Department of Computer Science at San Francisco State University

CSC 805 Data Visualization: Concepts, Tools, Techniques, and Paradigms Visualization

Project - Phase 3
Project Implementation

Fall 2024
Submitted to Dr. Shahrukh Humayoun

Submitted by

Dhvanil Bhagat
Divya Panchal
Chetas Nikunjbhai Parekh

Unveiling Aviation Safety: A Visual and Interactive Exploration of Accident Data

Purpose:

Since aviation accidents are infrequent from a statistical perspective, the consequences of an accident call for deep understanding of the complex factors that contribute to such an incident. We aim at developing, on an interactive basis, a visualization platform outlining the usually camouflaged risks in aviation.

The project will use advanced data visualization, and other techniques to derive meaningful insights from the given Aviation Accident Database for the following key objectives:

1. How are the causes of accidents distributed according to the Broad Phase of Flight, with categories such as takeoff, cruise, and landing, by aircraft category?
2. Does the severity of the weather conditions (using a combined metric of severity) positively relate to the severity of injury sustained in the accident, by aircraft category?
3. Over the years present in the dataset, do specific aircraft models-that is, combinations of make and model-present statistically significant increases or decreases in their proportion of total accidents, indicating perhaps an emerging risk or a successful safety improvement?
4. From the 'Purpose of Flight' column, is there a clear difference in the distribution of the different categories of 'Accident Damage' across 'Personal', 'Business', and 'Scheduled' flights?
5. Considering the trends and patterns highlighted in the data, what precisely are the safety improvement focal points? Are there specific sequences of contributing factors or accident types that might be amenable to policy changes, technological changes, or pilot training programs?

Dataset Description:

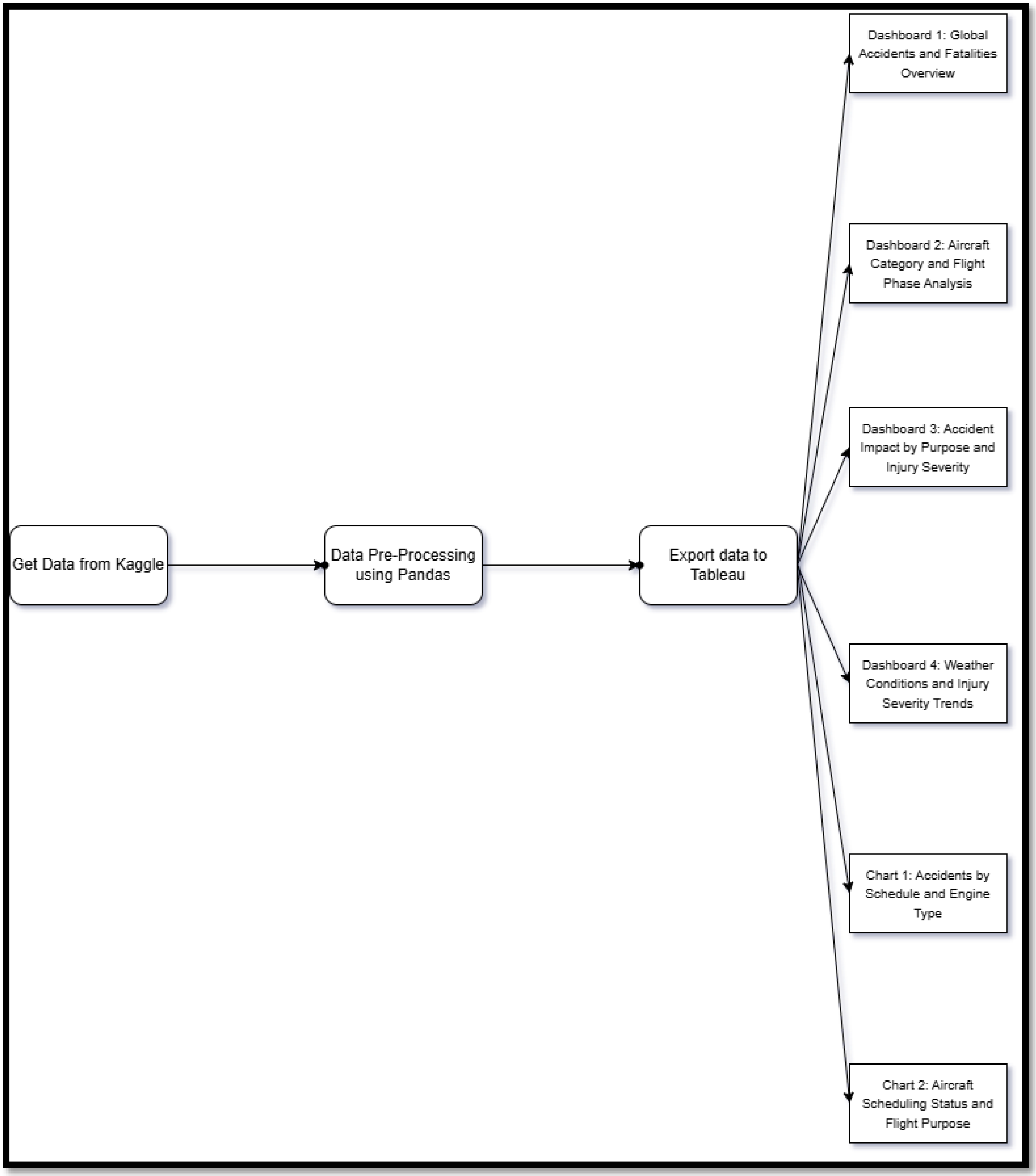
Aviation Accident Database & Synopses, up to 2023 (kaggle.com)

The dataset consists of the below details:

1. **AviationData.csv** => Detailed accident data includes a categorization of accidents by aircraft category, injury, and contributing factors, including weather and phase of flight. Data are presented on the frequency and trend in accident occurrence by specific aircraft types, injury distribution, and environmental and operational factors that contribute to aviation safety.

2. **USState_Codes.csv** => This consist of US States and their abbreviations.

System Architecture:



This architecture outlines the workflow for analyzing aircraft accident data and creating interactive dashboards. The process begins with acquiring the dataset from Kaggle, which serves as the foundational source of information. The data is then pre-processed using Pandas, a powerful Python library, to clean, transform, and structure it for analysis. This step ensures the data is accurate and ready for visualization.

Once the data was prepared, it was exported to Tableau, a leading visualization tool. From there, various dashboards and charts were developed to provide insightful analyses. These include:

- **Dashboard 1:** A global overview of accidents and fatalities, giving a high-level summary of key statistics.
- **Dashboard 2:** A detailed breakdown of aircraft categories and flight phases, exploring trends and patterns in operational contexts.
- **Dashboard 3:** Focused on the impact of accidents by purpose of flight and injury severity, revealing the most critical factors.
- **Dashboard 4:** Analyzing weather conditions and injury severity trends, shedding light on environmental factors affecting safety.
- **Chart 1:** Visualizing accidents by schedule and engine type, uncovering specific technical and operational correlations.
- **Chart 2:** Examining aircraft scheduling status and the purpose of flights, further contextualizing accident data.

Description:

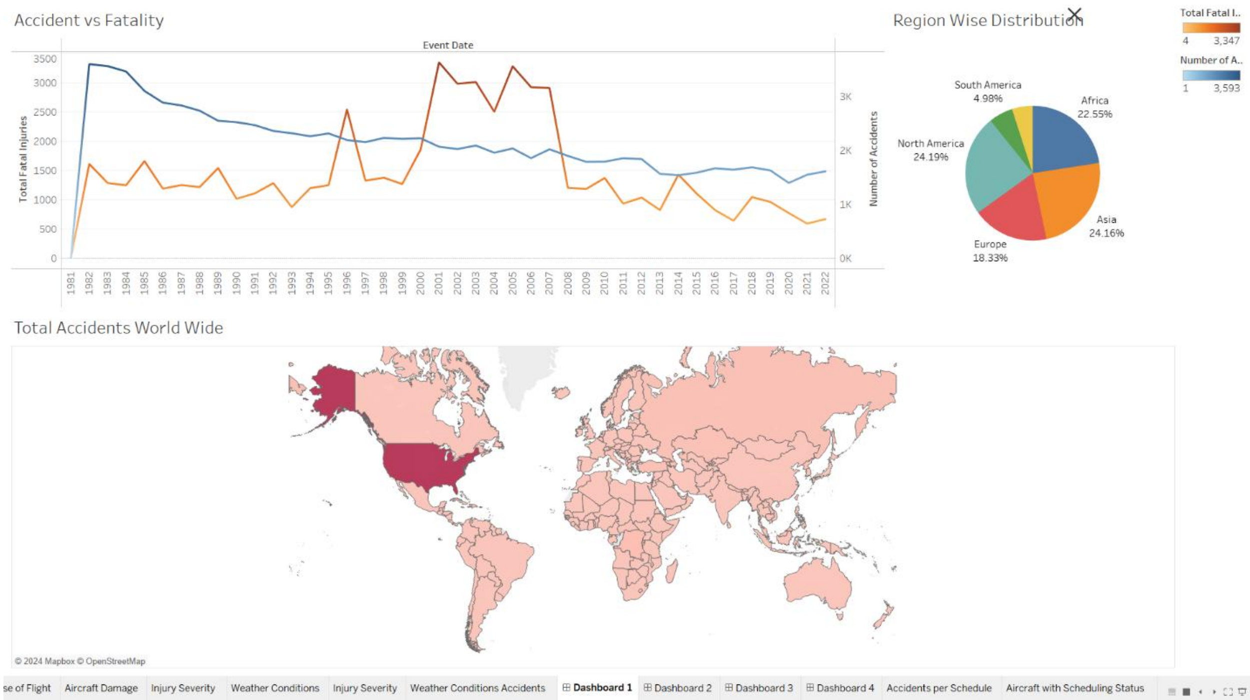
1. **Data Acquisition**:
   The first step involves collecting raw data from Kaggle, a well-known platform for sharing datasets. Kaggle provides reliable and diverse data sources, which form the foundation of the analysis. In this case, the dataset likely includes details about aviation accidents, such as dates, locations, weather conditions, aircraft types, and injury severities.

2. **Data Preprocessing with Pandas**:
   Once the raw data is collected, it undergoes cleaning and transformation using Pandas, a powerful Python library for data manipulation. This step ensures the dataset is structured and free of inconsistencies. Tasks such as removing duplicates, handling missing values, standardizing formats, and organizing columns make the data usable for further analysis. Proper preparation here is crucial to avoid errors and ensure the insights generated later are accurate.

3. **Export to Tableau**:
   After the data is cleaned and prepared, it is exported to Tableau. Tableau is an intuitive tool for creating visualizations and dashboards. At this stage, the cleaned data is ready to be used in Tableau for building charts and dashboards that provide insights into aviation accident trends, patterns, and contributing factors.

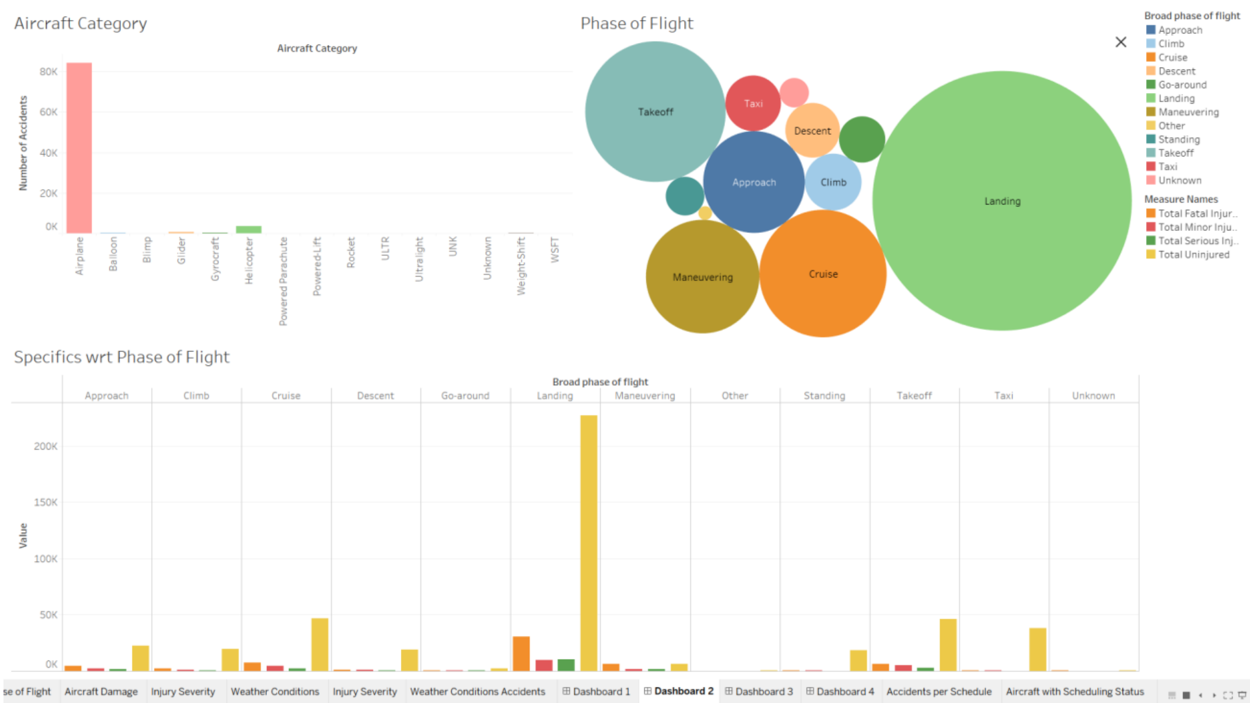
Dashboards and Descriptions

Global Accidents and Fatalities Overview:



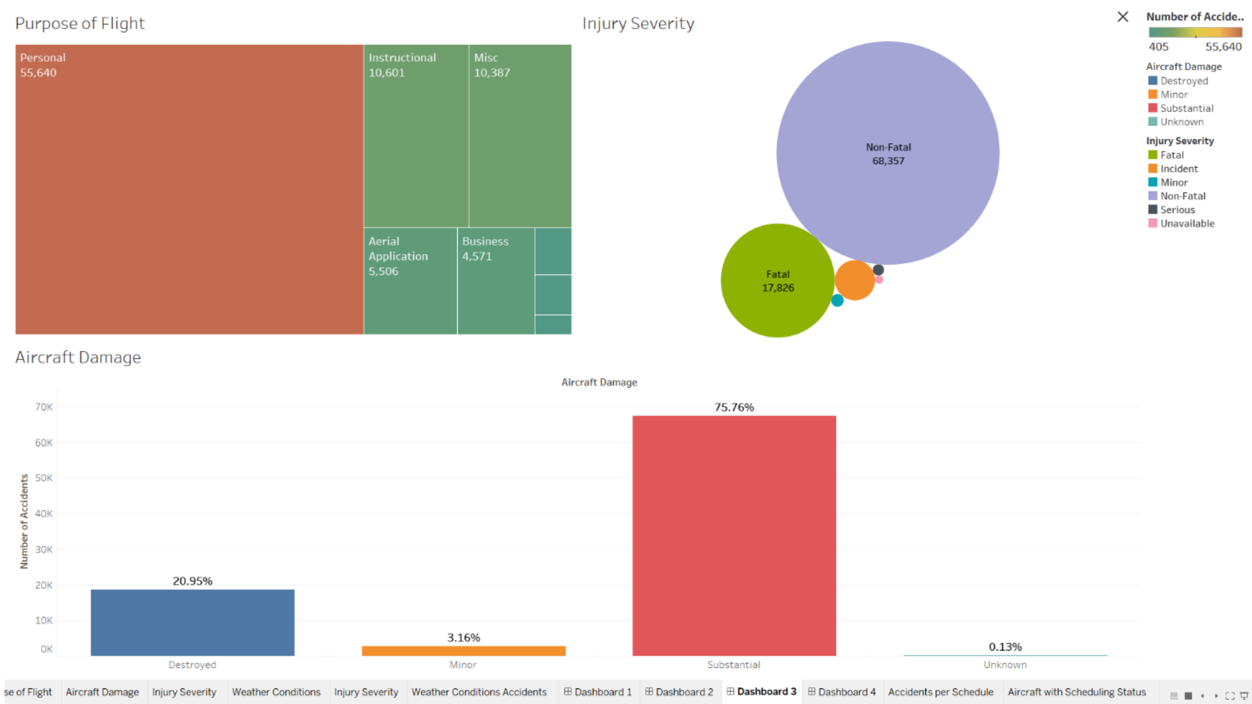
This is a Tableau dashboard, "Global Accidents and Fatalities Overview" visualizing aviation accident data across several dimensions. The top left section, "Accident vs Fatality," visualizes the trend of total accidents (blue) and total fatalities (orange) over time, from 1981 to 2022. The x-axis represents the year, while the y-axis represents the count of accidents and fatalities. This chart allows for a direct comparison of accident frequency and associated fatalities over the years. The top right visualization, "Region Wise Distribution," depicts the proportion of accidents across different regions by using a pie chart. Each segment of the pie stands for a region: South America, North America, Africa, Asia, and Europe; it corresponds to the percentage of its occurrence. The values and percentage of each are labeled with it for easy interpretation. A color legend matches each color to a region. Below these two charts, a filled map titled "Total Accidents World Wide" provides a geographical overview of aviation accidents. It is colored in shades of pink, with darker shades showing a greater concentration of accidents. Such visualization will provide the insights on the areas where higher frequencies of accidents are happening globally. In fact, the map is interactive, hovering on the countries will show the number of fatalities and accidents in that country. The deep red color of the map emphasizes the high number of accidents happening in the United States. It places the charts for better clarity of temporal and regional trends together with the geographic distribution of the accidents worldwide.

Aircraft Category and Flight Phase Analysis:



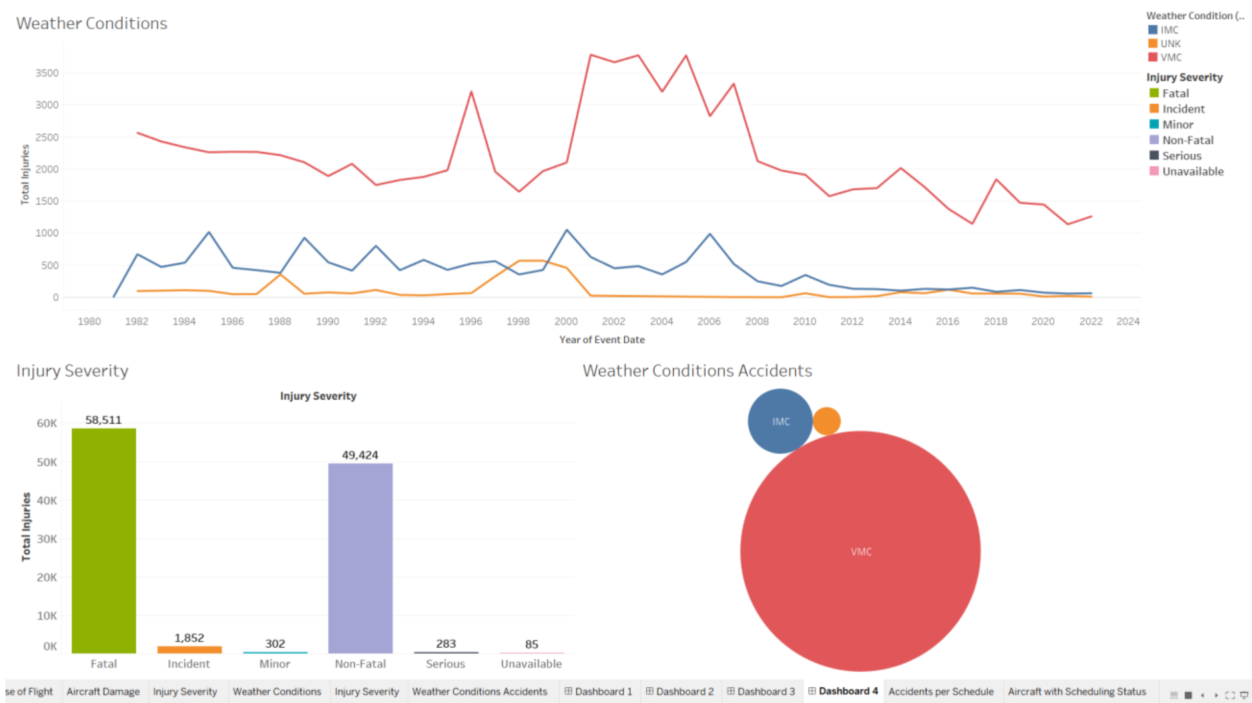
This dashboard in Tableau explores aviation accident data related to aircraft categories and phases of flight. The top left chart, “Aircraft Category,” is a bar chart showing the number of accidents in each aircraft category. The “Airplane” accidents are overwhelmingly dominant, with quite a few other categories, such as “Balloon,” “Blimp,” “Glider,” and “Gyrocraft,” having very few occurrences. A few “Helicopter” accidents round out the most prominent other categories. The top-right “Phase of Flight” visualization uses a bubble chart to show the distribution in the phases of flight for when the accidents happened. Each bubble's size indicates the total number of accidents for a particular phase. The phase that is most frequent to see accidents in is “Landing”, followed by “Cruising” and “Maneuvering”. The other phases are “Take-off,” “Descent,” “Approach,” “Taxi,” and “Climb.” Moreover, hovering on these bubbles can show detailed information about accidents and fatalities. Below these, a bar chart entitled “Specifics wrt Phase of Flight” seems to provide further breakdown of some metric (“Value”) with respect to the phases of flight, but since there are no measure names shown, we do not know what the label represents. The most prominent bars are for “Maneuvering” and “Landing”. The visualizations together on Dashboard 2 give an overview of the accidents based on aircraft type and phases of flight, with “Airplane” and “Landing” as the most critical areas for further investigation.

Accident Impact by Purpose and Injury Severity:



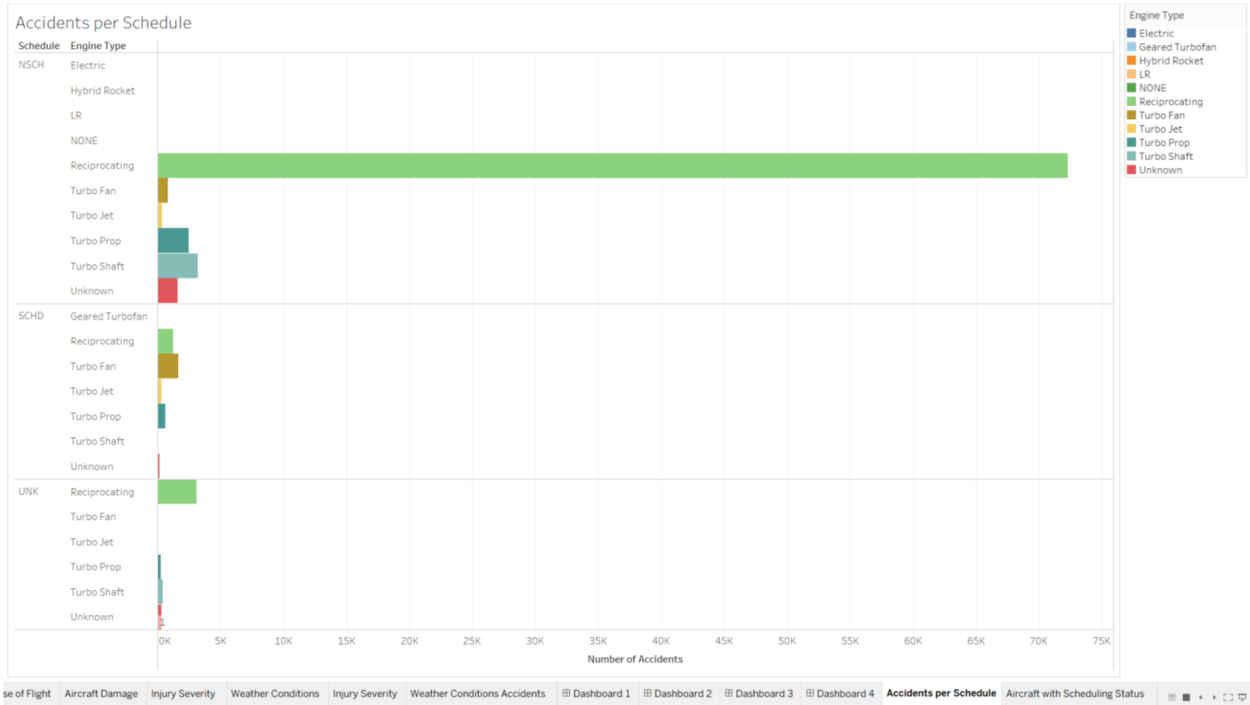
This is the Tableau dashboard 3, visualizing aviation accident data based on the purpose of flight, injury severity, and aircraft damage. The top-left chart, “Purpose of Flight,” is a treemap showing the distribution of accidents by flight purpose. “Personal” flights account for most accidents, followed by “Instructional” and “Misc.” “Aerial Application” and “Business” flights represent comparatively smaller portions. The numbers on each segment are the number of accidents. The bubble chart on the top right, “Injury Severity,” depicts the distribution of accidents by injury severity. “Non-Fatal” accidents are the most frequent and are the largest bubble, while “Fatal” accidents come in second. Incidents with “Minor,” “Serious,” “Incident,” and “Unavailable” injury information are represented by smaller bubbles. The bottom portion of the graph concerns “Aircraft Damage.” It depicts aircraft damage during accidents. The number one frequency is “Substantial.” It is very greatly different in occurrence from “Destroyed, Minor, and Unknown”. Percentage signs appear on each bar that underlines “Substantial” occurrences accounting for about 75 percent of all aviation mishaps. In a nutshell, the following dashboard effectively visualizes distributions of accidents by flight purpose, injury severity, and aircraft damage, pointing out the most outstanding categories: “Personal” flights, “Non-Fatal” injuries, and “Substantial” aircraft damages.

Weather Conditions and Injury Severity Trends:



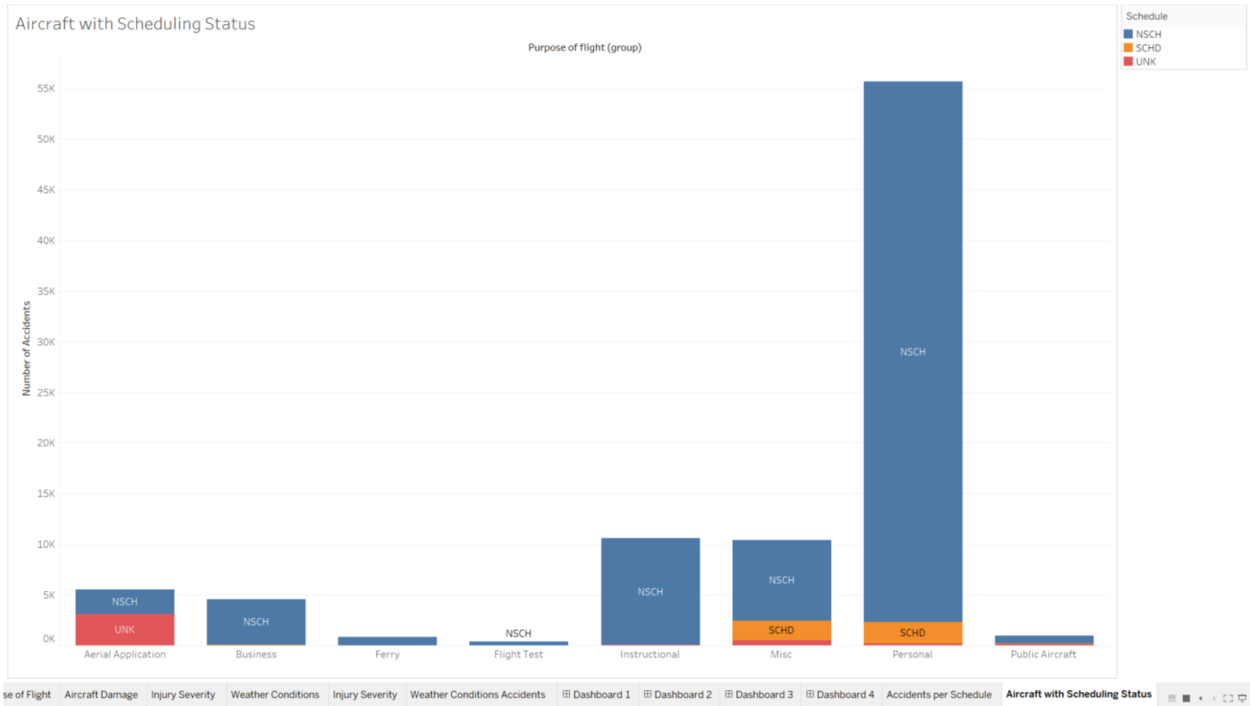
This is the Tableau dashboard analyzing aviation accident data with respect to weather conditions and injury severity. The top chart, "Weather Conditions," illustrates a dual-line graph that displays the trend in the number of total injuries, in red, against the counts of accidents contributed by different weather conditions - namely, VMC, or Visual Meteorological Conditions; IMC, or Instrument Meteorological Conditions; and UNK for Unknown - shown in shades of orange and blue over time from 1980 to 2022. The x-axis shows the year of the occurrence, while the y-axis displays the total injuries. This visualization allows the analysis of the relationship between weather conditions, total accidents, and the resulting injuries over time. The bottom-left chart, "Injury Severity," is a bar chart showing the total injuries categorized by injury severity (Fatal, Incident, Minor, Non-Fatal, Serious, Unavailable). "Fatal" injuries are the highest, followed by "Non-Fatal." This chart provides a breakdown of injury severity in weather-related accidents. The "Weather Conditions Accidents" bubble chart, bottom right, depicts the distribution of accidents due to either IMC or VMC. The size of the bubbles corresponds to the number of accidents. It depicts that the VMC accidents are much greater than the IMC accidents. This dashboard combines time-series analysis of injuries and weather-related accidents with breakdowns of injury severity and weather condition involvement, offering a comprehensive view of how weather impacts aviation accident outcomes.

Accidents by Schedule and Engine Type:



The following visualization in Tableau uses a horizontal bar chart to explore the distribution of aviation accidents by type of engine and flight schedule. Segmentation was done for three groups of NSCH (Non-Scheduled), SCHD (Scheduled), and UNK (Unknown). Within each category, bar colors represent engine types: “Reciprocating,” “Turbo Fan,” “Turbo Jet,” “Turbo Prop,” “Turbo Shaft,” “Unknown,” “Electric,” “Geared Turbofan,” “Hybrid Rocket,” “LR,” and “NONE.” Each bar's length corresponds to the number of accidents, enabling direct comparison of accident frequency across engine types and schedules. The NSCH category contains the largest number of accidents, overwhelmingly dominated by reciprocating engines. In the SCHD category, the share of accidents involving turbo fan engines is much smaller than for the reciprocating engines. The UNK category shows a small number of accidents, with the largest numbers once again in reciprocating engines. The plots reveals a strong association between non-scheduled flights and reciprocating engines regarding aviation accidents.

Aircraft Scheduling Status and Flight Purpose:



This is a Tableau visualization that uses a stacked bar chart to analyze the distribution of aviation accidents by purpose of flight and aircraft scheduling status-whether the aircraft was operating on a scheduled (SCHD), non-scheduled (NSCH), or unknown (UNK) basis. The x-axis is the categories of purpose of flight - Aerial Application, Business, Ferry, Flight Test, Instructional, Misc, Personal, Public Aircraft; and the y-axis is accidents. Each bar is partitioned by scheduling status, via color. The "Personal" category of flight purpose is the most frequent, and overwhelmingly NSCH. "Instructional" and "Misc" purposes also have a high number of accidents, mainly NSCH as well. "Aerial Application" and "Business" flights have a smaller number of accidents, with representation of both NSCH and UNK statuses. This visualization allows for comparison across different flight purposes, broken down by scheduling status. Given such dominance of NSCH flights especially within the "Personal" category, this is potentially an area for further investigation regarding aviation safety.

Demo Video

https://drive.google.com/file/d/1rghjGGmurFohOwY-Tr5FdLozjmNrBgGl/view?usp=sharing

Tableau Public Link

https://public.tableau.com/views/AviationCrashAnalysis_17329992895540/GlobalAccidentsandFatalitiesOverview?:language=en-US&:sid=&:redirect=auth&:display_count=n&:origin=viz_share_link