

# Exploração e análise de dados de crédito com SQL

*Esse notebook faz parte do projeto final do curso SQL para análise de dados da EBAC.*

## 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>

## Exploração de dados:

Tem como base o entendimento dos nosso dados.

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

Query: `SELECT count(*) FROM credito`

Reposta: 2564 linhas

**Quais os tipos de cada dado**

Query: `DESCRIBE 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
```

Como são os dados

Query: SELECT \* FROM credito LIMIT 10;

# ▾	idade ▾	sexo ▾	dependentes ▾	escolaridade ▾	estado_civil ▾	salario_anual ▾	tipo_cartao ▾	qtd_produtos ▾	iteracoes_12m ▾	meses_inativo_12m ▾	limite_credito ▾	valor_transacoes_12m ▾	qtd_transacoes_12m ▾
1	45	M	3	ensino medio	casado	\$60K - \$80K	blue	5	3	1	12691.51	1144.9	42
2	49	F	5	mestrado	solteiro	menos que \$40K	blue	6	2	1	8256.96	1291.45	33
3	51	M	3	mestrado	casado	\$80K - \$120K	blue	4	0	1	3418.56	1887.72	20
4	40	F	4	ensino medio	na	menos que \$40K	blue	3	1	4	3313.03	1171.56	20
5	40	M	3	sem educacao formal	casado	\$60K - \$80K	blue	5	0	1	4716.22	816.08	28
6	44	M	2	mestrado	casado	\$40K - \$60K	blue	3	2	1	4010.69	1088.07	24
7	51	M	4	na	casado	\$120K +	gold	6	3	1	34516.72	1330.87	31
8	32	M	0	ensino medio	na	\$60K - \$80K	silver	2	2	2	29081.49	1538.32	36
9	37	M	3	sem educacao formal	solteiro	\$60K - \$80K	blue	5	0	2	22352.5	1350.14	24
10	48	M	2	mestrado	solteiro	\$80K - \$120K	blue	6	3	3	11656.41	1441.73	32

Quais são os tipos de escolaridade disponíveis no dataset?

Query: SELECT DISTINCT escolaridade FROM credito

# ▾	escolaridade
1	graduacao
2	doutorado
3	na
4	ensino medio
5	mestrado
6	sem educacao formal

Quais são os tipos de estado\_civil disponíveis no dataset?

Query: SELECT DISTINCT estado\_civil FROM credito

# ▾	estado_civil
1	divorciado
2	solteiro
3	na
4	casado

**Quais são os tipos de salario\_anual disponíveis no dataset?**

**Query:** SELECT DISTINCT salario\_anual FROM credito

# ▾	salario_anual
1	\$60K - \$80K
2	\$80K - \$120K
3	na
4	\$40K - \$60K
5	\$120K +
6	menos que \$40K

**Quais são os tipos de cartão disponíveis no dataset?**

**Query:** SELECT DISTINCT tipo\_cartao FROM credito

# ▾	tipo_cartao
1	silver
2	platinum
3	blue
4	gold

---

## Análise de dados

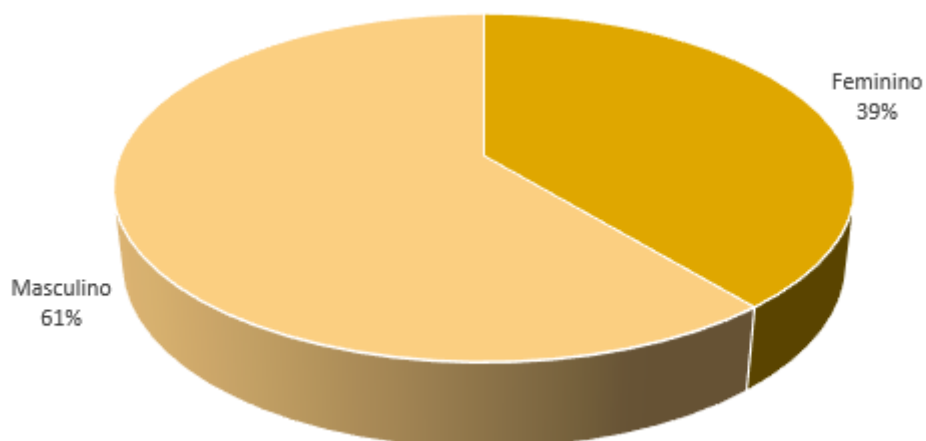
**Query:** select count(\*), salario\_anual from credito group by salario\_anual

# ▾	_col0 ▾	salario_anual
1	467	\$40K - \$60K
2	222	\$120K +
3	451	\$60K - \$80K
4	235	na
5	488	\$80K - \$120K
6	701	menos que \$40K

Nesse banco de dados, quantos clientes são homens e quantos são mulheres?

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

# ▾	_col0 ▾	sexo
1	1001	F
2	1563	M



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
1	46.42957042957043	26	67	F
2	45.57389635316699	26	73	M

Qual a maior e menor transação dos clientes?

**Query:** select min(valor\_transacoes\_12m) as transacao\_minima, max(valor\_transacoes\_12m) as transacao\_maxima from

credito				
#	transacao_minima			transacao_minima
1	510.16			4776.58

Quais as características dos clientes que possuem os maiores creditos?

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 order by limite\_credito desc limit 10

#	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

O salário impacta no limite?

Query: select avg(qtd\_produtos) as qts\_produtos, avg(valor\_transacoes\_12m) as media\_valor\_transacoes, avg(limite\_credito) as media\_limite, sexo, salario\_anual from credito where salario\_anual != 'na' group by sexo, salario\_anual order by avg(valor\_transacoes\_12m) desc

#	qts_produtos	media_valor_transacoes	media_limite	sexo	salario_anual
1	4.394230769230769	1963.6719	3452.4036	M	menos que \$40K
2	4.291970802919708	1878.2974	5753.31	M	\$40K - \$60K
3	4.408710217755444	1845.1332	4212.1987	F	menos que \$40K
4	4.352549889135255	1818.6364	9096.028	M	\$60K - \$80K
5	4.4404145077720205	1781.4299	4773.4463	F	\$40K - \$60K
6	4.329918032786885	1755.2499	14886.556	M	\$80K - \$120K
7	4.328828828828829	1701.4652	17801.488	M	\$120K +

Conclusão

Essas foram **algumas** análises extraídas do dataset de crédito.

Alguns insights interessantes:

- a maior parte dos clientes possui renda até 40K
- a maior parte dos clientes é masculino!
- os clientes com maiores limites são em sua maioria homens
- os clientes com menores limites são em sua maioria mulheres
- a faixa salarial impacta diretamente no limite de crédito
- nao existem clientes com salário anual acima de 60K do sexo feminino