Data Visualization of Bird Strike 2000-2011

Rishi Jhangili

Table of Contents

Project Recap

In our bird strike analysis project, we began by collecting extensive data from various aviation authorities. Using Jupyter Notebook, we imported the data with pandas and performed thorough cleaning, addressing missing values and identifying outliers with the help of libraries seaborn. Once cleaned, the data was loaded into SQL for deeper analysis. Our initial findings revealed a significant increase in bird strikes from 2000 to 2011, peaking in 2010. Geographical analysis showed that most data came from the USA, with only 277 records from Canada, providing a clearer view of the states involved. We found that 'SOUTHWEST AIRLINES' and 'BUSINESS' airlines, both based in the USA, encountered the most bird strikes. Financially, bird strikes cost $23,259,457 in 2001 alone.

Further analysis indicated that bird strikes were more frequent under neutral weather conditions and at altitudes below 1000 ft, making this the most vulnerable range. The approach phase of flight saw the most bird strikes, with some incidents also occurring when aircraft were parked. Our findings also highlighted that when no action was taken, bird strikes caused minimal damage. However, precautionary measures like engine shutdowns resulted in more severe damage. During an aborted take-off, bird strikes caused less damage. In the 'precautionary landing' phase, the probabilities of damage and no damage were nearly equal. Additionally, a large number of bird strikes below 1000 ft resulted in 42,650 incidents without any effect, while 5,976 caused damage. Analysis of pilot warnings showed almost the same result whether the pilot was warned or not about potential bird strikes. These insights underline the importance of effective bird strike mitigation strategies to enhance aviation safety and reduce related costs.

Project Description

Transport and communication are in the crucial domain in the field of analytics. Environmental impacts and safety are, nowadays, two major concerns of the scientific community with respect to transport scenarios and to the ever-growing urban areas. These issues gain more importance due to the increasing amount of vehicles and people. Seeking new solutions is reaching a point where available technologies and artificial intelligence, especially MAS, are being recognized as ways to cope with and tackle these kinds of problems in a distributed more appropriate way.

A bird strike is strictly defined as a collision between a bird and an aircraft which is in flight or on a take-off or landing roll. The term is often expanded to cover other wildlife strikes - with bats or ground animals. Bird Strike is common and can be a significant threat to aircraft safety. For smaller aircraft, significant damage may be caused to the aircraft structure and all aircraft, especially jet-engine ones, are vulnerable to the loss of thrust which can follow the ingestion of birds into engine air intakes. This has resulted in several fatal accidents.

Bird strikes may occur during any phase of flight, but are most likely during the take-off, initial climb, approach and landing phases due to the greater numbers of birds in flight at lower levels. To have a closer look the following document visually depicts the data collected on Bird Strikes by FAA between 2000-2011.

Case Study

- ❑ Visuals Depicting the Number of Bird Strikes
- ❑ Yearly Analysis & Bird Strikes in the US
- ❑ Top 10 US Airlines in terms of having encountered bird strikes

- ❑ Airports with most incidents of bird strikes – Top 50
- ❑ Yearly Cost Incurred due to Bird Strikes
- ❑ When do most bird strikes occur?
- ❑ Altitude of the Airplane at the time of the strike
- ❑ Phase of flight at the time of the strike.
- ❑ Average Altitude of the Airplane in different phases at the time of strike
- ❑ Effect of Bird Strikes & Impact on Flight
- ❑ Effect of Strike at Different Altitude
- ❑ Were Pilots Informed? & Prior Warning and Effect of Strike Relation

Process

Data Understanding

Data Cleaning

Data Modelling

Data Analysis

Uncover Insights

Tech Stack-Used

MySQL

Jupyter Notebook

Tableau

Microsoft Excel

Microsoft Word

Data Cleaning

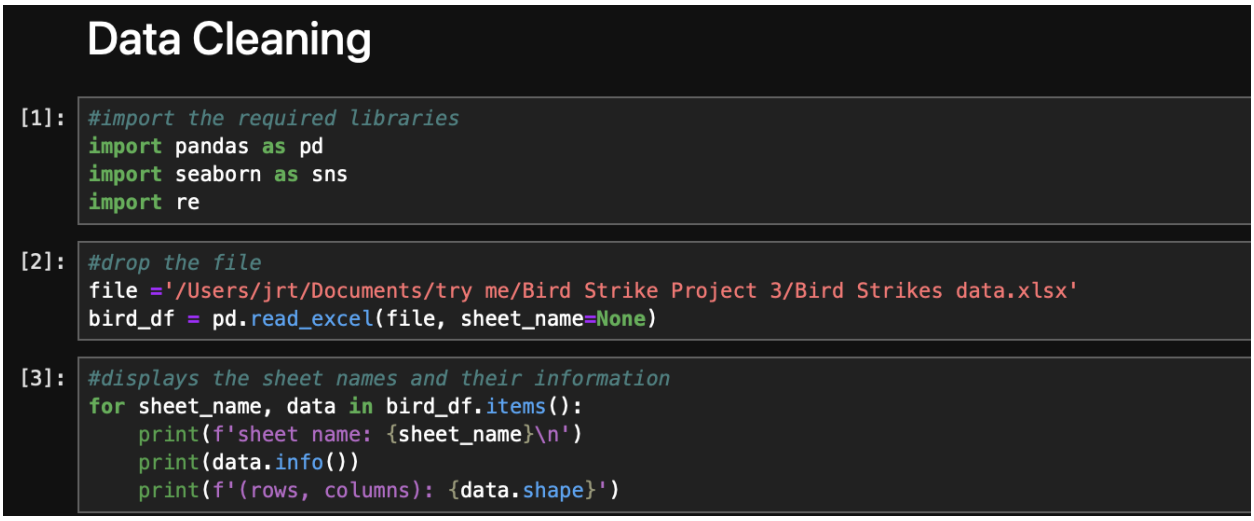
I am utilizing various Python libraries such as Pandas, Seaborn, and re for data cleaning tasks. Pandas provides powerful data manipulation and analysis tools, making it easier to handle and process large datasets. Seaborn, a statistical data visualization library, helps in identifying and rectifying data anomalies through comprehensive visualizations. The re library, which is used for regular expression operations, assists in cleaning and

Preprocessing text data by enabling efficient pattern matching and substitution. Together, these tools streamline the data preparation process, ensuring accuracy and efficiency in subsequent analyses.

The Data contains 26 columns and 25558 rows

Import the Pandas, Seaborn and re libraries

Read the Data from Excel and remove duplicates



Remove the unwanted columns from the data frame

Use ‘re’ library to fix column names

```
unwanted_columns = ['Aircraft: Type', 'Aircraft: Type', 'Wildlife: Number struck', 'Aircraft: Number of engines?',
                    'Remains of wildlife collected?', 'Remains of wildlife sent to Smithsonian', 'Remarks',
                    'Wildlife: Size', 'Number of people injured', 'Is Aircraft Large?']
```

```
df = df.drop(columns = unwanted_columns)
```

```
#function to clean column names
def clean_column_names(df):
    df.columns = [re.sub(r'[^\w\s]', '', col).replace(' ','_').lower() for col in df.columns]
    return df
```

```
df = clean_column_names(df)
```

```
df.info()
print('data shape:', df.shape)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25558 entries, 0 to 25557
Data columns (total 17 columns):
```

The remaining columns are 17

Find nulls and check before removing. If the null percentage is greater than 50 then we can drop the columns

The higher nulls are in 'effect_impact_to_flight' 91% and 'conditions_precipitation' 92%. The actual values are ‘none’ but pandas read nulls as. So, change those values as not null to none.

```
#<NA> values are 'none' where the impact of flight and conditions have no change {pandas read none as nulls/blanks}
for col in higher_nulls:
    df[col] = df[col].fillna('none')
```

```
#removing special character in aircraft_airlines
df['aircraft_airline'] = df['aircraft_airline'].str.replace(r'[^\w\s]','', regex = True)
```

```
#find out remaining null values also
remaining_nulls = ['airport_name', 'altitude_bin', 'aircraft_airline', 'origin_state', 'phase_of_flight',
                   'pilot_warned_of_wildlife']
```

```
total = df.isnull().sum()
percent = df.isnull().sum()/df.shape[0]*100
nulls = pd.DataFrame({'total_nulls': total,
                      'percentage': percent,
                      'data_type': df.dtypes})
nulls
```

Do the needful if there are any changes inside the column

Change to object to string data type as we perform analysis in MySQL so it will easy to understand and readable.

Find Outliers

Find outliers by describing the data frame and plot the graph by using seaborn for clear understanding.

```
sns.boxplot(df['wildlife_number_struck_actual'])
```

The min value is 1

The max value is 942. Which we can find the outliers in this like in a day the max bird strike is 942 which is impossible.

There are different methods to remove the outliers. I used percentile method.

By using ‘capping’ we can change the outliers to upper limit and lower limit vice versa

And there is another column (total_cost) with outliers. But in terms we cannot remove each and every outlier.

Missing Values

Find the null values along with the data typer numerical data types fill with median and for string types do mode. Fill the missing dates with forward fill or backward fill.

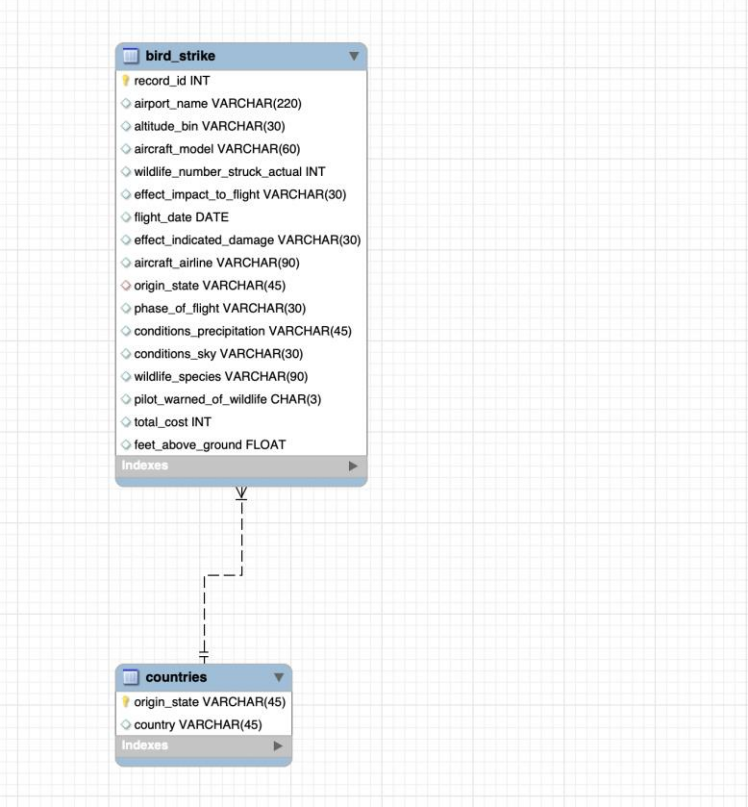
Save the file into CSV or you can choose your format.

Data Modelling

Once the database is selected, create a table within it. This initial table should contain all relevant columns with their respective data types. Use a load or import wizard provided by the database management system to import the necessary data into this table efficiently.

Next, create another table that will be used specifically for analysis and data modeling. The State column in this new table should correspond with the State column in the original “Bird Strike” table, ensuring that data can be accurately linked and analyzed. This second table will serve as the foundation for modeling, allowing for a structured and efficient examination of bird strike patterns and impacts across different states.

The common attribute between the “Bird Strike” table and the “Country” table is the “State” attribute. In the “Bird Strike” table, the “State” attribute represents the location where the bird strike incidents occurred. Meanwhile, in the “Country” table, the “State” attribute provides information about various states within a country. By linking these two tables through the “State” attribute, we can analyze bird strike incidents in the context of specific states, facilitating regional assessments and enabling targeted measures for wildlife hazard management and aviation safety.



We can see that origin state is the common attribute between these two tables.

By using countries tables we can find the “country name” of the state which is present in the bird strike table.

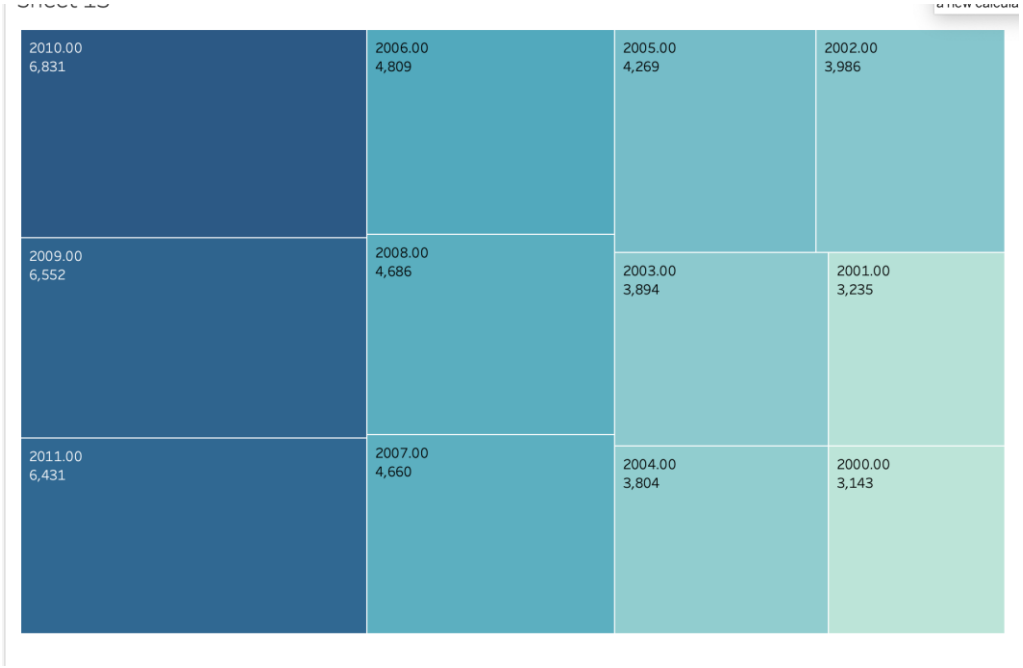
Data Analysis

```
SELECT * FROM bird_strike;
```

```
SELECT * FROM countries;
```

01. Visuals Depicting the Number of Bird Strikes

```
/*Visuals Depicting the Number of Bird Strikes
*/
with total as(
SELECT
    DISTINCT EXTRACT(YEAR FROM flight_date) AS `year`,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY EXTRACT(YEAR FROM flight_date)) AS total_birds_strike
FROM
    bird_strike) select sum(total_birds_strike) from total;
```

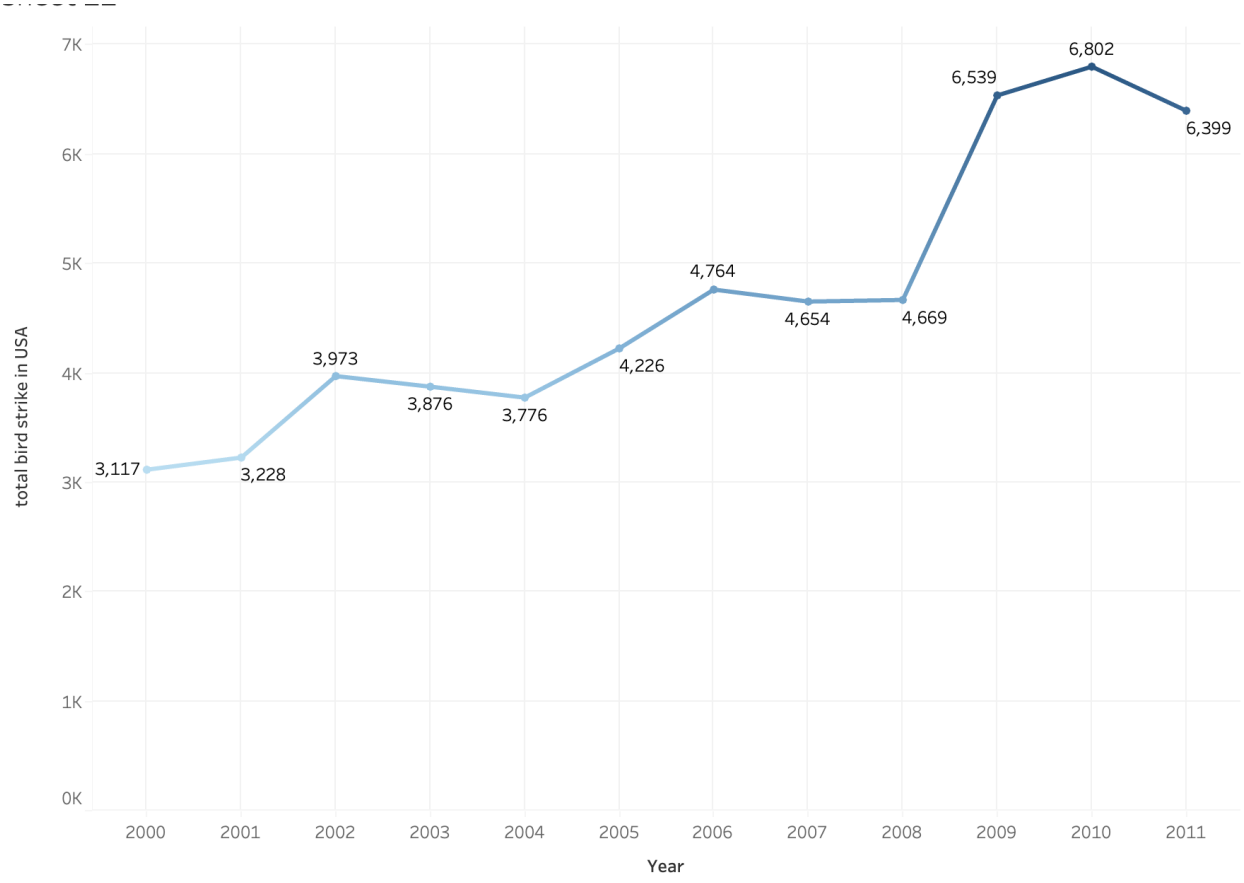


Insights

We can see that the number of bird strike is increased from year 2000 to 2011. And the highest bird strike counts in year 2010.

02. Yearly Analysis & Bird Strikes in the US

```sql
/*Yearly Analysis & Bird Strikes in the US
*/
with total as(
SELECT
    EXTRACT(YEAR FROM flight_date) AS `year`,
    SUM(wildlife_number_struck_actual) AS total_bird_strike_in_USA
FROM
    bird_strike bs
        INNER JOIN
    countries c ON bs.origin_state = c.origin_state
WHERE
    country = 'United States'
GROUP BY `year`
ORDER BY `year`) select sum(total_bird_strike_in_USA) from total;
```



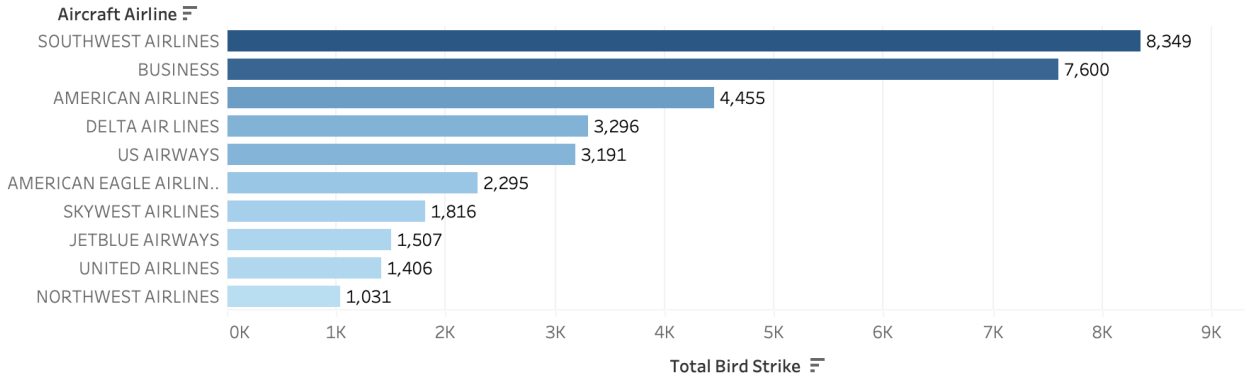
Insight

Here we are joining the country table to get which states belong to the USA. There will be a slight variation in the data from task 1 as most of the data is from the USA. Only 277 records belong to Canada.

03. Top 10 US Airlines in terms of having encountered bird strikes

```sql
/*Top 10 US Airlines in terms of having encountered bird strikes
*/

SELECT
    DISTINCT aircraft_airline,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY aircraft_airline) AS total_bird_strike
FROM
    bird_strike
ORDER BY
    total_bird_strike DESC
LIMIT 10;
```



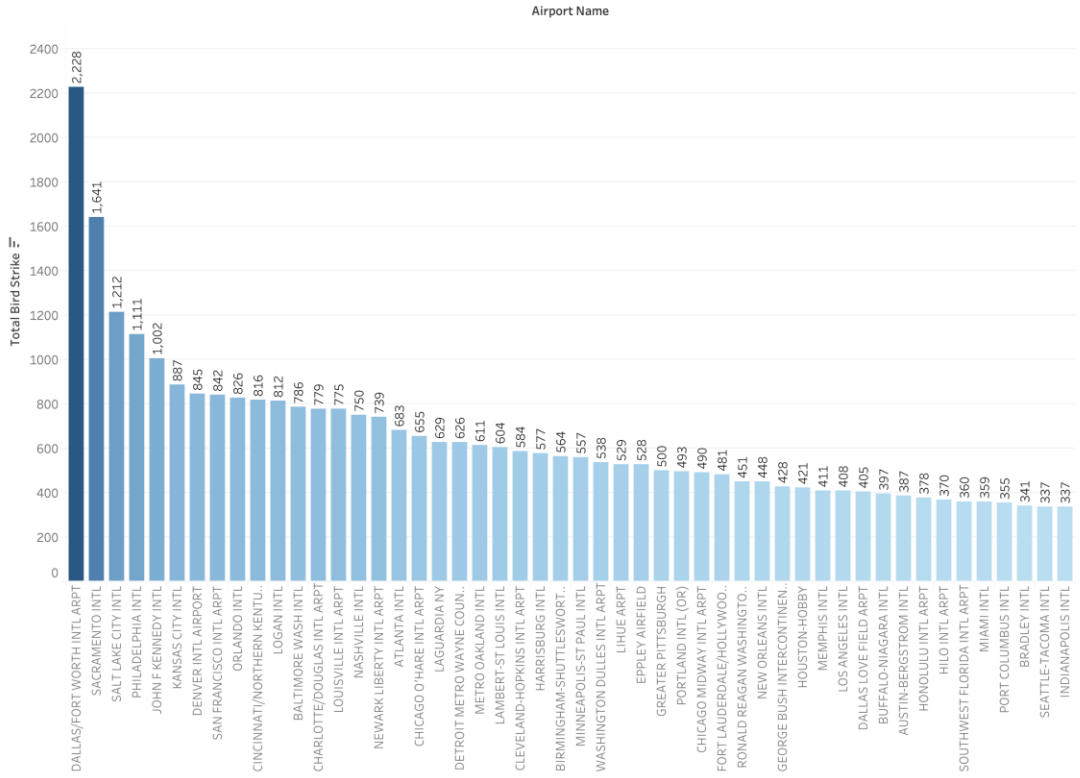
Insights

The airlines ‘SOUTHWEST AIRLINES’ and ‘BUSINESS’ have encountered the greatest number of birds strikes which are located in USA.

04. Airports with most incidents of bird strikes – Top 50

```sql
/*Airports with most incidents of bird strikes – Top 50
*/

SELECT
    DISTINCT airport_name,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY airport_name) AS total_bird_strike
FROM
    bird_strike
ORDER BY total_bird_strike DESC
LIMIT 50;
```



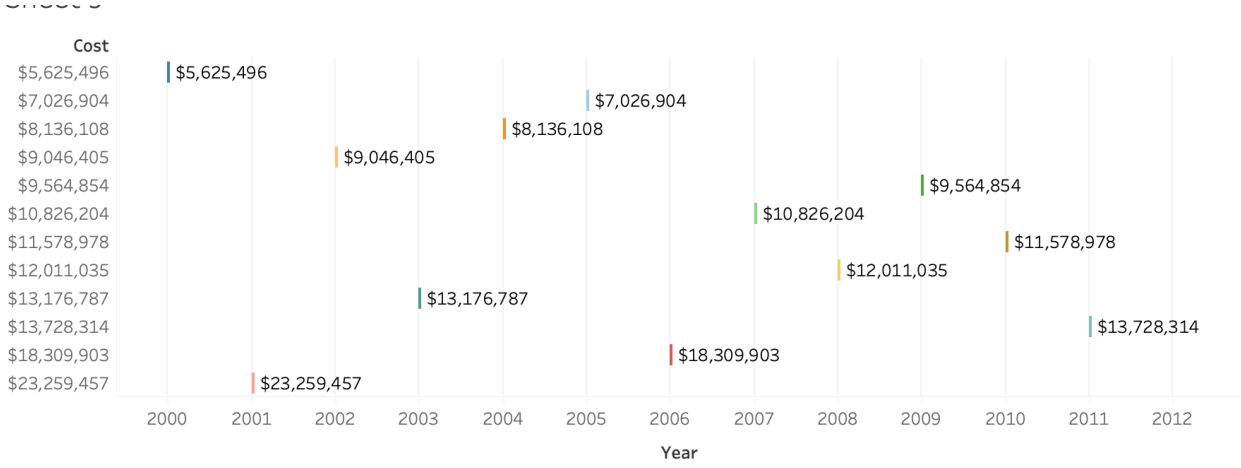
Insight

These are the top 50 airports with the greatest amount of bird strikes. The Dallas internation airport with 2200 above bird strikes is the first in the list.

05. Yearly Cost Incurred due to Bird Strikes

```sql
/*Yearly Cost Incurred due to Bird Strikes:
*/

WITH yearly_cost AS (
    SELECT
        DISTINCT EXTRACT(YEAR FROM flight_date) AS `year`,
        SUM(total_cost) OVER(PARTITION BY EXTRACT(YEAR FROM flight_date)) AS cost_per_year
    FROM
        bird_strike
    -- ORDER BY cost_per_year DESC
)
SELECT
    `year`,
    CONCAT('$', FORMAT(cost_per_year,0)) AS cost
FROM
    yearly_cost;
```

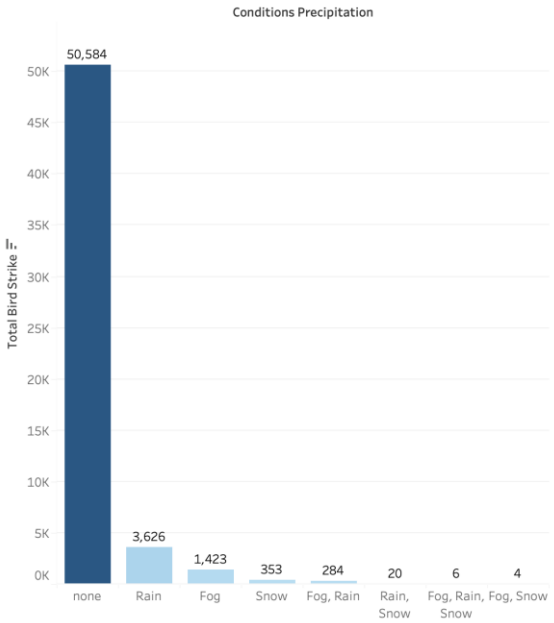


Insights

From the Gantt Bar, we can see that in the year 2001, the cost incurred due to bird strikes as $23,259,457.

06. When do most bird strikes occur?

```sql
SELECT
    DISTINCT conditions_precipitation,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY conditions_precipitation) AS total_bird_strike
FROM
    bird_strike
ORDER BY
    total_bird_strike DESC;
```

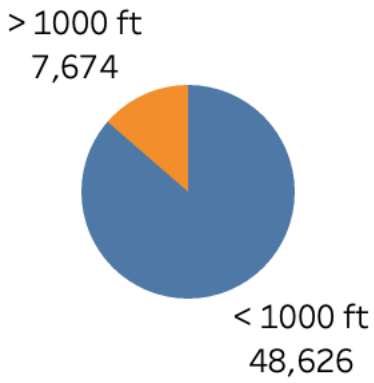


Insights

As we can see when the climatic condition is neutral (none) the bird strike is more compared to the other conditions.

07. Altitude of the Airplane at the time of the strike

```sql
SELECT
    DISTINCT altitude_bin AS altitude,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY altitude_bin) AS total_bird_strike
FROM
    bird_strike
ORDER BY
    total_bird_strike DESC;
```

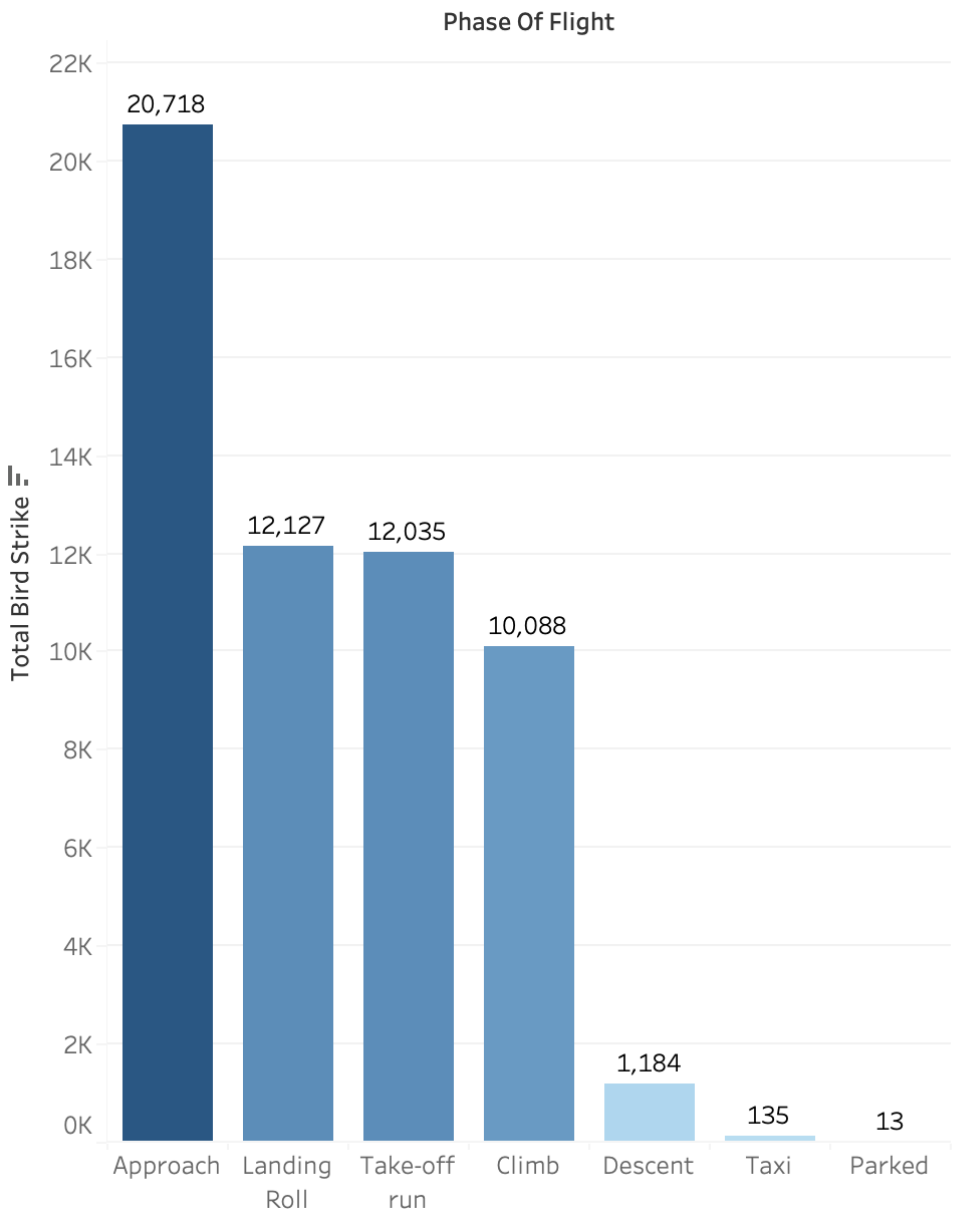


Insights

The majority of bird strikes occurred at an altitude of “< 1000 ft”.

08. Phase of flight at the time of the strike.

```sql
SELECT
    DISTINCT altitude_bin AS altitude,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY altitude_bin) AS total_bird_strike
FROM
    bird_strike
ORDER BY
    total_bird_strike DESC;
```

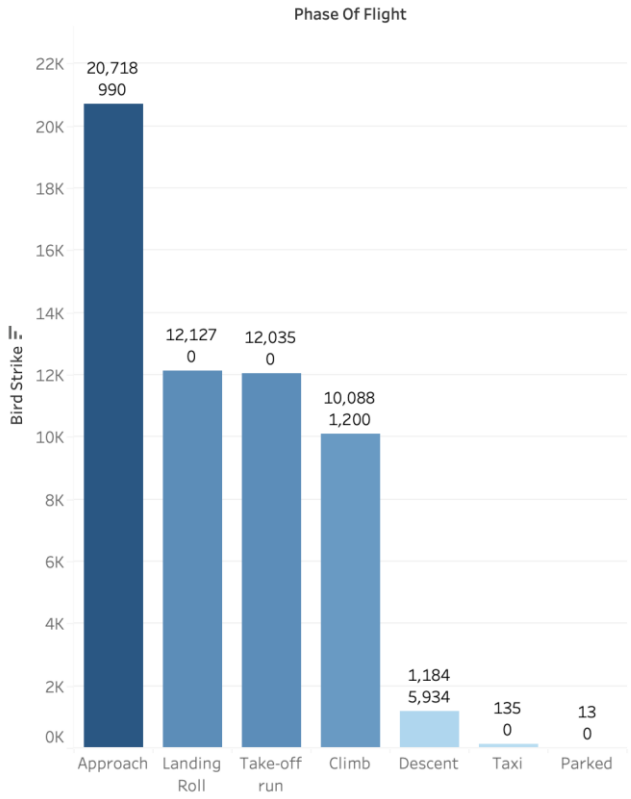


Insights

We observe that more bird strikes occur when the flight is approaching, and there are also some bird strikes when the flight is parked.

09. Average Altitude of the Aeroplan in different phases at the time of the strike

```sql
SELECT
    DISTINCT phase_of_flight,
    ROUND(AVG(feet_above_ground) OVER(PARTITION BY phase_of_flight),2) AS altitude_average,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY phase_of_flight) AS bird_strike
 FROM
    bird_strike
ORDER BY
    bird_strike DESC;
```

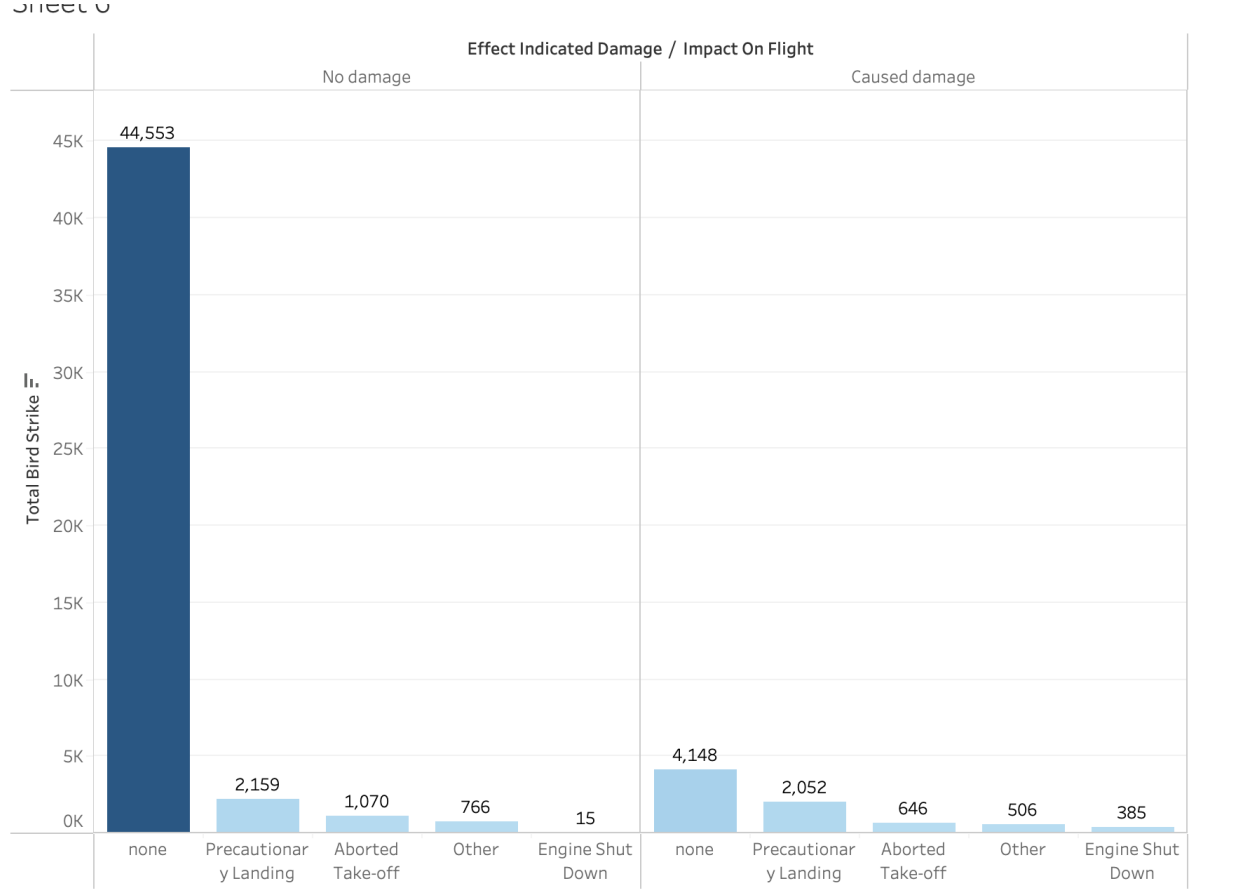


Insights

Before have a similar column chart. As we see now with the average altitude of each flight at each phase. During the 'Descent' phase of flight, the average altitude is high enough that most birds cannot approach and in the 'Approach' phase of flight, the average altitude is medium, so most bird strikes occur during this phase (this is the clear explanation for previous task).

10. Effect of Bird Strikes & Impact on Flight

```sql
SELECT
    DISTINCT effect_impact_to_flight AS impact_on_flight,
    effect_indicated_damage,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY effect_indicated_damage, effect_impact_to_flight) AS total_bird_strike
FROM
    bird_strike
ORDER BY
    total_bird_strike DESC;
```



Impact On Flight

Effect Indicated ..	Aborted Take-off	Engine Shut Down	none	Other	Precautionary Landing
No damage	1,070	15	44,553	766	2,159
Caused damage	646	385	4,148	506	2,052

Insights

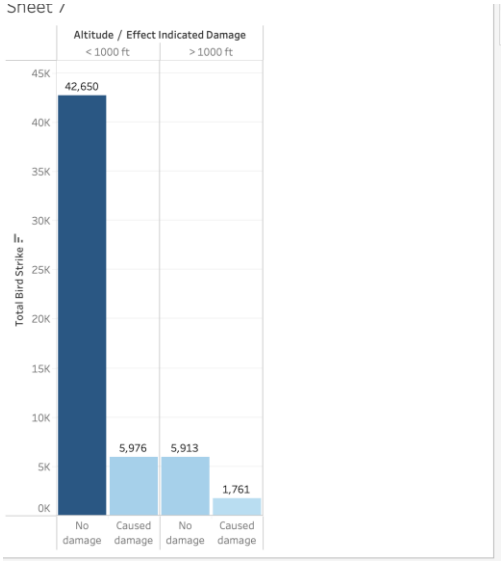
When no action is taken, bird strikes do not cause much damage.

However, when precautions like engine shutdown are taken, bird strikes cause more damage. In contrast, during an aborted take-off,

bird strikes cause less damage. Additionally, during the ‘precautionary landing’ phase, the probability of no damage and the probability, of damage have only a slight difference.

11. Effect of Strike at Different Altitude

```sql
SELECT
    DISTINCT altitude_bin AS altitude,
    effect_indicated_damage,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY effect_indicated_damage, altitude_bin) AS total_bird_strike
FROM
    bird_strike
ORDER BY
    altitude, total_bird_strike DESC;
```

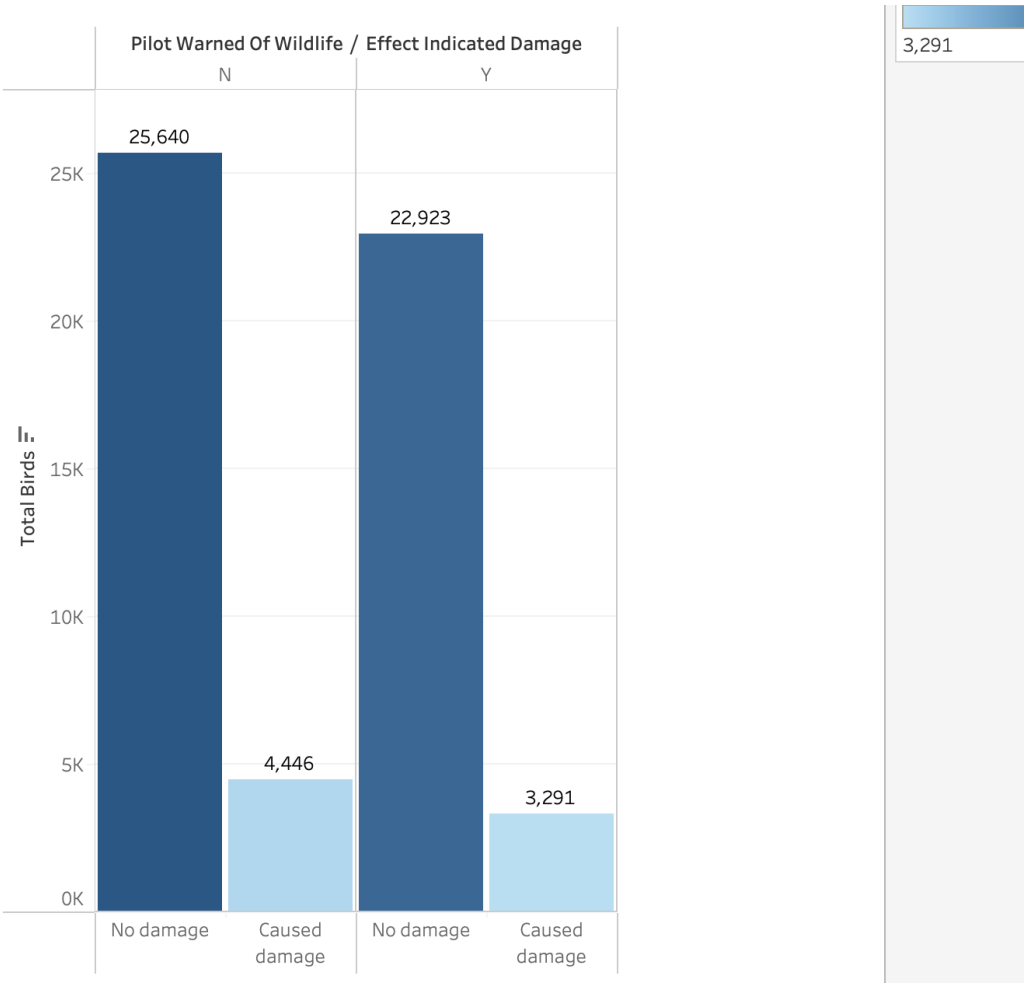


Insights

From task 7, we know that “a large number of bird strikes occur below 1000 ft”. In this case, we found that bird strikes do not impact the aero planes much; there were 42,650 bird strikes without any effect, while 5,976 caused damage.

12. Were Pilots Informed? & Prior Warning and Effect of Strike Relation

```sql
SELECT
    DISTINCT pilot_warned_of_wildlife,
    effect_indicated_damage,
    SUM(wildlife_number_struck_actual) OVER(PARTITION BY effect_indicated_damage, pilot_warned_of_wildlife) AS total_birds
FROM
    bird_strike
ORDER BY
    total_birds DESC;
```



Insights

From the graph, we can see that “there is almost the same result”, whether the pilot is “warned” or “not warned”.

Conclusion

In conclusion, our analysis of bird strike data revealed several key insights crucial for improving aviation safety. We observed a significant increase in bird strikes from 2000 to 2011, with the highest frequency occurring in 2010. The majority of incidents were concentrated in the USA, particularly affecting 'SOUTHWEST AIRLINES' and 'BUSINESS' airlines, and Dallas International Airport emerged as the most affected airport. Financially, bird strikes imposed a substantial cost, exemplified by the $23,259,457 incurred in 2001. Bird strikes were most frequent under neutral weather conditions and predominantly occurred at altitudes below 1000 ft, with the approach phase of flight being the most vulnerable. While higher altitude flights and the descent phase saw fewer incidents, precautionary measures such as engine shutdowns were found to increase damage severity. Aborted take-offs resulted in less damage, and the precautionary landing phase showed nearly equal probabilities of damage and no damage.

Additionally, bird strikes below 1000 ft often resulted in no significant effect, although some did cause damage. Pilot warnings had minimal impact on outcomes. These findings underscore the importance of implementing effective bird strike mitigation strategies to enhance safety and reduce economic losses in aviation.

Thank You