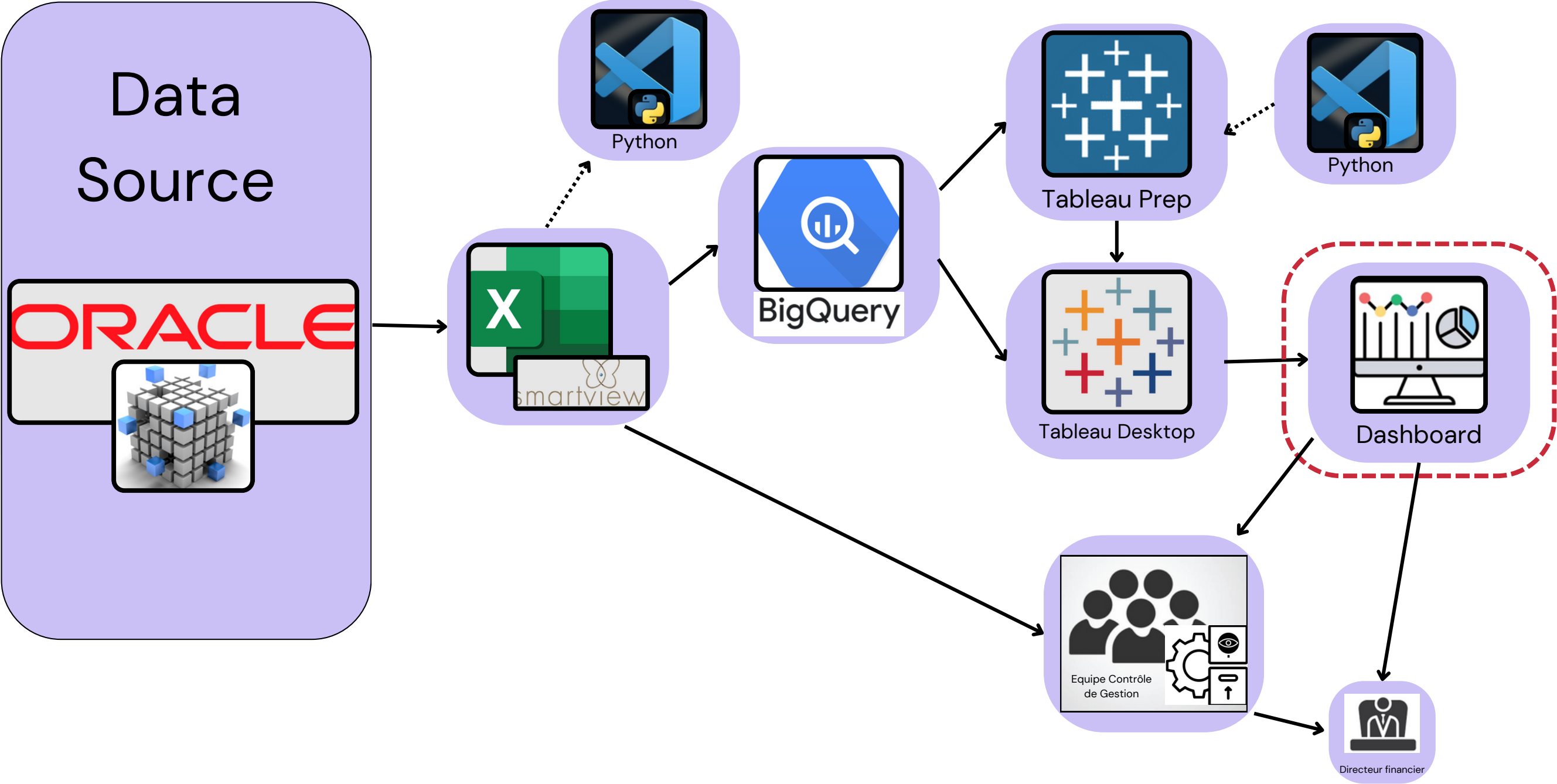


Pipeline de Données Automatisée : Extraction, Transformation et Visualisation via Tableau



Mémoire de Recherche

Réalisé par :
VERA COSIO Ray Leonardo
M2 LM DS1 Paris



Resulta General

Big Query



À l'aide de Python, le fichier Excel a été divisé en plusieurs CSV contenant les entités qui seront connectées à BigQuery.



Google Cloud

entrepot-sncf

Tapez / pour rechercher des ressources, des documents, des produits, etc

Recherche

Explorateur

Ajouter

Rechercher des ressources BigQuery

Afficher les ressources

N'afficher que les favoris

entrepot-sncf

entrepot_sncf_rle_margec

rle_cptl

rle_ctd

rle_cte

rle_ctf

rle_cti

rle_fwd_gmbh

rle_fwd_sas

rle_ibercargo

rle_rtx

Afficher plus

Résumé

Aucun élément actuellement sélectionné

Requête sans ...

Exécuter

Planifier

Enregistrer

Télécharger

Partager

Requête

1 SELECT * FROM `entrepot-sncf.entrepot_sncf_rle_margec.rle_cte` LIMIT 1000

Résultats de la requête

Enregistrer les résultats

Explorer les données

Informations sur le job

Résultats

Graphique

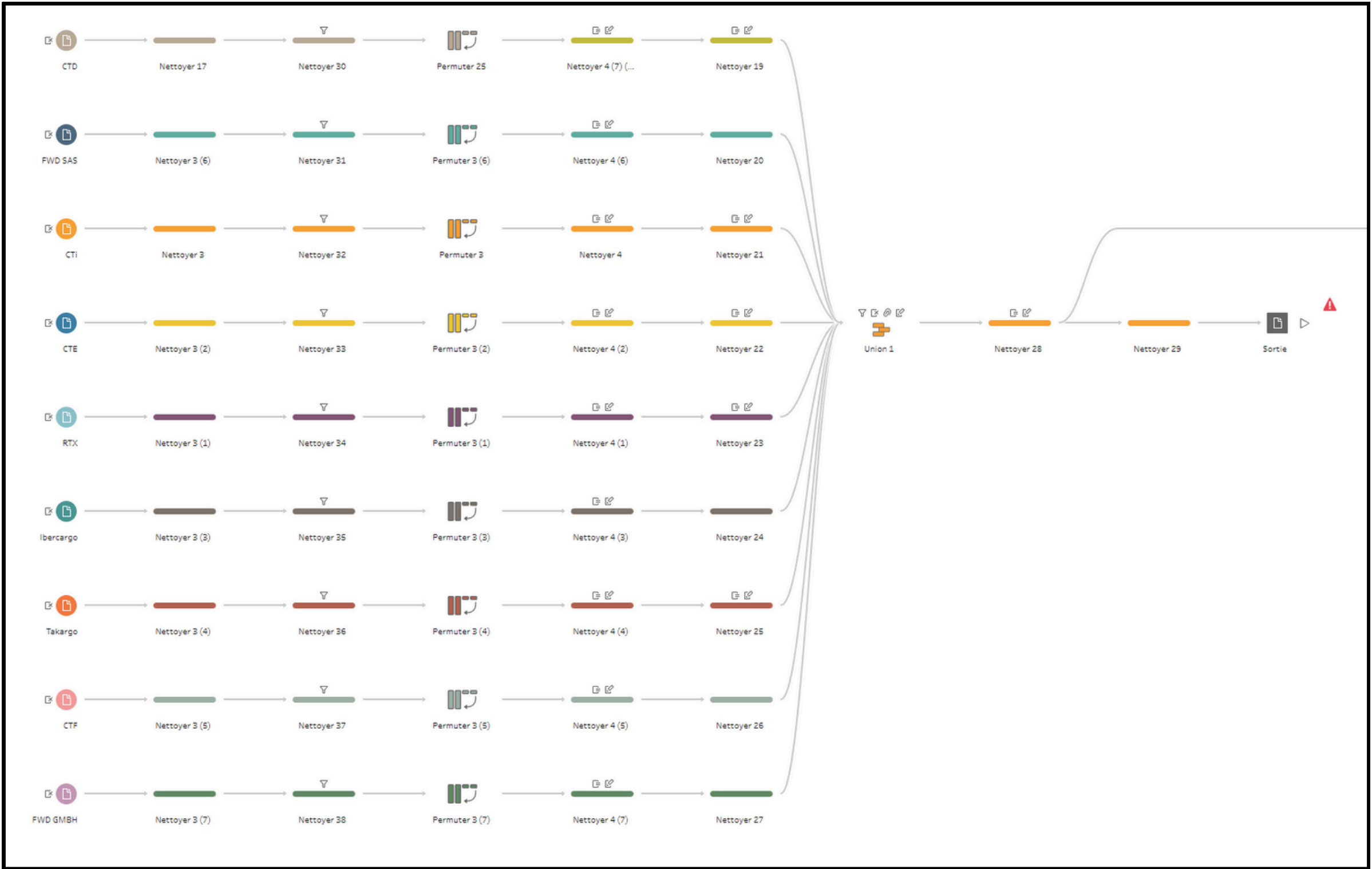
JSON

Détails de l'exécution

Graphique d'exécution

Ligne	string_field_0	string_field_1	string_field_2	string_field_3	string_field_4
1	null	null	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ESPANA,S.A.U.	CAPTRAIN ES
2	null	null	Current version	Current version	Current versio
3	null	null	Actual	Actual	Actual
4	null	null	2023	2023	2023
5	null	null	Social GAAP after allocation	Social GAAP after allocation	Social GAAP
6	null	null	Total functions	Total functions	Total fonction
7	null	null	Total Profit Center	Total Profit Center	Total Profit C
8	null	null	Total Partner	Total Partner	Total Partner
9	null	null	January	January	January
10	null	null	Correct : Total = 0,00, Somme =...	Correct : Total = 0,00, Somme =...	Correct : Tot
11	Unnamed: 0	Unnamed: 1	Total revenue (no subsidies)	RUS Subcontracting costs	Net Revenue
12	MET_4	4- Opérations de Proximité	0	0	0
13	New_1	New_1 - BLS - Hupac Luino Nov...	0	0	0
14	New_2	New_2 - Abs-Udine - Udine - Tor...	0	0	0
15	New_3	New_3 - Abs-Sisak - Udine/Os...	0	0	0
16	New_4	New_4 - Mis - VE Marghera - Lo...	0	0	0
17	New_5	New_5 - Hitachi - Pistoia - Mad...	0	0	0
18	New_6	New_6 - CD Cargo - Tarvisio - O...	0	0	0
19	New_7	New_7 - Sangritana - (Fossace...	0	0	0
20	New_8	New_8 - Railtraxx-Domo-Sacco...	0	0	0

Charger plus de données



Actions et resultats

Tableau Prep



- Ce code vérifie si la colonne [F1] n'est pas NULL et exclut les lignes où la valeur est vide. Cela permet de filtrer les données et de s'assurer que seules les lignes avec des valeurs valides sont utilisées pour l'analyse.

Modifier le filtre

Calcul

Doit renvoyer une valeur booléenne

```
NOT (((ISNULL([F1])) OR (ISNULL([F1])) OR  
(ISNULL([F1])) OR (ISNULL([F1])) OR (ISNULL([F1])) OR  
(ISNULL([F1])) OR (ISNULL([F1])) OR (ISNULL([F1])) OR  
(ISNULL([F1]))))
```

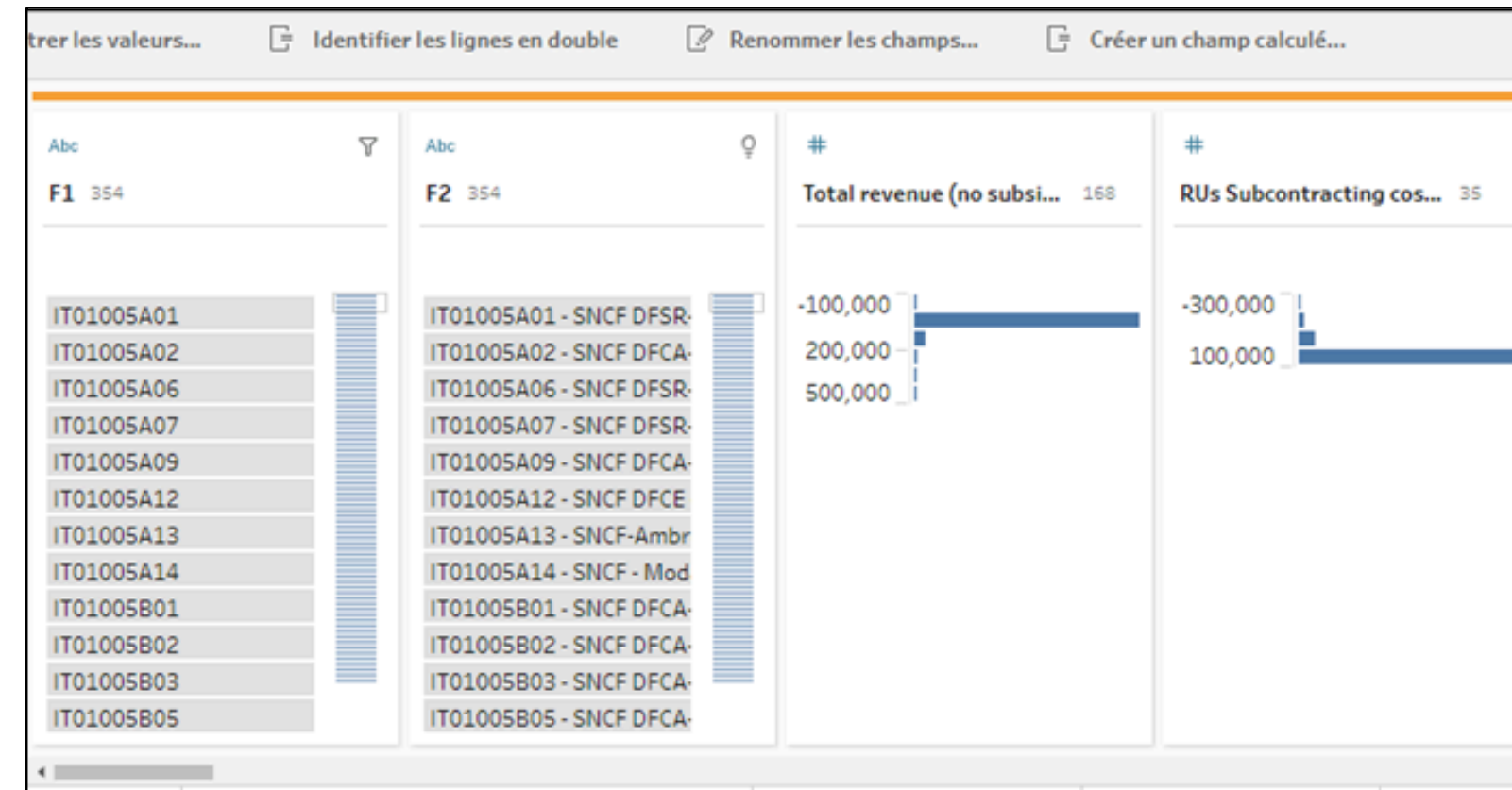
- Ce filtre exclut les lignes où la valeur de [F1] est exactement égale à "ANALYTIC_CTI" et où la colonne n'est pas NULL. Cela permet de nettoyer les données en éliminant des cas spécifiques pour une analyse plus précise.

Modifier le filtre

Calcul

Doit renvoyer une valeur booléenne

```
NOT ((([F1] == "ANALYTIC_CTI") AND NOT (ISNULL([F1]))))
```



Actions et resultats

Tableau Prep

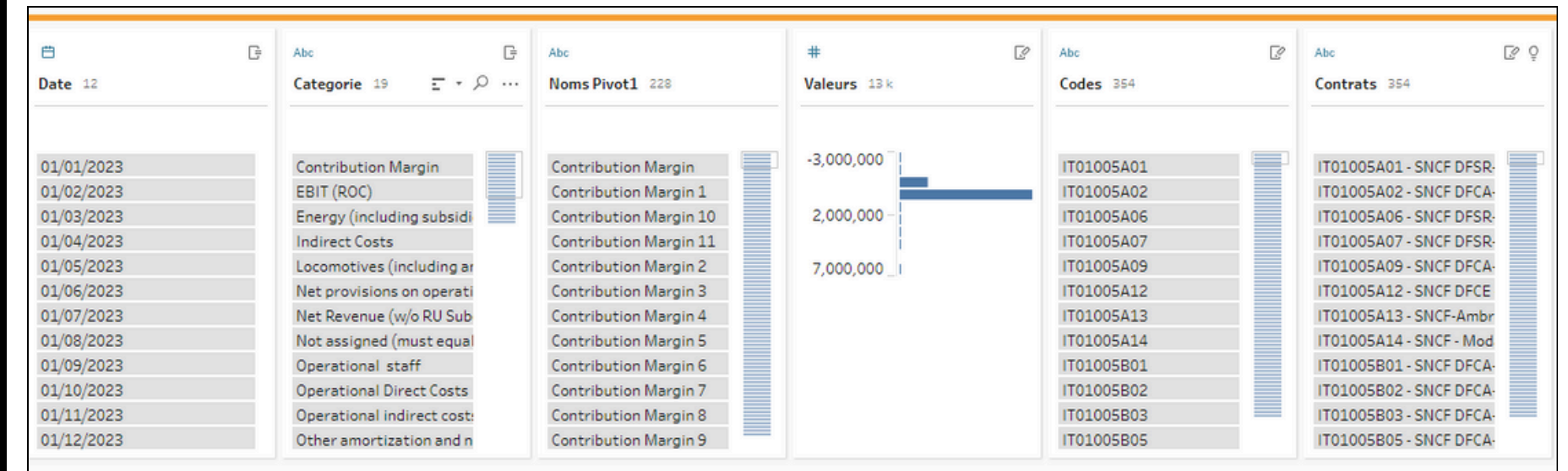


- Champ utilisée pour retirer des suffixes numériques ajoutés à des noms ou catégories et ainsi normaliser les données.

```
Modifier le champ  
Nom du champ  
Categorie  
  
IF ISNULL(INT(RIGHT([Noms Pivot1], 1))) THEN  
    [Noms Pivot1]  
ELSEIF ISNULL(INT(RIGHT([Noms Pivot1], 2))) THEN  
    TRIM(LEFT([Noms Pivot1], LEN([Noms Pivot1]) - 1))  
ELSE  
    TRIM(LEFT([Noms Pivot1], LEN([Noms Pivot1]) - 2))  
END
```

- Champ utilisée pour créer une colonne de date basée sur un suffixe numérique dans les noms, par exemple, pour attribuer une date à un élément en fonction de son rang ou son ordre dans une série

```
Modifier le champ  
Nom du champ  
Date  
  
IF RIGHT([Noms Pivot1], 2) = '11' THEN DATE("2023-12-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '1' THEN DATE("2023-02-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '2' THEN DATE("2023-03-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '3' THEN DATE("2023-04-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '4' THEN DATE("2023-05-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '5' THEN DATE("2023-06-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '6' THEN DATE("2023-07-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '7' THEN DATE("2023-08-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '8' THEN DATE("2023-09-01")  
ELSEIF RIGHT([Noms Pivot1], 1) = '9' THEN DATE("2023-10-01")  
ELSEIF RIGHT([Noms Pivot1], 2) = '10' THEN DATE("2023-11-01")  
ELSE DATE("2023-01-01")  
END
```





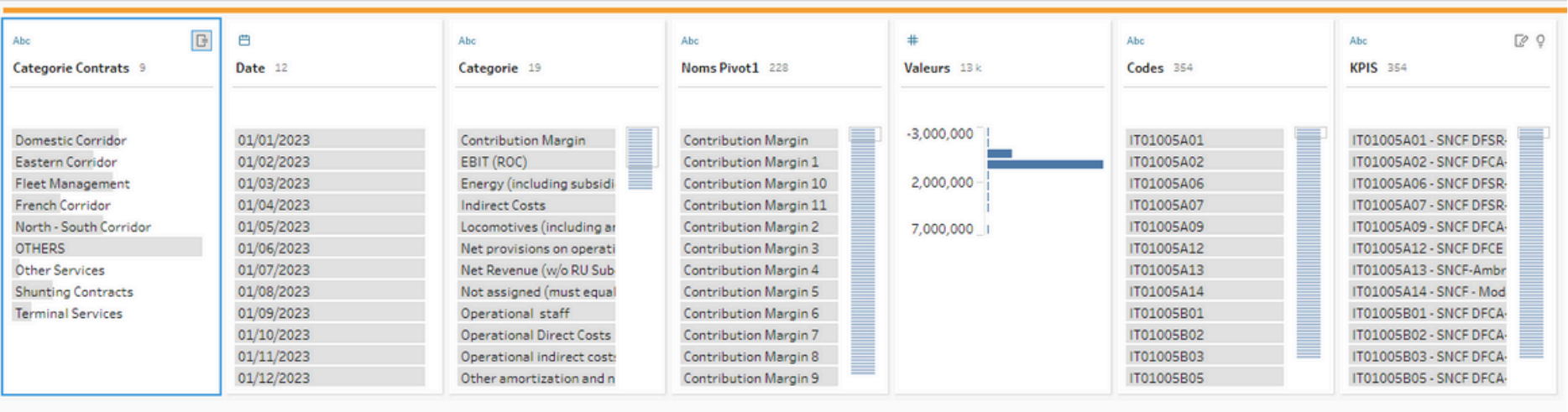
- Champ utilisée pour categoriser les contrats par corridor.
- Cette étape a été réalisée ici en raison du manque d'informations internes au moment de la collecte des données.

Modifier le champ

Nom du champ

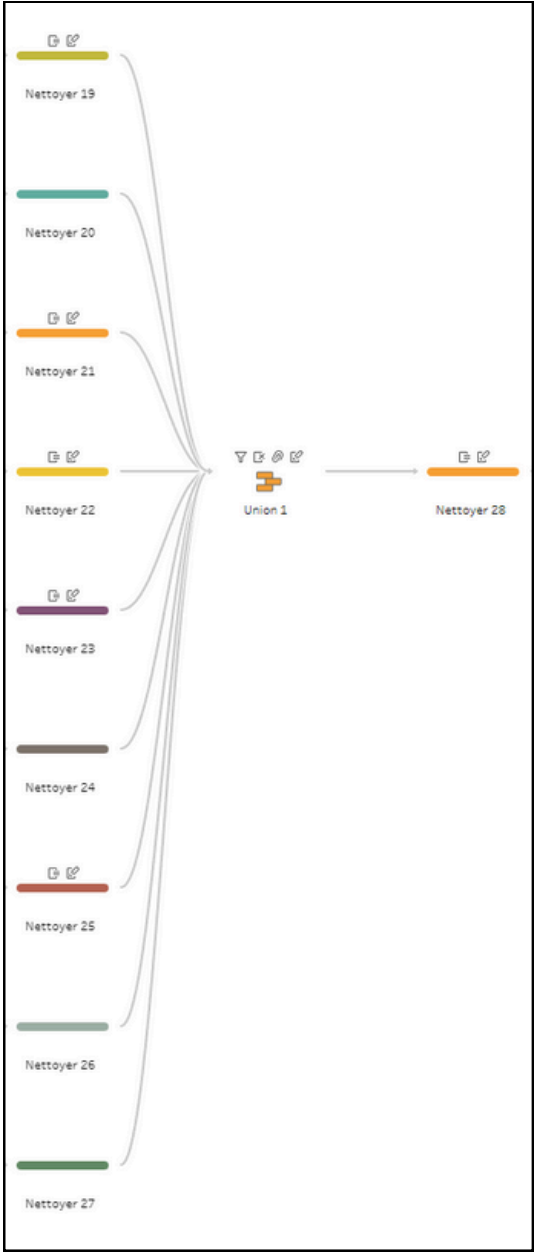
Categorie Contrats

```
IF LEFT([KPIS], 9) = "IT01TC010" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01029C01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01029C02" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01086A04" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01034A01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01034B01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01035B03" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01048B01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01048C01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01055A01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01055A02" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01055A03" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01064A01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01064A04" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01064B02" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01064B05" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01064B07" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01064B08" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01077A01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01082A01" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01INTERN" THEN "Domestic Corridor"
ELSEIF LEFT([KPIS], 10) = "IT01029B01" THEN "Domestic Corridor"
```



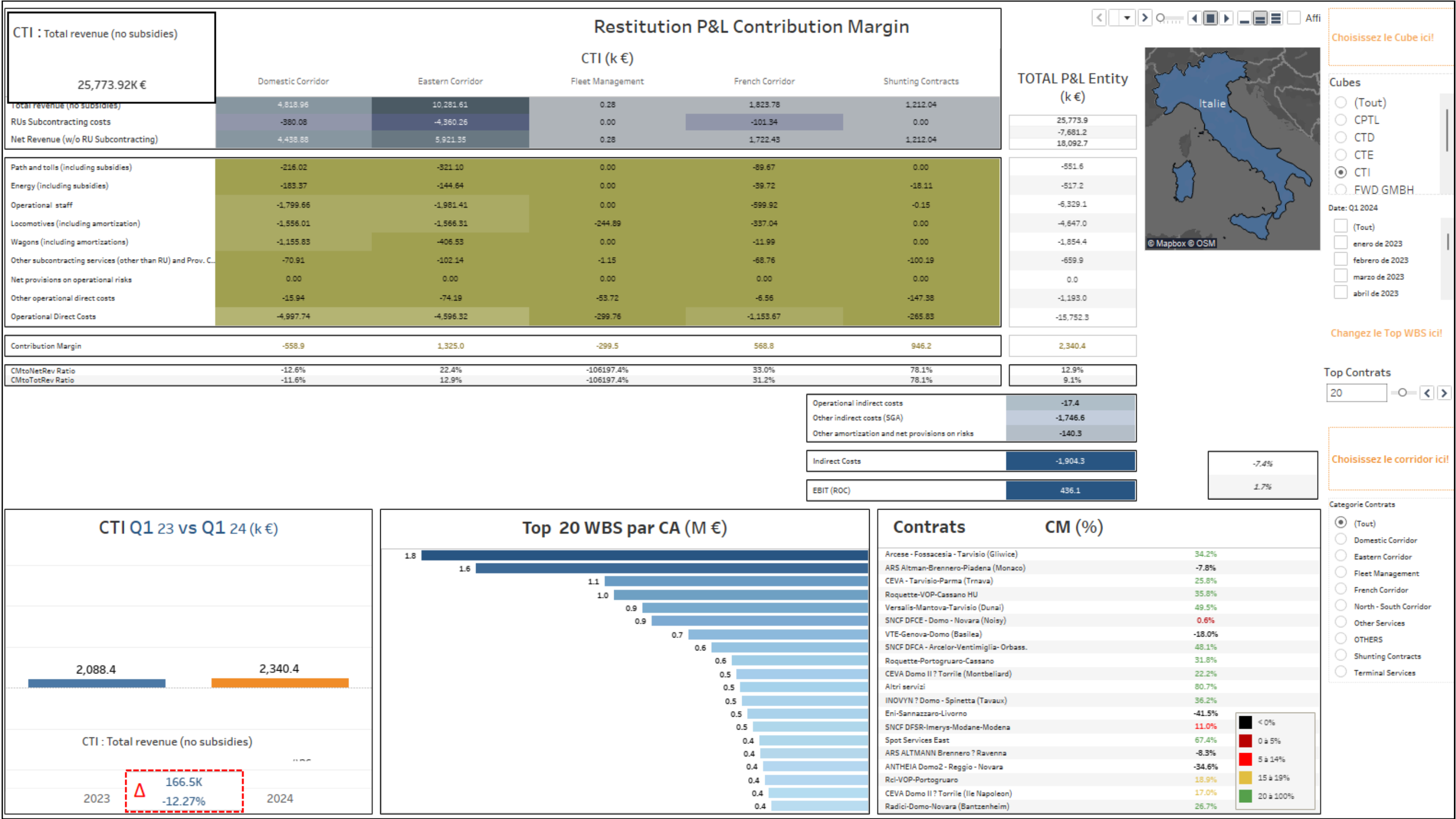


- Ici, nous joignons simplement toutes les données avec union et on organise le reste des colonnes



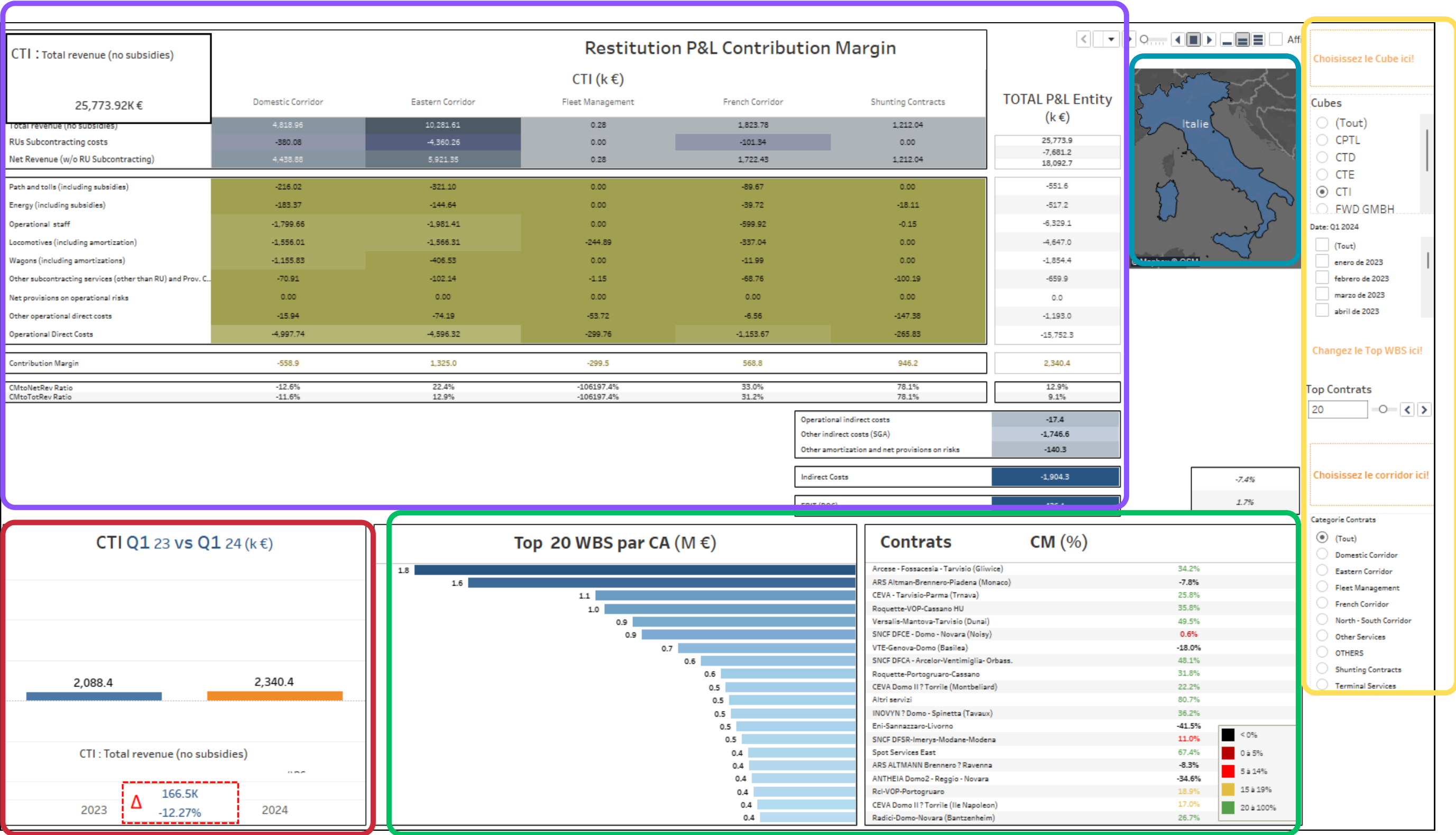
Date	Cubes	Categorie	Codes	Categorie Contrats	Valeurs
01/01/2023 01/01/2025	CPTL CTD CTE CTF CTI FWD GMBH FWD SAS IBERCARGO RTX	Contribution Margin EBIT (ROC) Energy (including subsidi Indirect Costs Locomotives (including ar Net Revenue (w/o RU Sub Net provisions on operati Not assigned (must equal Operational staff Operational Direct Costs Operational indirect cost Other amortization and n	003AB1 005AV1 005AV3 005BC1 005BD1 005BK2 005BN1 005BQ1 010AF1 010AG1 014AK1 027AA2	NULL Anciliary Activities Anciliary activities CTE Traction Captrain DE (ITL) Captrain NL DB Cargo Domestic Corridor Domestic Maritime Spain Domestic Spain Eastern Corridor Fleet Management	-0.002185856372623810 -0.01 -0.02 -0.031038780313510976 -0.040000000000000001 -0.059999999997671694 -0.059999999999999998 -0.08 -0.11 -0.14628595031010469 -0.17425238198703649 -0.20000000000000001

Date	Cubes	Categorie	Contrats	Codes	Categorie Contrats	Valeurs
01/01/2023 01/01/2025	CPTL CTD CTE CTI FWD GMBH FWD SAS IBERCARGO RTX	Contribution Margin EBIT (ROC) Energy (including subsidi Indirect Costs Locomotives (including ar Net Revenue (w/o RU Sub Net provisions on operati Not assigned (must equal Operational staff Operational Direct Costs Operational indirect cost Other amortization and n	ACC MASSY CLERMONT BRIFER CREIL CHANTIER HOURCADE Developpement Chantier Developpement ITE DIRECTION DU MATERIE EIFFAGE RAIL BPL ETF SPOT Formation continue FORMATION CONTINUE I FORWARDIS KLEBER POI Forwardis SAS - Buta - M	003AB1 005AV1 005AV3 005BC1 005BD1 005BK2 005BN1 005BQ1 010AF1 010AG1 014AK1 027AA2	NULL Anciliary Activities Anciliary activities CCW CTE Traction Captrain DE (ITL) Captrain NL DB Cargo Domestic Corridor Domestic Maritime Spain Domestic Spain Fleet Management	-0.002185856372623810 -0.01 -0.031038780313510976 -0.040000000000000001 -0.059999999997671694 -0.059999999999999998 -0.08 -0.14628595031010469 -0.40000000000000002 -0.48999999999999999 -0.69999999999999996 -0.89000000000000001



Parties du graphique

Tableau Desktop



Top 20 des contrats

Tableau Desktop

