read_csv('crash.csv')
```

In [3]:
```
df.head()
```

Out[3]:

	Date	Time	Location	Operator	Flight #	Route	Type	Registration	cn/In
0	09/17/1908	17:18	Fort Myer, Virginia	Military - U.S. Army	NaN	Demonstration	Wright Flyer III	NaN	1
1	07/12/1912	06:30	AtlantiCity, New Jersey	Military - U.S. Navy	NaN	Test flight	Dirigible	NaN	NaN
2	08/06/1913	NaN	Victoria, British Columbia, Canada	Private	-	NaN	Curtiss seaplane	NaN	NaN
3	09/09/1913	18:30	Over the North Sea	Military - German Navy	NaN	NaN	Zeppelin L-1 (airship)	NaN	NaN
4	10/17/1913	10:30	Near Johannisthal, Germany	Military - German Navy	NaN	NaN	Zeppelin L-2 (airship)	NaN	NaN

Showing Columns present in this dataset

In [4]:
```
df.columns
```

Out[4]:
```
Index(['Date', 'Time', 'Location', 'Operator', 'Flight #', 'Route', 'Type',
       'Registration', 'cn/In', 'Aboard', 'Fatalities', 'Ground', 'Summar
y'],
      dtype='object')
```

Defining shape of Above Data

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

Out[5]:
```
(5268, 13)
```

- There are 13 columns and 5268 rows

Finding Basic Information of Datasets

In [6]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5268 entries, 0 to 5267
Data columns (total 13 columns):
 #   Column        Non-Null Count  Dtype  
---  ------        --------------  -----  
 0   Date          5268 non-null   object 
 1   Time          3049 non-null   object 
 2   Location      5248 non-null   object 
 3   Operator      5250 non-null   object 
 4   Flight #      1069 non-null   object 
 5   Route         3562 non-null   object 
 6   Type          5241 non-null   object 
 7   Registration  4933 non-null   object 
 8   cn/In         4040 non-null   object 
 9   Aboard        5246 non-null   float64
 10  Fatalities    5256 non-null   float64
 11  Ground        5246 non-null   float64
 12  Summary       4878 non-null   object 
dtypes: float64(3), object(10)
memory usage: 535.2+ KB
```

- there is null data present in each column except date
- large number of null data is present in Flight column
- there is a total of 10 categorical and 3 numerical columns

Sum of Null Values in dataset

Finding null values from dataset

In [7]:
```
df.isnull().sum()
```

Out[7]:
```
Date               0
Time            2219
Location          20
Operator          18
Flight #        4199
Route           1706
Type              27
Registration     335
cn/In           1228
Aboard            22
Fatalities        12
Ground            22
Summary          390
dtype: int64
```

- by using this formula we can assume the null data present in columns

Finding percentage of null value from dataset

In [8]:
```
df.isnull().sum()/len(df)*100
```

Out[8]:
```
Date             0.000000
Time            42.122248
Location         0.379651
Operator         0.341686
Flight #        79.707669
Route           32.384207
Type             0.512528
Registration     6.359150
cn/In           23.310554
Aboard           0.417616
Fatalities       0.227790
Ground           0.417616
Summary          7.403189
dtype: float64
```

- Searching for columns having a large amount of null data to drop it
- if a column contains atleast 50% of null data it will be dropped
- we need Time column for making insights of situations

Dropping columns which has more than 50% of null data

In [9]:
```
df.drop(['Flight #', 'Registration', 'cn/In','Time'], axis=1,inplace=True)
```

- dropping Flight table as it contains a large amount of null data
- it contains around 79% of null data

Filling null data of remaining columns except time

In [10]:
```
df['Route'] = df['Route'].fillna('Not defined')
df['Summary'] = df['Summary'].fillna('No comments')
df['Operator'] = df['Operator'].fillna('Unknown')
df['Type'] = df['Type'].fillna('No Type')
```

In [11]:

```
df.head(20)
```

Out[11]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	der fl
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	diri ex
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	Th C
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	thu a
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hy be wa
5	03/05/1915	Tienen, Belgium	Military - German Navy	Not defined	Zeppelin L-8 (airship)	41.0	21.0	0.0	C at
6	09/03/1915	Off Cuxhaven, Germany	Military - German Navy	Not defined	Zeppelin L-10 (airship)	19.0	19.0	0.0	Ex b
7	07/28/1916	Near Jambol, Bulgeria	Military - German Army	Not defined	Schutte-Lanz S-L-10 (airship)	20.0	20.0	0.0	Cr
8	09/24/1916	Billericay, England	Military - German Navy	Not defined	Zeppelin L-32 (airship)	22.0	22.0	0.0	Sh Bri
9	10/01/1916	Potters Bar, England	Military - German Navy	Not defined	Zeppelin L-31 (airship)	19.0	19.0	0.0	Sh fla
10	11/21/1916	Mainz, Germany	Military - German Army	Not defined	Super Zeppelin (airship)	28.0	27.0	0.0	C
11	11/28/1916	Off West Hartlepool, England	Military - German Navy	Not defined	Zeppelin L-34 (airship)	20.0	20.0	0.0	Sh
12	03/04/1917	Near Gent, Belgium	Military - German Army	Not defined	Airship	20.0	20.0	0.0	ar

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	
13	03/30/1917	Off Northern Germany	Military - German Navy	Not defined	Schutte-Lanz S-L-9 (airship)	23.0	23.0	0.0	lig c
14	05/14/1917	Near Texel Island, North Sea	Military - German Navy	Not defined	Zeppelin L-22 (airship)	21.0	21.0	0.0	C th ar
15	06/14/1917	Off Vlieland Island, North Sea	Military - German Navy	Not defined	Zeppelin L-43 (airship)	24.0	24.0	0.0	Sh
16	08/21/1917	Off western Denmark	Military - German Navy	Not defined	Zeppelin L-23 (airship)	18.0	18.0	0.0	Sh
17	10/20/1917	Near Luneville, France	Military - German Navy	Not defined	Zeppelin L-44 (airship)	18.0	18.0	0.0	Sh F
18	04/07/1918	Over the Mediterranean	Military - German Navy	Not defined	Zeppelin L-59 (airship)	23.0	23.0	0.0	Ex c the
19	05/10/1918	Off Helgoland Island, Germany	Military - German Navy	Not defined	Zeppelin L-70 (airship)	22.0	22.0	0.0	Sh Bri cra

- it shows 1st 20 rows of data present in this file

finding out fatalities of people killed on ground

In [12]:
```
df['Ground'].value_counts()
```

Out[12]:
```
0.0       5027
1.0         53
2.0         27
3.0         21
4.0         15
5.0         10
8.0         10
7.0          8
11.0         6
6.0          6
22.0         5
13.0         4
24.0         3
10.0         3
44.0         3
20.0         3
14.0         2
2750.0       2
30.0         2
12.0         2
19.0         2
47.0         2
52.0         2
70.0         2
54.0         1
18.0         1
45.0         1
16.0         1
35.0         1
50.0         1
23.0         1
225.0        1
125.0        1
75.0         1
15.0         1
32.0         1
49.0         1
9.0          1
40.0         1
36.0         1
113.0        1
107.0        1
33.0         1
87.0         1
31.0         1
63.0         1
17.0         1
37.0         1
58.0         1
85.0         1
Name: Ground, dtype: int64
```

- finding out the fatalities by the planes with respect to ground
- as we can see there are some large amount of fatalities
- in 2 airplane crashes, nearly 2750 people died.

In [13]:
```
df['Survived'] = df['Aboard'] - (df['Fatalities'] + df['Ground'])
df.Survived = np.where(df.Survived < 0, 0, df.Survived)
df.head()
```

Out[13]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fl thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro wl bein was s

finding unique records present in columns

In [14]:
```
df.nunique()
```

Out[14]:
```
Date          4753
Location      4303
Operator      2477
Route         3245
Type          2447
Aboard         239
Fatalities     191
Ground          50
Summary       4674
Survived       165
dtype: int64
```

transfroming statistical information of certain columns for better understanding

In [15]:
```
df.describe().T
```

Out[15]:

	count	mean	std	min	25%	50%	75%	max
Aboard	5246.0	27.554518	43.076711	0.0	5.0	13.0	30.0	644.0
Fatalities	5256.0	20.068303	33.199952	0.0	3.0	9.0	23.0	583.0
Ground	5246.0	1.608845	53.987827	0.0	0.0	0.0	0.0	2750.0
Survived	5236.0	7.421314	28.108456	0.0	0.0	0.0	2.0	516.0

- maximum 516 people survived in airplane accidents
- minimum 0 people are died in plane crashes
- maximum 583 people died in plane crashes
- 75% of airplane crashes resembles death count more than 20
- maximum 644 people boraded a plane
- maxmimum 2750 people died on ground due to airplane crashes
- average 20 people are died in airplane crashes
- average of 27 people abroaded a flight
- nearly 50% of airplane crashed resembles death of 9 people more or less

Accident of most people abroaded a flight

In [16]:
```
df.loc[df['Aboard'] == 644]
```

Out[16]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	Summary	Su
2963	03/27/1977	Tenerife, Canary Islands	Pan American World Airways / KLM	Tenerife - Las Palmas / Tenerife - Las Palmas	Boeing B-747-121 / Boeing B-747-206B	644.0	583.0	0.0	Both aircraft were diverted to Tenerife becaus...	

In [17]:
```
df.loc[2963].Summary
```

Out[17]:
```
'Both aircraft were diverted to Tenerife because of a bombing at Las Palmas
Airport. After an extended delay,  both planes were instructed to back track
up the runway. The KLM plane reached its takeoff point while the Pan Am plan
e was still on the runway. The Pan Am plane continued up the runway missing
the taxiway turnout. There was heavy fog on the runway. The KLM plane began
its takeoff roll without permission with the Pan Am plane still on the runwa
y. The KLM plane hit the Pan Am plane just as it was taking off. Both planes
burst into flames.  KLM 234 + 14 crew,  Pan Am 326 + 9 crew killed. All aboa
rd the KLM plane were killed. The Pan Am aircraft was named Clipper Victor.
The KLM aircraft was named Rhine River.'
```

This incident of most people killed occured on 27 March 1977

Incident of most people survived

In [18]:
```
df.loc[df['Survived'] == 516]
```

Out[18]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	Summary	Su
4645	07/23/1999	Tokyo, Japan	All Nippon Airways	Tokyo, Japan - Chitose, Japan	Boeing B-747	517.0	1.0	0.0	Two minutes after taking off from Haneda Airpo...	

In [19]:
```
df.loc[4645].Summary
```

Out[19]:
```
'Two minutes after taking off from Haneda Airport, a man carrying a knife fo
rced a flight attendant to take him in the cockpit of the plane.  A fan of c
omputer flight-simulation games, he stated he just wanted to fly a real plan
e.  After forcing the co-pilot out of the cockpit he ordered the captain to
fly to a U.S. Air Force base in western Tokyo. When he refused, he stabbed t
he captain and seized the controls. After a sudden drop in altitude, the co-
pilot and an off duty crew member entered the cockpit and overpowered the hi
jacker.  A one point the plane plunged to within 984 feet of the ground. The
plane ultimately  landed safely but the captain died of his injuries.'
```

This incident of most people survived occured on 23 July 1999

Incident of most people died on ground

In [20]:
```
df.loc[df['Ground']==2750]
```

Out[20]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	Summary	Su
4803	09/11/2001	New York City, New York	American Airlines	Boston - Los Angeles	Boeing 767-223ER	92.0	92.0	2750.0	The aircraft was hijacked shortly after it lef...	
4804	09/11/2001	New York City, New York	United Air Lines	Boston - Los Angeles	Boeing B-767-222	65.0	65.0	2750.0	The aircraft was hijacked shortly after it lef...	

In [21]:
```
df.loc[4803].Summary
```

Out[21]:
```
'The aircraft was hijacked shortly after it left Logan International Airport
in Boston. The hijackers took control of the aircraft and deliberately crash
ed it into the north tower of the World Trade Center between the 94th and 99
th floors at approximately 450 mph.  After 102 minutes, the building collaps
ed. It was one of four planes that were hijacked the same day.'
```

In [22]:
```
df.loc[4804].Summary
```

Out[22]:
```
'The aircraft was hijacked shortly after it left Logan International Airport
in Boston. The hijackers took control of the aircraft and deliberately crash
ed it into the south tower of the World Trade Center between the 78th and 84
th floors at approximately 550 mph. After 56 minutes, the building collapse
d. It was one of four planes that were hijacked the same day.'
```

This two incidents of most people killed on the ground relates to same place which occured on 11 September 2001

In [23]:
```
df.head(5)
```

Out[23]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fle thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro wl bein was s

Average of people abroading in each flight

In [24]:
```
df['Aboard'].value_counts().mean()
```

Out[24]:
```
21.94979079497908
```

- On an average 22 count of people abroad each flight

Showing Most Accidental Locations

In [25]:
```
df['Location'].value_counts()
```

Out[25]:
```
Sao Paulo, Brazil                      15
Moscow, Russia                         15
Rio de Janeiro, Brazil                 14
Anchorage, Alaska                      13
Manila, Philippines                    13
                                       ..
Near Charana, Bolivia                   1
Monte Matto, Italy                      1
Misaki Mountain, Japan                  1
Angelholm, Sweden                       1
State of Arunachal Pradesh, India       1
Name: Location, Length: 4303, dtype: int64
```

- Most incidents are occured in Brazil,Russia,Alask and Philippines.

In [26]:
```
df.head()
```

Out[26]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fle thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro w bein was s

Finding data of most accidental area in World i.e Brazil(Sao Paulo)

In [27]:
```
df[(df['Location']=='Sao Paulo, Brazil')][['Location','Summary']]
```

Out[27]:

	Location	Summary
469	Sao Paulo, Brazil	The mail plane crashed while taking off.
664	Sao Paulo, Brazil	Crashed in fog.
836	Sao Paulo, Brazil	Crashed into the Solimoes extension of the Ama...
1148	Sao Paulo, Brazil	Crashed into a house shortly after taking off ...
1203	Sao Paulo, Brazil	Crashed while attempting to make an emergency ...
1269	Sao Paulo, Brazil	Crashed while on final approach to Sao Paulo. ...
1327	Sao Paulo, Brazil	The cargo plane crashed on takeoff. Elevator l...
1406	Sao Paulo, Brazil	Crashed on takeoff.
1619	Sao Paulo, Brazil	Crashed a few minutes after taking off from Sa...
1828	Sao Paulo, Brazil	The crippled airliner crashed into houses and ...
1848	Sao Paulo, Brazil	The aircraft returned to airport after the No....
3601	Sao Paulo, Brazil	The crew accidently tried to take off from a t...
4406	Sao Paulo, Brazil	The crew was advised they were too high and fa...
5159	Sao Paulo, Brazil	The jet airliner crashed while attempting to l...
5177	Sao Paulo, Brazil	The executive jet took off, banked to the righ...

In [28]:
```
df[(df['Fatalities']>20) & (df['Location']=='Sao Paulo, Brazil')]
```

Out[28]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	Summary	S
1406	12/19/1955	Sao Paulo, Brazil	Cruzeiro Do Sud	Sao Paulo - Belem	Douglas DC-3	26.0	26.0	0.0	Crashed on takeoff.	
1848	05/03/1963	Sao Paulo, Brazil	Cruzeiro	Sao Paulo - Rio de Janeiro	Convair CV-340-59	50.0	37.0	0.0	The aircraft returned to airport after the No....	
5159	07/17/2007	Sao Paulo, Brazil	TAM (Brazil)	Porto Alegre - Sao Paulo	Airbus A-320-233	187.0	187.0	12.0	The jet airliner crashed while attempting to l...	

- There are 3 main incidents to observe which caused more than 20 deaths in a single flight crash
- only one of 3 incidents people survived on a minimum count of 13
- The flight from Porto Alegre resulted atmost deaths in a single crash in which not a single person survived but caused a death of 187 people

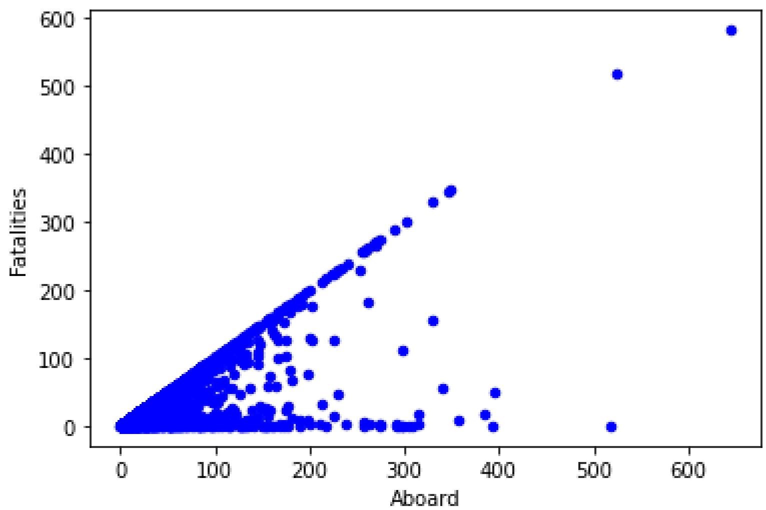
In [29]:
```
df.head()
```

Out[29]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fle thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro wl bein was s

Showing correlation between abroad and fatalities

In [30]:
```python
df.plot(x="Aboard", y="Fatalities", kind="scatter",color='blue')
plt.show()
```



1.There is a positive correlation between abroad and fatalities

2.As there is increase in abroading passengers there is increase in fatalities

In [31]:
```python
df.head()
```

Out[31]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fle thund and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro w bein was s

Finding out count of accidents on different routes

In [32]:
```
df['Route'].value_counts()
```

Out[32]:
```
Not defined                     1706
Training                          81
Sightseeing                       29
Test flight                       17
Test                               6
                                ... 
Manila - Lapu Lapu                 1
Saint Denis - Paris                1
Cork - London                      1
Peoria, IL - St. Louis, MO         1
Mechuka for Jorhat                 1
Name: Route, Length: 3245, dtype: int64
```

- Most accidents are occured while training, sighseeing and test flight

Creating New columns named Day,Month,Year,Decade,Date of Year

In [33]:
```
df['Month'] = pd.DatetimeIndex(df['Date']).month
df['Year'] = pd.DatetimeIndex(df['Date']).year
df['Decade'] = (df['Year']) // 10 * 10
df['Day of Week'] = pd.DatetimeIndex(df['Date']).day_name()
```

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

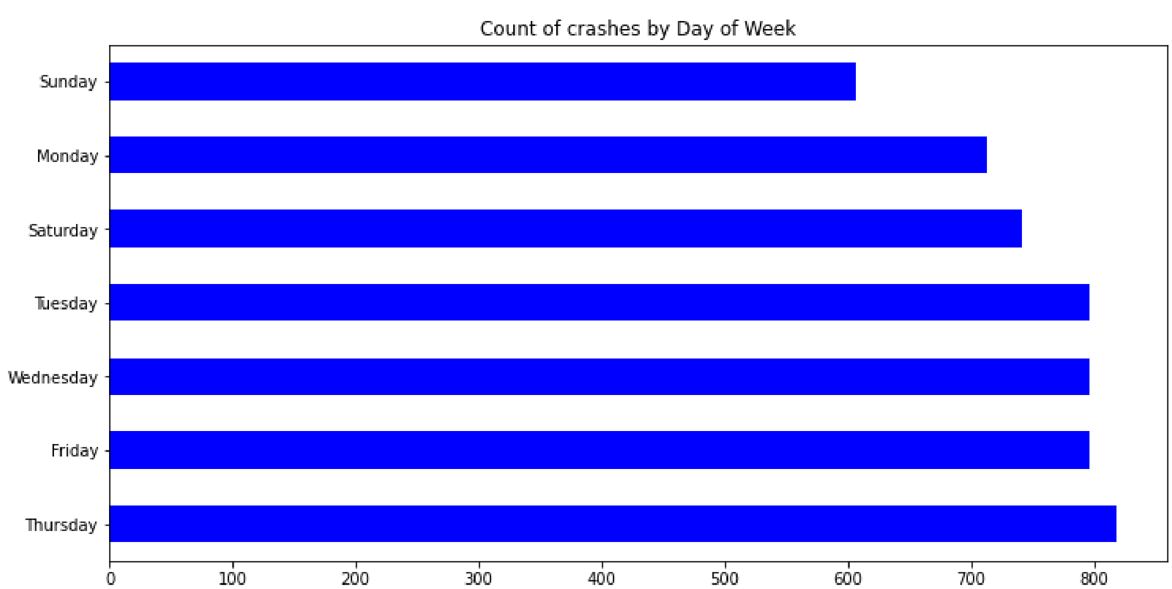
Out[34]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlanticCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fle thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro wl bein was s

Plot the graph of showing counts of accidents on weekdays

In [35]:
```
df['Day of Week'].value_counts().plot(kind='barh', figsize=[12, 6], title='Cou
```

Out[35]: `<AxesSubplot:title={'center':'Count of crashes by Day of Week'}>`



- From this bar graph we can assume that sunday has the least amount of accidents

- While Thursday has most accidents through the week
- Tuesday, Wednesday and Friday have same amount of accidents

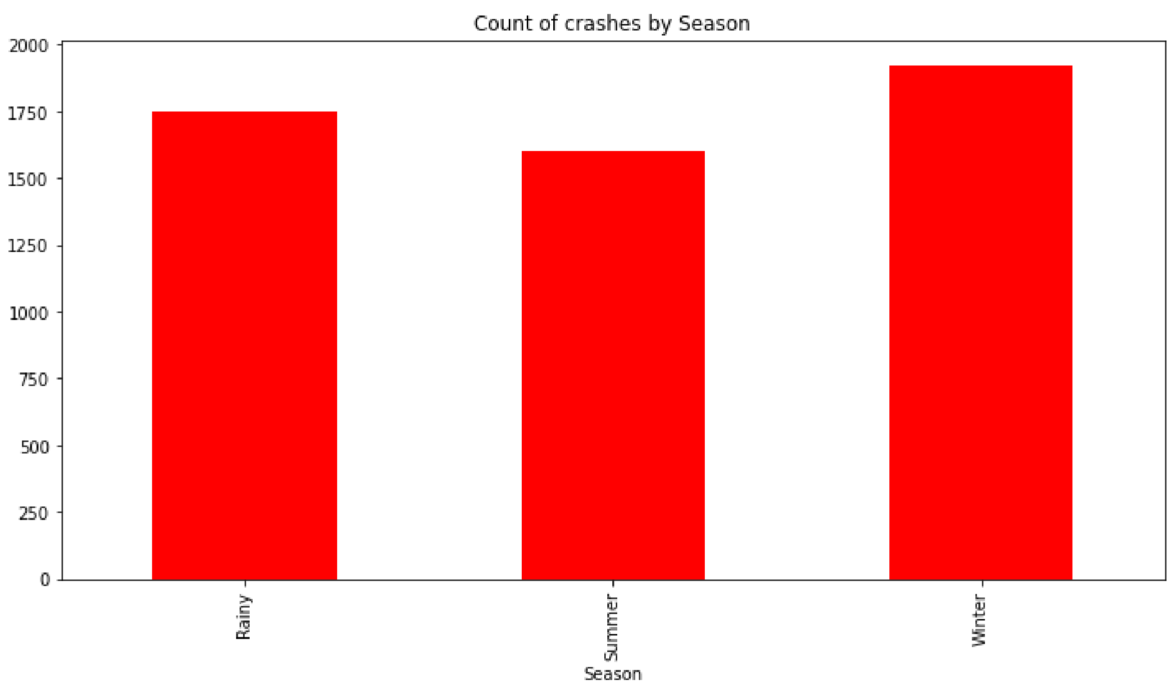
Plotting Bar graph to see Accidents with respect to Seasons

In [36]:
```
def get_season(month):
    if month >= 2 and month <= 5:
        return 'Summer'
    elif month >= 6 and month <= 9:
        return 'Rainy'
    else:
        return 'Winter'

df['Season'] = df['Month'].apply(get_season)

crashed_by_season = df['Season'].groupby(df['Season']).count()
crashed_by_season.plot(kind='bar', figsize=[12, 6], title='Count of crashes by
crashed_by_season
```

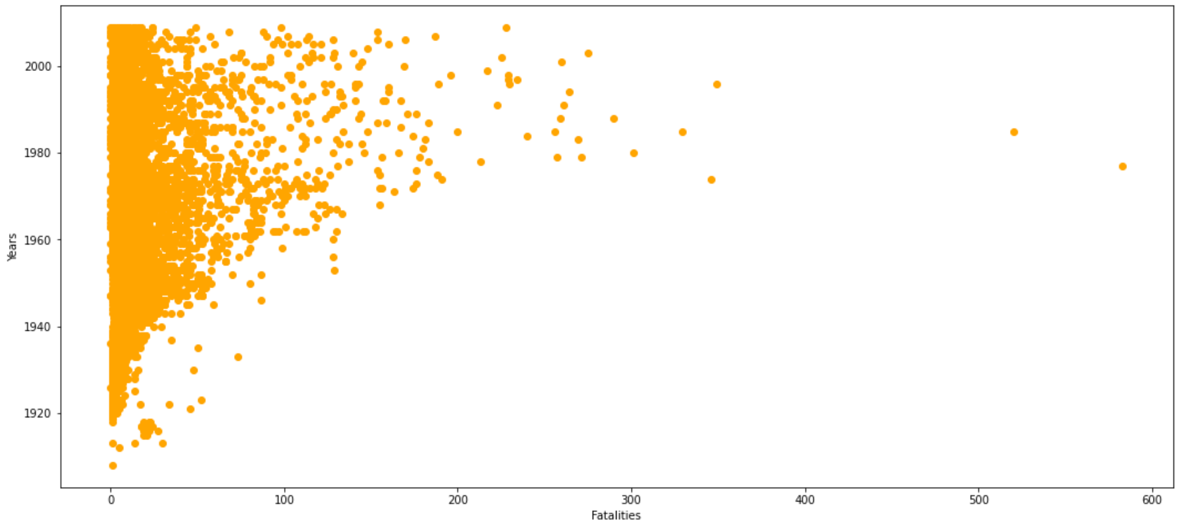
Out[36]:
```
Season
Rainy     1747
Summer    1599
Winter    1922
Name: Season, dtype: int64
```



- As compared there are more incidents in Winter season as compared to Rainy season and Summer season
- Summer season has the lowest amount of accidents through the other 2 seasons

Plotting to see Fatalities with respect to years

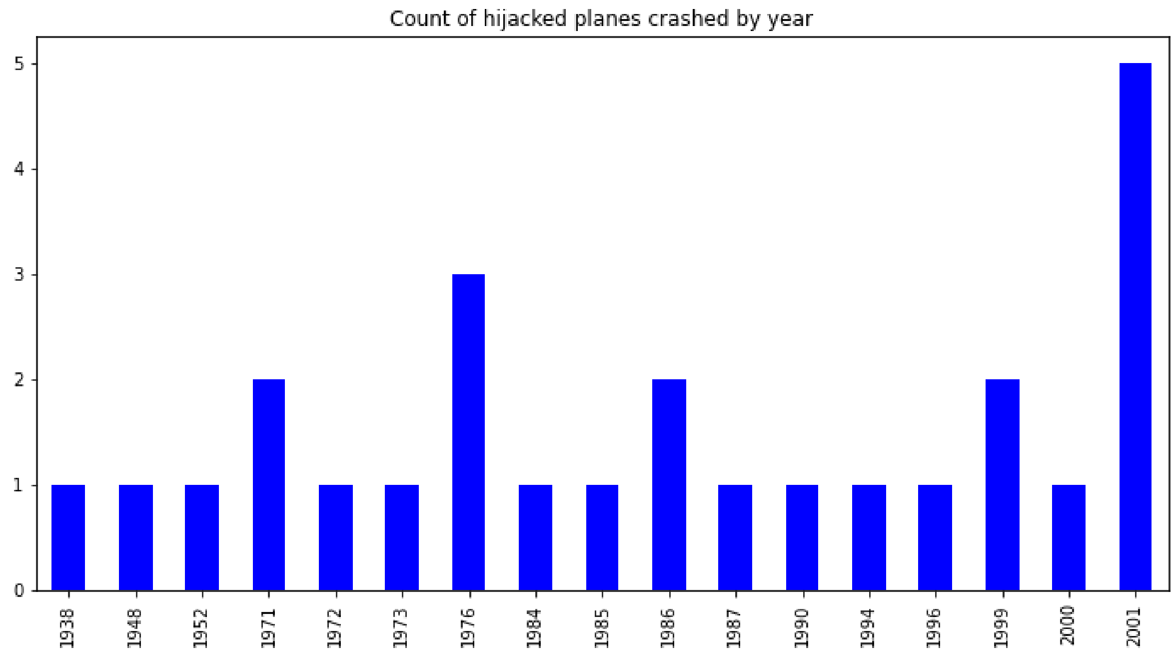
In [37]:
```
plt.figure(figsize= (18, 8))
years = df['Year']
plt.plot(df['Fatalities'], years, 'o',color='orange')
plt.xlabel('Fatalities')
plt.ylabel('Years')
plt.show()
```



- most airplane accidents occured after 1940
- There is a serious amount of fatalities in years 1960 to 2000

Plotting bar graph for incidents in which plane is highjacked with respect to year

In [38]:
```
df[df['Summary'].str.contains('hijacke')]['Year'].value_counts().sort_index().
plt.show()
```



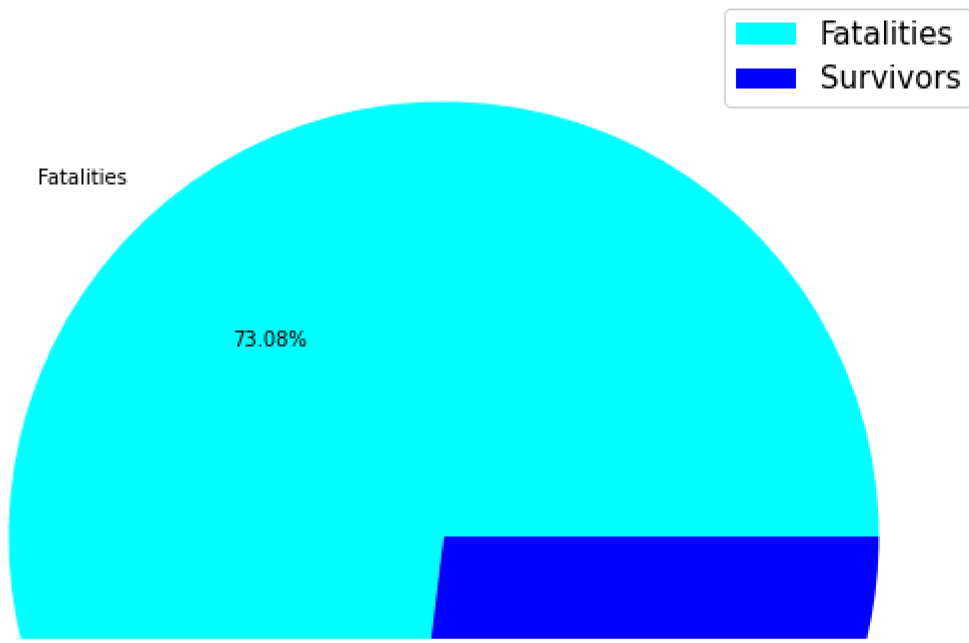
- Most airplanes are hijacked in the year 2001 followed by year 1976

Plotting pie diagram to see difference between fatalities and survivors

In [39]:
```
sns.set_palette('pastel')
plt.figure(figsize=(15,10))
Survived=df.Survived.sum()
Fatalities = df.Fatalities.sum()

y = np.array([Fatalities, Survived])
mylabels = ["Fatalities", "Survivors"]

plt.pie(y, labels = mylabels,autopct='%1.2f%%',colors=['cyan','blue'])
plt.legend(fontsize=15)
plt.show()
```

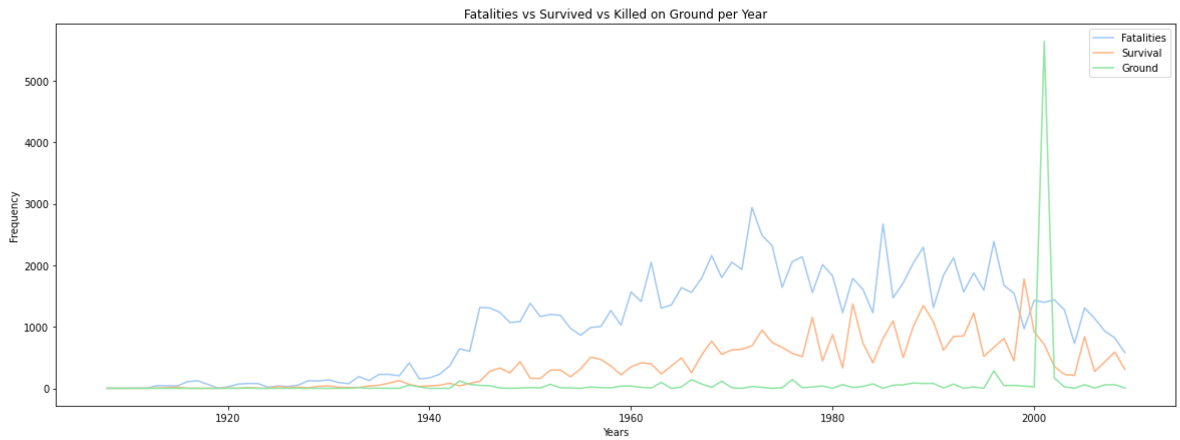


- There is a serious amount of fatalities as compared to survivors
- Around 73% of fatalities are recorded in accidents as compared to 27% of survivors
- It shows a guarantee of more fatalities in a single accident than survivors

Showing realtion between Fatalities, Survivors and People killed on Ground

In [40]:
```
FSG_per_year = df[['Year', 'Fatalities', 'Survived', 'Ground']].groupby('Year'
```

In [41]:
```
plt.figure(figsize=(20,7))
sns.lineplot(x = 'Year', y = 'Fatalities', data = FSG_per_year,palette='bright
sns.lineplot(x = 'Year', y = 'Survived', data = FSG_per_year)
sns.lineplot(x = 'Year', y = 'Ground', data = FSG_per_year)
plt.legend(['Fatalities', 'Survival', 'Ground'])
plt.xlabel('Years')
plt.ylabel('Frequency')
plt.title('Fatalities vs Survived vs Killed on Ground per Year')
plt.show()
```



- By assuming this data we clearly state that more fatalities are occured
- Survival count is very low with respect to fatalities
- In a certain there are few incidents on ground which caused highest amount of deaths
- Very few incidents have more survivals than fatalities

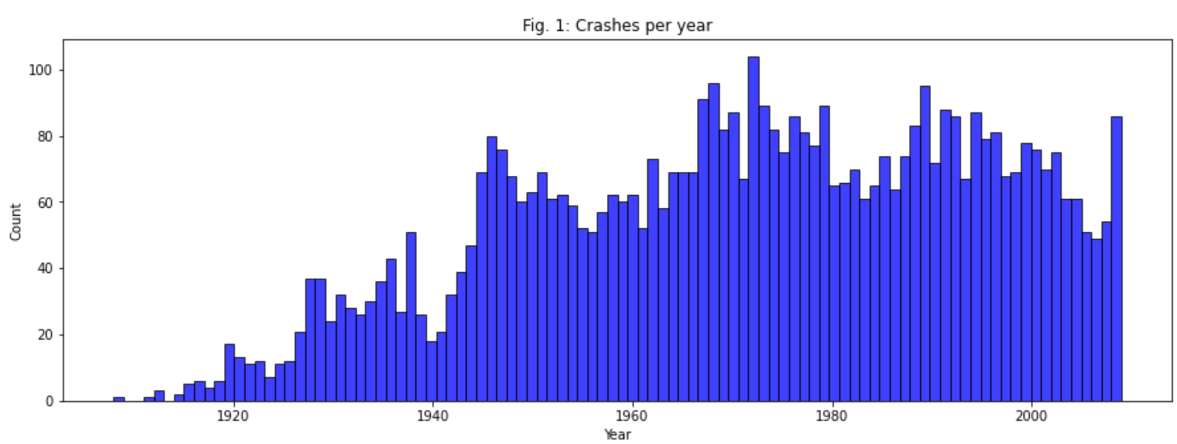
In [42]:
```
df.head()
```

Out[42]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlanticCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fl thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro wl bein was s

Plotting a bar graph to show Crashes per year

In [43]:
```
plt.figure(figsize=(15, 5))
per_year_plot = sns.histplot(data = df, x = 'Year', bins = 100, color = "blue'
per_year_plot.set(title = "Fig. 1: Crashes per year")
fig = per_year_plot.get_figure()
fig.savefig('per_year_plot.png')
```



- There are very much incidents occured in certain years
- After 1940 there is a medium amount of increase in Airplane crashes

- While on the other side, there is a tremendous growth in accidents from year 1960 which is near to constant uptil 2000

Showing fatalities with respect to model of airplanes

In [44]:
```
df_fatal = df.groupby('Type')[['Fatalities']].sum()
df_fatal
```

Out[44]:

	Fatalities
Type	
AAC-1 Toucan	23.0
AEGK	5.0
AT L98 Carvair	4.0
ATR 42-300	2.0
ATR-42-300	46.0
...	...
de Havilland Dove 1	22.0
de Havilland Dragon 1	3.0
de Havilland RU-6A Beaver /Bell UH-1H	18.0
de havilland Canada Twin Otter 200	11.0
deHavilland DH-86	9.0

2447 rows × 1 columns

- This shows the accidents in all certain models of airplanes which are present in data
- By analyzing it shows that AAC-1, Atr-42-300 and Bell UH-1H plane types have occured in such accidents the most

Showing Top 10 Type of airplanes as per accidents

In [45]:
```
df_fatal = df_fatal.rename(columns={'Type': 'Fatalities'})
df_type_fatal = df_fatal.sort_values(by='Fatalities')
df_type_fatal_top10 = df_fatal.head(10)
df_type_fatal_top10
```

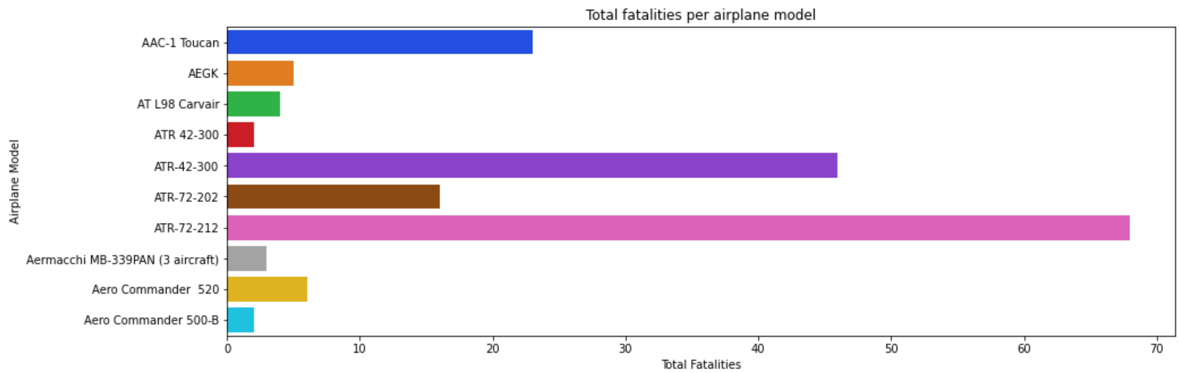
Out[45]:

	Fatalities
Type	
AAC-1 Toucan	23.0
AEGK	5.0
AT L98 Carvair	4.0
ATR 42-300	2.0
ATR-42-300	46.0
ATR-72-202	16.0
ATR-72-212	68.0
Aermacchi MB-339PAN (3 aircraft)	3.0
Aero Commander 520	6.0
Aero Commander 500-B	2.0

- The ATR models have faced more situations than other models
- As we can see ATR-42-300 and ATR-72-212 have most accidents followed type AAC-1 Toucan
- As the other side Aero commander types has the least accidents than other airplane types

Plotting bar graph to see in detail about Types of models of airplanes and fatalities caused by them

In [46]:
```
plt.figure(figsize=(15, 5))
sns.barplot(y=df_type_fatal_top10.index, x="Fatalities", data=df_type_fatal_to
plt.xlabel('Total Fatalities')
plt.ylabel('Airplane Model')
plt.title('Total fatalities per airplane model')
plt.show()
```



- By seeing the above graph we can state that ATR-72-212 has the highest amount of fatalities which is 68
- While compared Aero Commander 500-B and ATR-42-300 has the least amount of deaths i.e. 2

Creating a column Total death for further plottings

In [47]:
```
df["Total Death"] = df["Fatalities"] + df["Ground"]
df.head()
```

Out[47]:

	Date	Location	Operator	Route	Type	Aboard	Fatalities	Ground	S
0	09/17/1908	Fort Myer, Virginia	Military - U.S. Army	Demonstration	Wright Flyer III	2.0	1.0	0.0	demo fligh A
1	07/12/1912	AtlantiCity, New Jersey	Military - U.S. Navy	Test flight	Dirigible	5.0	5.0	0.0	F dirigib explo
2	08/06/1913	Victoria, British Columbia, Canada	Private	Not defined	Curtiss seaplane	1.0	1.0	0.0	The ac Can
3	09/09/1913	Over the North Sea	Military - German Navy	Not defined	Zeppelin L-1 (airship)	20.0	14.0	0.0	Th fle thun and
4	10/17/1913	Near Johannisthal, Germany	Military - German Navy	Not defined	Zeppelin L-2 (airship)	30.0	30.0	0.0	Hydro w bein was

- Total death is a total of fatals in airplane accident and people which died because of plane crash on ground

Showing Top 10 airline operators which result in most total death

In [48]:
```
df_death_airline = df.groupby('Operator')[['Total Death']].sum()
df_death_airline = df_death_airline.sort_values(by='Total Death', ascending=Fa
df_death_airline_top10 = df_death_airline.head(10)

df_death_airline_top10
```

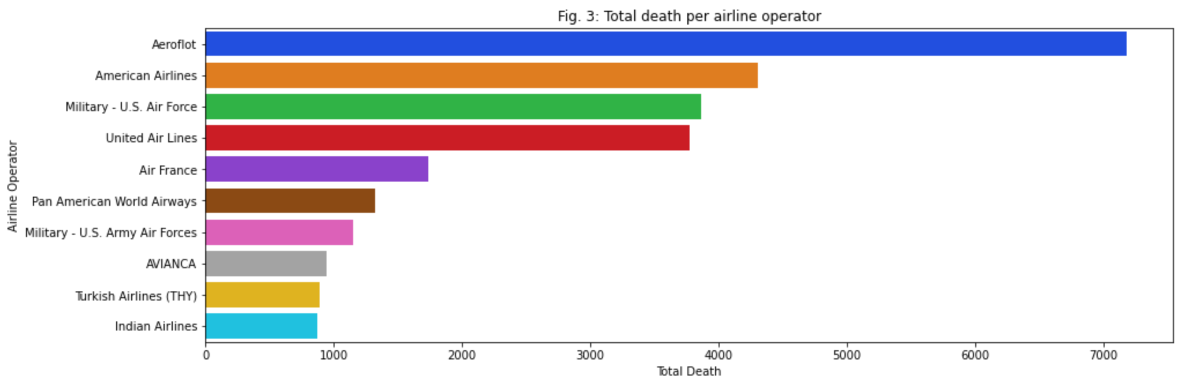
Out[48]:

	Total Death
Operator	
Aeroflot	7184.0
American Airlines	4310.0
Military - U.S. Air Force	3866.0
United Air Lines	3770.0
Air France	1739.0
Pan American World Airways	1322.0
Military - U.S. Army Air Forces	1150.0
AVIANCA	944.0
Turkish Airlines (THY)	891.0
Indian Airlines	870.0

- As we can see Aeroloft has caused most amount of deaths which 7184 followed by American Airline
- Also note the point that Indian Airlines has the least amount of deaths throughout all 10
- Also there is high amount of difference in total death between Aeroloft and American Airlines

Plotting Bar Graph for showing above data in detail

In [49]:
```
plt.figure(figsize=(15, 5))
sns.barplot(y=df_death_airline_top10.index, x="Total Death", data=df_death_air
plt.xlabel('Total Death')
plt.ylabel('Airline Operator')
plt.title('Fig. 3: Total death per airline operator')
plt.show()
```



- As we can see there is growth in total death in American company operators airplanes

In [ ]:
```
```