

Introdução a Bancos de Dados

Exercícios de SQL

Professor: Clodoveu Davis

Banco de dados CINEMICRO (CINE80)

Instalação e configuração do PostgreSQL:

- **Windows:** <https://fabridata.com/como-instalar-postgresql-13-no-windows/>
- **Mac/Linux/Windows:** <https://hcode.com.br/blog/o-que-e-o-postgresql-instalando-e-criando-primeiro-banco-de-dados>
- **Ubuntu (Linux):** <https://www.hostinger.com.br/tutoriais/instalar-postgresql-ubuntu>

Instalar também o pgadmin 4: <https://www.pgadmin.org/>

No pgadmin 4, restaurar o BD:

- Baixar o arquivo zip da página da disciplina
- Descompactar no disco local
- Iniciar o PostgreSQL, e ativar o pgadmin
- Na coluna da esquerda, em “Servidores”, clicar o botão direito e criar um banco de dados (Create Database...). Basta informar um nome.
- Clicando o botão direito sobre “Servidores”, executar “Refresh” (atualizar a lista de bancos de dados do servidor). Surgirá o banco criado.
- Clicar o botão direito sobre o nome do banco criado, e executar “Restore” (restaurar). Indicar o arquivo descompactado do backup.
- Clicar novamente o “Refresh”, agora para o banco de dados, e navegar abaixo: as tabelas recuperadas surgirão dentro do esquema “public”.

Consulte a documentação (página da disciplina), incluindo esquema ER e esquema relacional para este banco de dados.

Execute as seguintes consultas em SQL, levando em consideração as boas práticas:

1. Correção (resultado atende ao especificado)
2. Legibilidade (expressão compreensível, espaçamento, quebra de linhas, indentação, etc.)
3. Economia (expressão não envolve estruturas desnecessárias, resultado não inclui dados desnecessários e elimina repetições)
4. Documentação (com comentários, se for necessário)
5. Eficiência

Em cada questão, apresente a consulta realizada em SQL e, se necessário, apresente considerações adicionais que você porventura tenha feito na solução do que foi pedido.

Em várias questões, valores constantes como títulos de produções e nomes de pessoas foram adaptados para os que utilizarem a versão reduzida do banco de dados, a cine80)

Parte 1: Análise exploratória de dados, caracterização do conteúdo do BD (data science)

Os resultados refletem o BD completo; reexecute cada consulta na sua instalação, com o BD reduzido.

As tarefas a seguir são importantes para que o interessado passe a conhecer melhor tanto a estrutura quanto a composição do conteúdo do banco de dados e suas tabelas. Seriam atividades preliminares comuns em *data science*, para caracterizar a distribuição de dados de acordo com atributos categóricos, avaliar a distribuição de valores de atributos numéricos, e obter dados detalhados sobre o volume de dados que se tem à disposição.

O resultado esperado não é apenas uma expressão SQL, mas também uma avaliação do resultado encontrado, que pode ser apresentada usando tabelas e gráficos.

1. Caracterizar cada tabela do esquema por quantidade de registros.

(obs: comando para obter uma lista de tabelas do esquema public

```
select table_name
from information_schema.tables
where table_schema = 'public')
```

```
select count(*) from tabela;
```

Tabela	Contagem
AUTOR	7.600.952
CONHECIDO_POR	11.620.223
DIRETOR	4.513.558
EPISODIO	3.947.290
GENERO	7.127.785
PARTIC	32.496.240
PESSOA	6.612.378
PROD	5.039.589
PROD_ALT	19.293.111
PROFISSAO	8.515.758

2. Caracterizar as produções por tipo de produção: quantidade de produções em cada tipo, avaliação média por tipo, desvio-padrão da avaliação

```
select tipo_prod, count(*), avg(avaliacao),
stddev(avaliacao)
from prod
group by tipo_prod;
```

	tipo_prod character (20)	count bigint	avg numeric	stddev numeric
1	movie	578741	6.1404745595943159	1.3366554179180319
2	tvEpisode	4132810	7.4973358602833711	1.1016635489983424
3	tvMiniSeries	35983	7.1181068208879061	1.2665599527698821
4	tvMovie	131665	6.6428057099545225	1.3610542667140612
5	tvSeries	160390	6.9544683230017923	1.3459737835828070

3. Caracterizar as produções por ano de produção: quantidade por ano, avaliação média no ano, desvio-padrão no ano, porém considerar apenas produções de 1940 em diante, e produções cuja quantidade de votos recebidos seja acima de 10.000. Apresentar o resultado em ordem descendente da média da avaliação.

```
select ano_prod, count(*), avg(avaliacao) as
media_aval, stddev(avaliacao) as dp_aval
from prod
where num_votos > 10000
group by ano_prod
having ano_prod > 1940
order by avg(avaliacao) desc;
```

1	1946	13	7.8153846153846154	0.37158703297736733176
2	1962	31	7.7870967741935484	0.40557936788736428338
3	1948	16	7.7750000000000000	0.29776948578836392696
4	1945	12	7.7750000000000000	0.30785179433085536708
5	1952	11	7.7727272727272727	0.56584610821478500456
6	1950	17	7.7647058823529412	0.41373478290836908328
7	1951	14	7.7357142857142857	0.39146105584305348849
8	1941	13	7.7076923076923077	0.30946604138294522933
9	1959	24	7.7041666666666667	0.54332244143079424304
10	1960	30	7.6966666666666667	0.56841178209374628141

4. Caracterizar as produções por quantidade média de participantes, em cada tipo de produção.

```
select a.tipo_prod, num_prods, parts,
1.0*parts/num_prods as media
from
(select tipo_prod, count(*) as num_prods
from prod
group by tipo_prod) a
join
(select tipo_prod, count(*) as parts
from prod p, partic pa
where p.id_prod = pa.id_prod
group by tipo_prod) b
on a.tipo_prod = b.tipo_prod
```

	tipo_prod character (20)	num_prods bigint	parts bigint	media numeric
1	movie	578741	4094808	7.0753722304104945
2	tvEpisode	4132810	26321067	6.3688064537203501
3	tvMiniSeries	35983	197082	5.4770864019120140
4	tvMovie	131665	965226	7.3309231762427372
5	tvSeries	160390	918057	5.7239042334310119

5. Caracterizar as pessoas por profissão: quantidade de pessoas em cada profissão, quantidade que já faleceram em cada profissão, média de idade em 2021 dos que estão vivos em cada profissão.

```
select prof, count(*) as num_prof
from profissao p
group by prof
```

	prof character varying (50)	num_prof bigint
1	actor	1803878
2	actress	1094227
3	animation_department	125914
4	art_department	290688
5	art_director	54128
6	assistant	1152
7	assistant_director	159196
8	camera_department	438479
9	casting_department	38041
10	casting_director	20726

```
select prof, count(*) as mortos_ou_indefinidos
from pessoa a join profissao p on a.id_pessoa =
p.id_pessoa
where ano_morte is null
group by prof
```

	prof character varying (50)	mortos_ou_indefinidos bigint
1	actor	1729618
2	actress	1062251
3	animation_department	124424
4	art_department	287084
5	art_director	52197
6	assistant	1148
7	assistant_director	154699
8	camera_department	433672
9	casting_department	37672
10	casting_director	20284

```
select prof, avg(2021-ano_nasc) as media_idade_2021
from pessoa a join profissao p on a.id_pessoa =
p.id_pessoa
where ano_morte is null
-- obs.null p/pessoas cujas datas são desconhecidas
group by prof
```

	prof character varying (50)	media_idade_2021 numeric
1	actor	53.8598844618129301
2	actress	51.7189790154792121
3	animation_department	52.4062717770034843
4	art_department	53.0053269410041284
5	art_director	59.7109448082319925
6	assistant	43.5102040816326531
7	assistant_director	52.9154431465009109
8	camera_department	50.2308368752748058
9	casting_department	47.1227915194346290
10	casting_director	50.5140931372549020

6. Caracterizar os participantes de produções por tarefa e por categoria.

```
select categoria, tarefa, count(*) as num
from partic
group by categoria, tarefa
order by num desc
```

	categoria character (20)	tarefa character varying (300)	num bigint
1	actor	[null]	8412197
2	actress	[null]	6469239
3	director	[null]	3715629
4	self	[null]	2887032
5	producer	producer	2016567
6	writer	[null]	1994818
7	composer	[null]	1024037
8	cinematographer	[null]	898717
9	editor	[null]	846591
10	writer	writer	648021

7. Caracterizar as produções quanto à região, dado que consta de prod_alt.

```
select regiao, count(*) as num
from prod_alt
group by regiao
order by num desc
```

	regiao character (10)	num bigint
1	JP	2362376
2	FR	2348147
3	DE	2328850
4	IN	2313726
5	ES	2304145
6	IT	2296046
7	PT	2260499
8	[null]	814943
9	US	548917
10	GB	248151

8. Caracterizar as produções do tipo 'movie' por avaliação, em faixas de valores a cada 0.5 ponto. Apenas considere produções que tenham recebido pelo menos 10.000 votos, e desconsidere aquelas que não tenham avaliação (null).

```
select trunc(avaliacao*2)/2 as faixa, count(*) as
quant
from prod
where avaliacao is not null
and num_votos > 10000
and tipo_prod = 'movie'
group by faixa
```

Faixa	quant
1.0	6
1.5	15
2.0	17
2.5	16
3.0	36
3.5	61
4.0	110
4.5	235
5.0	538
5.5	1014
6.0	1617
6.5	1797
7.0	1876
7.5	1381
8.0	639
8.5	102
9.0	11

9. Determinar a quantidade de participações em que o nome do personagem é informado.

```
select count(*) from partic where personagem is not null
```

➤ 15.386.935

10. Caracterizar as produções por tempo de duração: determinar a quantidade de produções por faixa de tempo, com intervalos de 15 minutos (0-15, 16-30, ...).

```
select trunc(tempo_minutos/15)*15 as faixa, count(*)
as quant
from prod
where tempo_minutos is not null
group by trunc(tempo_minutos/15)
```

	faixa double precision	quant bigint
1	0	84473
2	15	292801
3	30	216341
4	45	186153
5	60	130006
6	75	113386
7	90	138731
8	105	42342
9	120	26216
10	135	11221

11. Verificar se existe ou não redundância total entre as tabelas partic e autor, e entre as tabelas partic e diretor. Deseja-se saber se todas as participações destacadas como autor ou diretor nas tabelas específicas estão refletidas em partic, categorias 'writer' e 'director' – caso positivo, isso poderia levar a uma redução de linhas na tabela partic, ou à eliminação das tabelas autor e diretor. Caso não exista redundância total, sugerir maneiras de melhorar a qualidade dos dados existentes através da eliminação de redundâncias.

```
-- participacoes sem autor correspondente
select p.id_prod, count(*) as quant
from partic p left join autor a on p.id_prod =
a.id_prod
where categoria = 'writer'
group by p.id_prod
having p.id_prod is null
```

```
-- autores sem participacao correspondente
select a.id_prod, count(*) as quant
from partic p right join autor a on p.id_prod =
a.id_prod
where categoria = 'writer'
group by a.id_prod
having a.id_prod is null
```

```
-- participacoes sem diretor correspondente
select p.id_prod, count(*) as quant
from partic p left join diretor d on p.id_prod =
d.id_prod
where categoria = 'director'
group by p.id_prod
having p.id_prod is null
```

```
-- diretores sem participacao correspondente
select d.id_prod, count(*) as quant
from partic p right join diretor d on p.id_prod =
d.id_prod
```

```
where categoria = 'director'  
group by d.id_prod  
having d.id_prod is null
```

Não há autores nem diretores sem a participação registrada como 'writer' ou 'director', nem o inverso. As tabelas autor e diretor são redundantes em relação a partic.

12. Analisar os dados de data de nascimento e data de morte de pessoas registradas, observando a quantidade de dados nulos. Perceba que nessa tabela e nessas colunas o valor nulo tanto reflete desconhecimento quanto indefinição (ex., pessoa viva).

```
select count(*)  
from pessoa  
where ano_nasc is null
```

>> 6129759

```
select count(*)  
from pessoa  
where ano_morte is null
```

>> 6432665

```
select count(*)  
from pessoa  
where ano_nasc is null  
and ano_morte is null
```

>> 6120532

```
select count(*)  
from pessoa  
where ano_nasc is null  
and ano_morte is not null
```

>> 9227

```
select count(*)  
from pessoa  
where ano_nasc is not null  
and ano_morte is not null
```

>> 170486

```
select max (ano_morte - ano_nasc)  
from pessoa  
where ano_nasc is not null
```

```
and ano_morte is not null
```

```
>> 122 -- provável inconsistência
```

```
select min (ano_morte - ano_nasc)
from pessoa
where ano_nasc is not null
and ano_morte is not null
```

```
>> -42 – inconsistência?
```

```
select * from pessoa
where ano_nasc is not null
and ano_morte is not null
and ano_morte - ano_nasc < 0
```

	id_pessoa [PK] character (12)	nome character varying (150)	ano_nasc integer	ano_morte integer
1	nm0515385	Titus Livius	59	17
2	nm0393598	Michael Hook	1946	1913
3	nm0202552	C. Daveillans	1962	1936

1. Autor romano, indicado como roteirista em um filme. As datas são A.C.
2. e 3. Inversão nascimento-morte?

Parte 2: Consultas simples

1. Selecionar todos os filmes (movie) produzidos no ano de 1939 e listar seu título principal e ano de produção.
2. Listar os nomes e anos de nascimento de todas as pessoas envolvidas em produções cujo primeiro nome é 'Jennifer'.
3. Selecionar todos os filmes (movie) que contaram com a participação de 'Tom Hanks' na década de 1990, e listar seu título, ano de produção e duração em minutos.
4. Selecionar todos os episódios de séries de TV produzidos entre 1990 e 1999, listando o título da série, o título do episódio, a temporada, o número do episódio na temporada, o ano de produção, e a duração do episódio.
5. Listar o título e o ano de produção das produções pelas quais a pessoa 'Michael J. Fox' é conhecida.
6. Listar todos os participantes de 'Jurassic Park' (cine80: 'Back to the Future') que são da categoria 'actor' ou 'actress', apresentando nome, ano de nascimento, profissão, categoria e tarefa de cada um deles.
7. Listar todos os participantes de 'Jurassic Park' (cine80: 'Back to the Future'), **exceto** os que são da categoria 'actor' ou 'actress', apresentando nome, ano de nascimento, profissão, categoria e tarefa de cada um deles.
8. Listar o título e o ano de produção de todos os filmes (movie) dos quais participaram simultaneamente 'Tom Hanks' e 'Meg Ryan' (cine80: 'Harrison Ford' e 'Carrie Fisher').
9. Listar todos os filmes do gênero 'Action' ou 'Adventure' dos quais participou como ator (actor/actress) a pessoa 'Bruce Willis', apresentando nome e ano de produção.
10. Listar o nome, ano de nascimento e ano de morte de todos os diretores que dirigiram filmes (movie) escritos por 'Dalton Trumbo' (cine80: 'Robert Zemeckis'), sem repetições.
11. Listar o nome, no idioma original (titulo_alt), e o título principal de todos os filmes cujo idioma é 'pt' produzidos desde 2000 (cine80: desde 1980).
12. Obter o título principal e o nome do roteirista do filme cujo nome em português é 'Suplício de uma Saudade' (cine80: 'The Return of Godzilla').
13. Listar os nomes dos atores que são conhecidos pelo filme cujo nome em português é 'Suplício de uma Saudade'.
14. Listar os títulos principais e anos de produção dos filmes pelos quais 'William Holden' (cine80: 'Dan Aykroyd') é conhecido e que têm avaliação igual ou acima de 7, em ordem de avaliação.
15. Listar os títulos principais e anos de produção dos filmes dos quais 'William Holden' (cine80: 'Dan Aykroyd') participou e que têm avaliação igual ou acima de 7, em ordem de avaliação.
16. Listar o título principal, título alternativo, ano de produção, avaliação e duração em minutos dos filmes dos quais participou 'Toshiro Mifune'.
17. Listar o nome da série, o título do episódio, o número da temporada, o número do episódio de todas as produções do tipo 'tvSeries' das quais participou 'Bryan Cranston' (cine80: 'Jerry Seinfeld'). Organizar a saída em ordem de ano de produção, e dentro deste por ordem de temporada e episódio.

18. Listar os títulos, ano de produção e ano de encerramento das séries de TV (tvSeries) cuja avaliação é acima de 8.5.
19. Gerar a quantidade de filmes do gênero 'War' a cada ano entre 1940 e 2000 (cine80: entre 1980 e 1989), em ordem decrescente.
20. Determinar a quantidade de produções em cada gênero, apresentando a lista em ordem decrescente de quantidade, e apenas para os gêneros com mais de 100.000 produções.
21. Determinar a avaliação média de filmes (movie) de cada gênero, e apresentar o resultado em ordem decrescente de avaliação.
22. Determinar a quantidade de produções em cada combinação entre tipo de produção e gênero.
23. Determinar a avaliação média de todas as produções das quais 'Steven Spielberg' participou.
24. Determinar a avaliação média de todas as produções cujo diretor foi 'Steven Spielberg'.
25. Determinar a quantidade de participantes da produção cujo título principal é 'Toy Story' (cine80:).
26. Determinar a média de idade das pessoas cuja profissão é 'actor' ou 'actress' na época de cada produção pelas quais são conhecidos.
27. Determinar a pessoa cuja profissão é 'actor' ou 'actress' que tem a maior quantidade de participações em produções.
28. Determinar a produção (título principal, ano de produção) que tem a maior quantidade de participantes registrados.
29. Determinar a produção (título principal, ano de produção) que tem a maior quantidade de diretores.
30. Determinar a produção (título principal, ano de produção) que tem a maior quantidade de roteiristas (autores).
31. Determinar a produção do tipo 'tvSeries' ou 'tvMiniSeries' que tem a maior quantidade de episódios.
32. Determinar a produção do tipo 'tvSeries' ou 'tvMiniSeries' que tem a menor quantidade de episódios.
33. Determinar a duração média, mínima e máxima dos episódios de 'tvSeries'.
34. Determinar o personagem que foi interpretado por mais atores.
35. Informar a série que teve mais temporadas.
36. Determinar a produção do tipo 'movie' que tem duração mais longa.
37. Determinar o total de votos apurados para avaliação dos filmes (movie) de cada gênero, e apresentar o resultado em ordem decrescente desse total.
38. Determinar a quantidade de produções por idioma. Considerar que o idioma é 'en' sempre que o atributo estiver nulo.

Parte 3: Consultas complexas

1. Listar as pessoas que participaram tanto de filmes (movie) quanto de séries de TV (tvSeries, e seus episódios) na década de 1990 (1990-1999) (cine80: década de 1980 (1980-1989)), em ordem alfabética.

2. Listar as pessoas que participaram de filmes (movie) com avaliação acima de 8.5, mas não de séries de TV, na década de 1990 (cine80: década de 1980), em ordem reversa de seu ano de nascimento.
3. Listar as pessoas que participaram simultaneamente, na mesma produção, como roteirista (tabela autor), como diretor (tabela diretor) e como ator (tabela partic, categoria 'actor' ou 'actress').

Parte 4: Visões

1. Criar uma visão em que se possa consultar apenas atores e atrizes (actor) e os títulos e anos dos filmes (movie) em que atuaram.
2. Criar uma visão em que dados básicos da produção (id, título, ano de produção) venham acompanhados do nome do diretor.

Parte 5: INSERT/DELETE/UPDATE

Obs: erros nesses comandos, ou mesmo os efeitos desejados deles, causarão alterações no conteúdo das tabelas, com efeito sobre os exercícios anteriores. Caso necessário, restaure novamente o backup do conteúdo do banco de dados, como na primeira página.

1. Alterar o título de cada episódio de série para incluir " (TV episode)" após o título.
2. Excluir as pessoas com mais de 120 anos de idade em 2021 e que não tenham data de morte registrada.
3. Incluir seus dados (ou dados fictícios criados por você) na tabela pessoa, e criar registro de participação em uma produção como "actor". Indique como nome de personagem algo aleatório, como "Brazilian student". Execute uma consulta como a questão 6 da parte 2.