

Datafest 2017 Data Dictionary

Notes:	Each row represents a site user's interaction with an Expedia site ("event"), which is either clicking on or booking a hotel, during a session.
	A session is defined as a series of user actions that are not interrupted for more than 30 minutes.

Name of field	Description
date_time	Timestamp time, date, year in user's local time
site_name	Expedia point of sale, i.e., the site (Expedia.com is the US, Expedia.ca is Canada, Expedia.de is Germany, though Expedia.com is often used by people outside the US)
user_location_country	Country the customer is located at the time of interaction with Expedia sites
user_location_region	Region the customer is located (State in the US, greater metro area for Europe, etc but these data can be messy!)
user_location_city	City the customer is located at the time of interaction with Expedia sites
user_location_latitude	Latitude of the city where the customer is located, i.e. the city center lat-long
user_location_longitude	Longitude of the city where the customer is located, i.e. the city center lat-long
orig_destination_distance	Physical distance between a hotel and a customer at the time of search, in miles
user_id	ID of a user (unique)
is_mobile	1 when a user connected from a mobile device (whether app or not), 0 otherwise
is_package	1 if the click/booking was generated as a part of a package search (i.e. a hotel combined with a flight and/or car rental), 0 otherwise
channel	ID of a marketing channel (anonymized), i.e. how a user may have arrived at an Expedia site, such as directly, a google search, a link from Tripadvisor, a link from Trivago, from a marketing email Expedia sent, etc.
srch_ci	Check-in date specified in the customer search
srch_co	Check-out date specified in the customer search
srch_adults_cnt	The number of adults specified to occupy the hotel room
srch_children_cnt	The number of (optional) children specified to occupy the hotel room
srch_rm_cnt	The number of hotel rooms specified in the search
srch_destination_id	ID of the destination where the hotel search was performed
hotel_country	Country the hotel is located in
is_booking	1 if a booking, 0 if a click
hotel_id	ID of the hotel (there is no particular significance to the ID number)
prop_is_branded	1 if the hotel is part of a major hotel chain (Hilton, Marriott, Sheraton, etc.), 0 if it is not
prop_starrating	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
distance_band	Banded distance of a hotel from the search destination center relative to other hotels in the same destination (VC = very close, C = close, M = medium close, F = far, VF = very far). Generally, hotels closer to the center are more desirable (but not always)
hist_price_band	Banded historical purchase price of a hotel relative to other hotels in the same destination (VL = very low, L = low, M = medium, H = high, VH = very high)
popularity_band	Banded hotel popularity relative to other hotels in the same destination, i.e. how often it is booked on Expedia (VL = very low, L = low, M = medium, H = high, VH = very high)
cnt	Number of similar events (clicks or bookings) in the context of the same user session

dest.txt
Destination File

Name of field	Description
srch_destination_id	ID of the destination where the hotel search was performed
srch_destination_name	Name of the destination where the hotel search was performed
srch_destination_type_id	Type of destination
srch_destination_latitude	Latitude of destination
srch_destination_longitude	Longitude of destination
popular_*	Popularity scores of travel related facets of destinations. The interpretation of the popularity score is log10 of the probability that customer would endorse a destination for specific facet. These data are from the reviews that Expedia customers wrote for hotels in each destination. As not all events are mentioned in every destination, some sparsity is to be expected.