

Frédéric La Rosa

# PORTFOLIO DATA ANALYST



## **Niveau requis :** Prépa scientifique ou bac+2 en mathématiques

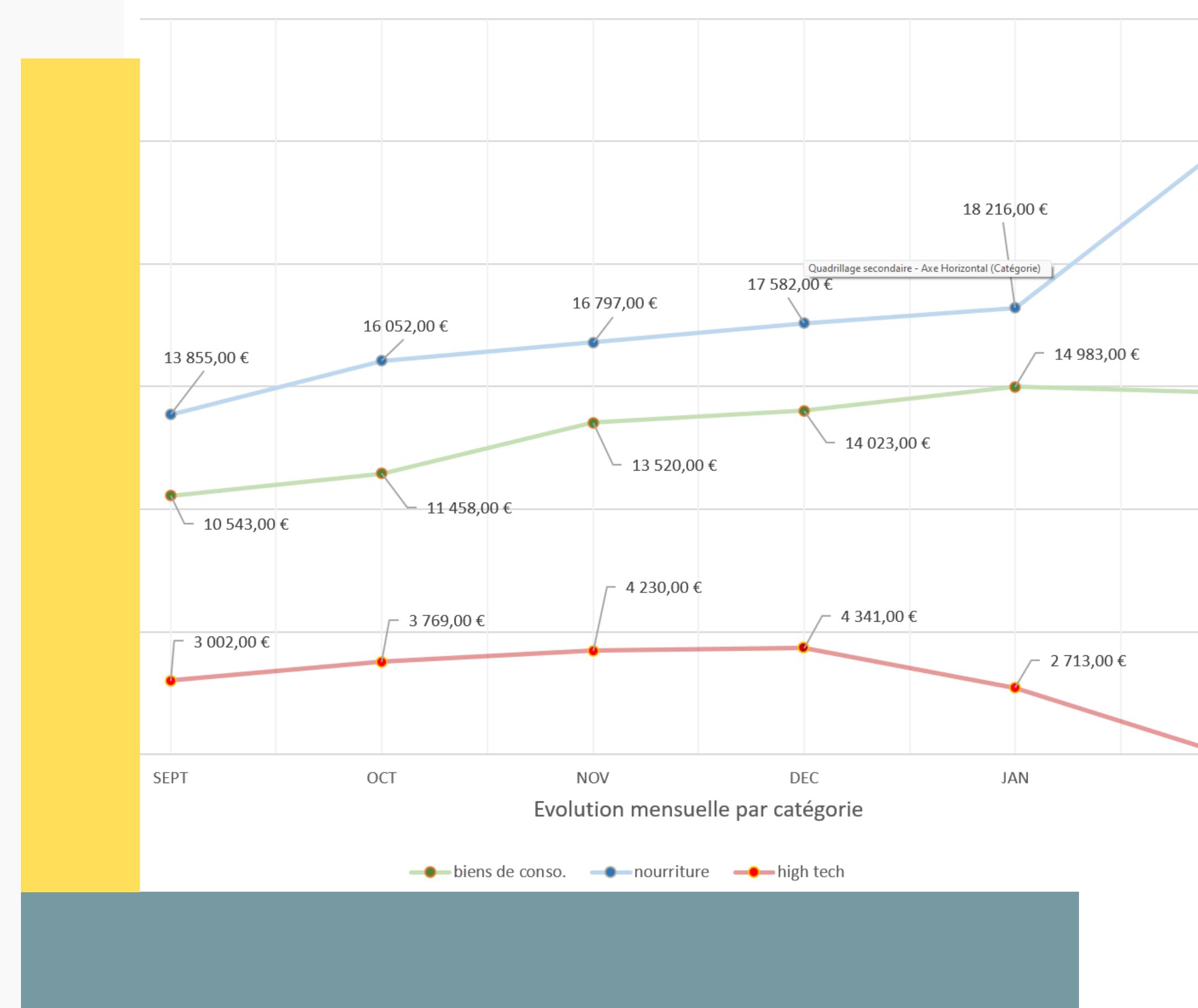
---

- Apprentissage durant deux mois intensif de l'algèbre linéaire et des statistiques
- Une note de 10/10 obtenue au test
- Formation refusée pour manque d'expérience
- Accepté en formation Data Analyst



“**FAITES UNE ANALYSE  
DES VENTES POUR UN  
E-COMMERCE**”

- Interpréter les informations provenant d'un dashboard
- Générer des graphiques adaptés aux types de données
- Synthétiser des résultats à destination d'un client



```

WITH
valeur_moyenne_par_ville AS (
  SELECT
    code_departement,
    nom_departement,
    nom_commune,
    AVG(vente.valeur_fonciere) AS valeur_fonciere
  FROM vente
  JOIN bien
    ON vente.bien_id = bien.id
  JOIN cadastre
    ON bien.cadastre_id = cadastre.id
  JOIN commune
    ON cadastre.commune_id = commune.id
  JOIN departement
    ON commune.departement_id = departement.id
  WHERE code_departement IN (6,13,33,59,69)
  GROUP BY code_departement, nom_commune)
SELECT
  CONCAT(nom_departement, "_", "(", code_departement, "))") AS departement,
  nom_commune AS commune,
  FORMAT(ROUND(valeur_fonciere), '###.###.###') AS prix_moyen
FROM (
  SELECT
    code_departement,
    nom_departement,
    nom_commune,
    valeur_fonciere,
    RANK() OVER (PARTITION BY code_departement ORDER BY valeur_fonciere
  FROM valeur_moyenne_par_ville) AS resultat_top_trois
  WHERE rang <= 3
);

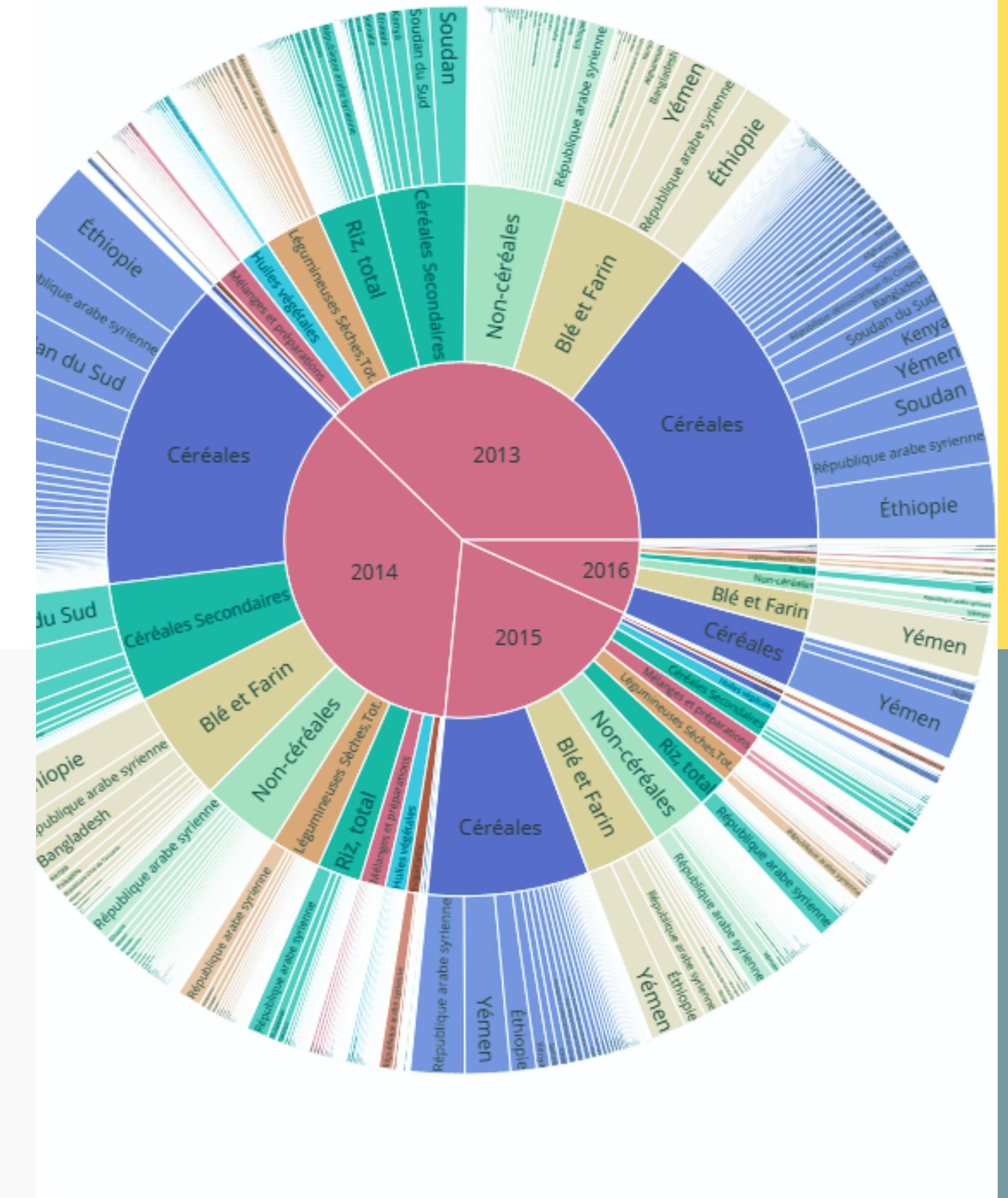
```

“ CRÉEZ ET UTILISEZ UNE BASE DE DONNÉES IMMOBILIÈRE AVEC SQL ”

- Créer le schéma relationnel d'une base de données
- Effectuer des requêtes SQL pour répondre à une problématique métier
- Mettre à jour un catalogue de données
- Créer des tables dans une base de données
- Charger des données dans une base de données
- Extraire des données avec des requêtes SQL

“ RÉALISEZ UNE ÉTUDE  
DE SANTÉ PUBLIQUE  
AVEC PYTHON ”

- Créer un environnement de développement
- Utiliser des librairies spécialisées pour les traitements data
- Manipuler des DataFrames

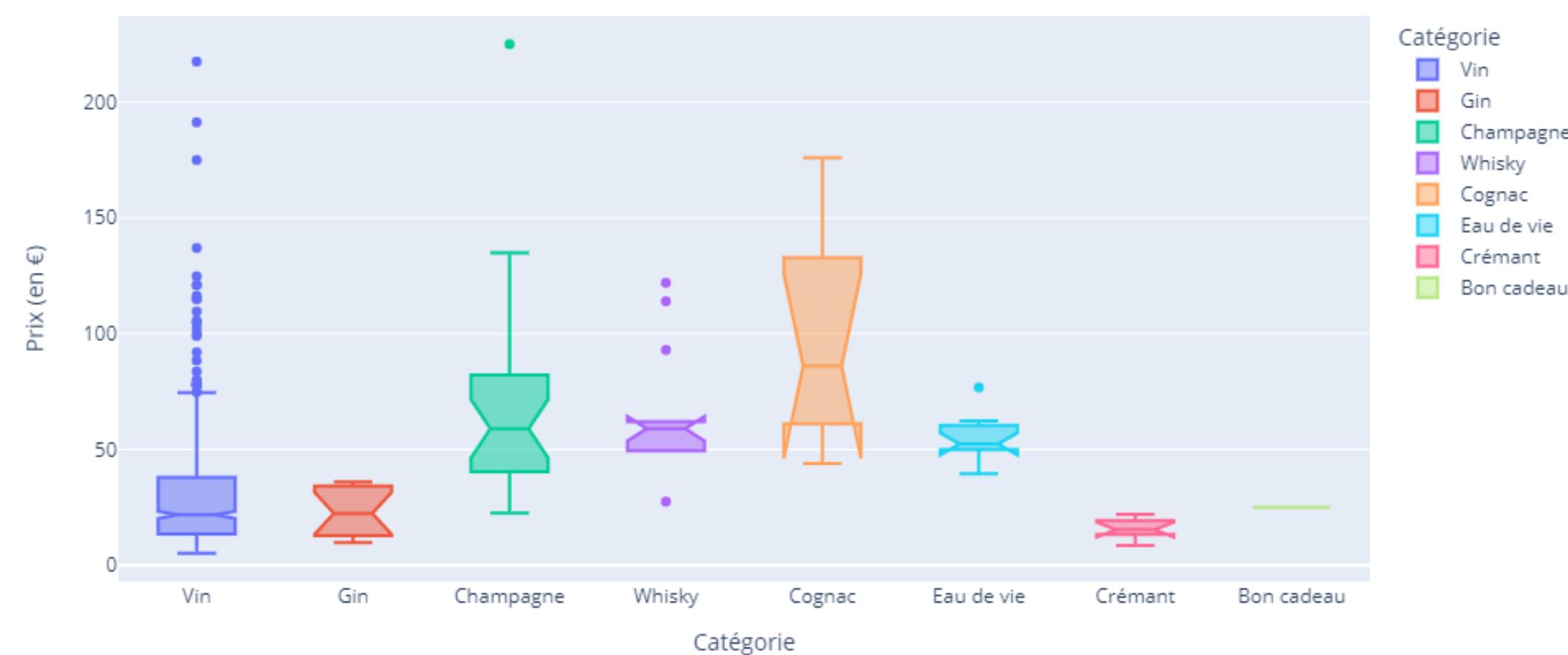


# “OPTIMISEZ LA GESTION DES DONNÉES D'UNE BOUTIQUE AVEC PYTHON

”

- Classifier différents types de données
- Gérer les erreurs et les incohérences présentes sur des données stockées
- Réaliser une analyse univariée pour interpréter des données

Quantiles des prix et outliers par catégorie de produits



“

# ANALYSEZ LES VENTES D'UNE LIBRAIRIE AVEC PYTHON

”

- Réaliser un test statistique
- Réaliser une analyse bivariée pour interpréter des données
- Analyser des séries temporelles

```

# Régression linéaire 17-44 ans
# Dataset
lien_age_frequence_17_44 = clients_non_outliers.loc[
    clients_non_outliers["Age - achat"] < 45].pivot_table(
    index="client_id",
    values=["session_id", "Age - achat"],
    aggfunc={
        "session_id": lambda session_id: len(session_id.unique()),
        "Age - achat": np.mean
    })

lien_age_frequence_17_44["session_id"] = round(
    lien_age_frequence_17_44["session_id"] /
    clients_non_outliers["Année et mois"].nunique(), 2)
lien_age_frequence_17_44 = lien_age_frequence_17_44.rename(
    columns={"session_id": "Fréquence d'achat (par mois)})

# Modèle de régression linéaire
X_17_44 = lien_age_frequence_17_44["Age - achat"].values.reshape(-1, 1)

# On crée un modèle de régression linéaire
model_17_44 = LinearRegression()
model_17_44.fit(X_17_44,
                 lien_age_frequence_17_44["Fréquence d'achat (par mois)"])

# On prédit y à partir de nos datas
x_range_17_44 = np.linspace(X_17_44.min(), X_17_44.max(), 100)
y_range_17_44 = model_17_44.predict(x_range_17_44.reshape(-1, 1))

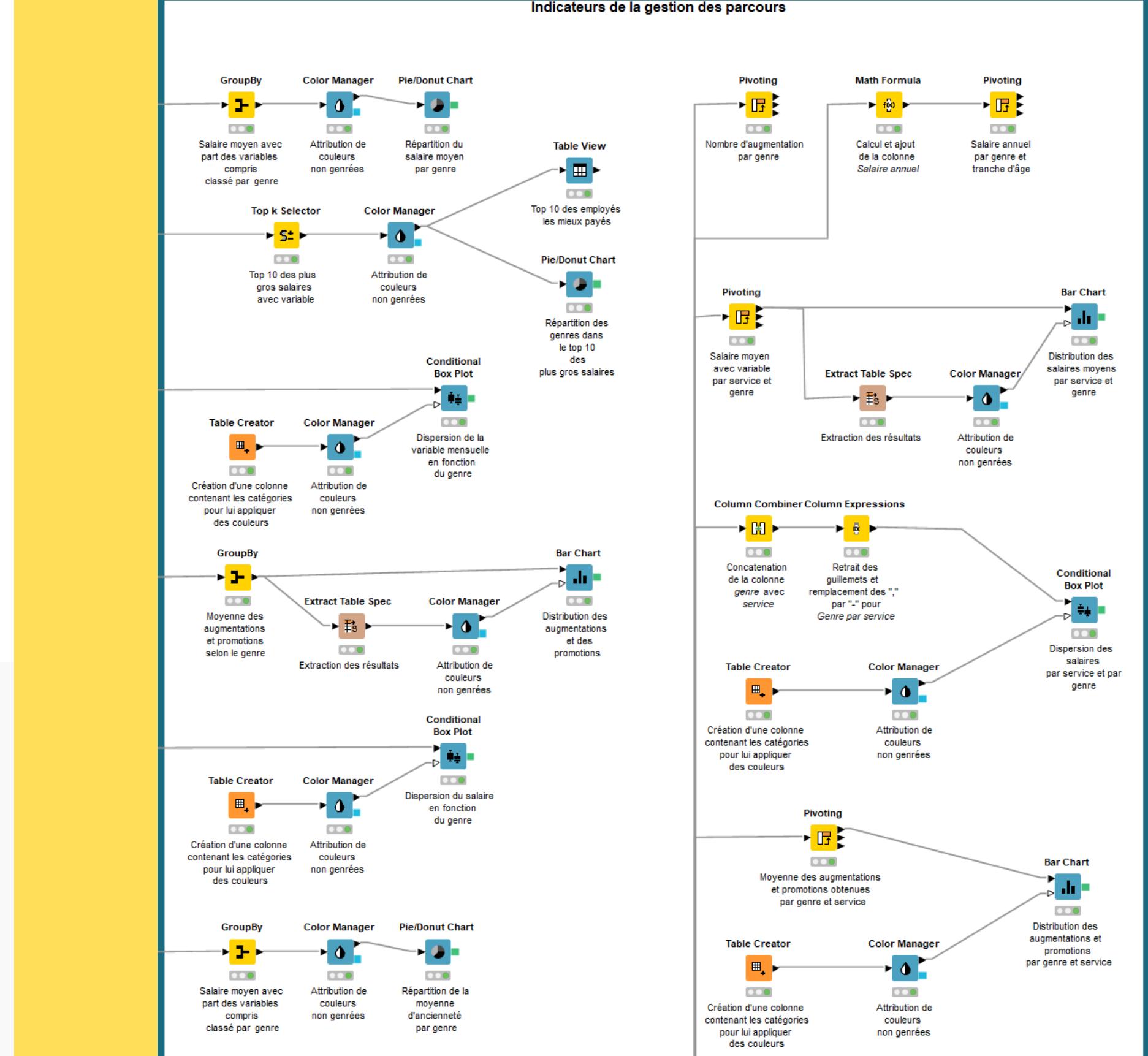
# Régression linéaire 45-94 ans
# Dataset
lien_age_frequence_45_94 = clients_non_outliers.loc[
    clients_non_outliers["Age - achat"] >= 45].pivot_table(
    index="client_id",
    values=["session_id", "Age - achat"],
    aggfunc={
        "session_id": lambda session_id: len(session_id.unique()),
        "Age - achat": np.mean
    })

lien_age_frequence_45_94["session_id"] = round(
    lien_age_frequence_45_94["session_id"] /
    clients_non_outliers["Année et mois"].nunique(), 2)

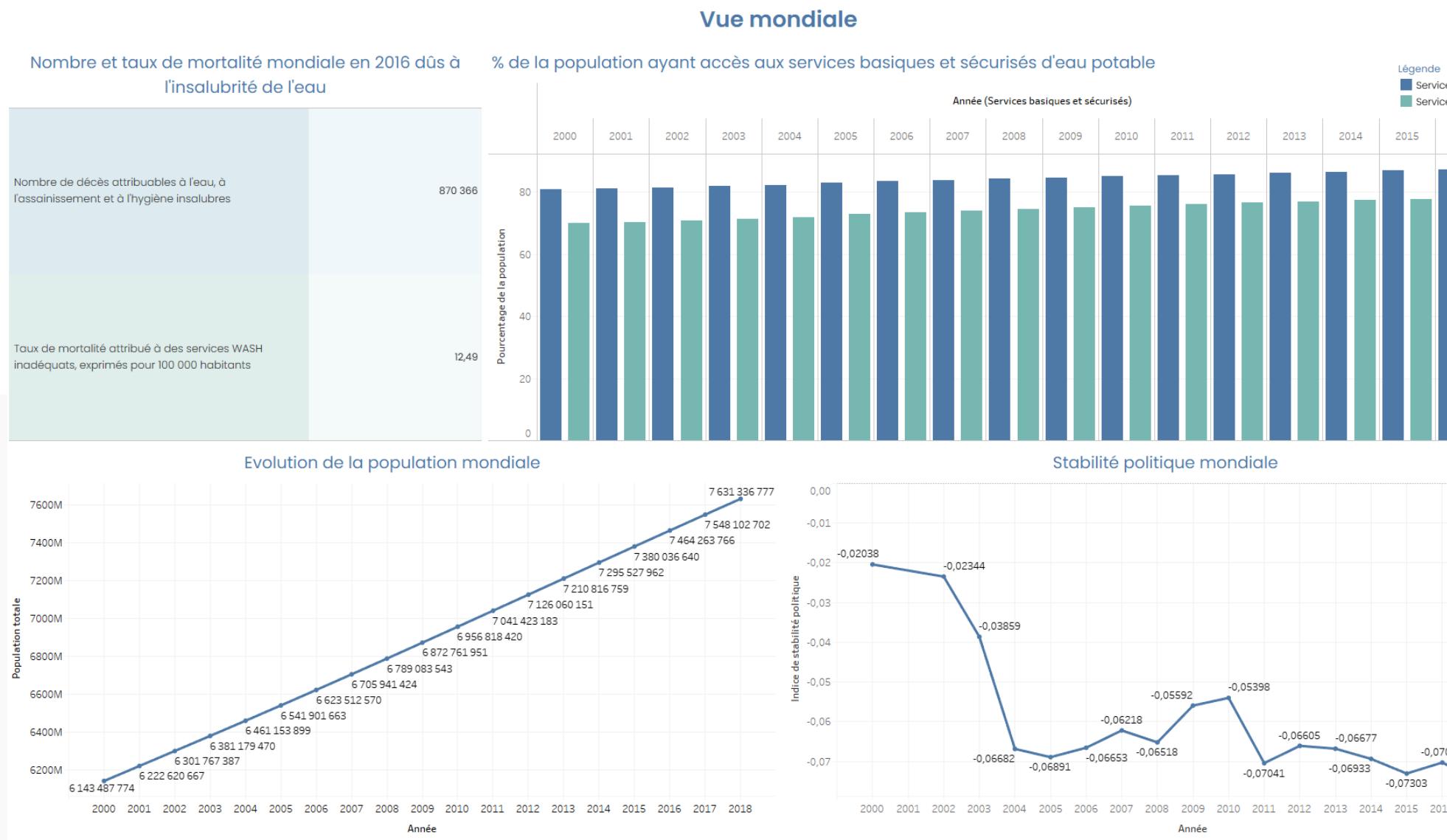
```

# “ANALYSEZ DES INDICATEURS DE L'ÉGALITÉ FEMME-HOMME AVEC KNIME”

- Préparer des données pour l'analyse en respectant les normes internes à l'entreprise
- Collecter des données en respectant le RGPD
- Transférer des données vers une zone de préparation



# “ FAITES UNE ÉTUDE SUR L'EAU POTABLE AVEC TABLEAU ”

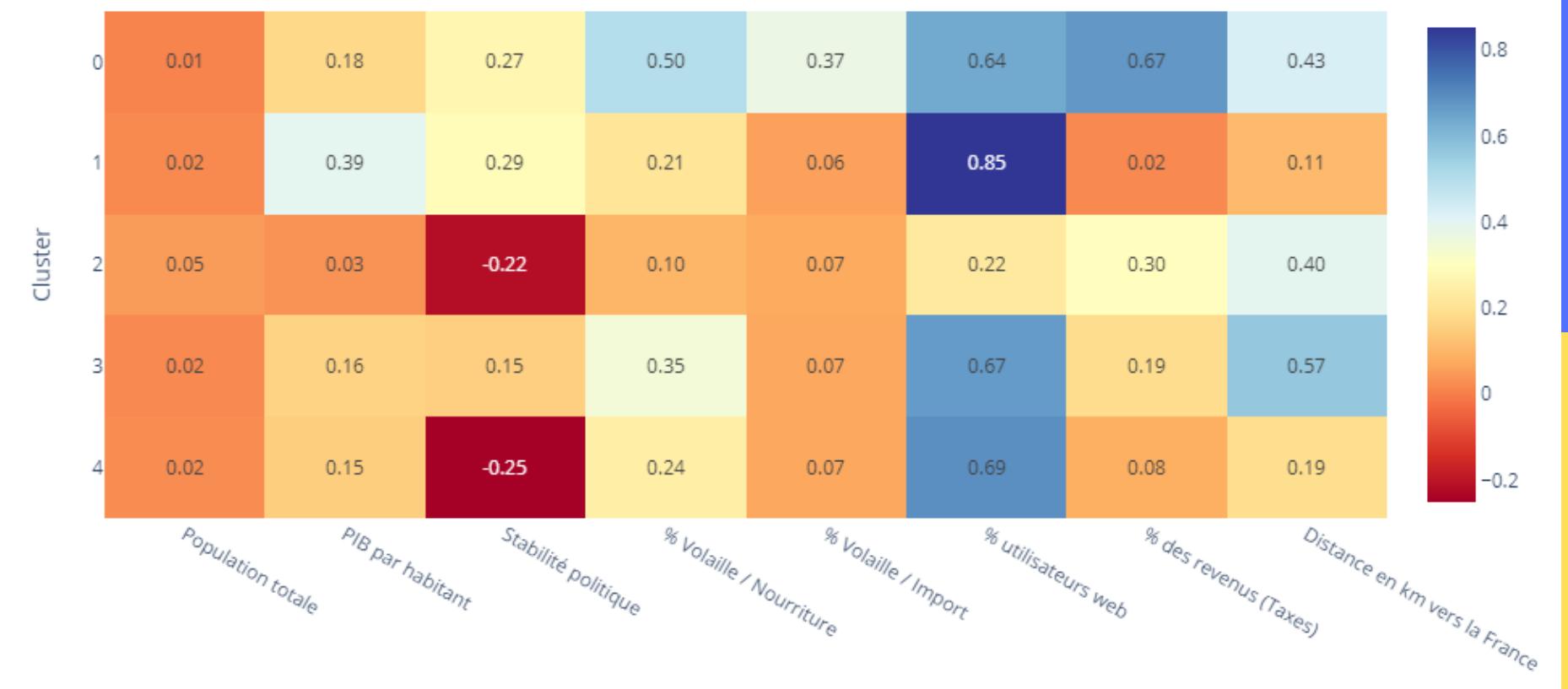


- Analyser un besoin client pour formuler des questions analytiques
- Créer un tableau de bord répondant à des questions analytiques
- Générer des graphiques adaptés aux types de données
- Synthétiser des résultats à destination d'un client

# “PRODUISEZ UNE ÉTUDE DE MARCHÉ”

”

Illustration des centroïdes en heatmap



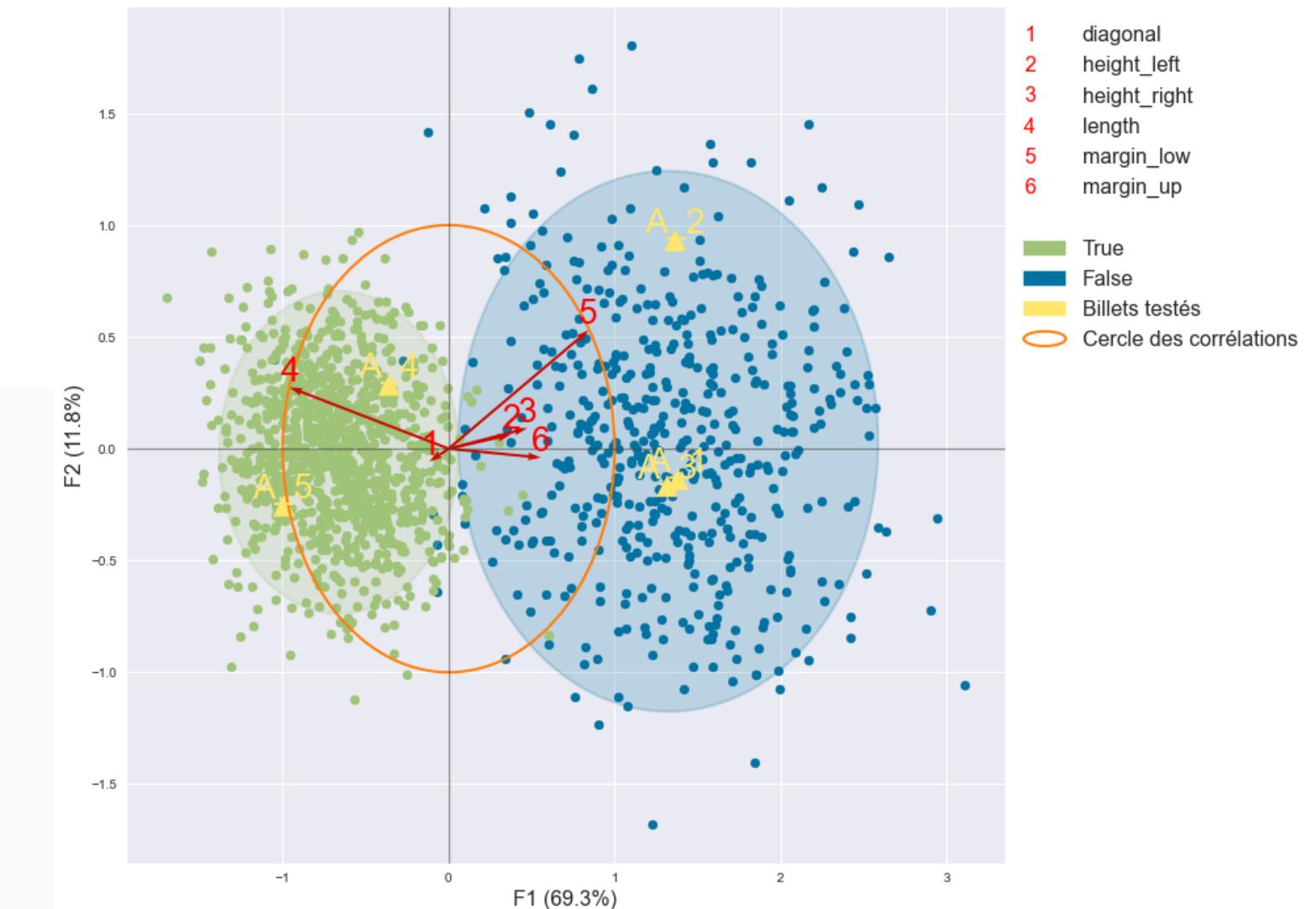
- Effectuer un clustering simple
- Explorer des données pour synthétiser des variables

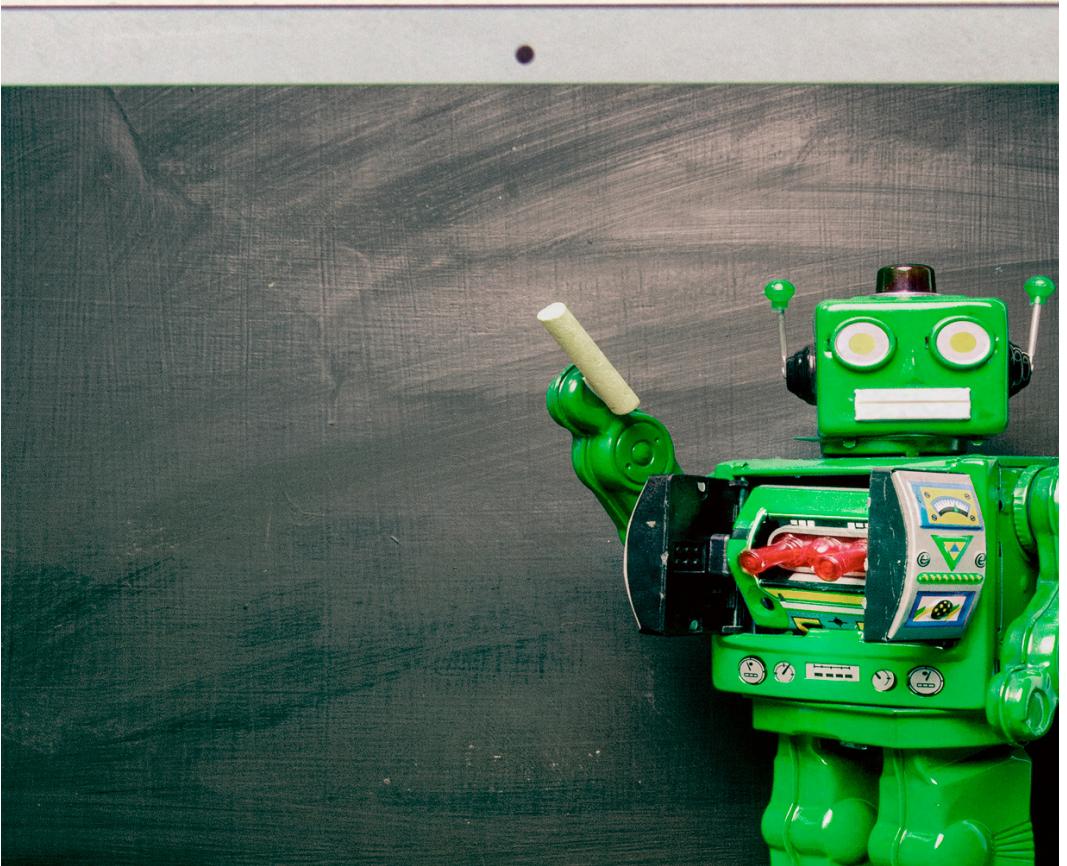
# “CRÉEZ UN ALGORITHME DE DÉTECTION AUTOMATIQUE DE FAUX BILLETS”

- Opérer des classifications automatiques pour partitionner les données
- Réaliser une régression linéaire
- Réaliser une régression logistique
- Réaliser une analyse prédictive

id	diagonal	height_left	height_right	margin_low	margin_up	length	Nature du billet	Proba Faux	Proba Vrai
	171.76	104.01	103.54	5.21	3.30	111.42	False	1.00	0.00
A_2	171.87	104.17	104.13	6.00	3.31	112.09	False	1.00	0.00
A_3	172.00	104.58	104.29	4.99	3.39	111.57	False	1.00	0.00
A_4	172.49	104.55	104.34	4.44	3.03	113.20	True	0.03	0.97
A_5	171.65	103.63	103.56	3.77	3.16	113.33	True	0.00	1.00

Billets à prédire sur le biplot des individus et des variables





# PROCHAIN OBJECTIF

**Réaliser en alternance le parcours ingénieur en Machine Learning**

Note: l'entièreté des projets sont disponibles sur mon **GitHub** (FredLaRosa) via le repository "[Portfolio](#)"

## Langages, librairies et logiciels utilisés:

- **SQL:** MySQL, Looping
- **Python:** Pandas; Numpy; Scikit-Learn; SciPy; StatsModels; Matplotlib; Plotly; Seaborn; Pingouin; ABC-Analysis; PCA; fanalysis; Prince; BeautifulSoup; Fuzzywuzzy; Datetime; Geopandas; RE; Math; NLTK; Streamlit; ...et plusieurs en cours d'apprentissage
- **KNIME** (Certification niveau L1)
- **Tableau**
- **Excel, Power Query**