Motivation

Hopper customers use the Hopper App to do more than just buy flight tickets and book hotel rooms. Hopper is a valuable tool for price sensitive travelers who want to plan their travel arrangements easily and alleviate the anxiety of booking travel at the best possible price.

As Data Scientists, we strive to combine unique behavioral and market data with algorithms and data-driven insights to create product features that enable our users to reach their travel goals. An example of such a feature are our "Buy or Wait" recommendations; when users search on Hopper, we forecast the lowest price one can expect for his/her trip, and tell him/her to wait for a better deal if we expect prices to drop. Only if we find a good deal do we recommend users to buy. Nearly half of searchers are recommended to "watch" a trip and wait for a better price, leading to savings averaging over $100.

We're on the hunt for new ways to help our customers find great deals for their upcoming trips. To do so effectively, we need to understand what a user's travel intent is when they search and watch, so that we can make recommendations that are in line with their intent. This begs the question:

> Given a user's search and watch history, can we understand the types of **trips** he/she is planning? If so, how can we use this knowledge to add value to this user?

Notice the emphasis on the **trips**; while a search is characterized by an origin airport, destination airport, departure date and return date, people do not always think in terms of these parameters when planning an upcoming trip.

For example: suppose a particular user has made the following two searches in short succession:

| Search Number | Origin | Destination | Departure Date | Return Date |
|---|---|---|---|---|
| 1 | Boston | Cancun, Mexico | Thursday, Dec. 26, 2019 | Saturday, Jan. 4, 2020 |
| 2 | Boston | Miami, USA | Friday, Dec 27, 2019 | Sunday, Jan 5, 2020 |

One way to interpret these searches is to consider each a distinct trip, and give travel advice for each search independently. However, if the user's intent is to travel someplace warm during Christmas time, we can provide advice that is more pertinent if we can correctly identify this broader trip intent.

## Assignment

In the `search_sample_data` directory, you'll find the search and watch history of a sample of Hopper users, which you'll use to complete this challenge. A description of the datasets provided can be found in `table-metadata.md`

Your task is to develop a proposal for how we can add value by automatically grouping user searches/watches into "trips," and to quantify the potential impact of your approach. How you choose to define these groupings and add value is up to you; some examples are alternative flight recommendations and helpful travel advice.

## Success Criteria

We do not expect a complete, production-quality solution. Instead, we're more interested in your approach and your understanding of the problem, and how a solution might impact our users for the better.

We do not enforce that you take use a particular approach to solving this problem, and encourage creativity. There are, however, several qualities we look for in submissions:

- **User focused**: We are looking for submissions that prove that you're passionate about understanding the nuances of user behavior and intent, and are thinking of how you identify and alleviate pain-points or deliver delight.
- **Action oriented**: Data Scientists at Hopper are responsible not only for finding opportunities, but also for building and shipping solutions. We value submissions that show that you not only have good ideas, but also demonstrate that you can prototype a simple solution, and quantify the impact of said solution.
- **Transparency and Storytelling**: We prefer simple solutions that we can explain to our colleagues and customers - even at the cost of performance/generality.

## Submission Format

Please submit a writeup describing your thought process and approach to solving this problem. Also, please submit any code you wrote (Jupyter notebooks, RMarkdown, scripts, etc) while working on this assignment.