Database Structure Proposal for Storing Tweets from CommBank Twitter Account

List of Tables

Users

Tweets

Replies

QuoteRetweets

Mentions

Items to Include in Each Data Set

1. Users Table

user_id (Primary Key): Unique identifier for each user.

username: The Twitter handle of the user.

display_name: The display name of the user.

bio: The bio of the user.

profile_image_url: URL of the user's profile image.

created_at: The date when the user account was created.

2. Tweets Table

tweet_id (Primary Key): Unique identifier for each tweet.

user_id (Foreign Key): The user who posted the tweet.

content: The content of the tweet.

created_at: The date and time when the tweet was posted.

retweet_count: Number of retweets.

like_count: Number of likes.

reply_count: Number of replies.

quote_count: Number of quote retweets.

3. Replies Table

reply_id (Primary Key): Unique identifier for each reply.

tweet_id (Foreign Key): The original tweet that is being replied to.

user_id (Foreign Key): The user who posted the reply.

content: The content of the reply.

created_at: The date and time when the reply was posted.

4. QuoteRetweets Table

quote_id (Primary Key): Unique identifier for each quote retweet.

tweet_id (Foreign Key): The original tweet that is being quoted.

user_id (Foreign Key): The user who posted the quote retweet.

content: The content of the quote retweet.

created_at: The date and time when the quote retweet was posted.

5. Mentions Table

mention_id (Primary Key): Unique identifier for each mention.

tweet_id (Foreign Key): The tweet in which the user is mentioned.

mentioned_user_id (Foreign Key): The user who is mentioned.

created_at: The date and time when the mention was made.

Primary Keys for Each Data Set

Users: user_id

Tweets: tweet_id

Replies: reply_id

QuoteRetweets: quote_id

Mentions: mention_id

Relationships Between Tables

Users and Tweets

One-to-Many: One user can have many tweets.

Relationship Field: user_id in Tweets references user_id in Users.

Tweets and Replies

One-to-Many: One tweet can have many replies.

Relationship Field: tweet_id in Replies references tweet_id in Tweets.

Tweets and QuoteRetweets

One-to-Many: One tweet can have many quote retweets.

Relationship Field: tweet_id in QuoteRetweets references tweet_id in Tweets.

Tweets and Mentions

One-to-Many: One tweet can have many mentions.

Relationship Field: tweet_id in Mentions references tweet_id in Tweets.

Users and Replies

One-to-Many: One user can post many replies.

Relationship Field: user_id in Replies references user_id in Users.

Users and QuoteRetweets

One-to-Many: One user can post many quote retweets.

Relationship Field: user_id in QuoteRetweets references user_id in Users.

Users and Mentions

One-to-Many: One user can be mentioned in many tweets.

Relationship Field: mentioned_user_id in Mentions references user_id in Users.

This structure ensures that we capture all necessary data points while maintaining the integrity and relationships of the data. This approach will help in querying and analyzing the data effectively for business insights.