

Modeling for Price Accuracy in the Airline Tickets Domain

Skyscanner

About Skyscanner:

Skyscanner is a leading global travel search site, a place where people are inspired to plan and book direct from millions of travel options at the best prices.

We are unbiased and free, which means that the 60 million people who use us every month can trust our comprehensive range of flight, hotel and car hire options.

We have a unique proprietary technology that connects people directly to everything the travel industry has to offer. We also power travel search for over 1200 partners through our Skyscanner for Business products.

Skyscanner is a "data company". It does not sell a travel product (e.g. flight ticket). The business model behind our product is to aggregate data from as many as possible travel product companies (Airlines, Online Travel Agencies, Hotel Chains, Car Hire companies) and allow our users to compare prices and choose the best options for their trips. Once a user finds her perfect choice she buys it from our partner that offers that option.

Context:

The Flights travel industry has a complex data ecosystem. This causes a number of problems that lead to direct impact with the end-users. The most significant one is the accuracy of the displayed flight ticket prices: often people see one price when they search for their trip and end up being asked for a different price when they reach the payment stage.

The reasons:

- There is no such thing as price for a ticket - it changes dynamically based on "seats left in a plane" (a.k.a. Availability).
- More than 3.5 billion people fly every year. Thousands of seats are sold every minute.
- The data is expensive (because of the computational complexity to keep it synchronized and derive it).
- The data takes a long way to reach the end user... and everyone caches it because of the cost.
- You never know through how many caches did a price go (how old is the data). You never know if it is still available.

The effect:

- Users of Skyscanner see a price that is not accurate.
- They don't have visibility about the complexity that leads to that. They should not and do not care if it is Skyscanner's fault.
- Skyscanner aims to be the most trusted brand in travel, thus WE need to make sure the price is correct.

This leads to the following problem we are actively trying to solve (a never-ending battle):

Out of all prices that we get from our partners, identify those that are not accurate and remove them from the list of options displayed to a customer. While doing that filter as few prices that are accurate as possible, because that limits the options to choose from for the end user.

To solve that problem, we track a subset of the user experiences OUTSIDE of Skyscanner's web site and app (we call that an "Exit"). That allows us to identify cases where the price changes, in what direction (up or down) and with how much.

The scale is such that tracking every Exit literally overloads (shuts down) the systems of most of our partners, thus we can only do that for a small subset of the Exits.

The task:

Based on the provided data for prices displayed on Skyscanner and later booked by our users, build a model that filters as many of the inaccurate prices as possible AND as little accurate prices as possible.

The definition of "accurate" comes from data about our user's perception of the problem and is set to USD +4.0, meaning that if a booked price is more than \$4 above the price we displayed, it is considered inaccurate.

Using the provided data, build a model that removes as many of the inaccurate prices AND as little as possible of the accurate prices.

Example:

We display a price for a flight from Sofia to New York (and back) in October that costs **\$474**, but after we "exit" to the partner who provides that price, we see a price of **\$481** (\$7 more expensive). That is considered an inaccurate price.

If in the same case we were to pay **\$478** (\$4 more expensive) that would have been considered an accurate price.

If the final payment is at **\$454** (\$20 cheaper) it would still be considered an accurate price (users are not complaining in that scenario).

Thus, for a price to be considered accurate it needs to be \$4 or less MORE expensive or LESS expensive than the originally displayed price.