



Módulo | SQL: Projeto

Caderno de **Informações**

Professor [Mariane Neiva](#)

Tópicos

1. Criação da tabela e algumas queries;
2. Links úteis;
3. Dicas de livros, podcasts e outros;

▼ O PROJETO

Crie um bloco de notas e insira o link do seu projeto para o tutor avaliar!

Não esqueça de deixar o portfólio público! Você pode postar o link com uma pequena descrição nas suas redes sociais (Facebook, LinkedIn, Twitter, ...), prática altamente recomendável

1. Criação da tabela e algumas queries

Para essa parte, utilizaremos os mesmo dados do módulo 3.

As informações de **credito8.csv** estarão disponíveis no material de aula.

Siga os seguintes passos para criação da tabela (desconsidere caso você já tenha os dados no seu S3):

- Crie uma pasta bucket-transacoes no seu S3 e carregue o arquivo **credito8.csv**
- Volta para o AWS Athena e execute o seguinte comando:

```
CREATEEXTERNAL TABLE IF NOTEXISTS default.credito (  
  `idade` int,  
  `sexo` string,  
  `dependentes` int,  
  `escolaridade` string,  
  `estado_civil` string,  
  `salario_anual` string,  
  `tipo_cartao` string,  
  `qtd_produtos` bigint,  
  `iteracoes_12m` int,  
  `meses_inativo_12m` int,  
  `limite_credito` float,  
  `valor_transacoes_12m` float,  
  `qtd_transacoes_12m` int  
)  
ROWFORMATSERDE'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'  
WITHSERDEPROPERTIES(  
  'serialization.format' = ',',  
  'field.delim' = ',',  
) LOCATION <sua-localizacao>  
TBLPROPERTIES ('has_encrypted_data'='false');
```

Utilizaremos a tabela para o projeto.

Abaixo vão algumas dicas de queries que podem ser utilizadas. Sugiro que crie novas, com outras funcionalidades não utilizadas aqui e faça novas perguntas para o dataset.

```
select * from credito limit 10;
```

```
select count(*) from credito
```

```
DESCRIBE credito
```

```
SELECTDISTINCT escolaridade FROMcredito
```

```
select count(*), salario_anual from credito group by salario_anual
```

```
select count(*), sexo from credito group by sexo
```

```
select max(limite_credito) as limite_credito, escolaridade, tipo_cartao, sexo from credito  
where escolaridade != 'na' and tipo_cartao != 'na'  
group by escolaridade, tipo_cartao, sexo  
order by limite_credito desc  
limit 10
```

```
select max(valor_transacoes_12m) as maior_valor_gasto, avg(valor_transacoes_12m) as media_valor_gast  
from credito  
group by sexo
```

```
select avg(qtd_produtos) as qts_produtos, avg(valor_transacoes_12m) as media_valor_transacoes, avg(l  
from credito  
where salario_anual != 'na'  
group by sexo, salario_anual  
order by avg(valor_transacoes_12m) desc
```

2. Links úteis

- Kaggle: <https://www.kaggle.com/>
- Portfólio projeto: <https://www.kaggle.com/marianeneiva/credit-eda-and-analysis>
- LinkedIn: <https://www.linkedin.com/feed/>
- GitHub: <https://github.com/>
- Dados originais - gitHub André Perez: <https://github.com/andre-marcos-perez/ebac-course-utils>
- Imagens do gitHub - Mariane Neiva: <https://github.com/marianeneiva/sqlEBAC>
- AWS Athena: <https://aws.amazon.com/pt/athena/?whats-new-cards.sort-by=item.additionalFields.postDateTime&whats-new-cards.sort-order=desc>

3. Dicas de livros, podcasts e outros

Livro data science para negócios: [aqui](#)

Livro Storytelling with data: [aqui](#)

Podcast Data Hackers: [aqui](#)

Medium Toward Data Science: [aqui](#)

Analytics Vidhya: [aqui](#)

MIT Tech Review: [aqui](#)

Não há necessidade de comprar/ouvir/ler! São apenas dicas para você!