Documentation: Optimisation Cost file, forecast audience

0. Vue d'ensemble

Ce projet Python de "proof of concept" vise à simplifier le traitement des calculs d'audience et de la manipulation des contrats de vente TV et radio pour l'entreprise Proximus. Cette application Windows prend en entrée des data de références, les centralise et génère des tables Excel raffinées, prêtes à être intégrées en base de données. L'objectif étant de remplacer les processus actuels s'appuyant sur une manipulation laborieuse de multiples feuilles de calculs. Le fonctionnement général de ce programme est le suivant:

- 1. L'utilisateur charge les données de références (actuals de l'audience, cost file, channel et product groupings Proximus).
- 2. Les emplacements de ces documents sont sauvés en configuration, celle-ci est visible depuis le menu Préférences.
- 3. L'utilisateur rentre les réglages voulus pour le forecast de l'audience (date maximale de référence, timespan visé, emplacement d'output, sélection de channels spécifiques ou bien de tout le dataset).
- 4. L'utilisateur dispose de boutons pour produire le forecast, voir le résultat dans Excel, valider que les réglages soient corrects, rafraichir les filtres des channels, inspecter les metadata du fichier d'audience.
- 5. L'onglet Cost offre une interface pour filtrer tous les contrats existants. Les Business models et les types d'allocations sont les deux variables principales à prendre en compte dans le traitement de ces contrats de vente.
- 6. L'utilisateur dispose d'une interface de display, avec une table dynamique adaptant ses colonnes selon le Business model choisi.
- 7. L'interface New Deal s'adapte selon les filtres du display, elle permet de configurer de nouveaux contrats.
- 8. L'onglet Cost génère les nouveaux contrats dans une fichier Excel distinct, séparé d'une feuille par Business model. Le display reprend à la fois les contrats existants du cost file ainsi que les nouveaux contrats produits dans ce dit fichier.

1. Première installation

Pre-requis:

Python

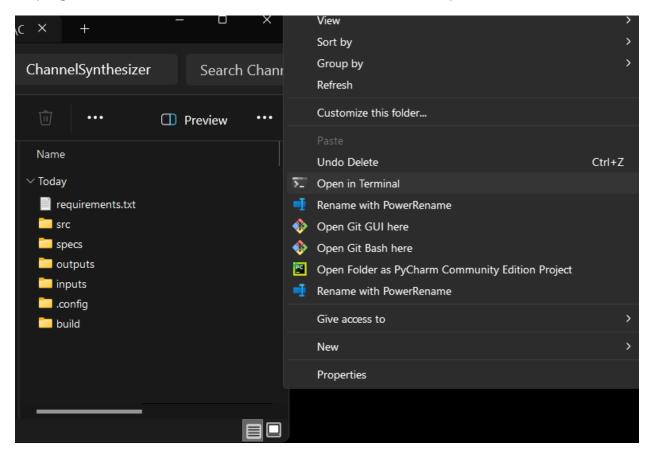
Téléchargez depuis le Microsoft Store à l'adresse suivante: https://www.microsoft.com/store/productId/9NRWMJP3717K?ocid=pdpshare. Pendant le processus d'installation, assurez-vous de cocher l'option Add python.exe to PATH, sinon Python ne sera pas reconnu comme une commande disponible.



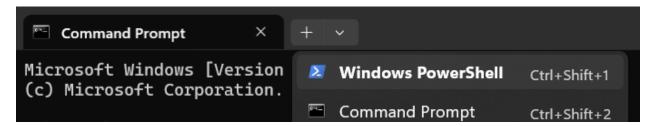
Tester la configuration correcte dans n'importe quel terminal Windows (Command Prompt ou PowerShell), tapez python --version ou python3 --version. L'une de ces commandes devrait retourner un numéro de version. Si vous avez déjà une autre version de Python installée, vous pouvez choisir de ne pas installer la version 3.11. Gardez cependant cela à l'esprit si le projet échoue à se configurer correctement.

Virtual Environment

Utilisez l'Explorateur Windows pour naviguer jusqu'au répertoire contenant tous les dossiers du programme, faites un clic droit à l'intérieur et sélectionnez Open in Terminal.



Dans le menu déroulant, sélectionnez PowerShell, comme ceci:



Maintenant, dans le terminal, copiez et collez ces commandes une par une :

python -m venv env

.\env\Scripts\activate

Si taper la commande pip seule retourne "commande non trouvée" ou "non disponible", vous devez l'installer en utilisant ces deux commandes :

```
Invoke-WebRequest -Uri https://bootstrap.pypa.io/get-pip.py -OutFile get-pip.py python .\get-pip.py
```

Sinon, nous mettons simplement à jour la version de pip de l'environnement virtuel et installons les dépendances :

```
pip install --upgrade pip ou python -m pip install --upgrade pip pip install -r requirements.txt
```

2. Lancer l'application

Ouvrez un terminal PowerShell à l'intérieur de la structure du projet comme dans l'étape 1. Activez l'environnement virtuel nécessaire au projet en collant la commande suivante:

```
.\env\Scripts\activate
python .\src\main.py
```

Assurez-vous qu'aucun fichier Excel impliqué dans cette application ne soit actuellement ouvert. Typiquement, le fichier de référence pour l'audience contient les colonnes suivantes: PERIOD_YEAR, PERIOD_MONTH, DATA_TYPE, PROD_NUM, BUS_CHANL_NUM, LIVE_TV_VIEWING_MINUTES, PVR_VIEWING_MINUTES, CUTV_VIEWING_MINUTES, OTT_VIEWING_MINUTES, VOD_VIEWING_MINUTES, PARK_PER_CHANNEL, LIVE_TV_ACTIVE_USERS, sum_eop_vol_2024,...

Vous devez avoir un fichier de référence Channel Grouping avec une feuille appelée Content_Channel_Grouping contenant une colonne CHANNEL_NAME et une colonne CHANNEL_NAME_GROUP. Le fichier de référence Product Grouping doit disposer d'une feuille de calcul appelée Content_Product_Grouping WS 241 comprenant la colonne PROD_NUM.

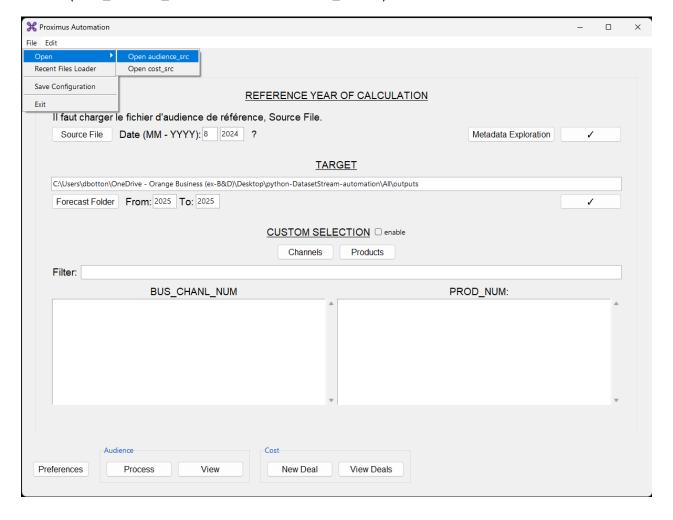
Ces valeurs exactes sont nécessaires pour le processus, il est recommandé de modifier les fichiers d'entrées plutôt que de changer le code source. Vous pouvez retrouver ces nomenclatures écrites en dur dans le fichier src\ui\tab_audience.py. Le cost file de référence doit disposer d'une feuille appelée all contract cost file avec les colonnes NETWORK_NAME, CNT_NAME_GRP, Business model.

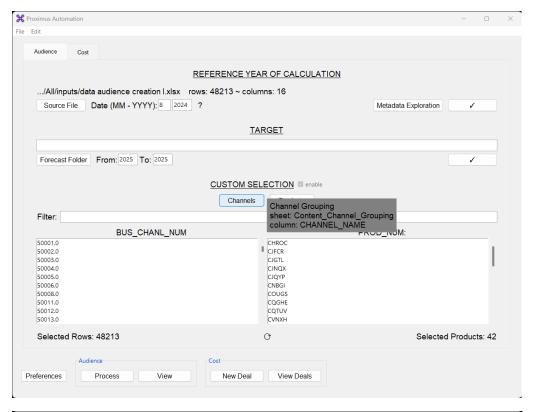
3. Workflow

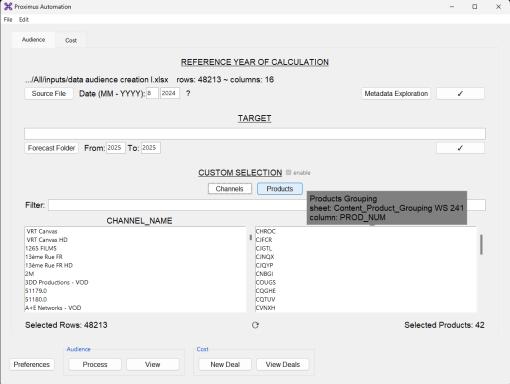
3.1. Configuration des données d'entrées

Il existe plusieurs méthodes pour charger les fichiers de références. Il y a le menu File et les différents boutons contenus dans l'interface (boutons Source File, Channels, Products, Cost File). Des tooltips (aides) apparaissent quand vous placez votre curseur sur ceux-ci, elles indiquent les noms exacts des feuilles et des colonnes essentielles au bon fonctionnement de l'application.

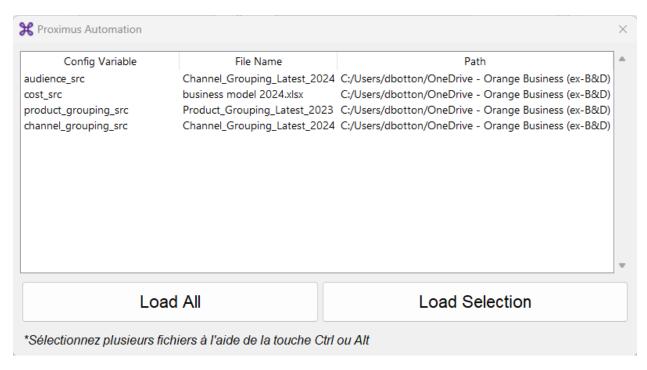
Il est possible de générer le calcul d'audience en spécifiant des channels et produits spécifiques sans charger les Product et Channel Groupings. Quand l'utilisateur charge ces derniers, un remapping des valeurs BUS_CHANL_NUM et PROD_NUM s'effectue (BUS_CHANL_NUM devient CHANNEL_NAME).



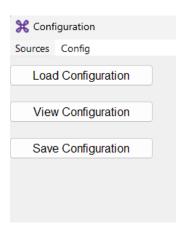




Une fois que tous les fichiers de référence sont chargés une première fois, il est recommandé de redémarrer l'application. Vous profitez ainsi de l'interface de chargement des fichiers récents qui s'inspire du fichier de configuration automatisé situé à l'emplacement .config\config.json.



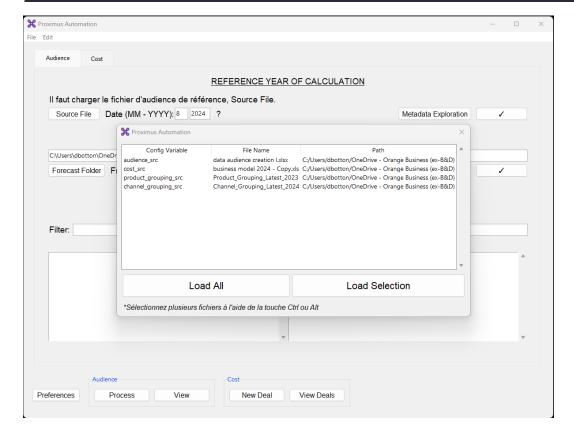
Ce popup apparait au démarrage si un fichier a récemment été chargé dans l'application. Au cours des différents workflows du programme, des données de références doivent être sourcées depuis des fichiers Excels. Les variables audience_src, cost_src, product_grouping_src, channel_grouping_src représentent l'emplacement de ces fichiers. Le bouton Preferences ainsi que le menu Edit permettent d'accéder à l'interface suivante, proposant de sauver, charger et voir la configuration.



La figure suivante est un exemple de configuration complète au format json, elle est bien sur directement modifiable pour plus de flexibilité.

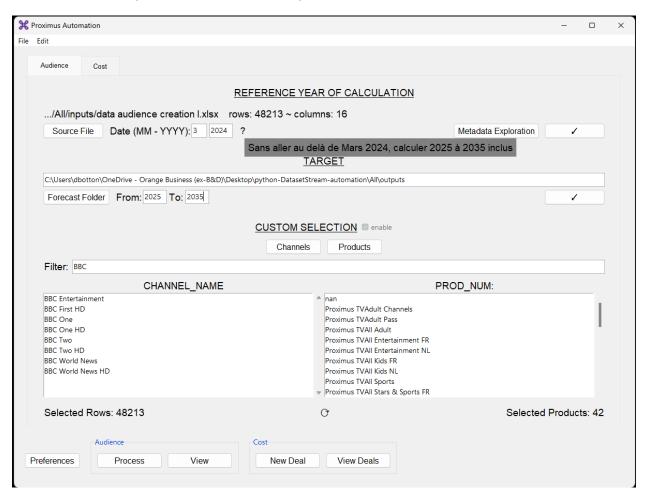
Cette configuration prévoit les cas de figure où elle serait vide, corrompue et gère automatiquement ses updates au fur et à mesure que de nouveaux référence files sont chargés. L'utilisateur n'a pas besoin d'interagir avec celle-ci.

```
"audience_src": "C:/Users/dbotton/OneDrive - Orange Business (ex-B&D)
   /Desktop/python-DatasetStream-automation/All/inputs/data audience creation l.xlsx",
   "audience_dest": "C:\\\\Users\\\\dbotton\\\\OneDrive - Orange Business (ex-B&D)
   \\\\Desktop\\\\python-DatasetStream-automation\\\\All\\\\\outputs",
   "cost_src": "C:/Users/dbotton/OneDrive - Orange Business (ex-B&D)
   /Desktop/python-DatasetStream-automation/All/inputs/business model 2024.xlsx",
   "cost_dest": "C:\\Users\\dbotton\\OneDrive - Orange Business (ex-B&D)
   \\\Desktop\\python-DatasetStream-automation\\All\\outputs",
   "product_grouping_src": "C:/Users/dbotton/OneDrive - Orange Business (ex-B&D)
   /Desktop/python-DatasetStream-automation/All/inputs/Product_Grouping_Latest_20231117.xlsx",
   "channel_grouping_src": "C:/Users/dbotton/OneDrive - Orange Business (ex-B&D)
   /Desktop/python-DatasetStream-automation/All/inputs/Channel_Grouping_Latest_20240607.xlsx"
```



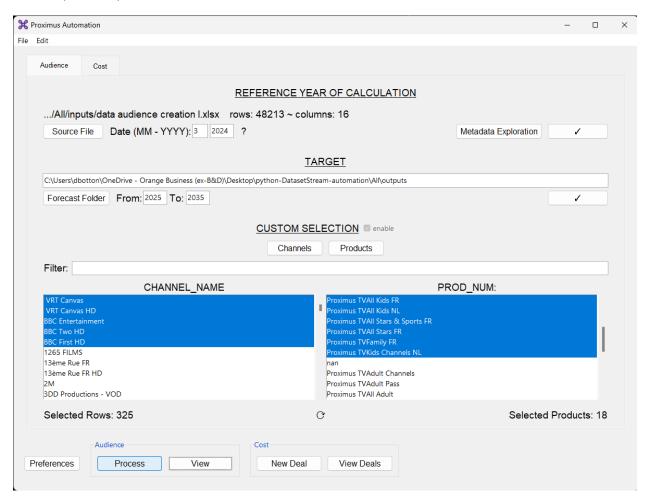
3.2. Calcul de l'audience

La qualité du travail produit est fortement dépendante du contenu des fichiers de références. L'utilisateur va régler quatre facteurs pour le calcul d'audience: la date limite sur laquelle baser les données de référence, la durée cible, les channels et produits spécifiques à inclure dans le forecast. Une aide dynamique explique le comportement attendu selon les dates entrées par l'utilisateur. La figure suivante illustre cette aide, en l'occurrence ici le programme va utiliser les douze derniers mois des actuals, précédant avril 2024, afin de calculer la période 2025 à 2035 en prédiction.



Les channels et produits spécifiques sont filtrables et sélectionnables. La figure suivante illustre ces comportements. Si rien n'est choisi le programme produit les prévisions d'audience pour toutes les entries comprises dans les périodes indiquées par l'utilisateur. Le calcule d'audience est le suivant (cfr. audience source file): les colonnes LIVE_TV_VIEWING_MINUTES, PVR_VIEWING_MINUTES, CUTV_VIEWING_MINUTES, OTT_VIEWING_MINUTES, Sont multipliées à la somme des

volumes de l'année suivante **sum_eop_vol_2025** puis divisées à **sum_eop_vol_2024** (actuals).



Voici les logs générés par le processus de prévision de l'audience avec les configurations présentées par les figures précédentes.

```
References Month: 3, Year: 2024

Target Start Year: 2025, End Year: 2035

File Path: C:/Users/dbotton/OneDrive - Orange Business (ex-B&D)/Desktop/python-DatasetStream-automation/All/inputs/data audience creation l.xlsx

Specifics Enabled: True

Selected BUS_CHANL_NUMs: ['50706.0', '51145.0', '50366.0', '50367.0', '50042.0']

Script Path: C:\Users\dbotton\OneDrive - Orange Business (ex-B&D)\Desktop\python-DatasetStream-automation\All\src\parser\parser_audience.py

Filtering reference data based on provided month and year...

Reference data after initial filter: 5861 rows

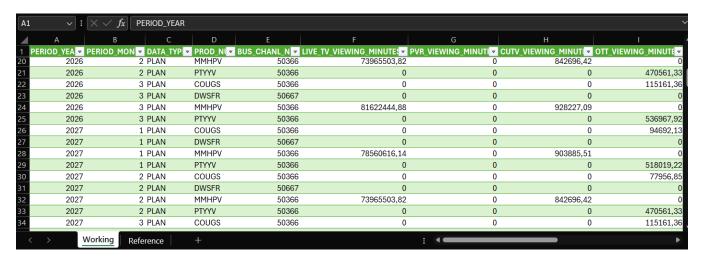
Filtering reference data based on specifics...

Selected PROD_NUMs: None

Selected BUS_CHANL_NUMs: ['50706.0', '51145.0', '50366.0', '50367.0', '50042.0']
```

```
Reference data after specifics filter: 89 rows
Checking for duplicates...
Calculating reference cop volumes...
Starting forecast calculation...
Forecast calculation completed. Total forecast rows: 979
INFO:root:Loading original workbook from C:/Users/dbotton/OneDrive - Orange Business (ex-B&D)/Desktop/python-DatasetStream-automation/All/inputs/data audience creation
INFO:root:Writing data to the Working sheet
INFO:root:Writing data to the Reference sheet
INFO:root:Adjusting column widths and applying styles
INFO:root:Saving workbook to C:/Users\dbotton/OneDrive - Orange Business (ex-B&D)\Desktop\python-DatasetStream-automation\All\outputs\forecast_audience.xlsx
INFO:root:Data saved to C:\Users\dbotton\OneDrive - Orange Business (ex-B&D)\Desktop\python-DatasetStream-automation\All\outputs\forecast_audience.xlsx
Output path: C:\Users\dbotton\OneDrive - Orange Business (ex-B&D)\Desktop\python-DatasetStream-automation\All\outputs\forecast_audience.xlsx
Result file path: C:\Users\dbotton\OneDrive - Orange Business (ex-B&D)\Desktop\python-DatasetStream-automation\All\outputs\forecast_audience.xlsx
```

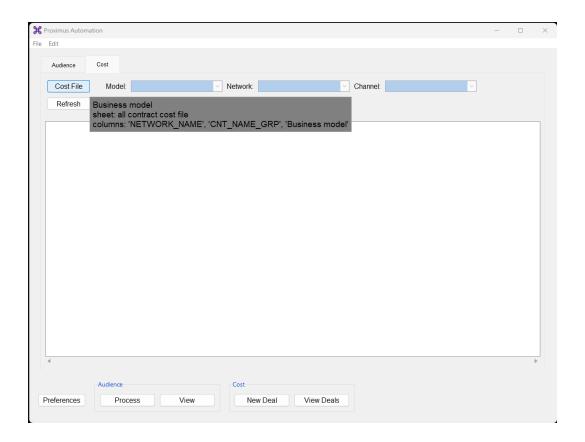
Le fichier Excel généré dispose d'une sheet Working avec les prévisions et d'une seconde sheet Reference qui reprend uniquement les actuals nécessaires au forecast (selon les configurations entrées par l'utilisateur).

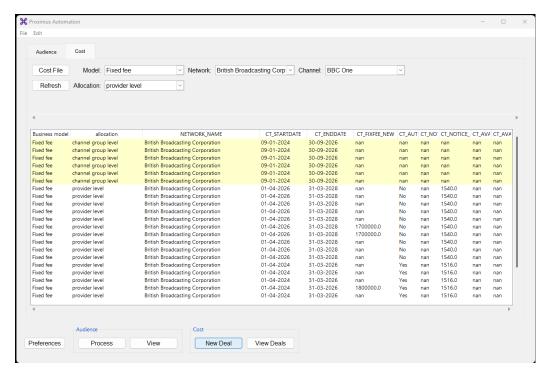


3.3. Optimisation du Cost file

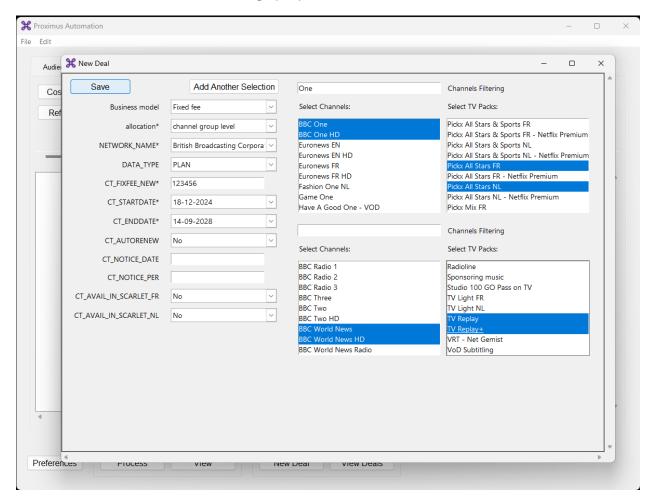
Le deuxième onglet se charge du display des contrats de vente. Si les champs sont bleutés c'est que aucun cost file n'est chargé en mémoire. Il est recommandé de charger tous les fichiers requis par l'application durant son démarrage à l'aide du Recent Files Loader.

Ce fichier Cost File doit comprendre la sheet *all contract cost file* ainsi que les colonnes NETWOK_NAME, CNT_NAME_GRP, Business model etc.. Le programme, dans l'étape Display du cost file, fait un agrégat des données du fichier de référence ainsi que des données du fichier produit. Les entries générées par le projet sont colorées en jaune pour mieux le distinguer.

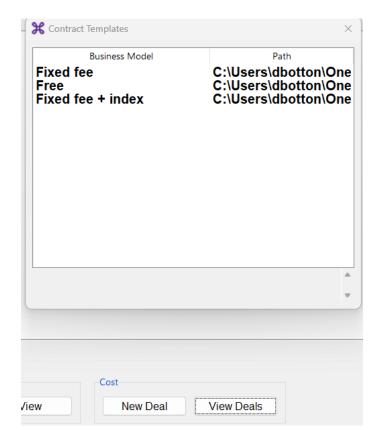


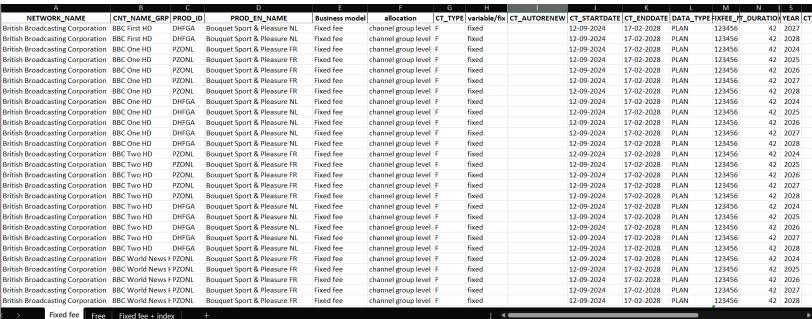


Le bouton New Deal va passer les différents filtres du Display directement dans le formulaire de création de contrat. Les champs à remplir sont dépendants du Business model choisi. Ces champs sont réduits au minimum requis par l'opérateur afin de diminuer le risque de rentrer des informations erronées. Le bouton Save va appliquer des modifications propres au Business model ainsi que différents calculs et des ajouts de colonnes qui seraient redondants de faire remplir par l'opérateur. Le bouton Add Another Selection va faire apparaître une nouvelle paire de correspondance Channel à Produit, avec leur cartouche de filtrage propre.



De retour sur l'onglet Cost, le bouton View Deals ouvre une fenêtre récapitulant chaque Business model auquel des contrats ont été ajoutés. Un double-clic sur une entrée va ouvrir le fichier Excel produit à la page correspondante.





Contact Information

If you need further assistance or have questions about the system, please contact the previous developer at david.botton@hainaut-promsoc.be. For urgent technical issues, reach out to the project lead at yasser.eljasouli-ext@bd-orange.com.