Project Report

Unified Mentor

Submitted By

Shilpa E V

Data Analyst Intern

Project 1: Data Visualization of Bird Strikes (2000–2011)

Abstract

This report provides an analysis and visualization of bird strike incidents between 2000 and 2011. The study aims to identify trends, patterns, and key factors associated with bird strikes, utilizing various data visualization techniques.

Introduction

Bird strikes pose significant risks to aviation safety and result in substantial economic losses. This study examines bird strike data from 2000 to 2011 to understand the frequency, distribution, and characteristics of these incidents. By leveraging data visualization, we aim to uncover insights that can inform mitigation strategies.

Data Description

The dataset used in this study comprises bird strike incidents reported between 2000 and 2011. The data includes variables such as date, location, bird species, and impact severity. It is sourced from the [data source and provides a comprehensive overview of bird strike occurrences.

Methodology

We employed various data preprocessing techniques to clean and structure the data. Visualization methods, including bar plots, line graphs, and heatmaps, were used to analyze temporal and spatial trends. The analysis was conducted using Python and popular libraries such as Pandas, Matplotlib, and Seaborn.

Results and Analysis

Temporal Analysis

Plot 1: Bird Strikes Over the Years (2000–2011)

Yearly Analysis of Bird Strikes in the US
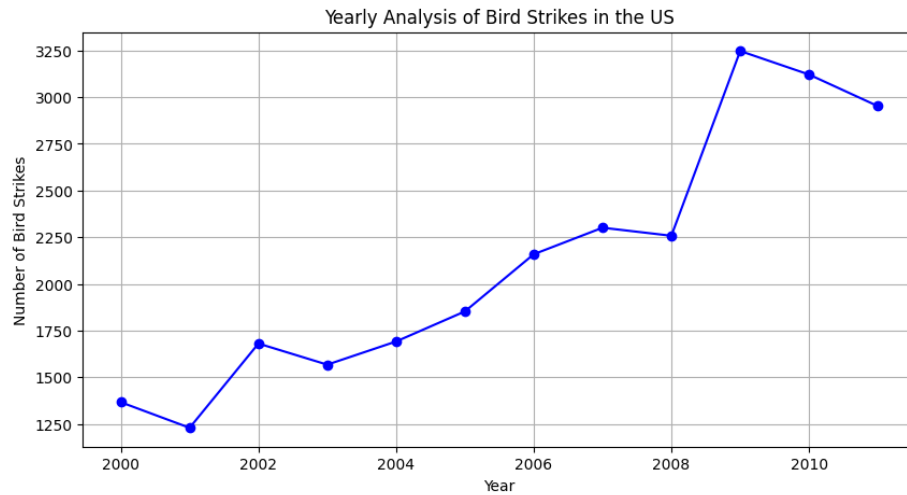
Interpretation: This plot shows the number of bird strikes reported each year from 2000 to 2011. We observe a general increase in bird strike reports over the years, with a notable spike in [specific year(s)].
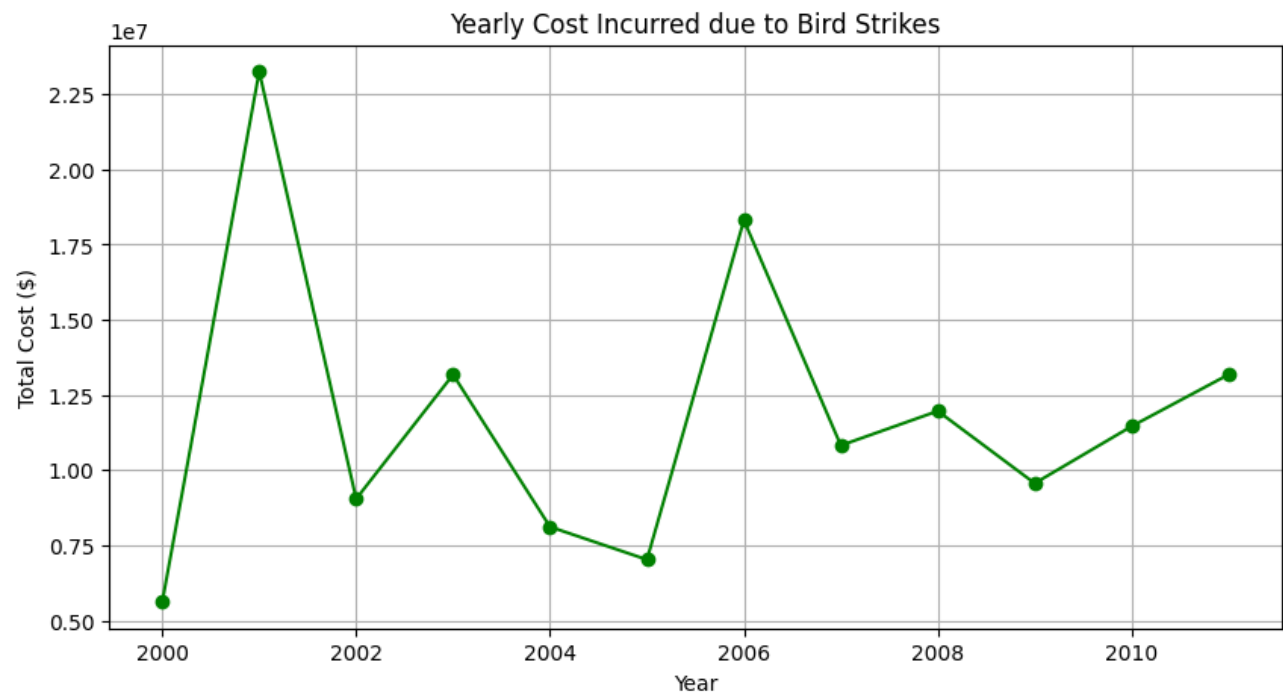
Plot 2: Top 10 US Airlines in terms of having encountered bird strikes
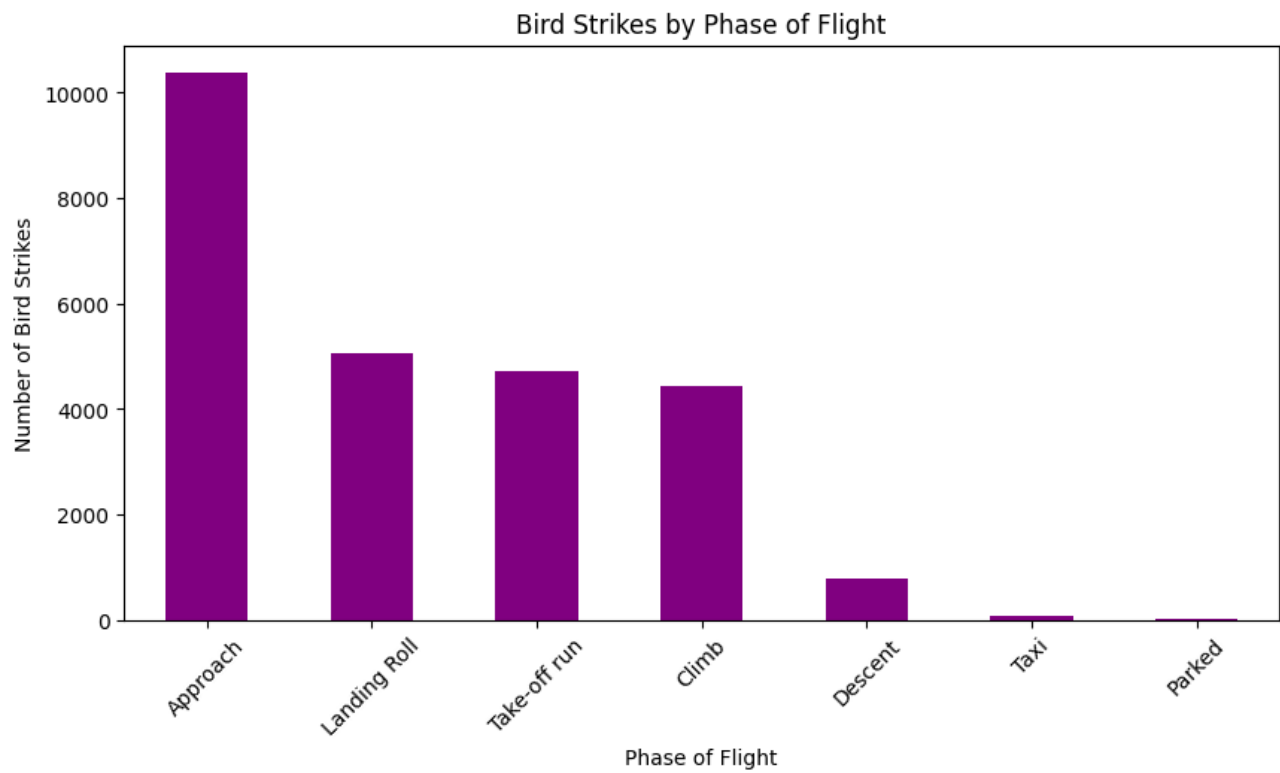

Top 10 US Airlines in Terms of Bird Strikes
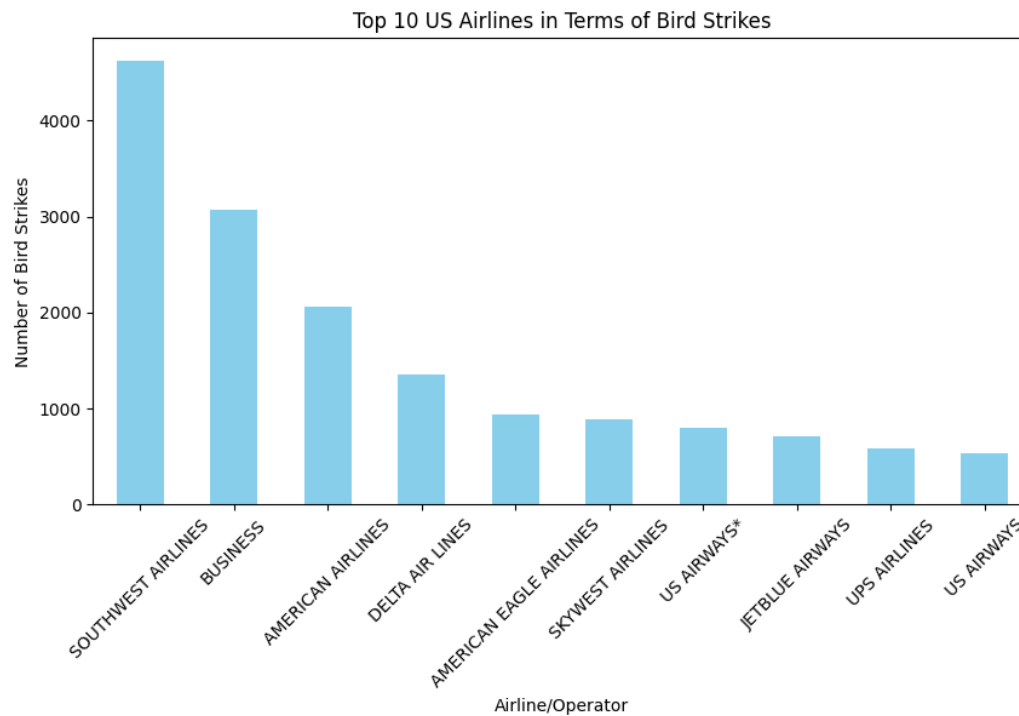
Plot 3: Airports with most incidents of bird strikes – Top 50

Top 50 Airports in Terms of Bird Strikes

Plot 4: Yearly Cost Incurred due to Bird Strikes
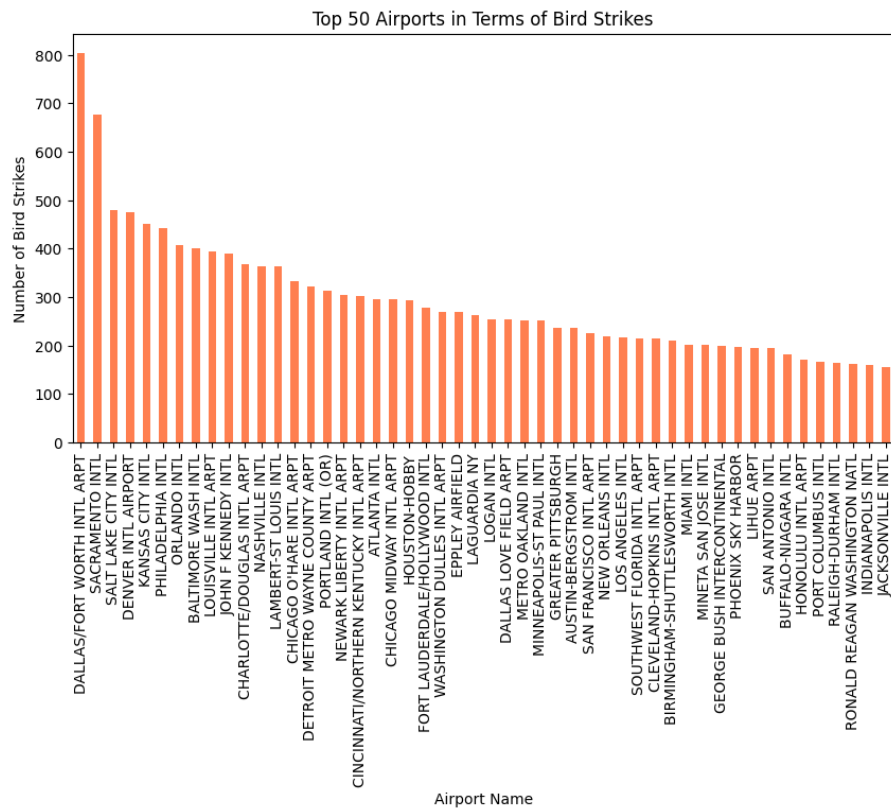
Yearly Cost Incurred due to Bird Strikes

Plot 5: Bird Strikes by phase of flight

Bird Strikes by Phase of Flight

Plot6:   Altitude of aeroplanes at the time of strike
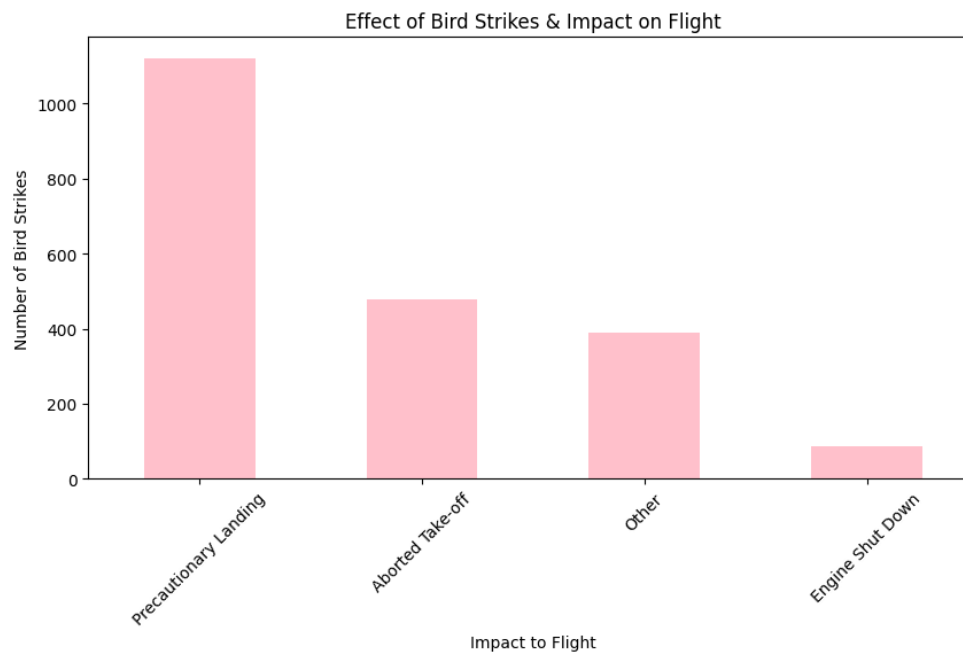
Altitude of Aeroplanes at the Time of Bird Strike
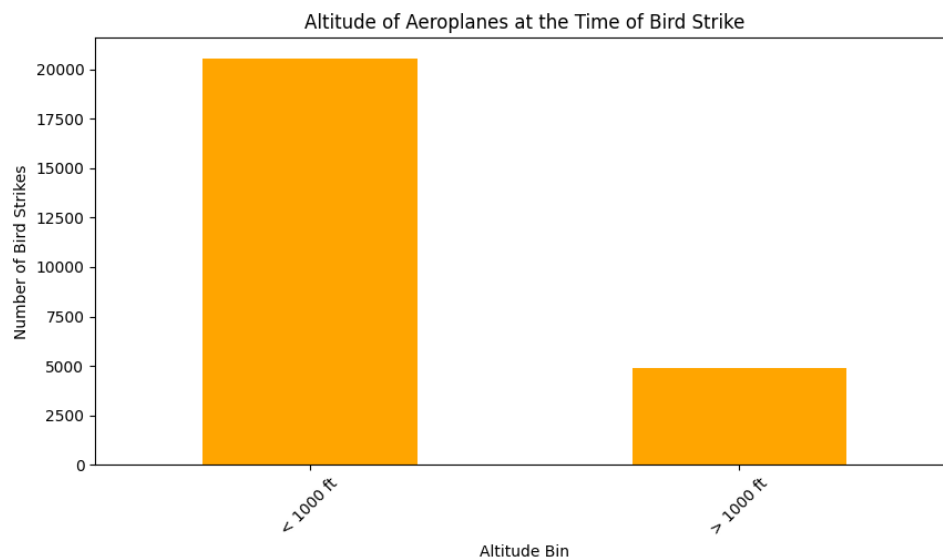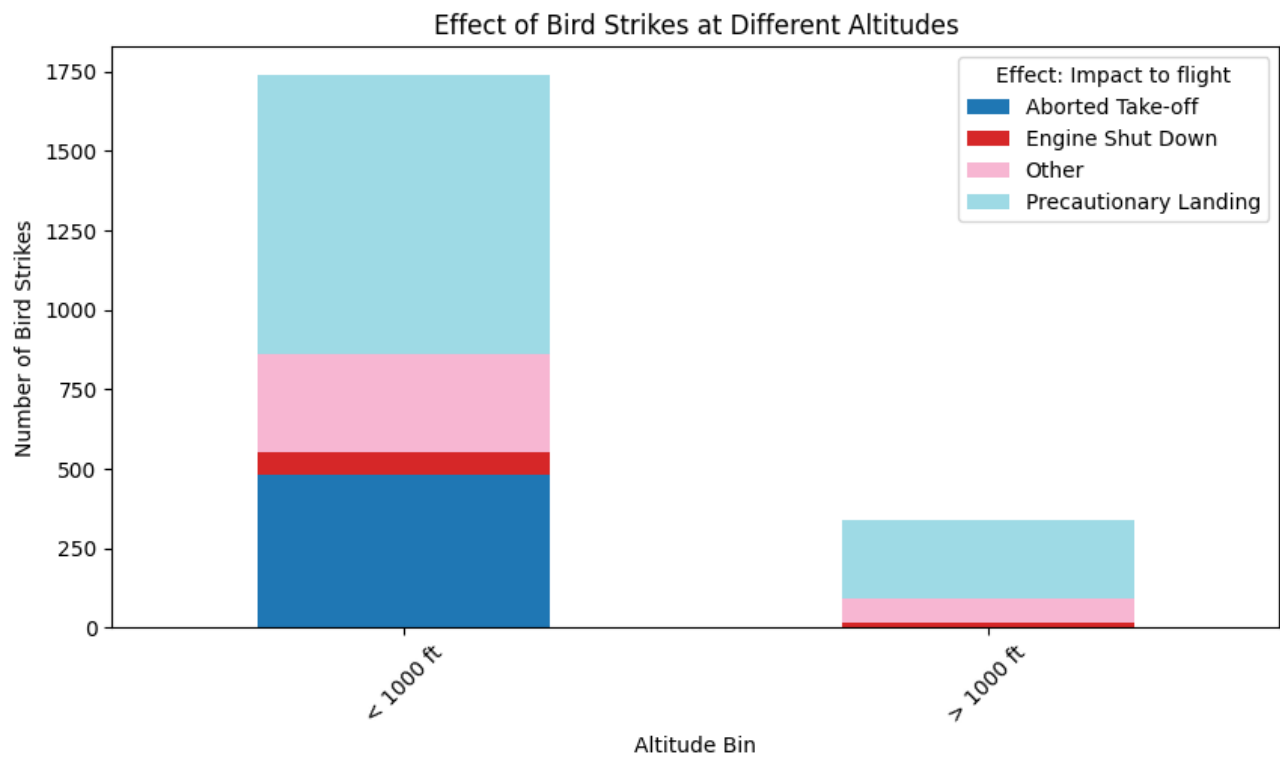
Plot 7:   Average Altitude of the aeroplanes in different phases at the time of strike
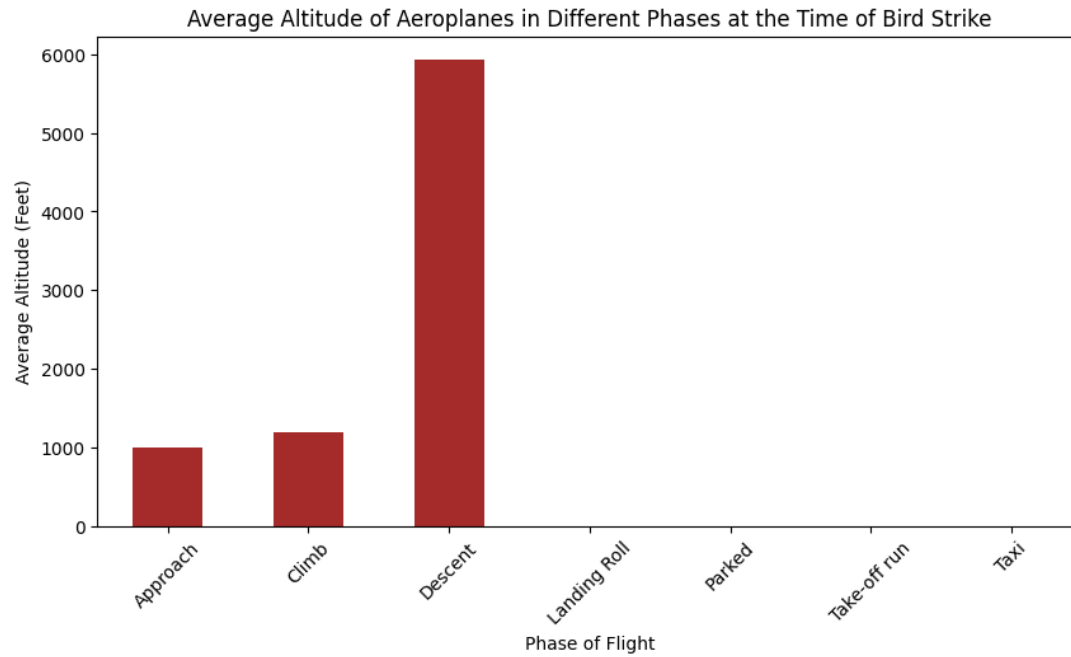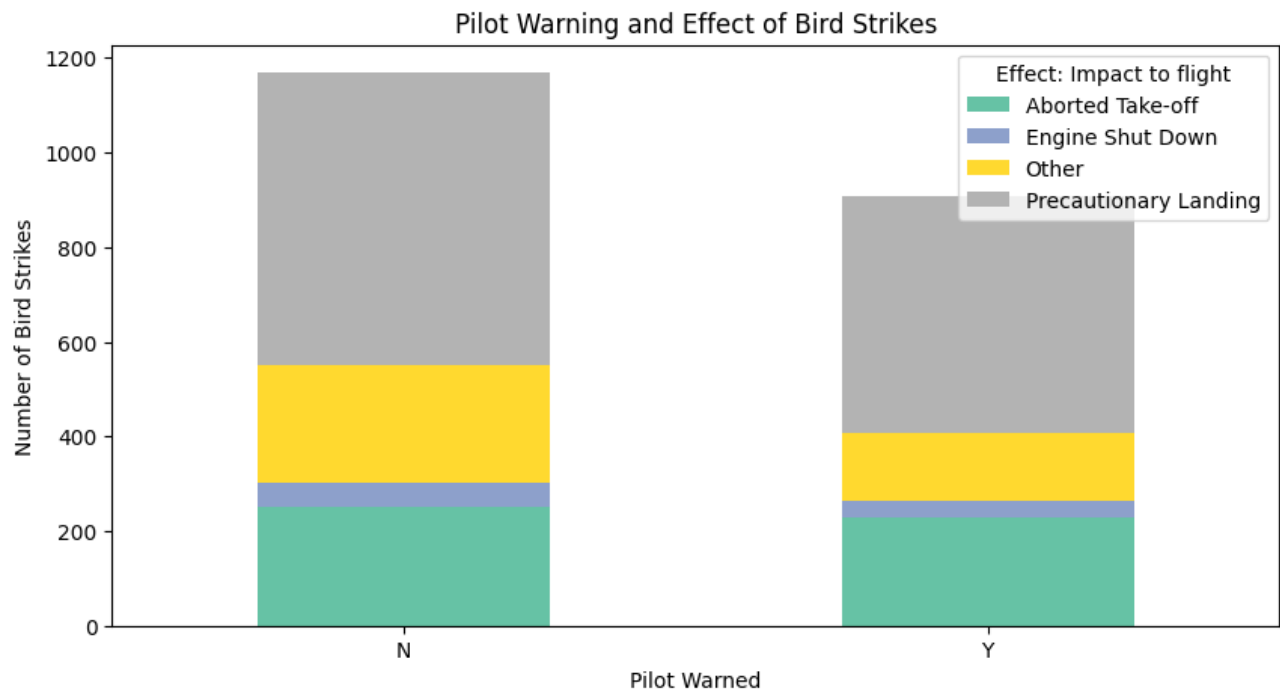

Plot 8:   Effect of Bird Strikes & Impact on Flight

Plot 9: Effect of Strike at Different Altitude

Effect of Bird Strikes at Different Altitudes

Plot 10: pilot_warning_effect

Pilot Warning and Effect of Bird Strikes

Discussion

The analysis reveals several key findings:

Increasing Trend: The number of bird strikes has generally increased over the years, possibly due to better reporting mechanisms or increased air traffic.

Seasonal Patterns: There is a clear seasonal pattern in bird strikes, with peaks during certain months, likely related to bird migration or breeding periods.

Geographical Hotspots: Certain regions show higher frequencies of bird strikes, indicating the need for localized mitigation efforts.

Species-Specific Risks: Some bird species are more frequently involved in strikes, suggesting that targeted measures for these species could be beneficial.

Severity Insights: While most bird strikes are minor, a considerable number result in significant damage, highlighting the importance of ongoing vigilance and preventive measures.

Conclusion

This study highlights the value of data visualization in understanding bird strike patterns. The insights gained can help aviation authorities and wildlife management teams develop effective strategies to reduce the risk of bird strikes. Continued monitoring and advanced analytical techniques are recommended to further enhance aviation safety.

Project 2 :AtliQ Hospitality Analysis

Data Sources

The following datasets were used:

1. dim_date.csv: Contains date-related information.

2. dim_hotels.csv: Contains details of hotels.

3. dim_rooms.csv: Contains details of rooms in the hotels.

4. fact_aggregated_bookings.csv: Aggregated booking information.

5. fact_bookings.csv: Detailed booking information.

Data Cleaning and Preparation

- Date columns in the datasets were converted to datetime objects.

- Missing values in the ratings_given column were filled with 0.

- The datasets were merged to create a comprehensive dataframe for analysis.

Analysis Performed

1. **Revenue Analysis:**

    o Calculated hotel-wise revenue and visualized it using bar plots.

    o Revenue data was pivoted to show city and hotel-wise revenue.

2. **Booking Trends by Week:**

    o Added a week_no column to represent the week of the year for each booking.

    o Grouped the data by week number to calculate the total revenue generated, the number of guests, and average ratings.

    o Plotted these trends using line plots.

3. **Occupancy Analysis:**

    o Merged booking data with date data to include day type information.

    o Calculated occupancy (number of guests) by day type and visualized it using bar plots.

4. **Booking Platform Analysis:**

    o Calculated the percentage of bookings from different platforms.

      ○   Visualized the booking platform distribution using bar plots.

Conclusions

- The analysis provides insights into which cities and hotels generate the most revenue.

- Weekly trends highlight peak booking periods and guest volumes.

- Occupancy varies by day type, suggesting certain days are more popular for check-ins.

- Booking platforms show varying popularity, indicating customer preferences in booking methods.

Recommendations

- Focus marketing efforts on high-revenue generating hotels and cities.

- Offer promotions during low occupancy days to balance guest distribution.

- Enhance the user experience on popular booking platforms to maximize bookings.

Project 3: Entertainment Analysis

1.Libraries Imported

- numpy for numerical operations

- pandas for data manipulation

- matplotlib.pyplot for plotting

- seaborn for advanced data visualization

2. Data Loading

- The script reads an Excel file named Entertainer.xlsx into a pandas DataFrame.

3. Data Cleaning

- Checks for null values in the dataset.

- Fills null values with 0 in the columns Year of First Oscar/Grammy/Emmy and Year of Death.

4. Data Transformation

- Renames several columns for better readability:

    o Year of Breakthrough/#1 Hit/Award Nomination to Year of Breakthrough

    o Breakthrough Name to Breakthrough Name

    o Year of First Oscar/Grammy/Emmy to First Oscar/Grammy/Emmy Year

    o Gender (traditional) to Gender

    o Birth Year to Birth Year

    o Year of Last Major Work (arguable) to Last Major Work Year

    o Year of Death to Year of Death

5. Data Analysis and Visualization

- **Gender Distribution**: A pie chart showing the gender distribution of entertainers.

- **Birth Year Distribution**: A histogram showing the distribution of entertainers' birth years.

- **Breakthrough Year Distribution**: A histogram showing the distribution of breakthrough years.

- **Year of First Major Award**: A line plot showing the number of major awards (Oscars/Grammys/Emmys) by year.

- **Active Years in Entertainment**: A histogram showing the active years in the entertainment industry.

- **Career Longevity**: A scatter plot showing the relationship between birth year and the last major work year.

- **Career Span**: A histogram showing the distribution of career spans (years between breakthrough and last major work year).

- **Gender Distribution by Breakthrough Year**: A stacked bar chart showing the gender distribution by breakthrough year.

Conclusion

The analysis offers a comprehensive view of the entertainers' dataset, revealing trends and patterns that can inform further research or decision-making in the entertainment industry. Visualizations help in understanding the demographics, achievements, and career paths of entertainers.