

# **PROJET SCIENCES DES DONNÉES 2 + BASES DE DONNÉES**

**LOUATI Chamss-Eddine**

**NDIAYE Ibrahima**

**EL OUALYDY Mohamed Amine**

**SARTORI Adrien**

# PLAN

- 01** CONTEXUALISATION
- 02** PROBLÉMATIQUE
- 03** QUESTIONS À ABORDER
- 04** BASE DE DONNÉES
- 05** REQUETES ET RÉSULTATS
- 06** CONCLUSION

# 1. CONTEXUALISATION

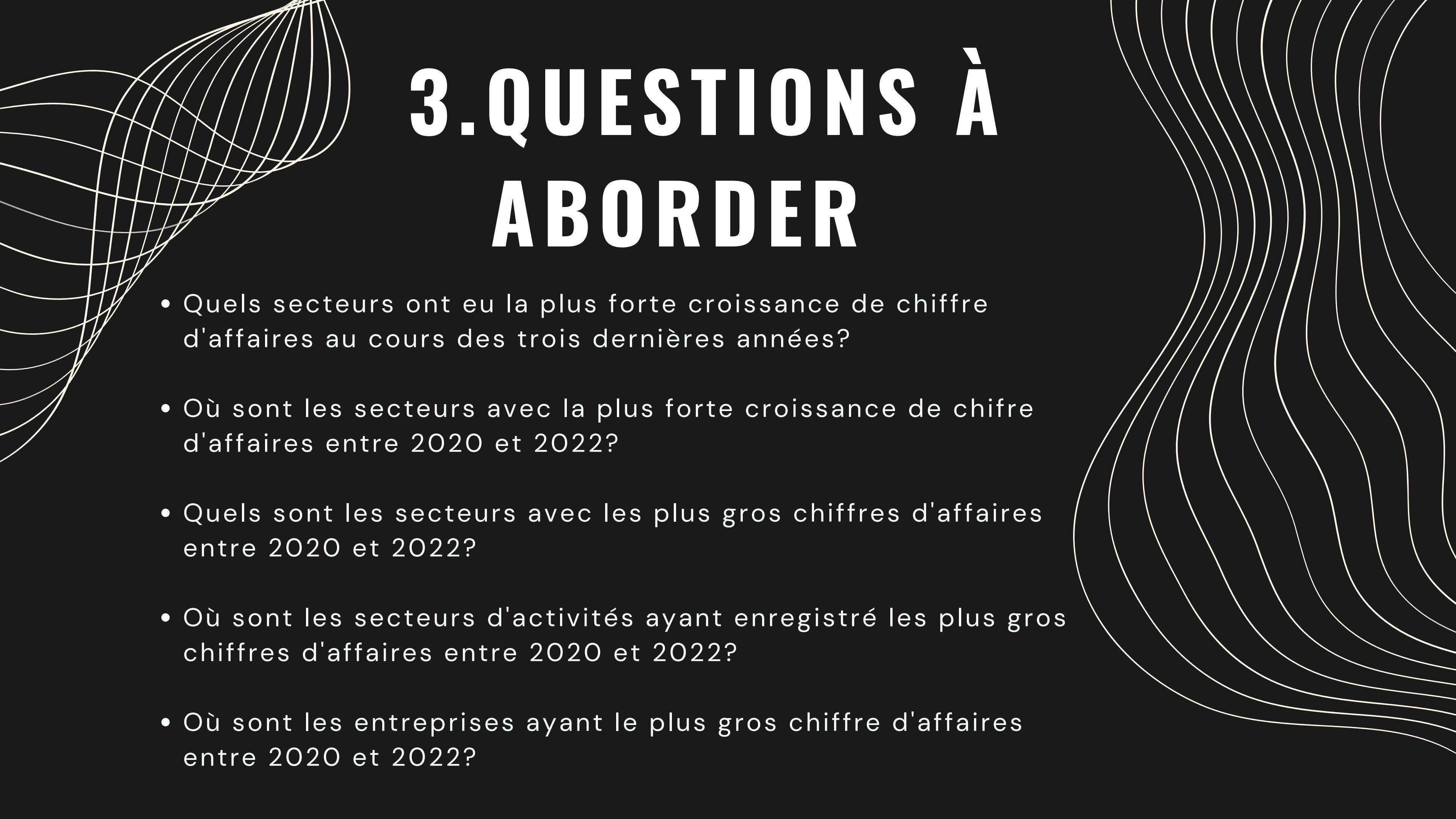
- Importance des secteurs avec un gros chiffre d'affaire.
- Mieux connaître les secteurs pour mieux investir.
- Les gouvernements régionaux ont besoin de ces informations.

- Les trois secteurs économiques principaux sont:
  1. Le secteur primaire : collecte et l'exploitation des ressources naturelles (matériaux, énergie, et certains aliments).
  2. Le secteur secondaire : industries de transformation des matières premières.
  3. Le secteur tertiaire : les industries du service;



## **2. PROBLÉMATIQUE**

**QUELS SONT LES SECTEURS D'ACTIVITÉS QUI  
ENREGISTRENT LES PLUS GROS CHIFFRES  
D'AFFAIRES ET DANS QUELLES RÉGIONS AU  
COURS DE CES TROIS DERNIÈRES  
ANNÉES(2020,2021,2022) ?**



# 3. QUESTIONS À ABORDER

- Quels secteurs ont eu la plus forte croissance de chiffre d'affaires au cours des trois dernières années?
- Où sont les secteurs avec la plus forte croissance de chiffre d'affaires entre 2020 et 2022?
- Quels sont les secteurs avec les plus gros chiffres d'affaires entre 2020 et 2022?
- Où sont les secteurs d'activités ayant enregistré les plus gros chiffres d'affaires entre 2020 et 2022?
- Où sont les entreprises ayant le plus gros chiffre d'affaires entre 2020 et 2022?

# **4. BASE DE DONNÉES**

- **Import et nettoyage des données**
- **Modélisation conceptuel des données (MCD)**
- **Modélisation organisationnelle des données (MOD)**
- **MOD MANUSCRIT**

# IMPORT ET NETTOYAGE DES DONNÉES

Importation des  
données

Nettoyage

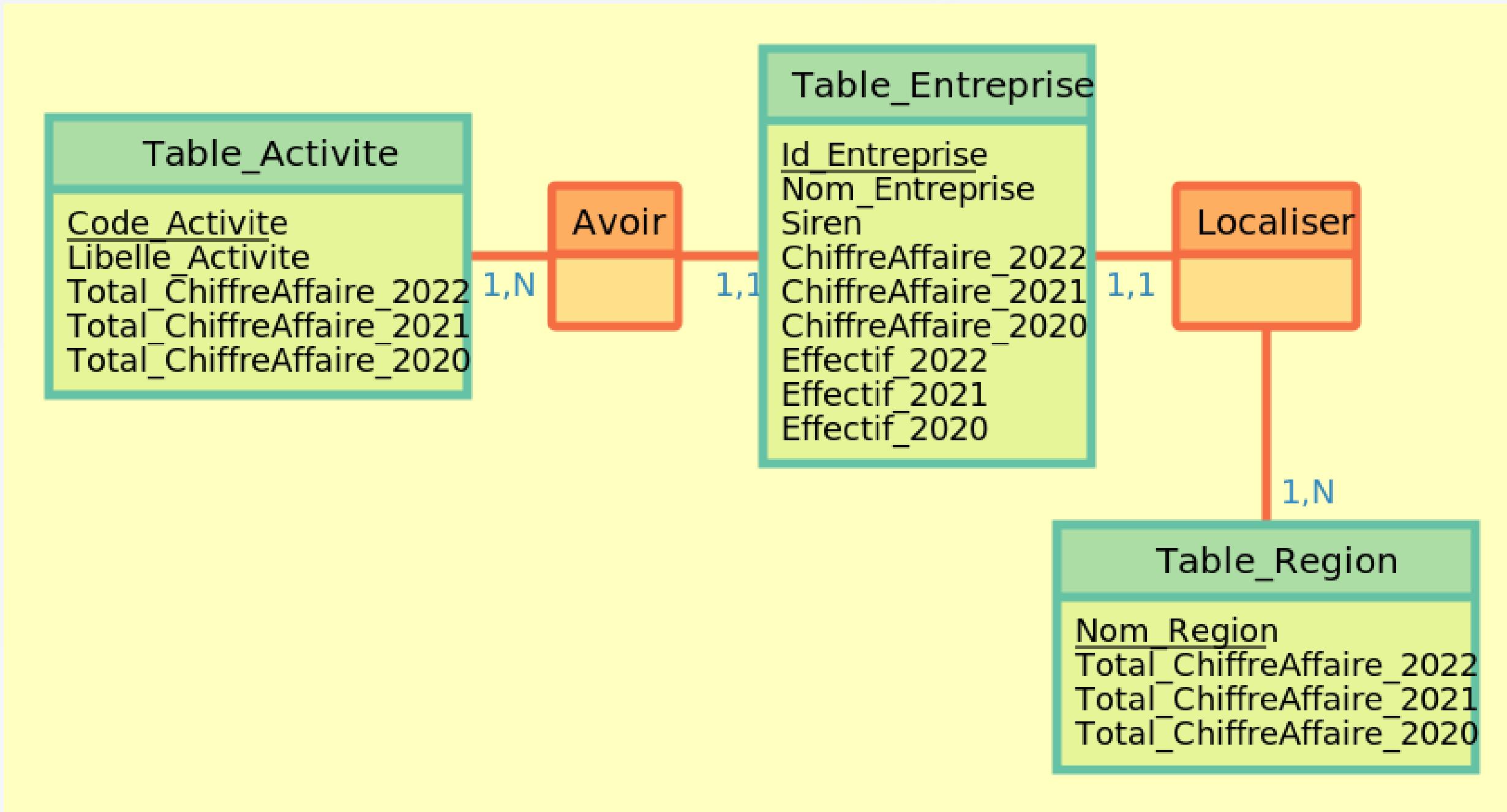
Pré-traitements

Récupération des  
données afin de les  
analyser

Nettoyage des  
données sur Excel  
et Python

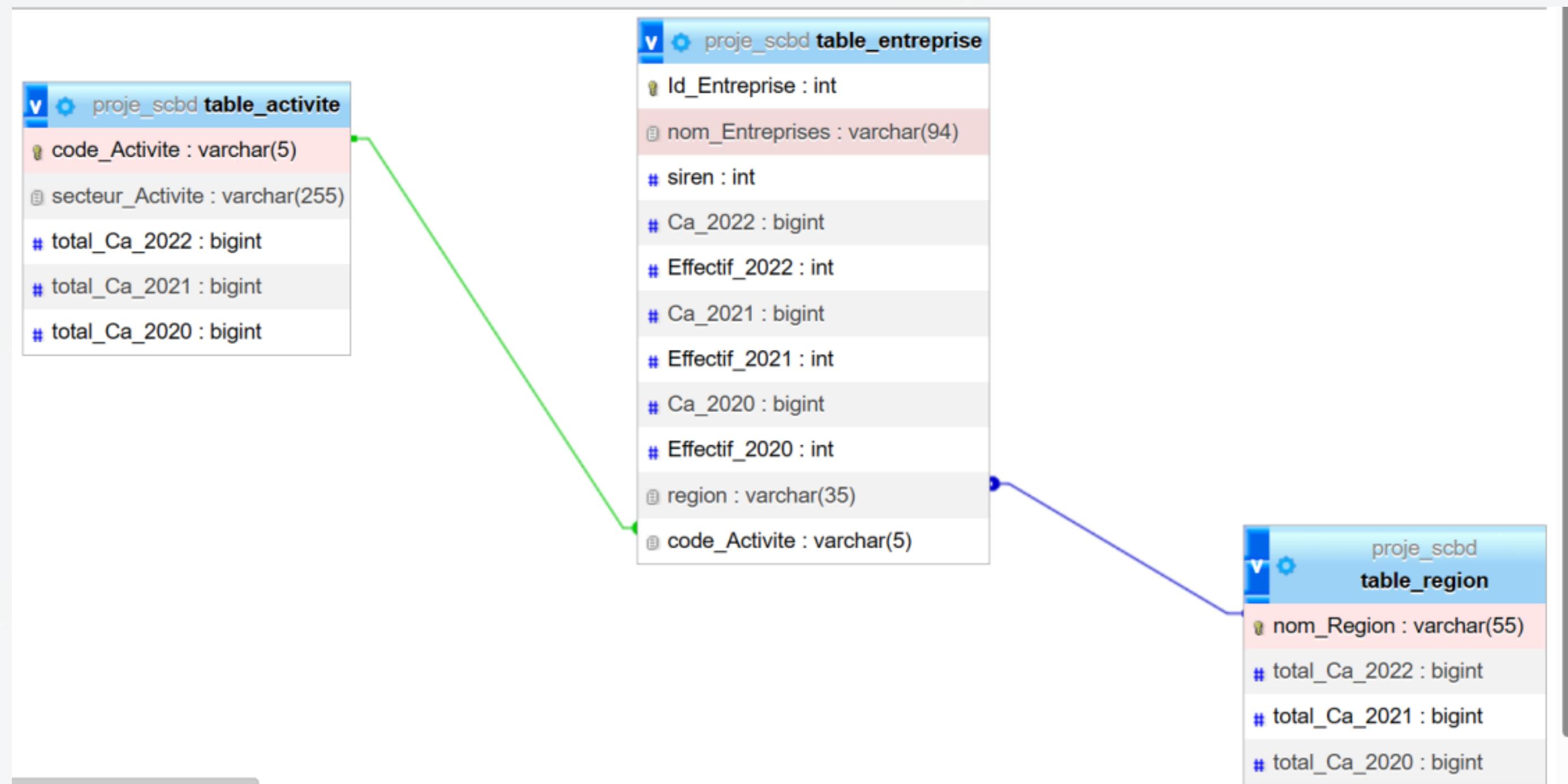
Filtrer le fichier  
afin d'avoir nos  
trois tables

# MCD



- Description des données utilisées et leurs relations

# MOD



- **Légende:** Dans l'explication de notre MOD le signe # représente les clés étrangères et les mots soulignés sont les clés primaires.

# MOD MANUSCRIT

## ▼ TABLE\_ACTIVITE ( Code\_Activite, Libelle\_Activite, Total\_ChiffreAffaire\_2022, Total\_ChiffreAffaire\_2021, Total\_ChiffreAffaire\_2020 )

- Le champ *Code\_Activite* constitue la clé primaire de la table. C'était déjà un identifiant de l'entité *Table\_Activite*.
- Les champs *Libelle\_Activite*, *Total\_ChiffreAffaire\_2022*, *Total\_ChiffreAffaire\_2021* et *Total\_ChiffreAffaire\_2020* étaient déjà de simples attributs de l'entité *Table\_Activite*.

## ▼ TABLE\_ENTREPRISE ( Id\_Entreprise, Nom\_Entreprise, Siren, ChiffreAffaire\_2022, ChiffreAffaire\_2021, ChiffreAffaire\_2020, Effectif\_2022, Effectif\_2021, Effectif\_2020, #Code\_Activite, #Nom\_Région )

- Le champ *Id\_Entreprise* constitue la clé primaire de la table. C'était déjà un identifiant de l'entité *Table\_Entreprise*.
- Les champs *Nom\_Entreprise*, *Siren*, *ChiffreAffaire\_2022*, *ChiffreAffaire\_2021*, *ChiffreAffaire\_2020*, *Effectif\_2022*, *Effectif\_2021* et *Effectif\_2020* étaient déjà de simples attributs de l'entité *Table\_Entreprise*.
- Le champ *Code\_Activite* est une clé étrangère. Il a migré par l'association de dépendance fonctionnelle *Avoir* à partir de l'entité *Table\_Activite* en perdant son caractère identifiant.
- Le champ *Nom\_Région* est une clé étrangère. Il a migré par l'association de dépendance fonctionnelle *Localiser* à partir de l'entité *Table\_Région* en perdant son caractère identifiant.

## ▼ TABLE\_REGION ( Nom\_Région, Total\_ChiffreAffaire\_2022, Total\_ChiffreAffaire\_2021, Total\_ChiffreAffaire\_2020 )

- Le champ *Nom\_Région* constitue la clé primaire de la table. C'était déjà un identifiant de l'entité *Table\_Région*.
- Les champs *Total\_ChiffreAffaire\_2022*, *Total\_ChiffreAffaire\_2021* et *Total\_ChiffreAffaire\_2020* étaient déjà de simples attributs de l'entité *Table\_Région*.

Version manuscrite du  
MOD

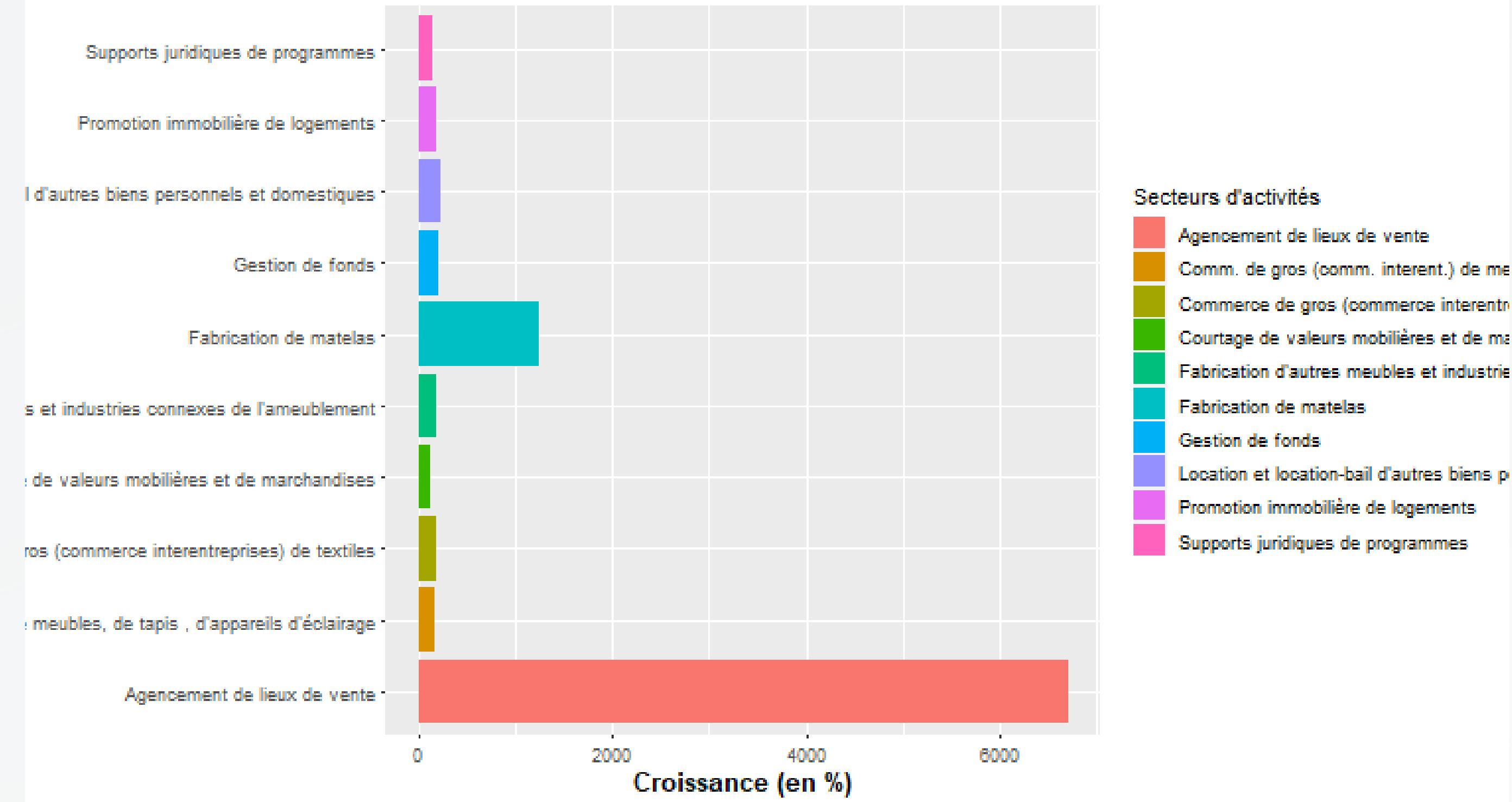
# 5. REQUETES ET RÉSULTATS

```
SELECT secteur_activite,  
((total_ca_2022 - total_ca_2020)/total_ca_2020)*100 AS croissance  
FROM table_activite  
ORDER BY croissance DESC  
LIMIT 10;"
```

- Cette requête permet de calculer le taux de croissance pour la période 2020 à 2022. Les résultats sont triés et limités aux 10 premiers résultats pour illustrer les secteurs d'activité les plus porteurs.

secteur_activite	croissance
Agencement de lieux de vente	6704.5386
Fabrication de matelas	1248.9513
Location et location-bail d'autres biens personnel...	231.5481
Gestion de fonds	200.1801
Promotion immobilière de logements	190.6370
Fabrication d'autres meubles et industries connexe...	188.9064
Commerce de gros (commerce interentreprises) de te...	186.6665
Comm. de gros (comm. interent.) de meubles, de tap...	160.7028
Supports juridiques de programmes	148.8556
Courtage de valeurs mobilières et de marchandises	134.4692

## Quels sont les secteurs d'activités qui ont enregistré la plus forte croissance entre 2020 et 2022 ?



1) Quels sont les secteurs d'activités ayant enregistré la plus forte croissance de chiffre d'affaires au cours des trois dernières années(2020,2021,2022)?

```

SELECT a.secteur_Activite, r.nom_region,
       (e.ca_2022 - e.ca_2020) AS croissance
  FROM Table_Entreprise e
  JOIN Table_Région r ON e.région = r.nom_region
 JOIN table_activite a on e.code_Activite=a.code_Activite
 ORDER BY croissance DESC
 LIMIT 10;

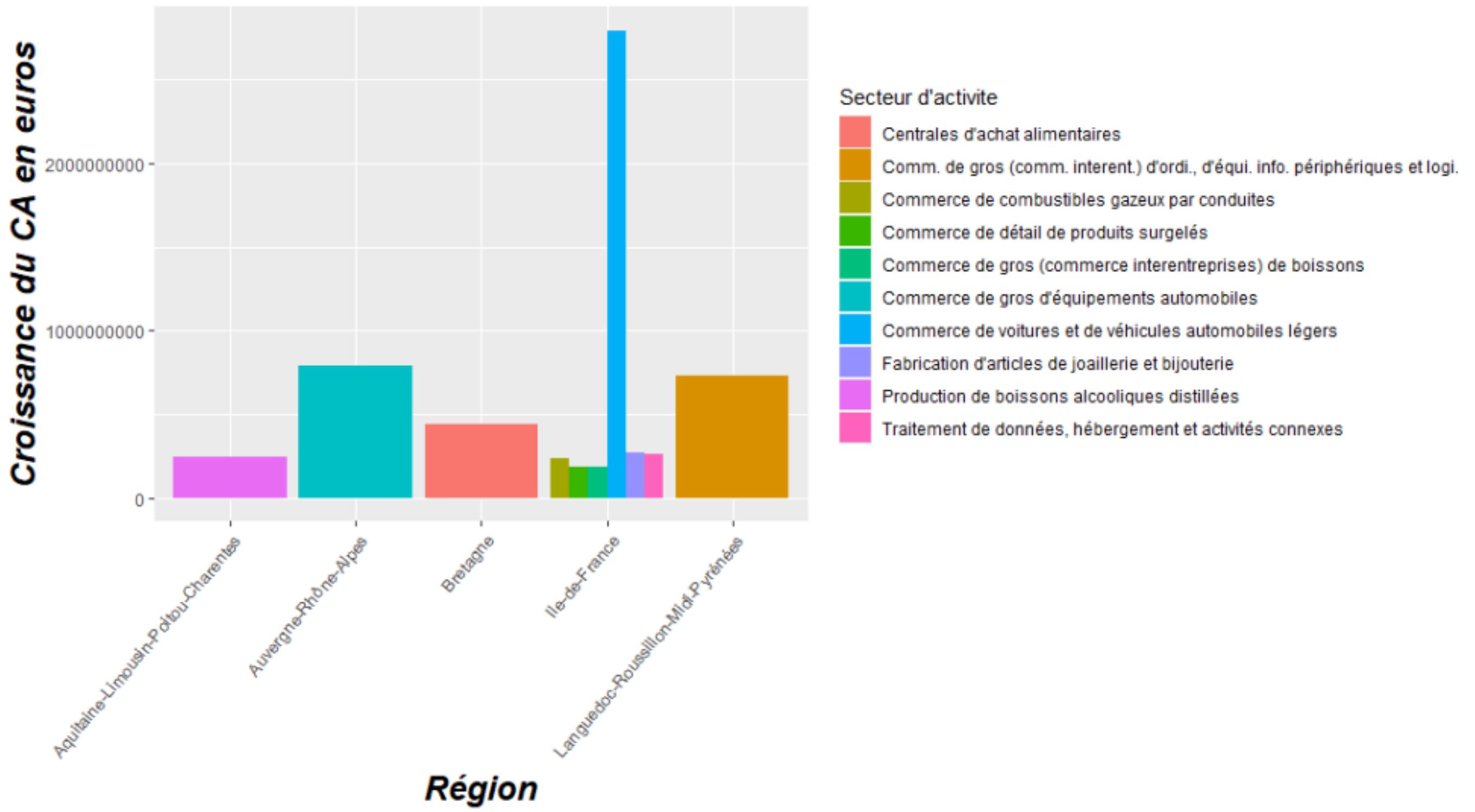
```

- Cette requête sélectionne les 10 premières entreprises avec la croissance de revenus la plus rapide entre 2020 et 2022.

- Elle relie nos trois tables pour obtenir le secteur d'activité de chaque entreprise et sa région.

secteur_Activite	nom_region	croissance_absolu
Commerce de voitures et de véhicules automobiles I...	Ile-de-France	2797434094
Commerce de gros d'équipements automobiles	Auvergne-Rhône-Alpes	785127506
Comm. de gros (comm. interent.) d'ordi., d'équi. i...	Languedoc-Roussillon-Midi-Pyrénées	733456394
Centrales d'achat alimentaires	Bretagne	439842175
Fabrication d'articles de joaillerie et bijouterie	Ile-de-France	271997592
Traitement de données, hébergement et activités co...	Ile-de-France	257601806
Production de boissons alcooliques distillées	Aquitaine-Limousin-Poitou-Charentes	242490547
Commerce de combustibles gazeux par conduites	Ile-de-France	231895611
Commerce de gros (commerce interentreprises) de bo...	Ile-de-France	183402449
Commerce de détail de produits surgelés	Ile-de-France	183041393

***Titre: Dans quelle région se trouvent les secteurs d'activités ayant enregistré la plus forte croissance de chiffre d'affaires***



2) Dans quelles régions se trouvent les secteurs d'activités ayant enregistré la plus forte croissance de chiffre d'affaires entre 2020 et 2022?

```

SELECT secteur_activite, SUM(total_ca_2020) + SUM(total_ca_2021) + SUM(total_ca_2022)
FROM table_activite
GROUP BY secteur_activite
ORDER BY chiffre_affaires_total DESC
LIMIT 10;

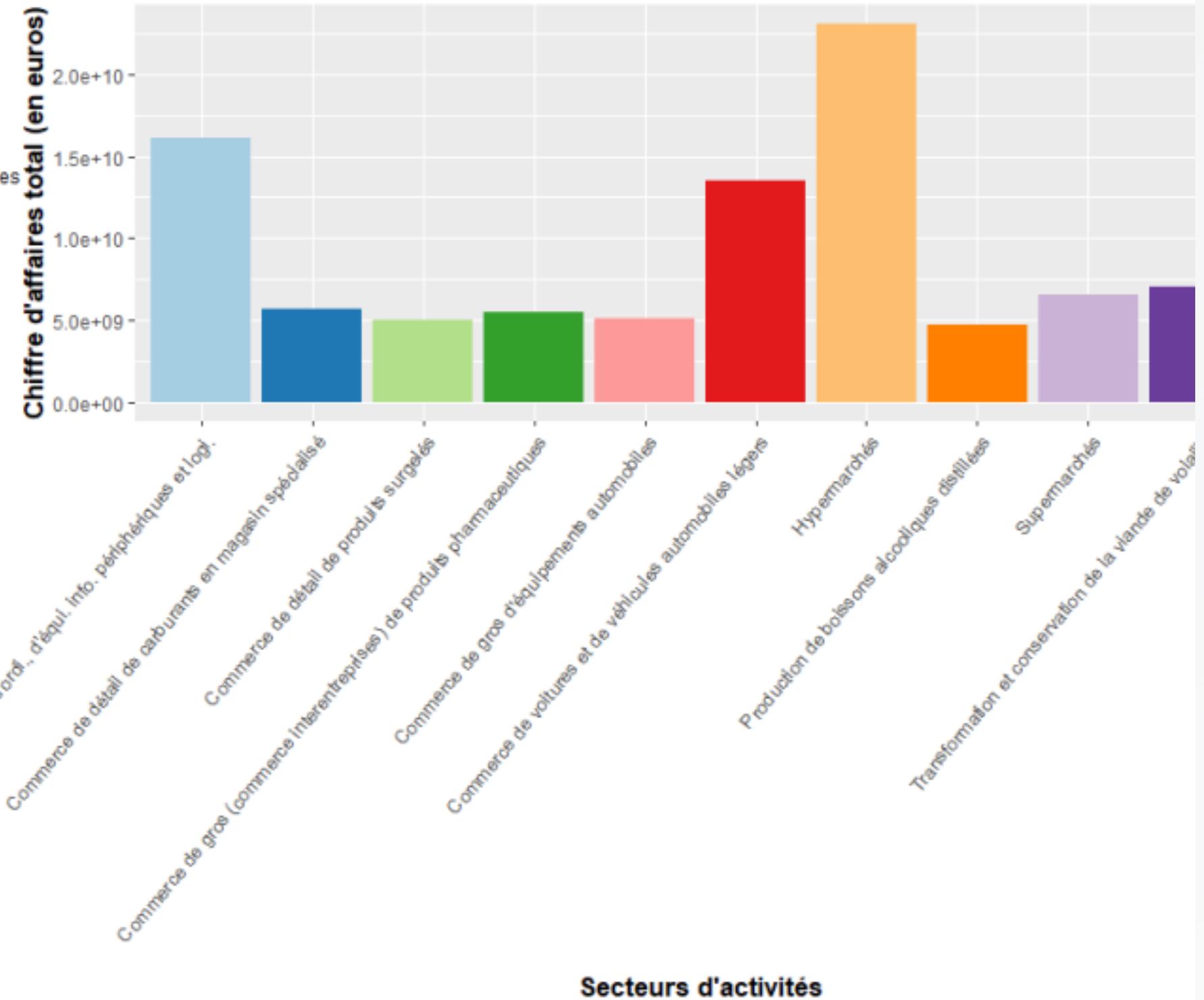
```

Cette requête calcule le chiffre d'affaires total pour les 10 secteurs d'activité les plus rentables en faisant la somme des chiffres d'affaires pour les années 2020, 2021 et 2022.

secteur_activite	chiffre_affaires_total
Hypermarchés	23094581075
Comm. de gros (comm. interent.) d'ordi., d'équi. i...	16112007262
Commerce de voitures et de véhicules automobiles I...	13498677744
Transformation et conservation de la viande de vol...	7029289253
Supermarchés	6537233019
Commerce de détail de carburants en magasin spécia...	5720266855
Commerce de gros (commerce interentreprises) de pr...	5525270607
Commerce de gros d'équipements automobiles	5115696924
Commerce de détail de produits surgelés	4967466344
Production de boissons alcooliques distillées	4695927275

### **Quels sont les secteurs d'activités qui ont enregistré les plus gros chiffres d'affaires entre 2020 et 2022 ?**

urs d'activités  
omm. de gros (comm. interent.) d'ordi., d'équi. info. périphériques et logi.  
commerce de détail de carburants en magasin spécialisé  
commerce de détail de produits surgelés  
commerce de gros (commerce interentreprises) de produits pharmaceutiques  
commerce de gros d'équipements automobiles  
commerce de voitures et de véhicules automobiles légers  
hypermarchés  
roduction de boissons alcooliques distillées  
upermarchés  
transformation et conservation de la viande de volaille



3) Quels sont les secteurs d'activités qui ont enregistré les plus gros chiffres d'affaires entre 2020 et 2022?

```

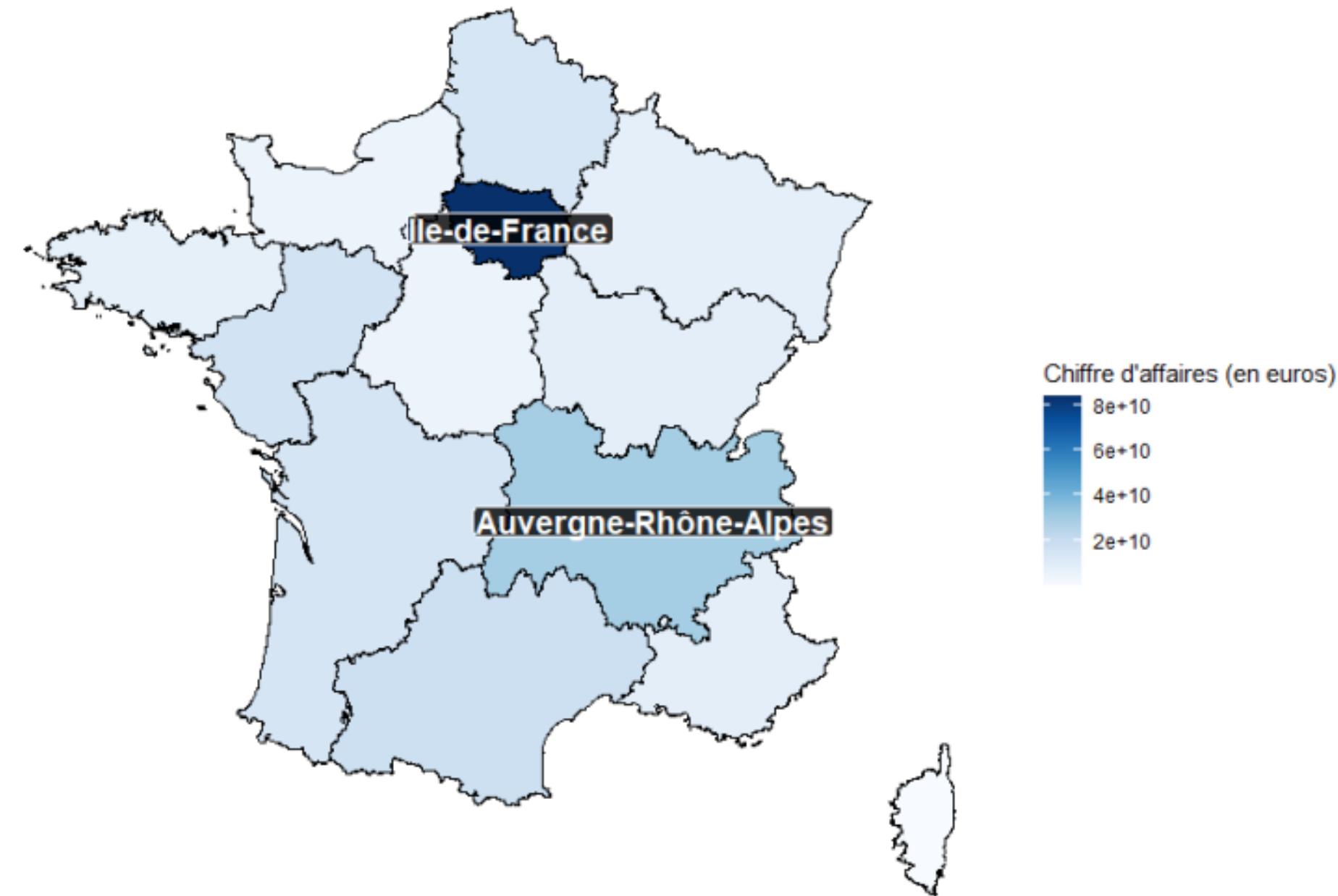
SELECT nom_region, SUM(total_ca_2020 + total_ca_2021
+ total_ca_2022) AS chiffre_affaires_total
FROM table_region
GROUP BY nom_region
ORDER BY chiffre affaires total DESC

```

- Cette requête calcule le chiffre d'affaires total pour chaque région en faisant la somme des chiffres d'affaires pour les années 2020, 2021 et 2022.

nom_region	chiffre_affaires_total
Ile-de-France	83839137861
Auvergne-Rhône-Alpes	29637867111
Languedoc-Roussillon-Midi-Pyrénées	17949033635
Pays-de-la-Loire	15616181147
Aquitaine-Limousin-Poitou-Charentes	14169070334
Nord-Pas-de-Calais-Picardie	13937525165
Bourgogne-Franche-Comté	8478873979
Provence-Alpes-Côte d'Azur	8308523140
Alsace-Champagne-Ardenne-Lorraine	7571729875
Bretagne	7521388532
Normandie	5232736055
Centre-Val de Loire	5209336028

**Chiffre d'affaires (en euros) accumulé par région entre 2020 et 2022 en France**



4) Dans quelle région se trouvent les secteurs d'activités ayant enregistré les plus gros chiffres d'affaires au cours des trois dernières années(2020, 2021 et 2022)?

```

SELECT region,
       MAX(total_ca_2022) AS ca_2022,
       MAX(total_ca_2021) AS ca_2021,
       MAX(total_ca_2020) AS ca_2020
  FROM Table_Region,table_entreprise
 where Table_Region.nom_region = Table_Entreprise.region
 GROUP BY region
 ORDER BY ca_2022 DESC, ca_2021 DESC, ca_2020 DESC
LIMIT 10;

```

- Dans cette requête, on utilise une jointure entre les tables pour regrouper les entreprises par région, puis on a calculé le chiffre d'affaire total par année selon les régions.

5) Quelles sont les régions où se trouvent les entreprises ayant le plus gros chiffre d'affaires sur les trois dernières années(2020, 2021, 2022)?

region	ca_2022	ca_2021	ca_2020
Île-de-France	32030960828	24888726192	26919450841
Auvergne-Rhône-Alpes	10949204267	9709795617	8978867227
Languedoc-Roussillon-Midi-Pyrénées	6635244454	5938282745	5375506436
Pays-de-la-Loire	5611136015	5046699992	4958345140
Aquitaine-Limousin-Poitou-Charentes	5250251762	4493803368	4425015204
Nord-Pas-de-Calais-Picardie	5086752271	4477083866	4373689028
Bourgogne-Franche-Comté	3024731487	2735781954	2718360538
Bretagne	2994007971	2194004587	2333375974
Provence-Alpes-Côte d'Azur	2915407551	2635227711	2757887878
Alsace-Champagne-Ardenne-Lorraine	2761721799	2373682523	2436325553

# 6. CONCLUSION



Investir dans des secteurs en fonction de l'évolution de leur chiffre d'affaires sur les trois dernières années.

**STRATÉGIE N°1**



Investir dans des secteurs en fonction du chiffre d'affaires que leur région a générée sur les trois dernières années.

**STRATÉGIE N°2**