



Projeto Credit EDA and Analysis

Professor [Mariane Neiva](#)

Aluno [Denner Martins](#)

Exploração e análise de dados de crédito com SQL

✓ O PROJETO

Este projeto em SQL visa analisar dados de crédito bancário para entender padrões de comportamento financeiro e suporte ao crédito entre os clientes. Utilizando um banco de dados de transações, ele investiga variáveis como salários, gênero dos usuários e o valor máximo de crédito disponível. O foco é identificar fatores que influenciam o uso de crédito, criando insights para decisões de concessão de crédito mais informadas. Além disso, o projeto aborda a construção de consultas e visualizações que facilitam o acompanhamento de tendências e comportamentos de usuários no sistema bancário.

Os Dados

Os dados representam informações de clientes de um banco e contam com as seguintes colunas:

idade = idade do cliente

sexo = sexo do cliente (F ou M)

dependentes = número de dependentes do cliente

escolaridade = nível de escolaridade do clientes

salario_anual = faixa salarial do cliente

tipo_cartao = tipo de cartao do cliente

qtd_produtos = quantidade de produtos comprados nos últimos 12 meses

iteracoes_12m = quantidade de iterações/transacoes nos ultimos 12 meses

meses_inativo_12m = quantidade de meses que o cliente ficou inativo

limite_credito = limite de credito do cliente

valor_transacoes_12m = valor das transações dos ultimos 12 meses

qtd_transacoes_12m = quantidade de transacoes dos ultimos 12 meses

A tabela foi criada no AWS Athena junto com o S3 Bucket com uma versão dos dados disponibilizados em: <https://github.com/andre-marcos-perez/ebac-course-utils/tree/main/dataset>

Criação da Tabelas e algumas Queries

Não há necessidade de entrega. Para esses exercícios, 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:

```
CREATE EXTERNAL TABLE IF NOT EXISTS 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  
)  
  
ROW FORMAT SERDE 'org.apache.hadoop.hive.serde2.lazy.LazySimpleSerDe'  
WITH SERDEPROPERTIES (  
  'serialization.format' = ',',  
  'field.delim' = ',',  
) LOCATION <sua-localizacao>  
TBLPROPERTIES ('has_encrypted_data'='false');
```

✚ Exploração de dados

A primeira fase da análise é entender o que tem na nossa matéria prima. Vamos a exploração de dados:

Qual a quantidade de informações temos na nossa base de dados?

```
Query: SELECT count(*) FROM credito  
  
Resposta: 2564 linhas
```

#	▼	_col0
1		2564

Quais são esse dados?

```
Query: SELECT * FROM credito LIMIT 10;
```

idade ▼	sexo ▼	dependentes ▼	escolaridade ▼	estado_civil ▼	salario_anual ▼	tipo_cartao ▼	qtd_produtos ▼	iteracoes_12m ▼	meses_inativo_12m ▼	lir	limite_credito ▼	valor_transacoes_12m ▼	qtd_transacoes_12m
45	M	3	ensino medio	casado	\$60K - \$80K	blue	5	3	1		12691.51	1144.9	42
49	F	5	mestrado	solteiro	menos que \$40K	blue	6	2	1		8256.96	1291.45	33
51	M	3	mestrado	casado	\$80K - \$120K	blue	4	0	1		3418.56	1887.72	20
40	F	4	ensino medio	na	menos que \$40K	blue	3	1	4		3313.03	1171.56	20
40	M	3	sem educacao formal	casado	\$60K - \$80K	blue	5	0	1		4716.22	816.08	28
44	M	2	mestrado	casado	\$40K - \$60K	blue	3	2	1		4010.69	1088.07	24
51	M	4	na	casado	\$120K +	gold	6	3	1		34516.72	1330.87	31
32	M	0	ensino medio	na	\$60K - \$80K	silver	2	2	2		29081.49	1538.32	36
37	M	3	sem educacao formal	solteiro	\$60K - \$80K	blue	5	0	2		22352.5	1350.14	24
48	M	2	mestrado	solteiro	\$80K - \$120K	blue	6	3	3		11656.41	1441.73	32

Ps: É possível notar que tem alguns valores nulos

✚ Análise de Dados

Segunda fase é analisar esses dados de uma forma completa e clara.

Bora para análise?!

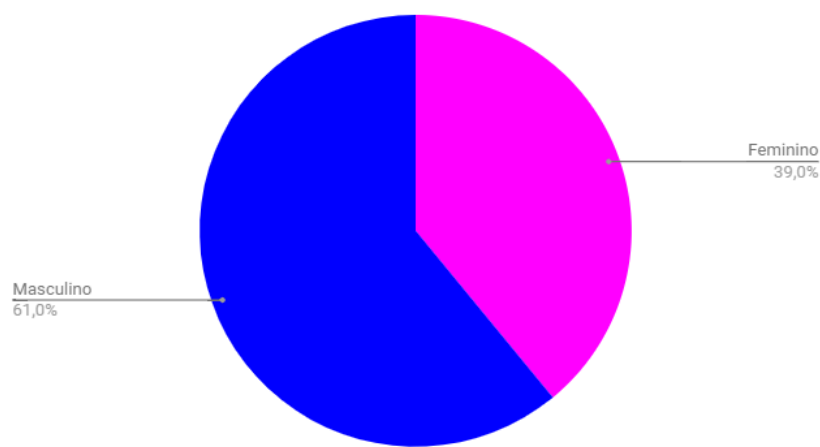
Será que a questão do sexo influencia no limite de crédito?

Quantos clientes são homens e quantos são mulheres?

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

#	_col0	sexo
2	1001	F
1	1563	M

Quantidade de cliente por sexo



Insight: A maioria dos clientes desse banco é homem! Do csv extraído dos dados é possível gerar o gráfico em pizza com para melhor a proporção de masculino/feminino

Vamos olhar os maiores limites de créditos!

```
Query:
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
```

#	limite_credito	escolaridade	tipo_cartao	sexo
1	34516.99	sem educacao formal	blue	M
2	34516.98	mestrado	gold	M
3	34516.97	mestrado	blue	M
4	34516.96	sem educacao formal	silver	M
5	34516.96	doutorado	platinum	F
6	34516.95	ensino medio	gold	M
7	34516.94	graduacao	gold	F
8	34516.94	graduacao	gold	M
9	34516.89	graduacao	silver	M
10	34516.74	ensino medio	blue	M

Agora vamos olhar os menores limites para tirarmos uma conclusão melhor.

```
Query:
```

```
select max(limite_credito) as limite_credito, escolaridade, tipo_cartao, sexo from credito where escolaridade != 'na' and tipo_cartao != 'na' grou
```

#	limite_credito	escolaridade	tipo_cartao	sexo
1	13882.37	ensino medio	silver	F
2	15534.85	ensino medio	gold	F
3	15987.78	mestrado	gold	F
4	23032.31	graduacao	blue	F
5	24936.22	mestrado	blue	F
6	26053.49	doutorado	blue	F
7	26181.89	sem educacao formal	blue	F
8	29543.1	ensino medio	blue	F
9	30310.45	mestrado	silver	F
10	34516.16	doutorado	gold	M

Insight 1: É visivelmente notável que os maiores limites de crédito são para o **sexo masculino**, **apenas 3 do Top 10 maiores limites são mulheres** e no Top 10 menores créditos temos apenas 1 do sexo masculino.

Insight 2: A escolaridade não é um critério para ter limite alto ou menor, podemos ver que clientes **Sem educação formal** tem mais crédito que clientes com **Mestrado**.

Qual sexo gasta mais créditos, Masculino ou Feminino?

```
Query:
select max(valor_transacoes_12m) as maior_valor_gasto, avg(valor_transacoes_12m) as media_valor_gasto, min(valor_transacoes_12m) as min_valor_gast
```

#	maior_valor_gasto	media_valor_gasto	min_valor_gasto	sexo
1	4686.93	1807.968	530.36	M
2	4776.58	1839.6226	510.16	F

Insight 1: As mulheres gastam bem mais crédito que os homens.

Insight 2: Acho que seria uma boa aumentar os limites do sexo feminino para que tenha mais giro de crédito.

Mas o que realmente impacta para gerar os Limites?

O salário impacta?

```
Query:
select avg(qtd_produtos) as qts_produtos, avg(valor_transacoes_12m) as media_valor_transacoes, avg(limite_credito) as media_limite, sexo, salario_
```

media_valor_transacoes	media_limite	sexo	salario_anual
1701.4652	17801.488	M	\$120K +
1755.2499	14886.556	M	\$80K - \$120K
1781.4299	4773.4463	F	\$40K - \$60K
1818.6364	9096.028	M	\$60K - \$80K
1845.1332	4212.1987	F	menos que \$40K
1878.2974	5753.31	M	\$40K - \$60K
1963.6719	3452.4036	M	menos que \$40K

Insight 1: Sim, o salário impacta no limite de crédito, As pessoas que tem menor faixa salarial também apresentam menor limite de credito.
Insight 2: E é notável que o máximo de salário de uma mulher é de 40k - 60k.

✦ **Vamos fazer o Marketing!**

Queremos focar o nosso marketing de maneira adequada para nossos clientes, qual será a idade deles?

Query:
select avg(idade) as media_idade, min(idade) as min_idade, max(idade) as max_idade, sexo from credito group by sexo

#	media_idade	min_idade	max_idade	sexo
2	46.42957042957043	26	67	F
1	45.57389635316699	26	73	M

Insight 1: Por meio dessa análise não foi possível extrair nenhuma informação relevante. A menor idade dos dois sexos é a mesma e a média é muito similar. A unica diferença é a idade máxima mas fato quase irrelevante por a diferença não é tão gritante.

Insight 2: Podemos fazer o Marketing para ambos o sexos, lembrando que a maioria dos clientes são homens e um grande ponto é que o sexo feminino gasta mais em crédito.

✦ **Conclusão**

Ao concluir este projeto de análise de crédito em SQL, foi possível identificar padrões e tendências importantes no comportamento financeiro dos clientes. A partir da análise de dados de transações, salários, gênero e limites de crédito, desenvolvemos insights valiosos que oferecem suporte à tomada de decisões estratégicas na concessão de crédito. Os resultados destacam a importância de uma gestão baseada em dados, permitindo ao banco ajustar políticas de crédito, reduzir riscos e atender melhor as necessidades dos clientes. Este projeto reforça o impacto do SQL como ferramenta essencial para transformar dados brutos em soluções práticas e orientadas a resultados no setor bancário.

Alguns insights interessantes:

- A maior parte dos clientes possui renda até 40K
- A maior parte dos clientes é masculino!
- A escolaridade não parece influenciar no limite nem no tipo do cartão
- Os clientes com maiores limites são em sua maioria homens
- Os clientes com menores limites são em sua maioria mulheres
- Dentre os menores limites não há presença de cartão platinum
- A faixa salarial impacta diretamente no limite de crédito
- Não existem clientes com salário anual acima de 60K do sexo feminino

Uma exploração maior dos dados pode explicar porque as mulheres tem menor crédito. Isso também pode ser um problema cultural que pode ser repensado!