Celebrity Flight Pattern

Group 5: Jared Willson, Katie Beale, Megan Palma, Jim Haugen, Anusha Balasubramanian, Mohin Grewal

Project Overview

Target Audience / Value Proposition:

Our target audience includes anyone interested in comparing the travel habits of celebrities to those of the average person.

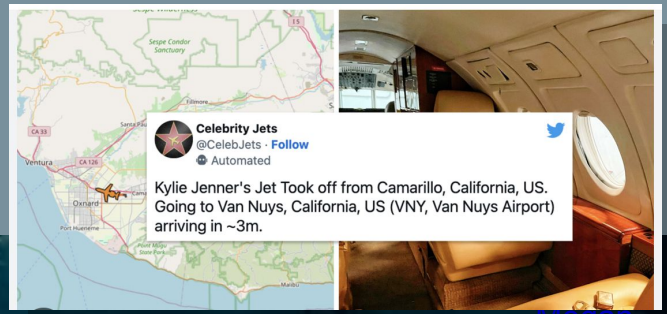
Additionally, our data can be the beginning steps for investigating how frequent celebrity travel might impact the areas they visit most often ie air quality changes and other environmental effects.

Motivations:

Our motivation was to gain an understanding of celebrity flight trends compared to the average person.

Tools Used:

BeautifulSoup, Plotly, Splinter, Flask, greatcircle.js, Leaflet



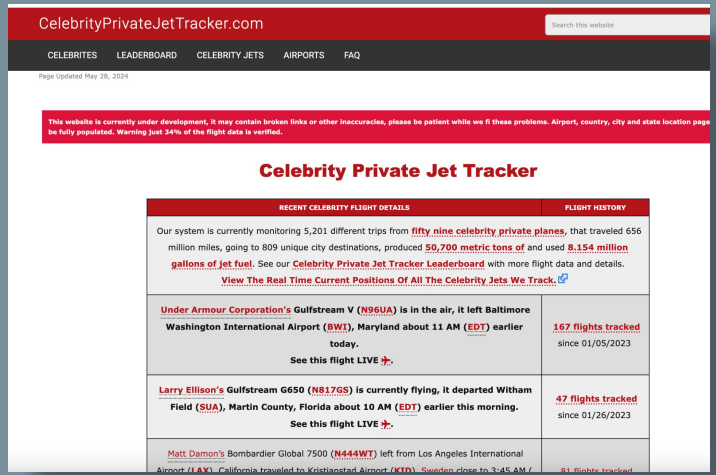
Megan

Data Overview

Data Source:
https://celebrityprivatejettracker.com/

- Website run by minimal resources (one person)
- Most data is based solely on unverified tail numbers
- Each plane has its own page, but not all pages have information
- Data is continually updated, but we grabbed a static snapshot

Site Homepage



Katie

Data Scraping / Processing Pipeline

- Goal: fetch all flight data for all celebrity planes
- Process:
  - Use the Leaderboard page as a starting block
  - Use Leaderboard table to grab links and tail numbers
  - Start two files: to hold links and tail numbers, and tail numbers and flight info
  - Looking for:
    - Flight date
    - Departure and arrival airports
    - Flight distance in miles
    - Flight time
    - Flight fuel usage in gallons
  - Write to flight info to CSV, adding corresponding tail number to each row of flight info

Katie

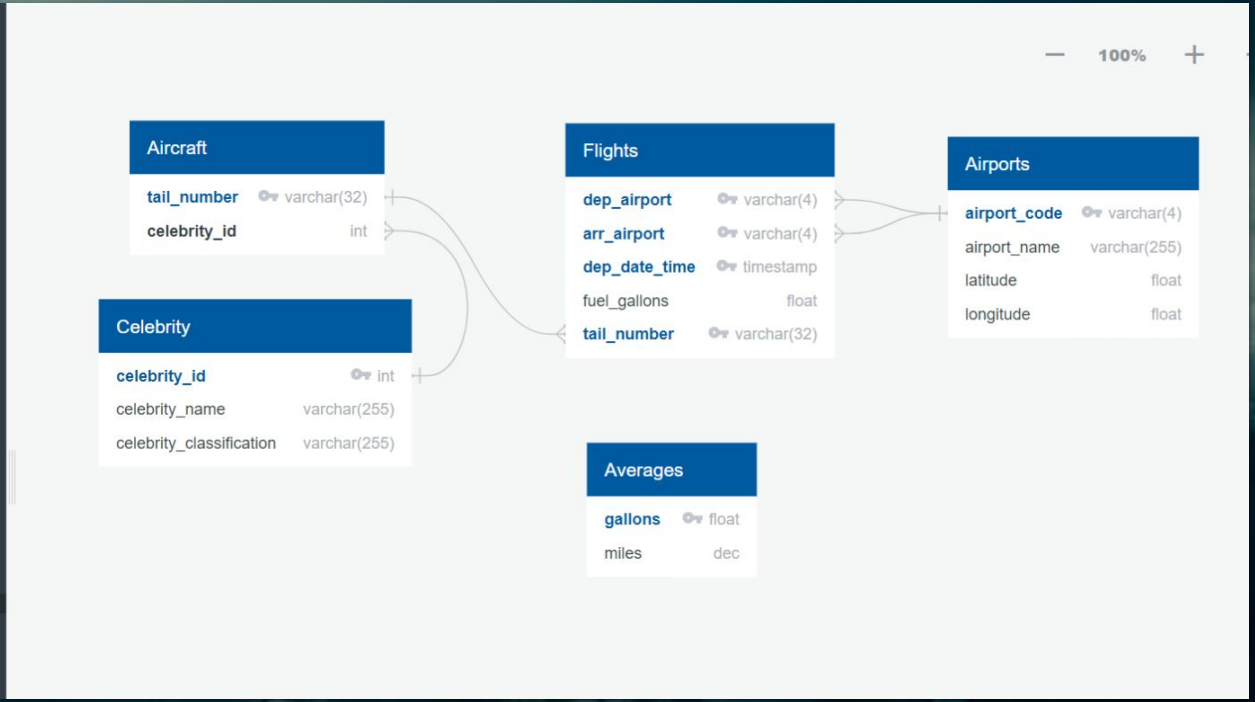
Database Architecture

- Database Used: PostgreSQL
  - PostgreSQL was used to ensure data integrity of the scraped data.
  - 5 Tables were created:
    - Aircraft - List of Aircrafts associated with different celebrities
    - Celebrity - List of Celebrities with ID and types: Business, Sports, Entertainer, Actor
    - Airports - List of Airports with airport code, name and location details
    - Averages - US averages for miles flown and Gallons used
    - Flights - List of flight data with date of travel, departure and arrival airport information
  - CSVs were imported into Postgres for use with FLASK APIs.
  - Primary and Foreign keys created and depicted in the following ERD diagram

Anusha

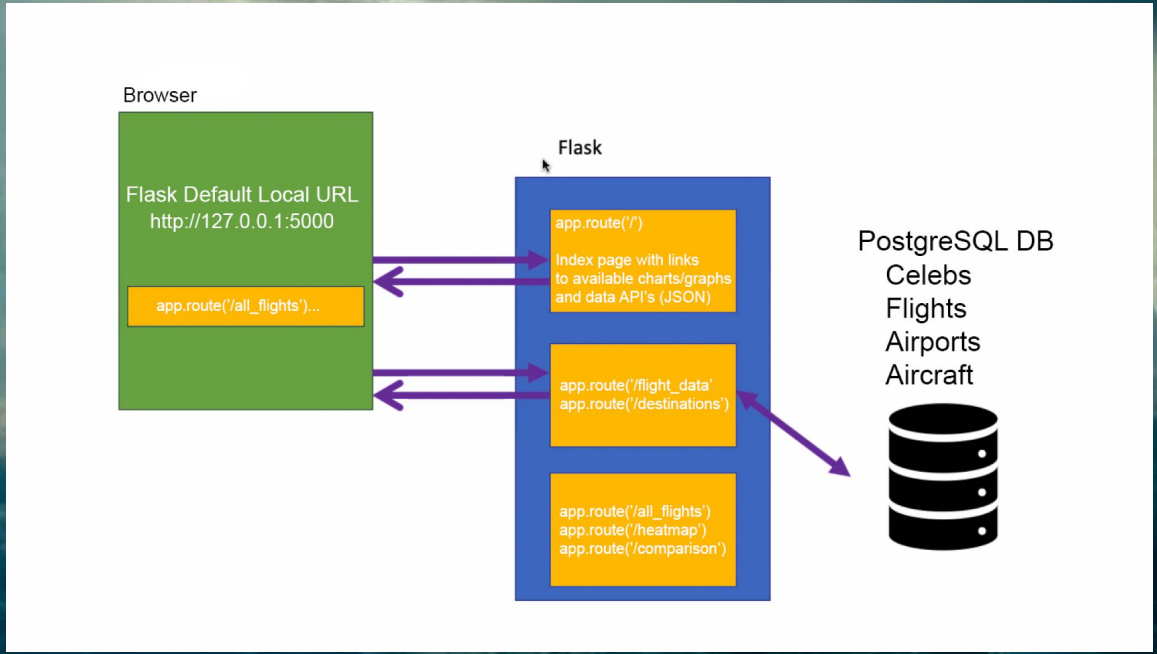
Entity Relationship Diagram

```
1  Celebrity
2  -
3  celebrity_id PK int
4  celebrity_name varchar(255)
5  celebrity_classification varchar(255)
6
7  Aircraft
8  -
9  tail_number PK varchar(32)
10 celebrity_id int FK >- Celebrity.celebrity_id
11
12 Airports
13 -
14 airport_code PK varchar(4)
15 airport_name varchar(255)
16 latitude float
17 longitude float
18
19 Averages
20 -
21 gallons PK float
22 miles dec
23
24 Flights
25 -
26 dep_airport PK varchar(4) FK >- Airports.airport_code
27 arr_airport PK varchar(4) FK >- Airports.airport_code
28 dep_date_time PK timestamp
29 fuel_gallons float
30 tail_number PK varchar(32) FK >- Aircraft.tail_number
31
```



Anusha

Application Architecture



Jared

Jared

Challenges and Next Steps

Challenges:

- Accuracy Determination
- Scraping Speed
- Data Integration from Multiple Sources

Next Steps:

- Proof or verification of accuracy
- Subscribe to real-time or near-real-time info services
- Attempt to track and collect real-time flight info

Jim