

Comprendre les données du défi 2

Guide de prise en main · STT-4230 / STT-6230

Le contexte

Une coopérative canadienne fictive achète trois lots de marchandises provenant des États-Unis. Un lot est un ensemble de produits auquel on associe une valeur totale en dollars canadiens. Pour prévoir le coût de sa commande, la coopérative veut connaître la surtaxe qui s'ajouterait à la valeur des marchandises.

Dans cette simulation, une surtaxe est un montant calculé en appliquant un taux en pourcentage à la valeur du lot. Le taux à retenir dépend du produit et de la date étudiée. La coopérative souhaite donc comparer la même commande avant et après un changement de tarifs.

Les produits et leurs valeurs restent les mêmes aux deux dates. On cherche à comprendre comment le changement de tarifs modifie la surtaxe totale de cette même commande.

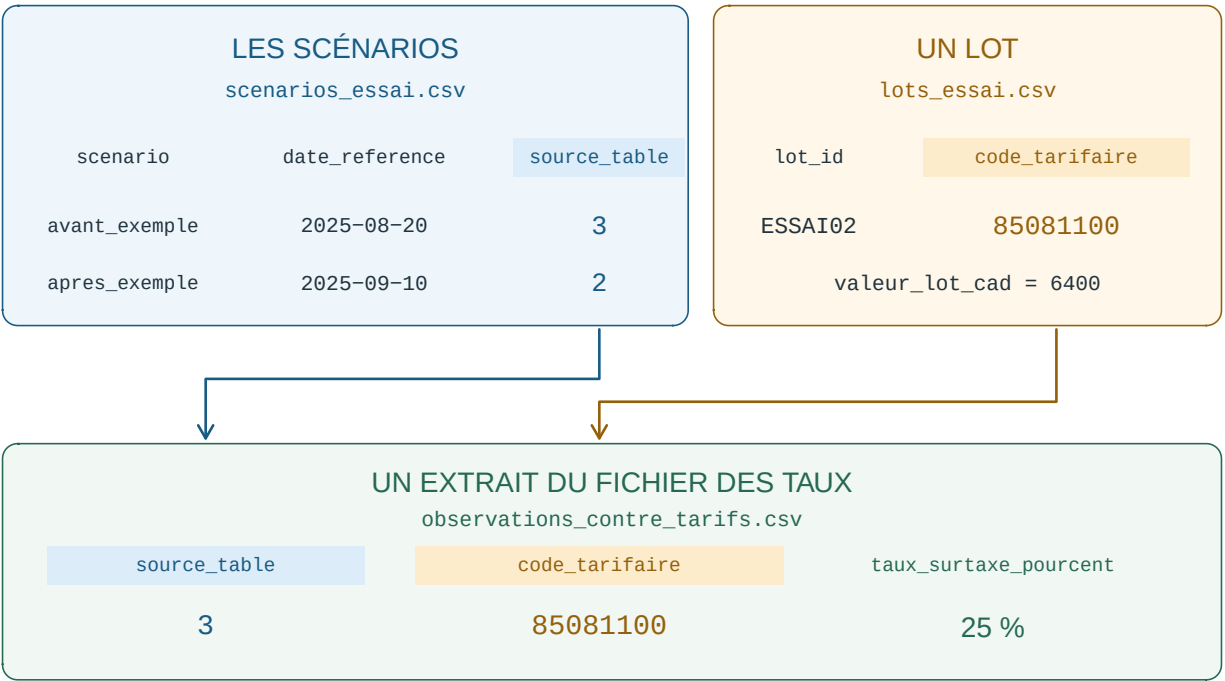
Les lots, les valeurs et les scénarios d'essai sont fictifs. Ils servent à comprendre les fichiers. Les lots et les scénarios de l'évaluation seront différents et vous seront fournis lors du défi.

Les trois fichiers

Chaque fichier répond à une question différente : quelles marchandises importe-t-on, à quelles dates les observe-t-on et quels taux faut-il utiliser ?

Fichier	Ce qu'il contient	Une ligne représente
lots_essai.csv	Les produits de la commande et leurs valeurs	Un lot de marchandises
scenarios_essai.csv	Les deux situations : avant et après	Un scénario daté
observations_contre_tarifs.csv	Les taux pour les produits, selon la période	Un code tarifaire pour une période donnée

Voici des extraits de ces fichiers. Les colonnes en couleur permettent de les relier.



Même nom de colonne, même couleur : les clés de liaison.

`source_table`, en bleu, relie chaque scénario aux lignes de taux qui lui correspondent. `code_tarifaire`, en orange, relie le produit d'un lot au même code dans le fichier des taux.

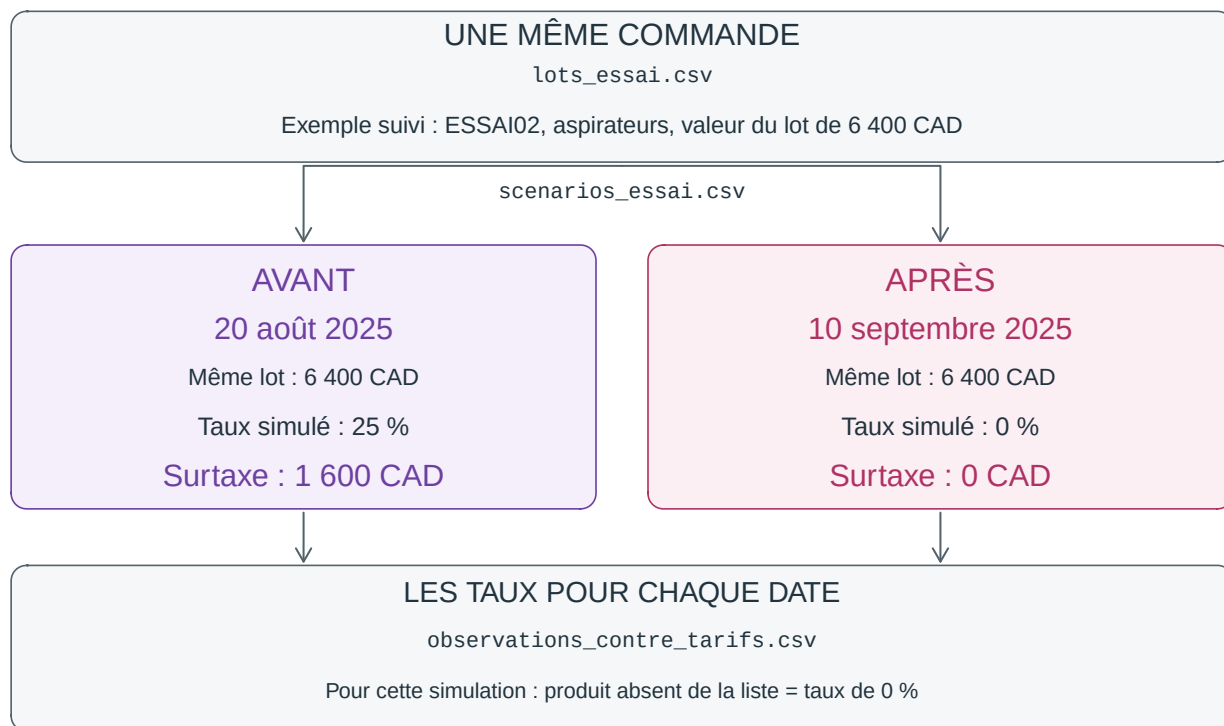
Pour retrouver le taux d'un lot dans un scénario, on utilise ces deux informations ensemble. Le fichier complet contient les autres lignes ; le dessin montre seulement un extrait.

Comparer une même commande à deux dates

Le fichier `scenarios_essai.csv` définit deux situations : « avant », le 20 août 2025, et « après », le 10 septembre 2025. Chaque scénario s'applique aux mêmes trois lots, avec les mêmes valeurs.

Scénario	Date de référence	source_table à utiliser
avant_exemple	2025-08-20	3
apres_exemple	2025-09-10	2

Pour illustrer la comparaison, suivons le lot **ESSAI02** : des aspirateurs d'une valeur totale de 6 400 CAD.



Le 20 août, on cherche le code de l'aspirateur dans le tableau 3, indiqué par le scénario « avant » : il y figure, au taux de 25 %. Le 10 septembre, on le cherche dans le tableau 2, indiqué par le scénario « après » : il n'y figure pas. Pour cette simulation, on lui attribue alors un taux de 0 %.

La surtaxe passe donc de 1 600 CAD à 0 CAD, alors que le lot vaut toujours 6 400 CAD. On compare ainsi deux situations tarifaires pour un même achat.

Quelques repères pour lire les données

Le fichier `lots_essai.csv` contient les trois lots suivants :

Lot	Code tarifaire	Libellé fictif	Valeur (CAD)
ESSAI01	07020010	Tomates, exemple fictif	1250
ESSAI02	85081100	Aspirateurs, exemple fictif	6400
ESSAI03	94013110	Sièges, exemple fictif	875

La valeur porte sur le lot entier. Par exemple, les 6 400 CAD de **ESSAI02** représentent la valeur totale du lot d'aspirateurs.

Conservez les codes comme du texte : le zéro au début de `07020010` fait partie du code. Les codes sont fournis ; vous n'avez pas à classer les marchandises.

Les dates des CSV suivent le format année-mois-jour. Les valeurs des lots sont en dollars canadiens et les taux sont des pourcentages. **NA** indique une information manquante dans le fichier des taux.

`observations_contre_tarifs.csv` contient 2774 lignes et 19 colonnes. Le [dictionnaire complet](#) explique chaque colonne. Certaines descriptions contiennent des retours à la ligne ; un éditeur de texte peut donc afficher plus de lignes que le nombre de lignes lues par R.

Ce que vous recevez

Avant le défi	Lors du défi
Le projet RStudio à ouvrir et à tester	Les lots et les scénarios de l'évaluation
Le fichier des taux et son dictionnaire	L'énoncé et les règles de calcul
Les fichiers d'essai et ce guide	Le rapport QMD à compléter et le code à examiner

Le modèle de déclaration `docs/ai-log.md` est aussi fourni dans votre dépôt. Vous le complétez pendant le défi, même sans IA.

Le code et le rapport évalués seront rédigés pendant le défi. Pour choisir les tableaux, vous pourrez lire le fichier des scénarios ou indiquer les numéros directement dans le code avec un commentaire explicatif. Les deux méthodes sont acceptées. Aucune connaissance douanière n'est nécessaire ; l'énoncé donnera les règles de simulation.

Avant le défi

1. Ouvrez `defi-2.Rproj`, puis les trois CSV. Repérez les lots et les deux dates d'essai.
2. Exécutez la vérification technique et testez le rendu HTML du guide.
3. Testez un commit et un push. Le README donne les étapes.

`guide-donnees.qmd` est la source de ce guide. Lors du défi, vous recevrez un autre QMD, `rapport-defi-2.qmd`, pour rédiger votre rapport.

Sources et limites

Les taux viennent du ministère des Finances du Canada, à partir des données préparées par [Données bleues](#). Leur provenance et les conditions de réutilisation figurent dans `data/SOURCE.md` et `data/DATA_LICENSE.md`.

Les montants du défi seront des simulations. Les données ne suffisent pas à établir une facture douanière réelle.