

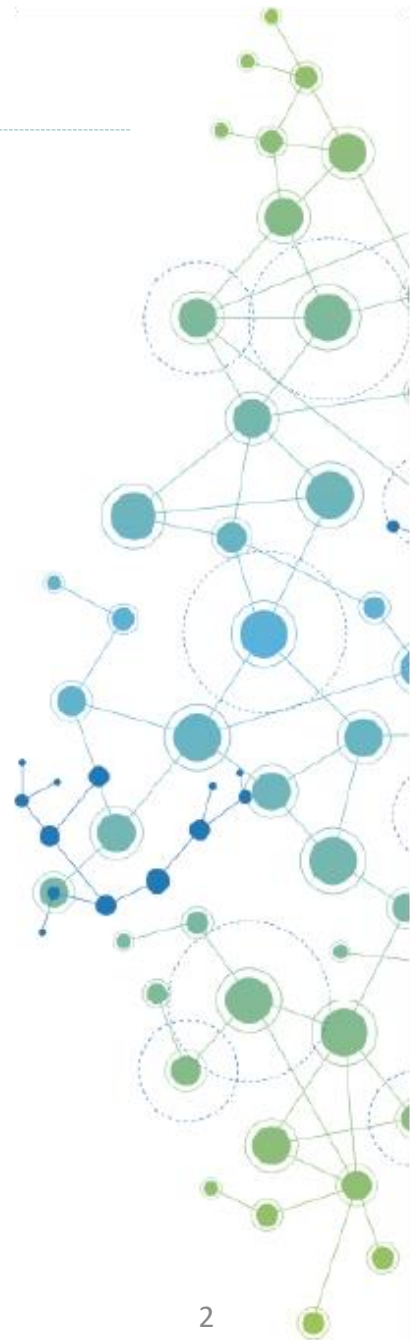


Preditiva.ai

SQL - Funções importantes e Tratamento de Dados **SUBQUERY** e Tabelas Temporárias

SQL – Funções Importantes

SUBQUERY



Extração de Dados



Predictiva.ai

Funções importantes para a Análise de Dados: **SUBQUERY**

Qual a taxa de sobrevivência entre as variáveis: classe, sexo e idade?

Para responder essa questão, o primeiro passo é fazer a análise exploratória de cada variável. Calcularemos a **frequência relativa** utilizando o recurso do SQL chamado de **subquery**. Com ele é possível encadear queries:

A contagem de observações resultantes dessa **subquery** é utilizada como denominador para o cálculo da frequência relativa.

```
SELECT
  Pclass,
  COUNT(passengerId) as n,
  CAST(COUNT(passengerId) AS FLOAT) / (SELECT COUNT(passengerId) FROM titanic) as freq_rel
FROM
  titanic
GROUP BY
  Pclass
ORDER BY
  Pclass
```

	123 Pclass	123 n	123 freq_rel
1	1	216	0,24
2	2	185	0,21
3	3	494	0,55

O comando **CAST** serve para converter uma variável para o tipo desejado. Neste caso, queremos converter o número inteiro para um com casas decimais (FLOAT).

Extração de Dados



Funções importantes para a Análise de Dados: SUBQUERY

Um dado interessante é a taxa de sobrevivência de cada classe. Para isso, vamos utilizar o fato da variável **Survived** ser binária, calculando a sua média para ter a proporção de sobreviventes:

```
SELECT
    Pclass,
    COUNT(passengerId) as n,
    CAST(COUNT(passengerId) AS FLOAT) / (SELECT COUNT(passengerId) FROM titanic) as freq_rel,
    AVG(CAST(Survived AS FLOAT)) as taxa_sobrevivencia
FROM
    titanic
GROUP BY
    Pclass
ORDER BY
    Pclass
```

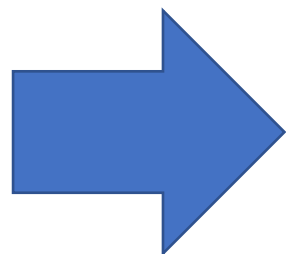
	123 Pclass	123 n	123 freq_rel	123 taxa_sobrevivencia
1	1	216	0,24	0,63
2	2	185	0,21	0,48
3	3	494	0,55	0,24

Extração de Dados

Funções importantes para a Análise de Dados: **SUBQUERY**

Incluindo o sexo conseguimos extrair mais informações interessantes:

```
SELECT
  Pclass,
  Sex,
  COUNT(passengerId) as n,
  CAST(COUNT(passengerId) AS FLOAT) / (SELECT COUNT(passengerId) FROM titanic) as freq_rel,
  AVG(CAST(Survived AS FLOAT)) as taxa_sobrevivencia
FROM
  titanic
GROUP BY
  Pclass, Sex
ORDER BY
  Pclass, Sex
```



	123 Pclass	ABC Sex	123 n	123 freq_rel	123 taxa_sobrevivencia
1	1	female	94	0,11	0,97
2	1	male	122	0,14	0,37
3	2	female	77	0,09	0,92
4	2	male	108	0,12	0,16
5	3	female	144	0,16	0,5
6	3	male	350	0,39	0,13

Aqui temos a taxa de sobrevivência pela combinação Classe e Sexo.

Extração de Dados

Funções importantes para a Análise de Dados: **HAVING**

Podemos destacar os grupos com mais de 50% de sobrevivência utilizando o **HAVING**.

SELECT

```
Pclass,  
Sex,  
COUNT(passengerId) as n,  
CAST(COUNT(passengerId) AS FLOAT) / (SELECT COUNT(passengerId) FROM titanic) as freq_rel,  
AVG(CAST(Survived AS FLOAT)) as taxa_sobrevivencia
```

FROM

titanic

GROUP BY

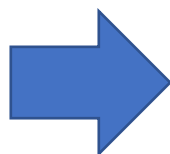
Pclass, Sex

HAVING

AVG(CAST(Survived AS FLOAT)) >= 0.5

ORDER BY

Pclass, Sex



	123 Pclass	ABC Sex	123 n	123 freq_rel	123 taxa_sobrevivencia
1	1	female	94	0,105	0,9681
2	2	female	77	0,086	0,9221
3	3	female	144	0,1609	0,5

HAVING: é utilizado para filtrar resultados com condições baseadas em funções de agregação.



Preditiva.ai