



escola
britânica de
artes criativas
& tecnologia

Módulo | SQL: Agregações

Caderno de **Exercícios**

Professor [Mariane Neiva](#)

▼ Tópicos

1. Função count e cláusula group by;
 2. Funções min / max / sum / avg;
 3. Cláusula having;
-

▼ Instruções

Neste exercício, vamos configurar e testar os serviços AWS S3 e AWS Athena da plataforma de computação em nuvem da AWS para utilizarmos durante todo o curso:

Na etapa de **configuração**, você deve seguir o passo a passo de configuração do ambiente, não há entregáveis.

Na etapa de **atividades**, você deve:

1. Executar a consulta SQL fornecida;
2. Exportar os resultados em um arquivo csv para sua máquina;
3. Renomear o arquivo como **query_<numero-da-query>.csv**;
4. Enviar para a avaliação do tutor na plataforma da EBAC.

Atenção: Substitua o **<numero-da-query>** número da consulta, exemplo:
query_1.csv.

Atenção: Os dados são baseados no dataset:

<https://www.kaggle.com/rashikrahmanpritom/heart-attack-analysis-prediction-dataset>

▼ Configuração

▼ 1. Configurando o AWS S3

1. Acesse AWS S3;
2. Crie um novo bucket:
 - **heart-attack-<seu-nome>-ebac**: para os dados;
3. Faça o upload do arquivo **heart.csv** no bucket de dados.

Atenção: Caso o bucket já exista, sinta-se a vontade para escolher o nome que desejar.

Atenção: O arquivo **heart.csv** está na plataforma da EBAC.

▼ Atividades

▼ 1. Criação da tabela

No console do AWS Athena, execute a seguinte query:

```
CREATE EXTERNAL TABLE IF NOT EXISTS default.heartattack (  
  `age` int,  
  `sex` int,  
  `cp` int,  
  `trtbps` int,  
  `chol` int,  
  `fbs` int,
```

```
`restecg` int,  
`thalachh` int,  
`exng` int,  
`oldpeak` double,  
`slp` int,  
`caa` int,  
`thall` int,  
`output` int  
)  
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'  
WITH SERDEPROPERTIES (  
  'serialization.format' = ',',  
  'field.delim' = ','  
) LOCATION 's3://heart-attack-<seu-nome>-ebac/'  
TBLPROPERTIES ('has_encrypted_data'='false');
```

No painel de resultados, você deve encontrar o seguinte resultado.

Query successful.

Atenção: No campo LOCATION, substitua o <seu-nome> pelo seu primeiro nome, exemplo: **heart-attack-mari-ebac**.

Atenção: Nesta atividade, você não precisa exportar os resultados.

▼ 2. Função COUNT e GROUP BY

▼ 2.1. Query 1

No console do AWS Athena, execute a seguinte query:

```
SELECT * FROM heartattack limit 10;
```

No painel de resultados você deve encontrar o seguinte resultado.

| | age ▾ | sex ▾ | cp ▾ | trtbps ▾ | chol ▾ | fbs ▾ | restecg ▾ | thalachh ▾ | exng ▾ | oldpeak ▾ | slp ▾ | caa ▾ | thall ▾ | output ▾ |
|----|-------|-------|------|----------|--------|-------|-----------|------------|--------|-----------|-------|-------|---------|----------|
| 1 | 63 | 1 | 3 | 145 | 233 | 1 | 0 | 150 | 0 | 2.3 | 0 | 0 | 1 | 1 |
| 2 | 37 | 1 | 2 | 130 | 250 | 0 | 1 | 187 | 0 | 3.5 | 0 | 0 | 2 | 1 |
| 3 | 41 | 0 | 1 | 130 | 204 | 0 | 0 | 172 | 0 | 1.4 | 2 | 0 | 2 | 1 |
| 4 | 56 | 1 | 1 | 120 | 236 | 0 | 1 | 178 | 0 | 0.8 | 2 | 0 | 2 | 1 |
| 5 | 57 | 0 | 0 | 120 | 354 | 0 | 1 | 163 | 1 | 0.6 | 2 | 0 | 2 | 1 |
| 6 | 57 | 1 | 0 | 140 | 192 | 0 | 1 | 148 | 0 | 0.4 | 1 | 0 | 1 | 1 |
| 7 | 56 | 0 | 1 | 140 | 294 | 0 | 0 | 153 | 0 | 1.3 | 1 | 0 | 2 | 1 |
| 8 | 44 | 1 | 1 | 120 | 263 | 0 | 1 | 173 | 0 | 0.0 | 2 | 0 | 3 | 1 |
| 9 | 52 | 1 | 2 | 172 | 199 | 1 | 1 | 162 | 0 | 0.5 | 2 | 0 | 3 | 1 |
| 10 | 57 | 1 | 2 | 150 | 168 | 0 | 1 | 174 | 0 | 1.6 | 2 | 0 | 2 | 1 |

Atenção: Extraia os resultados para o arquivo CSV através do botão de download e renomeie-o com o número da query. Você deve enviá-lo para os tutores de EBAC.

▼ 2.2. Query 2

No console do AWS Athena, execute a seguinte query para praticar a função de contagem **COUNT**:

```
SELECT COUNT(age) AS QUANTIDADE_LINHAS
FROM heartattack
```

No painel de resultados você deve encontrar o seguinte resultado. Esse resultado nos retorna a informação de quantas linhas temos na tabela.

| | QUANTIDADE_LINHAS ▾ |
|---|---------------------|
| 1 | 303 |

Atenção: Extraia os resultados para o arquivo CSV através do botão de download e renomeie-o com o número da query. Você deve enviá-lo para os tutores de EBAC.

▼ 2.3. Query 3

No console do AWS Athena, execute a seguinte query para praticar a função **GROUP BY**:

```
SELECT COUNT(age) AS QUANTIDADE,
CASE
```

```

WHEN output =1 THEN ' more chance of heart attack'
ELSE 'less chance of heart attack'
END AS output
FROM heartattack
GROUP BY output;

```

No painel de resultados você deve encontrar o seguinte resultado.

| | QUANTIDADE ▼ | output ▼ |
|---|--------------|-----------------------------|
| 1 | 138 | less chance of heart attack |
| 2 | 165 | more chance of heart attack |

Atenção: Extraia os resultados para o arquivo CSV através do botão de download e renomeie-o com o número da query. Você deve enviá-lo para os tutores de EBAC.

▼ 3. Funções MIN/MAX/SUM/AVG

▼ 3.1 Query 4

No console do AWS Athena, execute a seguinte query para praticar a função GROUP BY:

```

SELECT MAX(age), MIN(age), AVG(age), output
FROM heartattack
GROUP BY output

```

No painel de resultados você deve encontrar o seguinte resultado.

| | _col0 ▼ | _col1 ▼ | _col2 ▼ | output ▼ |
|---|---------|---------|-------------------|----------|
| 1 | 76 | 29 | 52.4969696969697 | 1 |
| 2 | 77 | 35 | 56.60144927536232 | 0 |

Atenção: Extraia os resultados para o arquivo CSV através do botão de download e renomeie-o com o número da query. Você deve enviá-lo para os tutores de EBAC.

▼ 3.2 Query 5

No console do AWS Athena, execute a seguinte query para praticar a função GROUP BY:

```
SELECT MAX(age), MIN(age), AVG(age), output ,sex
FROM heartattack
GROUP BY output, sex;
```

No painel de resultados você deve encontrar o seguinte resultado.

| | _col0 ▾ | _col1 ▾ | _col2 ▾ | output ▾ | sex ▾ |
|---|---------|---------|--------------------|----------|-------|
| 1 | 76 | 34 | 54.55555555555556 | 1 | 0 |
| 2 | 70 | 29 | 50.903225806451616 | 1 | 1 |
| 3 | 77 | 35 | 56.08771929824562 | 0 | 1 |
| 4 | 66 | 43 | 59.041666666666664 | 0 | 0 |

Atenção: Observe a diferença entre os resultados de saída da Query 4 e da Query 5.

Atenção: Extraia os resultados para o arquivo CSV através do botão de download e renomeie-o com o número da query. Você deve enviá-lo para os tutores de EBAC.

▼ 4. Função HAVING

▼ 4.1. Query 6

No console do AWS Athena, execute a seguinte query para praticar a função HAVING:

```
SELECT COUNT(output), output, sex
FROM heartattack
GROUP BY output, sex
having COUNT(output) > 25
```

No painel de resultados, verifique a saída:

| | _col0 ▾ | output ▾ | sex ▾ |
|---|---------|----------|-------|
| 1 | 93 | 1 | 1 |
| 2 | 114 | 0 | 1 |
| 3 | 72 | 1 | 0 |

Atenção: Extraia os resultados para o arquivo CSV através do botão de download e renomeie-o com o número da query. Você deve enviá-lo para os tutores de EBAC.

