CS411 Project Proposal
Rushika Kumarswamy, Chen Wang, Enya Sun, Sriya Mikkilineni

Flight Status Visualizing Map

Project Summary

Flight Delays and Cancellations are not uncommon. There may be patterns in these occurrences, and being able to visualize them in a simple way can help people determine what airlines to avoid. Our project will record data from DOT's 2015 airline statistics and plot them on an interactive map displaying major airports (As long as the data continues to be updated in the same format, our application should be able to model these changes so we are not limited to displaying data from 2015).

Color coding will be used to indicate the frequency of delayed or canceled flights for destination and origin airports, as well as airlines. Some other features will include the formation of tables that display the number of flights delayed per origin and departure airport, a running leaderboard ranking airlines with the fewest number of delays and cancellations to the ones with the most. Graphical representations of the data will also be presented along with the map and tables, which will predict future flight delays (based on either location or airport).

Data

The data we will be using consist of flight delays and cancellations from 2015 published by the DOT's Bureau of Transportation Statistics. We will extract IATA_code, airport, city, state, departure delay, and dates from the three CSV files of flight information. We plan on utilizing geographical data (airport, city, state, etc) to aid in the construction of our map, and other statistics (such as departure time, weather delays, cancelations, etc) to create our own tables, rankings, and predictions.

Basic Functions

The purpose of our website is to visualize flight delay and cancelation data in a comprehensive and user-friendly way. The main functions will consist of a map with all the airport locations marked in color-coordinated symbols indicating details like delayed, on-time, diverted, and canceled flights per location. Some simple functions include having multiple summarizing tables and displaying delay and cancelation frequencies for respective airlines and locations, allowing users to make better decisions when booking flights for the future. One of the more complicated functions is the prediction feature, which will allow users to input airlines or flight paths (origin and destination) to view a program-generated estimate of whether delays will occur.

Creative Component

Since we have a substantial amount of data to work with, we run into the possibility of our map becoming overcrowded. A way to combat this while adding a unique feature is to implement a toggling option, specifically, a hover display function where when the user hovers their mouse over a dot and more detailed information will be presented from the columns of the flights.csv. If possible, we would also like to make our map interactive, meaning that users can zoom in to see view specific locations in higher detail.

Description

      We can analyze which airport always makes the departure late. Then maybe we can suggest air companies to modify their departure times based on the table.

Usefulness

      Statistic-based applications exist, but they usually do not provide enough information about flights. An example includes the Bureau of Transportation Statistics' own website (https://www.transtats.bts.gov/ontime/Cancellation.aspx), which states that it provides detailed cancelation statistics, but as no visualization features are present, the interface becomes unappealing and inefficient to use. Instead, we plan to use aggregation to detail note how many hours the flight delays. Flightview by OAG also exists (https://www.flightview.com/traveltools/), which allows users to live track flights in real-time, but it provides little to know historical data. Our application will incorporate similar elements as Flightview's to create a working map, but we will attempt to implement prediction analysis as well.
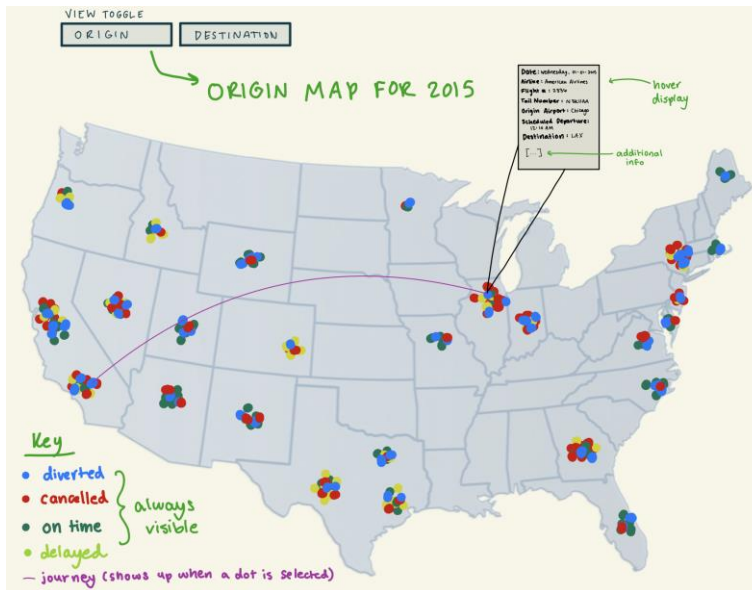
Realness

      For this project, we will be using the Kaggle dataset "2015 Flight Delays and Cancellations (https://www.kaggle.com/datasets/usdot/flight-delays?select=flights.csv). Here is a description of the dataset: "The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights are published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations." Some of the names of the relevant columns we will be using for this project are airline, origin, destination, year, month, day of the week, and departure delay.

Functionality

- Users will be able to add, delete, and edit any raw data (We are using flight data from 2015, but as long as raw data is altered in the same format/columns, then our website should accurately reflect that. This allows our application to be generalized beyond solely 2015 data)
- Users will be able to use keyword searching to easily access airport and airline rankings in both the map and tables

Low-fidelity UI mockup
Website map:

Website tables (dummy numbers for demonstration):

Project work distribution

Backend

The backend portion of our project is expected to be divided into 4 main subcategories. The first would be in filtering flights based on their origin or destination airport by directly selecting and analyzing these columns of the raw data. Sriya will be primarily responsible for this task. The second task is to analyze flight delays and cancellations per airline requiring airline-specific filtering of the data. Rushika will be primarily responsible for this task. The third task would be to generate statistics and new columns of data based on flights delayed per time of year. Enya will be primarily responsible for this task. The fourth task would be to analyze the data per week to understand the trends of cancellations/delays per day of the week. Chen will be primarily responsible for this task.

Frontend

Sriya and Enya will be responsible for the UI which includes designing the map visualization, color coding, and hover information display components. Rushika and Chen will be primarily responsible for working on integrating frontend and backend APIs to use derived information and connect it to what information to display. Possible addition (if extra time) of airlines ranking leaderboard and statistics relating to the amount of time delayed, which everyone will work on together.

(Additional Breakdown):
All visualizations will be per origin or destination respectively
- Rushika: (Airline Specific) Create a ranking of all airlines in terms of delays
- Chen: (Time of the year) Generate statistics and new columns of data based on flights delayed per time of year
- Sriya: (day of the week) Generate statistics to display flights delayed based on the day of the week
- Enya: Taking in user input and displaying relevant information
    - Ex: origin/destination
    - Airlines
    - Time of year

*Data cleaning/compression note:

Since we will not be using entire dataset, we plan to randomly select roughly 850 rows from each month to ensure accurate representation of compressed data (we are mainly focusing on the frequency of delays per airline instead of per month). Everyone will work together to write a function for this task.

---

1. Describe what data is stored in the database. (Where is the data from, and what attributes and information would be stored?)

1. 2015 Flight delays and cancellations (https://www.kaggle.com/datasets/usdot/flight-delays)
2. It includes IATA_code, airport, city, state, flight information

2. What are the basic functions of your web application? (What can users of this website do? Which simple and complex features are there?)
    Ideas:
    - data visualization somehow?
    - Color coordinating with dots on a map with details like delayed, on-time, diverted, and canceled flights per location
    - Predicting delays (https://www.kaggle.com/code/fabiendaniel/predicting-flight-delays-tutorial)

3. What would be a good creative component (function) that can improve the functionality of your application? (What is something cool that you want to include? How are you planning to achieve it?)
    - https://www.kaggle.com/datasets/usdot/flight-delays?select=flights.csv
    - A hover display function where when the user hovers their mouse over a dot, more detailed information will be presented from the columns of the flights.csv

4. Project Title: Flight delays and cancellation Visualized

5. Project Summary: It should be a 1-2 paragraph description of what your project is.

    Flight Delays and Cancellations are not uncommon. There may be patterns in these occurrences, and being able to visualize them in a simple way can help people determine what airlines to avoid. Our project will record data from DOT's 2015 airline statistics and plot them on an interactive map displaying major airports (As long as the data continues to be updated in the same format, our application should be able to model these changes so we are not limited to displaying data from 2015).

    Color coding will be used to indicate the frequency of delayed or canceled flights for destination and origin airports, as well as airlines. Some other features will include the formation of tables that display the number of flights delayed per origin and departure airport, a running leaderboard ranking airlines with the fewest number of delays and cancellations to the ones with the most. Graphical representations of the data will also be presented along with the map and tables, which will predict future flight delays (based on either location or airport).

6. **Description** of an application of your choice. State as clearly as possible what you want to do. What problem do you want to solve, etc.?
   We can analyze which airport always makes the departure late. Then maybe we can suggest air companies to modify their departure times based on the table.

7. **Usefulness**. Explain as clearly as possible why your chosen application is useful. Make sure to answer the following questions: Are there any similar websites/applications out there? If so, what are they, and how is yours different?
   https://www.transtats.bts.gov/ontime/Cancellation.aspx
   However, this website did not provide too much information about the flight. So, in our application, we would use aggregation to detail note how many hours the flight delays.

8. **Realness**. Describe what your data is and where you will get it.
   For this project, we will be using the Kaggle dataset "2015 Flight Delays and Cancellations. Here is a description of the dataset: "The U.S. Department of Transportation's (DOT) Bureau of Transportation Statistics tracks the on-time performance of domestic flights operated by large air carriers. Summary information on the number of on-time, delayed, canceled, and diverted flights is published in DOT's monthly Air Travel Consumer Report and in this dataset of 2015 flight delays and cancellations." Some of the names of the relevant columns we will be using for this project are: airline, origin, destination, year, month, day of week, departure delay.

9. Description of the **functionality** that your website offers. This is where you talk about what the website delivers. Talk about how a user would interact with the application (i.e., things that one could create, delete, update, or search for). Read the requirements for stages 4 and 5 to see what other functionalities you want to provide to the users. You should include:
   1. **A low-fidelity UI mockup**: What do you imagine your final application's interface might look like? A PowerPoint slide or a pencil sketch on a piece of paper works!

   2. **Project work distribution**: Who would be responsible for each of the tasks or subtasks?
      List of the person responsible for which exact functionalities in section 6. Explain how backend systems will be distributed across members. Be as specific as possible as this could be part of the final peer evaluation metrics.

The backend portion of our project is expected to be divided into 4 main subcategories. The first would be in filtering flights based on their origin or destination airport by directly selecting and analyzing these columns of the raw data. Sriya will be primarily responsible for this task. The second task is to analyze flight delays and cancellations per airline requiring airline specific filtering of the data. Rushika will be primarily responsible for this task. The third task would be to generate statistics and new columns of data based on flights delayed per time of year. Enya will be primarily responsible for this task. The fourth task would be to analyze the data per week to understand the trends of cancellations/delays per day of week. Chen will be primarily responsible for this task.

Sriya and Enya will be responsible for the UI which includes designing the map visualization, color coding and hover information display components. Rushika and Chen will be primarily responsible for working on integrating frontend and backend APIs to use derived information and connect it to what information to display. Possible addition (if extra time) of airlines ranking leaderboard and statistics relating to amount of time delayed, everyone works on it.

- Backend
- Note: will not be using entire dataset, will divide months equally and randomly sample rows for each month to ensure accurate representation of compressed data
    - Selecting relevant columns of data to use in raw form ex: origin and destination columns and determine total number of flights
        - Flight delayed per origin airport
        - Flights delayed per destination airport
    - Creating new tables containing information to generate secondary data such as statistics on ex: what percentage of alaska airlines flights departing from Seattle are delayed
    - Graphing/predicting future delays ?
    - All visualizations will be per origin or destination respectively
        - (Airline Specific) Create ranking of all airlines in terms of delays
        - (Time of the year) Generate statistics and new columns of data based on flights delayed per time of year
        - (day of the week) Generate statistics to display flights delayed based on the day of the week
        - Taking in user input and displaying relevant information
            - Ex: origin/destination
            - Airlines

- Time of year

  - UI (Sriya and Enya)
    - Color coded
    - Info pop up during hover
- Integrating Backend and UI Components (Chen and Rushika)
-

Rubric:
1. Updates in **TeamInfo.md** for teamName
2. Updates in **TeamInfo.md** for project summary
3. Creation of the project proposal placed in the **doc** folder.
4. The project proposal includes the title and project summary
5. The project proposal includes a detailed description of your application
6. The project proposal includes a detailed usefulness description of your application:
7. The project proposal includes a detailed realness description of your application
8. The project proposal includes a detailed functionality description of your application (2%)
9. The project proposal includes a detailed low-fidelity UI mockup of your application
10. The project proposal includes a detailed project work distribution across the team
11. Create a release with the correct tag for your submission and submit it on canvas