

Tabelas Temporárias



Extração de Dados

Funções importantes para a Análise de Dados: Tabelas Temporárias



Preditiva.ai

Vamos criar a tabela de frequência da variável Age, porém, por ser uma variável quantitativa, precisar criar faixas de valores, para isso vamos utilizar o CASE WHEN:

```
SELECT
*,
CASE WHEN Age IS NULL THEN NULL
      WHEN Age < 10 THEN 'A - Entre 0 e 9 anos'
      WHEN Age < 20 THEN 'B - Entre 10 e 19 anos'
      WHEN Age < 30 THEN 'C - Entre 20 e 29 anos'
      WHEN Age < 40 THEN 'D - Entre 30 e 39 anos'
      WHEN Age < 50 THEN 'E - Entre 40 e 49 anos'
      WHEN Age < 60 THEN 'F - Entre 50 e 59 anos'
      WHEN Age < 70 THEN 'G - Entre 60 e 69 anos'
      WHEN Age <= 100 THEN 'H - Acima de 70 anos'
      ELSE 'Z - Checar'
END AS Faixa_Etaria
FROM
titanic
```

	123 PassengerId	123 Survived	123 Pclass	ABC Name	ABC Sex	123 Age	ABC Faixa_Etaria
58	58	0	3	Novel, Mr. Mansouer	male	285	Z - Checar
59	59	1	2	West, Miss. Constance Mi	female	5	A - Entre 0 e 9 anos
60	60	0	3	Goodwin, Master. Willian	male	11	B - Entre 10 e 19 anos
61	61	0	3	Sirayanian, Mr. Orsen	male	22	C - Entre 20 e 29 anos
62	62	1	1	Icard, Miss. Amelie	female	38	D - Entre 30 e 39 anos
63	63	0	1	Harris, Mr. Henry Birkhar	male	45	E - Entre 40 e 49 anos
64	64	0	3	Skoog, Master. Harald	male	4	A - Entre 0 e 9 anos
65	65	0	1	Stewart, Mr. Albert A	male	[NULL]	[NULL]
66	66	1	3	Moubarek, Master. Gerio	male	[NULL]	[NULL]
67	67	1	2	Nye, Mrs. (Elizabeth Ram	female	29	C - Entre 20 e 29 anos

Extração de Dados

Funções importantes para a Análise de Dados: Tabelas Temporárias



Preditiva.ai

Porém vimos que precisamos criar uma variável chamada **Age_Tratada** para os casos acima de 100 anos. Vamos criar uma query adicionando essa variável aos campos de Titanic:

```
SELECT
    *,
    CASE WHEN Age > 100 THEN Age/10
    ELSE Age
    END AS Age_Tratada
FROM
    titanic
```

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked	Age_Tratada
1	1	0	3	Braund, Mr. Owen Harris	male	22	1	0	A/5 21171	725	[NULL]	S	22
2	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs)	female	38	1	0	PC 17599	712.833	C85	C	38
3	3	1	3	Heikkinen, Miss. Laina	female	26	0	0	STON/O2. 3101282	7.925	[NULL]	S	26
4	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	1	0	113803	531	C123	S	35
5	5	0	3	Allen, Mr. William Henry	male	35	0	0	373450	805	[NULL]	S	35
6	6	0	3	Moran, Mr. James	male	[NULL]	0	0	330877	84.583	[NULL]	Q	[NULL]
7	7	0	1	McCarthy, Mr. Timothy J	male	54	0	0	17463	518.625	E46	S	54
8	8	0	3	Palsson, Master. Gosta Leonard	male	2	3	1	349909	21.075	[NULL]	S	2
9	9	1	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina)	female	27	0	2	347742	111.333	[NULL]	S	27
10	10	1	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	1	0	237736	300.708	[NULL]	C	14

Extração de Dados

Funções importantes para a Análise de Dados: Tabelas Temporárias



Preditiva.ai

Utilizando **SUBQUERY** podemos incluir a variável Age_Tratada na análise, substituindo no **FROM** a query criada anteriormente, ao invés da base titanic (que não possui a variável Age_Tratada):

SELECT

```
*,
CASE WHEN Age_Tratada IS NULL THEN NULL
      WHEN Age_Tratada < 10 THEN 'A - Entre 0 e 9 anos'
      WHEN Age_Tratada < 20 THEN 'B - Entre 10 e 19 anos'
      WHEN Age_Tratada < 30 THEN 'C - Entre 20 e 29 anos'
      WHEN Age_Tratada < 40 THEN 'D - Entre 30 e 39 anos'
      WHEN Age_Tratada < 50 THEN 'E - Entre 40 e 49 anos'
      WHEN Age_Tratada < 60 THEN 'F - Entre 50 e 59 anos'
      WHEN Age_Tratada < 70 THEN 'G - Entre 60 e 69 anos'
      WHEN Age_Tratada <= 100 THEN 'H - Acima de 70 anos'
      ELSE 'Z - Checar'
END AS Faixa_Etaria_Tratada
```

FROM

```
(SELECT
  *,
  CASE WHEN Age > 100 THEN Age/10
        ELSE Age
  END AS Age_Tratada
FROM titanic) AS titanic_age_tratada
```

	PassengerId	Survived	Pclass	Name	Sex	Age	Age_Tratada	Faixa_Etaria_Tratada
58	58	0	3	Novel, Mr. Mansouer	male	285	28,5	C - Entre 20 e 29 anos
59	59	1	2	West, Miss. Constance Mi	female	5	5	A - Entre 0 e 9 anos
60	60	0	3	Goodwin, Master. Willian	male	11	11	B - Entre 10 e 19 anos
61	61	0	3	Sirayanian, Mr. Orsen	male	22	22	C - Entre 20 e 29 anos
62	62	1	1	Icard, Miss. Amelie	female	38	38	D - Entre 30 e 39 anos
63	63	0	1	Harris, Mr. Henry Birkhard	male	45	45	E - Entre 40 e 49 anos
64	64	0	3	Skoog, Master. Harald	male	4	4	A - Entre 0 e 9 anos
65	65	0	1	Stewart, Mr. Albert A	male	[NULL]	[NULL]	[NULL]
66	66	1	3	Moubarek, Master. Gerio	male	[NULL]	[NULL]	[NULL]

SELECT

```
*,
CASE WHEN Age > 100 THEN
  Age/10
      ELSE Age
END AS Age_Tratada
```

FROM

titanic

Extração de Dados

Funções importantes para a Análise de Dados: **Tabelas Temporárias**



Preditiva.ai

Podemos utilizar uma outra forma de realizar as queries em etapas, usando as **Tabelas Temporárias**. Elas são indicadas quando queremos deixar o código mais organizado, e também quando queremos criar uma variável/query que será utilizada em outras queries:

```
DROP TABLE IF EXISTS #titanic_age_tratada
```

```
SELECT
```

```
*,
```

```
CASE WHEN Age > 100 THEN Age/10
```

```
ELSE Age
```

```
END AS Age_Tratada
```

```
INTO
```

```
#titanic_age_tratada
```

```
FROM
```

```
titanic
```

Para utilizar novamente uma tabela temporária é necessário excluí-la antes:

```
DROP TABLE IF EXISTS #titanic_age_tratada
```

O comando INTO cria a tabela temporária, na qual definimos um nome iniciado com “#”

Extração de Dados

Funções importantes para a Análise de Dados: Tabelas Temporárias



Preditiva.ai

Através das tabelas temporárias, podemos criar queries em cadeia, dessa vez utilizando a variável **Age_Tratada** para criar a variável **Faixa_Etaria_Tratada**:

```
DROP TABLE IF EXISTS #titanic_faixa_etaria_tratada
```

```
SELECT
```

```
*,
```

```
  CASE WHEN Age_Tratada IS NULL THEN NULL
```

```
    WHEN Age_Tratada < 10 THEN 'A - Entre 0 e 9 anos'
```

```
    WHEN Age_Tratada < 20 THEN 'B - Entre 10 e 19 anos'
```

```
    WHEN Age_Tratada < 30 THEN 'C - Entre 20 e 29 anos'
```

```
    WHEN Age_Tratada < 40 THEN 'D - Entre 30 e 39 anos'
```

```
    WHEN Age_Tratada < 50 THEN 'E - Entre 40 e 49 anos'
```

```
    WHEN Age_Tratada < 60 THEN 'F - Entre 50 e 59 anos'
```

```
    WHEN Age_Tratada < 70 THEN 'G - Entre 60 e 69 anos'
```

```
    WHEN Age_Tratada <= 100 THEN 'H - Acima de 70 anos'
```

```
    ELSE 'Z - Checar'
```

```
  END AS Faixa_Etaria_Tratada
```

```
INTO
```

```
  #titanic_faixa_etaria_tratada
```

```
FROM
```

```
  #titanic_age_tratada
```

Estamos criando outra tabela temporária através da primeira que criamos

Extração de Dados

Funções importantes para a Análise de Dados: Tabelas Temporárias



Preditiva.ai

Agora conseguimos fazer a análise de maneira apropriada para a variável idade:

SELECT

```
Faixa_Etaria_Tratada,  
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_faixa_etaria_tratada
```

GROUP BY

```
Faixa_Etaria_Tratada
```

ORDER BY

```
Faixa_Etaria_Tratada
```

	Faixa_Etaria_Tratada	n	freq_rel	taxa_sobrevivencia
1	[NULL]	177	0,1978	0,2938
2	A - Entre 0 e 9 anos	55	0,0615	0,5636
3	B - Entre 10 e 19 anos	103	0,1151	0,4078
4	C - Entre 20 e 29 anos	223	0,2492	0,3453
5	D - Entre 30 e 39 anos	167	0,1866	0,4371
6	E - Entre 40 e 49 anos	90	0,1006	0,3889
7	F - Entre 50 e 59 anos	48	0,0536	0,4167
8	G - Acima de 60 anos	32	0,0358	0,4062



Preditiva.ai