

UNIVERSIDADE DE BRASÍLIA

Faculdade do Gama

Sistemas de Banco de Dados 2

Trabalho Final (TF)

Business Intelligence (BI) - C

Artur Handow Krauspenhar – 231034082

Carlos Eduardo Rodrigues – 221031265

Brasília, DF

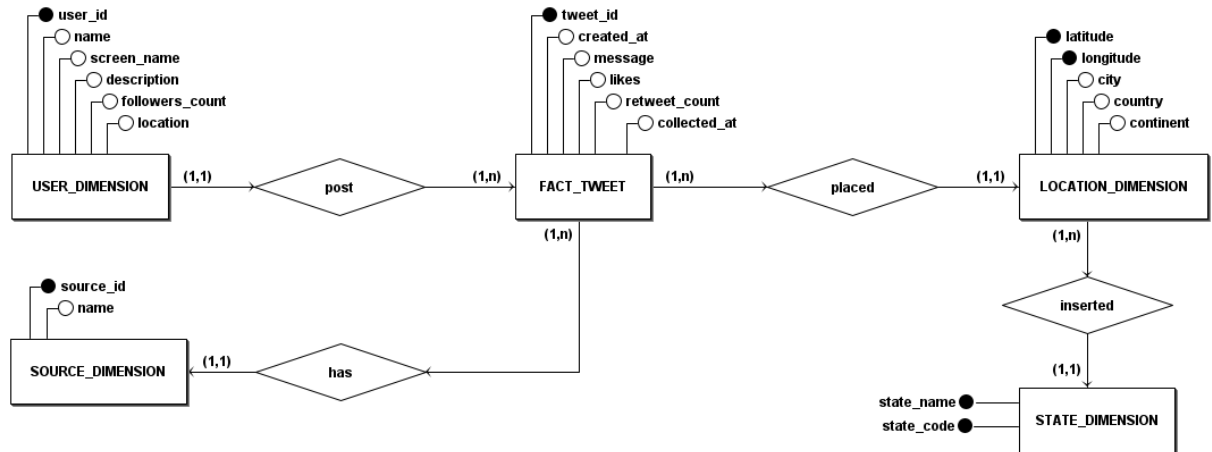
2025

BASE DE DADOS

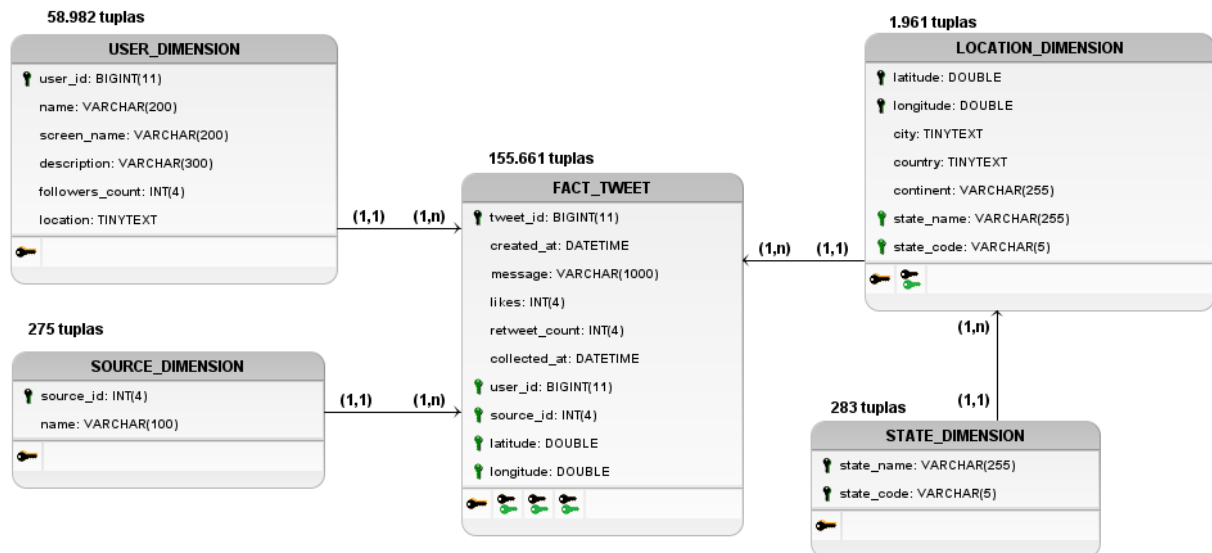
A base de dados escolhida para o trabalho foi proposta pelo estudante Artur Handow. A base denominada "US Election 2020 Tweets - Preliminary EDA" está disponível em: https://www.kaggle.com/code/manchunhui/us-election-2020-tweets-preliminary-eda/input?select=hashtag_joebiden.csv. O arquivo original é no formato CSV, porém, os dados foram normalizados para um BD MySQL através de um script em Python.

A base de dados contém tweets que utilizaram as hashtags #JoeBiden e #Biden, e que estavam relacionados às eleições de 2020 nos EUA. A escolha dessa base de dados foi feita estrategicamente pelos integrantes, sendo o principal motivo dessa escolha a possibilidade de investigar tendências e o engajamento político nas eleições de 2020 nos EUA. A base permite uma análise detalhada sobre como as redes sociais influenciaram a percepção pública e o discurso em torno da candidatura de Joe Biden, fornecendo insights sobre a interação dos eleitores naquele período.

DE-R (DIAGRAMA ENTIDADE-RELACIONAMENTO)



DLD (DIAGRAMA LÓGICO DE DADOS)



DICIONÁRIO DE DADOS

Entidade: USER_DIMENSION				
Descrição: Contém os dados relacionados aos usuários do X (antigo twitter)				
Atributo	Propriedades do atributo	Tipo de Dado	Tamanho	Descrição
user_id	chave primária obrigatório	BIGINT	11	Número que identifica unicamente o usuário no twitter
name	obrigatório	VARCHAR	200	Nome que aparece na página de perfil
screen_name	obrigatório	VARCHAR	200	Nome único de usuário
description	obrigatório	VARCHAR	300	Descrição do usuário
followers_count	obrigatório	INT	4	Quantidade de seguidores no twitter
location	obrigatório	TINYTEXT	-	Localização do usuário

Entidade: STATE_DIMENSION				
Descrição: Contém os Estados onde os tweets foram feitos.				
Atributo	Propriedade do atributo	Tipo de Dado	Tamanho	Descrição
state_code	chave primária obrigatório	VARCHAR	5	Código do estado
state_name	chave primária obrigatório	VARCHAR	255	Nome do estado

Entidade: LOCATION_DIMENSION				
Descrição: Contém dados da localização dos tweets				
Atributo	Propriedade do atributo	Tipo de Dado	Tamanho	Descrição
latitude	Chave primária obrigatório	DOUBLE	-	Coordenada da latitude onde foi feito o tweet

longitude	Chave primária obrigatório	DOUBLE	-	Coordenada da longitude onde foi feito o tweet
city	obrigatório	TINYTEXT	-	Nome da cidade onde foi feito o tweet
country	obrigatório	TINYTEXT	-	Nome do país onde foi feito o tweet
continent	obrigatório	VARCHAR	255	Nome do continente onde foi feito o tweet
state_code	Chave estrangeira obrigatória	VARCHAR	5	Código do estado onde foi feito o tweet
state_name	Chave estrangeira obrigatória	VARCHAR	255	Nome do estado onde foi realizado o tweet

Entidade: SOURCE_DIMENSION				
Descrição: Contém dados dos aparelhos utilizados para realizar os tweets				
Atributo	Propriedade do atributo	Tipo de Dado	Tamanho	Descrição
source_id	Chave primária obrigatório	INT	4	Número de identificação do aparelho utilizado para fazer o tweet.
name	obrigatório	VARCHAR	100	Nome do aparelho utilizado para fazer o tweet.

Entidade: FACT_TWEET				
Descrição: Contém os dados relacionados aos tweets com as hashtags #JoeBiden e #Biden.				
Atributo	Propriedade do atributo	Tipo de Dado	Tamanho	Descrição
tweet_id	Chave primária obrigatória	BIGINT	11	Número que identifica

				unicamente o tweet
created_at	obrigatório	DATETIME		Data e hora de quando o tweet foi criado
message	obrigatório	VARCHAR	1000	Mensagem que foi escrita no tweet
likes	obrigatório	INT	4	Número de likes do tweet
retweet_count	obrigatório	INT	4	Número de compartilhamentos do tweet
source_id	Chave estrangeira obrigatória	INT	4	Número de identificação do aparelho utilizado para realizar o tweet
user_id	Chave estrangeira obrigatória	BIGINT	11	Número de identificação do usuário que fez o tweet
latitude	Chave estrangeira obrigatória	DOUBLE	-	Latitude de onde foi feito o tweet
longitude	Chave estrangeira obrigatória	DOUBLE	-	Longitude de onde foi feito o tweet
collected_at	optativo	DATETIME	-	Data e hora de quando os dados do tweet foram extraídos

Visão: TWEETS_BY_STATE				
Descrição: Dados relacionados a quantidade de tweets por estado				
Atributo	Propriedade do atributo	Tipo de Dado	Tamanho	Descrição
state_name	Chave primária obrigatório	VARCHAR	255	Nome do estado onde foram realizados os tweets

tweet_count	Derivado	BIGINT	-	Soma da quantidade de tweets registrados em cada estado
-------------	----------	--------	---	---

Visão: TOP_NEWSPAPER_PROFILES				
Descrição: Contém dados dos 20 perfis jornalísticos com maior engajamento (soma do total de likes com total de retweets).				
Atributo	Propriedade do atributo	Tipo de Dado	Tamanho	Descrição
name	obrigatório	VARCHAR	200	Nome que aparece na página de perfil
screen_name	obrigatório	VARCHAR	200	Nome de usuário único.
followers_count	obrigatório	INT	-	Total de seguidores do perfil no twitter
total_Likes	Derivado de likes	DECIMAL	(32,0)	Total de likes nas publicações que possuíam as hashtags #JoeBiden e #Biden
total_Retweets	Derivado de retweet_count	DECIMAL	(32,0)	Total de retweets nas publicações que possuíam as hashtags #JoeBiden e #Biden
engajamento_Total	Derivado de likes e retweet_count	DECIMAL	(33,0)	Soma do total_Likes + total_Retweets