

Equivalências (Modelagem própria → Airbnb)

Artur De Vlieger Lima [13671574], Gabriel da Costa Merlin [12544420],
Gustavo Gabriel Ribeiro [13672683], Lucas de Souza Brandão [13695021]

1. Mapeamento do nosso trabalho:

- a. Usuario(nome, sobrenome, num_tel, data_nasc, endereco, sexo, email, senha, eh_locatario, eh_anfitriao, id_local)
- b. Localizacao(id_local, pais, estado, cidade, bairro)
- c. Propriedade(nome, endereco, tipo, qtd_quartos, qtd_banheiros, max_hospedes, min_noites, max_noites, preco_noite, taxa_limpeza, data_inicio, data_fim, horario_checkIn, horario_checkOut, nomeUsu, sobrenomeUsu, num_telUsu, id_local)
- d. Quarto(id_quarto, nomeProp, endProp, qtd_camas, tipo_cama)
- e. Conta(num_conta, nomeUsu, sobrenomeUsu, num_telUsu, num_roteamento, tipo_conta)
- f. Avaliacao(tempo, nomeUsu, sobrenomeUsu, num_telUsu, nomeProp, endProp, data_checkIn, fotos, rate_limpeza, rate_comunicacao, rate_localizacao, mensagem)
- g. Locacao(data_checkIn, nomeUsu, sobrenomeUsu, num_telUsu, nomeProp, endProp, data_checkOut *, data_reserva, qtd_hospedes, imposto, preco_estadia, cod_promocional, desconto, preco_total, condicao)
- h. Pts_Interesse(ponto, id_local)
- i. Comodidades(comodidade, nomeProp, endProp)
- j. Regras(regra, nomeProp, endProp)

* data_checkOut poderia ser usada no lugar de data_checkIn. Assim, data_checkOut, nomeUsu, sobrenomeUsu, num_telUsu, nomeProp, endProp são chave candidata.

2. Tabelas do airbnb:

- a. **Listings:** ['id', 'name', 'host_id', 'host_name', 'neighbourhood_group', 'neighbourhood', 'latitude', 'longitude', 'room_type', 'price', 'minimum_nights', 'number_of_reviews', 'last_review', 'reviews_per_month', 'calculated_host_listings_count', 'availability_365', 'number_of_reviews_ltm', 'license']
- b. **Calendar:** ['listing_id', 'date', 'available', 'price', 'adjusted_price', 'minimum_nights', 'maximum_nights']
- c. **Review:** ['listing_id', 'id', 'date', 'reviewer_id', 'reviewer_name', 'comments']

3. Tabela de correspondências:

Aqui, temos uma tabela com os elementos para os quais encontramos uma equivalência entre o nosso mapeamento e as tabelas do AirBnB, e logo abaixo estão listados os dados para os

quais não encontramos nenhuma equivalência no nosso mapeamento e nos dados disponibilizados pelo Airbnb.

Nosso Mapeamento	Tabelas AirBNB
Usuario:nome	Listings: host_name ou Reviews:nomeUsu
Usuario:sobrenome	Embora em Listings: host_name e Reviews:nomeUsu, algumas pessoas colocam o nome completo aqui nessa coluna, usaremos ela só em Usuario:nome
Usuario:eh_locatario	Sempre será 1 se vier de reviews
Usuario:eh_anfitriao	Sempre será 1 se vier de listings
Localizacao:pais	Tabela vem do Brasil, então todos Brasil
Localizacao:estado	Tabela é do Rio de Janeiro, então Rio de Janeiro
Localizacao:cidade	Tabela é do Rio de Janeiro, então Rio de Janeiro
Localizacao:bairro	Listings:neighbourhood (válido apenas se for de propriedade)
Propriedade:nome	Listings:name(multivalorado no entanto)
Propriedade:endereco	Vamos considerar o Listings:neighbourhood, mas na verdade ele é só o bairro
Propriedade:tipo	Listings:room_type
Propriedade:qtd_quartos	Listings:name(multivalorado no entanto)
Propriedade:min_noites	Listings:minimum_nights
Propriedade:max_noites	Calendar:maximum_nights
Propriedade:preco_noite	Listings:price, existe também o calendar: price, mas esse seria para cada data específica
Propriedade:data_inicio Propriedade:data_fim	Calendar:date (aqui tem um data para cada dia, no entanto, os dados para uma dada propriedade estão completamente sequenciais, assim a disponibilidade será da data mais cedo da propriedade na tabela até a mais tarde)
Propriedade:qtd_banheiros	Listings:name(multivalorado no entanto)
Propriedade:nomeUsu	Listings: host_name

Quarto:nomeProp	Listings:name(multivalorado no entanto)
Quarto:endProp	Listings:neighbourhood
Quarto:qtd_camas	Listings:name(multivalorado no entanto)
Conta:nomeUsu	Listings: host_name
Avaliacao:tempo	Review:date
Avaliacao:nomeUsu	Review:reviewer_name
Avaliacao:nomeProp e Avaliação:endProp	A chave estrangeira aqui eh Review:listings_id
Avaliacao:mensagem	Review:comments
Locacao:data_reserva *	Calendar:date. Aqui temos uma coluna para cada dia da locação, e essa locação é só de datas futuras(todas começam a partir do dia 26 de dez de 2023), vamos usar elas como se fossem locações quando Calendar:available for falso continuamente
Locacao:nomeProp, Locacao:endProp *	A chave estrangeira aqui eh Calendar:listing_id
Locacao:preco_estadia *	Calendar:price
Locacao:imposto *	Talvez fosse o Calendar:adjusted_price, mas ele sempre é NaN
Locacao:preço_total *	Seria a soma dos calendar Price de uma mesma locação
Comodidades:nomeProp	Listings:name(multivalorado no entanto)
Comodidades:endProp	Listings:neighbourhood
Regras:nomeProp	Listings:name(multivalorado no entanto)
Regras:endProp	Listings:neighbourhood

* Pelas orientações do trabalho, não é necessário popular a tabela de Locações. Portanto, não faremos uso das equivalências supracitadas.

4. Dados sem equivalência:

a. Nosso mapeamento:

- i. Usuario(sobrenome, num_tel, data_nasc, endereco, sexo, email, senha, id_local)
- ii. Localizacao(id_local)

- iii. Propriedade(max_hospedes, max_noites, taxa_limpeza, horario_checkIn, horario_checkOut, sobrenomeUsu, num_telUsu, id_local)
- iv. Quarto(id_quarto, tipo_cama)
- v. Conta(num_conta, sobrenomeUsu, num_telUsu, num_roteamento, tipo_conta)
- vi. Avaliacao(sobrenomeUsu, num_telUsu, data_checkIn, fotos)
- vii. Locacao(data_checkIn, nomeUsu, sobrenomeUsu, num_telUsu, data_checkOut, qtd_hospedes, cod_promocional, desconto, condicao)
- viii. Pts_Interesse(ponto, id_local)
- ix. Comodidades(comodidade)
- x. Regras(regra)

b. Mapeamento AirBnB:

- i. **Listings:** ['id', 'host_id', 'neighbourhood_group', 'latitude', 'longitude', 'number_of_reviews', 'last_review', 'reviews_per_month', 'calculated_host_listings_count', 'availability_365', 'number_of_reviews_ltm', 'license']
- ii. **Calendar:** ['listing_id', 'adjusted_price', 'minimum_nights', 'maximum_nights']
- iii. **Review:** ['id', 'reviewer_id']

5. Alterações feitas para tornar a carga de dados possível:

Para as buscas pedidas, serão necessárias as tabelas de Usuário, Propriedade e Locações.

- **Usuários:** Aqui enfrentamos o problema de encontrar uma chave para usuário com base na tabela do Airbnb, pois usuário pode vir de duas tabelas, o único atributo chave que tinha equivalência entre o nosso mapeamento e o do Airbnb era o nome do usuário (atributo que não consegue ser chave sozinho). Por isso, decidimos usar o host_id da tabela Listings e o reviewer_id da tabela Review, e pelo que verificamos, um mesmo host_id e reviewer_id se referem à mesma pessoa nas duas tabelas (assim conseguimos usar ele como chave sozinho, pois não existe a possibilidade de termos diferentes valores de host_id e reviewer_id apontando para pessoas diferentes), portanto teremos como chave user_id (que equivale ao host_id ou reviewer_id).
- **Propriedade:** como nome e endereço do Airbnb não respeitam a unicidade da chave no Airbnb, tivemos que fazer uma mudança em nosso mapeamento. Para isso, usamos o atributo id em listings, que já funciona por si só como uma chave, para ser o atributo chave do nosso mapeamento. Decidimos chamar esse atributo de prop_id.
- **** Locações:** Em nosso mapeamento, a chave de locação dependia da data_checkin, nome_usu, sobrenomeUsu, num_telUsu, nomeProp e de endProp. Mas no mapeamento do Airbnb, não temos nenhum atributo do usuário na tabela calendar (que seria a mais próxima da nossa locação), o que nos faz ter que descartar todos os atributos de usuário da chave (nome_usu, sobrenomeUsu, num_telUsu). Além disso, como estamos usando o id como chave das propriedades, no lugar de nomeProp, endProp que são chaves estrangeiras usaremos o id da propriedade. Por fim, não

temos a equivalência para uma `data_checkin`, mas resolvemos usar a primeira data do `data_reserva` no lugar como equivalente.

- *Avaliacao*: Aqui, a nossa formulação para chave primária dessa relação mostrou-se ineficaz quando usada no escopo dos dados fornecidos pelo Airbnb, dado que, usávamos, como chave, *tempo*, *nomeUsu*, *sobrenomeUsu*, *num_telUsu*, *nomeProp*, *endProp*. Assim, precisávamos trocar as chaves estrangeiras de usuário e propriedade pelas novas, que têm fomato de id. Mesmo assim, essa chave, que identifica uma avaliação por meio da identificação de usuário, propriedade e timestamp da avaliação, não foi suficiente para identificar uma única avaliação nos dados do Airbnb. Isso ocorre por dois motivos: nos dados do Airbnb, não se usa timestamp, e sim uma data referente a um dia do ano; e um usuário pode avaliar a mesma propriedade mais de uma vez no mesmo dia. Dessa forma, tivemos que trocar nossa chave por `id_avaliacao`, chave utilizada pela empresa de hospedagem.

Assim teremos como chave das relações:

- Usuários: (`user_id`)
- Propriedade: (`prop_id`)
- **Locações: (`prop_id`, `data_checkin`)
- Avaliação: (`id_avaliacao`)

** Novamente, como não é necessário popular a tabela de Locações, essas afirmações, aqui presentes, sobre essa tabela servem apenas como esclarecimento teórico do que faríamos caso fosse necessário popular essas tabelas.