

Objectif du test

Contexte d'affaires

Descriptions des champs disponibles

- **abo.csv**
- **visionnements.csv**
- **cms.csv**

Objectif du test

Votre mission, dans le cadre de ce test, est de réaliser un projet de segmentation libre à partir des données fournies. Ce projet vous permettra de démontrer vos compétences en exploration de données, pré-traitement, modélisation et en interprétation des résultats.

Contexte d'affaires

Chez Radio-Canada, nous souhaitons améliorer le funnel de conversion pour le compte RC en comprenant mieux le comportement de nos utilisateurs. Pour ce faire, nous menons une analyse de segmentation afin d'identifier des tactiques d'optimisation du parcours de visionnement et de mieux cerner les différents profils.

Détails de la mission

1. Explorer et préparer les données

- Charger et analyser les deux fichiers, en identifiant les relations entre les contenus et les visionnements.
- Effectuer un nettoyage et une transformation des données (gestion des valeurs manquantes, agrégation éventuelle, etc.).

2. Définir et construire les variables de segmentation

- Vous pouvez vous inspirer des variables liées au comportement des utilisateurs (nombre de visites, pages vues, temps passé, consommation audio/vidéo, nombre de sections consultées, etc.).

3. Analyse de segmentation

- Appliquer une ou plusieurs techniques de clustering afin de segmenter les visiteurs en groupes.

4. Interpréter les segments

- Analyser les caractéristiques principales de chaque segment et discuter de leur signification en termes de comportement utilisateur.

5. Communication des résultats

- Présenter vos analyses, résultats et interprétations dans un rapport clair et synthétique.

Descriptions des champs disponibles

abo.csv

Identification de la date de début et de fin d'abonnement des clients ayant réalisé au moins un visionnement en 2019. L'équipe Tou.TV nous indique que TOUS les clients ayant fait un visionnement en 2019 devraient se retrouver dans ce fichier.

rcid_hash : identifiant du compte abonné (seulement disponible quand l'utilisateur est abonné et identifié sur la plateforme)

subscribe_on : date de début de l'abonnement

cancelled_on : date de fin de l'abonnement (vide si l'abonné est encore abonné)

visionnements.csv

Identification de tous les visionnements ayant été réalisés en 2019. Le fichier étant volumineux (plus de 800 Mo), tu peux utiliser le fichier de 100 000 lignes afin de faire tes premiers tests.

date : jour du visionnement

rcid_hash : identifiant du compte Radio-Canada sous lequel l'utilisateur est authentifié lorsqu'il réalise son visionnement. Pour utiliser un abonnement Extra, un utilisateur doit être sur le site en mode authentifié avec son compte Radio-Canada. Ce champ devrait être nul ou avec la valeur « anonyme » lorsque l'utilisateur n'est pas authentifié lors de son visionnement

visitor_id_hash : identifiant de la machine selon Adobe Analytics

statut_connexion : lors du visionnement, l'utilisateur était-il authentifié (True/False)

titre : concaténation des valeurs émissions:saisons:numéro_épisode

modele : statut du contenu lorsqu'il est visionné:

- Valeurs « payant » ou « extra » : contenu qui pour cette date ne peut être visionné que par les abonnés extra
- Valeur « gratuit » : contenu qui pour cette date peut être visionné par tous

videoinitiate : nombre de visionnement réalisés

content_time_spent : durée du visionnement réalisé (en seconde)

enchainement : est-ce que le visionnement de ce contenu fait suite à un enchaînement automatique depuis un précédent épisode ?

reprise_media : est-ce que le visionnement de ce contenu est une reprise d'un précédent visionnement ?

type_declenchement : est-ce que la vidéo a été déclenché automatiquement ou suite à un clic sur un bouton play ?

progress_marker_75_percent : est-ce que l'utilisateur a franchi l'étape de 75 % de la durée du contenu

progress_marker_95_percent : est-ce que l'utilisateur a franchi l'étape de 90 % de la durée du contenu

cms.csv

Identification du thème et de l'audience pour tous les contenus disponibles sur ICI TOU.TV

emission : le nom de l'émission (devrait correspondre à la valeur émission contenue dans le champ titre de la table visionnements.csv)

Theme : le thème/genre principal de l'émission

Audience : quel est le public cible de l'émission. Les contenus jeunesse sont identifiés par les valeurs « pour les petits », « pour les plus grands » et « ados »