

# Big Data, Data Science e Machine Learning com Google Cloud Platform

Data Science Academy

# Capítulo 6

## Google BigQuery

### Parte 3



■ New Visitor ■ Returning Visitor



# Introdução

Data Wrangling com Python e Pandas para conversão de dataset com colunas repetidas aninhadas

Funções Analíticas e Window Frames

Google BigQuery Machine Learning

# Importando dados públicos com estrutura aninhada repetida no BigQuery

<http://www.tse.jus.br/eleicoes/estatisticas/repositorio-de-dados-eleitorais-1>

# Funções Analíticas

```
analytic_function_name ( [ argument_list ] )  
  OVER (  
    [ PARTITION BY partition_expression_list ]  
    [ ORDER BY expression [{ ASC | DESC }] [, ...] ]  
    [ window_frame_clause ]  
  )
```

Uma função analítica (também chamada de window function) calcula valores agregados em um grupo de linhas

Diferentemente de funções de agregação como max, min, sum, avg, count, etc., que retornar um único valor de agregação para cada GRUPO de linhas, uma função analítica retorna UM ÚNICO valor para para cada linha, calculando a função em um grupo de linhas de entrada. Por exemplo, cálculo do ranking de um aluno com base em sua idade

# Funções Analíticas

Uma função analítica requer a cláusula OVER para definir a “janela” em que a função analítica será avaliada

A cláusula OVER pode ser vazia, por exemplo, OVER(), indicando que a “janela” deverá incluir todas as linhas de entrada

A cláusula OVER também pode ter a cláusula PARTITION BY para dividir as linhas de entrada em partições e a cláusula ORDER BY para especificar a ordem das linhas dentro de cada partição

# Funções Analíticas

```
SELECT firstname, department, startdate,  
       RANK() OVER ( PARTITION BY department ORDER BY startdate ) AS rank  
FROM Employees;
```

PARTITION BY department

ORDER BY startdate

RANK()

firstname	department	startdate
Andrew	1	1/23/1999
Jacob	1	7/11/1990
Daniel	2	6/24/2004
Anna	1	10/7/2001
Pierre	1	2/22/2009
Ruth	2	6/6/1998
Anthony	1	11/29/1995
Isabella	2	9/28/1997
Jose	2	3/17/2013

firstname	department	startdate
Andrew	1	1/23/1999
Jacob	1	7/11/1990
Anna	1	10/7/2001
Pierre	1	2/22/2009
Anthony	1	11/29/1995

firstname	department	startdate
Jacob	1	7/11/1990
Anthony	1	11/29/1995
Andrew	1	1/23/1999
Anna	1	10/7/2001
Pierre	1	2/22/2009

firstname	department	startdate	rank
Jacob	1	7/11/1990	1
Anthony	1	11/29/1995	2
Andrew	1	1/23/1999	3
Anna	1	10/7/2001	4
Pierre	1	2/22/2009	5

firstname	department	startdate
Ruth	2	6/6/1998
Daniel	2	6/24/2004
Jose	2	3/17/2013
Isabella	2	9/28/1997

firstname	department	startdate
Isabella	2	9/28/1997
Daniel	2	6/24/2004
Jose	2	3/17/2013
Ruth	2	6/6/2013

firstname	department	startdate	rank
Isabella	2	9/28/1997	1
Daniel	2	6/24/2004	2
Jose	2	3/17/2013	3
Ruth	2	6/6/2013	4

# Window Frames

```
analytic_function_name ( [ argument_list ] )  
    OVER { window_name | ( [ window_specification ] ) }
```

```
window_specification:  
    [ window_name ]  
    [ PARTITION BY partition_expression_list ]  
    [ ORDER BY expression [{ ASC | DESC }] [, ...] ]  
    [ window_frame_clause ]
```

```
window_frame_clause:  
{ ROWS | RANGE }  
{  
    { UNBOUNDED PRECEDING | numeric_expression PRECEDING | CURRENT ROW }  
    |  
    { BETWEEN window_frame_boundary_start AND window_frame_boundary_end }  
}
```

```
window_frame_boundary_start:  
{ UNBOUNDED PRECEDING | numeric_expression { PRECEDING | FOLLOWING } | CURRENT ROW }
```

```
window_frame_boundary_end:  
{ UNBOUNDED FOLLOWING | numeric_expression { PRECEDING | FOLLOWING } | CURRENT ROW }
```



# Window Frames

Uma “window\_frame\_clause” define uma “window frame” na linha atual dentro de uma partição, sobre qual a função analítica é avaliada.

Os

Exemplos de cláusulas de frame de janela:

```
ROWS BETWEEN UNBOUNDED PRECEDING AND UNBOUNDED FOLLOWING
```

- inclui a partição inteira.
- exemplo de uso: cálculo de um total geral na partição.

```
ROWS BETWEEN UNBOUNDED PRECEDING AND CURRENT ROW
```

- inclui todas as linhas na partição antes da linha atual ou incluindo-a.
- exemplo de uso: cálculo de uma soma cumulativa.

```
ROWS BETWEEN 2 PRECEDING AND 2 FOLLOWING
```

- inclui todas as linhas entre duas antes e duas depois da linha atual.
- exemplo de uso: cálculo de uma média móvel.

# BigQuery

## User Defined Functions

# User Defined Functions

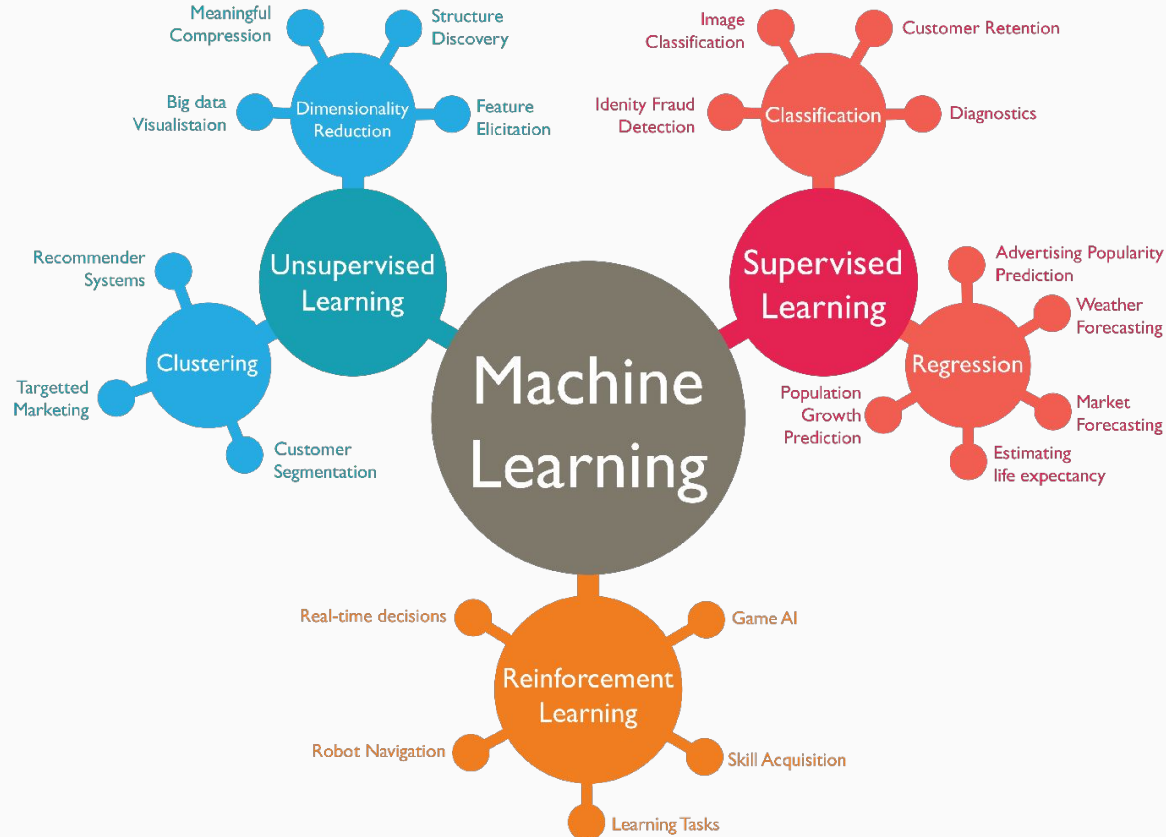
Permite a criação de funções usando expressões SQL ou linguagens de programação como JavaScript

UDF's são temporárias, assim você pode utilizá-las na consulta atual ou em uma mesma sessão via linha de comando

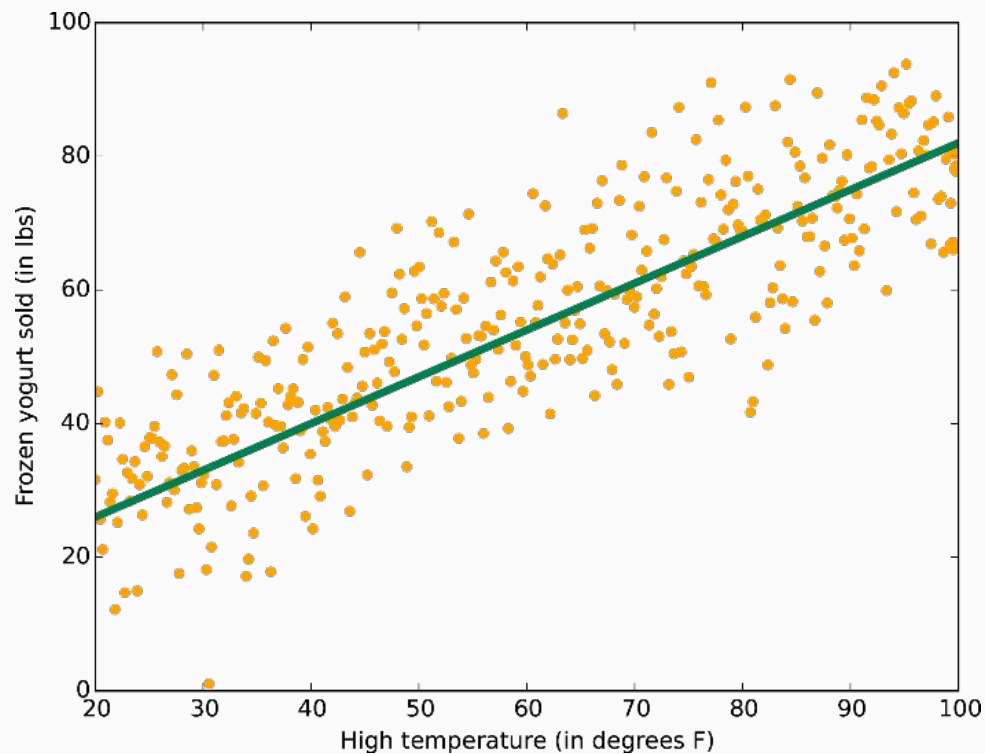
<https://cloud.google.com/bigquery/docs/reference/standard-sql/user-defined-functions>

# BigQuery Machine Learning

# Machine Learning - Divisão dos algoritmos

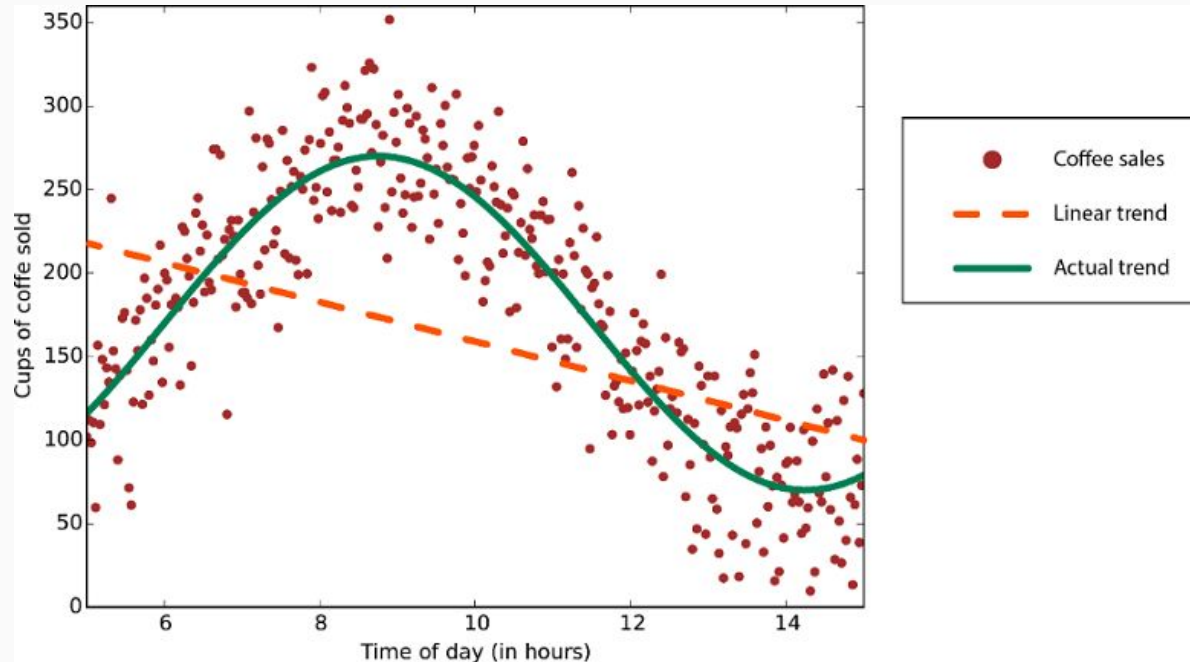


# Machine Learning - Regressão

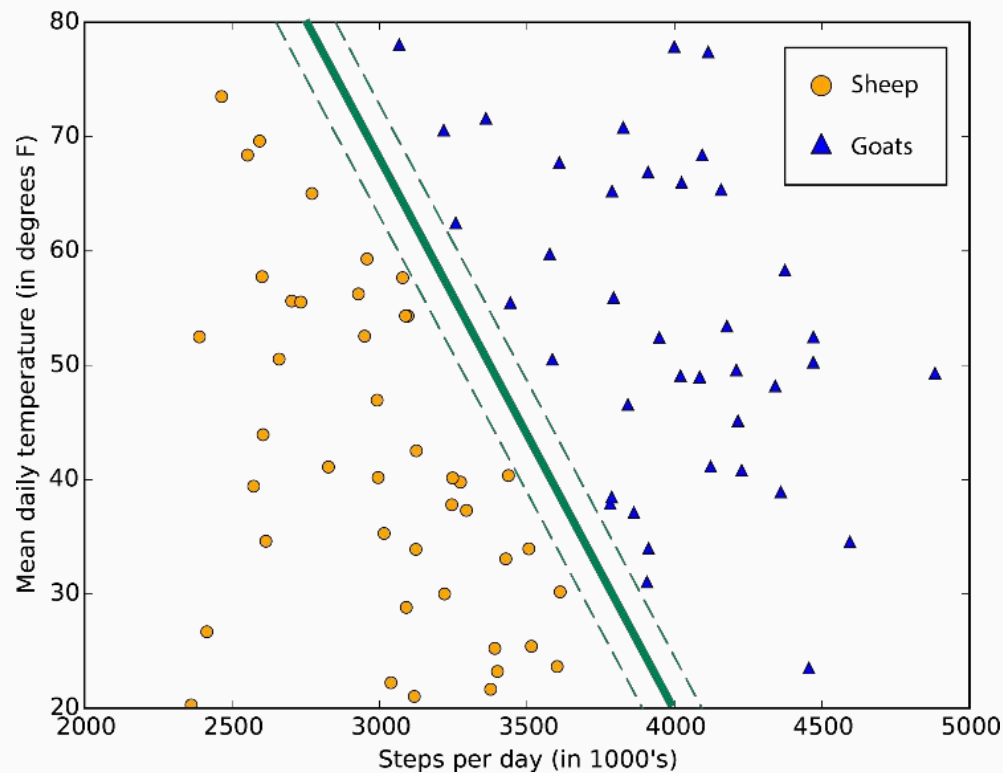


<https://docs.microsoft.com/en-us/azure/machine-learning/studio/algorithm-choice>

# Machine Learning - Regressão

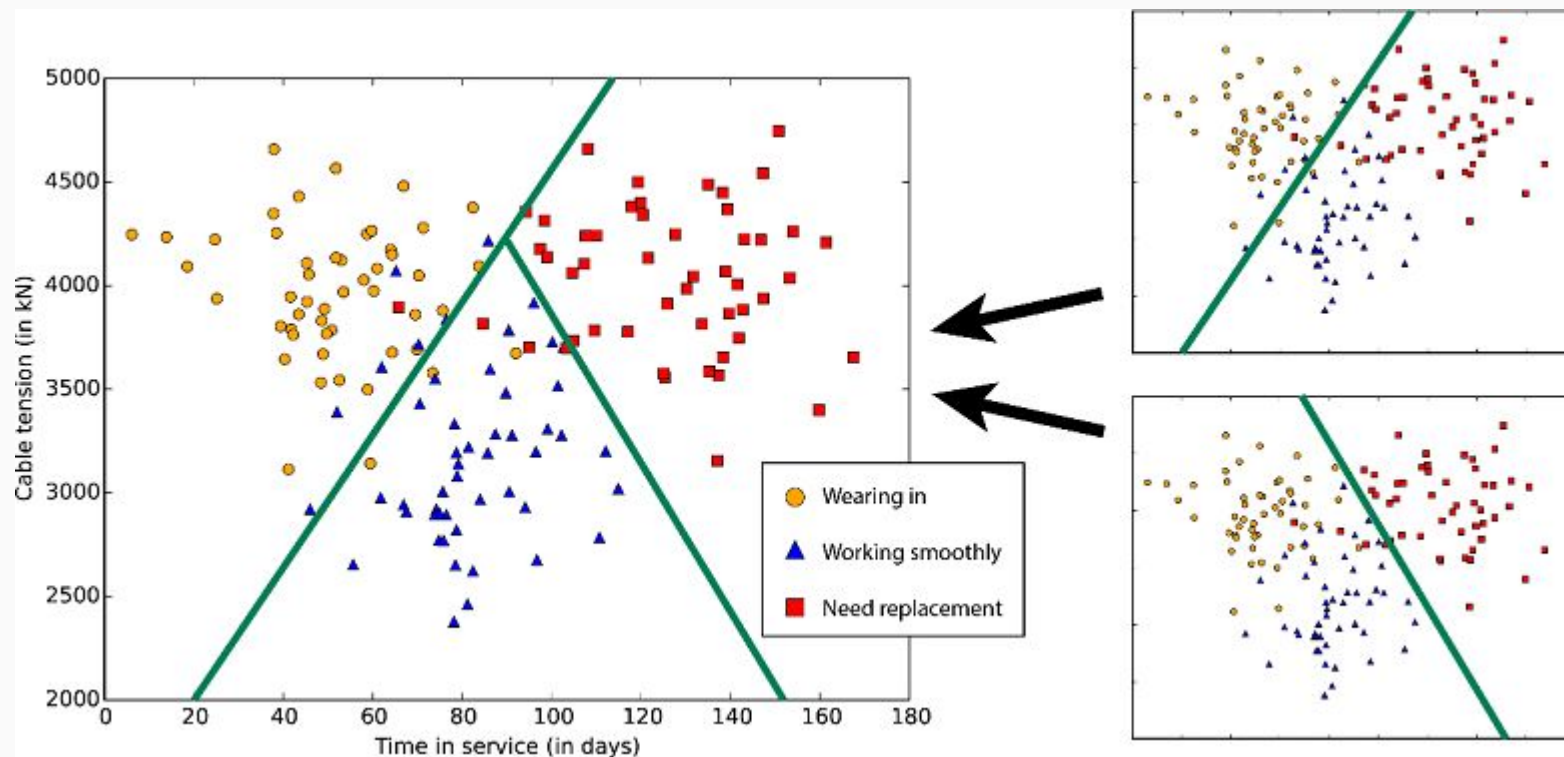


# Machine Learning - Classificação Binária





# Machine Learning - Classificação Binária



# BigQuery ML

Possibilita criação e execução de modelos de Machine Learning utilizando SQL

Algoritmos suportados:

- Regressão Linear
- Regressão Logística Binária
- Regressão Logística Multiclasse

# BigQuery ML

Disponível via Web UI, ferramenta de linha de comando “bq”, REST API e ferramentas externas como Jupyter Notebook

Vantagens:

- Criação de modelos de Machine Learning utilizando grandes volumes de dados
- Aceleração do desenvolvimento inclusive através de analistas de dados

# Os 7 passos de Machine Learning

<https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e>



1. Levantamento de dados
2. Preparação dos dados
3. Escolha do modelo
4. Treinamento
5. Avaliação
6. Tuning
7. Predição

# BigQuery ML

- Criação do modelo:  
<https://cloud.google.com/bigquery/docs/reference/standard-sql/bigqueryml-syntax-create>
- Inspeção do modelo:  
<https://cloud.google.com/bigquery/docs/reference/standard-sql/bigqueryml-syntax-train>

# BigQuery ML

- Avaliação do modelo:

<https://cloud.google.com/bigquery/docs/reference/standard-sql/bigquery-ml-syntax-evaluate>

- Predição:

<https://cloud.google.com/bigquery/docs/reference/standard-sql/bigquery-ml-syntax-predict>

# Obrigado

Data Science Academy