

```

1
%sql
-- Qual é o perfil predominante dos animais que chegam ao abrigo em termos de espécie, idade e condição?
-- Respostas: As principais espécies que chegam ao abrigo são cachorros e gatos. Em sua maioria, cachorros maiores que uma ano. A condição não parece influenciar.

SELECT
    da.especie,
    f.idade_entrada,
    dc.condicao_entrada,
    COUNT(*) AS qtd_animais
FROM fato_entrada f
JOIN dim_animal da
    ON f.id.dim_animal = da.id.dim_animal
JOIN dim_condicao dc
    ON f.id.dim_condicao = dc.id.dim_condicao
GROUP BY
    da.especie,
    f.idade_entrada,
    dc.condicao_entrada
ORDER BY qtd_animais DESC
LIMIT 10

```

> \_sqldf: pyspark.sql.connect.DataFrame = [especie: string, idade\_entrada: string ... 2 more fields]

Table				
	especie	idade_entrada	condicao_entrada	qtd_animais
1	Dog	1 year	Normal	8488
2	Dog	2 years	Normal	7704
3	Cat	1 month	Normal	4576
4	Dog	3 years	Normal	3678
5	Cat	1 year	Normal	2857
6	Dog	4 years	Normal	2270
7	Dog	1 month	Normal	2234
8	Dog	5 years	Normal	1999
9	Cat	2 years	Normal	1996
10	Cat	2 months	Normal	1937

```

2
%sql
-- Qual é o perfil predominante dos animais que saem do abrigo em termos de espécie, idade, sexo e castração?
-- Resposta: Cães são as principais espécies que chegam aos abrigos. Em sua maioria, jovens (até um ano) e castrados.

SELECT
    das.especie,
    das.castrado,
    fs.idade_saida,
    dtis.tipo_saida,
    COUNT(*) AS qtd_animais
FROM fato_saida fs
JOIN dim_animal_saida das
    ON fs.id.dim_animal_saida = das.id.dim_animal_saida
JOIN dim_tipo_saida dtis
    ON fs.id.dim_tipo_saida = dtis.id.dim_tipo_saida
GROUP BY
    das.especie,
    das.castrado,
    fs.idade_saida,
    dtis.tipo_saida
ORDER BY qtd_animais DESC
LIMIT 10

```

> \_sqldf: pyspark.sql.connect.DataFrame = [especie: string, castrado: string ... 3 more fields]

Table					
	especie	castrado	idade_saida	tipo_saida	qtd_animais
1	Dog	Yes	1 year	Adoption	4684
2	Cat	Yes	2 months	Adoption	4636
3	Dog	Yes	2 years	Adoption	3532
4	Dog	Yes	2 months	Adoption	2453
5	Dog	Yes	2 years	Return to Owner	1933
6	Cat	Yes	3 months	Adoption	1716
7	Dog	Yes	1 year	Return to Owner	1603
8	Dog	Yes	3 years	Adoption	1483
9	Dog	Yes	3 years	Return to Owner	1213
10	Dog	Yes	1 year	Transfer	1109

↓ 10 rows

3

```
%sql
-- Qual é a distribuição dos principais tipos de entradas dos animais?
-- Principais são cachorros de ruas e devolvidos pelo dono. Devolvidos pelo dono é um caso crítico quando pensamos em adoções. Vamos analisar melhor.

SELECT
    dti.tipo_entrada,
    COUNT(*) AS qtd
FROM foto_entrada f
JOIN dim_tipo dti
    ON f.id_dim_tipo = dti.id_dim_tipo
GROUP BY dti.tipo_entrada
ORDER BY qtd DESC
```

> \_sqldf: pyspark.sql.connect.DataFrame [tipo\_entrada: string, qtd: long]

	tipo_entrada	qtd
1	Stray	52789
2	Owner Surrender	14744
3	Public Assist	4788
4	Euthanasia Request	232
5	Wildlife	2

5 rows

4

```
%sql
-- Perfil predominante dos animais devolvidos pelo dono?
-- O perfil predominante dos animais devolvidos são cachorros que já passaram pela fase de filhote (animais maiores que 1 ano)

SELECT
    da.especie,
    f.idade_entrada,
    COUNT(*) AS qtd_animais
FROM foto_entrada f
JOIN dim_animal da
    ON f.id_dim_animal = da.id_dim_animal
JOIN dim_tipo dti
    ON f.id_dim_tipo = dti.id_dim_tipo
WHERE dti.tipo_entrada = 'Owner Surrender'
GROUP BY
    da.especie,
    f.idade_entrada
ORDER BY qtd_animais DESC
LIMIT 10
```

> \_sqldf: pyspark.sql.connect.DataFrame [especie: string, idade\_entrada: string ... 1 more field]

	especie	idade_entrada	qtd_animais
1	Dog	1 year	1679
2	Dog	2 years	1374
3	Dog	3 years	796
4	Cat	1 year	713
5	Cat	1 month	634
6	Cat	2 months	565
7	Dog	4 years	526
8	Cat	2 years	507
9	Dog	5 years	469
10	Dog	1 month	469

10 rows

```
%sql
-- Os animais devolvidos, em sua maioria, estão normais.

SELECT
    dc.condicao_entrada,
    COUNT(*) AS qtd_animais
FROM fato_entrada f
JOIN dim_animal da
    ON f.id_dim_animal = da.id_dim_animal
JOIN dim_condicao dc
    ON f.id_dim_condicao = dc.id_dim_condicao
JOIN dim_tipo dti
    ON f.id_dim_tipo = dti.id_dim_tipo
WHERE dti.tipo_entrada = 'Owner Surrender'
GROUP BY
    dc.condicao_entrada
ORDER BY qtd_animais DESC
LIMIT 10
```

> \_sqldf: pyspark.sql.connect.DataFrame = [condicao\_entrada: string, qtd\_animais: long]

Table					
	A <sup>0</sup> c condicao_entrada	I <sup>0</sup> g qtd_animais			
1	Normal	13830			
2	Sick	461			
3	Injured	259			
4	Nursing	110			
5	Aged	67			
6	Feral	9			
7	Pregnant	8			

↓ 7 rows

```
%sql
-- Qual é a distribuição dos principais desfechos dos animais (adoção, retorno, transferência, eutanásia)?

SELECT
    dtis.tipo_saida,
    COUNT(*) AS qtd
FROM fato_saida fs
JOIN dim_tipo_saida dtis
    ON fs.id_dim_tipo_saida = dtis.id_dim_tipo_saida
GROUP BY dtis.tipo_saida
order by qtd DESC
```

> \_sqldf: pyspark.sql.connect.DataFrame = [tipo\_saida: string, qtd: long]

Table					
	A <sup>0</sup> c tipo_saida	I <sup>0</sup> g qtd			
1	Adoption	33944			
2	Transfer	20988			
3	Return to Owner	14920			
4	Euthanasia	2803			
5	Died	480			
6	Missing	45			

↓ 6 rows

7

```
%sql
-- Como o tempo médio de permanência no abrigo varia de acordo com espécie, raça e idade?
-- Especie

SELECT
    da.especie,
    round(AVG(DATEDIFF(dts.data_saida,dt.data_entrada)),0) AS tempo_medio_dias
FROM fato_entrada f
JOIN fato_saida fs
ON f.id_dim_animal = fs.id_dim_animal_saida
JOIN dim_tempo dt
ON f.id_dim_data = dt.id_dim_data
JOIN dim_tempo_saida dts
ON fs.id_dim_data_saida = dts.id_dim_data_saida
JOIN dim_animal da
ON f.id_dim_animal = da.id_dim_animal
WHERE dts.data_saida >= dt.data_entrada
GROUP BY
    da.especie
ORDER BY tempo_medio_dias DESC
```

&gt; \_sqldf: pyspark.sql.connect.DataFrame = [especie: string, tempo\_medio\_dias: double]

Table		
	A <sup>B</sup> c especie	1.2 tempo_medio_dias
1	Livestock	994
2	Dog	546
3	Cat	520
4	Bird	420

↓ 4 rows

8

```
%sql
-- Como o tempo médio de permanência no abrigo varia de acordo com espécie, raça e idade?
-- Raça

SELECT
    da.raca,
    round(AVG(DATEDIFF(dts.data_saida,dt.data_entrada)),0) AS tempo_medio_dias
FROM fato_entrada f
JOIN fato_saida fs
ON f.id_dim_animal = fs.id_dim_animal_saida
JOIN dim_tempo dt
ON f.id_dim_data = dt.id_dim_data
JOIN dim_tempo_saida dts
ON fs.id_dim_data_saida = dts.id_dim_data_saida
JOIN dim_animal da
ON f.id_dim_animal = da.id_dim_animal
WHERE dts.data_saida >= dt.data_entrada
AND da.especie = 'Livestock'
GROUP BY
    da.raca
ORDER BY tempo_medio_dias DESC
```

&gt; \_sqldf: pyspark.sql.connect.DataFrame = [raca: string, tempo\_medio\_dias: double]

Table		
	A <sup>B</sup> c raca	1.2 tempo_medio_dias
1	Pig Mix	1061
2	Goat Mix	860

↓ 2 rows

9

```
%sql
-- Como o tempo médio de permanência no abrigo varia de acordo com espécie, raça e idade?
| -- Raça
SELECT
| da.raça,
| round(AVG(DATEDIFF(dts.data_saida,dt.data_entrada)),0) AS tempo_medio_dias
FROM fato_entrada f
JOIN fato_saida fs
ON f.id_dim_animal = fs.id_dim_animal_saida
JOIN dim_tempo dt
ON f.id_dim_data = dt.id_dim_data
JOIN dim_tempo_saida dts
ON fs.id_dim_data_saida = dts.id_dim_data_saida
JOIN dim_animal da
ON f.id_dim_animal = da.id_dim_animal
WHERE dts.data_saida >= dt.data_entrada
AND da.especie = 'Dog'
GROUP BY
| da.raça
ORDER BY tempo_medio_dias DESC
LIMIT 5
```

> `_sqldf: pyspark.sql.connect.DataFrame = [raça: string, tempo_medio_dias: double]`

Table	
A <sup>o</sup> raça	1.2 tempo_medio_dias
1 Border Collie/American Staffordshire Terrier	1590
2 Shetland Sheepdog/Irish Greyhound	1569
3 Border Collie/Anatolian Shepherd	1562
4 Australian Kelpie/Shiba Inu	1539
5 Plott Hound/Carolina Dog	1515

↓ 5 rows

10

```
%sql
-- Como o tempo médio de permanência no abrigo varia de acordo com espécie, raça e idade?
| -- Raça
SELECT
| da.raça,
| round(AVG(DATEDIFF(dts.data_saida,dt.data_entrada)),0) AS tempo_medio_dias
FROM fato_entrada f
JOIN fato_saida fs
ON f.id_dim_animal = fs.id_dim_animal_saida
JOIN dim_tempo dt
ON f.id_dim_data = dt.id_dim_data
JOIN dim_tempo_saida dts
ON fs.id_dim_data_saida = dts.id_dim_data_saida
JOIN dim_animal da
ON f.id_dim_animal = da.id_dim_animal
WHERE dts.data_saida >= dt.data_entrada
AND da.especie = 'Cat'
GROUP BY
| da.raça
ORDER BY tempo_medio_dias DESC
LIMIT 5
```

> `_sqldf: pyspark.sql.connect.DataFrame = [raça: string, tempo_medio_dias: double]`

Table	
A <sup>o</sup> raça	1.2 tempo_medio_dias
1 Devon Rex	1253
2 Snowshoe/Ragdoll	1053
3 American Wirehair ...	1042
4 Bengal	995
5 Bengal Mix	944

↓ 5 rows

11

```
%sql
-- Como o tempo médio de permanência no abrigo varia de acordo com espécie, raça e idade?
-- Raça
SELECT
    da.raca,
    round(AVG(DATEDIFF(dts.data_saida,dt.data_entrada)),0) AS tempo_medio_dias
FROM fato_entrada f
JOIN fato_saida fs
    ON f.id_dim_animal = fs.id_dim_animal_saida
JOIN dim_tempo dt
    ON f.id_dim_data = dt.id_dim_data
JOIN dimtempo_saida dts
    ON fs.id_dim_data_saida = dts.id_dim_data_saida
JOIN dim_animal da
    ON f.id_dim_animal = da.id_dim_animal
WHERE dts.data_saida >= dt.data_entrada
AND da.especie = 'Bird'
GROUP BY
    da.raca
ORDER BY tempo_medio_dias DESC
LIMIT 5
```

> \_sqldf: pyspark.sql.connect.DataFrame = [raca: string, tempo\_medio\_dias: double]

Table		
	A% raca	1.2 tempo_medio_dias
1	Rhode Island ...	1415
2	Parrot Mix	1159
3	Chicken	957
4	Parakeet Mix	743
5	Silkie Mix	457

↓ 5 rows

12

```
%sql
-- Como o tempo médio de permanência no abrigo varia de acordo com espécie, raça e idade?
-- Animais muito jovens não podem sair do abrigo pois precisam de cuidados especiais até atingir certa idade para adoção.
-- Animais mais velhos tendem a ficar mais tempo nos abrigos sem serem adotados.

SELECT
    f.idade_entrada,
    round(AVG(DATEDIFF(dts.data_saida,dt.data_entrada)),0) AS tempo_medio_dias
FROM fato_entrada f
JOIN fato_saida fs
    ON f.id_dim_animal = fs.id_dim_animal_saida
JOIN dim_tempo dt
    ON f.id_dim_data = dt.id_dim_data
JOIN dimtempo_saida dts
    ON fs.id_dim_data_saida = dts.id_dim_data_saida
JOIN dim_animal da
    ON f.id_dim_animal = da.id_dim_animal
WHERE dts.data_saida >= dt.data_entrada
GROUP BY
    f.idade_entrada
ORDER BY tempo_medio_dias DESC
LIMIT 10
```

> \_sqldf: pyspark.sql.connect.DataFrame = [idade\_entrada: string, tempo\_medio\_dias: double]

Table		
	A% idade_entrada	1.2 tempo_medio_dias
1	6 days	652
2	22 years	589
3	9 months	587
4	20 years	586
5	14 years	583
6	11 months	567
7	3 years	559
8	7 years	557
9	13 years	557
10	15 years	555

↓ 10 rows

```
%sql
-- Existem padrões sazonais no volume de entradas e adoções ao longo do ano?
--- Entradas
SELECT
    MONTH(dt.data_entrada) AS mes_entrada,
    COUNT(*) AS qtd_animais
FROM fato_entrada f
JOIN dim_tempo dt
    ON f.id_dim_data = dt.id_dim_data
GROUP BY
    MONTH(dt.data_entrada)
ORDER BY
    qtd_animais DESC
LIMIT 3;
```

> \_sqldf: pyspark.sql.connect.DataFrame = [mes\_entrada: integer, qtd\_animais: long]

Table ▾

	mes_entrada	qtd_animais
1	5	7141
2	10	6991
3	6	6701

↓ 3 rows



ⓘ This result is stored as \_sqldf and can be used in other Python and SQL cells.

```
%sql
-- Existem padrões sazonais no volume de entradas e adoções ao longo do ano?
--- Saidas: Datas especiais (dia das crianças, ferias e natal)
SELECT
    MONTH(dts.data_saida) AS mes_saida,
    COUNT(*) AS qtd_animais
FROM fato_saida fs
JOIN dim_tempo_saida dts
    ON fs.id_dim_data_saida = dts.id_dim_data_saida
GROUP BY
    MONTH(dts.data_saida)
ORDER BY
    qtd_animais DESC
LIMIT 3;
```

> \_sqldf: pyspark.sql.connect.DataFrame = [mes\_saida: integer, qtd\_animais: long]

Table ▾

	mes_saida	qtd_animais
1	10	7107
2	7	6835
3	12	6630

↓ 3 rows



ⓘ This result is stored as \_sqldf and can be used in other Python and SQL cells.