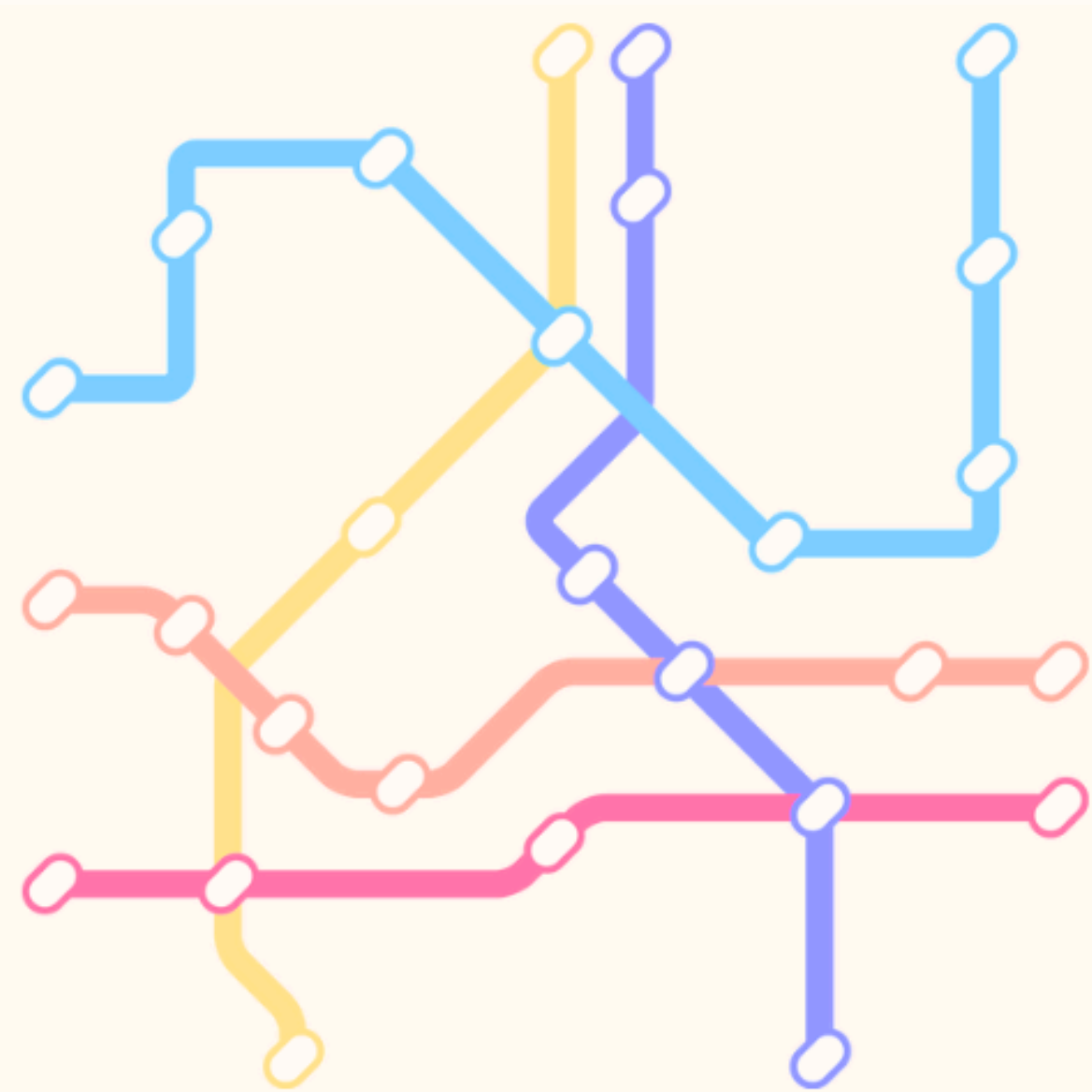


Présentation de mon expérience de stage



Stage de fin d'étude MSc Business Analytics à RATP CAP Île-de-France
(Juin à Novembre 2024)



RATP CAP
ÎLE-DE-FRANCE

Vincent Plateau



Contenu



1. Description de mon expérience professionnelle (slide 3 à 5)
2. Focus sur une mission (slide 6 et 7)
3. Apprentissages professionnels et personnels (slide 8 et 9)
4. Liens avec les cours du MSc Business Analytics (slide 10)
5. Conclusion (slide 11)
6. Annexe (slide 12)

RATP CAP IDF - Présentation générale

Contexte : RATP CAP Île-de-France est une filiale du groupe RATP créée en 2021 pour répondre à l'ouverture à la concurrence du marché des transports en Île-de-France. Cette entité est spécialisée dans la gestion et l'exploitation des réseaux de transport public (bus principalement) et se positionne pour répondre aux appels d'offres d'Île-de-France Mobilités (IdFM) sur les différents lots de transport franciliens.

Place dans le groupe RATP

RATP CAP en chiffres



12 millions de voyageurs
transportés chaque année
(20 millions de voyages en 2023)



1700 salariés
ayant pour mission de répondre aux
appels d'offres et d'opérer les réseaux

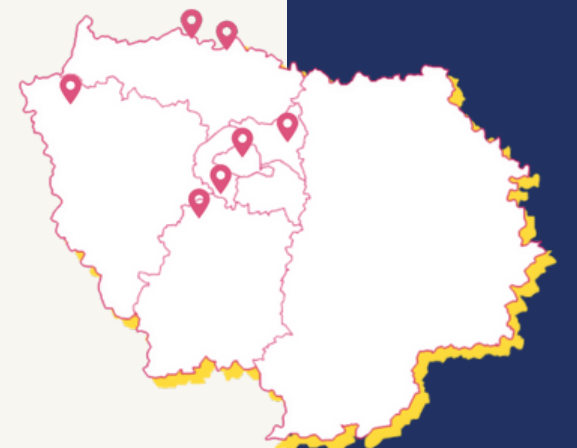
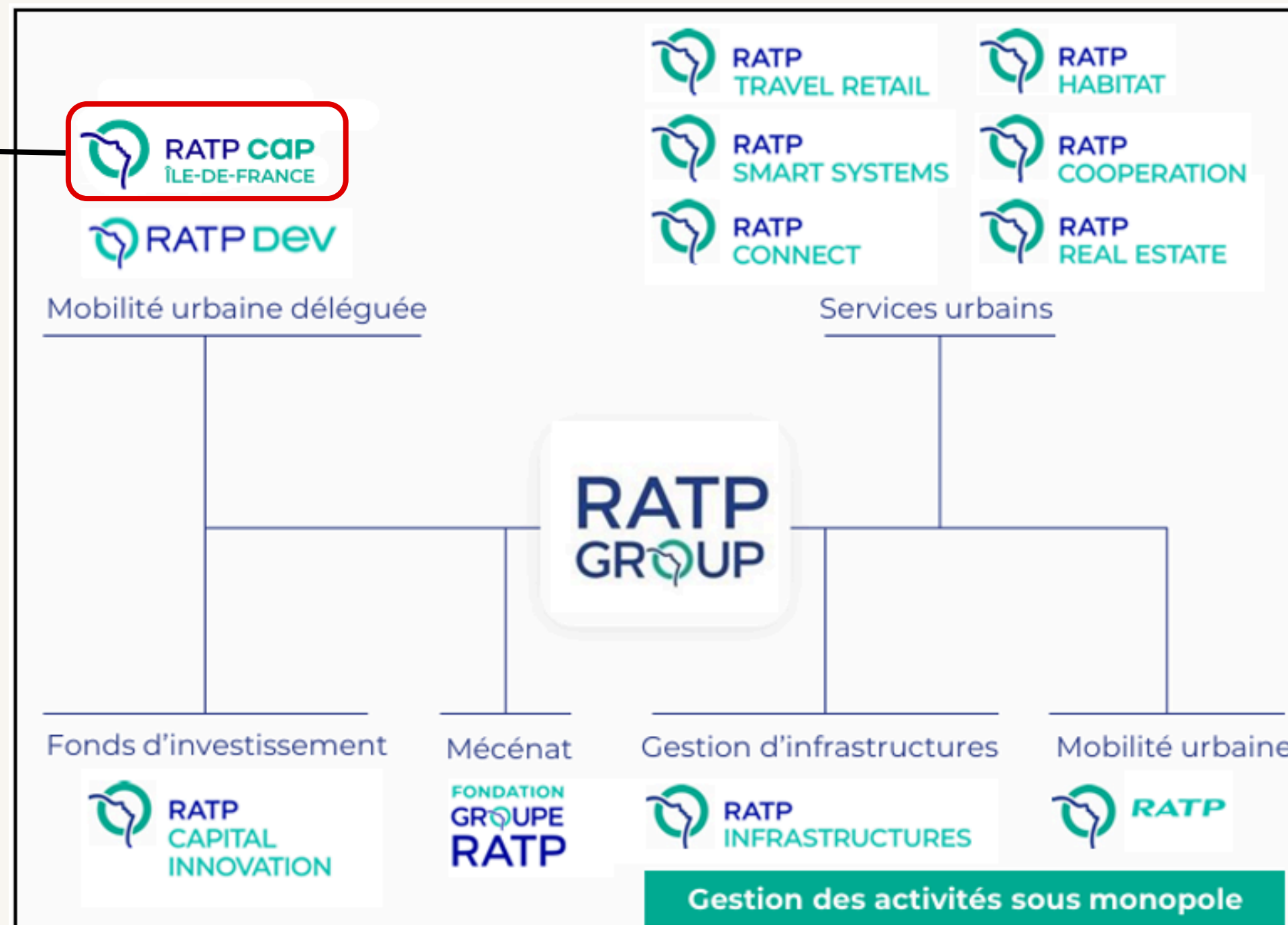


+ 750 véhicules exploités
et 15 millions de km parcourus
par an



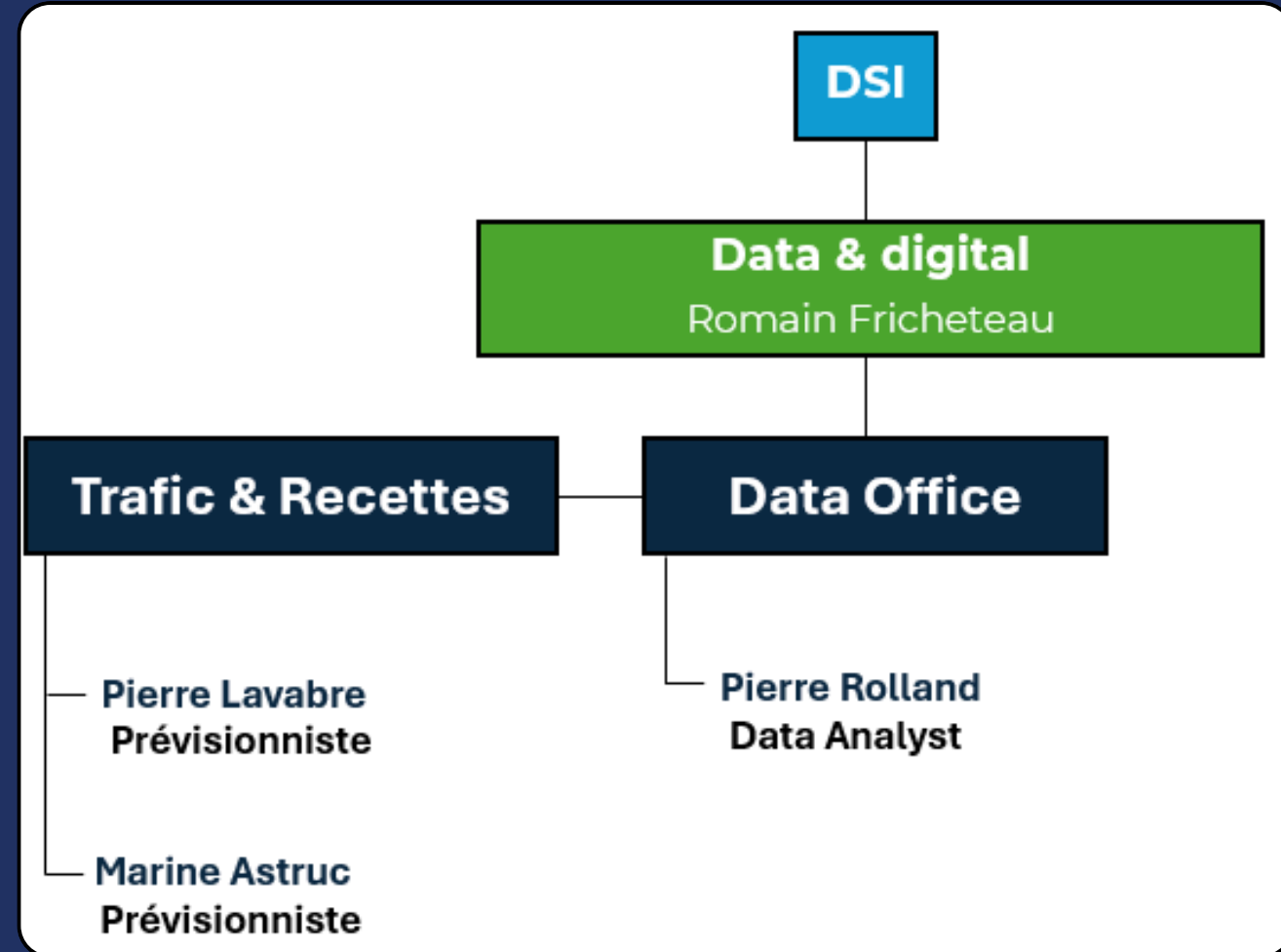
150 M€
de chiffre d'affaires

12 filiales en Île-de-France
(1 filiale dédiée par contrat)

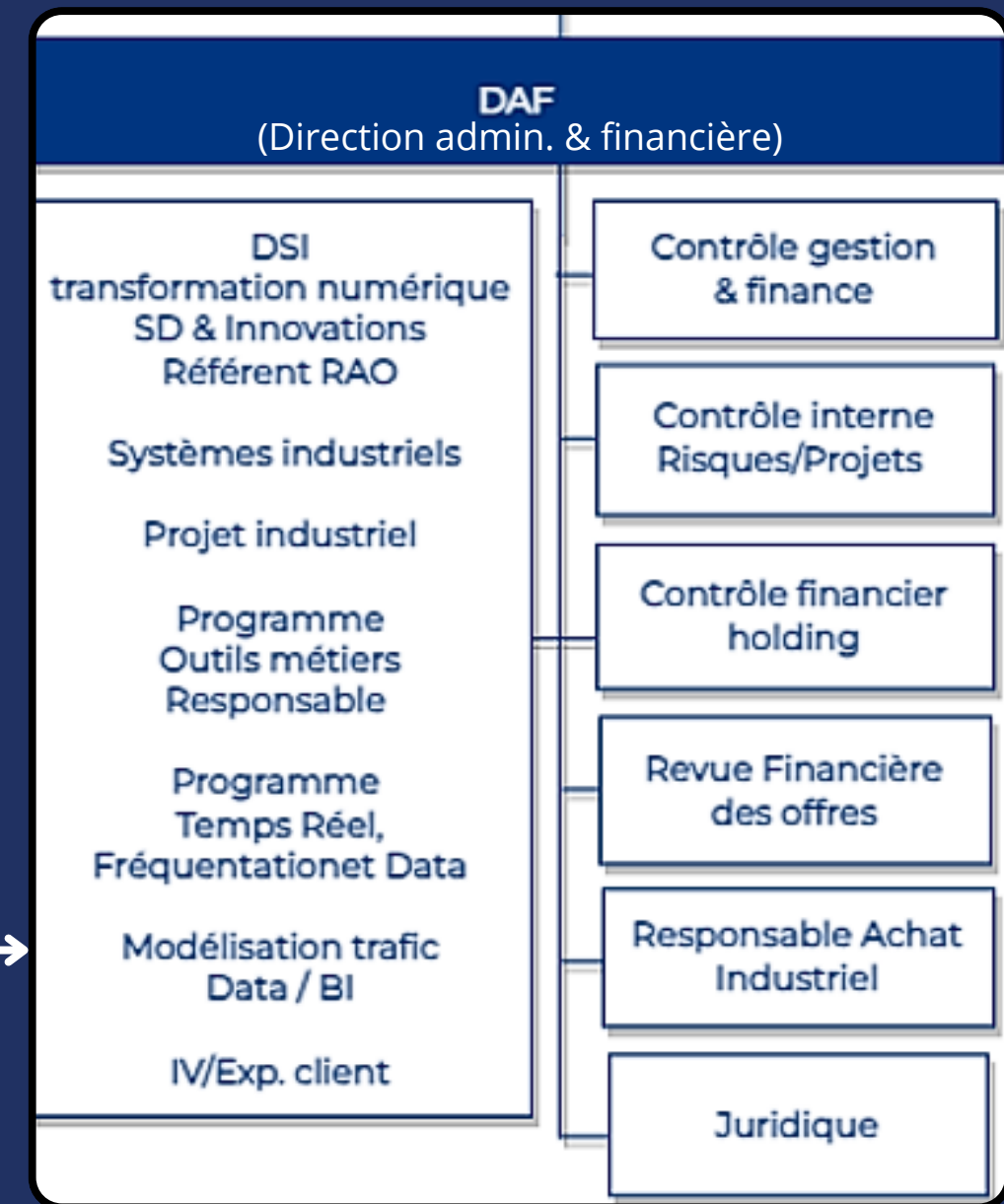


Mon équipe

Au sein de la Direction des Systèmes d'Information, composant de la DAF, l'équipe **Data & digital** apporte son soutien à divers acteurs en produisant des livrables répondant à des besoins exprimés en interne.



J'ai eu la chance de rejoindre cette équipe en tant que Data Analyst et de travailler avec **Marine Astruc, Chargée d'études Prévisions de Trafic**, qui m'a détaillé les projets sur lesquels j'allais travailler, m'a aidé et encadré tout au long du stage. Tous les membres de l'équipe on également su m'aider à un moment ou à un autre.



À l'échelle de l'entreprise, tous les projets ont un objectif premier : répondre à un appel d'offre ou participer à la bonne exploitation des réseaux gagnés. Les prévisionnistes passent la majorité de leur temps sur les RAO. Des sujets de fond émergent, tels que le suivi des validations télé-billettiques et la compréhension des tendances de mobilité en Île-de-France.

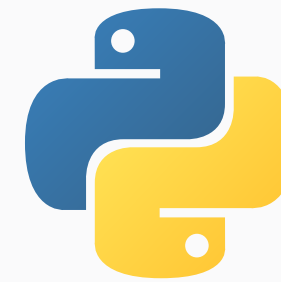


Mes missions

Deux projets ont été au centre de ce stage.

- Le premier, le plus important, consista en la création d'une application Qlik de visualisation, nommée "**DatAOd**esign" des données de validations télébillettiques remontées par les bus dans toute la région par le SIDV. Ces dashboards sont à destination des designers d'offres et ont été conçu pour répondre à leurs besoins. La version finale de ce projet a été livrée au bout de 4 mois, en septembre.
- Le deuxième projet entamé fin septembre a pour but d'exploiter et de valoriser les données de l'**enquête de mobilité par GPS (EMG)** réalisée par l'Institut Paris Région et achetées par l'équipe data au début de l'année 2024. Le but est de comprendre les prochaines grandes tendances qui façonneront la mobilité des Franciliens dans les années à venir. Le support, encore en développement, est une application Streamlit développée en Python.
- Enfin, j'ai également apporté mon soutien aux prévisionnistes de manière ponctuelle sur des sujets transversaux à l'équipe tels que le calcul de l'élasticité de l'offre, le requêtage du data lake ou des contributions personnelles comme la création de cartes à des fins d'analyse.

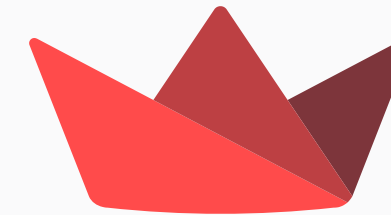
Les outils utilisés :



Python



Qlik



Streamlit



IBM Cognos / SIDV



Excel



Snowflake

Dashboard
QlikSense

Enquête EMG

Petites
contributions

Focus sur

Contexte

- **Remplacement d'un ancien outil** Power BI pour la visualisation des données de validations télébilletiques.
- **Besoin** critique des designers d'offre **d'accéder aux données de validation de bus** pour optimiser l'offre de transport en vue des RAO. Le but est de donner accès à des données déjà transformées et rapidement exploitables.
- Projet stratégique dans le contexte d'ouverture à la concurrence qui a vocation est être amélioré.

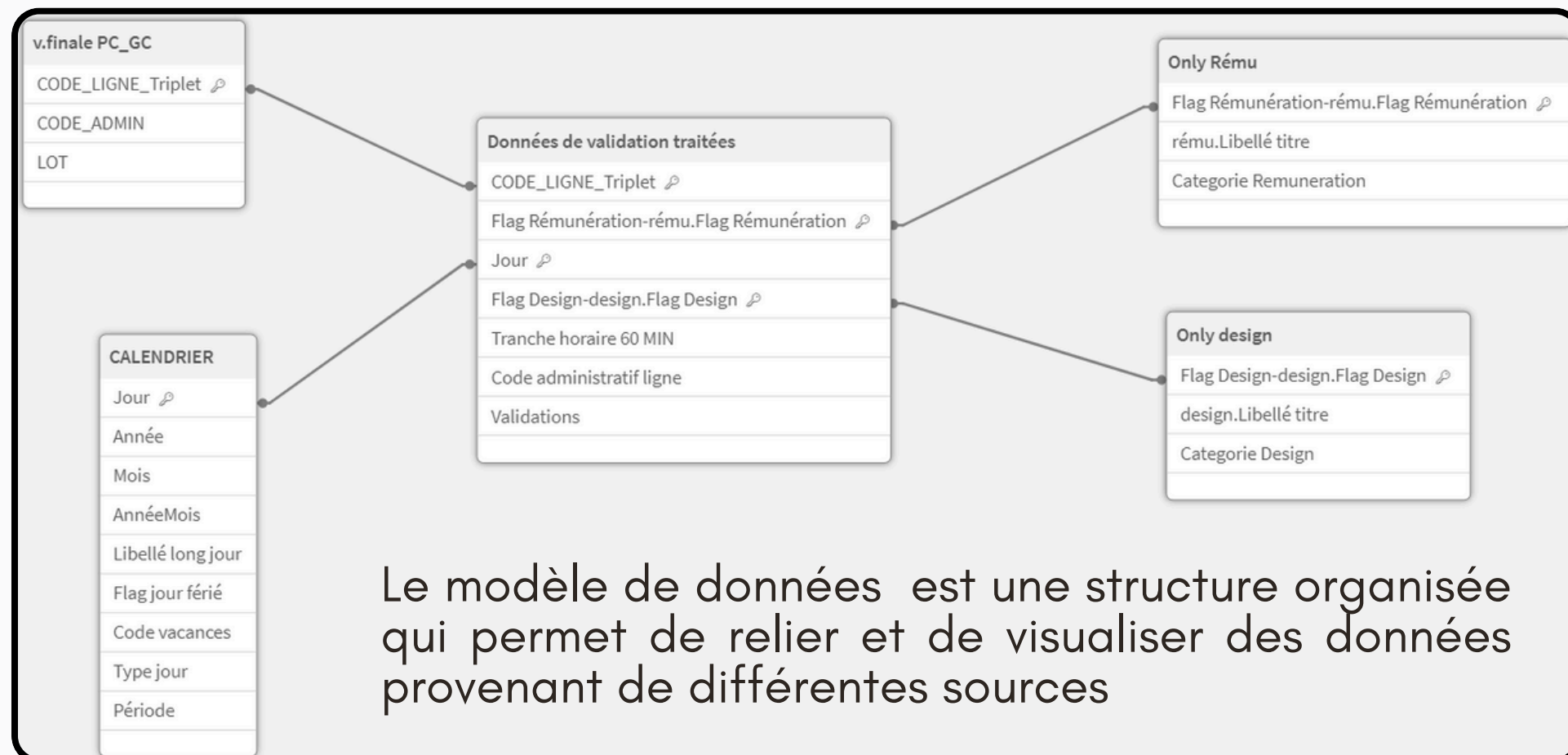
10 utilisateurs formés
19 Go de données traitées
500 heures de travail

DatAOdeshign

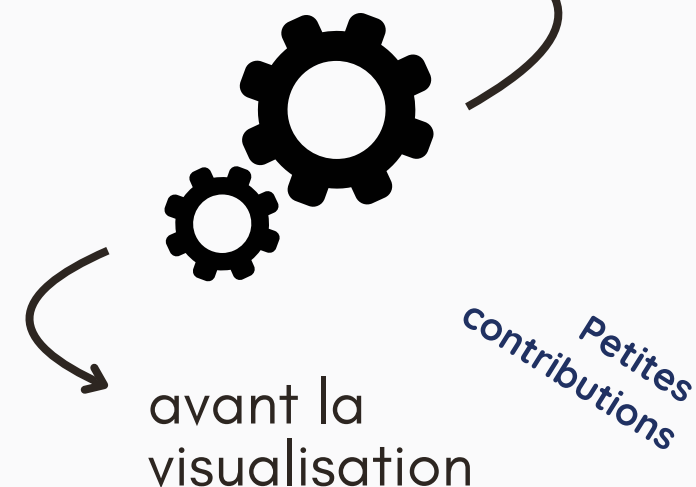
Architecture du projet



À l'intérieur de la machine...



Les données intégrées dans Qlik sont chargées via un script de chargement



Dashboard
QlikSense

Enquête EMG

Focus sur DatAOdésign

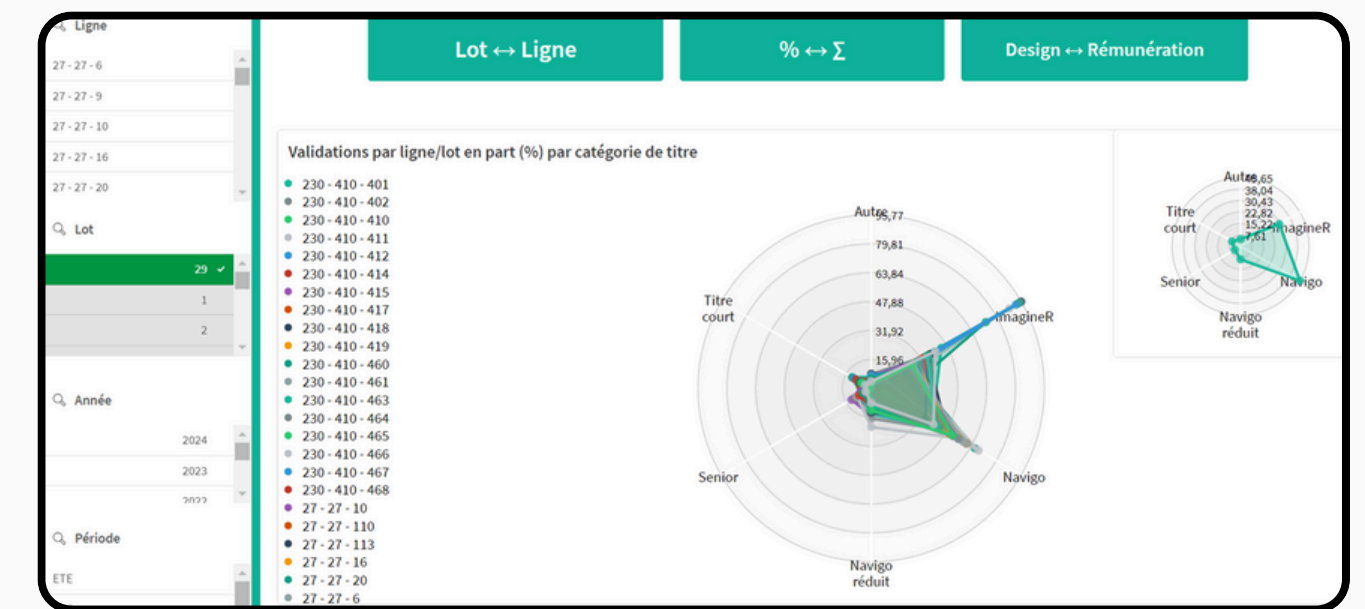
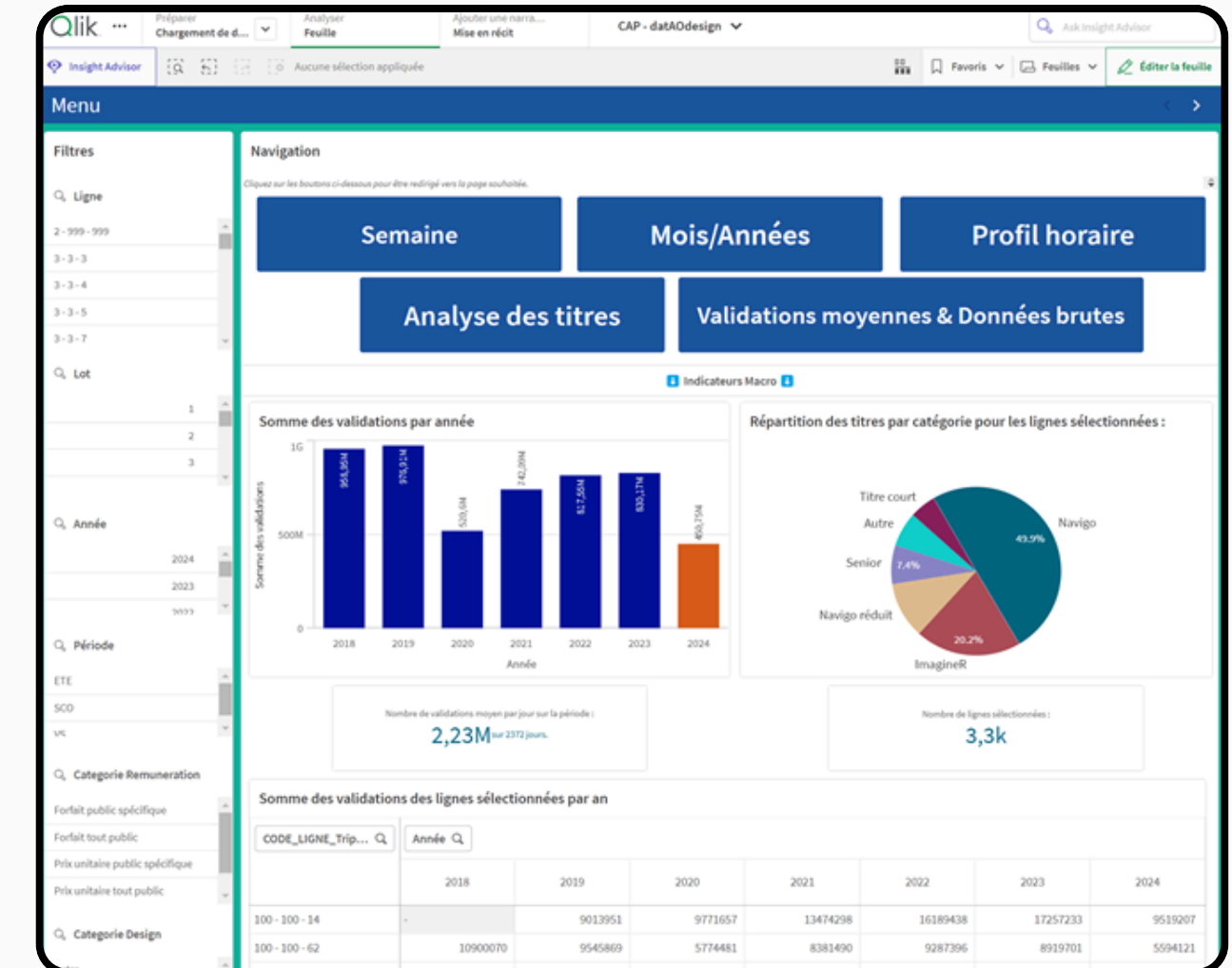
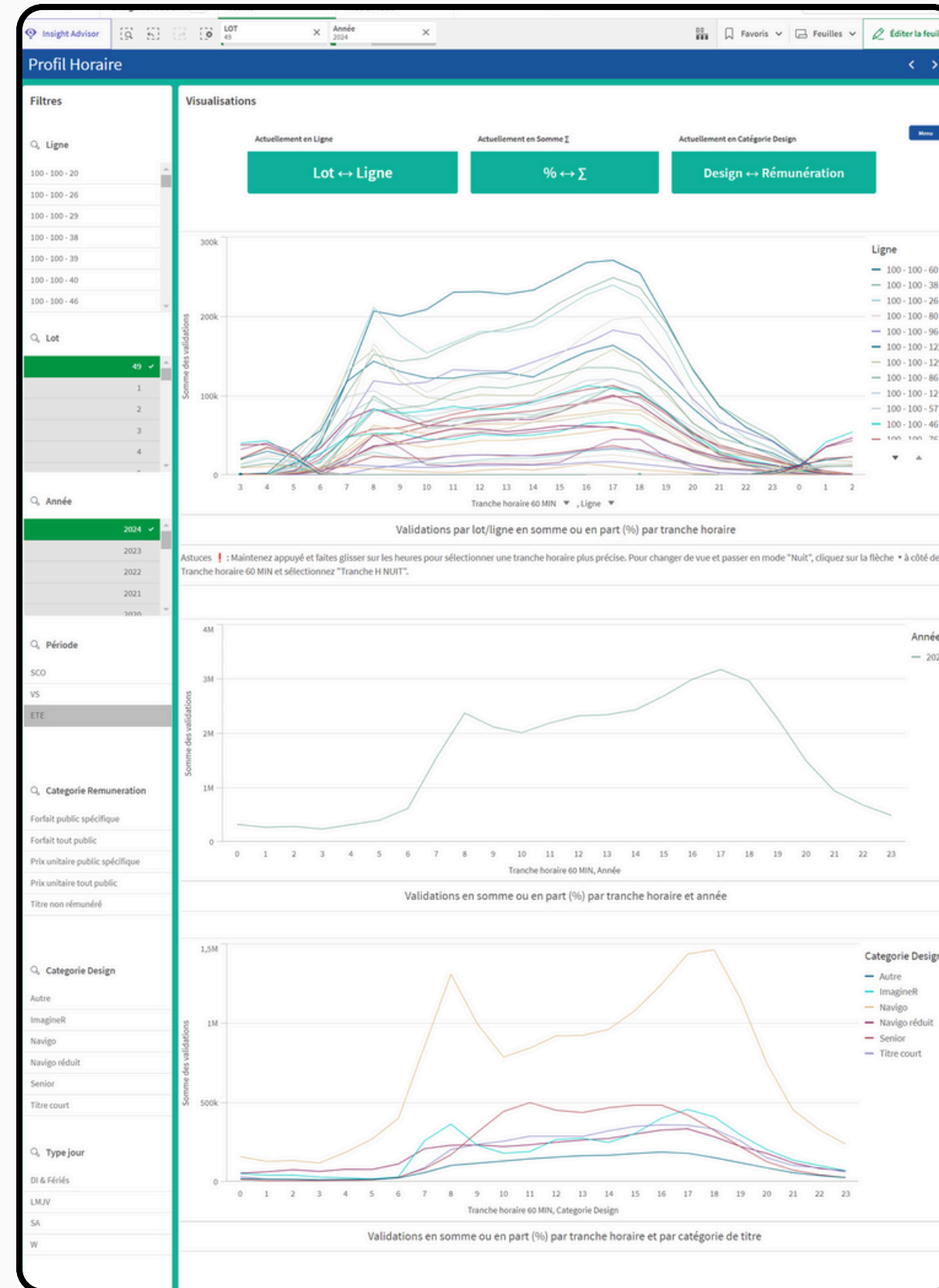
Fonctionnalités

- **Analyse temporelle** : profils horaires, évolution hebdomadaire/mensuelle.
- **Segmentation par type de titre** et catégorie d'usagers.
- Comparaison multi-périodes (avant/après Covid, scolaire/vacances).
- Export de données personnalisé.

Résultats & impact

- Application adoptée par les designers d'offre.
- **Retours positifs** suite aux formations et à l'utilisation.
- Base solide pour la migration vers QlikCloud.
- **Contribution à l'optimisation de l'offre bus** RATP CAP IDF.

Interface utilisateur



Ce que j'ai appris

Compétences Techniques approfondies

Découverte de **QlikSense** :

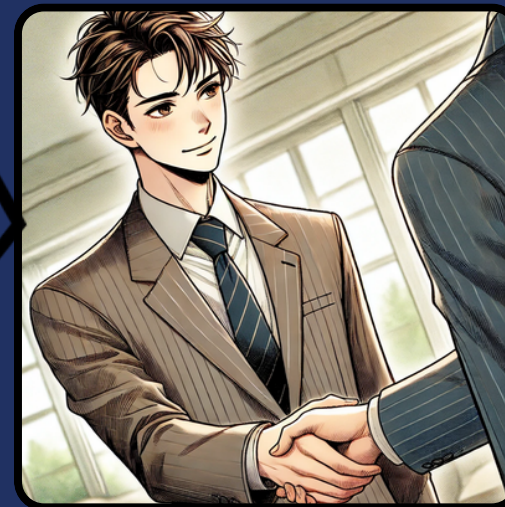
- Développement de dashboards complexes, avec une interface utilisateur avancée.
- Scripts de chargement et modélisation de données. Utilisation du langage **SQL** pour automatiser la lecture des données, créer les jointures et boutons.

Renforcement des compétences **Python** :

- Traitement et transformation de données massives
- Automatisation des tâches répétitives
- Création d'une app **Streamlit**

Meilleure connaissance du **secteur des transports publics** :

- Analyse des comportements de mobilité grâce aux données de validation
- Manipulation de données spatio-temporelles
- Compréhension de l'écosystème transport en IDF et des enjeux de l'ouverture à la concurrence



Soft Skills Développés

Gestion de projet autonome :

- Planification et priorisation
- Gestion des délais
- Adaptabilité face aux imprévus, aux demandes de dernières minutes

Communication transverse :

- Formation des utilisateurs
- Collaboration avec différents profils (Prévisionniste, Data Analyst...)
- Documentation technique et fonctionnelle



Ce que j'ai appris

Challenges surmontés

Gestion de la complexité :

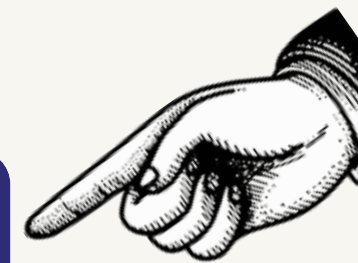
- Prise en main de nouvelles technologies
- Adaptation aux contraintes et résolution de problèmes techniques (mémoire, serveur)

Défis métier :

- Maîtrise du jargon technique
- Appréhender les subtilités des données du SIDV (codes lignes, catégories de titres Design/Rémunération, les périodes...)

★ Points d'Amélioration Identifiés

- 1 - Importance d'une conception robuste dès le début du projet.
- 2 - Anticiper le devenir du projet (Migration, mais aussi maintenance sur le long terme...).
- 3 - Valeur d'une documentation claire et complète.
- 4 - Communiquer



“La persévérance est la noblesse de l’obstination.”
- Adrien Decourcelle

Découvertes sur moi-même

J’ai tendance à complexifier certaines choses au lieu d’aller au plus simple.

J’aime travailler en autonomie, mais je peux m’égarer par perfectionnisme ou obstination...

Je suis capable de monter rapidement en compétences, de me remettre en question pour progresser.

Plutôt à l’aise dans la communication technique.

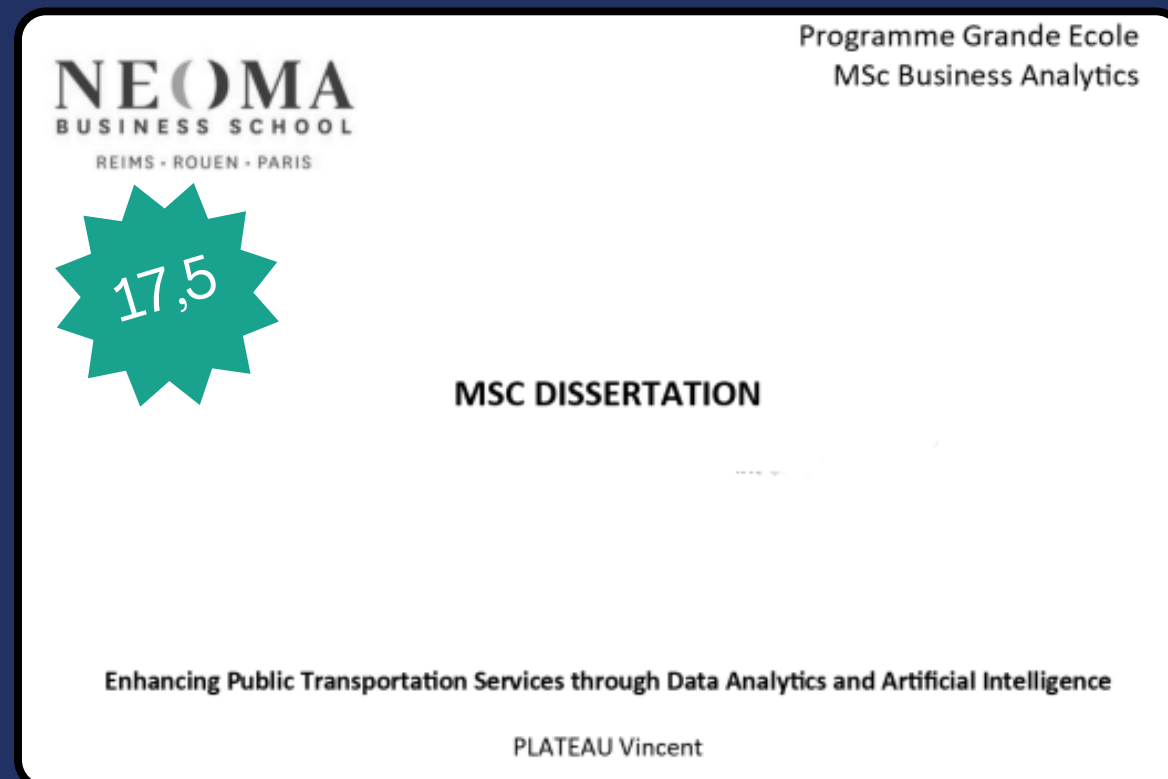
Goût pour les projets data à fort impact business !



Liens avec le MSc Business Analytics

La boucle est bouclée

Mon mémoire de fin d'études 'Enhancing Public Transportation Services through Data Analytics and Artificial Intelligence' s'est nourri de cette expérience terrain, créant une synergie parfaite entre théorie et pratique dans le domaine des transports publics.



Retour sur investissement confirmé !

Data Visualization & Dashboarding : Application directe pour la conception des tableaux de bord

- Principes de design d'interface utilisateur
- Best practices en visualisation de données

Data Preparation : Transformation des données avec Python et Qlik :

- Construction du modèle de données : comprendre les jointures
- Extraire, transformer, charger les données
- EDA (Exploratory Data Analysis).

Statistical Learning : Analyse des patterns de validation

- Concept de dimensionnalité
- Statistiques descriptives

Les pièces manquantes

Le MSc m'a appris à parler Tableau et R, mais sur le terrain, il a fallu devenir bilingue QlikSense et Python ! Heureusement : "les fondamentaux restent les mêmes : la vraie compétence est de savoir s'adapter aux outils." – ChatGPT

Conclusion : what's next ?



Confirmation de ma passion pour la data

- Gérer des projets data de bout en bout
- Transformer des données complexes en insights actionnables tout en travaillant sur des outils que j'aime
- Former et accompagner les utilisateurs métier
- Travailler en équipe, à plusieurs sur un même projet

Compétences clefs acquises, un tremplin pour la suite !

Objectif court terme :

- Poste de Data/Business Analyst (dans un secteur que j'aime)
- Continuer de développer mes compétences en Business Intelligence, data visualization...
- Focus sur les projets utiles et qui ont du sens

Vision moyen/long terme :

- Évolution vers des postes à responsabilité : être en charge d'un projet ?
- Contribution à la transformation digitale (du secteur mobilité ?)

Merci de votre attention !

Annexe

1. Code de l’application Streamlit “EMG”

2. Vue de l’application Streamlit “EMG”

3. Script de chargement dans QlikSense

```
import pandas as pd
import re
from folium.plugins import MarkerCluster

st.set_page_config(
    page_title="EMG App",
    page_icon="EMG/IPR.ico", # ancien logo
    layout="wide")

# Fonction pour charger les données avec mise en cache
@st.cache_data
def load_data():
    indiv_data = pd.read_excel(r"C:\Users\vincent.plateau\OneDrive - RATP Dev\Bureau\OneDrive_1_11-09-2024\EMG_BD")
    deplacement_data = pd.read_excel(r"C:\Users\vincent.plateau\OneDrive - RATP Dev\Bureau\OneDrive_1_11-09-2024\EMG_BD")

    # Prétraitement des données
    deplacement_data['Date_EMG'] = pd.to_datetime(deplacement_data['Date_EMG'], errors='coerce')
    deplacement_data = deplacement_data.dropna(subset=['Date_EMG'])

    return indiv_data, deplacement_data

# Chargement des données au démarrage de l'application
indiv_data, deplacement_data = load_data()

# Fonction pour filtrer les données
@st.cache_data
```

ancien script automatique

config qlik

ancien Load Conversion/Cale...

Load Conversion/Calendrier

Load Validation

```
17 For i = 0 to vNoOfRowsFiles - 1
18
19 LET vUniqueId = Peek('UniqueId', i, 'Office365Connector_ListFiles');
20 LET vFile = Peek('Name', i, 'Office365Connector_ListFiles');
21
22 trace $(vUniqueId);
23 //trace $(vFile);
24
25
26 [Données de validation traitées] :
27 Load
28 [Jour],
29 [Tranche horaire 60 MIN],
30 [Flag Design] as [Flag Design-design.Flag Design] ,
31 [Flag Rémunération] as [Flag Rémunération-rému.Flag Rémunération],
32 [Code administratif ligne],
33 [Code ligne] as [CODE_LIGNE_Triplet],
34 [Validations]
35
36 FROM [$(vQwcConnectionName)]
37 (URL IS [http://ratqliks2:5555/data?connectorID=Office365Connector&tab1]
38 parquet);
39
40 next i ;
41 // fin chargement
42
```

