

Spark – projekt – zestaw danych airbnb

Tematyka

Dane zawierają informacje o apartamentach z Airbnb oraz ich użytkowników z trzech europejskich miast.

Pochodzenie danych

Projekt realizowany jest tylko i wyłącznie w oparciu o dane dostarczone przez prowadzącego. Poniższe informacje wyjaśniają jedynie pochodzenie oryginalnych danych.

Dane zostały uzyskane poprzez przetworzenie zbiorów pochodzących z:

<http://insideairbnb.com/get-the-data.html>

Charakterystyka danych

Dane dotyczące apartamentów i ich recenzji z serwisu Airbnb pochodzą z 3 zestawów danych symulujące 3 niezależne bazy danych o podobnym schemacie. Każdy z zestawów dotyczy jednego z następujących miast: Berlina, Madrytu, Paryża.

Każdy zestaw składa się z trzech plików, w tym dwóch plików CSV (symulujących tabele) zawierające informacje o apartamentach oraz jednego pliku TXT zawierającego recenzje.

1. <nazwa miasta>Calendar.csv – dane dotyczące dostępności apartamentów
 - a. listing_id
 - b. date
 - c. available
 - d. price
2. <nazwa miasta>Listings.csv – dane dotyczące dostępnych apartamentów
 - a. id
 - b. listing_url
 - c. scrape_id
 - d. last_scraped
 - e. name
 - f. summary
 - g. space
 - h. description
 - i. experiences_offered
 - j. neighborhood_overview
 - k. notes
 - l. transit
 - ...
 - m. requires_license
 - n. license
 - o. jurisdiction_names
 - p. instant_bookable
 - q. is_business_travel_ready
 - r. cancellation_policy
 - s. require_guest_profile_picture
 - t. require_guest_phone_verification
 - u. calculated_host_listings_count
 - v. reviews_per_month
3. <nazwa miasta>Reviews.txt – plik tekstowy zawierające recenzje w których jest między innymi informacja o id apartamentu, dacie recenzji, id recenzenta.