

Data Overview

Expedia is a travel searching and booking platform. This data specifically showcases property searches on the platform [1]. “Property” refers to hotels, vacation rentals, apartments, B&Bs, hostels or other rental-housing options appearing on the brand Expedia’s websites. The data span a period from 2021-06-01 to 2021-07-31 and contain searches for a random sample of consumers who made at least one click during the above time frame.

For your project you will investigate the search features and recommendations of Expedia consumers. Data provided to students will have amenity information on the first three listed properties. Please note that some of the data was de-identified in order to protect the privacy of the consumers.

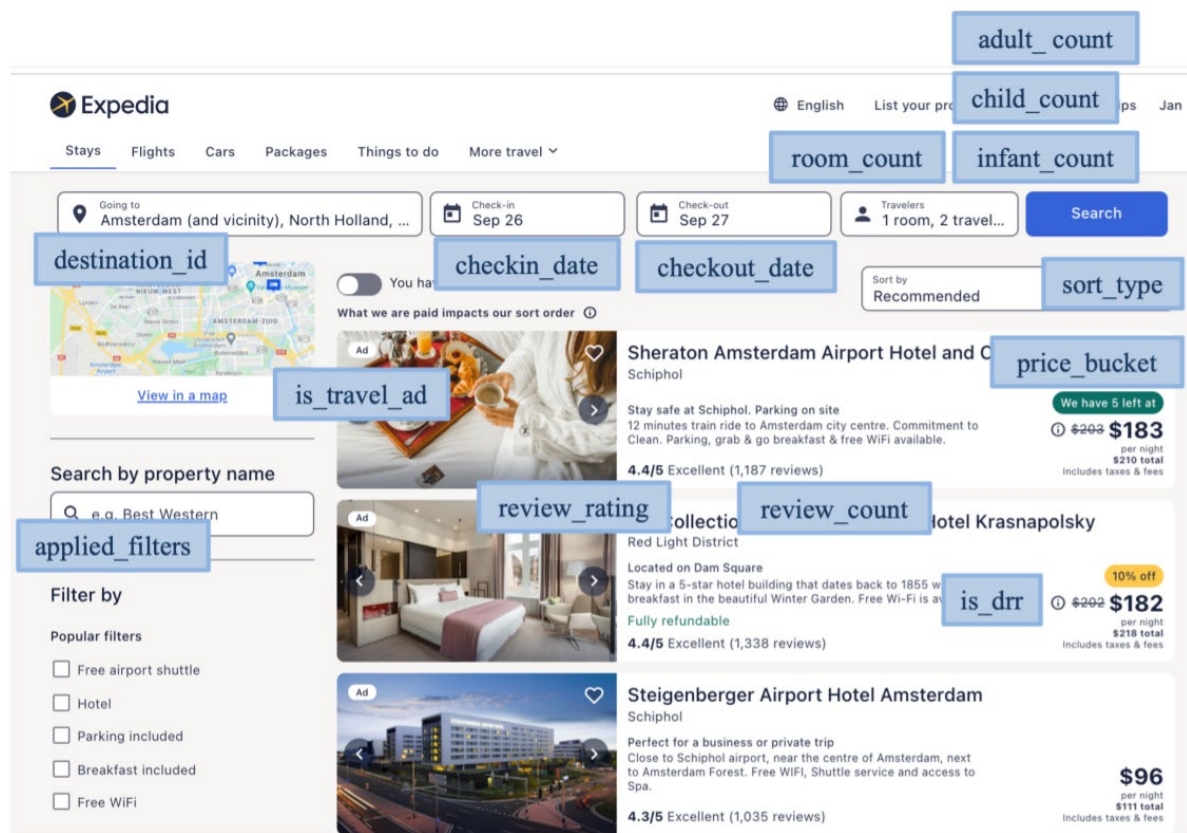


Figure 1: Data labels as seen on Brand Expedia sites.

Reference:

1. Adam Woznica and Jan Krasnodebski. 2021. Expedia Group RecTour Research Dataset. <http://ceur-ws.org/Vol-2974/invited1.pdf>.

Data Dictionary

Coded Variable Name	Type	Description
search_timestamp	date	Date and time of the search (rounded to the nearest minute)
geo_location_country	dbl	A numerical ID for the country/region where the consumer is located
is_mobile	dbl	Indicates whether the search was made from a mobile device. (1 = Yes, 0 = No)
destination_id	dbl	A numerical ID for the destination/region the booking property is located in
checkin_date	date	Stay start date
checkout_date	date	Stay end date
adult_count	dbl	The number of adults specified in the search
child_count	dbl	The number of children specified in the search
infant_count	dbl	The number of infants specified in the search
room_count	dbl	The number of rooms specified in the search
is_travel_ad1	dbl	If the first listed property is a travel ad (labelled "Ad", pay per click advertisement) (1 = Yes, 0 = No)
review_rating1	dbl	The mean customer review score for the first listed property on a scale out of 5, rounded to nearest integer. A 0 means there have been no reviews.
review_count1	dbl	The number of reviews for the first listed property rounded to the nearest 25
star_rating1	dbl	The star rating of the first listed booking, from 1 to 5.
is_free_cancellation1	dbl	If the first listed booking can be cancelled without extra fees (1 = Yes, 0 = No)
is_drr1	dbl	If the first listed property had a discount price reduction specifically displayed ("strikeout" price) (1 = Yes, 0 = No)
price_bucket1	chr	Price bucket (1-5) of the first listing, based on percentile of the distribution of impressed prices; lower values of price_bucket correspond to lower prices.
num_clicks1	dbl	Number of clicks, made by the consumer, on the first listed property within 180 minutes of the search timestamp
is_trans1	dbl	If there was a transaction made on the first listing, within 180 minutes of the search timestamp (1 = Yes, 0 = No)
AirConditioning1	dbl	Indicates whether the first listing has air conditioning (1 = Yes, 0 = No)
FreeBreakfast1	dbl	Indicates whether the first listing has free breakfast (1 = Yes, 0 = No)
FreeParking1	dbl	Indicates whether the first listing has free parking (1 = Yes, 0 = No)

FreeWiFi1	dbl	Indicates whether the first listing has free WiFi (1 = Yes, 0 = No)
Gym1	dbl	Indicates whether the first listing has access to a gym (1 = Yes, 0 = No)
HotTub1	dbl	Indicates whether the first listing has access to a hot tub (1 = Yes, 0 = No)
LaundryFacility1	dbl	Indicates whether the first listing has access to a laundry facility (1 = Yes, 0 = No)
PetsAllowed1	dbl	Indicates whether the first listing allows pets (1 = Yes, 0 = No)
SwimmingPool1	dbl	Indicates whether the first listing has access to a swimming pool (1 = Yes, 0 = No)
is_travel_ad2	dbl	If the second listed property is a travel ad (labelled "Ad", pay per click advertisement) (1 = Yes, 0 = No)
review_rating2	dbl	The mean customer review score for the second listed property on a scale out of 5, rounded to nearest integer. A 0 means there have been no reviews
review_count2	dbl	The number of reviews for the second listed property rounded to the nearest 25
star_rating2	dbl	The star rating of the second listed booking, from 1 to 5.
is_free_cancellation2	dbl	If the second listed booking can be cancelled without extra fees (1 = Yes, 0 = No)
is_drr2	dbl	If the second listed property had a discount price reduction specifically displayed ("strikeout" price) (1 = Yes, 0 = No)
price_bucket2	chr	Price bucket (1-5) of the second listing, based on percentile of the distribution of impressed prices; lower values of price_bucket correspond to lower prices.
num_clicks2	dbl	Number of clicks, made by the consumer, on the second listed property within 180 minutes of the search timestamp
is_trans2	dbl	If there was a transaction made on the second listing, within 180 minutes of the search timestamp (1 = Yes, 0 = No)
AirConditioning2	dbl	Indicates whether the second listing has air conditioning (1 = Yes, 0 = No)
FreeBreakfast2	dbl	Indicates whether the second listing has free breakfast (1 = Yes, 0 = No)
FreeParking2	dbl	Indicates whether the second listing has free parking (1 = Yes, 0 = No)
FreeWiFi2	dbl	Indicates whether the second listing has free WiFi (1 = Yes, 0 = No)
Gym2	dbl	Indicates whether the second listing has access to a gym (1 = Yes, 0 = No)
HotTub2	dbl	Indicates whether the second listing has access to a hot tub (1 = Yes, 0 = No)

LaundryFacility2	dbl	Indicates whether the second listing has access to a laundry facility (1 = Yes, 0 = No)
PetsAllowed2	dbl	Indicates whether the second listing allows pets (1 = Yes, 0 = No)
SwimmingPool2	dbl	Indicates whether the second listing has access to a swimming pool (1 = Yes, 0 = No)
is_travel_ad3	dbl	If the third listed property is a travel ad (labelled "Ad", pay per click advertisement) (1 = Yes, 0 = No)
review_rating3	dbl	The mean customer review score for the third listed property on a scale out of 5, rounded to nearest integer. A 0 means there have been no reviews
review_count3	dbl	The number of reviews for the second listed property rounded to the nearest 25
star_rating3	dbl	The star rating of the second listed booking, from 1 to 5.
is_free_cancellation3	dbl	If the second listed booking can be cancelled without extra fees (1 = Yes, 0 = No)
is_drr3	dbl	If the second listed property had a discount price reduction specifically displayed ("strikeout" price) (1 = Yes, 0 = No)
price_bucket3	chr	Price bucket (1-5) of the second listing, based on percentile of the distribution of impressed prices; lower values of price_bucket correspond to lower prices.
num_clicks3	dbl	Number of clicks, made by the consumer, on the second listed property within 180 minutes of the search timestamp
is_trans3	dbl	If there was a transaction made on the second listing, within 180 minutes of the search timestamp (1 = Yes, 0 = No)
AirConditioning3	dbl	Indicates whether the third listing has air conditioning (1 = Yes, 0 = No)
FreeBreakfast3	dbl	Indicates whether the third listing has free breakfast (1 = Yes, 0 = No)
FreeParking3	dbl	Indicates whether the third listing has free parking (1 = Yes, 0 = No)
FreeWiFi3	dbl	Indicates whether the third listing has free WiFi (1 = Yes, 0 = No)
Gym3	dbl	Indicates whether the third listing has access to a gym (1 = Yes, 0 = No)
HotTub3	dbl	Indicates whether the third listing has access to a hot tub (1 = Yes, 0 = No)
LaundryFacility3	dbl	Indicates whether the third listing has access to a laundry facility (1 = Yes, 0 = No)
PetsAllowed3	dbl	Indicates whether the third listing allows pets (1 = Yes, 0 = No)
SwimmingPool3	dbl	Indicates whether the third listing has access to a swimming pool (1 = Yes, 0 = No)