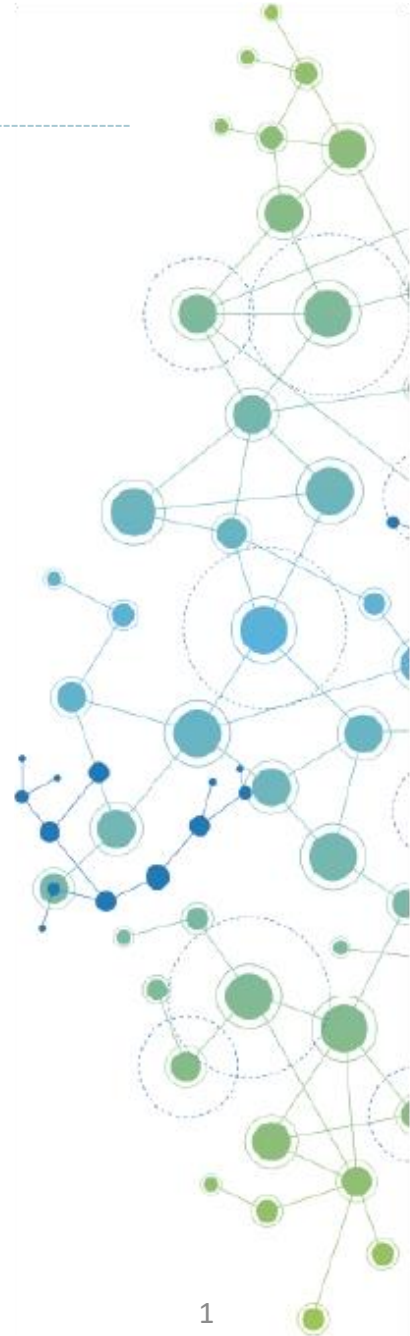


Transformações em campo Texto



Preparação de Dados

Transformações em campo Texto



Uma das primeiras transformações que fazemos quando começamos a trabalhar com dados em formato texto é converter todas as letras para minúsculo ou maiúsculo. No **SQL** temos a disposição as funções **LOWER** e **UPPER**.

```
SELECT
    t.Name,
    LOWER(t.Name) as nome_minusculas,
    UPPER(t.Name) as nome_maiusculas
FROM
    titanic t
```



Name	nome_minusculas	nome_maiusculas
Braund, Mr. Owen Harris	braund, mr. owen harris	BRAUND, MR. OWEN HARRIS
Cumings, Mrs. John Bradley (Florence Briggs Thayer)	cumings, mrs. john bradley (florence briggs thayer)	CUMINGS, MRS. JOHN BRADLEY (FLORENCE BRIG
Heikkinen, Miss. Laina	heikkinen, miss. laina	HEIKKINEN, MISS. LAINA
Futrelle, Mrs. Jacques Heath (Lily May Peel)	futrelle, mrs. jacques heath (lily may peel)	FUTRELLE, MRS. JACQUES HEATH (LILY MAY PEEL)
Allen, Mr. William Henry	allen, mr. william henry	ALLEN, MR. WILLIAM HENRY
Moran, Mr. James	moran, mr. james	MORAN, MR. JAMES
McCarthy, Mr. Timothy J	mccarthy, mr. timothy j	MCCARTHY, MR. TIMOTHY J
Palsson, Master. Gosta Leonard	palsson, master. gosta leonard	PALSSON, MASTER. GOSTA LEONARD

Preparação de Dados

Transformações em campo Texto



Anteriormente criamos a variável **Título** contendo parte da informação contida na variável **Name**, utilizando a função **CASE WHEN**. Uma forma **mais eficiente** é utilizando as funções específicas para tratamento de campos do tipo texto.

Podemos perceber que o título está localizado entre uma vírgula e um ponto final. Logo, o primeiro passo é obter a localização dos sinais de pontuação ",", " e ".". Para isso vamos utilizar a função **charindex**:

```
Drop table if exists #temp
select
    t.Name,
    CHARINDEX(',', t.Name) as pos_virgula,
    CHARINDEX('.', t.Name) as pos_ponto,
    CHARINDEX('.', t.Name) - CHARINDEX(',', t.Name) as diferenca
INTO #temp
from
    titanic t
```

	ABC Name	123 pos_virgula	123 pos_ponto	123 diferenca
1	Braund, Mr. Owen Harris	7	11	4
2	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	8	13	5
3	Heikkinen, Miss. Laina	10	16	6
4	Futrelle, Mrs. Jacques Heath (Lily May Peel)	9	14	5
5	Allen, Mr. William Henry	6	10	4
6	Moran, Mr. James	6	10	4
7	McCarthy, Mr. Timothy J	9	13	4
8	Palsson, Master. Gosta Leonard	8	16	8
9	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	8	13	5

Na função **charindex** indicamos qual o texto procurado e em qual variável. A função retorna a localização do texto procurado na variável.

Preparação de Dados


Transformações em campo Texto



Para extrair apenas o título que está entre essas duas posições vamos utilizar a função **SUBSTRING**. Nela devemos indicar o **nome da variável** queremos extrair uma parte do texto, a **posição inicial** e o **comprimento**.

```
select  
SUBSTRING(Name, pos_virgula, diferença)  
from  
#temp
```


1



	ABC
1	, Mr
2	, Mrs
3	, Miss
4	, Mrs
5	, Mr
6	, Mr

```
select  
SUBSTRING(Name, pos_virgula+2, diferença)  
from  
#temp
```

2



	ABC
1	Mr.
2	Mrs.
3	Miss.
4	Mrs.
5	Mr.
6	Mr.
7	Mr.
8	Master.

Preparação de Dados

Transformações em campo Texto



Outra função bastante interessante é a **CONCAT_WS**, que é uma variação da função **CONCAT**. Ambas têm o mesmo objetivo: **concatenar strings**. A diferença é que na 1ª você pode especificar um separador enquanto a 2ª concatena apenas as strings indicadas.

```
SELECT TOP 10
    t.Pclass,
    t.Sex,
    CONCAT_WS(' - ', t.Pclass, t.Sex) as pclass_sex
FROM
    titanic t
```

123 Pclass	ABC Sex	ABC pclass_sex
3.00	male	3 - male
1.00	female	1 - female
3.00	female	3 - female
1.00	female	1 - female
3.00	male	3 - male
3.00	male	3 - male
1.00	male	1 - male



Preditiva.ai