

Análise Exploratória usando o SQL Server: Boston Airbnb

O *dataset* está dividido em:

- propriedades_ofertas;
- propriedades_descricao;
- propriedades_reviews;
- propriedades_calendario;
- proprietarios;
- reviews;
- geolocalizacao;

Os resultados dos comandos são apresentados a seguir.

-- TOP 10 propriedades com maiores preços e qual é o tipo de propriedade de cada uma delas

```
SELECT TOP 10 po.id, pd.property_type, po.price
FROM propriedades_ofertas po
INNER JOIN propriedades_descricao pd
    ON po.id = pd.id
ORDER BY po.price DESC;
```

id	property_type	price
12972378	Apartment	4000.00
12679021	Apartment	3000.00
6093430	Apartment	1400.00
6451456	Apartment	1372.00
14798673	Apartment	1345.00
8303267	House	1300.00
12962229	Apartment	1275.00
10882267	Apartment	1250.00
14291722	Apartment	1250.00
14758045	Apartment	1235.00

-- Preço médio por tipo de propriedade

```
SELECT pd.property_type AS tipo_propriedade, ROUND(AVG(po.price),2) As
preço_médio
FROM propriedades_ofertas po
INNER JOIN propriedades_descricao pd
    ON po.id = pd.id
WHERE pd.property_type IS NOT NULL
GROUP BY pd.property_type
ORDER BY preço_médio DESC;
```

tipo_propriedade	preço_médio
Guesthouse	289.00
Boat	268.83

Other	229.94
Condominium	190.25
Villa	189.50
Loft	185.41
Apartment	181.97
Townhouse	149.04
House	131.60
Bed & Breakfast	129.37
Entire Floor	122.00
Dorm	62.50
Camper/RV	60.00

-- Desconto no preço de cada propriedade caso o hospede decida alugar o imóvel por uma semana ou por um mês

```
SELECT id, price,
       price*7 AS preço_bruto_semanal,
       weekly_price, weekly_price - price*7 AS desconto_semanal,
       price*30 AS preço_bruto_mensal,
       monthly_price, monthly_price - price*30 AS desconto_mensal
FROM propriedades_ofertas;
```

Obs: a tabela original possui 3.585 linhas

id	price	preço_bruto_semanal	weekly_price	desconto_mensal	preço_bruto_mensal	monthly_price	desconto_mensal
12147973	250	1750	NULL	NULL	7500	NULL	NULL
3075044	65	455	400	-55	1950	NULL	NULL
6976	65	455	395	-60	1950	1350	-600
1436513	75	525	NULL	NULL	2250	NULL	NULL
7651065	79	553	NULL	NULL	2370	NULL	NULL
12386020	75	525	NULL	NULL	2250	NULL	NULL
5706985	100	700	NULL	NULL	3000	NULL	NULL
2843445	75	525	NULL	NULL	2250	NULL	NULL
753446	58	406	NULL	NULL	1740	NULL	NULL
849408	229	1603	NULL	NULL	6870	NULL	NULL
12023024	60	420	NULL	NULL	1800	NULL	NULL
1668313	57	399	350	-49	1710	1125	-585
2684840	93	651	NULL	NULL	2790	NULL	NULL
13547301	150	1050	NULL	NULL	4500	NULL	NULL
5434353	145	1015	950	-65	4350	3000	-1350
225979	60	420	350	-70	1800	1100	-700
3420384	165	1155	750	-405	4950	NULL	NULL

1351 2930	75	525	NULL	NULL	2250	NULL	NULL
7482 195	49	343	NULL	NULL	1470	NULL	NULL
7252 607	49	343	280	-63	1470	950	-520

 -- Média de preço por política de cancelamento

```
SELECT pd.property_type, po.cancellation_policy, AVG(price) AS preço_médio
FROM propriedades_ofertas po
INNER JOIN propriedades_descricao pd
    ON po.id = pd.id
WHERE property_type IS NOT NULL
GROUP BY pd.property_type, po.cancellation_policy
ORDER BY property_type, cancellation_policy;
```

Property_type	cancellation_policy	preço_médio
Apartment	flexible	156.7043
Apartment	moderate	175.1422
Apartment	strict	190.8421
Apartment	super_strict_30	324.1411
Bed & Breakfast	flexible	77.70
Bed & Breakfast	moderate	92.60
Bed & Breakfast	strict	171.4761
Boat	flexible	345.00
Boat	moderate	229.00
Boat	strict	265.20
Camper/RV	moderate	60.00
Condominium	flexible	165.25
Condominium	moderate	161.0967
Condominium	strict	222.6952
Dorm	flexible	62.50
Entire Floor	flexible	100.00
Entire Floor	moderate	129.3333
Guesthouse	strict	289.00
House	flexible	115.2252
House	moderate	114.9865
House	strict	155.2164
Loft	flexible	170.10
Loft	moderate	157.5384
Loft	strict	217.625
Other	flexible	280.00
Other	moderate	400.00
Other	strict	205.3076
Townhouse	flexible	117.00
Townhouse	moderate	134.50
Townhouse	strict	177.4166
Villa	flexible	179.40
Villa	moderate	240.00

 -- Classifique as propriedades por faixa de preço e mostre a quantidade de casas por faixa de preço

```
WITH faixa_preço_count AS
(
SELECT id,
```

```

CASE
    WHEN price <= 100
        THEN 'Preço Baixíssimo'
    WHEN price > 100 AND price <= 200
        THEN 'Preço Baixo'
    WHEN price > 200 AND price <= 500
        THEN 'Preço Moderado'
    WHEN price > 500 AND price <= 1000
        THEN 'Preço Alto'
    WHEN price > 1000
        THEN 'Preço Altíssimo'
END faixa_preço
FROM propriedades_ofertas
)

```

```

SELECT faixa_preço, COUNT(faixa_preço) AS quantidade
FROM faixa_preço_count
GROUP BY faixa_preço
ORDER BY quantidade DESC;

```

Faixa_preço	quantidade
Preço Baixo	1337
Preço Baixíssimo	1252
Preço Moderado	930
Preço Alto	56
Preço Altíssimo	10

-- Número de propriedades por tipo

```

SELECT property_type, COUNT(property_type) AS quantidade_tipo
FROM propriedades_descricao
GROUP BY property_type
ORDER BY quantidade_tipo DESC;

```

property_type	quantidade_tipo
Apartment	2612
House	562
Condominium	231
Townhouse	54
Bed & Breakfast	41
Loft	39
Other	17
Boat	12
Villa	6
Entire Floor	4
Dorm	2
Guesthouse	1
Camper/RV	1
NULL	0

-- Média de acomodações, banheiros, quartos e camas por tipo de propriedade e faixa de preço

```

WITH faixa_preço_count AS
(
    SELECT id,
        CASE
            WHEN price <= 100

```

```

        THEN 'Preço Baixíssimo'
    WHEN price > 100 AND price <= 200
        THEN 'Preço Baixo'
    WHEN price > 200 AND price <= 500
        THEN 'Preço Moderado'
    WHEN price > 500 AND price <= 1000
        THEN 'Preço Alto'
    WHEN price > 1000
        THEN 'Preço Altíssimo'
    END faixa_preço
FROM propriedades_ofertas
)
SELECT property_type, fc.faixa_preço, ROUND(AVG(pd.accommodates), 2) acomodações,
       ROUND(AVG(pd.bathrooms),2) banheiros, ROUND(AVG(pd.bedrooms),2) quartos,
       ROUND(AVG(pd.beds),2) camas
FROM propriedades_descricao pd
INNER JOIN faixa_preço_count fc
    ON pd.id = fc.id
WHERE property_type IS NOT NULL
GROUP BY property_type, faixa_preço
ORDER BY property_type, faixa_preço;

```

property_type	faixa_preço	acomodações	banheiros	quartos	camas
Apartment	Preço Altíssimo	3	1.5	1.33	1.67
Apartment	Preço Alto	5.69	1.81	2.51	3.03
Apartment	Preço Baixíssimo	1.9	1.07	0.98	1.16
Apartment	Preço Baixo	3.08	1.1	1.08	1.5
Apartment	Preço Moderado	4.11	1.31	1.63	2.03
Bed & Breakfast	Preço Baixíssimo	2.23	1.29	1	1.24
Bed & Breakfast	Preço Baixo	2.08	1.13	1.08	1.23
Bed & Breakfast	Preço Moderado	2.33	1	1	1.33
Boat	Preço Baixo	5.67	1.33	1.67	4.33
Boat	Preço Moderado	5.89	1.33	1.89	3.67
Camper/RV	Preço Baixíssimo	7	0	0	3
Condominium	Preço Alto	4.63	1.75	2	2.63
Condominium	Preço Baixíssimo	2.16	1.12	1.18	1.25
Condominium	Preço Baixo	2.99	1.14	1.1	1.51
Condominium	Preço Moderado	4.25	1.39	1.84	2.24
Dorm	Preço Baixíssimo	2.5	1	1	1
Entire Floor	Preço Baixíssimo	2	1	1	1.5
Entire Floor	Preço Baixo	6	1.5	2	3
Guesthouse	Preço Moderado	4	1	2	2
House	Preço Altíssimo	8	4.5	5	6
House	Preço Alto	7.73	2.59	3.55	4.64

House	Preço Baixíssimo	1.98	1.34	1.01	1.21
House	Preço Baixo	3.45	1.32	1.46	1.98
House	Preço Moderado	6.44	1.99	2.82	3.76
Loft	Preço Baixíssimo	2.1	1.05	0.9	1.2
Loft	Preço Baixo	2.93	1.1	0.87	1.33
Loft	Preço Moderado	3.36	1.36	1.5	1.79
Other	Preço Baixíssimo	1.67	3.67	1	1
Other	Preço Baixo	2	5.6	1	1
Other	Preço Moderado	2.56	1.22	1	1.33
Townhouse	Preço Alto	5.5	2.5	3	3.5
Townhouse	Preço Baixíssimo	1.82	1.23	0.96	1.18
Townhouse	Preço Baixo	2.56	1.28	0.89	1.39
Townhouse	Preço Moderado	4.83	1.83	2.17	2.67
Villa	Preço Baixíssimo	2.33	1.5	1	1
Villa	Preço Moderado	4	1	1	1.67

 -- Número de propriedades por rua/avenida

```
SELECT SUBSTRING(street, 1, CHARINDEX(',', street) - 1),
       COUNT(id) propriedades_quantidade
FROM geolocalizacao
WHERE CHARINDEX(',', street) - 1 > 0
GROUP BY street
ORDER BY propriedades_quantidade DESC;
```

Obs: tabela original possui 1.229 linhas

avenida_rua	propriedades_quantidade
Boylston Street	64
Beacon Street 50	
Garrison Street	44
Brookline Avenue	41
Columbus Avenue	31
Boston 30	
Commonwealth Avenue	30
Massachusetts Avenue	30
Seaport Boulevard	28
Commonwealth Avenue	28
Newbury Street	26
Washington Street	24
Charles Street	24
Emerson Place 22	
Harrison Avenue	22
Tremont Street	22
Tremont Street	21
Washington Street	21

Marlborough Street 21
Park Drive 19

```
-----  
-----  
-- Preço médio dos tipos de propriedades em cada rua/avenida  
  
-- DROP TABLE #id_rua;  
CREATE TABLE #id_rua (id NUMERIC, rua VARCHAR(MAX));  
  
INSERT INTO #id_rua  
SELECT id, SUBSTRING(street, 1, CHARINDEX(',', street) - 1)  
FROM geolocalizacao  
WHERE CHARINDEX(',', street) - 1 > 0;  
  
SELECT * FROM #id_rua;  
  
SELECT rua, pd.property_type, AVG(price) AS preço_médio  
FROM propriedades_ofertas po  
INNER JOIN propriedades_descricao pd  
    ON po.id = pd.id  
INNER JOIN #id_rua ir  
    ON po.id = ir.id  
GROUP BY ir.rua, pd.property_type  
ORDER BY ir.rua;
```

Obs: tabela original possui 1.367 linhas

rua	property_type	preço_médio
13th St	Apartment	225
13th Street	Apartment	149
13th Street	Boat	229
13th Street	Condominium	234
1st Avenue	Apartment	236
1st Avenue	Condominium	150
8th Street	Apartment	1000
A Garden St	Apartment	229
A Mount Auburn Street	House	85
A St	Loft	170
A Street	Apartment	305.2857
A Street	Condominium	390
Aberdeen Street	Apartment	98.3333
Academy Hill Rd	House	70
Adair Road	House	55
Adams Street	Apartment	67.5
Adams Street	House	176.6666
Adams Terrace	Condominium	150
Adamson Street	Condominium	120
æ³¢å£«éj¿	House	100

-- Os 20 proprietários com mais reviews e suas notas

```
SELECT TOP 20 p.host_id, SUM(pr.number_of_reviews) reviews_contagem,
AVG(pr.review_scores_rating) AS score_médio
FROM proprietarios p
LEFT JOIN propriedades_descricao pd
    ON pd.host_id = p.host_id
INNER JOIN propriedades_reviews pr
    ON pd.id = pr.id
GROUP BY p.host_id
ORDER BY reviews_contagem DESC, score_médio;
```

host_id	reviews_contagem	score_médio
324630	1448	76.71429
3637081	1333	90.5
6608084	1223	97.14286
119019	1179	87.22222
1407005	904	86.75
9410008	778	97.73333
1480518	772	97
85770	761	88.4
1444340	668	87
8139434	629	94.9
28820575	555	89.09091
25188	551	88.64063
22264304	535	88.66667
1651480	527	96.4
9663343	498	84.33333
68870	464	88
6103425	431	88.6
5618949	423	79.72727
5578043	404	88.25
12243051	390	91.25641

-- Média dos scores por tipo de propriedade

```
SELECT pd.property_type, AVG(pr.review_scores_rating) AS scores_média
FROM propriedades_descricao pd
INNER JOIN propriedades_reviews pr
    ON pd.id = pr.id
GROUP BY pd.property_type
ORDER BY scores_média DESC;
```

property_type	scores_média
Entire Floor	100
Guesthouse	100
Villa	99.5
Loft	96.3125
Townhouse	94.7608695652174
Condominium	94.4519774011299
House	92.2045929018789

Bed & Breakfast	91.8
Apartment	91.5141414141414
Boat	91.4545454545455
Other	88.1111111111111
Dorm	86.5
NULL	72.5
Camper/RV	NULL

-- Quais ruas/avenidas receberam os melhores notas de reviews?

```
SELECT g.street, AVG(pr.review_scores_location) AS location_review
FROM geolocalizacao g
INNER JOIN propriedades_reviews pr
ON g.id = pr.id
WHERE pr.review_scores_location IS NOT NULL
GROUP BY g.street
ORDER BY location_review DESC;
```

Obs: tabela original possui 1.073 linhas

street	location_review
, MA 02467, United States	10
1st Avenue, Boston, MA 02129, United States	10
A Garden St, Boston, MA 02114, United States	10
A Mount Auburn Street, Cambridge, MA 02138, United States	10
Adams Street, Boston, MA 02122, United States	10
Agassiz Park, Boston, MA 02130, United States	10
Alaric Street, Boston, MA 02132, United States	10
Albemarle Street, Boston, MA 02115, United States	10
Aldworth St, Boston, MA 02130, United States	10
Alleyne Street, Boston, MA 02132, United States	10
Amory Street, Boston, MA 02130, United States	10
Anderson Pl, Boston, MA 02114, United States	10
Anderson St, Boston, MA 02114, United States	10
Anderson Street, Boston, MA 02114, United States	10
Appleton St, Boston, MA 02116, United States	10
Alta Crest Road, Boston, MA 02132, United States	10
Archstone Avenir, 101 Canal Street, Boston, MA 02114, United States	10
Arlington Street, Boston, MA 02116, United States	10
Atlantic Ave, Boston, MA 02109, United States	10
Atlantic Avenue, Boston, MA 02110, United States	10