

Business Intelligence

3ème année LNBIS

Mini Projet

Institut Supérieur de Gestion de Tunis

Sliti Manel

Nous désirons stocker les tweets des utilisateurs afin de procéder à une analyse des sujets tendances sur le web. À cette fin, nous capturons le flux Twitter et stockons les informations des tweets intéressants dans un entrepôt de données.

Pour chaque tweet que nous stockons : l'heure et la date ; le message lui-même ; les hashtags (par exemple #DataWarehousing) ; mentions d'utilisateurs (par exemple @dreamer) ; s'il s'agit de retweets ou non, et si oui, de quel tweet ; et l'utilisateur qui a envoyé le tweet.

De plus, nous effectuons un traitement du langage naturel sur les tweets, ce qui permet de caractériser le sentiment du tweet (euphorique, heureux, neutre, triste, en colère et scandaleux sont les valeurs possibles). Pour chaque utilisateur, nous disposons du nom complet (tel qu'enregistré par Twitter). Pour certains utilisateurs, nous disposons aussi des caractéristiques démographiques, de la tranche d'âge, du sexe et de la nationalité.

Chaque tweet est associé à exactement un utilisateur et un sentiment, mais peut avoir 0 ou plusieurs mentions et 0 ou plusieurs hashtags. Un tweet peut être un retweet d'exactly un autre tweet, ou être un tweet original.

Le datawarehouse stockant les tweets sera ensuite utilisé pour répondre à des questions analytiques, comme (liste non exhaustive !) :

- Pour le mois de décembre 2023, donner pour chaque catégorie de sentiment le nombre de tweets qui contenaient # ceasefire ;
- Donner par mois le nombre de tweets contenant le hashtag #gaza;
- Indiquer par mois combien de fois les tweets de l'utilisateur @alquds sont retweetés en moyenne.

Travail à faire :

1. Identifiez le sujet d'analyse principal de l'entrepôt de données
2. Proposez une modélisation multidimensionnelle de l'ED, en précisant pour chaque table sa nature dimensionnelle (table de faits ou table de dimension), ses clés, ses attributs, ainsi que la nature des champs.

3. Identifiez le type du schéma de donnée obtenu.

Partie power BI

1. Créez la table de faits et des dimensions manuellement dans Power BI en utilisant l'option « Enter Data »
2. **Configurez des relations entre les tables , Allez dans Model (vue modèle) pour gérer les relations :**

- Faits_Tweets[Date_ID] → Date[Date_ID] (relation un-à-plusieurs).
- Faits_Tweets[ID_Utilisateur] → Utilisateur[ID_Utilisateur] (relation un-à-plusieurs).
- Faits_Tweets[ID_Sentiment] → Sentiment[ID_Sentiment] (relation un-à-plusieurs).
- Faits_Tweets[ID_Hashtag] → Hashtag[ID_Hashtag] (relation un-à-plusieurs).

3. Ajout des colonnes calculées

3.1 Nom complet de l'utilisateur (concatenation) :

- Dans la table Utilisateur, créez une nouvelle colonne :

Nom_Complet_Concat = CONCATENATE(Utilisateur[Nom_Complet], " (" & Utilisateur[Sexe] & ")")

3.2 Type du tweet (original ou retweet) :

- Dans Faits_Tweets, ajoutez :

Type_Tweet = IF(Faits_Tweets[Est_Retweet] = TRUE, "Retweet", "Original")

4. Mesures calculées :

4.1. Nombre total de tweets :

- Ajoutez une mesure :

Nb_Tweets = COUNT(Faits_Tweets[ID_Tweet])

4.2 Tweets contenant un hashtag spécifique (#ceasefire) :

- Ajoutez une mesure :

Tweets_Hashtag_Ceasefire = CALCULATE(COUNT(Faits_Tweets[ID_Tweet]),
Hashtag[Hashtag_Texte] = "#ceasefire")

5. Création du rapport dans Power BI

Passez à l'onglet Report pour concevoir des visuels interactifs.

Visualisation 1 : Histogramme des sentiments

- Visual : Histogramme empilé.
- Axes :
- Catégorie X : Sentiment[Sentiment_Description].
- Valeurs Y : Nb_Tweets_Sentiment.
- Filtre : Configurez le filtre pour inclure uniquement les tweets de décembre 2023.

Visualisation 2 : Graphique en courbes (Tweets avec #gaza)

- Visual : Ligne.
- Axes :
- X : Date[Mois].
- Y : Mesure Tweets_Hashtag_Gaza.

Notez qu'il n'existe pas une seule solution correcte et que plusieurs faits peuvent être nécessaires. Pour chaque fait de votre modèle, expliquez succinctement ce qu'il représente. Toutes les informations contenues dans la description ne sont pas nécessairement pertinentes pour le modèle de faits dimensionnel. Votre modèle doit représenter la description le plus fidèlement possible.