Group Members: Chris Whamond, Shawn Lee, Jackie Chan, Aditya Karan

Project Description

Title: Flight Pricing and User Profiles

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

   We will have a database that contains the following: (1) flight information including the carrier, time of departure/arrival on each part of the itinerary, (2) information about the user and their frequently visited websites, and (3) the price that a given user saw for a flight as well as the seller of that flight. The data comes from sock puppet accounts that were tasked to scrape data from Kayak, an online travel agency website, over the past year to determine whether web browsing history affects the prices shown to end users. For this project, users can also supply information about their own history and the prices they see on Kayak to provide additional data.

2. **What are the basic functions of your web application? (What can users of this website do? Which simple and complex features are there?)**

   Our web application will allow users to access an existing database of historic flight prices for consideration when making their travel plans. The user will be able to query the database to view results for specified airport pairings, selected airlines, and preferred time of year. The available data will be returned to the user, both summarized and visualized in graphs. In addition, the user will be able to "save" particular search parameters into their list of "saved flight profiles" for future reference, using their unique user ID. The user can view, add, delete, and update their saved flight patterns as their travel plans evolve. Finally, we introduce additional complexity to the application through the inclusion of "user attribute profiles" when viewing the flight data. Based on our predefined user attribute profiles, the user will be able to update their search queries to see how historic flight prices differed between users with specified attributes, like gender and age.

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?)**

   The core functions of the website will allow the user to access our database of historic flight price data and refine search parameters to see changes in price between the predefined user attribute profiles. We also plan to implement functionality to allow the user to contribute individual flight prices they obtain on their own from direct services (like GoogleFlights, airline

websites, etc.) and add this information to our database. In addition, when providing data in the application, the user will be able to specify some of their personal attributes, expanding the available set of user attribute profiles. These components are additive both in (1) adding additional rows into our database to reflect current market conditions and (2) giving users more user attribute profiles to use when comparing flight prices. A user who finds themselves not represented well with our predefined users can effectively insert their general profile into our database for future reference, both by that user and by others using the application.

Summary

The price of airline tickets often seems like an arbitrary market set by carriers. Users purchasing tickets are presented with a single price for a single flight, and make an isolated decision about whether to purchase a ticket. Behind this single price, however, airlines are often presenting different users with different prices in an effort to maximize the amount they can collect for each ticket. Since users can only see the price they are being offered at that instant, there is little comparison to be done. Our application seeks to empower users with a database of historic flight prices to use when benchmarking current ticket offerings. Further, our application will allow the user to see how prices on an individual flight were offered differently to mock users with various gender-age combinations.

Our application will create real value for the user each time they are able to reference our database for a flight they are interested in. In addition, we want users to be able to contribute their own flight price data. This helps our database maintain current records, expand the repository of historical data, and gradually expand the set of attribute profiles available to other users when viewing pricing differences on a single ticket. Overall, our intention is for our application to help users make more informed decisions about airline ticket purchases. By leveraging an existing database of scraped data and allowing users to contribute their own data, our application will accomplish this goal.

Description

Flight ticket prices are subject to change based on numerous factors including the time of day prices are checked, the location of the user checking flights, and even the websites the user has previously visited. Without insider knowledge, it is impossible to know all the parameters that make up the final price users see on flight listings. That structure leaves people vulnerable to the predatory marketing strategies ticket sellers use to maximize the price people are willing to pay for a flight. Our website, which presents historical flight ticket prices, allows users to see price fluctuations from a variety of different carriers as well as sock puppet accounts that can highlight pricing differences influenced by user profile data. This website is informative because it allows users to see how much pricing changes to allow users to make better-informed decisions when it comes to travel planning.

Websites like Kayak make it hard to come by this type of historical data for individual users, so our website fulfills that need by allowing users to search through our database of historical flight prices. We

overcome the challenge that users have with finding this type of data through multiple sock puppet accounts that scrape Kayak for relevant pricing information for multiple months. In addition to basic querying, we will provide visualizations that will better inform users of the shape of the data to, again, better inform travelers of the pricing landscape, and allow the user to enter flight prices they are offered into our database.

Usefulness

Our application offers many use cases. First, giving users access to the historic ticket price data itself is valuable, since neither carrier websites nor other aggregators provide this level of information. The data will be provided in a user-friendly way with visual tools such as graphs so that users can analyze the data in a more convenient way. Also, our application uniquely allows users to compare flight prices that they see relative to the price others may see for a comparable flight, based on their own profile. This gives users a view into pricing schemes used by airline ticket sellers and can help the user save money if they guide their purchasing with the data. Lastly, the website will allow users to contribute their own data so that the database becomes more valuable to users as traffic grows and more flight data is compiled.

Realness

The source of our data is detailed more completely in the first section of the **Project Description**. Our database contains prices for actual flights, scraped from the aggregator site Kayak. The value of the application comes from both the realness of the data and from the profile-level insights the user can gain into how prices may appear differently to users with other attributes. The motivation for initial data collection stems from a research initiative to understand how online behavior influences prices; we seek to expand the scope of this work by (1) making the data more accessible to end users and (2) allowing users to contribute their own data to expand the database of flight prices and available user attribute profiles.
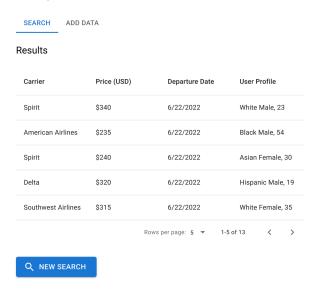
Low-Fidelity UI Mockup

Historical Flight Database

Query historical data from Kayak through multiple pre-defined profiles. Provide the necessary fields below a hit search!

SEARCH   ADD DATA

Flight Options

```
┌ Carrier ─────────────────┐
│ Carrier              ▾   │
└──────────────────────────┘
```

```
┌ Departure ───────────────┐
│ Departure            ▾   │
└──────────────────────────┘
```

```
┌ Destination ─────────────┐
│ Destination          ▾   │
└──────────────────────────┘
```

Flight Date Options

```
┌ Departure Time ──────────┐
│ Time                 🕐  │
└──────────────────────────┘
```

```
┌ Departure Date ──────────┐
│ Date                 📅  │
└──────────────────────────┘
```

```
┌ Arrival Time ────────────┐
│ Time                 🕐  │
└──────────────────────────┘
```

```
┌ Arrival Date ────────────┐
│ Time                 📅  │
└──────────────────────────┘
```

```
┌ User Profile ────────────┐
│ All                  👤  │
└──────────────────────────┘
```

🔍 SEARCH

Mockup 1: Default query page with basic parameters to modify the query.

Historical Flight Database

Query historical data from Kayak through multiple pre-defined profiles. Provide the necessary fields below a hit search!

SEARCH   ADD DATA

Results

Carrier	Price (USD)	Departure Date	User Profile
Spirit	$340	6/22/2022	White Male, 23
American Airlines	$235	6/22/2022	Black Male, 54
Spirit	$240	6/22/2022	Asian Female, 30
Delta	$320	6/22/2022	Hispanic Male, 19
Southwest Airlines	$315	6/22/2022	White Female, 35

Rows per page: 5 ▾   1-5 of 13   ‹   ›

🔍 NEW SEARCH

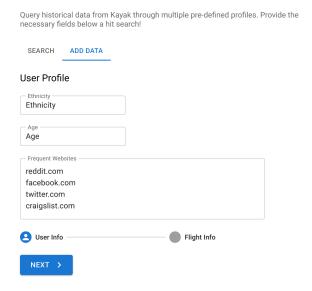**Mockup 2:** Results page following a submitted query.

Historical Flight Database

Query historical data from Kayak through multiple pre-defined profiles. Provide the necessary fields below a hit search!

SEARCH   **ADD DATA**

User Profile

Ethnicity
> Ethnicity

Age
> Age

Frequent Websites
> reddit.com
> facebook.com
> twitter.com
> craigslist.com

👤 **User Info** ————————— ⚪ **Flight Info**

[NEXT >]

Mockup 3: Submit data page where users can input their flight price data.

Project Work Distribution

While group members will share the implementation of each project component, we have given each member a broad primary area of ownership, with the understanding that each group member will contribute to each component of the project where possible to ensure an even workload distribution.

Here are our assignments of primary areas of responsibility:
- **Chris Whamond:** Project management, visualization of data.
- **Shawn Lee:** Lead back-end.
- **Jackie Chan:** Lead front-end.
- **Aditya Karan:** Lead database.

We will continue to allocate more specific responsibilities as the scope of the work in each area becomes more clear during implementation. We will also be sure to break down the database work so that everyone gets a chance to work on the database and is familiar with it by the end of the project.