

Dataset

TravelTide stores its data in a relational database, which you can access here:

postgres://Test:bQNxVzJL4g6u@ep-noisy-flower-846766.us-east-2.aws.neon.tech/TravelTide?sslmode=require

Connected via PostgreSQL

Description of each table and its columns:

users: user demographic information

- user_id: unique user ID (P-key, int)
- birthdate: user date of birth (datetime)
- gender: user gender (nominal)
- married: user marriage status (binary)
- has_children: whether or not the user has children (binary)
- home_country: user's resident country (nominal)
- home_city: user's resident city (nominal)
- home_airport: user's preferred hometown airport (nominal)
- home_airport_lat: geographical north-south position of home airport (decimal)
- home_airport_lon: geographical east-west position of home airport (decimal)
- sign_up_date: date of TravelTide account creation (datetime)

sessions: information about individual browsing sessions (note: only sessions with at least 2 clicks are included)

- session_id: unique browsing session ID (P-key, string)
- user_id: the user ID (F-key, int)
- trip_id: ID mapped to flight and hotel bookings (F- key, string)
- session_start: time of browsing session start (timestamp)
- session_end: time of browsing session end (timestamp)
- flight_discount: whether or not flight discount was offered (binary)
- hotel_discount: whether or not hotel discount was offered (binary)
- flight_discount_amount: percentage off base fare (decimal)
- hotel_discount_amount: percentage off base night rate (decimal)
- flight_booked: whether or not flight was booked (binary)
- hotel_booked: whether or not hotel was booked (binary)
- page_clicks: number of page clicks during browsing session (int)
- cancellation: whether or not the purpose of the session was to cancel a trip (binary)

flights: information about purchased flights

- trip_id: unique trip ID (P-key, string)
- origin_airport: user's home airport (nominal)
- destination: destination city (nominal)
- destination_airport: airport in destination city (nominal)
- seats: number of seats booked (int)
- return_flight_booked: whether or not a return flight was booked (binary)
- departure_time: time of departure from origin airport (timestamp)
- return_time: time of return to origin airport (timestamp)
- checked_bags: number of checked bags (int)
- trip_airline: airline taking user from origin to destination (nominal)
- destination_airport_lat: geographical north-south position of destination airport (decimal)
- destination_airport_lon: geographical east-west position of destination airport (decimal)
- base_fare_usd: pre-discount price of airfare (decimal)

hotels: information about purchased hotel stays

- trip_id: unique trip ID (P-key, string)
- hotel_name: hotel brand name (nominal)
- nights: number of nights stayed at the hotel (int)
- rooms: number of rooms booked with hotel (int)
- check_in_time: time user hotel stay begins (timestamp)
- check_out_time: time user hotel stay ends (timestamp)
- hotel_per_room_usd: pre discount price of hotel stay per room per night (decimal)

TravelTidefilteredSessionLvl.csv

- Sessions Level Aggregation
- Filters applied: session_start > '2023-01-04', only users with more than 7 sessions
- 49,211 rows
- 44 columns

TravelTideFinal.csv

- User Level Aggregation
- Filters applied: session_start > '2023-01-04', only users with more than 7 sessions
- Engineered Features
- 5,998 rows
- 21 columns