

Projeto SQL - Comparativo de gastos anuais entre homens e mulheres.

O projeto abaixo tem como principal objetivo fazer um comparativo entre os gastos anuais dos nossos clientes e as possíveis causas para a discrepância nos valores entre homens e mulheres.

Para isso fiz uma pequena análise do perfil dos nossos clientes com os seguintes critérios:

- idade;
- escolaridade;
- estado civil;
- salário anual;
- dependentes;
- valor das transações nos últimos 12 meses.

Utilizei o SQLite para executar as queries, iniciei com a alteração do nome da coluna:

```
ALTER table credito RENAME COLUMN c1 to idade;
ALTER table credito RENAME COLUMN c2 to sexo;
ALTER table credito RENAME COLUMN c3 to dependentes;
ALTER table credito RENAME COLUMN c4 to escolaridade;
ALTER table credito RENAME COLUMN c5 to estado_civil;
ALTER table credito RENAME COLUMN c6 to salario_anual;
ALTER table credito RENAME COLUMN c7 to tipo_cartao;
ALTER table credito RENAME COLUMN c8 to qtde_produtos;
ALTER table credito RENAME COLUMN c9 to interacoes_12m;
ALTER table credito RENAME COLUMN c10 to meses_inativo_12m;
ALTER table credito RENAME COLUMN c11 to limite_credito;
ALTER table credito RENAME COLUMN c12 to valor_transacoes_12m;
ALTER table credito RENAME COLUMN c13 to qtd_transacoes_12m;
```

Dei um SELECT* para verificar se as colunas foram renomeadas corretamente:

```
SELECT*FROM credito;
```

Abaixo, fiz uma análise mais geral dos dados de cada coluna, calculando a média, valor mínimo e máximo das colunas onde temos números e um SELECT DISTINCT nas colunas que possuem strings e números:

```
SELECT DISTINCT idade FROM credito ORDER BY idade ASC;
```

*# O resultado desta query se encontra no arquivo Query 1 - Projeto, que se encontra no mesmo material deste projeto.
#Resumidamente, temos os valores 26 À 68, 70 e 73.*

#Média idade:

```
SELECT AVG (idade) FROM credito;
```

#Resultado: 46 anos.

```
SELECT DISTINCT escolaridade FROM credito;
```

O resultado desta query se encontra no arquivo Query 2 - Projeto, que se encontra no mesmo material deste projeto.

Ensino médio, Mestrado, sem educação formal, na, graduacao e doutorado.

Para descobrirmos a quantidade de clientes referente a escolaridade, executei os comandos abaixo:

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade = 'ensino medio';
```

#545

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade = 'mestrado';
```

#810

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade = 'mestrado' AND sexo = 'F';
```

#323 mulheres com mestrado

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade = 'mestrado' AND sexo = 'M';
```

#487 homens com mestrado

Escolaridade - Mestrado.png

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade = 'sem educacao formal';
```

#367

```
SELECT COUNT(escolaridade) FROM credito WHERE escolaridade = 'sem educacao formal' AND sexo = 'M';
```

#210 homens nao possuem uma educação formal.

Sem educação formal.png

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade = 'na';
```

#346

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade =  
'graduacao';
```

#261

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade =  
'doutorado';
```

#235

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade =  
'doutorado' AND sexo = 'F';
```

#98 mulheres com doutorado

```
SELECT COUNT (escolaridade) FROM credito WHERE escolaridade =  
'doutorado' AND sexo = 'M';
```

#137 homens com doutorado

Escolaridade - Doutorado.png

Podemos concluir que possuímos uma maior quantidade de clientes com mestrado e uma menor quantidade de clientes com doutorado

```
SELECT DISTINCT estado_civil FROM credito;
```

*# 0 resultado desta querie se encontra no arquivo Query 3 - Projeto, que se encontra no mesmo material deste projeto.
#Casado, solteiro, na, divorciado.*

Para descobrirmos a quantidade de clientes referente ao estado civil, executei os comandos abaixo:

```
SELECT COUNT (estado_civil) FROM credito WHERE estado_civil =  
'casado';
```

#1521

```
SELECT COUNT (estado_civil) FROM credito WHERE estado_civil =  
'solteiro';
```

#766

```
SELECT COUNT (estado_civil) FROM credito WHERE estado_civil = 'na';
```

#116

```
SELECT COUNT (estado_civil) FROM credito WHERE estado_civil =  
'divorciado';
```

#161

Podemos concluir que a maioria dos nossos clientes são casados.

```
SELECT DISTINCT salario_anual FROM credito;
```

O resultado desta query se encontra no arquivo Query 4 - Projeto, que se encontra no mesmo material deste projeto.

Na, menos de \$40K, \$40K à \$60K, \$60K à \$80K, \$80K à \$120K, mais de \$120K.

Para descobrirmos a quantidade de clientes referente a cada tipo de salario, executei os comandos abaixo:

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = 'menos  
que $40K';
```

#701

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = 'menos  
que $40K' AND sexo = 'F';
```

#597 mulheres ganham abaixo de 40K

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = 'menos  
que $40K' AND sexo = 'M';
```

#104 homens ganham abaixo de 40K

#É perceptível a discrepância nos valores, a maioria dos clientes que ganham menos de \$40K são mulheres.

_ Salario Anual _40K.png

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = '$40K  
- $60K';
```

#467

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = '$60K  
- $80K';
```

#451

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = '$80K  
- $120K';
```

#488

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = '$120K+';
```

#222

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = '$120K+' AND sexo = 'M'
```

#222

#Não há mulheres ganhando acima de \$120K anualmente.

```
SELECT COUNT (salario_anual) FROM credito WHERE salario_anual = 'na';
```

#235

Executei o comando abaixo para verificar a quantidade de dependentes:

```
SELECT DISTINCT dependentes FROM credito ORDER BY dependentes ASC;
```

#0,1,2,3,4,5

```
SELECT COUNT (dependentes) FROM credito WHERE dependentes = '0'
```

#337

```
SELECT COUNT (dependentes) FROM credito WHERE dependentes = '1'
```

#545

```
SELECT COUNT (dependentes) FROM credito WHERE dependentes = '2'
```

#680

```
SELECT COUNT (dependentes) FROM credito WHERE dependentes = '3'
```

#612

```
SELECT COUNT (dependentes) FROM credito WHERE dependentes = '4'
```

#324

```
SELECT COUNT (dependentes) FROM credito WHERE dependentes = '5'
```

#66

```
SELECT SUM (dependentes) FROM credito WHERE sexo = 'F';
```

#2069

```
SELECT SUM (dependentes) FROM credito WHERE sexo = 'M';
```

#3298

Dependentes.png

Conclui-se que os homens possuem mais dependentes que as mulheres.

```
SELECT SUM (dependentes) FROM credito WHERE sexo = 'F' and dependentes  
>= 3;
```

#1339

```
SELECT SUM (dependentes) FROM credito WHERE sexo = 'M' and dependentes  
>= 3;
```

#2123

Dependentes_=3.png

A maioria dos clientes possui 3 dependentes, sendo muito provavelmente conjuge e filhos. Neste cenário os homens também lideram.

Criei uma nova tabela com os valores das transações em ordem crescente, renomeei a coluna e executei as queries abaixo para identificar a mediana, visto que há uma discrepância entre o valor minimo e maximo no valor das transações dos ultimos 12 meses, conforme pode ser observado abaixo:

```
SELECT MIN (valor_transacoes_12m) FROM credito;
```

#\$510,16

```
SELECT MAX (valor_transacoes_12m) FROM credito;
```

#\$4776,58

```
SELECT AVG (valor_transacoes_12m) FROM credito;
```

#\$1820,32

```
SELECT SUM (valor_transacoes_12m) FROM credito WHERE sexo = 'F';
```

#\$1.841.462,13

```
SELECT SUM (valor_transacoes_12m) FROM credito WHERE sexo = 'M';
```

#\$2.825.853,99

```
SELECT valor_transacoes_12m FROM credito ORDER BY valor_transacoes_12m  
ASC;
```

#Tabela criada e extraída em formato .csv, incluída também no material referente a este projeto.

```
ALTER TABLE VALOR_TRANSACOES_12M RENAME COLUMN c1 TO VALORES;
```

Os primeiros valores em #azul correspondem a quantidade de linhas equivalente a porcentagem calculada, considerando o total de linhas na coluna valores.

```
SELECT count (valores),  
0.25 * COUNT(valores) AS vinte_e_cinco_porcento FROM  
VALOR_TRANSACOES_12M;
```

#641

#1352.81 (Q1)

```
SELECT count (valores),  
0.50 * COUNT(valores) AS cinquenta_porcento FROM VALOR_TRANSACOES_12M;
```

#1282

#1665.91 (Q2-médiana)

```
SELECT count (valores),  
0.75 * COUNT(valores) AS setenta_e_cinco_porcento FROM  
VALOR_TRANSACOES_12M;
```

#1923

#2156.6 (Q3)

Para confirmar a quantidade de linhas, vamos executar o abaixo:

```
select COUNT (valores) FROM VALOR_TRANSACOES_12M;
```

#2564

#Considerando que 50% deste valor é 1282, podemos assegurar de que as queries foram feitas e executadas corretamente.

Para calcular os quartis, executei os comandos abaixo, incluindo em frente ao limit os primeiros valores em #azul:

```
SELECT valores FROM VALOR_TRANSACOES_12M LIMIT 641;  
SELECT valores FROM VALOR_TRANSACOES_12M LIMIT 1282;  
SELECT valores FROM VALOR_TRANSACOES_12M LIMIT 1923;
```

Podemos concluir que não há tanta discrepância nos valores, comparando o valor da média (1820,32) e mediana(1665,91).

Considerações finais:

- É perceptível que o valor das transações feitas pelos clientes do sexo masculino foram praticamente 1 milhão a mais em comparação ao valor das transações feitas pelas clientes do sexo feminino;
- Podemos concluir que tal diferença está relacionada diretamente ao ganho anual de cada um, onde os homens possuem salários maiores que as mulheres (todos os clientes com salário superior a \$120K são homens). A maioria do publico que ganha menos de 40K/ano são mulheres;
- Tal diferença também pode estar relacionada a quantidade de dependentes, onde neste cenário possui uma maior quantidade de clientes homens;
- A discrepância dos salários não está relacionada a escolaridade, pois não há uma diferença significativa entre homens e mulheres;
- Fato interessante: A quantidade de homens sem educação formal é maior em comparação as mulheres.