**About:**

We were tasked with created a custom and creative D3.js project with the purpose of showing travels and airlines where accidents occur most, and what time of the year is perhaps safer than others. Viewing these trends can provide valuable information that could help with safety precautions for both travelers and airlines. This project tested all of our capabilities and skills and pushed us all to think outside the box. We found our dataset on [Kaggle](https://www.kaggle.com/prathamsharma123/aviation-accidents-and-incidents-ntsb-faa-waas) and downloaded the CSV’s to our local drives to begin the initial code and analysis. We had one GitHub repository with several branches that we organized our different parts of analysis into. We began with attempting to merge 4 separate airline datasets; Airline\_Accidents, Faa\_Incidents\_Data, NTSB\_Aviation\_Data, and World\_Aircraft\_Accidents. However, we realized not all of the datasets shared a common attribute to merge on so we decided to keep them separate and use the dataset that was appropriate for each analysis.

Each one of us took on a separate responsibility; Creating a SQL Database through PgAdmin, creating a Flask App to host the database and connecting it to Heroku, a calendar visualization, a map visualization, a world tour visualization, and data comparisons. Below is a breakdown of each element, the challenges we came across, and the conclusions we realized through our analysis.

**Creating a SQL Database through PgAdmin: Jenny/Andrew**

Creating the SQL Database with Flask through PgAdmin and Heroku was a team effort that had many challenges. One of the initial challenges that Jenny came across was cleaning the data which needed to be done in order to import into sql schemas successfully. Through this process Jenny learned a lot about how best to clean csv files and identify what needs to be cleaned. Unfortunately, we ended up not being able to utilize the joined tables in the sql database because our dataset was too big to host in Heroku.

Andrew and Tanlin worked on the Flask app which also presented many challenges due to the limitations that Heroku has on large datasets. Because the Heroku app was limiting our data set size through SQL Database, we all felt it would be detrimental to the analysis if we cut down our dataset by 97% just to use the Heroku app. Instead, we decided to host the files through Github directly and allow for the main index to host the website we created. So even though it won't be in the final version, Andrew learned a lot about procfiles, requirement files and building web apps.

**Calendar visualization: Andrew baker**

Andrew spent many hours working on the calendar visualization and working through many challenges along the way. Andrew felt that the calendar visualization was the perfect way to represent the large amount of data and allow users to interact with the data at the same time. The calendar shows two view, one for FAA flights and the other for NTSB flights. The calendar has a gradation that represents the amount of accidents, the darker the color the more accidents that occurred. As you scroll through the calendar you can see the color scale getting less dark, which allows us to visually see that the amount of airline accidents decreases over time. You can also see that the center of the calendar is darker than the edges, showing us that more accidents occur during the middle months of the year.

**Map visualization: Romi Sinnigen**

For the map visualization, Romi used the data set “ Airline\_Accidents” which had over 150,000 rows of data. Romi felt that a map with points would be a great way to visualize where in the world accidents were occurring, and provide some information about each of those accidents. The first attempt at the marker map it became clear that 100,000 points on the map was overwhelming, so she converted the points to a cluster map and implemented several map layers for satellite, light, and dark. The map is interactive, so the further you zoom in or click on a cluster, the map will adjust and further expand the markers. Each marker has a popup with information on the Airline code and the location in which the accident occurred. Romi’s conclusion from this visualization is that a majority of accidents happen in the USA. However, this could be because America has some of the largest airports in the world, as well as a large majority of smaller domestic airports.

**World tour visualization: Tanlin Hung**

The world tour visualization was created by Tanlin and works in junction with Andrew’s calendar map. Tanlin originally wanted to use cities to point to places on the map but a large portion of the data was missing values for cities. Instead she used country of origin for the map. The world tour is initiated when a date on the calendar is selected. The globe will spin and color the country of origin red where the accident occurred. Because a majority of the accidents occurred in the United States, the globe hovers over the United States for many of the calendar days chosen.

**Creating website: All**

When we were all done with our visualizations we worked together on creating the website using a bootstrap template. We had to make a lot of modifications with the multiple index files, app files and JS files. Andrew had the most experience with HTML so he lead the group with reconfiguring the template html to fit our visualizations and vision. All said and done, we are very happy with the story our data tells.