



Personalize Expedia Hotel Searches - ICDM 2013

Learning to rank hotels to maximize purchases

\$25,000 · 337 teams · 4 years ago

[Overview](#)[Data](#)[Discussion](#)[Leaderboard](#)[More](#)[Submit Predictions](#)

Training Data

4 files

basicPythonBenchmark...

data.zip

randomBenchmark.zip

testOrderBenchmark.z...

data.zip

File size 414.44 MB

[Download File](#)

Data Introduction

[Sample code to create the benchmarks is available on Github.](#)

- **train.csv** - the training set
- **test.csv** - the test set (this contains data for both the public leaderboard and the final evaluation, which is randomly split between the two sets)

Note: **test.csv** does not contain the following columns: position, click_bool, gross_bookings_usd, nor booking_bool

You can refer to www.expedia.com to better understand hotel search.

“Hotel” refers to hotels, apartments, B&Bs, hostels and other properties appearing on Expedia’s websites. Room types are not distinguished and the data can be assumed to apply to the least expensive room type.

Most of the data are for searches that resulted in a purchase, but a small proportion are for searches not leading to a purchase.

Usage of outside data is prohibited and modeling should focus fully on the given data.

Column Name	Data Type	Description
srch_id	Integer	The ID of the search
date_time	Date/time	Date and time of the search
site_id	Integer	ID of the Expedia point of sale (i.e. Expedia.com, Expedia.co.uk, Expedia.co.jp, ..)
visitor_location_country_id	Integer	The ID of the country the customer is located
visitor_hist_starrating	Float	The mean star rating of hotels the customer has

visitor_hist_adr_usd	Float	<p>previously purchased; null signifies there is no purchase history on the customer</p> <p>The mean price per night (in US\$) of the hotels the customer has previously purchased; null signifies there is no purchase history on the customer</p>
prop_country_id	Integer	The ID of the country the hotel is located in
prop_id	Integer	The ID of the hotel
prop_starrating	Integer	The star rating of the hotel, from 1 to 5, in increments of 1. A 0 indicates the property has no stars, the star rating is not known or cannot be publicized.
prop_review_score	Float	The mean customer review score for the hotel on a scale out of 5, rounded to 0.5 increments. A 0 means there have been no reviews, null that the information is not available.
prop_brand_bool	Integer	+1 if the hotel is part of a major hotel chain; 0 if it is an independent hotel
prop_location_score1	Float	A (first) score outlining the desirability of a hotel's location
prop_location_score2	Float	A (second) score outlining the desirability of the hotel's location
prop_log_historical_price	Float	The logarithm of the mean price of the hotel over the last trading period. A 0 will occur if the hotel was not sold in that period.
position	Integer	Hotel position on Expedia's search results page. This is only provided for the training data, but not the test data.
price_usd	Float	Displayed price of the hotel for the given search. Note that different countries have different conventions regarding displaying taxes and fees and the value may be per night or for the whole stay
promotion_flag	Integer	+1 if the hotel had a sale price promotion specifically displayed
gross_booking_usd	Float	Total value of the transaction. This can differ from the price_usd due to taxes, fees, conventions on multiple day bookings and purchase of a room type other than the one shown in the search
srch_destination_id	Integer	ID of the destination where the hotel search was performed
srch_length_of_stay	Integer	Number of nights stay that was searched
srch_booking_window	Integer	Number of days in the future the hotel stay started from the search date
srch_adults_count	Integer	The number of adults specified in the hotel room
srch_children_count	Integer	The number of (extra occupancy) children specified in the hotel room
srch_room_count	Integer	Number of hotel rooms specified in the search
srch_saturday_night_bool	Boolean	+1 if the stay includes a Saturday night, starts from Thursday with a length of stay is less than or equal to 4 nights (i.e. weekend); otherwise 0
srch_query_affinity_score	Float	The log of the probability a hotel will be clicked on in Internet searches (hence the values are negative) A null signifies there are no data (i.e. hotel did not register in any searches)
orig_destination_distance	Float	Physical distance between the hotel and the customer at the time of search. A null means the distance could not be calculated.
random_bool	Boolean	+1 when the displayed sort was random, 0 when the normal sort order was displayed
comp1_rate	Integer	+1 if Expedia has a lower price than competitor 1 for the hotel; 0 if the same; -1 if Expedia's price

is higher than competitor 1; null signifies there is no competitive data

comp1_inv	Integer	+1 if competitor 1 does not have availability in the hotel; 0 if both Expedia and competitor 1 have availability; null signifies there is no competitive data
comp1_rate_percent_diff	Float	The absolute percentage difference (if one exists) between Expedia and competitor 1's price (Expedia's price the denominator); null signifies there is no competitive data
comp2_rate		
comp2_inv		(same, for competitor 2 through 8)
comp2_rate_percent_diff		
comp3_rate		
comp3_inv		
comp3_rate_percent_diff		
comp4_rate		
comp4_inv		
comp4_rate_percent_diff		
comp5_rate		
comp5_inv		
comp5_rate_percent_diff		
comp6_rate		
comp6_inv		
comp6_rate_percent_diff		
comp7_rate		
comp7_inv		
comp7_rate_percent_diff		
comp8_rate		
comp8_inv		
comp8_rate_percent_diff		



site_id

Home

Vacation Packages

Hotels

Cars

Flights

Cruises

Things to Do

DEALS

PLAN YOUR TRIP ON EXPEDIA

- ☐ Flight
- ☒ Hotel
- ☐ Car
- ☐ Activities
- ☐ Cruise

- ☐ Flight + Hotel
- ☐ Flight + Car
- ☐ Flight + Hotel + Car
- ☐ Hotel + Car



Hotel

Find hotels near:

A city, airport or attraction

What City?

New York (and vicinity), New York, United States of America

Check-in:

10/18/2013

Check-out:

10/20/2013

Rooms:

1

srch_booking_
window

srch_length_of_stay

srch_destination_id

srch_room_count

Room 1

2

0


srch_adults_count

srch_children_count



**BEST PRICE
GUARANTEE**

SEARCH FOR HOTELS



prop_id

Pod 39

4.3 out of 5

New York

Map

1-866-267-9053

Most Popular

296 people booked this hotel in the last 48 hours

prop_starrating

prop_review_score
(rounded to 0.5)

promotion_flag

Sale

Best Price

Only 5 rooms left at this price

price_usd

\$235
avg/night

Trip Summary



Pod 39
New York, NY



1 Room: Queen Pod

2 Nights: Oct/18/2013 - Oct/20/2013

Best Price

Room 1: 2 Adults

2 Nights

avg./night

\$275.00

\$44.07

gross_booking_USD

Trip Total: \$638.14