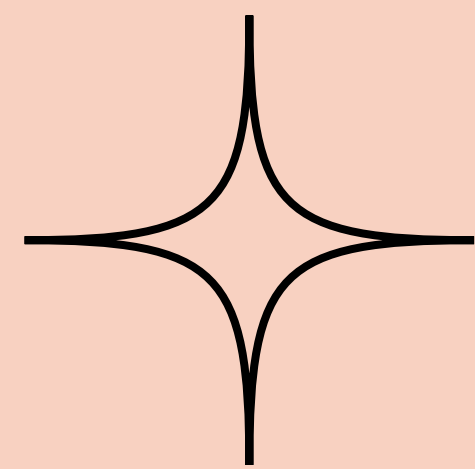
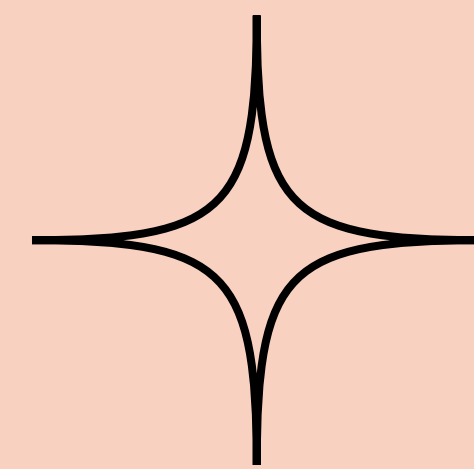


27 de Julho, Belém - PA

2022



Coda



Amazônia

→ Quando a planilha trava ←

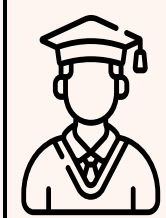
Trabalhando com bases de dados massivas

Agenda

01	Sobre: Jailson Pereira
02	Contexto: o mundo dos dados
03	Ferramenta: SQL
04	Referências
05	Agradecimentos
06	Contatos

Sobre: Jailson Pereira

Graduação



Sistemas de Informação - UEA

Mestrado



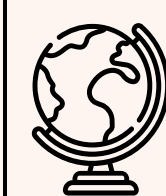
PPGI - UFAM

Profissional



Engenheiro de Dados - CI&T

Comunidade



PyData Manaus

Interesses e Pesquisa

NLP, Classificação de Imagens, Detecção de eventos sonoros e sistemas de recomendação

NLP, Detecção de viés político através de artigos de notícias

Curiosidades

Tecnologia, forró, anime, série e futevôlei

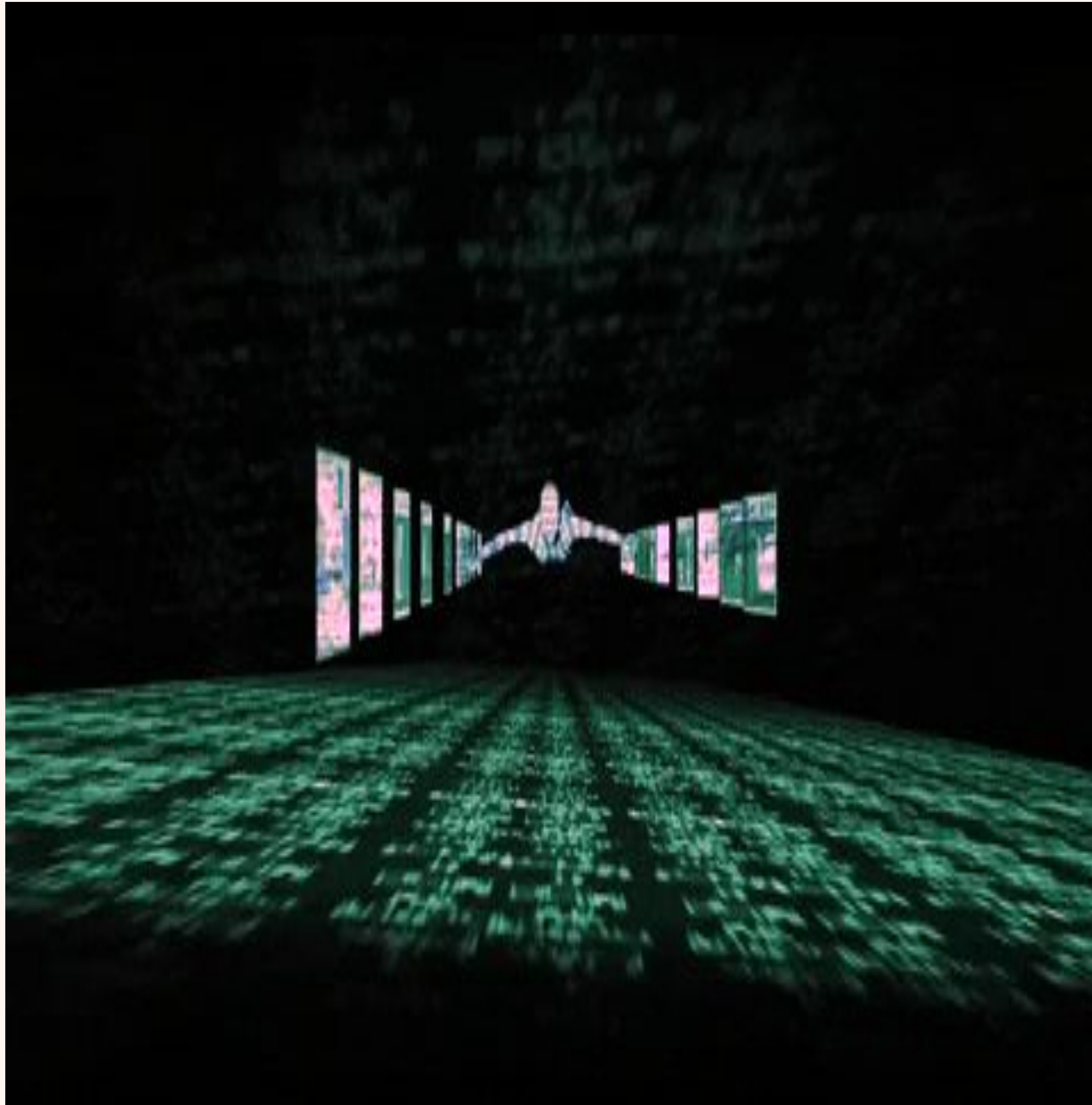
Contexto: o mundo dos dados



Introdução:

- Vivemos em um mundo com vasto volume de dados
- Dados disponíveis on-line, por exemplo, na web
- Grandes empresas como Google, Facebook e startups dos mais diversos segmentos baseiam seus negócios em dados coletados da web ou redes sociais
- Hoje o volume de dados pode ser tão grande que é simplesmente descartado por não ter condições de tratá-lo e utilizá-lo adequadamente

Contexto: o mundo dos dados



Introdução:

- Muito conhecimento importante sobre o negócio pode ser derivado a partir de dados registrados.
- Conhecimento pode ser estratégico para decisões e pode trazer vantagens competitivas
- Business Intelligence:
 - Tecnologias relacionadas, incluindo:
 - Data Warehousing
 - Data Mining
 - Data Analysis
 - Machine Learning

Contexto: o mundo dos dados



Big Data?

- Google
 - Processa 20 PB por dia (2008)
 - Coleta 20B Páginas web pages por dia (2012)
 - Índice de busca tem 100+ PB (2014)
 - Bigtable serve 2+EB, 600M QPS (2014)



Contexto: o mundo dos dados

Big Data?



- Netflix

- Milhões de membros em todo mundo, de milhares de dispositivos, visitam a Netflix e geram milhões de horas de visualização
- “milhões de membros assistindo mais de dois bilhões de horas todos os meses” (2015)
- Recomendação em tempo real



Contexto: o mundo dos dados



Ferramenta: SQL

Introdução:

- As linguagens de programação vêm e vão constantemente
- Na área de banco de dados, temos SQL
- O SQL foi criado inicialmente para ser a linguagem para gerar, manipular e recuperar dados de bancos de dados relacionais

Ferramenta: SQL

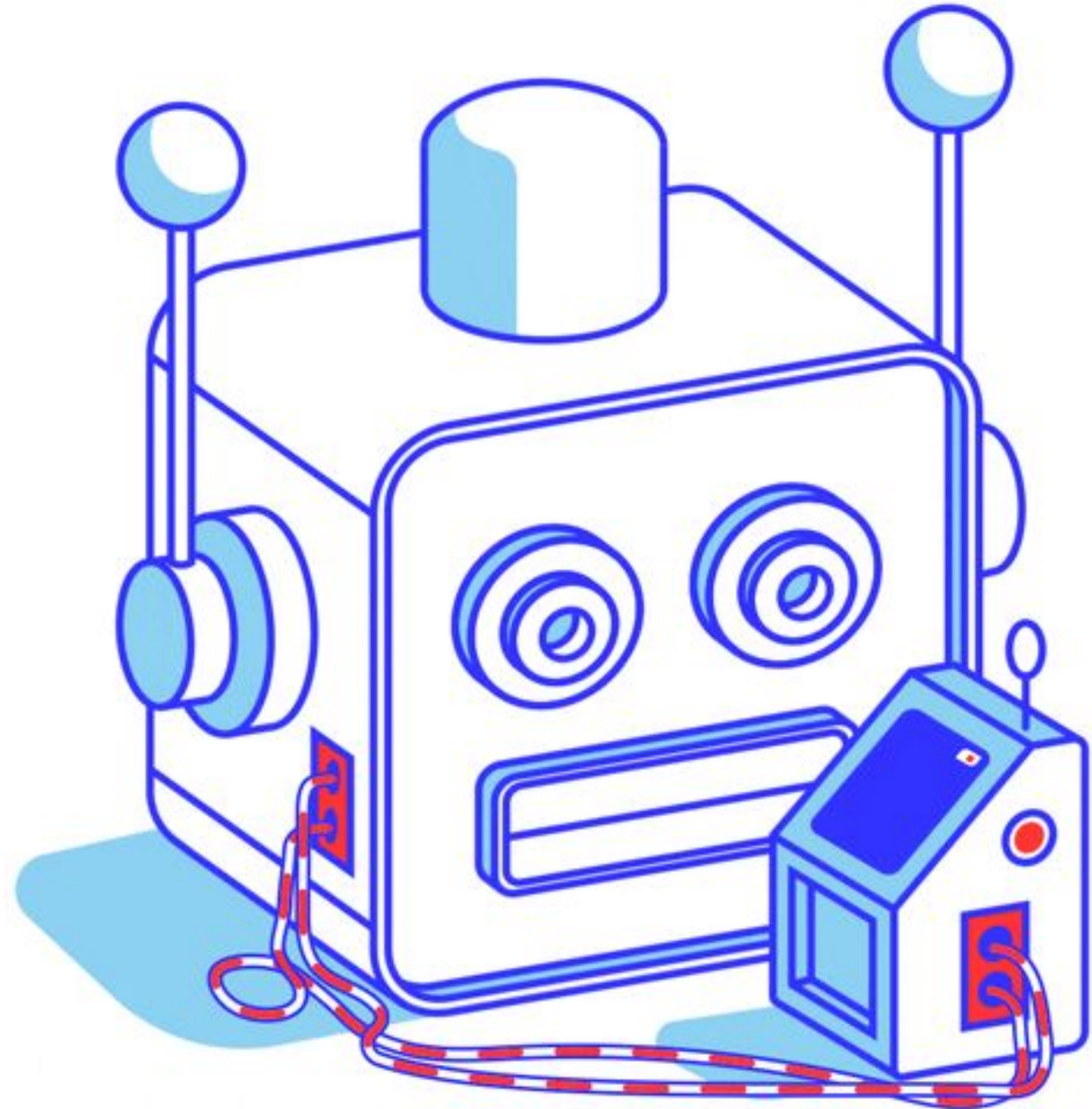
Introdução - Por quê estudar SQL?

- Usuários de banco de dados relacional ou não
- Pessoas que trabalham com Ciência de Dados, Business Intelligence ou Análise de Dados
- Combo SQL + Python ou SQL + R

Ferramenta: SQL

Introdução - Sistemas de Banco de dados:

- Mecanismos informatizados de armazenamento e recuperação de dados
- É um conjunto de informações relacionadas
 - Ex: Lista telefônica (nomes, números de telefone e endereço)
- Imprecisões:
 - Grande número de entradas
 - Indexação
 - Informações imprecisas ou desatualizadas



Ferramenta: SQL

Introdução - Modelo Relacional:

- Em 1970, Dr.E.F Codd, do laboratório de pesquisa da IBM publicou um artigo intitulado:
 - “A Relational Model of Data for Large Shared Data Banks”

Customer		
cust_id	fname	lname
1	George	Blake
2	Sue	Smith

Account			
account_id	product_cd	cust_id	balance
103	CHK	1	\$75.00
104	SAV	1	\$250.00
105	CHK	2	\$783.64
106	MM	2	\$500.00
107	LOC	2	0

Product	
product_cd	name
CHK	Checking
SAV	Savings
MM	Money market
LOC	Line of credit

Transaction				
txn_id	txn_type_cd	account_id	amount	date
978	DBT	103	\$100.00	2004-01-22
979	CDT	103	\$25.00	2004-02-05
980	DBT	104	\$250.00	2004-03-09
981	DBT	105	\$1000.00	2004-03-25
982	CDT	105	\$138.50	2004-04-02
983	CDT	105	\$77.86	2004-04-04
984	DBT	106	\$500.00	2004-03-27

Ferramenta: SQL - prática



Ferramentas: SQL - prática



Ferramenta: SQL - prática

Criação de Tabela:

- `CREATE TABLE databa_name.table (column type);`
- Tipos de dados:
 - Character Data: `char`, `varchar`
 - Text Data: `tinytext`, `text`, `mediumtext`, `longtext`
 - Numeric Data: `tinyint`, `smallint`, `mediumint`, `int`, `bigint`, `float`, `float`, `double`
 - Temporal Data: `date`, `datetime`, `timestamp`, `year`, `time`

Descrição da Tabela:

- `PRAGMA table_info("nome_da_tabela");`
- Ou `desc "nome_da_tabela";`

Ferramenta: SQL - prática

Inserir dados na Tabela:

- O nome da tabela na qual adicionar os dados
- Os nomes das colunas na tabela a serem preenchidas
- Os valores com os quais preencher as colunas

Comando para inserir dados na tabela:

- `INSERT INTO nome_da_tabela
(nome_da_coluna) VALUES
(valor_da_coluna);`

Comando para inserir dados na tabela:

- `SELECT * FROM nome_da_tabela;`

Ferramenta: SQL - prática

Query Cláusula:

- SELECT: Determina quais colunas incluir no conjunto de resultados da consulta
- FROM: Identifica as tabelas das quais recuperar dados e como as tabelas devem ser unidas
- WHERE: Filtra dados
- GROUP BY: Usado para agrupar linhas por valores de coluna comuns
- HAVING: Filtra grupos indesejados
- ORDER BY: Ordena as linhas do resultado final definido por uma ou mais colunas

Comando para uma query SELECT:

- `SELECT nome_da_coluna FROM nome_da_tabela;`

Combinações com SELECT:

- Alias
 - `SELECT nome_da_coluna AS novo_nome_da_coluna_consulta FROM nome_da_tabela;`
- Ordenação
 - `SELECT nome_da_coluna FROM nome_da_tabela ORDER BY nome_da_coluna_por_onde_ordenar;`
- Duplicados
 - `SELECT DISTINCT nome_da_coluna FROM nome_da_tabela; ou COUNT(DISTINCT)`

Ferramenta: SQL - prática

Subconsultas:

- Uma subconsulta é uma consulta contida em outra consulta
- As subconsultas são cercadas por parênteses e podem ser encontradas em várias partes de uma instrução select

Comando para uma subconsulta:

- SELECT + concat
 - SELECT concat(sub_consulta.col1, ',', sub_consulta.col2) nome FROM (SELECT col1, col2 FROM WHERE col3 = valor) sub_consulta;

Ferramenta: SQL - prática

Cláusula Where:

- A cláusula where é o mecanismo para filtrar linhas indesejadas do seu conjunto de resultados.

Comando para inserir dados na tabela:

- `SELECT nome_coluna FROM nome_da_tabela
WHERE nome_coluna = valor AND
nome_coluna2 >= valor;`

Ferramenta: SQL - prática

Cláusula GROUP BY:

- O GROUP BY gera o resultado de uma tendência com algum tipo de manipulação

Consulta com GROUP BY:

- ```
SELECT tb1.coluna_tabela_1 FROM tabela1
tb1 INNER JOIN tabela2 tb2 ON
tb1.coluna = tb2.coluna GROUP BY
tb1.coluna_tabela_1;
```

# Ferramenta: SQL - prática

## Cláusula ORDER BY:

- A cláusula ORDER BY é o mecanismo para ordenar seu conjunto de resultados usando dados de coluna brutos ou expressões baseadas em dados de coluna.

## Consulta com ORDER BY:

- ```
SELECT tb1.coluna_tabela_1 FROM tabela1  
tb1 INNER JOIN tabela2 tb2 ON  
tb1.coluna = tb2.coluna GROUP BY WHERE  
tb2.coluna = valor ORDER BY  
tb2.coluna2;
```
- Ascending ou Descending Sort Order

Ferramenta: SQL - prática

Consultas com INNER JOINS:

- Consultas em uma única tabela certamente não são raras, a maioria das consultas exige duas, três ou até mais tabelas.

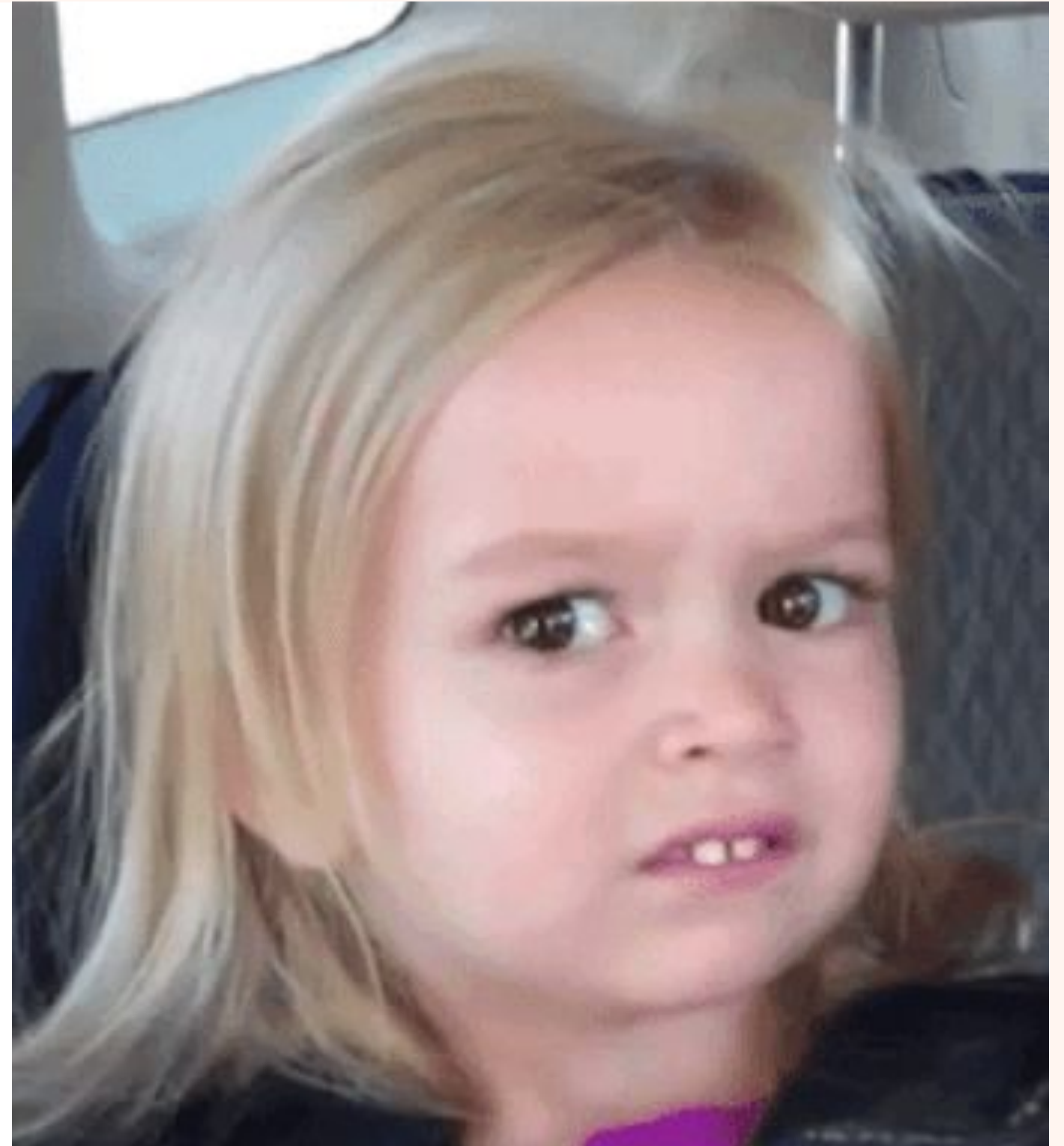
Consulta com INNER JOIN:

- ```
SELECT tb1.coluna FROM tabela1 tb1
INNER JOIN tabela2 tb2 ON tb1.coluna =
tb2.coluna WHERE tb1 = valor;
```

# Ferramenta: SQL - prática

Consultas com UNION:

- União
- Interseção
- Exeção
- Diferença





# Ferramenta: SQL - prática

Consultas com União:

- ALL
  - `SELECT tb1.coluna FROM tabela1 tb1  
UNION ALL SELECT tb2.coluna FROM  
tabela2 tb2`
- INTERSECT
  - `SELECT tb1.coluna FROM tabela1 tb1  
INTERSECT SELECT tb2.coluna FROM  
tabela2 tb2`

Consultas com União:

- EXCEPT
  - `SELECT tb1.coluna FROM tabela1 tb1  
EXCEPT SELECT tb2.coluna FROM  
tabela2 tb2`

# Referências

Livro: Alan Beaulieu - Learning SQL Generated, Manipulate and Retrieve Data - O'Reilly Media (2020)

Aulas Big Data/PPGI UFAM:

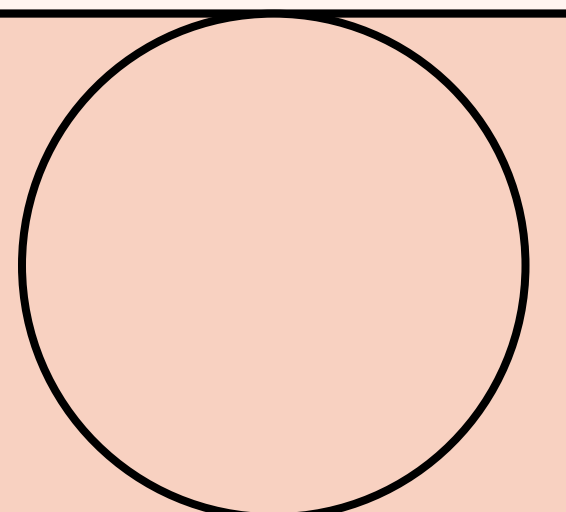
[https://www.youtube.com/watch?v=1BshZsP1UQQ&list=PLhyf4Hm3tvmh5fj35bB1W8cIV6mTiSn\\_u](https://www.youtube.com/watch?v=1BshZsP1UQQ&list=PLhyf4Hm3tvmh5fj35bB1W8cIV6mTiSn_u)

Base de dados: <https://basedosdados.org/>

Github Workshop: <https://github.com/jailsonpj/CodaAmazonia>



- LinkedIn: <https://www.linkedin.com/in/jailson-pereira-574a09115/>
- Email: [jailsoncolares@gmail.com](mailto:jailsoncolares@gmail.com)
- Github: jailsonpj
- Instagram: @jailson\_pj





# Agradecimentos ←

2022

OBRIGADO!!!

