

Correspondência

Correlação entre os atributos existentes nas tabelas geradas pelo mapeamento e pela carga de dados.

Listings

A tabela Listings armazena anúncios de Propriedades no Airbnb. Assim, ela armazena diversas informações que serão correspondentes a Propriedades e Usuários, mas não diretamente a nenhuma dessas relações.

Para mapear a relação precisamos entender quais são as chaves primárias de Listings. Inicialmente, vemos que **name**, **host__id** é um forte candidato. Porém, **name** possui um nome em constante mudança devido as avaliações. Assim, efetuamos um recorte na string a fim de torná-lo chave:

```
UPDATE Listings
SET name_ = CONCAT(
    SPLIT_PART(name_, '.', 1),
    SPLIT_PART(name_, '.', 3),
    SPLIT_PART(name_, '.', 4),
    SPLIT_PART(name_, '.', 5)
)
```

Dessa forma, Condo in Rio de Janeiro · *4.70 · 2 bedrooms · 2 beds · 1 bath se transforma em Condo in Rio de Janeiro 2 bedrooms 2 beds 1 bath e podemos criar uma tabela em SQL para Listings.

Para mapear as informações pra propriedade, será usado as duas relação oriundas de Listings original:

| Atributo em Property | Atributo em Propriedade |
|-----------------------|-------------------------------------|
| name | Nome |
| host_id | IDAnfitriao |
| neighbourhood_group | NULL |
| neighbourhood | (indiretamente)IDLocalização→Bairro |
| latitude | NULL |
| longitude | NULL |
| room_type | CasaInteira |
| number_of_reviews | NULL |
| last_review | NULL |
| reviews_per_month | NULL |
| number_of_reviews_ltm | NULL |
| license | NULL |

| Atributo em Listings | Atributo em Propriedade |
|----------------------|-------------------------|
| id | NULL |
| name | Nome |
| host_id | IDAnfitriao |
| price | ValorDiaria |
| minimum_nights | MinNoites |

| Atributo em Listings | Atributo em Propriedade |
|----------------------|-------------------------|
| availability_365 | NULL |

Hosts

| Atributo em Hosts | Correspondência em Usuários |
|---------------------------------|-----------------------------|
| host_id | ID |
| host_name | nome |
| calculated_host_listings_counts | NULL |

A maior parte dos atributos de usuários não têm correspondência com a tabela de hosts obtida pela normalização, justamente pela falta de atributos que foi necessário adicionar **ID** a Usuários.

Calendar

Calendar só possui correspondência quando **available** é verdadeiro. A correspondência ocorre apenas em dois atributos.

| Atributo em Calendar | Correspondência em DatasDisponiveis |
|----------------------|-------------------------------------|
| listing_id | IDPropriedade |
| date | data |
| available | se True |
| price | NULL |
| adjusted_price | NULL |
| minimum_nights | NULL |
| maximum_nights | NULL |

Reviews

| Atributo em Reviews | Correspondência em Mensagem |
|---------------------|-----------------------------|
| listing_id | IDPropriedade |
| date | DataCriacao |

Todos os atributos que não foram colocados nas tabelas não tem correspondência entre eles, a exceção é **localização**, que pelo dataset ser exclusivo da cidade do Rio de Janeiro possui todos os atributos iguais, apenas variando o bairro.