

Table des matières

1. Récolte & exploration des données.....	2
2. Exercice 1 : Création d'un Storage Account.....	7
3. Exercice 2 : Créer une base de données SQL.....	10
4. Exercice 3 : Connexion à Azure Blob Storage et chargement des fichiers dans Azure SQL Database via ADF.....	16
5. Exercice 4 : Transformation des données avec Azure Data Factory.....	26
6. Problème de connexion avec la base de donnée.....	33
7. Test de cluster dans Azure machine learning Workspace :	37
8. Exercice 5 : Utilisation de PowerBI pour créer la visualisation de nos données.....	47
9. Questions de réflexion.....	49

1. Récolte & exploration des données

Récupération du jeu de données dans kaggle :

Jeux de donnée brut : **synthetic_health_lifestyle_dataset.csv**

Health And Lifestyle Dataset

Health And Lifestyle Prediction Dataset

Column Name	Description
ID	Unique identifier for each individual
Age	Age of the individual (in years)
Gender	Gender identity (Male , Female , Other)
Height_cm	Height in centimeters
Weight_kg	Weight in kilograms
BMI	Body Mass Index (calculated as weight in kg / height in m ²)
Smoker	Indicates whether the person is a smoker (Yes , No)
Exercise_Freq	Frequency of physical exercise (None , 1-2 times/week , 3-5 times/week , Daily)
Diet_Quality	Self-rated diet quality (Poor , Average , Good , Excellent)
Alcohol_Consumption	Level of alcohol intake (None , Low , Moderate , High)
Chronic_Disease	Whether the person has a chronic illness (Yes , No)
Stress_Level	Self-reported stress level on a scale from 1 (low) to 10 (high)
Sleep_Hours	Average hours of sleep per night

Créer un notebook jupyter pour commencer à faire de l'exploration de données, de nettoyer les données, ..

1. Visualiser les lignes
2. Identifier les données abberantes
3. Identifier les valeurs manquantes et des outliers
4. Standardiser des valeurs

Action réalisées dans le note book :

Importation et visualisation de la table des données et vérification des valeurs nulles :

```
import pandas as pd
import numpy as np
# Chemin du fichier source
fichier_source = r"C:\Users\Etudiant\Desktop\Cours\cloud\enregistrement_github\synthetic_health_lifestyle_dataset.csv"

# Chargement du CSV
df = pd.read_csv(fichier_source)
print(df.shape) # pour voir combien de Lignes/colonnes
print(df.head())
print(df.isnull().sum())

(7500, 13)
   ID  Age  Gender  Height_cm  Weight_kg    BMI Smoker  Exercise_Freq \
0    1   56    Other     177.6    37.3  11.8    Yes           NaN
1    2   69    Other     169.3    78.7  24.7    No  1-2 times/week
2    3   46  Female     159.1    69.0  27.3    No      Daily
3    4   32    Male     170.6    76.4  26.3    No  3-5 times/week
4    5   68    Male     158.4    60.4  24.1    No  3-5 times/week

   Diet_Quality Alcohol_Consumption Chronic_Disease  Stress_Level  Sleep_Hours
0        Poor            NaN          No             9            8.5
1       Good            High         No             2            5.9
2  Excellent        Moderate        No             3            4.8
3  Excellent        Moderate        No             9            6.6
4  Excellent           Low         Yes             6            6.1
ID
Age
Gender
Height_cm
Weight_kg
BMI
Smoker
Exercise_Freq
Diet_Quality
Alcohol_Consumption
Chronic_Disease
Stress_Level
Sleep_Hours
dtype: int64
```

Vérification des valeurs uniques :

```
[15]: df['Gender'].unique()
[15]: array(['Other', 'Female', 'Male'], dtype=object)

[17]: df['Smoker'].unique()
[17]: array(['Yes', 'No'], dtype=object)

[19]: df['Exercise_Freq'].unique()
[19]: array([nan, '1-2 times/week', 'Daily', '3-5 times/week'], dtype=object)

[21]: df['Diet_Quality'].unique()
[21]: array(['Poor', 'Good', 'Excellent', 'Average'], dtype=object)

[23]: df['Alcohol_Consumption'].unique()
[23]: array([nan, 'High', 'Moderate', 'Low'], dtype=object)

[25]: df['Chronic_Disease'].unique()
[25]: array(['No', 'Yes'], dtype=object)

[27]: df['Stress_Level'].unique()
[27]: array([ 9,  2,  3,  6,  1,  7, 10,  4,  5,  8], dtype=int64)

[29]: df['Sleep_Hours'].unique()
[29]: array([ 8.5,  5.9,  4.8,  6.6,  6.1,  5.8,  6.8,  6.3,  5.6,  9.7,  7.7,
  7.2,  6.9,  8. ,  7.8,  5.4,  4.7,  3.6,  7.9,  7.3,  8.4,  5. ,
  6.5,  4.6,  5.7,  5.2,  8.3,  3.9,  6. ,  7.5,  9.2,  4.4,  5.5,
  5.3,  9. ,  8.6,  4.9,  8.2,  6.2,  10.2,  6.7,  7.1,  5.1,  4.2,
  7. ,  8.9,  10. ,  8.7,  9.8,  7.6,  7.4,  8.1,  8.8,  10.3,  6.4,
  9.3,  9.4,  9.5,  9.6,  10.1,  3.7,  9.1,  10.6,  10.7,  3.8,  2.1,
  10.8,  4. ,  4.5,  11.3,  4.3,  4.1,  3.5,  3.3,  10.5,  9.9,  11. ,
  11.6,  10.4,  10.9,  3.1,  3. ,  11.2,  2.4,  2.9,  3.4,  11.5,  3.2,
  11.4,  12.1,  2.6,  2.5,  2. ,  2.7,  11.1,  12.6,  1.9,  2.2,  2.8,
  11.7,  11.8,  2.3])
```

Remplacement de certaine catégorie par des valeurs numériques :

```
[31]: df['Smoker']=df['Smoker'].replace({'Yes':1, 'No':0})
df['Diet_Quality']=df['Diet_Quality'].replace({'Poor':0, 'Good':2, 'Excellent':3, 'Average':1})
df['Chronic_Disease']=df['Chronic_Disease'].replace({'Yes':1, 'No':0})
```

Dans Alcohol_Consumption remplacement des Nan par ‘0’ puis remplacement des catégories par des entiers :

```
[33]: # Remplacement des NaN par '0' (string)
df['Alcohol_Consumption'] = df['Alcohol_Consumption'].fillna('0')

# Puis remplacement des catégories par des entiers
df['Alcohol_Consumption'] = df['Alcohol_Consumption'].replace({
    '0': 0,
    'Low': 1,
    'High': 3,
    'Moderate': 2
}))
```

Dans Exercise_Freq remplacement des Nan par ‘0’ puis remplacement des catégories par des entiers :

```
[47]: # D'abord remplacer les NaN par '0' (string)
df['Exercise_Freq'] = df['Exercise_Freq'].fillna('0')

# Puis remplacer les catégories par des entiers
df['Exercise_Freq'] = df['Exercise_Freq'].replace({
    '0': 0,
    '1-2 times/week': 1,
    'Daily': 3,
    '3-5 times/week': 2
}))
```

Classement des heures de sommeil par catégorie :

```
[52]: def classer_sommeil(h):
    if h < 4:
        return 'Très peu'
    elif 4 <= h < 6:
        return 'Peu'
    elif 6 <= h < 8:
        return 'Moyen'
    else:
        return 'Beaucoup'

df['Sleep_Category'] = df['Sleep_Hours'].apply(classer_sommeil)

print(df[['Sleep_Hours', 'Sleep_Category']].head())
Sleep_Hours Sleep_Category
0      8.5      Beaucoup
1      5.9        Peu
2      4.8        Peu
3      6.6      Moyen
4      6.1      Moyen
```

Vérification qu'il n'y a plus de valeur nulle :

```
[55]: df.isnull().sum()
[55]:   ID          0
      Age         0
      Gender       0
      Height_cm    0
      Weight_kg    0
      BMI          0
      Smoker        0
      Exercise_Freq 0
      Diet_Quality 0
      Alcohol_Consumption 0
      Chronic_Disease 0
      Stress_Level 0
      Sleep_Hours   0
      Sleep_Category 0
      dtype: int64
```

Classement des IMC par catégorie :

```
[60]: def classer_imc (imc):
    if imc < 16.5:
        return 'Dénutrition'
    elif 16.5 <= imc < 18.4:
        return 'Maigreur'
    elif 18.4 <= imc < 24.9:
        return 'Normal'
    elif 25 <= imc < 29.9:
        return 'Surpoids'
    elif 30 <= imc < 34.9:
        return 'Obésité classe 1'
    elif 35 <= imc < 39.9:
        return 'Obésité classe 2'
    else:
        return 'Obésité classe 3'

df['IMC'] = df['BMI'].apply(classer_imc)

[63]: print(df[['IMC', 'BMI']].head())
      IMC   BMI
0  Dénutrition  11.8
1      Normal  24.7
2     Surpoids  27.3
3     Surpoids  26.3
4      Normal  24.1
```

Test pour déterminer si il y a une corrélation entre Chronic_Disease et les valeurs numériques :

Si p-value < 0.05 → il y a un lien statistique entre la variable catégorielle et la présence de maladie chronique.

```
[74]: from scipy.stats import pointbiserialr

variables = ['Stress_Level', 'Diet_Quality', 'Alcohol_Consumption', 'Smoker', 'Exercise_Freq', 'Age']

for var in variables:
    corr, pval = pointbiserialr(df[var], df['Chronic_Disease'])
    print(f"{var} + Corrélation = {corr:.3f}, p-value = {pval:.4f}")

Stress_Level + Corrélation = 0.018, p-value = 0.3802
Diet_Quality + Corrélation = -0.002, p-value = 0.8487
Alcohol_Consumption + Corrélation = 0.001, p-value = 0.9233
Smoker + Corrélation = -0.003, p-value = 0.8036
Exercise_Freq + Corrélation = 0.001, p-value = 0.9226
Age + Corrélation = 0.007, p-value = 0.5695
```

Dans le cas présent il n'y a aucune relation linéaire entre 'Stress_Level', 'Diet_Quality', 'Alcohol_Consumption', 'Smoker', 'Exercise_Freq', 'Age' et la maladie chronique.

Test pour déterminer si il y a une corrélation entre Chronic_Disease et les valeurs catégorielles :

Si p-value < 0.05 → il y a un lien statistique entre la variable catégorielle et la présence de maladie chronique.

```
[77]: from scipy.stats import chi2_contingency

cat_vars = ['IMC', 'Sleep_Category']

for var in cat_vars:
    table = pd.crosstab(df[var], df['Chronic_Disease'])
    chi2, p, dof, expected = chi2_contingency(table)
    print(f"{var} + Test du chi² : p-value = {p:.4f}")

IMC + Test du chi² : p-value = 0.3513
Sleep_Category + Test du chi² : p-value = 0.5654
```

Dans le cas présent il n'y a aucune relation linéaire entre 'IMC', 'Sleep_Category' et la maladie chronique

Test avec un model non linéaire avec random forest :

```
[88]: import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import classification_report, confusion_matrix, roc_auc_score
from sklearn.preprocessing import StandardScaler

features = ['Age', 'Stress_Level', 'Diet_Quality', 'Alcohol_Consumption',
           'Smoker', 'Exercise_Freq', 'Sleep_Hours']

X = df[features]
y = df['Chronic_Disease']

# Standardisation des variables numériques
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# Séparation train/test
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.2, random_state=42)

# Création et entraînement du modèle Random Forest
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)

# Prédictions
y_pred = rf_model.predict(X_test)
y_proba = rf_model.predict_proba(X_test)[:, 1]

# Évaluation du modèle
print("Classification Report:\n", classification_report(y_test, y_pred))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("AUC-ROC Score:", roc_auc_score(y_test, y_proba))

# Importance des variables
importances = rf_model.feature_importances_
feature_importance_df = pd.DataFrame({'Feature': features, 'Importance': importances})
feature_importance_df = feature_importance_df.sort_values(by='Importance', ascending=False)

print("\nImportance des variables :\n", feature_importance_df)

Classification Report:
              precision    recall   f1-score   support
          0       0.80      0.98      0.88     1199
          1       0.14      0.01      0.02      301

          accuracy                           0.79      1500
         macro avg       0.47      0.50      0.45      1500
      weighted avg       0.67      0.79      0.71      1500

Confusion Matrix:
 [[1175  24]
 [ 297  4]]
AUC-ROC Score: 0.5006954854405249

Importance des variables :
          Feature  Importance
6      Sleep_Hours  0.310933
0        Age        0.278188
1  Stress_Level   0.144725
3  Alcohol_Consumption  0.078730
2  Diet_Quality   0.078558
5  Exercise_Freq   0.078215
4      Smoker       0.030651
```

Matrice de confusion :

	Prédit 0	Prédit 1	Total réel
Réel 0	1175	24	1199
Réel 1	297	4	301

1175 Vrais négatif bien prédits comme non malade
 24 faux positif non malade prédit comme malades
 297 faux négatif malade non détectés
 4 vrais positif correctement détecté comme malade

Le modèle n'est pas capable de discriminer les malades des non-malades.
Il y a un problème de déséquilibre des classes car il y a plus d'exemples de personne non malade (80 %) que de personne malade (20%).
On ne peut donc pas conclure sur des relations significatives.

Jeux de donnée transformé pour travailler dans Microsoft azure : **IMC.csv**

```
[#9]: df.to_csv("C:\\Users\\\\Etudiant\\\\Desktop\\\\Cours\\\\cloud\\\\enregistrement_github\\\\IMC.csv", index=False)
```

2. Exercice 1 : Création d'un Storage Account

a) Étapes :

1. Connexion à Azure :

o Allez sur portal.azure.com et connectez-vous.

2. Créez un Storage Account :

- Recherchez Storage Accounts/compte de stockage dans la barre de recherche et sélectionnez-le.
- Cliquez sur Créer.
- Sélectionnez votre abonnement et créez un groupe de ressources (ex : test).
- Donnez un nom unique à votre compte de stockage (ex: storageidmckillian) et choisissez la région.
- Dans la section Charge de travail principale, sélectionnez Storage Blob Azure ou Azure Data Lake Storage Gen 2
- Dans la section Charge de travail principale, sélectionnez Analytique Big Data
- Dans la section Redondance, sélectionnez Locally-redundant storage (LRS).

Créer un compte de stockage

Informations de base Avancé Mise en réseau Protection des données Chiffrement Balises Vérifier + créer

Stockage Azure est un service géré par Microsoft qui offre un stockage cloud hautement disponible, sécurisé, durable, scalable et redondant. Stockage Azure comprend le stockage Blob Azure (objets), Azure Data Lake Storage Gen2, Azure Files, Files d'attente Azure et Tables Azure. Le coût de votre compte de stockage dépend de l'utilisation et des options que vous choisissez ci-dessous. [En savoir plus sur les comptes de stockage Azure](#)

Détails du projet

Sélectionnez l'abonnement dans lequel créer le compte de stockage. Choisissez un groupe de ressources nouveau ou existant pour organiser et gérer votre compte de stockage avec d'autres ressources.

Abonnement *	Azure for Students
Groupe de ressources *	NetworkWatcherRG
	Créer nouveau

Détails de l'instance

Nom du compte de stockage * ⓘ	storagewebjcf
Région * ⓘ	(Europe) France Central
	Déployer sur une zone étendue Azure
Service principal ⓘ	Stockage Blob Azure ou Azure Data Lake Storage Gen 2
Performance * ⓘ	<input checked="" type="radio"/> Standard: Recommandé pour la plupart des scénarios (compte universel v2) <input type="radio"/> Premium: Recommandé pour les scénarios nécessitant une faible latence.
Redondance * ⓘ	Stockage localement redondant (LRS)

3. Finaliser :

- Cliquez sur Vérifier + Créer, puis sur Créer pour lancer le déploiement.

Services Azure

Créer une ressource Cost Management Azure Machine Learning Azure AI Foundry Abonnements Microsoft Ent... Accès conditionnel.. Bases de données SQL Sécurité Autres services

Ressources

Récent Favori

Nom	Type	Dernier affichage
mabasededonnée	Base de données SQL	il y a 8 heures
storagewebjcf	Compte de stockage	il y a 2 jours
tets	Azure Machine Learning workspace	il y a 2 jours
dashboards	Groupe de ressources	il y a 2 jours
testsante	Azure AI Foundry	il y a 2 jours
managed-testclustergenaiwebappinsights-ws	Espace de travail Log Analytics	il y a 2 jours
managed-testclusterbackendwebappinsights-ws	Espace de travail Log Analytics	il y a 2 jours
managed-testclusterbackendwebappinsights-ws	Espace de travail Log Analytics	il y a 2 jours
Azure for Students	Abonnement	il y a 2 jours
datafactoryjcf	Fabrique de données (V2)	il y a 3 jours
faber	Serveur SQL	il y a 3 jours
NetworkWatcherRG	Groupe de ressources	il y a une semaine

Tout afficher

b) Création d'un conteneur Blob :

1. Une fois le compte de stockage créé, allez dans celui-ci.
 2. Dérouler le menu Stockage des données et cliquez Conteneurs puis + Conteneur.
 3. Nommez le conteneur donné (ex : data) et laissez les paramètres d'accès à Privé.
 4. Cliquez sur le conteneur et appuyez sur Charger pour ajouter le fichier

The screenshot shows the Microsoft Storage Explorer interface. The left sidebar has a tree view with nodes like 'Vue d'ensemble', 'Journal d'activité', 'Etiquettes', 'Diagnostiquer et résoudre les problèmes', 'Contrôle d'accès (IAM)', 'Migration des données', 'Événements', 'Navigateur de stockage', 'Storage Mover', 'Solutions de partenaire', 'Visualiseur de ressources', 'Stockage des données', and 'Conteneurs'. The 'Conteneurs' node is selected and highlighted with a green box. The main pane displays a table of storage containers:

	Nom	Dernière modification	Niveau d'accès anonyme	État du bâil
<input type="checkbox"/>	slogs	15/07/2025 11:09:10	Privé	Disponible
<input type="checkbox"/>	data	15/07/2025 11:13:09	Privé	Disponible
<input type="checkbox"/>	datasante	22/07/2025 20:11:41	Privé	Disponible

At the top of the main pane, there are buttons for 'Ajouter un conteneur', 'Charger', 'Actualiser', 'Supprimer', 'Modifier le niveau d'accès', 'Restaurer des conteneurs', and 'Modifier les colonnes'. A search bar at the top says 'Rechercher les conteneurs par préfixe'.

3. Exercice 2 : Créer une base de données SQL

Exercice 2 : Créer une base de données SQL

1. Dans la barre de recherche du portail Azure, recherchez SQL Database/Base de données SQL et cliquez sur Créer.
2. Sélectionnez le groupe de ressources précédemment créé (ex : test).
3. Donnez un nom à votre base de données (ex: XYZDB).

Accueil > Bases de données SQL > Crée une base de données SQL >

Créer un serveur SQL Database

Microsoft

Détails du serveur

Entrez les paramètres obligatoires pour cette serveur, notamment fournir un nom et un emplacement. Ce serveur sera créé dans le même abonnement et le même groupe de ressources que votre base de données.

Nom du serveur * faber .database.windows.net

Emplacement * (Europe) France Central

Authentification

Info Azure Active Directory (Azure AD) est désormais Microsoft Entra ID. [En savoir plus](#)

Sélectionnez vos méthodes d'authentification préférées pour accéder à ce serveur. Créez un identifiant serveur et un mot de passe administrateur pour accéder à votre serveur avec authentification SQL, sélectionnez uniquement l'autheutification Microsoft Entra. [En savoir plus](#) en utilisant un utilisateur, un groupe ou une application Microsoft Entra existant en tant qu'administrateur Microsoft Entra [En savoir plus](#), ou sélectionnez à la fois l'authentification SQL et l'autheutification Microsoft Entra.

Méthode d'authentification Utiliser l'autheutification Microsoft Entra uniquement Utilisez à la fois l'autheutification SQL et Microsoft Entra Utiliser l'autheutification SQL

Définir l'administrateur Microsoft Entra e18893u@etu.univ-lorraine.fr
ID d'objet/d'application d'administration :2d8143e6-235c-4f95-bf31-5d427491dd3d [Définir l'administrateur](#)

Services Azure



Ressources

Récent Favori

Nom	Type	Dernier affichage
mabasededonnée (faber/mabasededonnée)	Base de données SQL	il y a quelques secondes

The screenshot shows the Azure SQL Database portal for the 'mabasededonnée' database. The left sidebar includes options like 'Vue d'ensemble', 'Journal d'activité', 'Étiquettes', 'Diagnostiquer et résoudre les problèmes', 'Éditeur de requêtes (version préliminaire)' (which is selected), 'Base de données miroir dans Fabric (préversion)', 'Visualiseur de ressources', and 'Paramètres'. The main area displays a query editor titled 'Requête 1' with a single row of data:

```

1

```

A tooltip in the editor area states: "Affichage de l'Explorateur d'objets limité ici. Pour bénéficier d'une fonctionnalité complète, veuillez cliquer ici pour ouvrir Azure Data Studio."

4. Créez un nouveau serveur SQL avec un nom de serveur (ex: serveurxyz), un login d'administrateur et un mot de passe. Choisissez la même région que votre compte de stockage.

- 1. Méthode d'authentification : Utilisation de l'authentification SQL et entra ID
- 2. Rechercher son nom d'utilisateur pour le définir en tant qu'admin
- 3. Définir le nom et le mot de passe de l'admin SQL

The screenshot shows the Azure Services dashboard. The top navigation bar includes 'Services Azure' (selected), 'Mes services', 'Mes ressources', and 'Centre d'aide'. Below this, the 'Services Azure' section lists various services: 'Créer une ressource', 'Cost Management', 'Azure Machine Learning', 'Azure AI Foundry', 'Abonnements', 'Microsoft Ent...', 'Accès conditionnel...', 'Bases de données SQL', 'Sécurité', and 'Autres services'. The 'Ressources' section shows a table of recent resources:

Nom	Type	Dernier affichage
faber	Serveur SQL	il y a quelques secondes

The screenshot shows the configuration page for the Azure SQL Server 'faber'. The left sidebar includes 'Vue d'ensemble' (selected), 'Journal d'activité', 'Contrôle d'accès (IAM)', 'Étiquettes', 'Démarrage rapide', 'Diagnostiquer et résoudre les problèmes', and 'Visualiseur de ressources'. The main area displays server details and settings:

Informations sur le serveur

- Groupe de res... ([déplacer](#)) : NetworkWatcherRG
- Statut : Disponible
- Emplacement : France Central
- Abonnement ([déplacer](#)) : Azure for Students
- ID d'abonnement : ba4dc137-e617-425b-ae0f-8da48309d70b
- Étiquettes ([modifier](#)) : Ajouter des étiquettes

Administrateur du serveur : Cloud5A3f69a833
Mise en réseau : [Paramètres réseau](#)
Administrateur Microsoft... : [e18893u@tu.univ-lorraine.fr](#)
Nom du serveur : faber.database.windows.net

Notifications (0) **Fonctionnalités (6)**

5. Environnement de travail -> Développement

6. Calcul + Stockage -> Niveau de Service (Modèle d'achat Basé sur DTU) -> De Base

Accueil > Bases de données SQL >

Configurer ...



Niveau de service et de calcul

Sélectionnez les niveaux disponibles en fonction des besoins de votre charge de travail. Le modèle vCore fournit une large gamme de contrôles de configuration et offre les options Hyperscale et Serverless afin de mettre à l'échelle automatiquement votre base de données en fonction de vos besoins en matière de charge de travail. Sinon, le modèle DTU fournit des ensembles prix/performances que vous pouvez choisir pour faciliter la configuration. [En savoir plus](#)

SQL Database Hyperscale : un prix bas, une scalabilité élevée et le meilleur ensemble de fonctionnalités. [En savoir plus](#)

Niveau de service

[Comparer les niveaux de service](#)

UDBD [Comparer les options DTU](#)

5 (Basic)

Taille maximale des données (Go)

Récapitulatif des coûts
De base (Basic)
Cout par DTU (dans USD) 1.22
UDBD sélectionnés x 5
COÛT ESTIMÉ PAR MOIS 6.11 USD

7. Redondance du stockage de sauvegarde -> Stockage de sauvegarde redondant localement

Accueil > Bases de données SQL >

Créer une base de données SQL ...

Microsoft

Entrez les paramètres obligatoires pour cette base de données, choisissez notamment un serveur logique et configuez les ressources de calcul et de stockage

Nom de la base de données *

mabasededonnée

Serveur *

(nouveau) faber (France Central)

[Créer nouveau](#)

Vous souhaitez utiliser un pool élastique SQL ?

Oui Non

Environnement de charge de travail

Développement

Production

Paramètres par défaut fournis pour les charges de travail Development. Les configurations peuvent être modifiées selon les besoins.

Calcul + stockage *

De base

Stockage de 2 Go

[Configurer la base de données](#)

Redondance du stockage de sauvegarde

Choisissez comment répliquer vos sauvegardes PITR et LTR. La géorestauration ou la possibilité de récupération après une panne régionale est disponible seulement si le stockage géoredondant est sélectionné.

Redondance du stockage de sauvegarde

Stockage de sauvegarde redondant localement

Stockage de sauvegarde redondant interzone

Stockage de sauvegarde géoredondant

Stockage de sauvegarde géoredondant interzone

8. Mise en réseau -> Méthode de connectivité -> Point de terminaison public

9. Mise en réseau -> Règles de pare-feu ->

- 1. Autoriser les services et les ressources Azure à accéder à ce serveur -> Oui
- 2. Ajouter l'adresse IP actuelle du client -> Oui

Accueil > Bases de données SQL >

Créer une base de données SQL

Microsoft

De base **Mise en réseau** Sécurité Paramètres supplémentaires Balises Vérifier + créer

Configurez l'accès réseau et la connectivité de votre serveur. La configuration sélectionnée ci-dessous s'applique au serveur sélectionné « faber » et à toutes les bases de données qu'il gère. [En savoir plus](#)

Connectivité réseau

Choisissez une option pour configurer la connectivité de votre serveur via un point de terminaison public ou privé. Si vous choisissez Aucun accès, les valeurs par défaut sont utilisées et vous pouvez configurer la méthode de connexion après la création du serveur. [En savoir plus](#)

- Méthode de connectivité * ⓘ
- Aucun accès
 - Point de terminaison public
 - Point de terminaison privé

Règles de pare-feu

La définition du paramètre « Autoriser les services et les ressources Azure à accéder à ce serveur » sur Oui autorise les communications de toutes les ressources situées dans la limite Azure, qu'elles fassent ou non partie de votre abonnement. [En savoir plus](#)

La définition du paramètre « Ajouter l'adresse IP actuelle du client » sur Oui ajoute au pare-feu du serveur une entrée pour l'adresse IP du client.

- | | | |
|---|---------------------------|--------------------------------------|
| Autoriser les services et les ressources Azure à accéder à ce serveur * | <input type="radio"/> Non | <input checked="" type="radio"/> Oui |
| Ajouter l'adresse IP actuelle du client * | <input type="radio"/> Non | <input checked="" type="radio"/> Oui |

10. Cliquez sur Vérifier + Créer, puis Créer pour déployer la base de données.

11. Après la création de la ressource, se rendre dans le volet Sécurité -> Mise en réseau de la ressource serveur SQL et que la case à cocher est bien cochée permettant d'autoriser pour autoriser les services Azure.

mabasededonnée (faber/mabasededonnée)

Base de données SQL

Rechercher Copier Restaurer Exporter Définir un pare-feu de serveur

Vue d'ensemble

- Journal d'activité
- Étiquettes
- Diagnostiquer et résoudre les problèmes
- Éditeur de requêtes (version préliminaire)
- Base de données miroir dans Fabric (préversion)
- Visualiseur de ressources
- Paramètres
- Gestion des données
- Intégrations
- Power Platform
- Sécurité
- Performances intelligentes
- Supervision
- Automatisation
- Aide

Bases de données miroir dans Microsoft Fabric Répliquez facilement vos bases

Bases

Groupe de ressources (déplacer)	:	NetworkWatcherRG
État	:	Online
Emplacement	:	France Central
Abonnement (déplacer)	:	Azure for Students
ID d'abonnement	:	ba4dc137-e617-425b-ae0f-8da48309d70b
Étiquettes (modifier)	:	Ajouter des étiquettes

Prise en main Monitoring Propriétés Fonctionnalités Notifications

Configurer l'accès
Configurer l'accès réseau à votre SQL Server. [Découvrir plus d'informations](#)

Configurer

faber | Mise en réseau

Accès public Accès privé Connectivité

Accès au réseau public

Les points de terminaison publics autorisent l'accès à cette ressource via l'internet en utilisant une adresse IP publique. Une application ou une ressource à laquelle on accorde l'accès avec les règles réseau suivantes nécessite toujours une autorisation d'accès appropriée. [En savoir plus](#)

Accès au réseau public

Désactivé Réseaux sélectionnés

Les connexions provenant des adresses IP configurées dans la section Règles de pare-feu ci-dessous ont accès à cette base de données. Par défaut, aucune adresse IP publique n'est autorisée. [En savoir plus](#)

Réseaux virtuels

Authorisez les réseaux virtuels à se connecter à votre ressource en utilisant des points de terminaison de service. [En savoir plus](#)

Ajouter une règle de réseau virtuel

Règles de pare-feu

Authorisez certaines adresses IP Internet publiques à accéder à votre ressource. [En savoir plus](#)

Ajouter l'adresse IPv4 de votre client (91.168.170.226) Ajouter une règle de pare-feu

Nom de la règle	Adresse IPv4 de début	Adresse IPv4 de fin
ClientIp-2025-7-13_12-19-11	91.168.170.226	91.168.170.226

Exceptions

Autoriser les services et les ressources Azure à accéder à ce serveur

12. Se rendre dans Editeur de requêtes pour se connecter et visualiser les tables

Pour exemple Table chargée dbo.IMC.csv :

The screenshot shows the SSMS interface with the following details:

- Left pane (Object Explorer):** Shows the database structure with 'Tables' expanded, displaying 'dbo.avceraladiechronique', 'dbo.IMC.csv', 'dbo.sansmaladiechronique', and 'dbo.sansmodification'.
- Center pane (Query Editor):** The query 'SELECT * FROM [dbo].[IMC.csv]' is entered in the 'Requête 1' tab.
- Right pane (Results):** The results table contains the following data:

ID	Age	Gender	Height_cm	Weight_kg	BMI	Smoker
1	56	Other	177	37	11	1
2	69	Other	169	70	24	0
3	46	Female	159	69	27	0
4	32	Male	170	76	26	0

Partie 2 : Utilisation d'Azure Data Factory pour orchestrer et traiter les données

Dans cette partie, nous allons prendre en main Azure Data Factory (ADF) pour automatiser le processus de déplacement, transformation et chargement des données depuis Azure Blob Storage vers Azure SQL Database.

4. Exercice 3 : Connexion à Azure Blob Storage et chargement des fichiers dans Azure SQL Database via ADF

Objectif : Utiliser Azure Data Factory pour charger les fichiers CSV depuis un Blob Storage et les insérer dans une Azure SQL Database.

Étapes :

1. Création d'un Azure Data Factory :

- Dans le portail Azure, créez un service Azure Data Factory/Fabrique de données.
- Sélectionner « France Central » pour l'option région
- Cliquez sur Vérifier + Créer, puis Créer pour déployer le service.

Créer Data Factory

Informations de base Configuration Git Réseau Avancé Étiquettes Vérifier + créer

Un clic pour créer une fabrique de données avec un exemple de pipeline et de jeux de données. [Essayez-le](#)

Détails du projet

Selectionnez l'abonnement pour gérer les coûts et les ressources déployées. Utilisez les groupes de ressources comme les dossiers pour organiser et gérer toutes vos ressources.

Abonnement * ⓘ Azure for Students

Groupe de ressources * ⓘ NetworkWatcherRG [Créer nouveau](#)

Détails de l'instance

Nom * ⓘ datafactoryjcf

Région * ⓘ France Central

Microsoft.DataFactory-20250715152413 | Vue d'ensemble

Déploiement

Rechercher Supprimer Annuler Redéployer Télécharger Actualiser

Votre déploiement a été effectué

Nom du déploiement : Microsoft.DataFactory-20250715152413
Abonnement : Azure for Students
Groupe de ressources : NetworkWatcherRG

Heure de début : 15/07/2025 15:27:38
ID de corrélation : 60aa51c5-08ac-4874-aea1-921c106184e2

Détails du déploiement

Ressource	Type	Statut	Détails de l'opération
datafactoryjcf	Fabrique de données (V2)	OK	Détails de l'opération

Étapes suivantes

Accéder à la resource

Envoyer des commentaires
Partagez votre expérience avec le déploiement

- Cliquez sur Lancer le studio
- Accédez à l'interface de création de pipelines (via le bouton "Author & Monitor").

Services Azure



Ressources

Récent	Favori	
datafactoryjcf	Fabrique de données (V2)	il y a quelques secondes

2. Création d'un pipeline ADF :

- Dans Azure Data Factory, créez un pipeline pour extraire les données depuis le Blob Storage et les insérer dans la Azure SQL Database.

The screenshot shows the Azure Data Factory pipeline editor interface. On the left, the 'Ressources de fabrique' sidebar lists various resources like Pipelines, Capture des changements de données, J eux de données, Flux de données, and Power Query. The main workspace displays a pipeline named 'pipeline1' with a single activity: 'Copier les données' (Copy data1). A modal dialog titled 'Copier les données' is open, showing the 'Source' tab. In this tab, 'Stockage Blob Azure' is selected as the data source. To the right of the modal, a separate window titled 'Sélectionner un magasin de données' (Select a data store) shows a list of storage options, with 'Stockage Blob Azure' highlighted.

- Étapes principales du pipeline :**
- Source : Ajoutez une activité de type Copy Data/Copier les données.

- Sélectionnez le fichier ventes.csv depuis le conteneur Storage Blob Storage comme Source.

The screenshot shows the Azure Data Factory pipeline configuration. On the left, the 'Ressources de fabrique' sidebar lists 'Pipelines' (pipeline1), 'Capture des changements de do...', 'Jeux de données' (sante selected), 'Flux de données', and 'Power Query'. In the main pane, 'pipeline1' is selected, and under it, the 'sante' dataset is chosen. The 'Connexion' tab is active, showing the connection details: 'Service lié' is set to 'AzureBlobStorage1', 'Chemin d'accès au fichier' is 'data/IMC.csv', and 'Type de compression' is 'Aucune compression'. Other settings like 'Séparateur de colonne', 'Délimiteur de ligne', and 'Encodeage' are also visible. To the right, the 'Propriétés' panel shows the dataset name 'sante' and a 'Description' field.

Test de la connexion :

This screenshot shows the same pipeline configuration as above, but the 'Connexion' tab now displays a green checkmark indicating a successful connection test. The status message says 'Connexion établie' (Connection established). The rest of the interface remains the same, showing the dataset configuration and the Azure Blob Storage connection settings.

Aperçu des données :

This screenshot shows the 'Aperçu des données' (Preview) section of the pipeline configuration. It displays a preview of the 'IMC.csv' dataset with 10 rows of data. The columns are labeled: ID, Age, Gender, Height_cm, Weight_kg, BMI, Smoker, Exercise_Freq, Diet_Quality, and Alcohol_Consumption. The data includes various demographic and health-related information for different individuals.

ID	Age	Gender	Height_cm	Weight_kg	BMI	Smoker	Exercise_Freq	Diet_Quality	Alcohol_Consumption
1	56	Other	177.6	37.3	11.8	1	0	0	0
2	69	Other	169.3	70.7	24.7	0	1	2	3
3	46	Female	159.1	69.0	27.3	0	3	3	2
4	32	Male	170.6	76.4	26.3	0	2	3	2
5	60	Male	158.4	60.4	24.1	0	2	3	1
6	25	Female	167.7	62.0	22.0	0	3	2	2
7	38	Other	152.5	88.0	37.8	0	1	2	3
8	56	Male	165.3	65.0	23.8	0	3	2	2
9	36	Male	176.9	41.3	13.2	0	3	1	0
10	40	Other	162.0	77.4	20.5	1	2	1	1

- Destination : Configurez une Azure SQL Database comme destination (sous la forme d'un Linked Service) en utilisant l'authentification SQL (login/mot de passe).

Autorisation authentication SQL par désactivation authentication microsoft entra uniquement

Ajouter une attribution de rôle

Pour résoudre le problème je me suis mis en System-assigned managed identity pour le type d'authentification

3. Création des Linked Services :

- Créez un Linked Service pour Azure Blob Storage afin de se connecter au conteneur où les fichiers CSV sont stockés.

The screenshot shows the 'Services liés' (Linked Services) blade in the Azure Data Factory portal. A green box highlights the 'Nom' (Name) field where 'AzureblobStorage' is entered. The 'Type' dropdown is set to 'Stockage Blob Azure'. Other configuration options like 'Se connecter via un runtime d'intégration' (Connect via integration runtime), 'Clé d'authentification' (Authentication key), and 'Abonnement Azure' (Azure subscription) are visible. At the bottom right, there's a 'Connexion établie' (Connected) status indicator.

The screenshot shows the 'Services liés' (Linked Services) blade in the Azure Data Factory portal. A green box highlights the 'Nom' (Name) field where 'blobconnection' is entered. The 'Type' dropdown is set to 'Stockage Table Azure'. Other configuration options like 'Se connecter via un runtime d'intégration' (Connect via integration runtime), 'Clé de compte' (Account key), and 'Nom du compte de stockage' (Storage account name) are visible. At the bottom right, there's a 'Connexion établie' (Connected) status indicator.

- Créez un Linked Service pour

Azure SQL Database afin de se connecter à votre base de données en fournissant votre nom d'utilisateur et mot de passe défini pendant la création de la base de données.

4. Mapping des données :

- Mappez les colonnes du fichier CSV avec celles de la table SQL correspondante (par exemple, le fichier ventes.csv avec la table Ventes dans SQL).

The screenshot shows the Azure Data Factory pipeline editor. On the left, the 'Ressources de fabrique' sidebar lists various resources like Pipelines, Datasets, and Activities. In the center, a pipeline named 'pipeline1' is selected. A 'Copier les données' (Copy data) activity is highlighted. The 'Récepteur' (Sink) tab is active, showing the target dataset 'AzureSqlTable1' and the 'Insertion' (Insert) write behavior. A green box highlights the 'Récepteur' tab and the target dataset configuration.

This screenshot shows the same pipeline editor interface, but the 'Mappage' (Mapping) tab is now active for the Copy activity. It displays a table mapping from source columns to destination columns. A green box highlights this mapping table.

Source	Type	Destination	Type
ID	abc String	→ ID	abc nvarchar
Age	abc String	→ Age	abc nvarchar
Gender	abc String	→ Gender	abc nvarchar
Height_cm	abc String	→ Height_cm	abc nvarchar
Weight_kg	abc String	→ Weight_kg	abc nvarchar
BMI	abc String	→ BMI	abc nvarchar

This screenshot shows the pipeline editor with the 'Sortie' (Output) tab active. It displays the execution status of the pipeline 'pipeline1'. The status shows an error ('Echoué') for the 'Copy data1' activity. A green box highlights the 'Sortie' tab and the error message.

État du pipeline: Echoué

ID d'exécution de pipeline: 4048a369-9120-48ad-bdd9-4cfa7f2ed87c

Statut des éléments: Affichage des éléments 1-1 sur 1

Nom de l'activité	Statut de l'activité	Type d'activité	Début de l'exécution	Durée	Runtime d'intégration
Copy data1	Echoué	Copier les données	7/16/2025, 10:23:42 PM	20s	AutoResolve/integration/runtime (France Central)

The screenshot shows the 'Exécutions de pipeline' (Pipeline Executions) page in the Azure Data Factory portal. A single execution named 'pipeline1' is listed, which failed at 10:23:39 PM on 7/16/2025. A modal window titled 'Erreur' (Error) displays the following message:

```
Operation on target Copy data1 failed: Failure happened on 'Sink' side. Type=Microsoft.Data.SqlClient.SqlException,Message=CREATE TABLE permission denied in database 'mabasededonnee'.Source=Framework Microsoft SqlClient Data Provider.'
```

l'utilisateur que la Data Factory utilise pour se connecter à la base de données n'a pas la permission de CREATE TABLE (CRÉER des tables) dans la base de données mabasededonnée.

Il faut donner la permission dans SQL : avec le compte admin.

The screenshot shows the 'Éditeur de requêtes (version préliminaire)' (Query Editor) in the Azure portal. The query window contains the following SQL commands:

```
1 USE mabasededonnee;
2 ALTER ROLE db_ddladmin ADD MEMBER [datafactoryjcf];
```

5. Exécution du pipeline :

- Lancez le pipeline pour importer les données du fichier (ex : ventes.csv) dans la table Ventes.

The screenshot shows the 'Ressources de fabrique' (Fabric Resources) blade in the Azure portal. On the left, under 'Pipelines', the 'pipeline1' is selected. The main area shows the pipeline structure with activities: 'DelimitedText1' (source), 'AzureSqlTable1' (sink), and 'pipeline1' (parent). The 'Copier les données' (Copy data) activity is highlighted. At the bottom, the 'Etat du pipeline' (Pipeline Status) section shows the pipeline is successful ('Opération réussie').

5. Exercice 4 : Transformation des données avec Azure Data Factory

Objectif : Effectuer des transformations de données (par exemple, filtrer des lignes ou effectuer des agrégations) avant de charger les données dans la base de données.

Étapes :

1. Ajout d'une activité de transformation :

- Dans Azure Data Factory, ajoutez une activité Data Flow dans votre pipeline pour effectuer une transformation sur les données avant de les charger dans SQL.

J'ai fait 3 flux de transformations :

Premier flux de transformation :



Importation de donnée, passage en integer des valeurs, **filtre sur les personne non malade**, transfert de la table vers ma base de donnée

Deuxième flux de transformation :



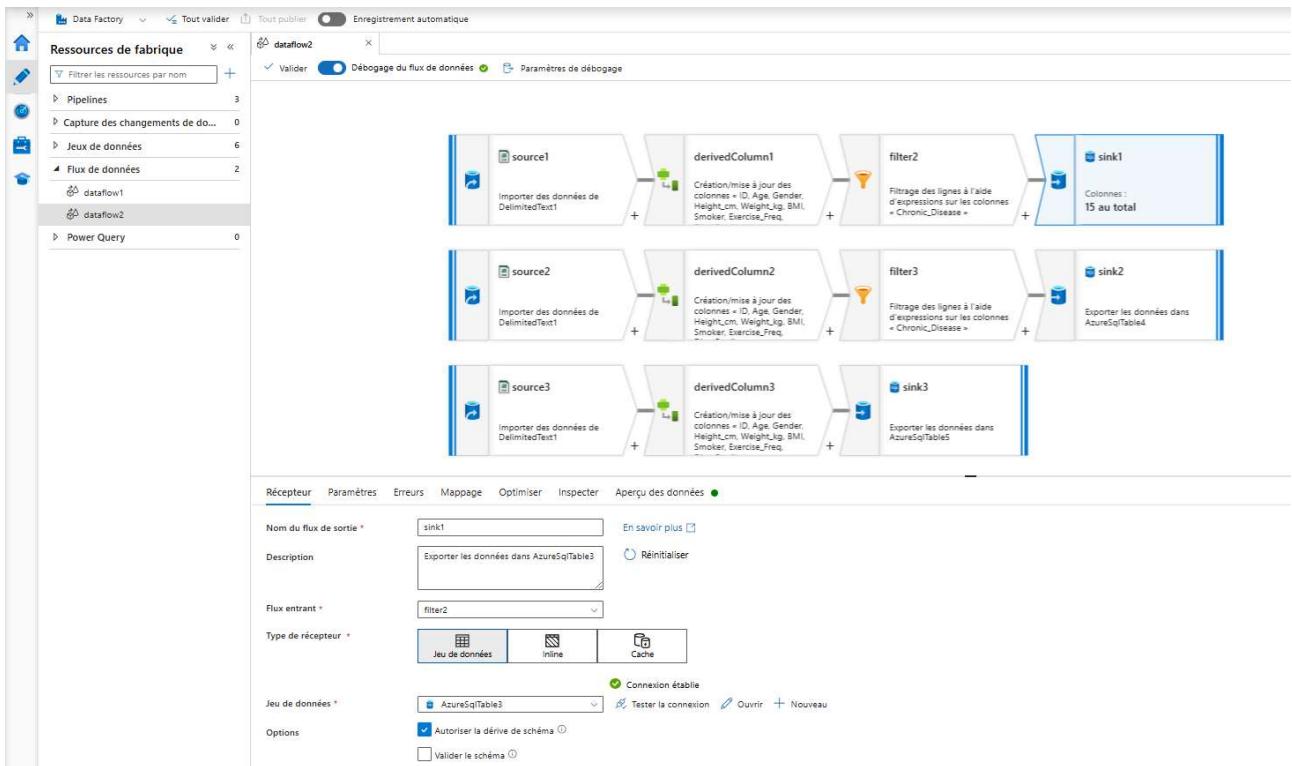
Importation de donnée, passage en integer des valeurs, **filtre sur les personne malade**, transfert de la table vers ma base de donnée

Troisième flux de transformation :



Importation de donnée, passage en integer des valeurs, transfert de la table vers ma base de donnée

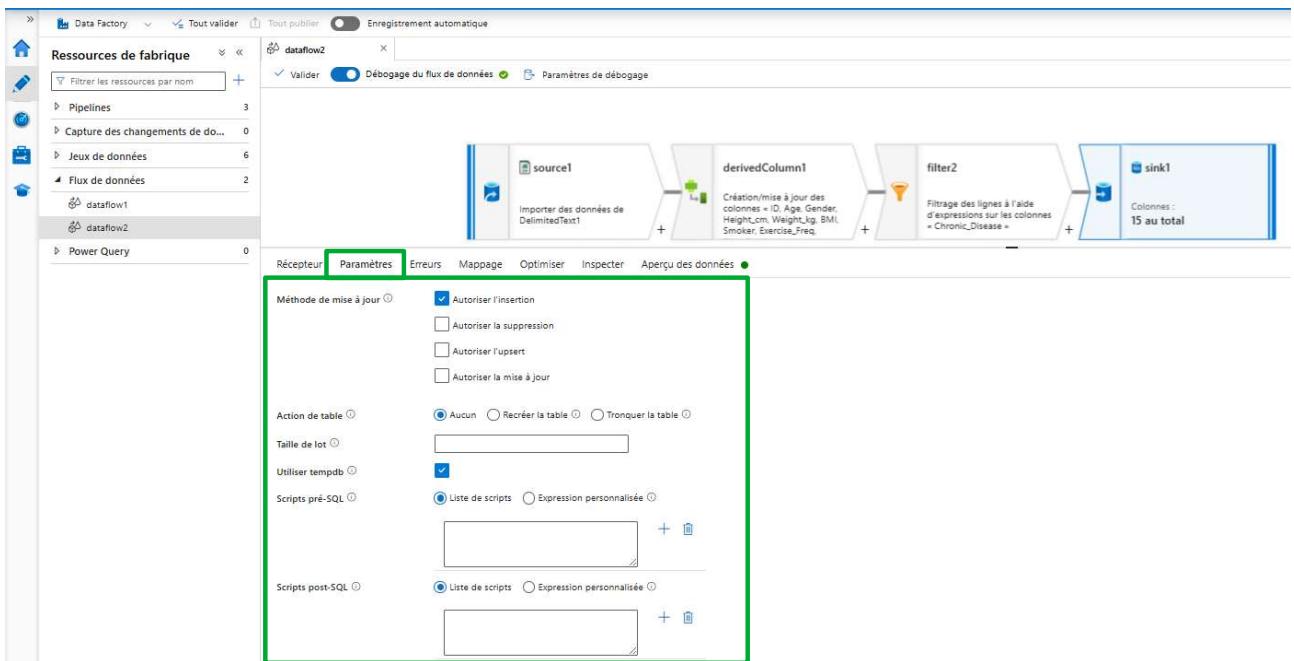
Flux global :



Explication en détail du premier flux de transformation et vérification des connexions :

The screenshot shows the Azure Data Factory interface with the following details:

- Pipeline:** pipeline3
- Source:** source1 (DelimitedText1)
- Configuration de la source:**
 - Nom du flux de sortie:** source1
 - Description:** Importer des données de DelimitedText1
 - Type de source:** Jeu de données
 - Jeu de données:** DelimitedText1 (Connexion établie)

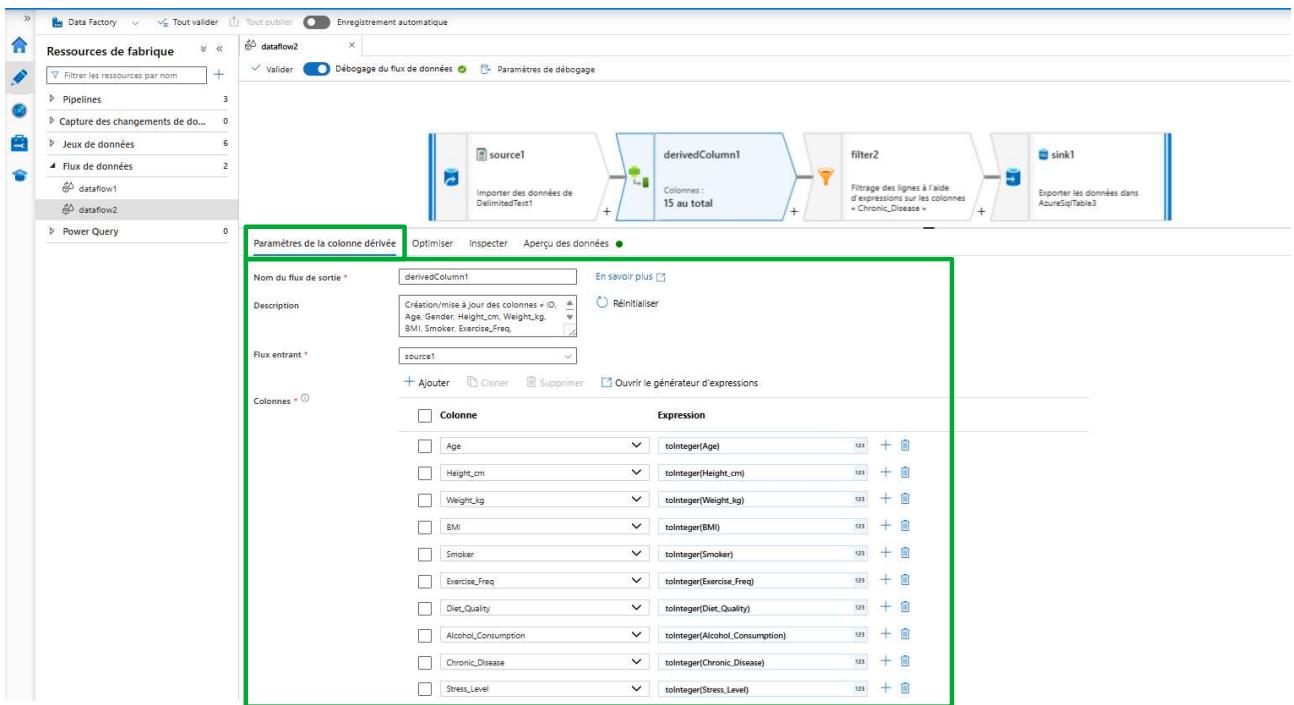


The screenshot shows the Azure Data Factory interface with the 'dataflow2' resource selected. The data flow diagram includes:

- source1**: Importer des données de DelimitedText
- derivedColumn1**: Création/mise à jour des colonnes = [D. Age, Gender, Height_cm, Weight_kg, BMI, Smoker, Exercise_Freq]
- filter2**: Filtrage des lignes à l'aide d'expressions sur les colonnes + Chronic_Disease =
- sink1**: Colonnes : 15 au total

A green box highlights the preview pane for the sink, which displays the following data:

ID	Age	Gender	Height_cm	Weight_kg	BMI	Smoker	Exercise_Freq	Diet_Quality	Alcohol_Consumption	Chronic_Disease	Stress_Level	Sleep_Hours	Sleep_Category	IMC
1	56	Other	177	37	11	1	0	0	0	0	9	8.5	Beaucoup	Dénutrition
2	69	Other	169	70	24	0	1	2	3	0	2	5.9	Peu	Normal
3	46	Female	159	69	27	0	3	3	2	0	3	4.8	Peu	Surpoids
4	32	Male	170	76	26	0	2	3	2	0	9	6.6	Moyen	Surpoids
6	25	Female	167	62	22	0	3	2	2	0	1	5.8	Peu	Normal
8	56	Male	165	65	23	0	3	2	2	0	7	6.8	Moyen	Normal
9	36	Male	176	41	13	0	3	1	0	0	6	6.3	Moyen	Dénutrition
11	28	Other	175	82	26	1	1	0	1	0	10	9.7	Beaucoup	Surpoids
12	28	Male	195	82	21	0	2	1	2	0	2	7.7	Moyen	Normal
13	41	Other	164	60	22	0	2	0	3	0	6	7.2	Moyen	Normal
14	53	Male	171	70	23	0	1	3	1	0	6	6.9	Moyen	Normal
15	57	Other	169	78	27	0	0	1	3	0	6	8.0	Beaucoup	Surpoids

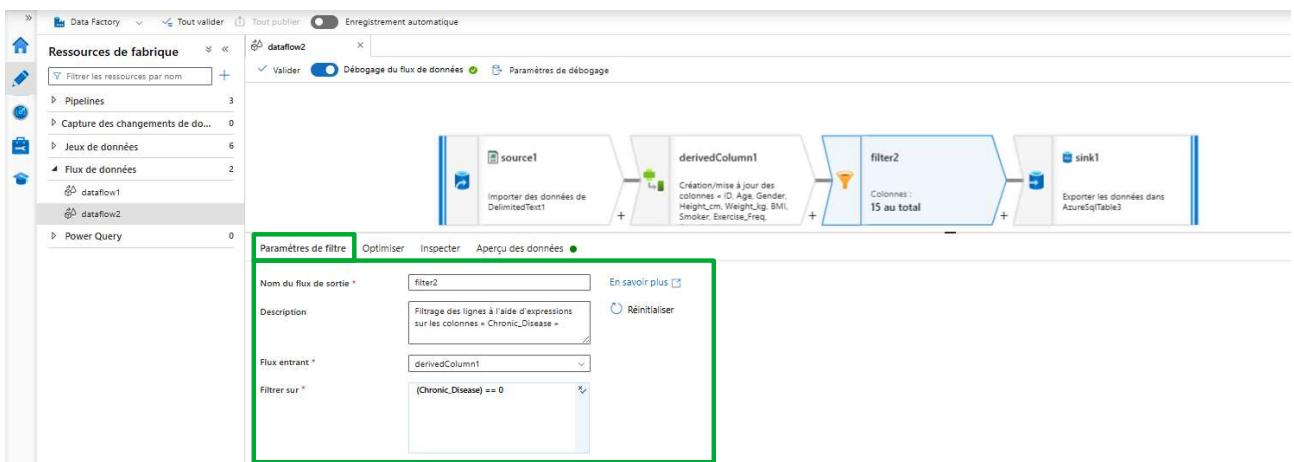


dataflow2

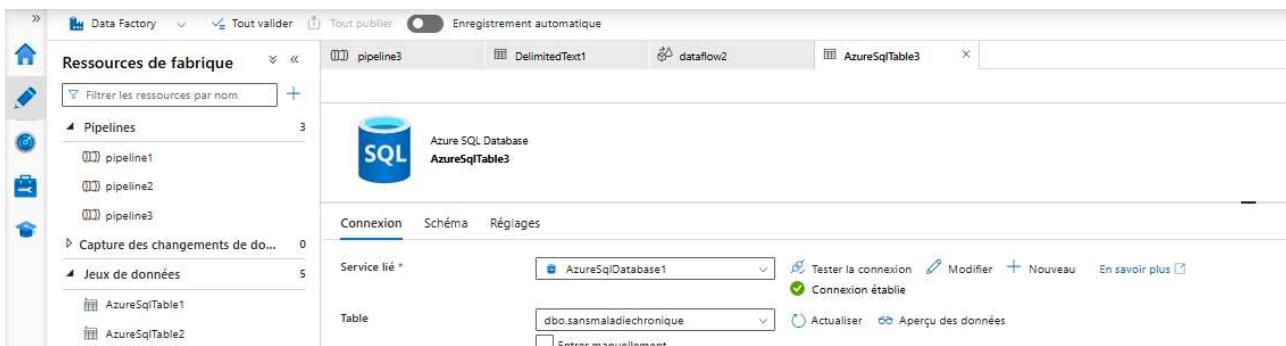
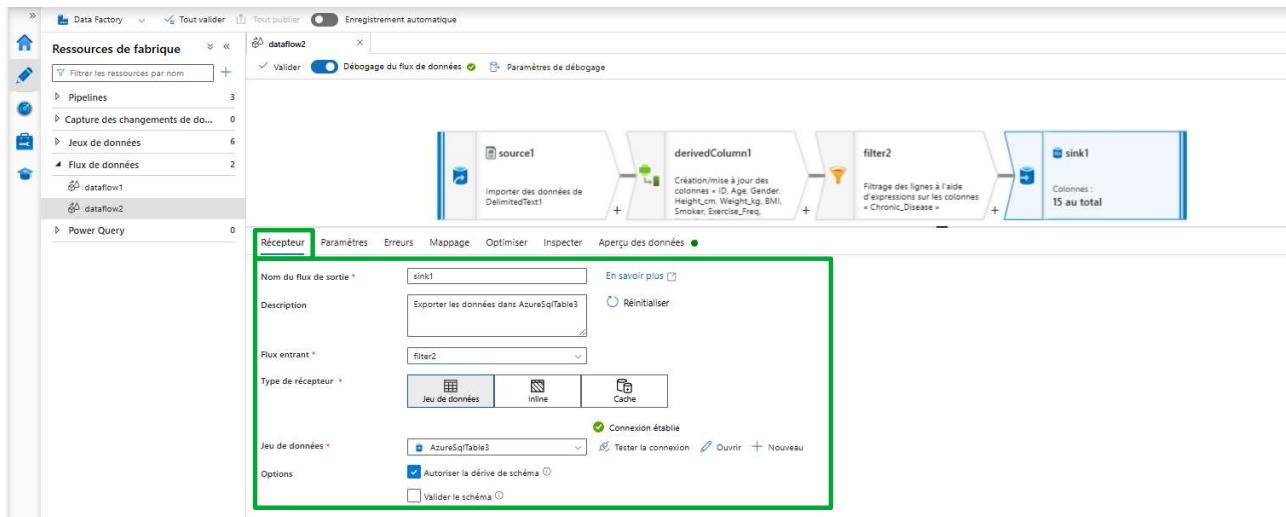
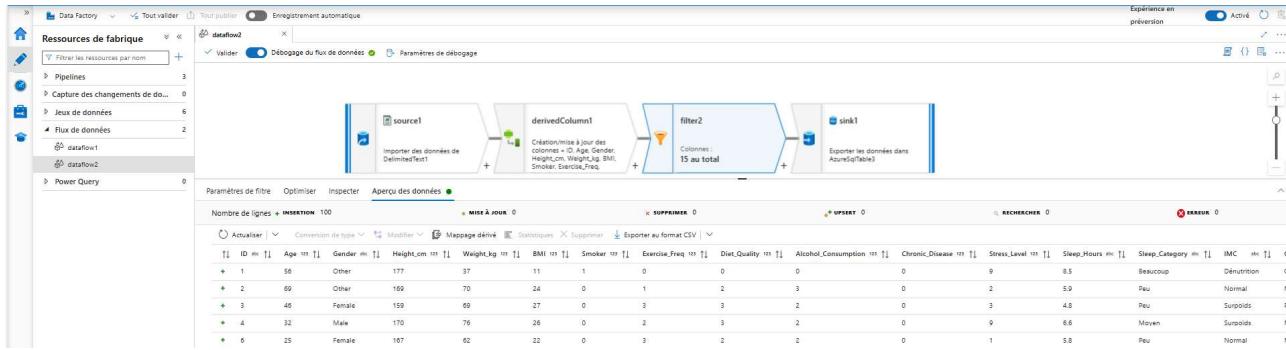
Paramètres de la colonne dérivée

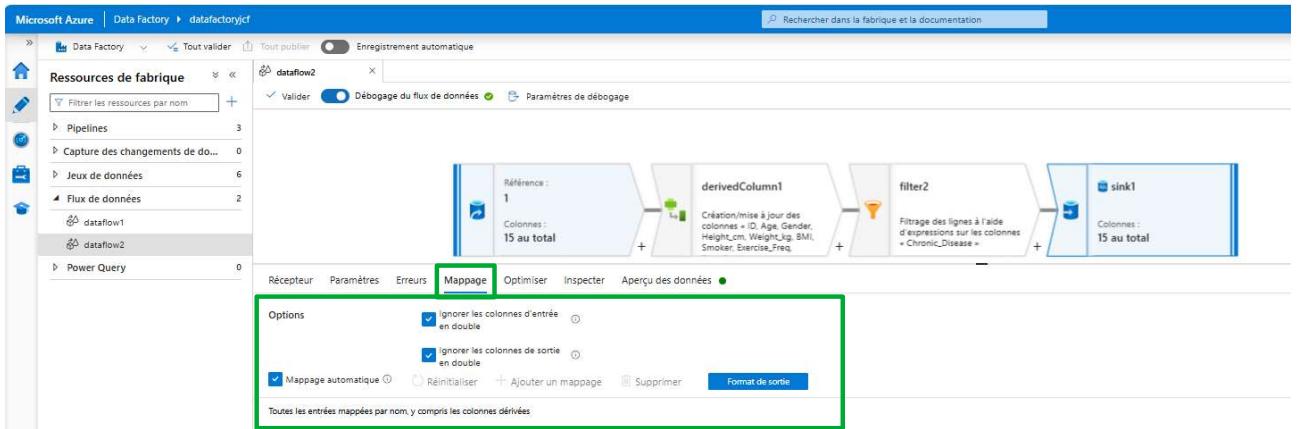
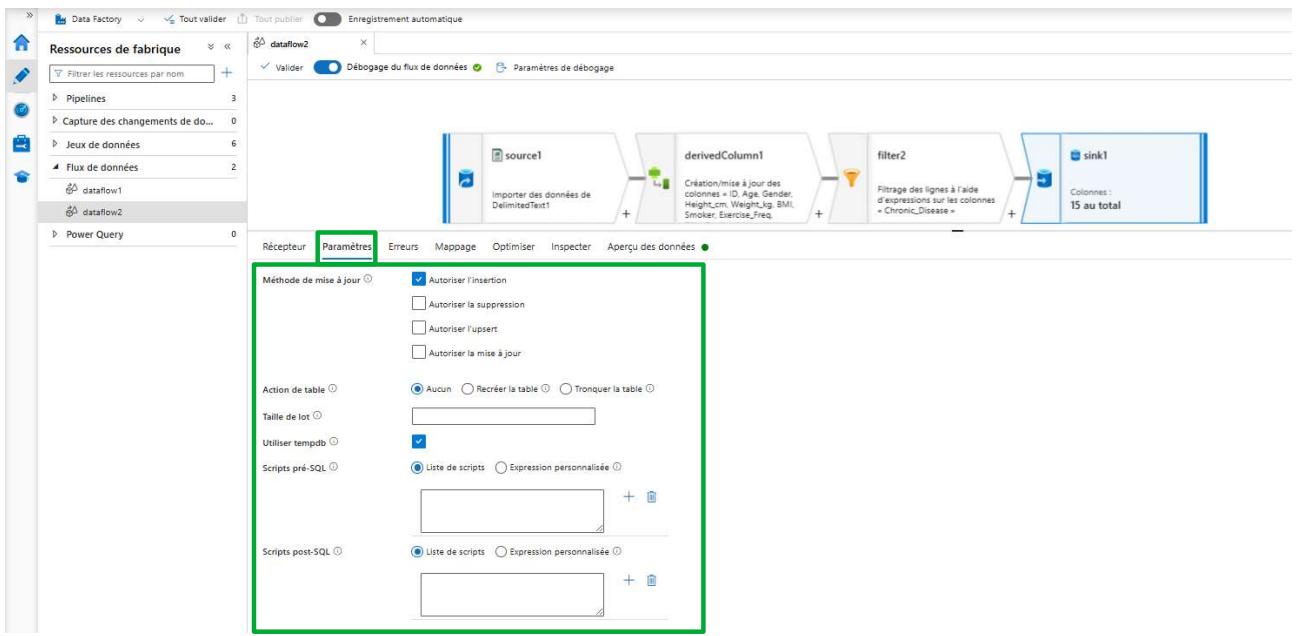
Nombre de lignes: 100

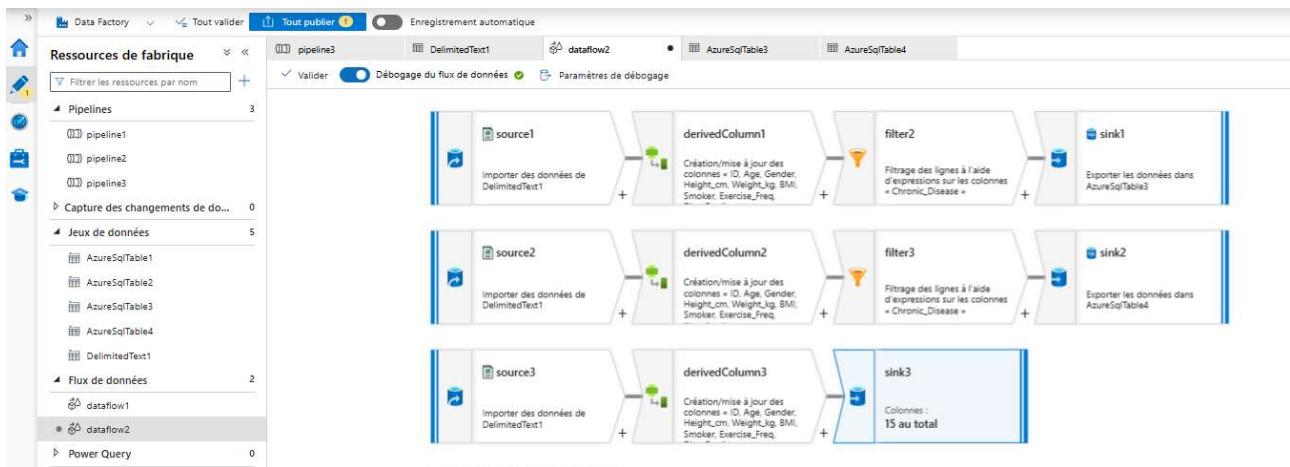
ID	Age	Gender	Height_cm	Weight_kg	BMI	Smoker	Exercise_Freq	Diet_Quality	Alcohol_Consumption	Chronic_Disease	Stress_Level	Sleep_Hours	Sleep_Category	IMC
1	56	Other	177	37	11	1	0	0	0	0	9	8.5	Beaucoup	Dénutrition
2	69	Other	169	70	24	0	1	2	3	0	2	5.9	Peu	Normal
3	46	Female	159	69	27	0	3	3	2	0	3	4.8	Peu	Surpoids
4	32	Male	170	76	26	0	2	3	2	0	9	6.6	Moyen	Surpoids
5	60	Male	158	60	24	0	2	3	1	1	6	6.1	Moyen	Normal



Le filtre fonctionne bien les valeur 0 sont bien sélectionnées :







Vérification des connections pour sink 2 et sink 3 :

Lancement du dataflow 2 dans le pipe 3 :

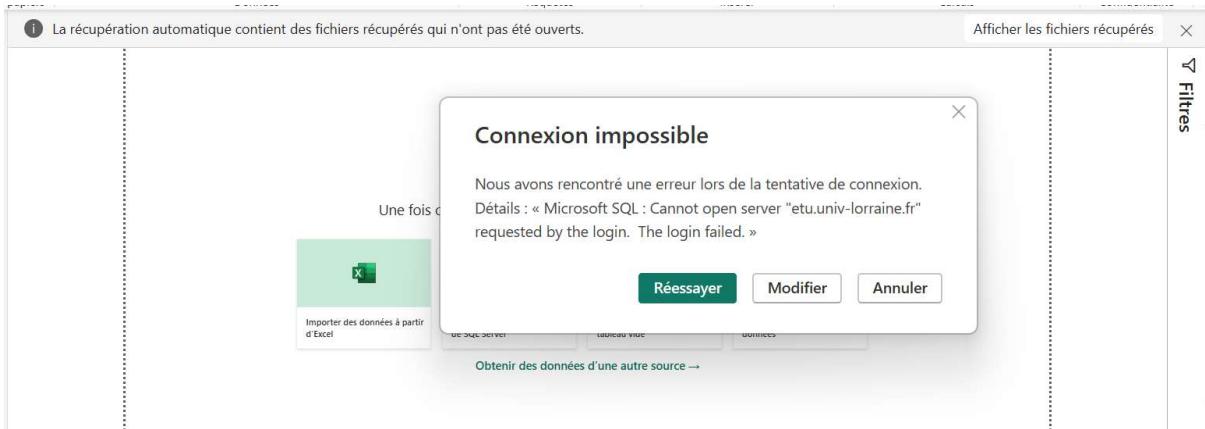
Les tables se retrouvent bien dans la base de donnée SQL :

The screenshot shows the Microsoft Azure portal interface. At the top, there's a blue header bar with the Microsoft Azure logo, a search bar, and a 'Copilot' button. Below the header, the URL 'Accueil > mabasededonnée (faber/mabasededonnée)' is visible. The main content area shows a 'Base de données SQL' named 'mabasededonnée'. On the left, a sidebar lists various options like 'Vue d'ensemble', 'Journal d'activité', 'Étiquettes', 'Diagnostiquer et résoudre les problèmes', and 'Éditeur de requêtes (version préliminaire)'. The 'Éditeur de requêtes' option is selected and highlighted with a grey background. The main pane displays the database structure, specifically the 'Tables' section, which is also highlighted with a green box. The tables listed are: dbo.avecmaladiechronique, dbo.IMC.csv, dbo.sansmaladiechronique, and dbo.sansmodification. To the right of the database structure, there's a 'Requête 1' editor window with a single digit '1' in it.

- Par exemple, vous pouvez :
 - Filtrer les ventes avec un montant supérieur à 100.
 - Effectuer des agrégations (par exemple, la somme des ventes par client).
2. Configuration du Data Flow :
- Sélectionnez la source de données (le fichier ventes.csv depuis le Blob Storage).
 - Ajoutez des transformations (par exemple, une activité Filter pour ne garder que les ventes au-dessus d'un certain seuil).
 - Configurez la destination pour insérer les données transformées dans une table SQL dédiée dans Azure SQL Database.
3. Exécution du pipeline avec transformation :
- Exécutez le pipeline et vérifiez que les données transformées ont bien été insérées dans la base de données.

6. Problème de connexion avec la base de donnée

J'ai un problème de connexion avec la base de donnée pour la connexion avec power-by:



J'ai vérifié mon pare-feu microsoft azure :
Mon adresse IP est bien indiquée

Nom de la règle	Adresse IPv4 de début	Adresse IPv4 de fin
ClientIP-2025-7-15_12-19-11	91.168.170.226	91.168.170.226
ClientIPAddress_2025-7-16_11-13-6	91.168.170.226	91.168.170.226
ClientIPAddress_2025-7-18_14-58-7	193.50.135.205	193.50.135.205

Quand je me connecte avec `jean-claude.faber1@etu.univ-lorraine.fr`,
Power BI transmet mon identité Azure Active Directory (AAD) à Azure SQL Server.

Mais Azure SQL appartient au compte universitaire.

mon compte `@etu.univ-lorraine.fr` n'est pas connu .

Résultat : le serveur SQL pense que "etu.univ-lorraine.fr" est un *autre serveur* d'où le message :

"Cannot open server 'etu.univ-lorraine.fr'"

C'est une interprétation SQL d'un utilisateur AAD inconnu.

Dans Microsoft Entra ID, je ne peux pas m'autoriser à accéder à la base de donnée avec les informations suivantes

- **Serveur SQL :** faber.database.windows.net
- **Base de données :** mabasededonnee
- **Mon identifiant étudiant :** jean-claude.faber1@etu.univ-lorraine.fr

Vous n'avez pas accès

Copiez les détails de l'erreur et envoyez-les à votre ou vos administrateurs pour accéder à cette page.

Copier les détails de l'erreur

Détails de l'erreur	Nom du groupe de ressources
ID de session 759f4762460245458f67b651ae66397d	
ID d'abonnement	ID de ressource
Code d'erreur 401	Détails Priviléges insuffisants pour effectuer l'opération.
Nom de la ressource	

Contribuer à l'amélioration de cette page

Envoyer des commentaires

J'ai envoyé un mail au service informatique :

Bonjour,

Dans le cadre de mon projet/cours, je dois me connecter à la base de données Azure SQL via Power BI.

Je rencontre actuellement une erreur de connexion ("Login Failed").

Comme mon compte est géré par l'Université de Lorraine, pourriez-vous s'il vous plaît effectuer les deux actions suivantes pour m'autoriser l'accès ?

Mes informations :

- **Serveur SQL :** faber.database.windows.net
- **Base de données :** mabasededonnee
- **Mon identifiant étudiant :** jean-claude.faber1@etu.univ-lorraine.fr

Les actions requises de votre part sont les suivantes

1. **Inviter mon compte** (`jean-claude.faber1@etu.univ-lorraine.fr`) dans votre annuaire Microsoft Entra ID en tant qu'utilisateur invité (externe).
2. **Exécuter les commandes SQL suivantes** sur la base `mabasededonnee` pour créer mon utilisateur et lui donner les droits de lecture :

CREATE USER [jean-claude.faber1@etu.univ-lorraine.fr] FROM EXTERNAL PROVIDER;

ALTER ROLE db_datareader ADD MEMBER [jean-claude.faber1@etu.univ-lorraine.fr];

Je vous remercie par avance pour votre aide et reste à votre disposition si besoin.

Cordialement,

Jean-Claude Faber

The screenshot shows a ticketing interface for a user named Jean-Claude Faber. At the top, it says "Demande Utilisateur ~ R-0075540 créée(e)". Below this, the title "DEMANDES EN COURS (1)" is displayed. A table lists one open request:

Nom complet	Titre	Dernière mise à jour	Statut	Service	Catégorie	Demandeur
R-0075540	Demande d'accès Power BI à la base Azure SQL (étudiant : jean-claude.faber1)	21/07/2025 14:02	Affectée	Mon environnement de travail	Mon équipe numérique de proximité	Jean-Claude Faber

Ci-joint la réponse.

DEMANDE D'ACCÈS POWER BI À LA BASE AZURE SQL (ÉTUDIANT : JEAN-CLAUDE.FABER1)

Un commentaire a été ajouté à votre ticket

21/07/2025 14:40 - Charles Monnier:

Bonjour,

Votre demande est des plus surprenantes, je vais me renseigner sur sa faisabilité.
En attendant, noté que nos infrastructures subissent quelques difficultés en ce moment. Certains services, y compris ceux d'authentifications, sont instables. Cela pourrait être lié à votre problème ("Login Failed")

Tout ce que je peux vous proposer pour le moment, c'est de patienter...

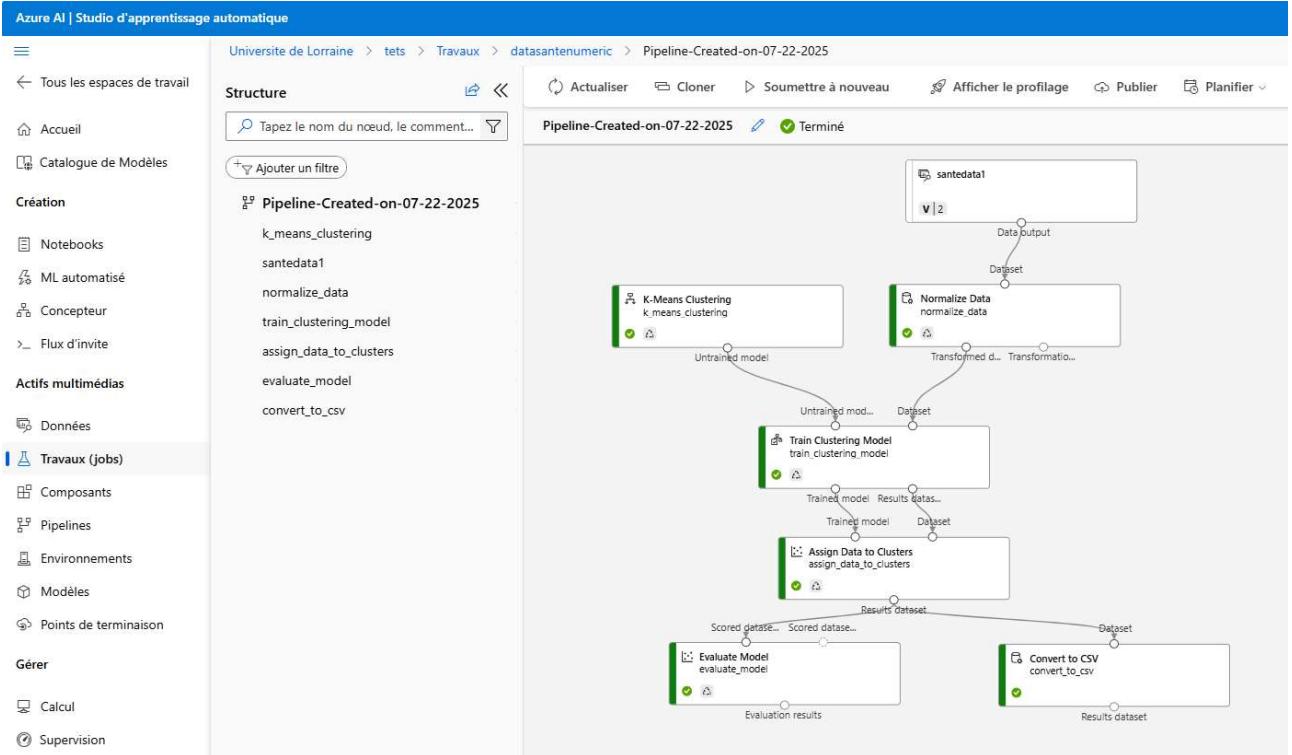
Bien à vous,
Charles

7. Test de cluster dans Azure machine learning Workspace :

The screenshot shows the Microsoft Azure portal interface. At the top, there's a search bar and a Copilot button. Below the header, a navigation bar titled "Services Azure" contains icons for creating resources, cost management, Azure Machine Learning, subscriptions, conditional access, SQL databases, security, and other services. Under the "Ressources" section, a table lists a single item: "tets" (Azure Machine Learning workspace). The "Dernier affichage" column indicates it was viewed "il y a quelques secondes".

This screenshot shows the detailed configuration page for the "tets" workspace. On the left, a sidebar lists options like "Vue d'ensemble", "Journal d'activité", "Contrôle d'accès (IAM)", "Étiquettes", "Diagnostiquer et résoudre les problèmes", "Visualiseur de ressources", "Événements", "Paramètres", "Supervision", "Automatisation", and "Support + dépannage". The main content area displays basic workspace details: Group resource: dashboards, Location: France Central, Subscription: Azure for Students, Storage: tets988426148, Provisioning State: Succeeded, Studio web URL: https://ml.azure.com/tid=158716cf-46b9-4bca-8c49-c7bb67e575f3&wsid=/subscriptions/..., Container Registry: clustersante1, Key Vault: tets3887166331, Application Insights: tets4652671837, and MLflow tracking URI: azureml://francecentral.api.azurerm.ms/mlflow/v1.0/subscriptions/ba4dc137-e617-425b-a... . A "Vue JSON" link is also present. At the bottom, a callout box encourages users to "Work with your models in Azure Machine Learning Studio" and provides a "Launch studio" button.

Création d'un pipe pour test de clustering :



Les données ont été retravaillées sur Jupyter Notebook pour avoir uniquement des valeurs numériques :

Ci joint le Notebook :

Transformation du fichier pour clustering dans Azure Ai studio, passage en numérique du genre :

```
[89]: def genre(g):
    if g=='Male':
        return 0
    elif g=='Female':
        return 1
    elif g=='Other':
        return 2

df['Gender_category'] = df['Gender'].apply(genre)

print(df[['Gender_category', 'Gender']].head())
   Gender_category  Gender
0                2    Other
1                2    Other
2                1   Female
3                0    Male
4                0    Male
```

Suppression des colonnes non numériques et redondantes :

```
[92]: df = df.drop(['ID','Gender', 'Sleep_Category', 'IMC', 'Categorie_taille','Height_cm','Weight_kg'], axis = 1)
df = df.rename(columns={'BMI':'IMC'})
print (df.head())
|
```

	Age	IMC	Smoker	Exercise_Freq	Diet_Quality	Alcohol_Consumption	Chronic_Disease	Stress_Level	Sleep_Hours	Gender_category
0	56	11.8	1	0	0	0	0	0	8.5	2
1	69	24.7	0	0	1	2	0	0	5.9	2
2	46	27.3	0	0	3	3	0	0	4.8	1
3	32	26.3	0	0	2	3	0	0	6.6	0
4	60	24.1	0	0	2	3	0	1	6.1	0

Vérification qu'il n'y a que des valeurs numériques :

```
[94]: df.dtypes
```

[94]:	Age	int64
	IMC	float64
	Smoker	int64
	Exercise_Freq	int64
	Diet_Quality	int64
	Alcohol_Consumption	int64
	Chronic_Disease	int64
	Stress_Level	int64
	Sleep_Hours	float64
	Gender_category	int64
	dtype:	object

Vérification qu'il n'y a pas de valeur nulle :

```
: df.isnull().sum()
```

:	Age	0
	IMC	0
	Smoker	0
	Exercise_Freq	0
	Diet_Quality	0
	Alcohol_Consumption	0
	Chronic_Disease	0
	Stress_Level	0
	Sleep_Hours	0
	Gender_category	0
	dtype:	int64

Exportation du fichier transformé :

```
[99]: df.to_csv("C:\\\\Users\\\\Etudiant\\\\Desktop\\\\Cours\\\\cloud\\\\sujet_a_rendre_web\\\\donneesantepourcluster.csv", index=False)
```

Ainsi les données **donneesantepourcluster.csv** ont été utilisées pour le pipe sous azure AI.

Création du pipe :

The screenshot shows the 'Configurer la tâche de pipeline' (Configure Pipeline Task) dialog box. Under 'Informations de base', the 'Nom de l'expérience' dropdown is set to 'Sélectionner une expérience existante' (Select existing experience), with 'Pipeline-Created-on-07-22-2025' selected. The 'Description du travail' field contains the placeholder 'Pipeline created on 20250722'. In the 'Balises de travail' section, there is a 'Nom' input field with 'Valeur' next to it, and an 'Ajouter' (Add) button.

The screenshot shows the 'Créer une ressource de données' (Create a data resource) dialog box. The left sidebar shows the 'Données' (Data) category is selected. The main pane shows the 'Type de données' (Data type) step, with 'Source de données' (Data source) highlighted. The right pane displays the 'Choisir une source pour votre ressource de données' (Choose a source for your data resource) section, which includes options for 'À partir d'un URI' (From a URI) and 'À partir de fichiers locaux' (From local files). A note at the bottom states: 'Créez une ressource de données à partir des services de stockage de données inscrits, notamment Azure Blob Storage, le partage de fichiers Azure et Azure Data Lake.'

Azure AI | Studio d'apprentissage automatique

Créer une ressource de données

Sélectionner un magasin de données

Choisissez un type de stockage et un magasin de données dans lequel charger vos données à l'étape suivante. Vous pouvez également créer d'abord un magasin de données pour vos données.

Type de magasin de données *

- Stockage Blob Azure
- Créer un magasin de données

Rechercher dans le magasin

Nom	Nom du stockage	Créé le
workspaceblobstore	tets9884226148	Jul 22, 2025 5:10 PM
workspaceartifactstore	tets9884226148	Jul 22, 2025 5:10 PM

Filtrer Colonnes

Azure AI | Studio d'apprentissage automatique

Créer une ressource de données

Choisir un dossier ou des fichiers

Choisissez les fichiers ou dossiers à charger à partir de votre lecteur local. Avec le type de uri_folder, vous pouvez charger un seul dossier, ou vous pouvez charger plusieurs fichiers ou dossiers et ils sont stockés dans un dossier conteneur.

Chemin de chargement

azureml://subscriptions/ba4dc137-e617-425b-ae0f-8da48309d70b/resourcegroups/dashboard...

Charger des fichiers ou un dossier

Remplacer s'il existe déjà

Télécharger la liste

datasante.csv ✓ 406.79 Ko/406.79 Ko

Azure AI | Studio d'apprentissage automatique

Créer une ressource de données

Vérifier

Passez en revue les paramètres de votre ressource de données et apportez les modifications nécessaires.

Type de données

Nom	santedata
Description	...
Type	uri_folder

Source de données

Type	Local
------	-------

Sélection du dossier

Chemin de chargement
azureml://subscriptions/ba4dc137-e617-425b-ae0f-8da48309d70b/resourcegroups/dashboard/resourcegroups/workspaceblobstore/paths/UI/2025-07-22_184516_UTC/

Fichiers chargés
datasante.csv

Stockage

Type de magasin de données	AzureBlob
Nom du magasin de données	workspaceblobstore

Azure AI | Studio d'apprentissage automatique

Université de Lorraine > tets > Concepteur > Création

Annuler Rétablir Valider Afficher le lignage ...

Rechercher par nom, étiquette et descript... Ajouter un filtre

Étiquettes : All

Données Composant

Succès : L'élément de données santedata1_... Dernière mise... |

1 + Des exemples de données prédefinies se trouvent sous l'onglet Composant. Cliquez ici

santedata1 Version 1 Jean-Claude Faber 7/22/2025

Configurer la tâche de pipeline

Informations de base Entrées & Sorties Paramètres d'exécution Vérifier + envoyer

Informations de base

Nom de l'expérience Sélectionner un groupe existant Crée nouveau

Expérience existante * Sélectionner une expérience

Nom d'affichage de l'emploi Pipeline-Created-on-07-22-2025

Description du travail Pipeline created on 20250722

Balises de travail Nom : Valeur Ajouter

Azure AI | Studio d'apprentissage automatique

Université de Lorraine > tets > Travaux > datasantenumeric > Pipeline-Created-on-07-22-2025

Structure Actualiser Cloner Soumettre à nouveau Afficher le profilage Publier Planifier

Tapez le nom du noeud, le comment... Ajouter un filtre

Pipeline-Created-on-07-22-2025 Terminé

Données

Travaux (jobs)

Composants

Pipelines

Environnements

Modèles

Points de terminaison

Gérer

Calcul

Supervision

Étiquetage des données

Services liés PRÉVISION

Pipeline-Created-on-07-22-2025

```

graph TD
    KMeans[K-Means Clustering k_means_clustering] -- "Untrained model" --> Train[Train Clustering Model train_clustering_model]
    Normalize[Normalize Data normalize_data] -- "Transformed data..." --> Train
    Train -- "Trained model" --> Assign[Assign Data to Clusters assign_data_to_clusters]
    Assign -- "Results dataset" --> Evaluate[Evaluate Model evaluate_model]
    Evaluate -- "Evaluation results" --> Convert[Convert to CSV convert_to_csv]
    Convert -- "Results dataset" --> Output[Dataset]
    
```

Explication du pipe :

Le jeu de données est normalisé pour que toutes les caractéristiques aient la même importance.

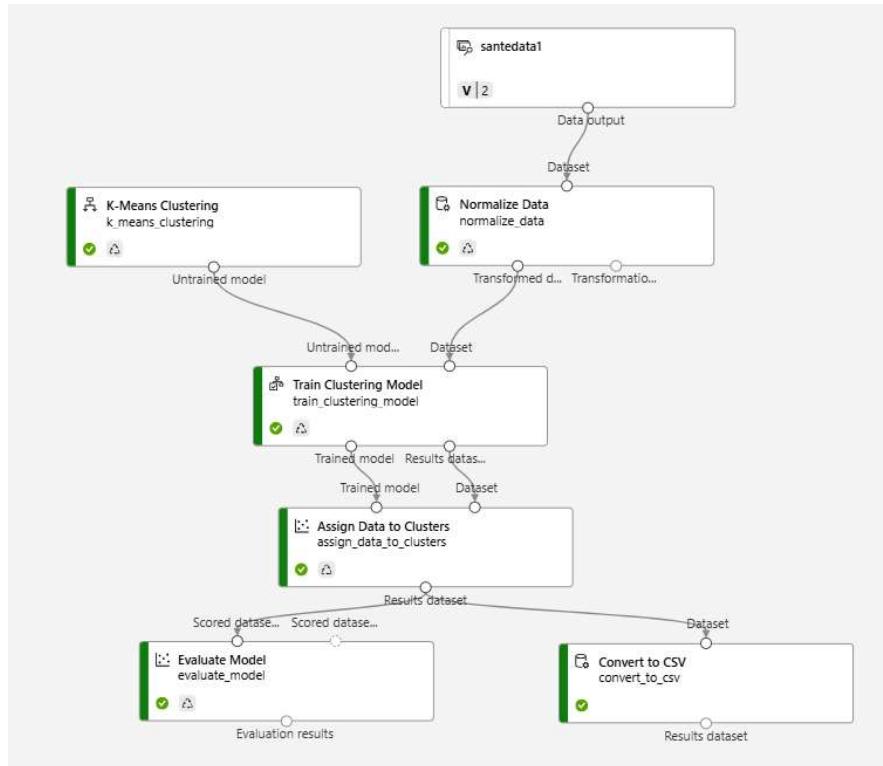
Le K-Means Clustering permet de définir l'algorithme de Clustering qui sera utilisé.

Le Train Clustering Model entraîne le modèle K-means en utilisant le jeu de données normalisé et le modèle non entraîné (K-Means Clustering), ce qui donne un modèle de clustering entraîné.

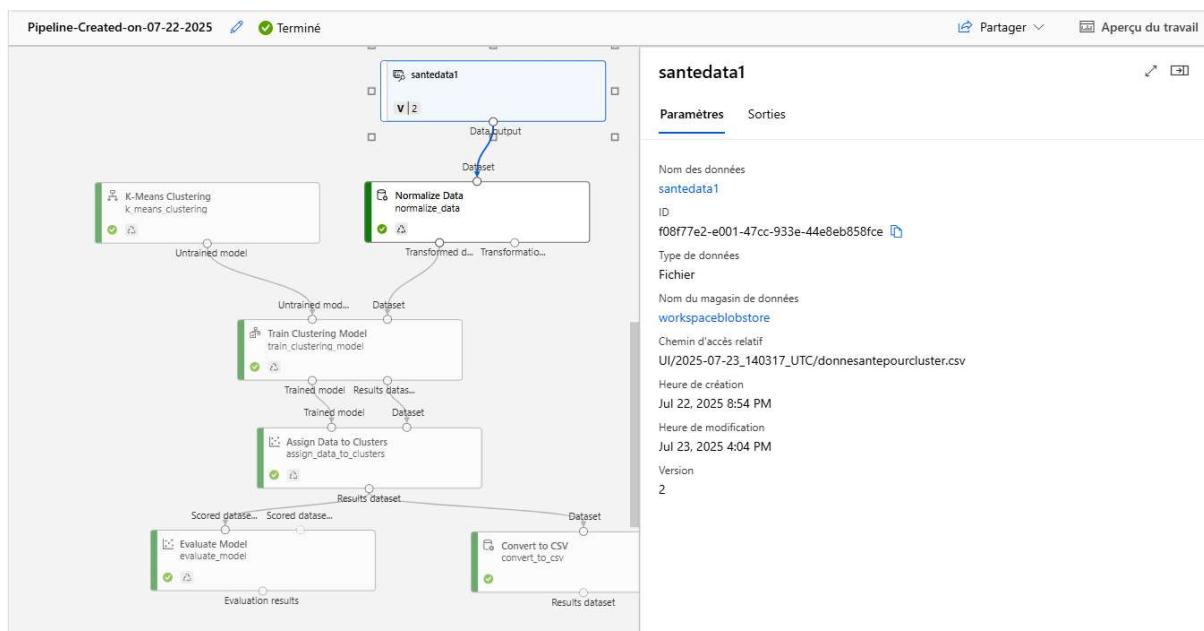
Assign Data to clusters permet au modèle entraîné d'assigner chaque données à un cluster.

Evaluate Model permet d'évaluer la qualité du clustering.

Convert to csv permet de convertir le jeu de données avec les cluster pour l'exportation et la visualisation qui sera réalisée dans Jupyter Notebook.



Paramètre du pipe :



Pipeline-Created-on-07-22-2025 Terminé

Normalize Data

- Paramètres**
- Transformation method**: ZScore
- Use 0 for constant columns when checked**: True
- Columns to transform**: Types de colonne: Numeric

Paramètres de sortie

- Regénérer la sortie**: Non

Transformed dataset

- Mode**: Upload
- Magasin de données**: workspaceblobstore

Pipeline-Created-on-07-22-2025 Terminé

K-Means Clustering

- Number of centroids**: 2
- Initialization**: K-Means++
- Random number seed**:
- Metric**: Euclidean
- Normalize features**: True
- Iterations**: 100

Pipeline-Created-on-07-22-2025 Terminé

Train Clustering Model

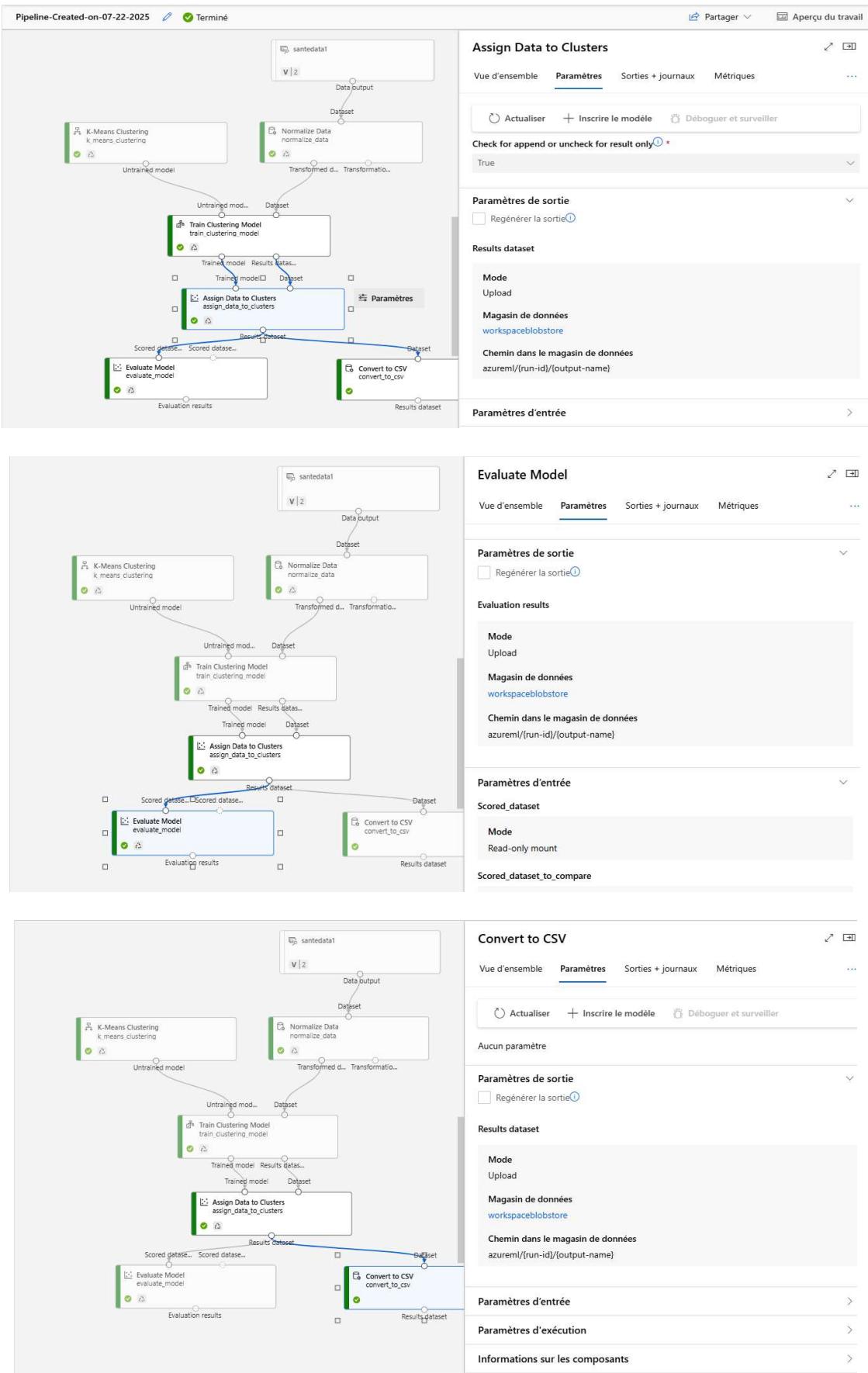
- Column set**: Types de colonne: Numeric
- Check for append or uncheck for result only**: True

Paramètres de sortie

Paramètres d'entrée

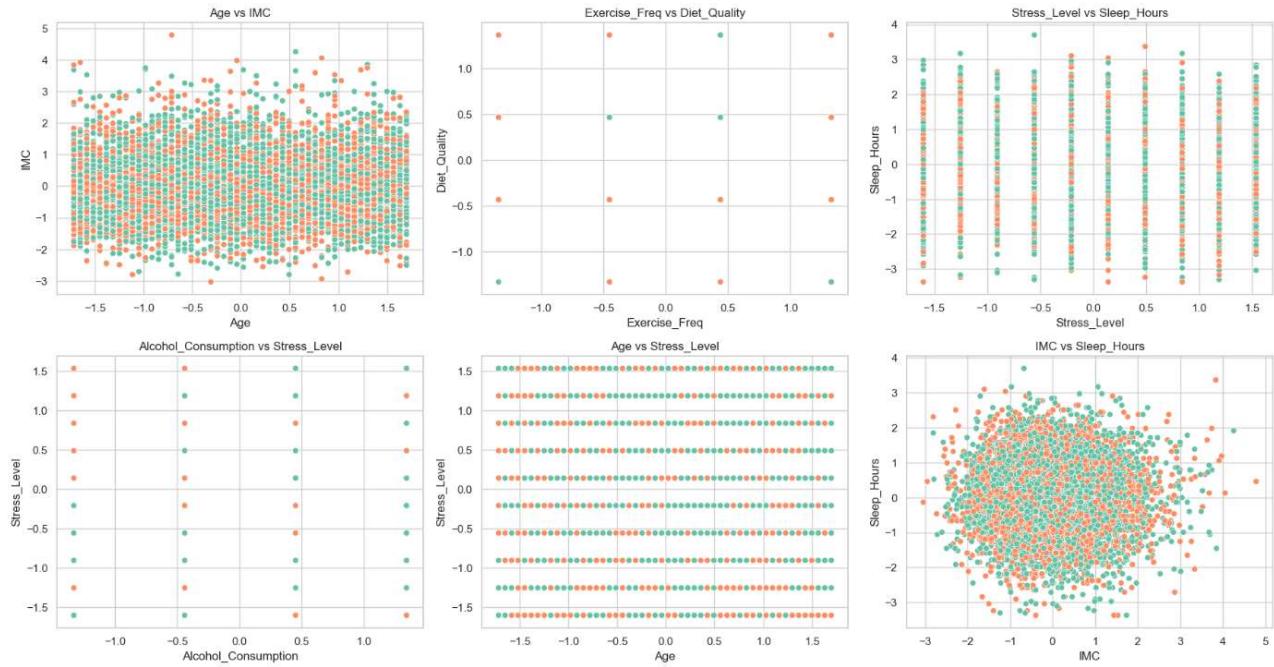
Paramètres d'exécution

Informations sur les composants



Les données de sortie du pipe ont été récupérées dans le fichier **clustersante.csv** pour visualisation dans Jupyter Notebook.

Visualisation des graphiques obtenus :



Conclusion :

Il n'y a pas de séparation nette des groupes, il n'y a pas de cluster qui se dégage.

8. Exercice 5 : Utilisation de PowerBI pour créer la visualisation de nos données

Objectif : Visualiser les données dans PowerBI après les avoir chargés dans Azure SQL Database.

1. Connexion à Azure SQL Database depuis Power BI :

- Dans Power BI, cliquer sur SQL Server dans l'onglet Accueil.
Il faut saisir l'adresse du serveur comme ci-dessous.

- A la suite de votre étape, vous pouvez sélectionner l'options de saisir les identifiants de votre serveur SQL renseigné lors de la création de la BDD pour y accéder.
- Vous pouvez récupérer la ou les tables qui vous intéressent.

Vu que je n'ai pas réussi à me connecter au serveur SQL j'ai extrait la table **dbo.sansmodification** qui m'a donnée **Requête3.csv** pour la travailler dans power by :

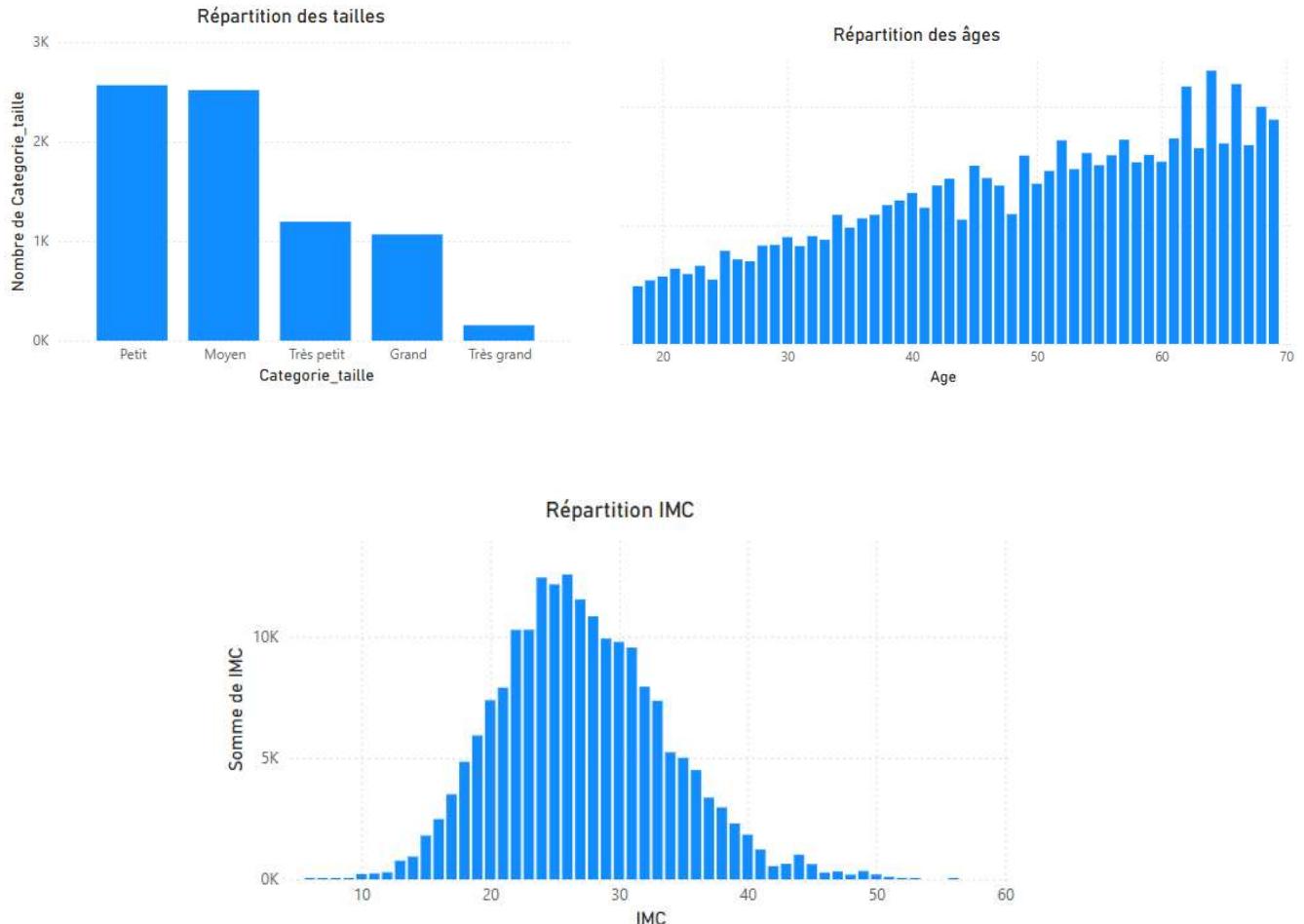


The screenshot shows the Microsoft Data Studio interface. On the left, there's a sidebar with various tools: 'Vue d'ensemble', 'Journal d'activité', 'Etiquettes', 'Diagnostiquer et résoudre les problèmes', 'Éditeur de requêtes (version préliminaire)' (which is selected), 'Base de données miroir dans Fabric (préversion)', 'Visualiseur de ressources', 'Paramètres', 'Calcul + stockage', 'Chaines de connexion', and 'Propriétés'. The main area shows the database 'mabasededonnée' selected. In the center, there's a tree view of the database structure under 'Tables': 'dbo.avecmaladiechronique', 'dbo.IMC.csv', 'dbo.sansmaladiechronique', and 'dbo.sansmodification'. The 'dbo.sansmodification' table is highlighted. At the bottom, there's a query editor window with the SQL command: 'SELECT * FROM dbo.sansmodification'.

2. Créations de rapports :

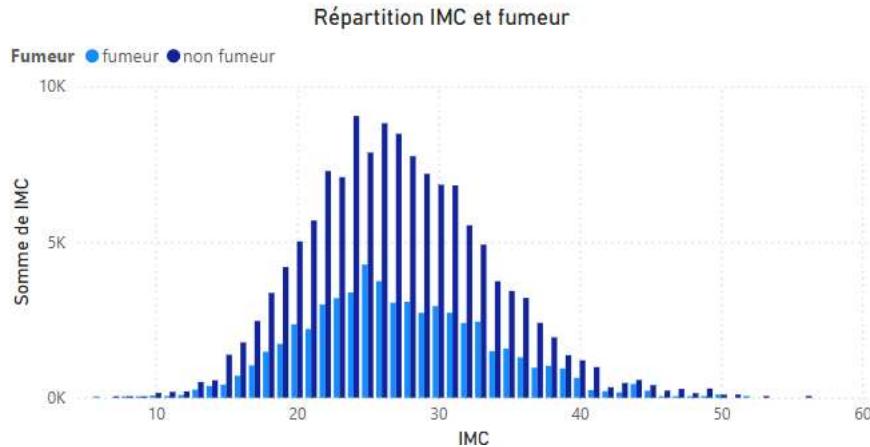
- Création d'un tableau de bord à partir de vos données.

Répartition des données initiales :

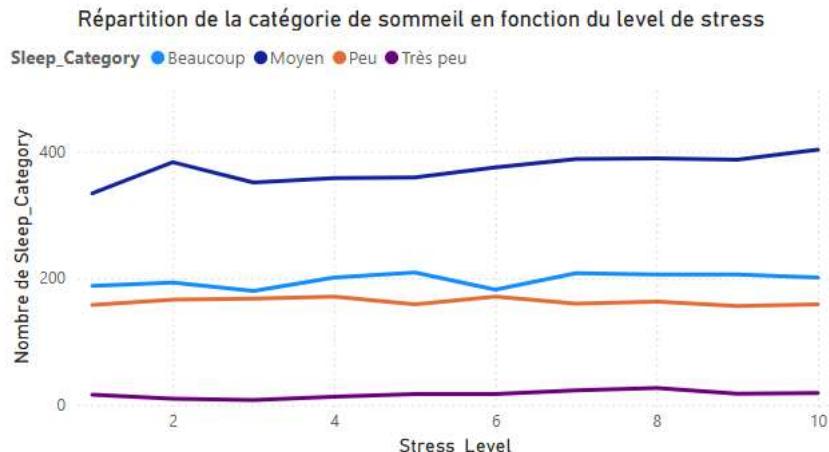


La taille de la population est plutôt petite et moyenne, pour l'âge nous sommes sur une majorité d'adulte et de personnes agée et pour l'IMC nous avons une distribution normale avec une fourchette majoritaire entre 20 et 30.

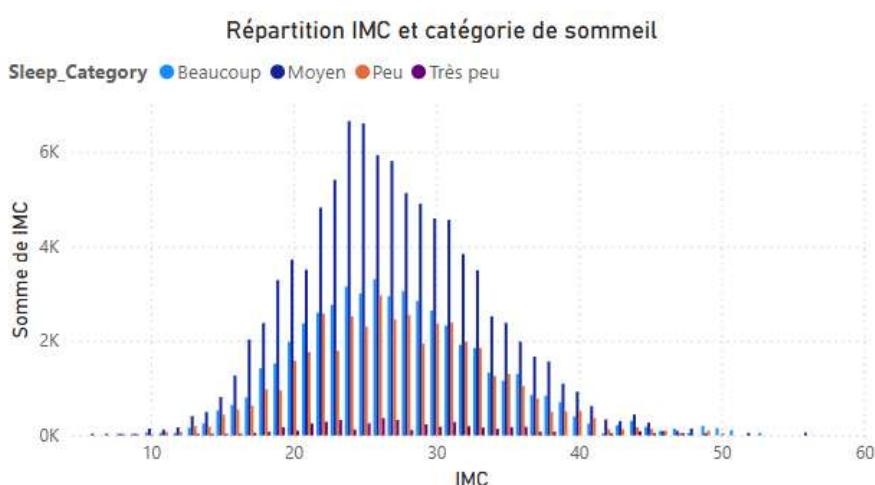
Relation entre les données :



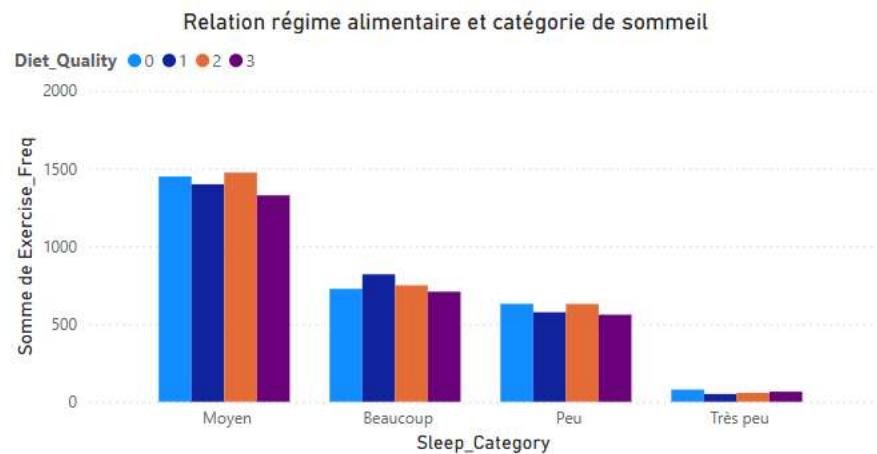
L'IMC n'a pas d'impact sur le fait de fumer ou de ne pas fumer car dans les deux cas la répartition est proche il n'y a pas de différence significative.



Le niveau de stress n'a pas d'impact sur les catégories de sommeil car les valeurs sont stables.



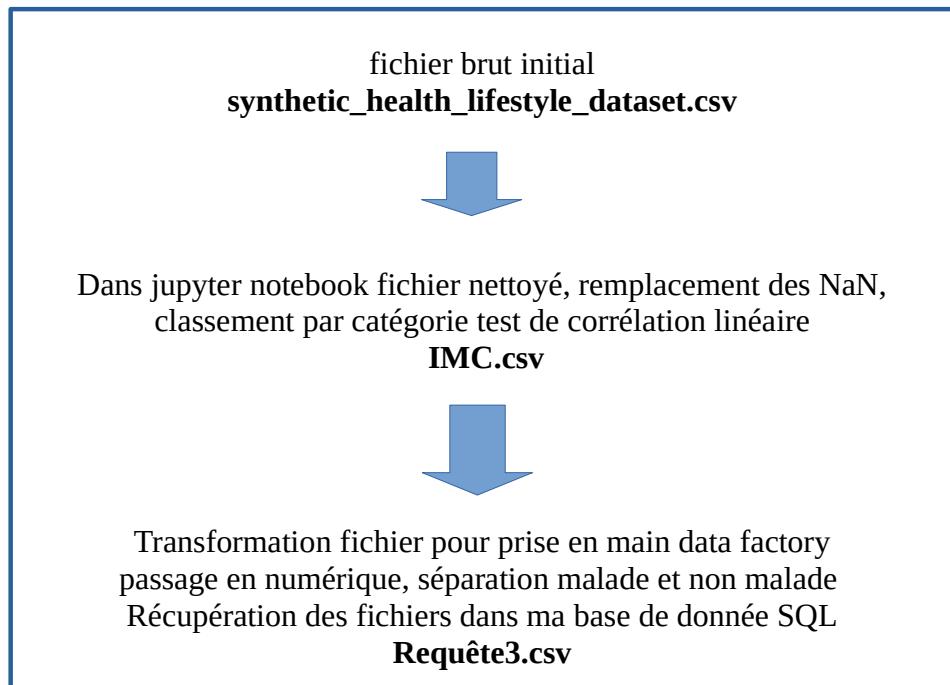
L'IMC n'a pas d'impact sur les catégories de sommeil car les répartitions sont proches il n'y a pas de différence significative.



Le régime alimentaire n'a pas d'impact sur les catégorie de sommeil car il n'y a pas de tendance qui ressortent.

9. Questions de réflexion

1. Comment avez-vous structuré votre démarche ETL dans cet exercice, depuis l'extraction des données des fichiers plats jusqu'au chargement final dans Azure SQL Database ou le Data Warehouse ?



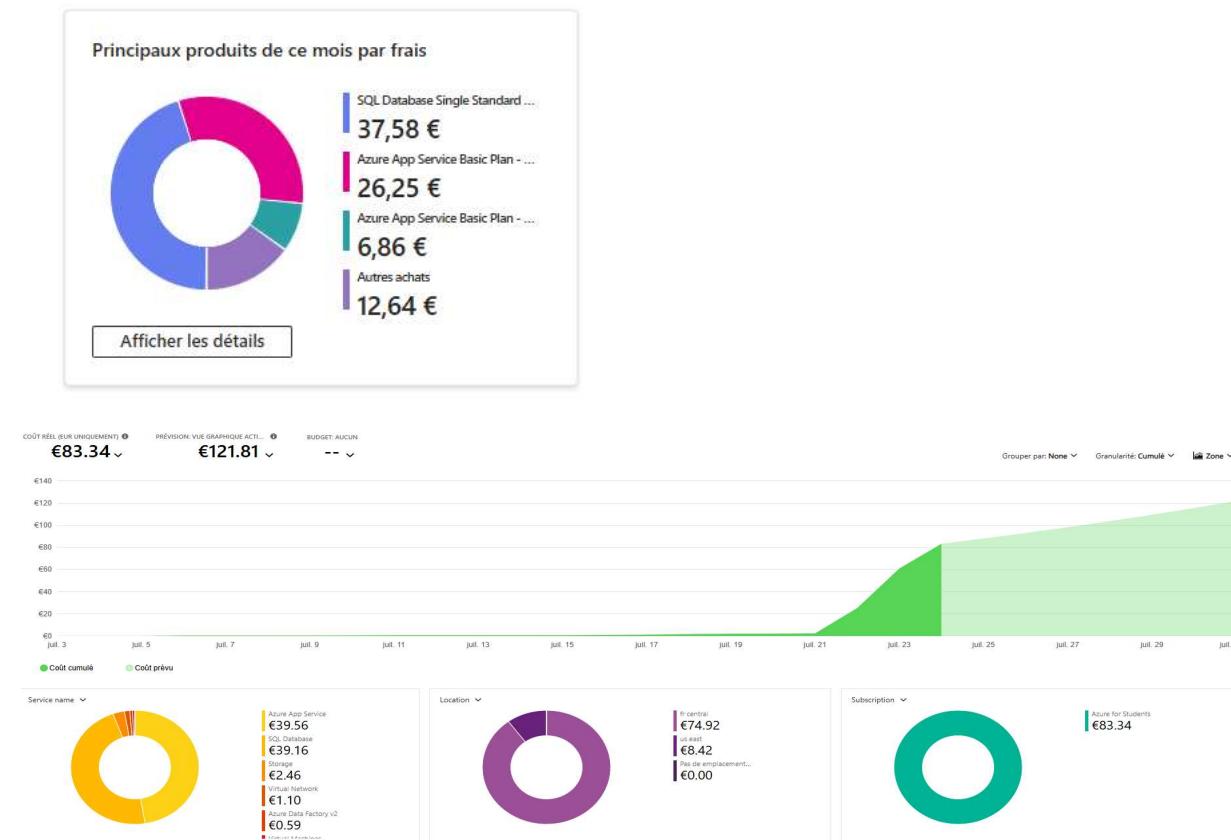
2. Quelle est la différence entre un Blob Storage et une Azure SQL Database en termes d'usage dans les pipelines de données ? Quelles considérations avez-vous prises en compte pour choisir les services de stockage dans ce TP ?

Le Blob Storage est utilisé pour les données non structurées, dans mon cas les données brut traités d'entrée.

Azure SQL Database est utilisé pour des données structurées dans mon cas les données transformées pour être ensuite interrogées via des requêtes SQL ou traitement par powerbi

3. Quels défis pensez-vous rencontrer en déployant un pipeline ADF de cet exercice dans un environnement de production réel ? Comment pourriez-vous vous assurer que le pipeline est robuste, scalable et facile à maintenir ?

Les paramètres à surveiller sont le contrôle des coûts pour maîtriser sont budget et ne pas faire de dépassement pour exemple :



Pour la fiabilité et robustesse,

Dans un premier temps, il est nécessaire de travailler sur un jeu de données propres, nettoyée et transformé correctement, ceci est réalisé avec des étapes de nettoyage des données. Des alertes et logs peuvent être mis en place pour surveiller l'exécution.

Pour la fiabilité qui est la capacité de résistance et de récupération d'une charge de travail face aux défaillances et pannes. Azure offre un certain nombre de fonctionnalités de résilience telles que :

- Les zones de disponibilité.
- La prise en charge multirégion.
- La réPLICATION des données.
- Les fonctionnalités de sauvegarde et de restauration.

Ces fonctionnalités doivent être prises en compte au moment de concevoir une charge de travail afin de répondre à ses impératifs de continuité d'activité.

La scalabilité sera réalisée lorsque l'on constate que les charges de travail atteignent une limite de performances.

Pour un scale-out horizontal le partitionnement de base de données permettra d'ajouter des ressources telles que des machines virtuelles pour gérer les données volumineuses et qui évoluent rapidement.

Les services Hyperscale tels que Microsoft Azure SQL database hyperscale et Azure Database pour PostgreSQL permettent aux utilisateurs de mettre à l'échelle rapidement le stockage jusqu'à 100 To. La mise à l'échelle automatique peut également être envisagée, cela va permettre de mettre en correspondance automatique et dynamique les ressources pour répondre aux besoins de performances du système et ainsi maintenir le système.

4. Dans quelles situations auriez-vous besoin de traiter les données en temps réel plutôt qu'en mode batch ?

J'aurais besoin de traiter des données en temps réel pour les productions industrielles en continu : Par exemple, lors de la fabrication du Keppra (un médicament antiépileptique) par procédé chromatographique continu, un suivi en temps réel est indispensable.

Cela permet de :

- garantir la qualité constante du produit.
- réagir immédiatement en cas d'anomalie.
- respecter les normes GMP (Good Manufacturing Practice) qui exigent une traçabilité des données de production.

En matière de cybersécurité, la détection de fraude nécessite une analyse instantanée des transactions financières :

- Pour bloquer les transactions suspectes avant qu'elles ne soient finalisées,
- Et protéger les comptes contre des activités malveillantes.

Le traitement en temps réel est incontournable dans toutes les situations où :

- Les décisions doivent être prises immédiatement.
- Et les erreurs ou retards peuvent entraîner des conséquences graves (santé, sécurité, perte financière).