

Rapport de gestion de projet SAE S6

Professeure : Rafaele BONDAZ

Léo COSTE

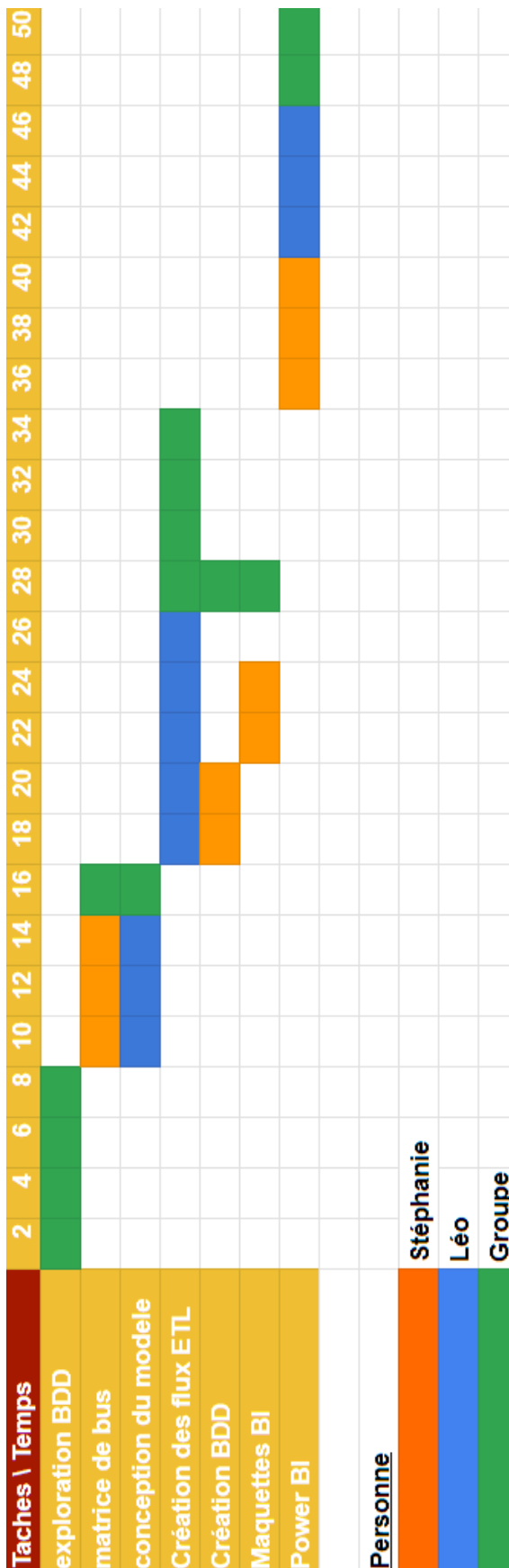
et

Stéphanie AGBANGLANON

Sommaire

1. Rétro-planning des tâches réalisées.....	3
2. Détail du travail réalisé.....	4
2.1. Conception.....	4
2.1.1. Choix des outils.....	4
2.1.2. Modélisation.....	5
2.1.3. Maquettage.....	6
2.2. Réalisation.....	7
2.2.1. Flux ETL et chargement du datawarehouse.....	7
2.2.1.1. Mappings.....	7
2.2.1.2. Flux.....	10
2.2.1.3. Modifications opérées.....	12
2.2.2. Visualisation.....	14
2.2.2.1. Indicateurs calculés.....	14
2.2.2.2. Graphiques choisis.....	16
2.2.2.3. Rendu visuel.....	17
Conclusion.....	22

1. Rétro-planning et organisation des tâches réalisées



Pour ce projet, nous avons réparti les tâches de manière assez naturelle après avoir établi une liste des différentes étapes nécessaires dès le début. Voici comment nous avons distribué et coordonné notre travail :

Identification des Tâches :

Dès le début du projet, nous avons identifié et défini toutes les tâches nécessaires. Cela nous a permis de structurer notre approche et de nous assurer que chaque étape était bien couverte.

Répartition des Tâches :

Comme le montre notre planning, nous avons pris en charge différentes parties du projet en fonction de nos compétences et points forts.

Revue Collective :

À la fin de chaque grande étape, représentée par les cases vertes dans le planning, nous avons fait une grosse revue de ce que chacun avait réalisé. Cela nous a permis de vérifier l'intégrité et la cohérence de notre travail, de faire des ajustements si nécessaire, et de valider ensemble chaque tâche avant de la clore.

En résumé, notre approche collaborative et structurée nous a permis de mener à bien notre projet de manière efficace, avec une vérification rigoureuse à chaque étape clé pour assurer la qualité et la cohérence de notre travail.

2. Détail du travail réalisé

Les tâches définies plus haut sont globalement celles qui ont servi de base pour effectuer le travail. Nous sommes d'abord passés par une phase de conception qui a permis de nous faire une idée claire de ce que nous voulions faire et de comment nous voulions le faire, avant de passer à la réalisation.

2.1. Conception

La phase de conception englobe notamment les étapes de choix divers tels que le choix des outils, celui de la modélisation et de la présentation à adopter pour les graphiques du tableau de bord.

2.1.1. Choix des outils

Les outils choisis l'ont été en général sur base de notre aisance à les manipuler. Cependant l'outil d'ETL Pentaho Data Integrator a été un compromis entre familiarité et découverte puisqu'il n'est pas celui qu'aucun d'entre nous utilise dans la vie professionnelle, mais il a été abordé en cours.

Voici donc ci-après les outils et plateformes que nous avons utilisés :

- **Maquettage** : Figma
- **Modèle de données** : Power AMC
- **ETL** : Pentaho Data Integrator
- **Stockage des données** : Oracle SQL Server
- **Visualisation** : Power BI

L'outil de visualisation Power BI a été choisi aussi bien pour sa connaissance plus large que les autres outils connus tels que Tableau, mais aussi pour certaines fonctionnalités telles que son éditeur Power Query, même s'il a fallu au départ régler les problèmes d'incompatibilité avec la base oracle.

Nous avons choisi une modélisation en flocon qui nous a permis de créer des hiérarchies et rendre plus clair et compréhensible notre modèle notamment pour la dimension géographique. Cependant, pour la dimension temporelle, nous avons jugé, finalement, qu'il serait inintéressant, dans le cadre de notre projet de créer une hiérarchie, dans la mesure où l'analyse se fait sur un seul niveau de hiérarchie qui est l'année. Par conséquent, bien que nous ayons généré la dimension temps avec les années, les mois et les jours, nous avons jugé plus sage de se limiter à une seule table pour cette dimension et de ne pas générer de hiérarchie.

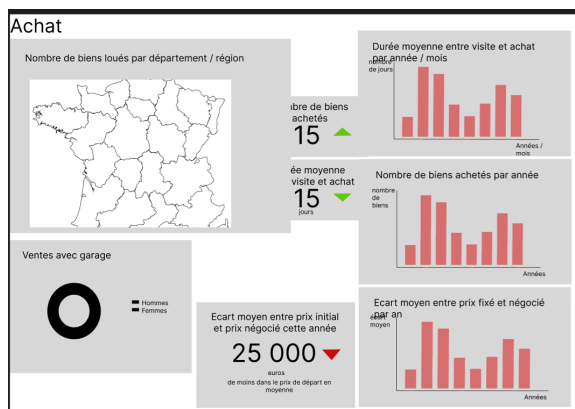
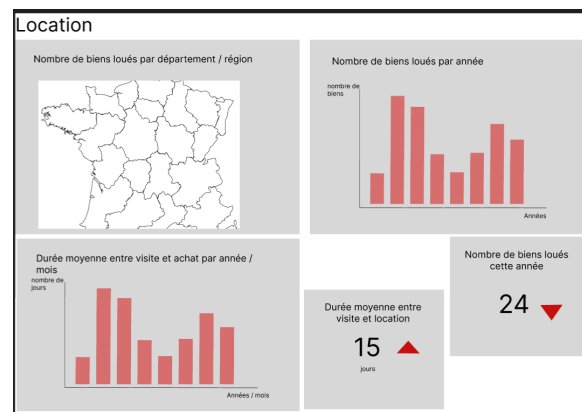
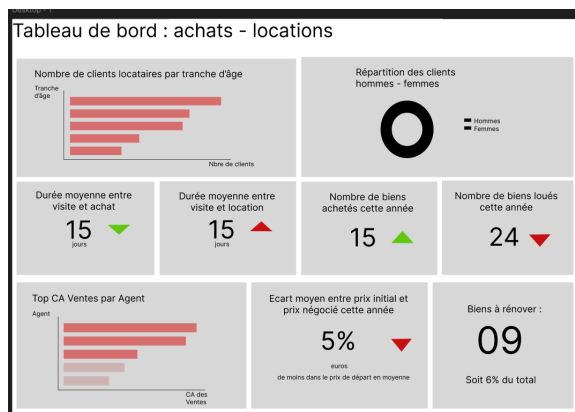
Un autre choix important que nous avons eu à faire est celui de créer deux tables de faits. Etant donné que les faits analysés sont très similaires, mais quand même distincts, et que même si à un moment une analyse jointe des deux types de faits que sont la location et la vente de biens immobiliers était nécessaire, il aurait fallu les distinguer pour pouvoir en étudier les spécificités séparément, nous avons décidé de les scinder en deux tables de faits autour desquelles gravitent les mêmes tables de dimension. D'abord par souci de logique, parce que ce sont quand même des faits différents avec certains attributs différents, mais aussi par souci de simplicité de compréhension et d'évolutivité du modèle dans le cas par exemple où il faudrait rajouter un indicateur qui serait uniquement pertinent pour un seul des faits, pour ne pas avoir à remplir la moitié de la table de nulls.

La dimension lots a finalement été supprimée et l'attribut Id_lot a été intégré à la table de faits des achats.

Cette modélisation a permis de générer le script SQL de création de la base de données qui a ensuite été remplie grâce aux flux ETL.

2.1.3. Maquettage

Nous avons décidé de passer par des maquettes pour visualiser le résultat que nous voulions obtenir pour aller plus vite dans la création du reporting, mais aussi pour choisir les graphiques appropriés à chaque indicateur. Voici un aperçu des maquettes réalisées.



Même si nous avons opéré des changements aussi bien dans la disposition et la répartition logique des graphiques, et que certains ont dû être éliminés ou modifiés parce que nous n'avions pas de moyens de les réaliser face à la réalité de l'outil ou des données (graphique indisponible ou autres), cette étape a été importante pour nous donner la direction dans laquelle nous voulions aller.

2.2. Réalisation

La phase de réalisation est relativement plus longue et plus riche que celle de conception, mais il faut noter que sa complexité est assez réduite par le temps passé à concevoir qui donne une certaine clarté sur cette étape. Elle est notamment composée de la création des flux ETL et du chargement du datawarehouse, et de la mise en place du rapport sur Power BI.

2.2.1. Flux ETL et chargement du datawarehouse

2.2.1.1. Mappings

La plupart des champs sources sont restés tels quels à part quelques légères modifications (voir flux). Les modifications majeures sont le plus souvent des indicateurs calculés. Cette table est le récapitulatif de tous les champs et de leur source

Table	Champ DWH	Champ source
dim_agent	Id_agent	id_agent
	nom_agent	nom_agent
	prenom_agent	prenom_agent
	Ranking	calculé
	nom_complet	calculé
dim_bien	id_bien	id_bien
	id_type	id_type
	id_etat	id_etat
	id_ville	id_ville
	nb_pieces	nb_pieces
	surface_bien	surface_bien
	prix_bien	prix_bien
	adresse_bien	adresse_bien
	codepostal_bien	codepostal_bien
	nbrebiensrenover	calculé

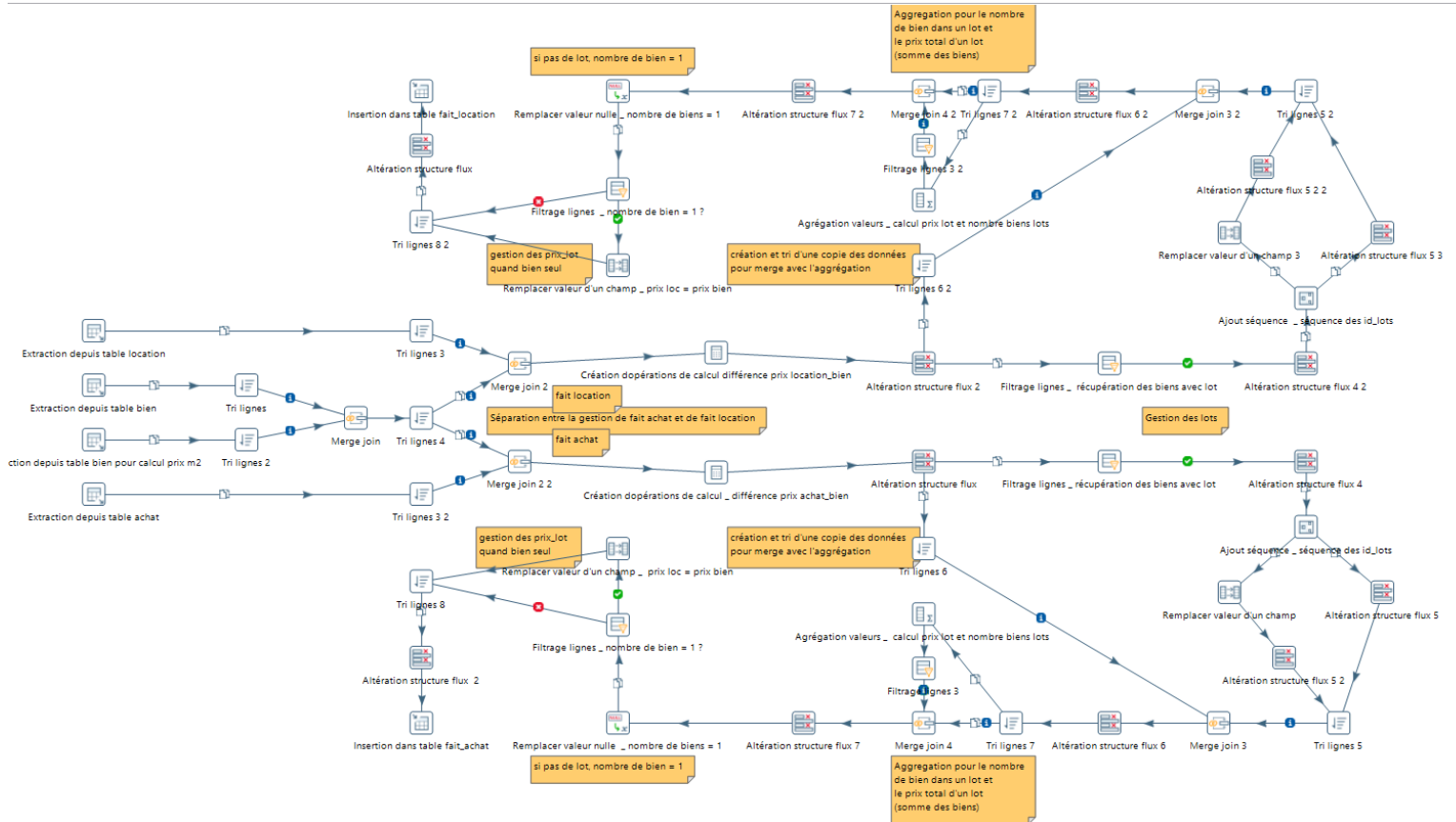
dim_ville	id_ville	id_ville
	id_departement	id_departement
	nom_ville	nom_ville
dim_departement	id_departement	id_departement
	id_region	id_region
	nom_departement	nom_departement
dim_region	id_region	id_region
	nom_region	nom_region
dim_etat	id_etat	id_etat
	nom_etat	nom_etat
dim_type	Id_type	Id_type
	nom_type	nom_type
dim_client	id_client	id_client
	id_civ	id_civilite
	id_ville	id_ville
	nom_client	nom_client
	prenom_client	prenom_client
	age	calculé
	tranche_age	calculé
	date_naissance	date_naissance
dim_civilite	id_civilite	id_civilite
	nom_civilite	nom_civilite
dim_date	date	générée
	mois	
	mois_text	
	jour	
	annee	

fait_achat	date_visite	date_visite
	date_achat	date_achat
	id_bien	id_bien
	id_lot	id_lot
	id_agent	id_agent
	id_client	id_client
	prix_achat	prix_achat
	prix_lot	calculé
	duree_visite_achat	calculé
	nombre_bien_lot	calculé
	ecart_prixachat_prixbien	calculé
	prix_m2	calculé
	nb_biens_lot	calculé
	prix_bien	calculé
	surface	calculé
fait_location	date_visite	date_visite
	date_location	date_location
	id_bien	id_bien
	id_lot	id_lot
	id_agent	id_agent
	id_client	id_client
	prix_location	prix_location
	prix_lot	prix_lot
	duree_visite_achat	calculé
	ecart_prixloc_prixbien	calculé
	prix_m2	calculé

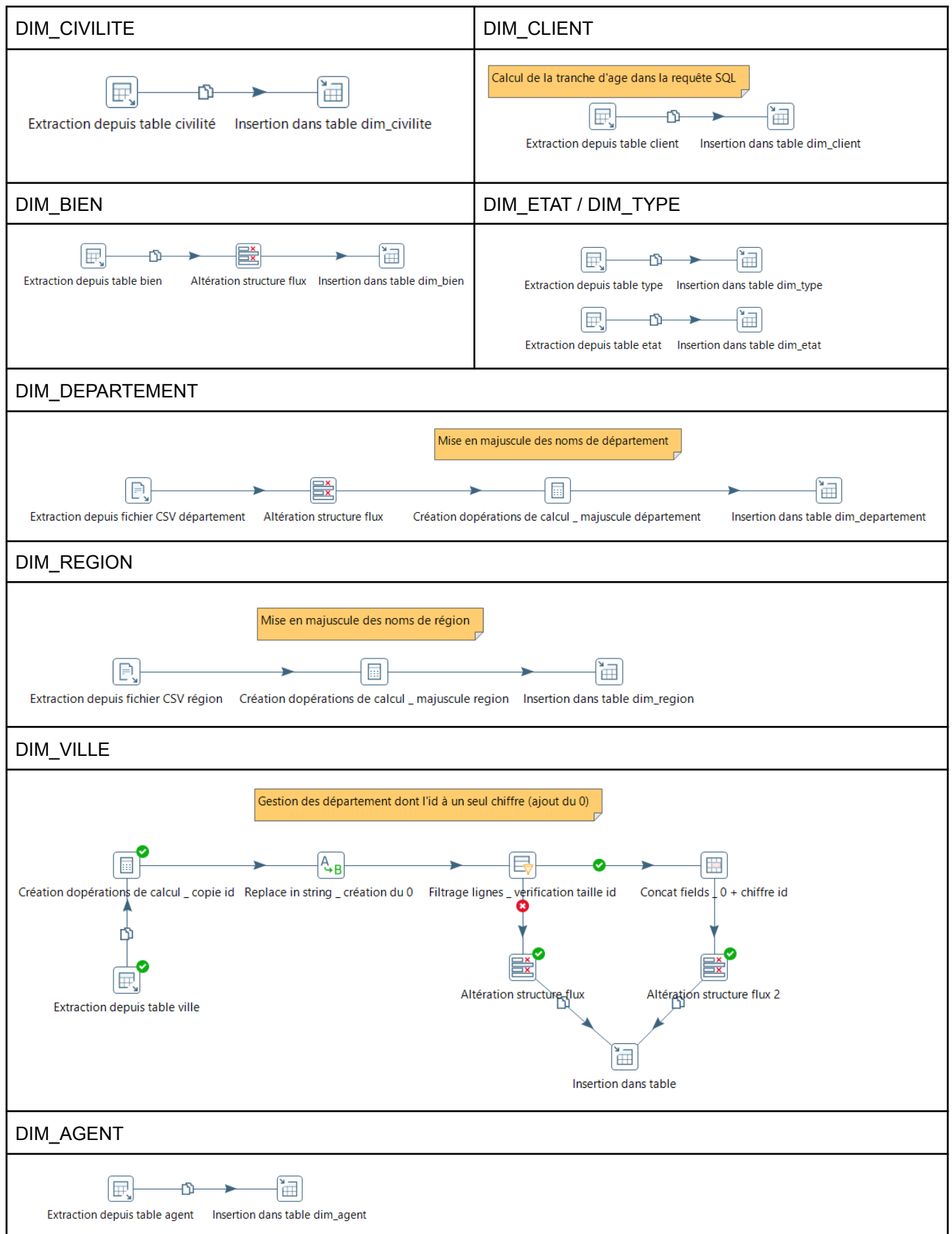
2.2.1.2. Flux

Tous les flux ont été exécutés le jour de la soutenance de présentation du projet soit le mercredi 29 mai à midi.

Flux d'alimentation des tables de faits

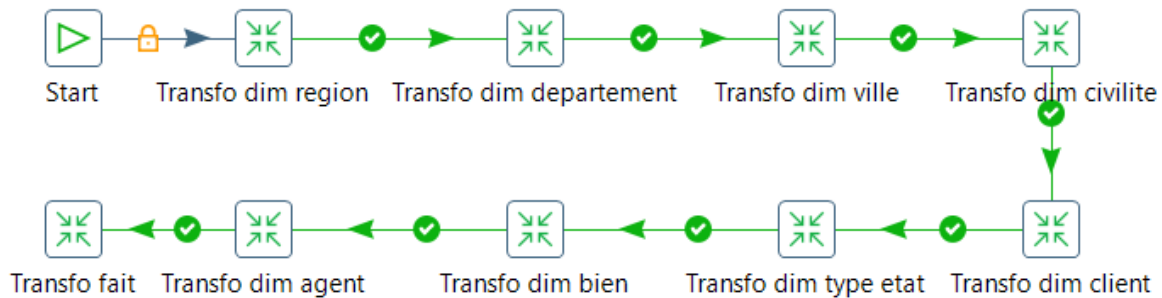


Flux d'alimentation des tables de dimensions



Job d'exécution des tâches

Ce job se charge d'exécuter les uns après les autres les différentes transformations d'alimentation des tables (dimensions et faits).



2.2.1.3. Modifications opérées

Les modifications effectuées présentées dans cette section l'ont été grâce à l'outil d'ETL. Nous avons choisi de calculer certains attributs dans l'ETL et d'autres sur power bi en fonction du besoin d'agrégation qu'ils nécessitent. Les indicateurs qui ne nécessitent pas d'agrégation, qui se calculent donc pour chaque ligne sont calculés dans l'etl tandis que les indicateurs et les mesures qui impliquent une agrégation de plusieurs lignes sont calculées sur Power BI.

a. Calcul de l'âge et ajout de la tranche d'âge

Pour la tranche d'âge, nous avons décidé de la calculer via une requête SQL qui va chercher l'âge dans la table des clients (soustraction de l'année de naissance et de l'année courante) et qui va, grâce à un switch case, placer dans les différentes catégories en fonction de cet âge calculé. Nous avons émis l'hypothèse que la tranche d'âge serait calculée en fonction de l'année courante et l'année de naissance sans prendre en compte les mois.

Script SQL :

```
SELECT
    EXTRACT(YEAR FROM current_date) - EXTRACT(YEAR FROM date_naissance) as
    age,
    (CASE WHEN EXTRACT(YEAR FROM current_date) - EXTRACT(YEAR FROM
    date_naissance) < 31 AND EXTRACT(YEAR FROM current_date) - EXTRACT(YEAR
    FROM date_naissance) >20 THEN '20-30'
        WHEN EXTRACT(YEAR FROM current_date) - EXTRACT(YEAR FROM
    date_naissance) > 30 and EXTRACT(YEAR FROM current_date) - EXTRACT(YEAR
    FROM date_naissance) < 51 THEN '31-50'
        ELSE 'plus de 50' END)
    AS tranche_age
FROM
    rbondaz.client;
```

b. Ajout de la dimension région et gestion de la dimension départements

Pour les dimensions de régions et de départements, nous avons décidé de repartir de 0 et de ne pas utiliser les données présentes en base (car régions pas présente et donc impossibilité de faire le lien entre départements et régions). Pour récupérer ces données, nous avons décidé d'utiliser comme source un CSV proposé par data.gouv et de le transférer dans les dimensions via l'ETL.

c. Ajout des indicateurs durée entre visite et location / achat;

Pour les indicateurs de durée entre la visite et la location ou l'achat nous avons fait la différence entre la date d'achat/location et la date de visite directement dans la requête SQL de récupération des informations de location et d'achat.

calcul : $\text{date_loc} - \text{date_visite_loc}$

d. Ecart entre prix de vente et négocié;

Pour les indicateurs de différence de prix entre le bien et l'achat ou la location, nous avons utilisé une "création d'opération de calcul" sur l'ET et nous avons fait la différence entre le prix du bien et le prix de l'achat / location.

Calcul : $\text{Prix_bien} - \text{prix_achat}$ (ou $\text{Prix_bien} - \text{prix_location}$)

e. Prix au m2 :

Pour l'indicateur "prix au m2", nous avons créé une nouvelle colonne dans les table de fait et nous avons divisé le prix du bien (et pas le prix de location ni d'achat, c'est un choix que l'on a fait pour avoir le prix au m2 du bien en prenant en compte sa vraie valeur, non négociée) par la surface en m2

Calcul : $\text{Prix_bien} / \text{surface}$

f. Lots :

Pour la gestion des lots, nous avons décidé de ne pas suivre ce qui était fait dans les données initiales (mettre en id lot l'id d'un autre bien qui est lié au bien initial). Nous avons pris la décision de créer nous même une séquence pour créer des id et attribuer les même id pour tous les biens d'un même lot (exemple : un garage et un appartement appartenant au même lot auront tous les deux le même id_lot (1 par exemple)).

2.2.2. Visualisation

2.2.2.1. Indicateurs calculés

Ces indicateurs calculés sur power BI le sont dans le cadre d'une agrégation ou d'une étape intermédiaire pour arriver à une agrégation. En plus de ces indicateurs, nous avons décidé de créer la dimension temporelle sur Power BI. Les fonctions sont écrites en dax. Certains indicateurs qui ne nécessitent pas d'agrégation complexe (moyennes, sommes) tels que la répartition en pourcentage homme/femme du nombre de clients, ou le prix moyen des biens, sont calculés automatiquement par Power BI. Voici les calculs d'indicateurs réalisés sur Power BI :

a. Nom des agents

Ce champ a été calculé dans la dimension des agents afin d'afficher le nom complet de chaque agent pour les graphiques et les filtres au lieu d'afficher uniquement le prénom ou le nom.

b. Rang des agents

Ce champ a été calculé dans la dimension des agents pour arriver à afficher dans l'ordre décroissant les chiffres d'affaires pour chaque agent en mettant en évidence le top 3.

Voici ci-dessous le résultat sans la colonne classement suivi du résultat avec la colonne classement.

Le calcul effectué est le suivant :

```
1 ranking = RANKX(all(DIM_AGENT), CALCULATE(SUM(FAIT_ACHAT[PRIX_ACHAT])),,DESC)
```

La fonction utilisée est la fonction RANKX. nous ne la connaissions pas; nous avons donc dû faire des recherches pour la découvrir et comprendre son fonctionnement.

c. Nombre de biens à rénover

```
1 Count of ID_BIEN for 4 =
```

```
2 CALCULATE(COUNTA('DIM_BIEN'[ID_BIEN]), 'DIM_BIEN'[ID_ETAT] IN { 4 })
```

Cette mesure a été réalisée en réponse au besoin exprimé de connaître le nombre de biens à rénover.

d. Nombre de biens à vendre avec un lot supplémentaire

Nous avons calculé cette mesure dans le but de pouvoir aboutir à la suivante. Pour ce faire, nous avons compté le nombre d'id_lot distincts (car 1 id_lot distinct = un groupement de biens en lots)

e. Nombre de biens ayant un garage

Pour calculer le nombre de bien ayant un garage, nous avons calculé le nombre de id_bien qui ont un id_lot (donc lié à un bien) et nous avons filtré par type "garages" (via le lien avec la dimension type).

f. Total des biens loués et vendus

```
1 total biens loués et vendus =  
2 COUNTA('FAIT_ACHAT'[ID_BIEN]) + COUNTA('FAIT_LOCATION'[ID_BIEN])
```

Cet indicateur n'était pas un besoin exprimé, mais une initiative ; nous avons jugé intéressant de fournir ce nombre sur une carte et montrer sa décomposition grâce à un graphique (voir graphiques choisis et rendu visuel.)

g. Dimension temps

Nous avons essayé de créer cette dimension sur l'ETL, mais sans succès. Pour ne pas perdre trop de temps, nous avons décidé de la créer sur Power BI puisque nous savions déjà comment le faire. La formule utilisée pour ce faire est la suivante :

```
1 DIM_DATE = CALENDAR(  
2   IF(MIN(FAIT_ACHAT[DATE_VISITE_ACHAT_TRIMMED]) < MIN(FAIT_LOCATION[DATE_LOC_TRIMMED]), MIN(FAIT_ACHAT[DATE_VISITE_ACHAT_TRIMMED]), MIN  
   (FAIT_LOCATION[DATE_LOC_TRIMMED])),  
3   IF(MAX(FAIT_ACHAT[DATE_VISITE_ACHAT_TRIMMED]) < MAX(FAIT_LOCATION[DATE_LOC_TRIMMED]), MAX(FAIT_LOCATION[DATE_LOC_TRIMMED]), MAX(FAIT_ACHAT  
   [DATE_VISITE_ACHAT_TRIMMED])))
```

La fonction utilisée est la fonction Calendar qui prend en paramètres deux dates et génère les dates comprises entre ces deux dates. Pour choisir ces dates, une comparaison est faite entre les dates pour trouver la plus antérieure et la plus récente et les définir comme bornes.

Ensuite il a fallu générer les autres attributs à savoir l'année, le mois et le jour pour chaque date dans le but d'éventuellement générer une hiérarchie automatique. Les attributs sont calculés grâce aux fonctions year, month et day qui extraient l'année, le mois ou le jour

```
1 annee = YEAR(DIM_DATE[Date])
```

```
1 mois = MONTH(DIM_DATE[Date])
```

```
1 jour = DAY(DIM_DATE[Date])
```

```
1 mois_text = SWITCH (  
2   [mois],  
3   1, "Janvier",  
4   2, "Février",  
5   3, "Mars",  
6   4, "Avril",  
7   5, "Mai",  
8   6, "Juin",  
9   7, "Juillet",  
10  8, "Août",  
11  9, "Septembre",  
12  10, "Octobre",  
13  11, "Novembre",  
14  12, "Décembre",  
15  "Unknown month number"  
16 )
```

2.2.2.2. Graphiques choisis

- a. Le nombre de clients locataires par tranche d'âge (20 à 30 ans, 31 à 50 ans et plus de 50 ans)
→ Diagramme en barres horizontales ordonnée par ordre décroissant
Ce graphique a été choisi dans l'optique de représenter un classement des tranches d'âge de la plus représentée à la moins représentée parmi les clients locataires.
- b. Le nombre de biens à rénover
→ card
Etant donné qu'il s'agit d'une agrégation qui peut varier en fonction de la dimension géographique, nous avons choisi de montrer sa valeur de telle sorte qu'elle varie avec les actions de sélection et de désélection des régions et/ou départements sur la carte qui est sur la même page (voir rendu visuel)
- c. La répartition des clients homme femmes (en pourcentage)
→ donut
Ce graphique est typique pour représenter la répartition d'un indicateur en fonction d'attributs, c'est pourquoi nous l'avons choisi.
- d. Le nombre de biens achetés ou loués par département et par région (avec un alerteur pour les valeurs au-dessus de 5)
→ carte
Pour cet indicateur, nous avons choisi une carte puisqu'il est représenté en fonction de la dimension géographique, pour une compréhension plus rapide de la donnée.
- e. Evolution chiffres d'affaires des biens achetés par année
→ diagramme en barres
Etant donné que le but ici est de montrer une évolution, et que les valeurs, représentées par année sont discrètes et non continues, nous avons opté pour un diagramme en barres verticales. Si les valeurs avaient été continues, nous aurions choisi une courbe d'évolution.
- f. La durée moyenne (en jours) entre la date de visite et la date de location ou d'achat et
- g. Le nombre de biens à la vente proposant la vente d'un lot supplémentaire de type garage
→ card
Nous avons choisi ce format pour afficher les indicateurs de manière claire mais aussi parce qu'il s'agissait de valeurs uniques car agrégées pour lesquelles ce format est approprié et permet de voir les variations de valeurs en fonction des filtres appliqués.

Par année :

- Le chiffre d'affaires des biens vendus par agent en mettant évidence le Top 3
→ diagramme en barres horizontales par ordre décroissant
Ce graphique a été choisi pour mettre en évidence le classement.
- L'écart moyen entre le prix de vente du bien et le prix d'achat négocié par le client,
- Le prix moyen des biens loués,
- Le prix moyen de vente au mètre carré
→ diagramme en barres car représentation par année
- Le nombre de biens vendus
et
- Le nombre de biens loués
→ card

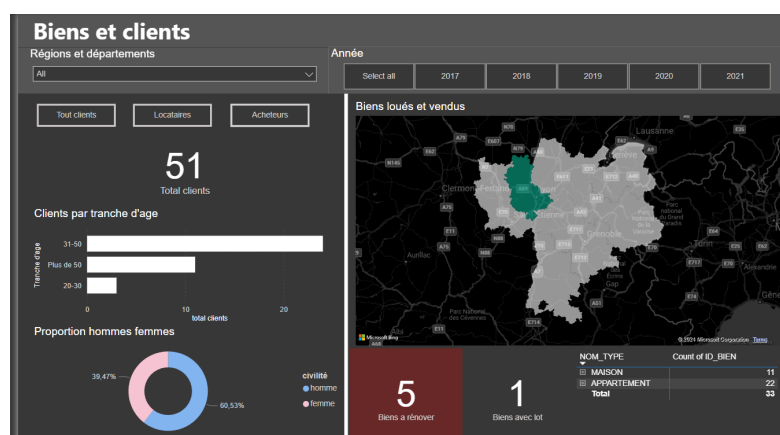
2.2.2.3. Rendu visuel

Notre application est constituée de 3 pages qui représentent chacune des sujets différents :

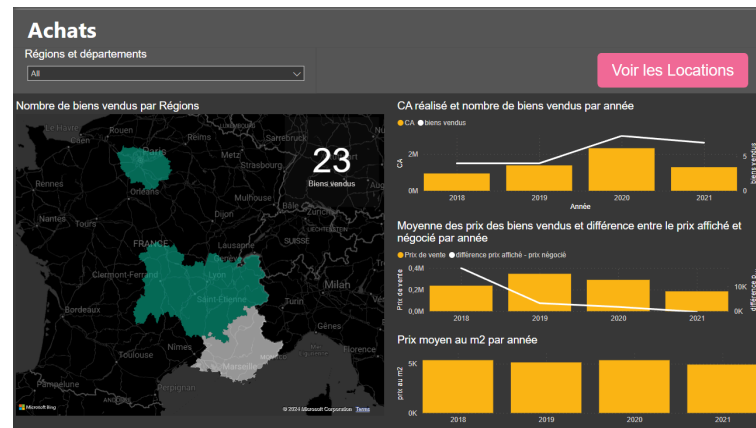
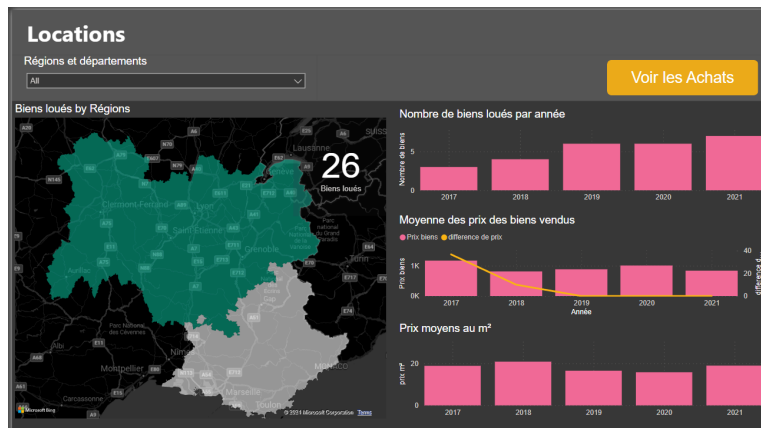
La première montre les informations relatives aux agents et aux chiffres réalisés par l'agence;



La deuxième concerne plutôt les informations relatives aux biens et aux clients

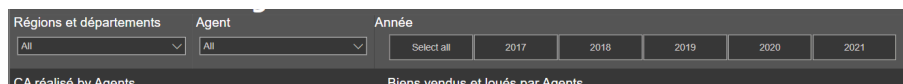


Et la troisième montre les détails des analyses des locations et achats en deux vues navigables à l'aide d'un bouton :

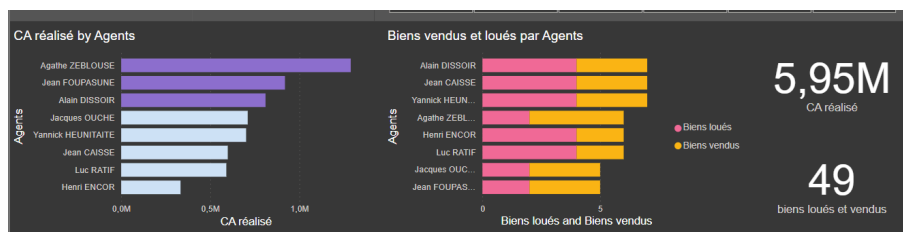


Dernière exécution : 04/06/2024

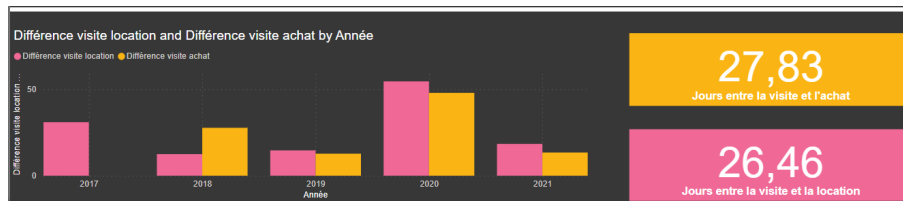
a. Performances et agents



Les filtres mis en place pour cette page sont basés sur la dimension agents, les hiérarchies spatiales et temporelles ; car ils sont en lien avec les thématiques de la page.



Nous avons choisi de mettre côte à côte deux classements des agents différents mais complémentaires. Le premier, qui se base sur le chiffre d'affaires réalisé, fait partie du besoin exprimé alors que le second se base sur le nombre de biens vendus et loués et provient de notre initiative. Nous avons également jugé intéressant de montrer le CA total et le nombre total de biens loués et vendus qui est un indicateur calculé de notre initiative également, qui sont également en cohérence avec la thématique des performances des agents.



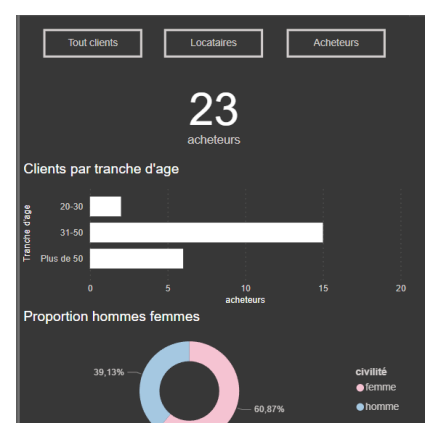
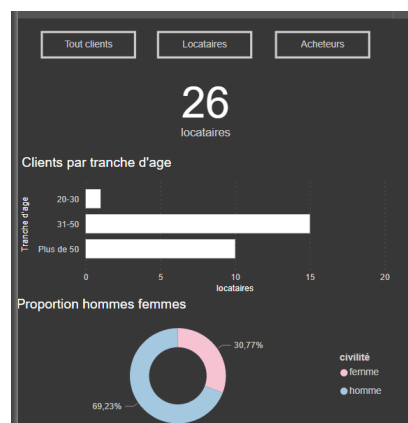
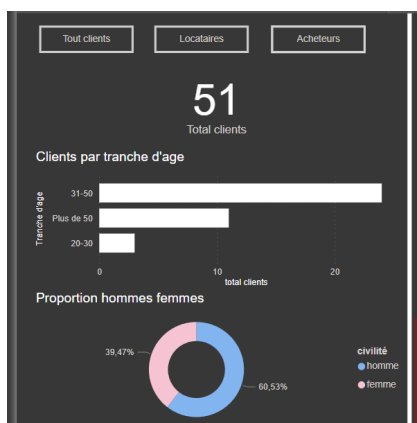
Parmi les performances nous avons les durées moyennes et par année entre la visite et l'achat ou la location que nous avons décidé de placer sur la même page que les agents pour pouvoir les voir varier en fonction des agents sans en faire un graphique particulier.

Concernant la charte graphique, nous avons décidé de représenter les achats par une couleur et les locations par une autre sur l'ensemble du reporting à savoir l'orange et le rose.

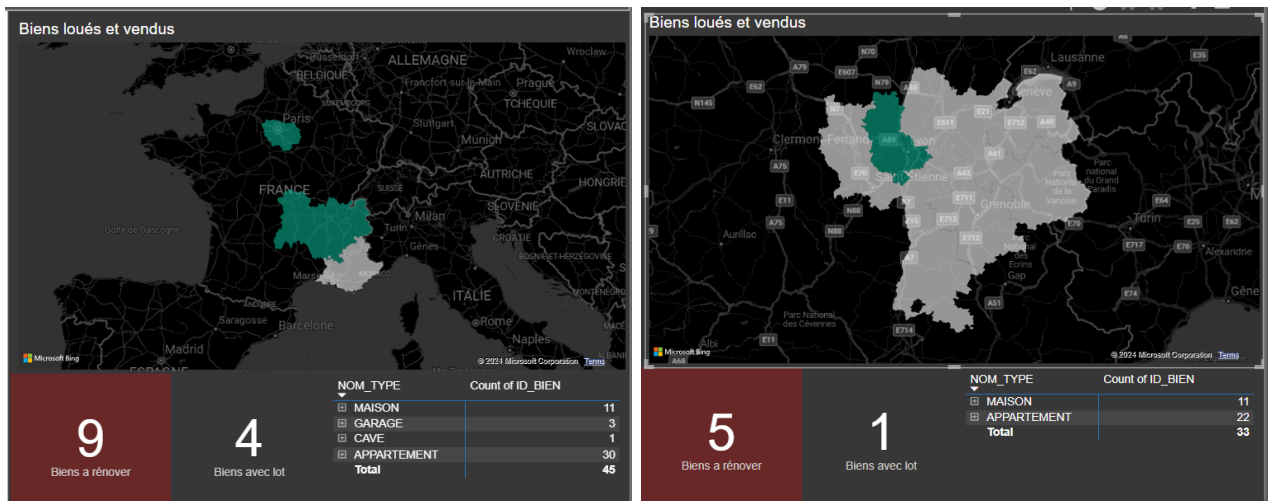


b. Biens et clients

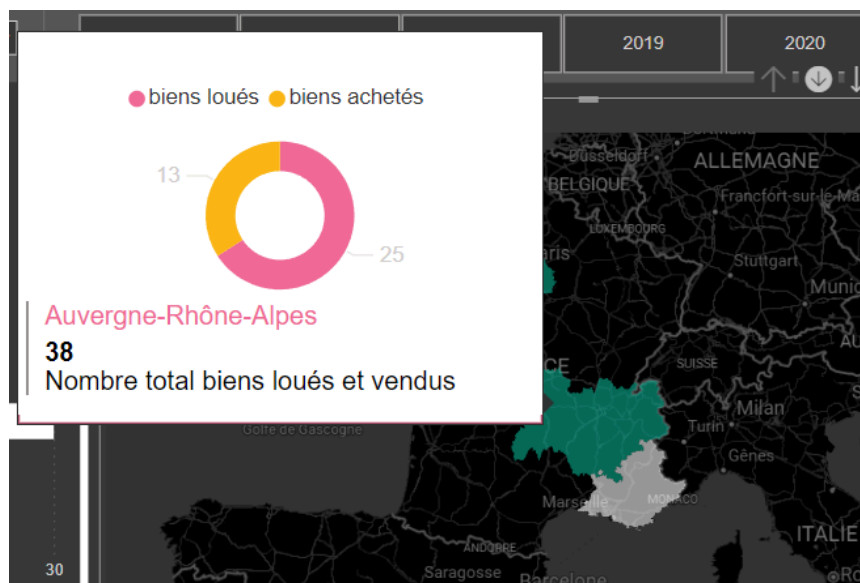
Pour cette partie les filtres sont presque pareils, basés sur les dimensions temporelle et spatiale. Le tableau de bord est divisé en deux parties qui interagissent quand même : les clients et les biens. L'analyse des clients se fait sur trois axes principaux : la répartition en fonction de la tranche d'âge, le nombre, et le genre, tout cela selon un filtre basé sur leur statut de locataire ou de propriétaire grâce aux boutons du dessus.



L'analyse des biens est axée sur le type et l'état. Elle est enrichie par la carte interactive des biens loués et vendus qui fait varier ces indicateurs selon la région ou le département.



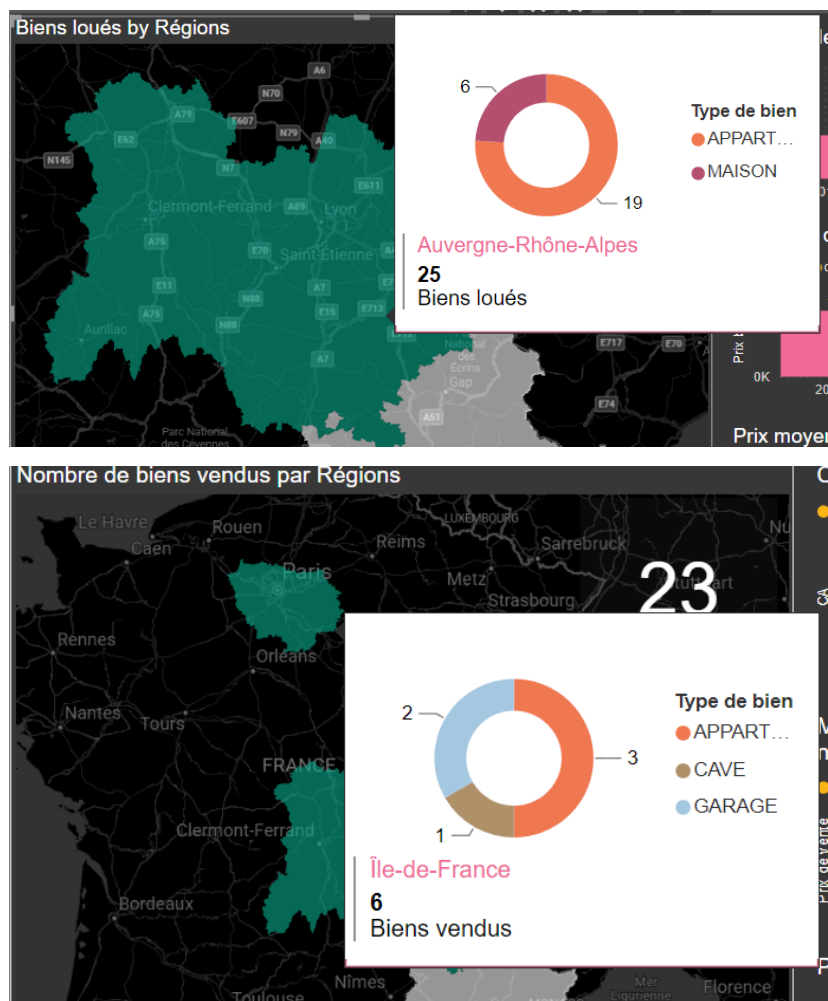
Nous avons également personnalisé l'effet au survol de la carte pour afficher à l'aide d'un donut la proportion d'achats et de locations



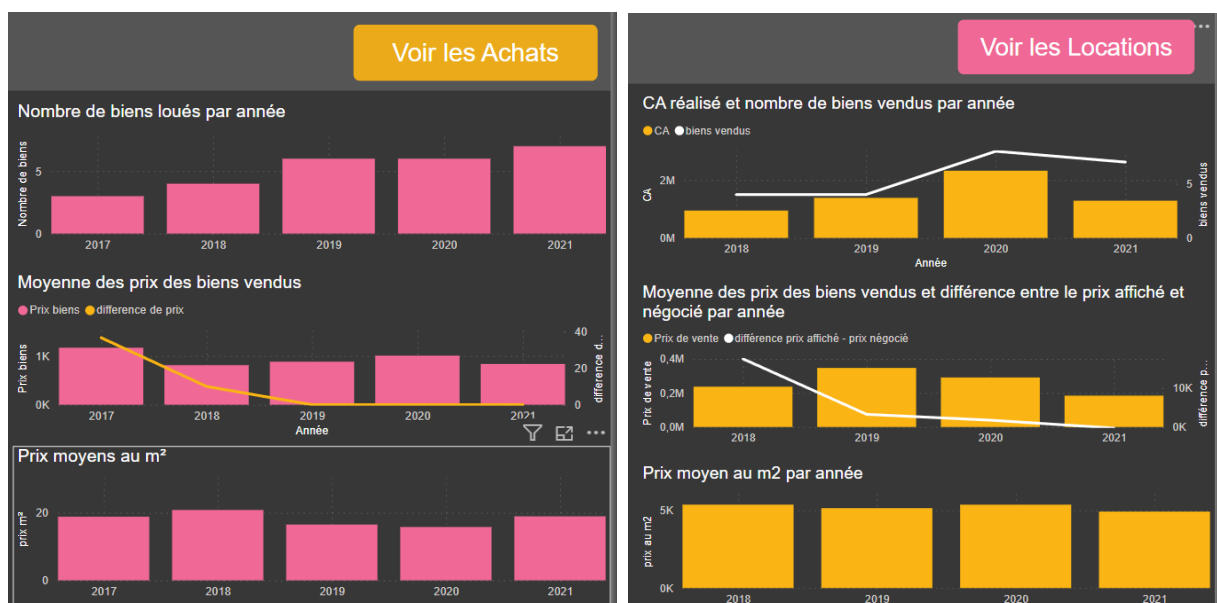
c. Location et achat

Pour l'analyse des deux faits nous avons opté pour un affichage similaire tout en conservant les spécificités de chaque fait aussi bien du point de vue graphique que logique. La carte que nous avons décidé d'inclure représente selon le cas les achats ou les locations avec une coloration pour les régions/départements au dessus de 5 achats ou locations.

Nous avons, là aussi, personnalisé l'effet au survol de la carte pour afficher à l'aide d'un donut la proportion occupée par chaque type de bien.



Les autres indicateurs représentés sont ceux qui par année représentent le nombre de biens, la moyenne des prix et les prix moyens au m2.



Conclusion

Ce projet a été une expérience enrichissante et nous a permis de mettre en pratique différentes compétences acquises au cours de l'année. En travaillant ensemble, nous avons pu explorer plusieurs aspects techniques, notamment l'initialisation de bases de données, la création de flux ETL, et la configuration d'outils BI.

En conclusion, ce projet nous a offert une opportunité précieuse de relier plusieurs matières de notre cursus, tout en développant des compétences pratiques et en consolidant notre capacité à travailler en équipe de manière productive et harmonieuse. Nous sommes satisfaits du travail accompli et des résultats obtenus.