

Etap 1

Autorzy:

- Mateusz Matukiewicz
 - Jakub Kryczka
-

1. Problem

"Nie do końca rozumiemy, jakimi kryteriami kierują się klienci, którzy rezerwują dłuższe noclegi. Taka informacja bardzo pomogłaby naszym konsultantom."

2. Założenia

- Długie noclegi to takie, które trwają co najmniej **7 dni** (> 6.5).

Zadaniem jest zrozumienie, jakimi kryteriami kierują się klienci dokonujący długich rezerwacji, aby zwiększyć dochody i zmaksymalizować zajętość apartamentów.

W celu rozwiązania zadania zostanie zaimplementowany **model analityczny**, który pomoże sklasyfikować, czy dla danych wejściowych użytkownik dokona dłuższej rezerwacji.

3. Dane

1. Informacje o sesjach użytkowników oraz o tym, które apartamenty były tylko oglądane, a które faktycznie zarezerwowane.
 2. Szczegółowe dane o ofertach akomodacji, które mogą wpływać na decyzje klientów.
 3. Dane o użytkownikach.
 4. Recenzje użytkowników dotyczące konkretnych ofert.
-

4. Definicja długości noclegu

Długość noclegu wyliczamy jako różnicę z dwóch atrybutów w danych **sessions** dla akcji rezerwacji (**action = book_listing**):

(*aktualnie przy tej analizie danych nie obsługujemy logiki anulowania rezerwacji*)

- **booking_date – początek rezerwacji**
 - **booking_duration – koniec rezerwacji**
-

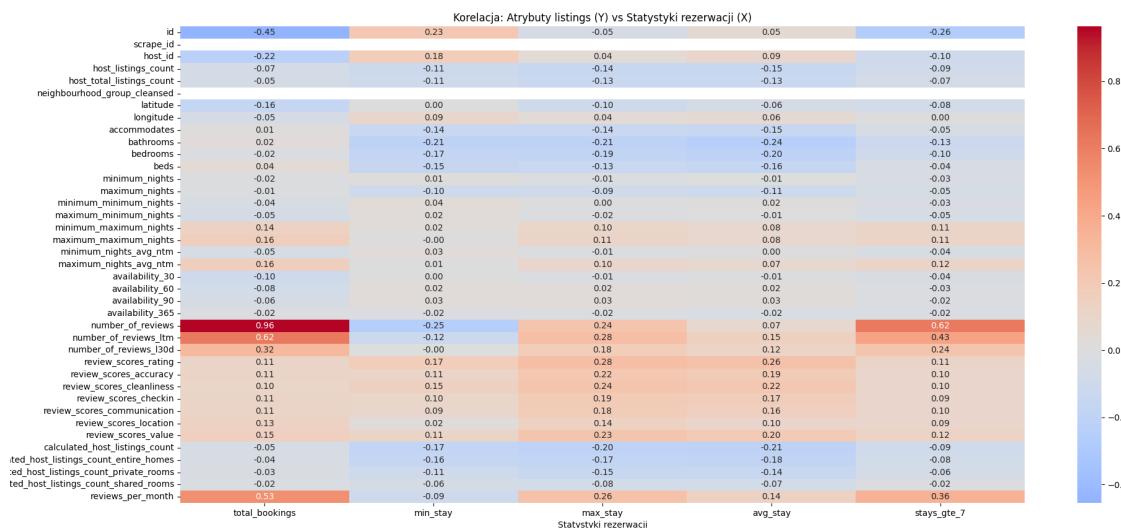
5. Na podstawie plików **sessions.csv** oraz **listings.csv** policzyliśmy:

1. Łączną liczbę rezerwacji danego obiektu.
2. Najkrótszy i najdłuższy pobyt.

3. Średnią długość pobytu — pozwalającą analizować wpływ parametrów obiektu na długość rezerwacji.
 4. Liczbę wynajęć obiektu na więcej niż 7 dni.
-

6. Macierz korelacji

Macierz korelacji istotnych atrybutów z danych **listings.csv** dla modelowania
(aktualnie wszystkich liczbowych; później zostanie podjęta decyzja o wyborze dokładnych atrybutów):



7. Wnioski

- **Największy wpływ** na łączną liczbę rezerwacji oraz liczbę długich rezerwacji ma **liczba opinii**.
- Natomiast na **maksymalną długość pobytu** największy wpływ ma **ocena**, a nie liczba opinii.