

Data Visualization & Storytelling : Analyse des contenus les plus populaires de Social Buzz

Réalisé par : Blandine JATSA NGUETSE

Contexte du projet	1
Processus de gestion du projet	2
Résultats	3
Conclusion et Recommandations	4

Contexte du projet

Social Buzz est une licorne technologie du domaine des médias sociaux. Elle a été créée en 2008 par deux anciens ingénieurs. Social Buzz est une nouvelle plateforme où le contenu occupe une place centrale.

Elle met l'accent sur le contenu en préservant l'anonymat de tous les utilisateurs et en suivant uniquement les réactions des utilisateurs sur chaque élément de contenu. Il existe plus de 100 façons pour les utilisateurs de réagir à un contenu, au-delà des réactions traditionnelles de type "j'aime", "je n'aime pas" et "commentaires", ce qui permet de s'assurer que le contenu en vogue, et non les utilisateurs individuels, est au premier plan des flux d'utilisateurs.

Au cours des cinq dernières années, Social Buzz a atteint plus de 500 millions d'utilisateurs actifs chaque mois. L'entreprise s'est développée plus rapidement que prévu et a besoin de l'aide d'une société de conseil pour superviser efficacement son processus de développement. **Autrement dit, Social buzz a besoin de s'adapter rapidement à sa croissance rapide.**

En raison de sa croissance rapide et de la nature numérique de son produit principal, la quantité de données qu'elle crée, recueille et doit analyser est énorme. Chaque jour, plus de 100 000 éléments de contenu, qu'il s'agisse de textes, d'images, de vidéos ou de GIF, sont publiés.

Avec :

- Plus de Plus de 100 000 messages par jour
- 36 500 000 contenus par an

Social Buzz veut apprendre les meilleures pratiques en matière de Big Data.

Mais comment en tirer profit quand il y en a autant ?

Pour cela, nous allons analyser les 5 catégories de contenu les plus populaires de Social Buzz. Notre tâche consiste à identifier les catégories de contenu les plus populaires. La popularité est déterminée en fonction des scores de réaction

Processus de gestion du projet

- Compréhension des données

Nous avons reçu 7 data sets et le modèle de données. La première étape consiste donc à utiliser ce modèle de données pour identifier les data sets qui seront nécessaires pour répondre à votre question commerciale, à savoir déterminer les 5 catégories les plus populaires.

Nous avons la table :

- User

ID: unique de l'utilisateur (généré automatiquement)

Name: Nom complet de l'utilisateur

Email: Adresse électronique de l'utilisateur

- Profile

User ID: ID unique d'un utilisateur qui existe dans la table des utilisateurs

Interests: Intérêts de l'utilisateur associé

Age: Âge de l'utilisateur associé

- Location

User ID: ID unique d'un utilisateur existant dans la table des utilisateurs

Address: Adresse complète de l'utilisateur

- Session

User ID: ID unique d'un utilisateur existant dans la table des utilisateurs

Device: Appareil mobile utilisé pour cette session sur l'application

Duration: Durée en minutes pendant laquelle cet utilisateur est resté actif sur l'application au cours de cette session

- Content

ID: ID unique du contenu qui a été téléchargé (généré automatiquement)

User ID: ID unique d'un utilisateur qui existe dans la table des utilisateurs

Type: Chaîne indiquant le type de contenu qui a été téléchargé : Chaîne détaillant la catégorie à laquelle appartient ce contenu
URL : Lien vers l'emplacement où ce contenu est stocké

- Reaction

Content ID: ID unique d'un élément de contenu qui a été téléchargé

User ID: ID unique d'un utilisateur existant dans la table des utilisateurs qui a réagi à ce contenu.

Type: Chaîne de caractères décrivant le type de réaction de l'utilisateur.

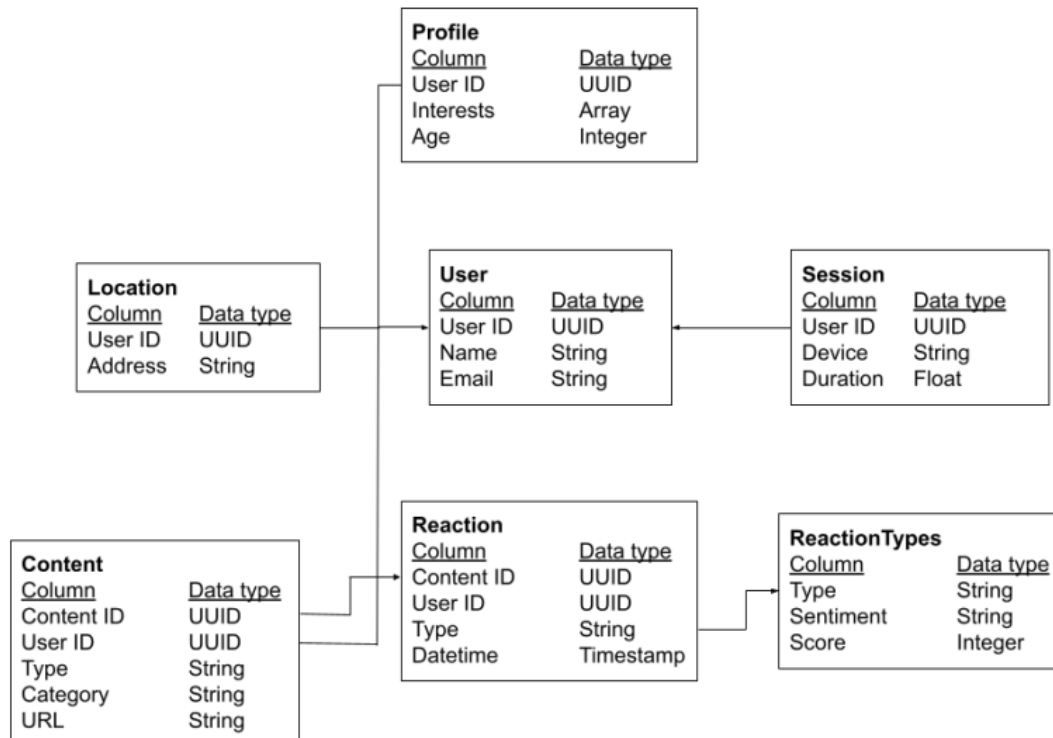
Datetime: La date et l'heure de cette réaction

- ReactionTypes

Type: Une chaîne de caractères détaillant le type de réaction de cet utilisateur.

Sentiment: Chaîne indiquant si ce type de réaction est considéré comme positif, négatif ou neutre.

Score: Il s'agit d'un nombre calculé par Social Buzz qui quantifie la "popularité" de chaque réaction. Un type de réaction ayant un score plus élevé doit être considéré comme une réaction plus populaire.



modèle de données

- Prétraitement des données

Réaction, contenu, type de réaction sont trois datasets qui seront nécessaires et pertinents pour compléter notre analyse. En effet, le client souhaitait voir "une analyse de leurs catégories de contenu montrant les 5 catégories les plus populaires".

Comme l'explique le modèle de données, la popularité est quantifiée par le "score" attribué à chaque type de réaction.

Nous avons donc besoin de données indiquant l'ID du contenu, la catégorie, le type de contenu, le type de réaction et le score de réaction.

Ainsi, pour déterminer la popularité, nous devons additionner les catégories de contenu qui ont le score le plus élevé.

Nous avons Nettoyé les données en supprimant les lignes dont les valeurs sont manquantes.

Nous avons modifier le type de données de certaines valeurs d'une colonne, et supprimer les colonnes qui ne sont pas pertinentes pour cette tâche et pour la question commerciale que nous étudions.

- Modélisation de données

Pour la modélisation des données, nous avons suivi les étapes suivantes :

1. Créez un ensemble de données final en fusionnant les trois tables : Formule utilisée RECHERCHEV (VLookUp).

2. Déterminez les 5 catégories les plus performantes en additionnant les scores totaux de chaque catégorie, Formule utilisée SOMME.SI (Sum If).

Le résultat final est une feuille de calcul qui contient un ensemble de données nettoyé et les 5 catégories les plus performantes.

Task 3_Final Content Data set.csv - LibreOffice Calc

Fichier Édition Affichage Insertion Format Styles Feuille Données Outils Fenêtre Aide

0,00 cm Aucun(e)

	B	C	D	E	F	G	H
	Content ID	Reaction Type	Datetime	Content Type	Category	Sentiment	Score
1	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	7/11/20 9:43	photo	studying	negative	0
2	97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	17/6/21 12:22	photo	studying	negative	10
3	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	18/4/21 5:13	photo	studying	negative	15
4	97522e57-d9ab-4bd6-97bf-c24d952602d2	disgust	6/1/21 19:13	photo	studying	negative	0
5	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	23/8/20 12:25	photo	studying	positive	30
6	97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	7/12/20 6:27	photo	studying	neutral	35
7	97522e57-d9ab-4bd6-97bf-c24d952602d2	cherish	11/4/21 17:35	photo	studying	positive	70
8	97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	27/1/21 8:32	photo	studying	negative	5
9	97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	1/4/21 22:54	photo	studying	neutral	35
10	97522e57-d9ab-4bd6-97bf-c24d952602d2	love	4/8/20 5:05	photo	studying	positive	65
11	97522e57-d9ab-4bd6-97bf-c24d952602d2	indifferent	7/11/20 8:36	photo	studying	neutral	20
12	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	2/11/20 6:28	photo	studying	negative	15
13	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	1/11/20 1:16	photo	studying	positive	30
14	97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	7/10/20 18:39	photo	studying	negative	5
15	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	3/9/20 18:51	photo	studying	negative	15
16	97522e57-d9ab-4bd6-97bf-c24d952602d2	super love	24/2/21 5:09	photo	studying	positive	75
17	97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	23/9/20 6:24	photo	studying	neutral	35
18	97522e57-d9ab-4bd6-97bf-c24d952602d2	indifferent	24/2/21 11:37	photo	studying	neutral	20
19	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	22/5/21 19:44	photo	studying	positive	30
20	97522e57-d9ab-4bd6-97bf-c24d952602d2	intrigued	31/1/21 16:03	photo	studying	positive	45
21	97522e57-d9ab-4bd6-97bf-c24d952602d2	peeking	20/11/20 17:26	photo	studying	neutral	35
22	97522e57-d9ab-4bd6-97bf-c24d952602d2	worried	11/4/21 8:47	photo	studying	negative	12
23	97522e57-d9ab-4bd6-97bf-c24d952602d2	like	13/6/21 16:46	photo	studying	positive	50
24	97522e57-d9ab-4bd6-97bf-c24d952602d2	heart	11/4/21 14:29	photo	studying	positive	60
25	97522e57-d9ab-4bd6-97bf-c24d952602d2	worried	2/3/21 19:21	photo	studying	negative	12
26	97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	27/6/20 18:44	photo	studying	negative	5
27	97522e57-d9ab-4bd6-97bf-c24d952602d2	hate	9/5/21 21:05	photo	studying	negative	5
28	97522e57-d9ab-4bd6-97bf-c24d952602d2	cherish	4/8/20 23:58	photo	studying	positive	70
29	97522e57-d9ab-4bd6-97bf-c24d952602d2	super love	28/8/20 3:57	photo	studying	positive	75
30	97522e57-d9ab-4bd6-97bf-c24d952602d2	intrigued	25/8/20 8:21	photo	studying	positive	45
31	97522e57-d9ab-4bd6-97bf-c24d952602d2	scared	20/4/21 23:48	photo	studying	negative	15
32	97522e57-d9ab-4bd6-97bf-c24d952602d2	interested	7/3/21 12:40	photo	studying	positive	30
33	97522e57-d9ab-4bd6-97bf-c24d952602d2	dislike	17/6/21 12:22	photo	studying	negative	10

Task 3_Final Content Data set Top 5 popular Category

- Analyse des données

Nous avons utilisé Excel pour produire les graphiques qui répondent aux questions suivantes :

Combien y a-t-il de catégories uniques ?

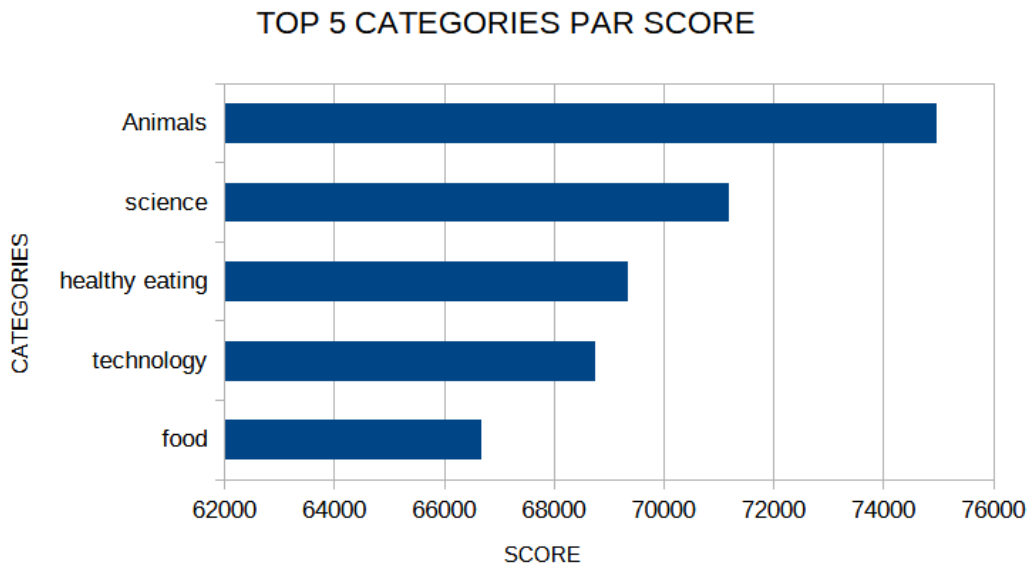
Combien de réactions à la catégorie la plus populaire ?

Quel est le mois où il y a eu le plus de messages ?

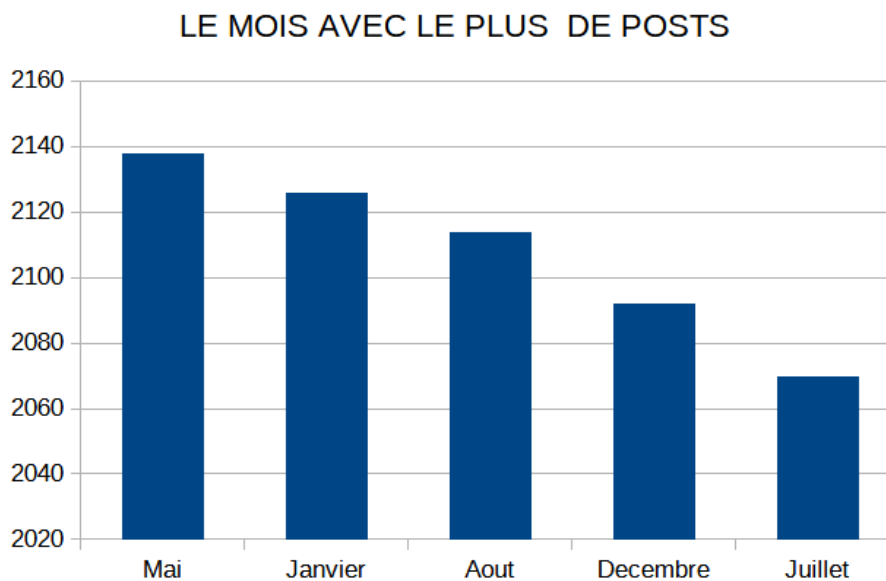
- Présentation des résultats

Résultats

Social Buzz propose 16 réactions de contenus : **heart, want, disgust, hate, interested, indifferent, love, super love, cherish, adore, like, dislike, intrigued, peeking, scared, worried.**



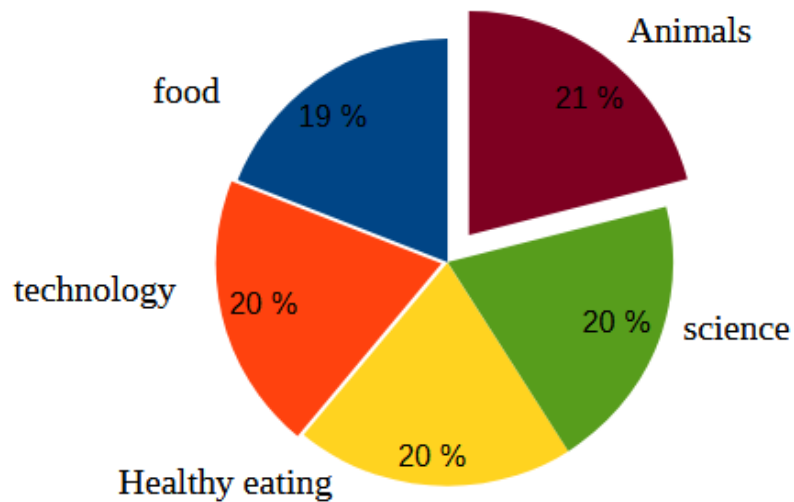
Pour la catégorie “Animals”, nous avons 1897 réactions, soit 7,7% du total des réactions. Les animaux et la science sont les deux catégories de contenu les plus populaires, montrant que les gens apprécient le contenu "réel" et "factuel" le plus.



Mai est le mois avec le plus grand nombre de messages. Nous pouvons pousser la réaction plus loin, y'a t'il un événement particulier qui poussent les clients à réagir à des pages? Cela pourrait vous aider à planifier des campagnes promotionnelles ou des activités spécifiques pendant cette période.

Nous pouvons aussi analyser les interactions spécifiques avec les clients en mai. Y a-t-il eu une augmentation des commentaires, des réponses aux sondages, ou d'autres formes d'engagement ? Cela pourrait indiquer une opportunité de renforcer la relation client tout au long de l'année.

POURCENTAGE DE POPULARITÉ DES 5 PREMIÈRES CATÉGORIES



La nourriture est un thème courant parmi les cinq principales catégories, avec "une alimentation saine" en tête. Cela peut donner une indication à l'audience au sein de votre base d'utilisateurs. Vous pourriez utiliser cette insight pour créer une campagne et collaborer avec des marques axées sur une alimentation saine afin de stimuler l'engagement des utilisateurs.

Conclusion et Recommandations

Cette analyse ad hoc est perspicace, mais il est temps de passer à une production à grande échelle pour une compréhension en temps réel de l'entreprise social buzz.