Flight Delay Prediction

Sky High Predictions Grounded in Data.

How bad is the problem?

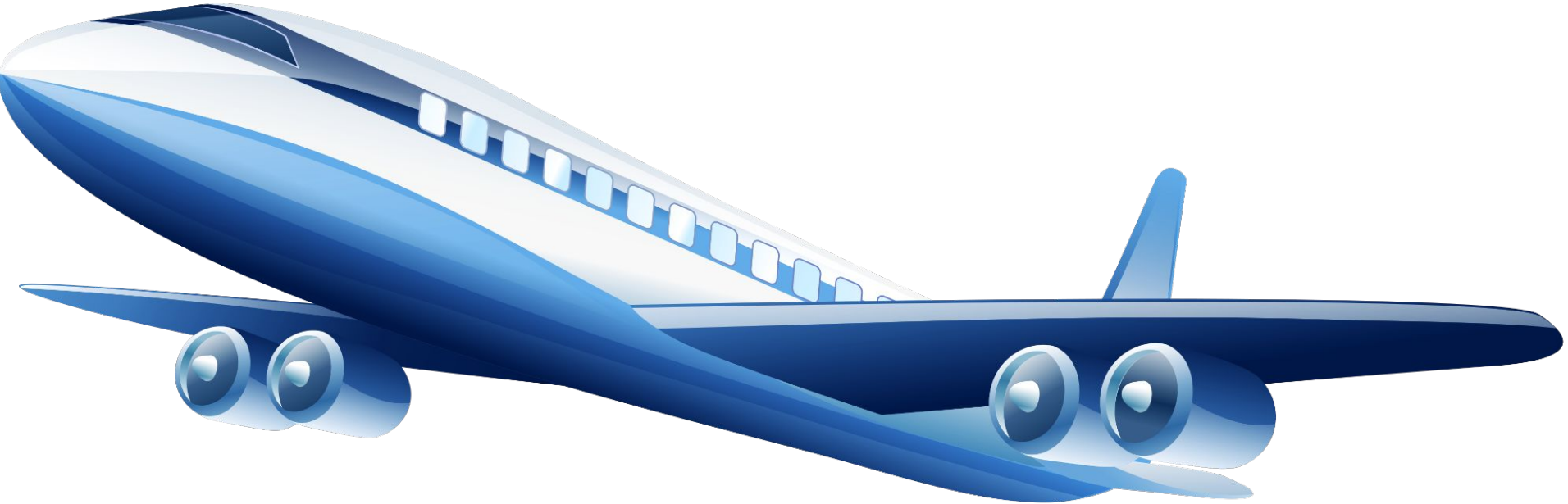
- According to the Bureau of Transportation, around 20% of flights are delayed
- The average length of delay in 2023 has been 53 minutes
- In 2017 there were ~ 740,000,000 passengers in ~8,000,000 flights
- On average there are 91 passengers/flight.

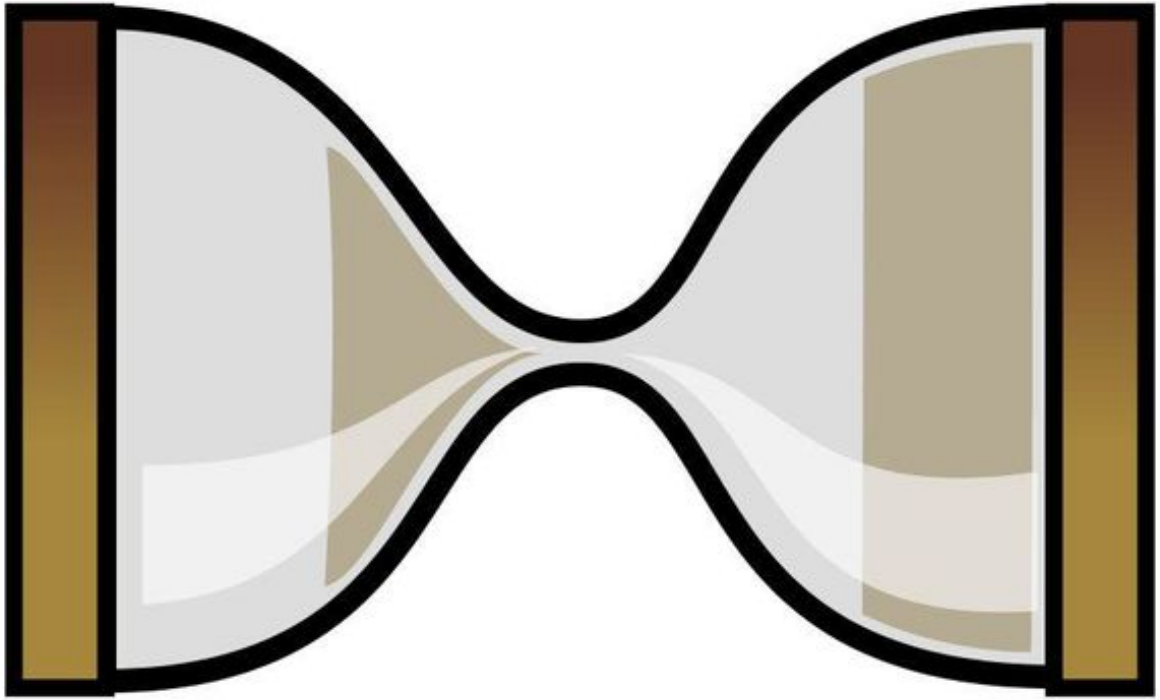
FLIGHT DELAY

PREDICTION

Sky High Predictions

Grounded in Data





How bad is the problem?

- According to the Bureau of Transportation, around 20% of flights are delayed
- The average length of delay in 2023 has been 53 minutes
- In 2017 there were ~ 740,000,000 passengers in ~8,000,000 flights
- On average there are 91 passengers/flight.

3 MILLION WORK HOURS LOST. EVERY. YEAR.

How bad is the problem?

- According to the Bureau of Transportation, around 20% of flights are delayed
- The average length of delay in 2023 has been 53 minutes
- In 2017 there were ~ 740,000,000 passengers in ~8,000,000 flights
- On average there are 91 passengers/flight.

3 MILLION WORK HOURS LOST. EVERY. YEAR.

That’s 342 work years per year

Overview:

Allow users to make more informed flight purchases to decrease the amount of delays they experiences.

Platform

Build out the core Chrome extension for extracting, processing, and displaying flight data

Analysis

Implement several machine learning analysis models on flight datasets

Data Visualization

Work on design aspects of the Chrome extension and data visualizations of model outputs

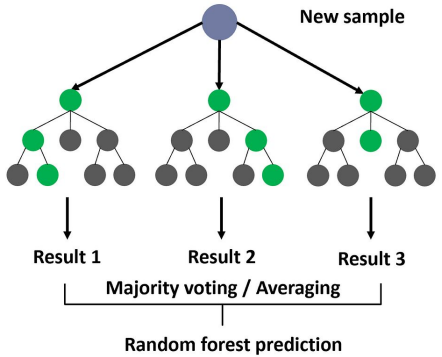
Our Team

Project Leads: Ari Nair & Gabe Ragy

Analysis:	Platform:	Data Viz:
Andrew Romitti	Pranav Eega	Jinseok (Jason) Hwang
Parth Parikh	Heeba Merchant	Kate Jeong
Sarvesh Tiku	Ian Wilson	Riyan Patel
Aayush Patel	Varenya Amagowni	
Samiya Pathak	Joie Yeung	
Adithya Rajesh		

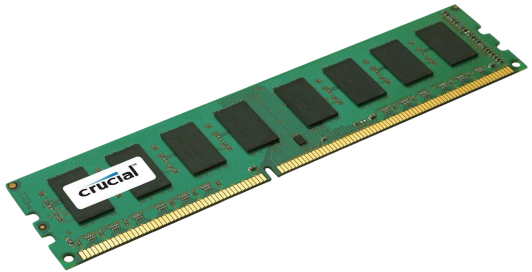
How did we approach the problem?

- **Data:** List of all flight data from 2023 to 2024. Includes flight schedule and delay time.
- **Model:** Random Forest
- **Inputs:** Distance, Origin Coordinates, Destination Coordinates, Day of the Week, Airline, Various Weather Factors (Cloud cover, temp, etc..)
- **Output:** Expected Delay (On Time, 0-30 Minutes, 30-60 Minutes, 60+ Minutes)



Issues Encountered

- **RAM Management:** working with a very large dataset (60m+ rows)
- **Weather Data Collection:** Excessive API calls for gathering weather data (100,000+)
- **Skewed Data:** The data we had was primarily skewed to one bin, this presented a big challenge that once solved improved our model significantly



Improvements Made

Fall 2024

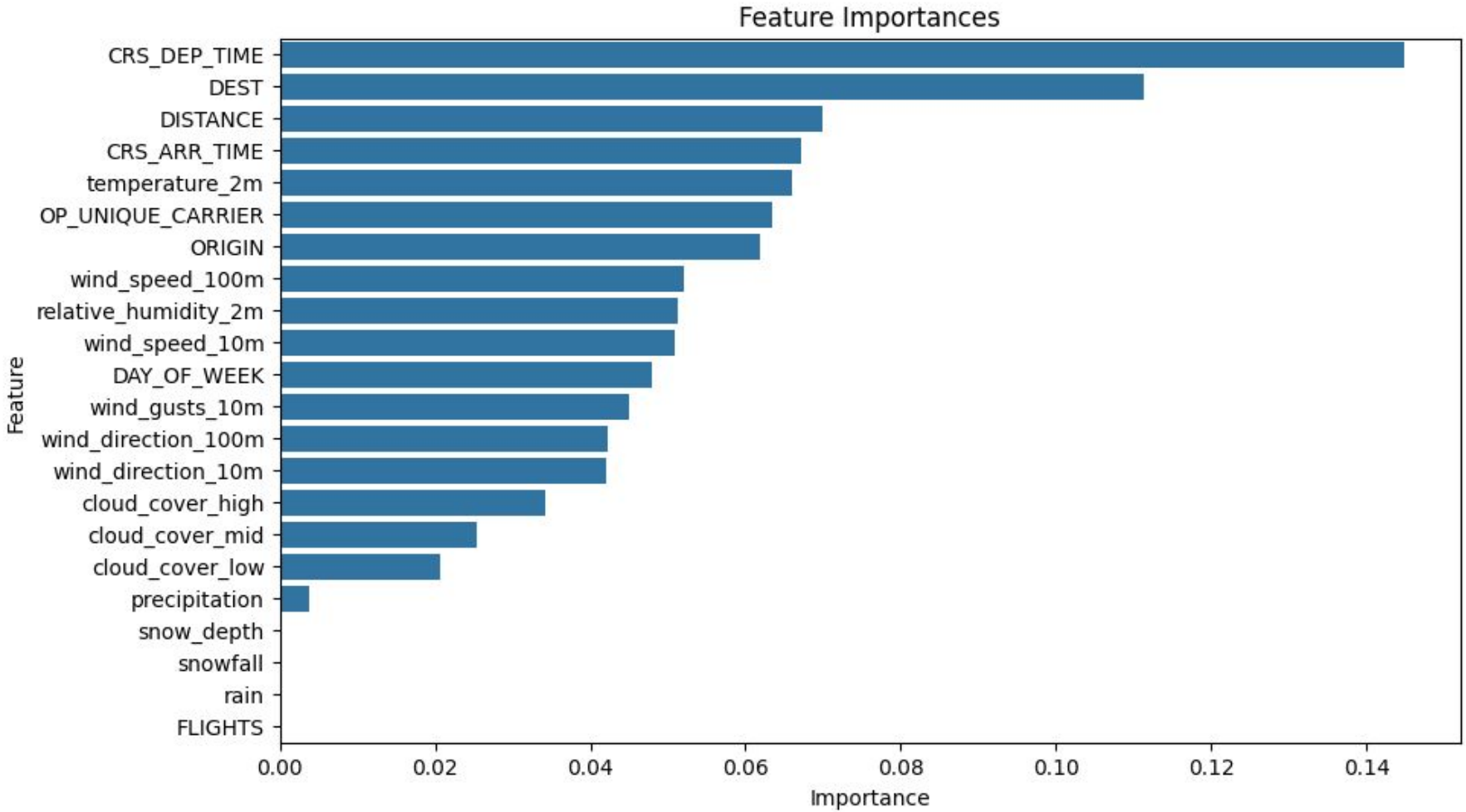
- Updated Dataset to 2023-2024 data
  - Improved training time & accuracy
- Gathered weather data for all departing flights
- Utilized Random Forest, XGBoost, and MLP model
- Incorporated Explainability
  - Most important feature & probability bins

Spring 2025

- Explored new columns
  - Arrival weather & time between flights
- Replaced one hot encoding with label encoding

Classification Model

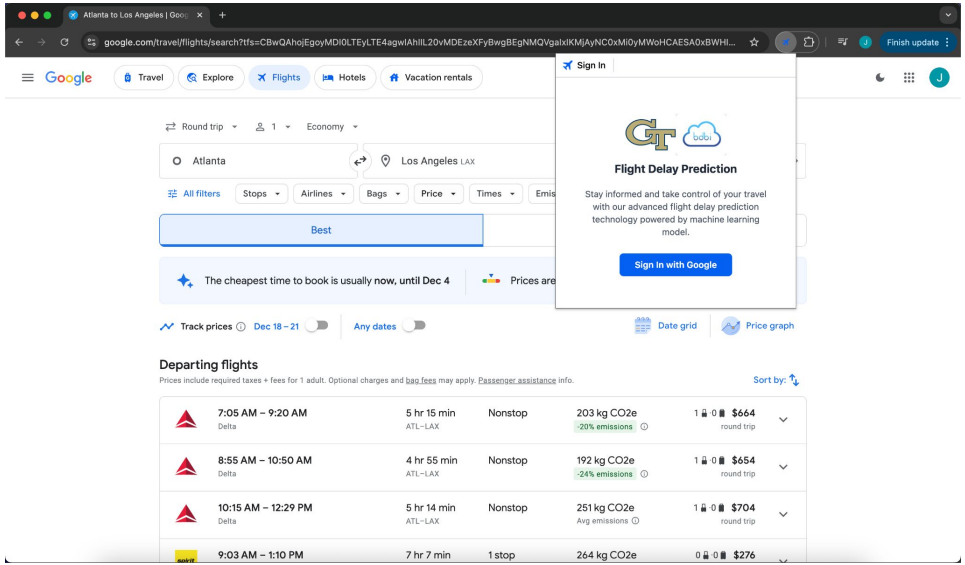
- Bins flight delays into categories based on the delay: On Time, 0-30 minutes, 30-60 minutes, 60+ minutes.
- Random Forest trained on 50 estimators
- Added metric of Binary Accuracy
- **Accuracy: 70% ← 6% more accurate than last semester in predicting delays**


Feature Importances
CRS_DEP_TIME
DEST
DISTANCE
CRS_ARR_TIME
temperature_2m
OP_UNIQUE_CARRIER
ORIGIN
wind_speed_100m
relative_humidity_2m
wind_speed_10m
DAY_OF_WEEK
wind_gusts_10m
wind_direction_100m
wind_direction_10m
cloud_cover_high
cloud_cover_mid
cloud_cover_low
precipitation
snow_depth
snowfall
rain
FLIGHTS
Feature
0.00
0.02
0.04
0.06
0.08
0.10
0.12
0.14
Importance

Future of Analysis

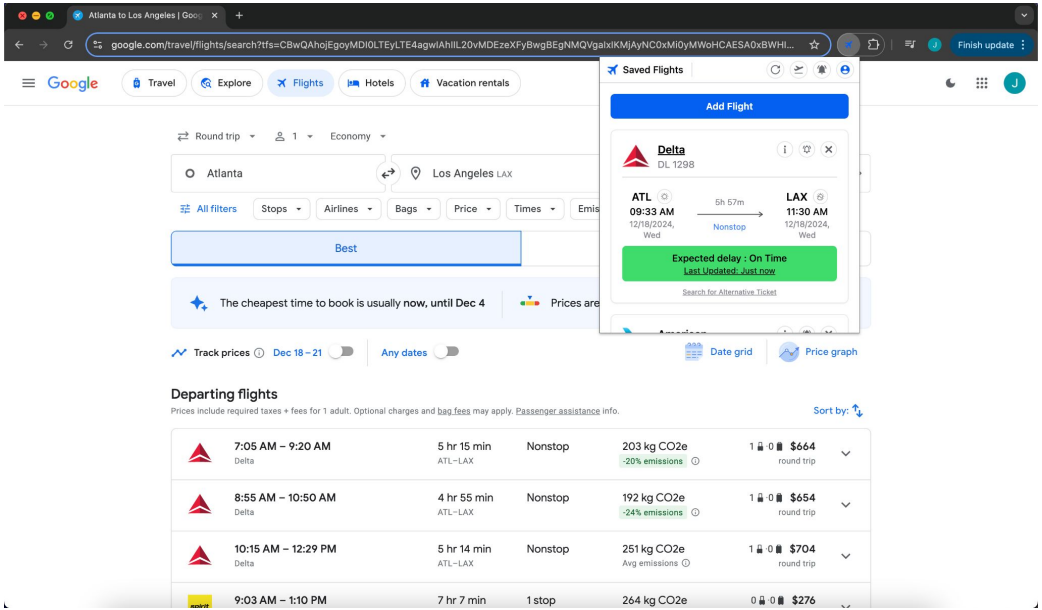
- **Try Stronger Models:** We tried MLP, LSTM, and Random Forest, XGBoost, and a Voting Classifier with decision tree & gradient boosting and they all resulted in the same accuracy, but stronger models could produce better results, such as an RNN
- **Improve Data Analysis:** We have many weather columns; keeping only the most important ones could help improve accuracy
- **Host Model on Vertex AI:** Create an endpoint on Vertex AI to integrate with the platform and streamline the prediction process

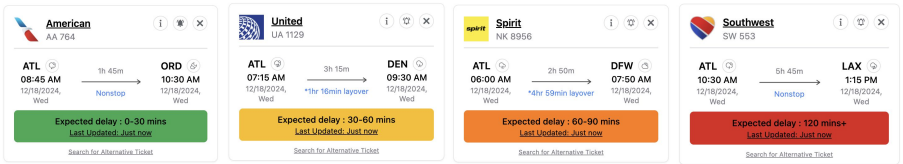
Frontend - Tech Stack & Sign In



- **Frontend Tech Stack:**
  - Built with **React** and **TailwindCSS**
  - Integrated **Google OAuth** for secure sign-in
- **Design:**
  - Google-like interface that goes well with the Google Flights
- **Features:**
  - **Sign-in** page with Google authentication
  - **Saved Flights** page for delay prediction
  - **Add Flights** page for manual flight addition
  - **Notification** page for delay notification
  - **Profile** page for user profile management

Frontend - Saved Flights

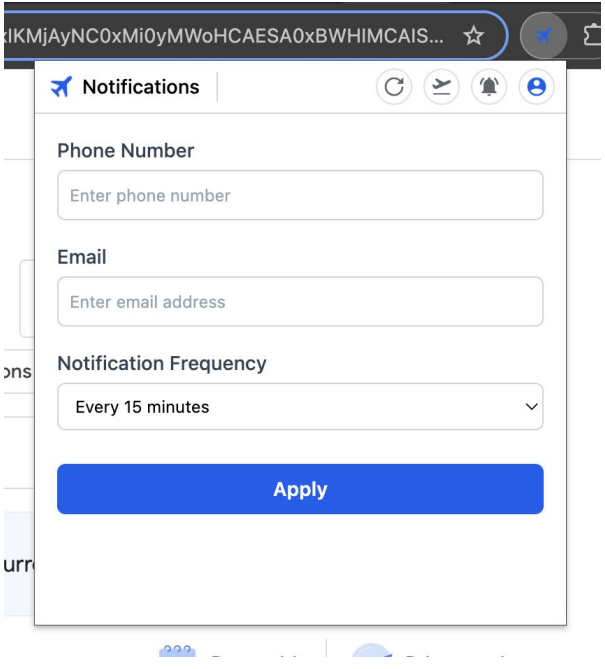




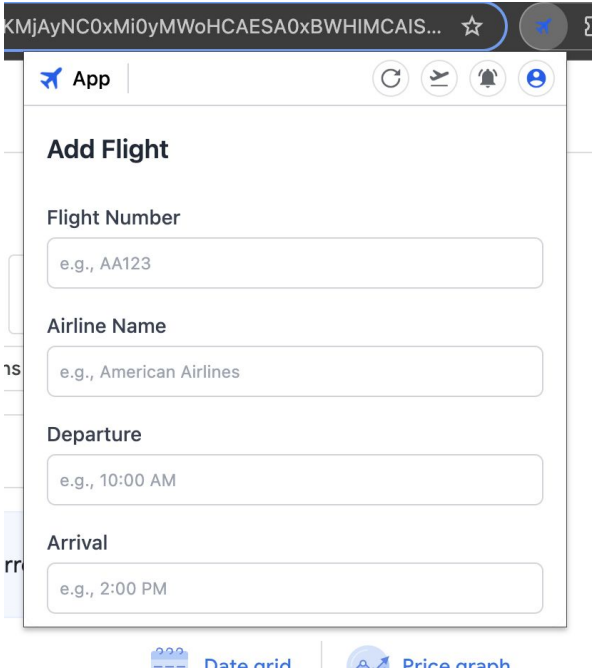
Saved Flights Page

- Flight tickets that user bought are added to the saved flights page for continuous delay prediction.
- User can also manually add the tickets by clicking ‘Add Flight’ button
- Users can turn on/off notification for each ticket by clicking the notification button
- Flight prediction section has different colors depend on delay times

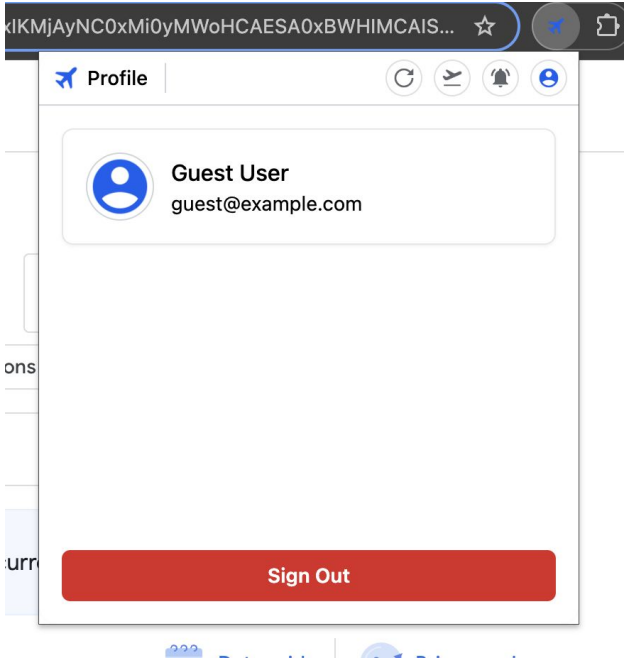
Frontend - Add Flight & Notification & Profile



Notification Page



Add Flight Page



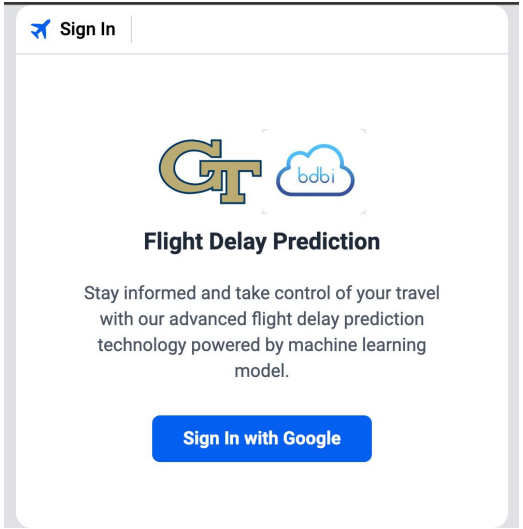
Profile Page

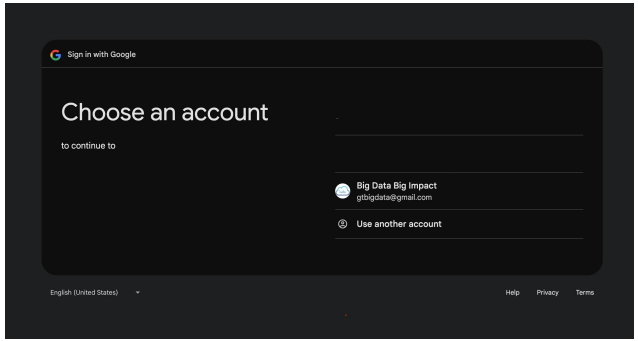
Future of Data-Viz

- **Polish Frontend Design:** Refine the interface to be more visually appealing, user-centric, and intuitive for all users.
- **Dark Mode Feature:** Add dark mode toggle feature to enhance user experience and accessibility.
- **Integration with Calendars:** Allow users to sync flight schedules with their Google or Outlook calendars.
- **Data Insights:** Add charts or graphs to visualize flight trends, such as delay patterns by airline and route.

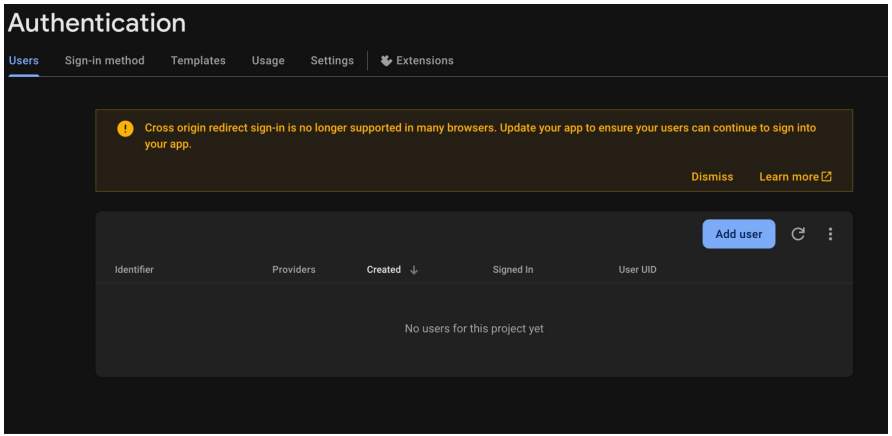
Platform Google Oauth

- Uses Firebase & Firestore to hold users emails
- The user will be prompted to login with their saved gmail accounts
- When we refresh Firebase, their account will be saved.

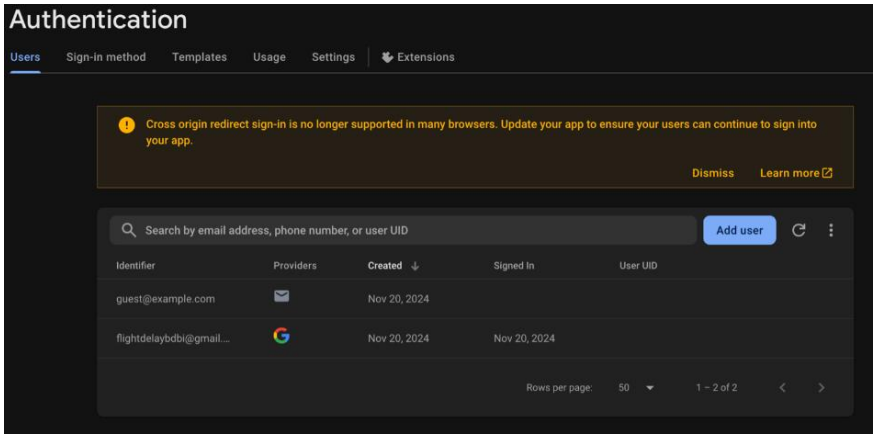




Firebase & Firestore

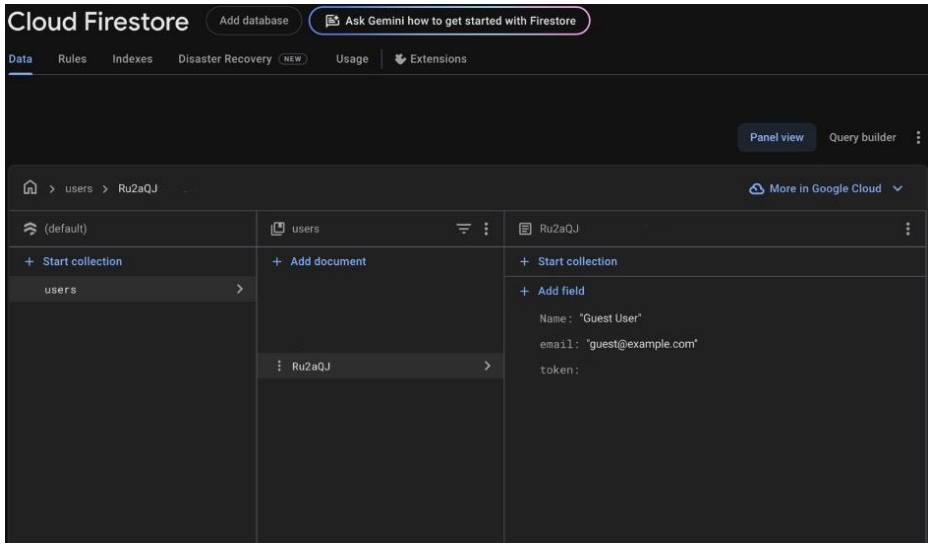


Before User Sign in



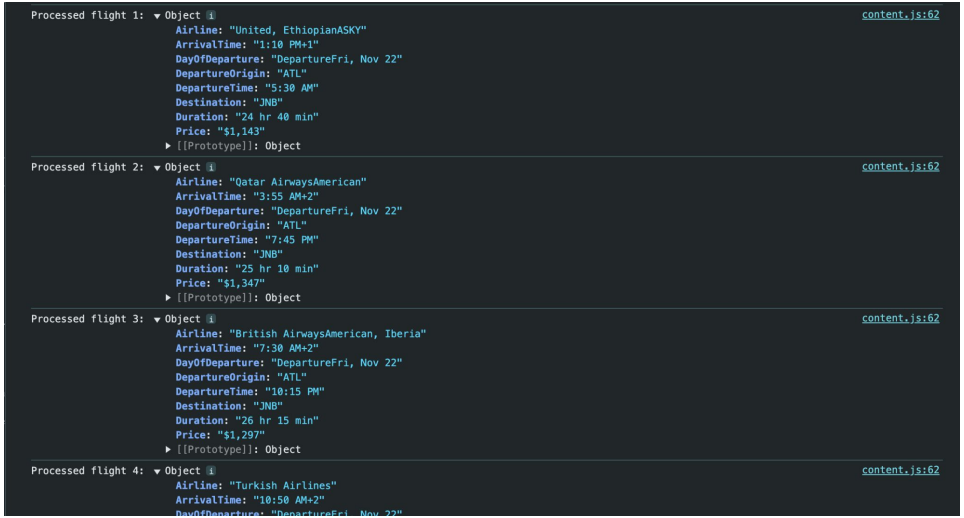
After User Sign in

Firestore



- Uses the User UID for each email that signed in
- Creates new database for each user
  - This is how we store flight information for each user

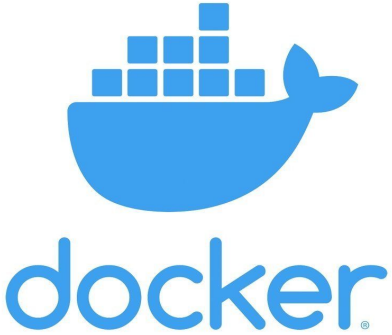
Google Flights Web Scraping



- Uses Data from Google Flights
  - Scrapes all the data and stores it in an array
- We use the data to confirm the flight exists when the user requests to add flight

Flight Delay Prediction API

- Flask-based REST API that serves our trained machine learning model to predict flight delays
- Containerized using Docker and deployed via Google Cloud Run
- Predicts flight delays using our pre-trained ML model



Future of Platform

- **Email Update:** Send email alerts whenever the model detects any delays with the flight that they added.
- **Connecting Model with Google Scraping:** Make sure the model will be able to take in all the data from Google Scraping and make predictions in real time.
- **Bug Fixes:** Fix any issues with the code itself so there are no problems during publication

Coming Spring 2025...

- First half
  - Model improvements
  - Connecting our model to the frontend Chrome Extension to display predictions
  - Email alerts
- Second half
  - Publish, publish publish!