



# CLOUD AVANT C

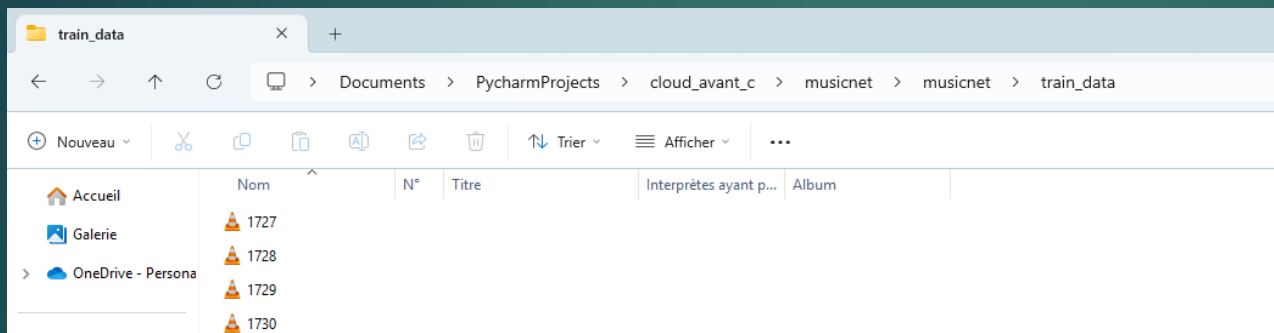
CLASSIFICATION D'UN COMPOSITEUR DE MUSIQUE

# Objectifs

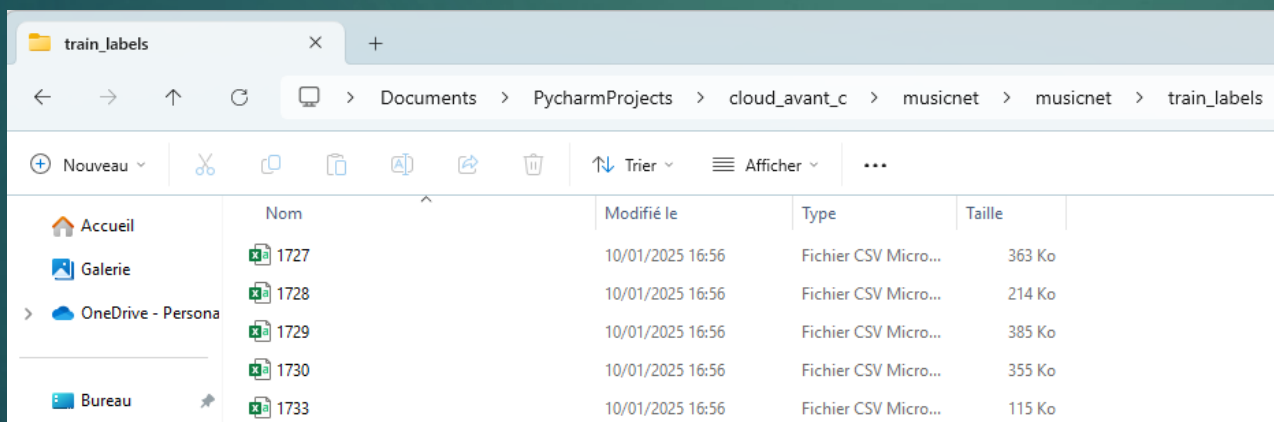
- Données musicales :
  - Fichiers sons
  - Fichiers csv : données sur les musiques
  - Fichiers midis
  - Métadonnées
- **Prédiction du compositeur**

# La donnée (1)

- Fichiers sons

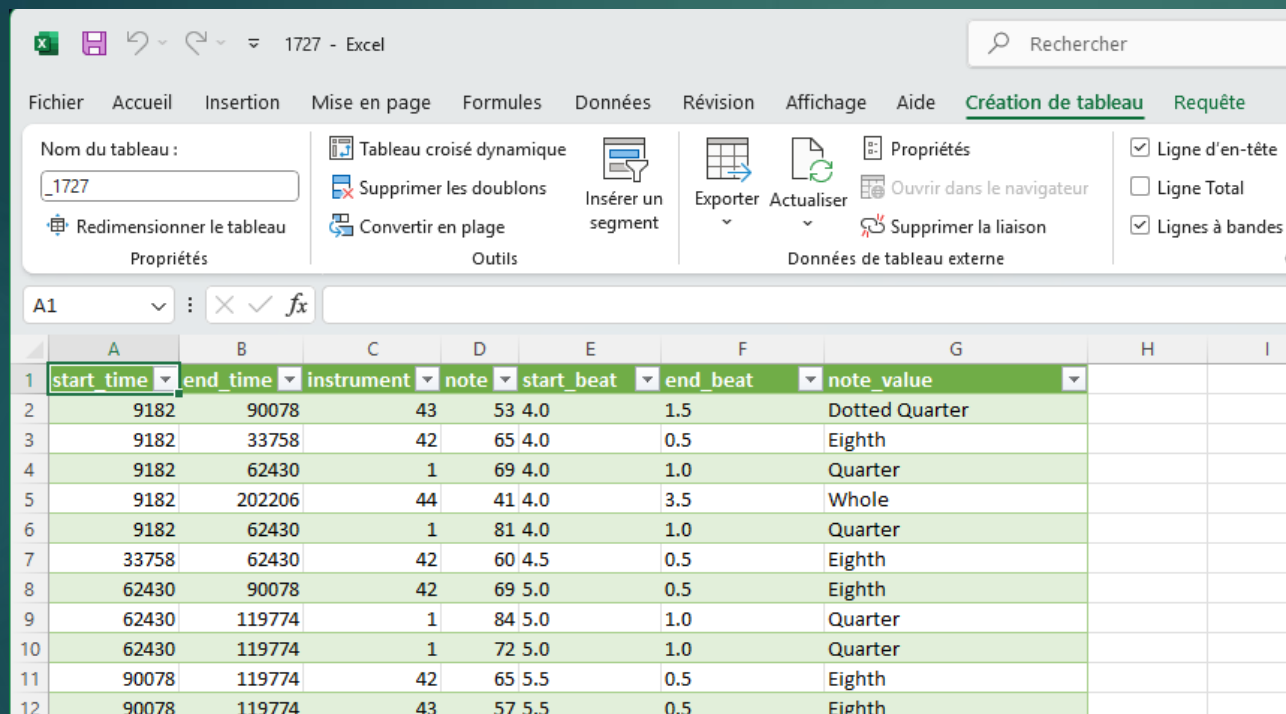


- Fichiers csv



# La donnée (2)

## Fichiers csv



1727 - Excel

Rechercher

Fichier Accueil Insertion Mise en page Formules Données Révision Affichage Aide **Création de tableau** Requête

Nom du tableau : 1727

Redimensionner le tableau

Propriétés

Tableau croisé dynamique

Supprimer les doublons

Convertir en plage

Outils

Insérer un segment

Exporter

Actualiser

Propriétés

Ouvrir dans le navigateur

Supprimer la liaison

Données de tableau externe

☒ Ligne d'en-tête

☐ Ligne Total

☒ Lignes à bandes

A1

	A	B	C	D	E	F	G	H	I
1	start_time	end_time	instrument	note	start_beat	end_beat	note_value		
2	9182	90078	43	53	4.0	1.5	Dotted Quarter		
3	9182	33758	42	65	4.0	0.5	Eighth		
4	9182	62430	1	69	4.0	1.0	Quarter		
5	9182	202206	44	41	4.0	3.5	Whole		
6	9182	62430	1	81	4.0	1.0	Quarter		
7	33758	62430	42	60	4.5	0.5	Eighth		
8	62430	90078	42	69	5.0	0.5	Eighth		
9	62430	119774	1	84	5.0	1.0	Quarter		
10	62430	119774	1	72	5.0	1.0	Quarter		
11	90078	119774	42	65	5.5	0.5	Eighth		
12	90078	119774	43	57	5.5	0.5	Eighth		

# La donnée (2)

- Métadonnées



1727 - Excel

Rechercher

Nom du tableau : musicnet\_metadata

Tableau croisé dynamique

Supprimer les doublons

Redimensionner le tableau

Propriétés

Insérer un segment

Outils

Exporter

Actualiser

Données de tableau externe

Propriétés

Ouvrir dans le navigateur

Supprimer la liaison

Options de style de tableau

Styles de tableau

	A	B	C	D	E	F	G	H	I
1	id	composer	composition	movement	ensemble	source	transcriber	catalog_name	seconds
2	1727	Schubert	Piano Quintet in A major	2. Andante	Piano Quintet	European Archive	<a href="http://tirolmusic.blogspot.com/">http://tirolmusic.blogspot.com/</a>	OP114	447
3	1728	Schubert	Piano Quintet in A major	3. Scherzo: Presto	Piano Quintet	European Archive	<a href="http://tirolmusic.blogspot.com/">http://tirolmusic.blogspot.com/</a>	OP114	251
4	1729	Schubert	Piano Quintet in A major	4. Andantino - Allegretto	Piano Quintet	European Archive	<a href="http://tirolmusic.blogspot.com/">http://tirolmusic.blogspot.com/</a>	OP114	444
5	1730	Schubert	Piano Quintet in A major	5. Allegro giusto	Piano Quintet	European Archive	<a href="http://tirolmusic.blogspot.com/">http://tirolmusic.blogspot.com/</a>	OP114	368
6	1733	Schubert	Piano Sonata in A major	2. Andantino	Solo Piano	Museopen	Segundo G. Yogore	D959	546
7	1734	Schubert	Piano Sonata in A major	3. Scherzo. Allegro vivace	Solo Piano	Museopen	Segundo G. Yogore	D959	325
8	1735	Schubert	Piano Sonata in A major	4. Rondo. Allegretto	Solo Piano	Museopen	Segundo G. Yogore	D959	714
9	1739	Schubert	Piano Trio in B-flat major	4. Rondo. Allegro vivace	Piano Trio	European Archive	harfesoft.de	OP99	490
10	1742	Schubert	String Quintet in C major	2. Adagio	Viola Quintet	European Archive	harfesoft.de	OP163	924
11	1749	Schubert	Piano Sonata in A minor	1. Moderato	Solo Piano	Museopen	Segundo G. Yogore	D845	696
12	1750	Schubert	Piano Sonata in A minor	2. Andante poco mosso	Solo Piano	Museopen	Segundo G. Yogore	D845	784

# Problème

- Manque d'expertise métier sur les données musicales :/
- Conséquences :
  - Exploitation des fichiers csv et métadonnées uniquement
  - Aggrégation de ces données pour rentrer dans un modèle de ML

# Aggrégation des données

- Pour chaque musique :
  - Aggrégations sur les notes :
    - Moyenne, écart type, médiane
    - Moyenne et écart type de la durée des notes
    - Proportion de notes graves et aigues dans la musique
  - Variable catégorielle de "l'ensemble" des instruments dans le morceau
  - Durée du morceau

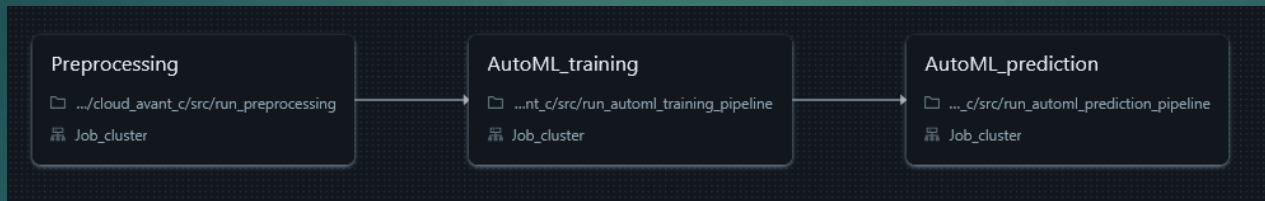
# Mise en production (1)

- Architecture :
  - Préprocessing des données brutes : entraînement et prédiction
    - Aggrégation des données en vue d'un ML
    - Sauvegarde des données dans des tables SQL
  - En parallèle :
    - Data Visualisation
    - Entraînement des modèles
  - Prédictions
    - Résultats sauvés dans une table SQL

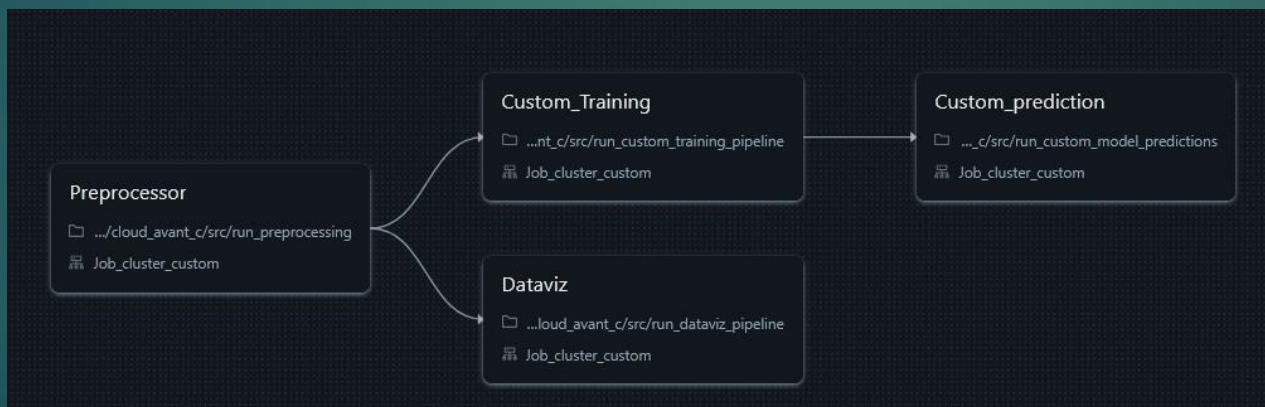


# Mise en production (2)

- 2 Pipelines
  - AutoML

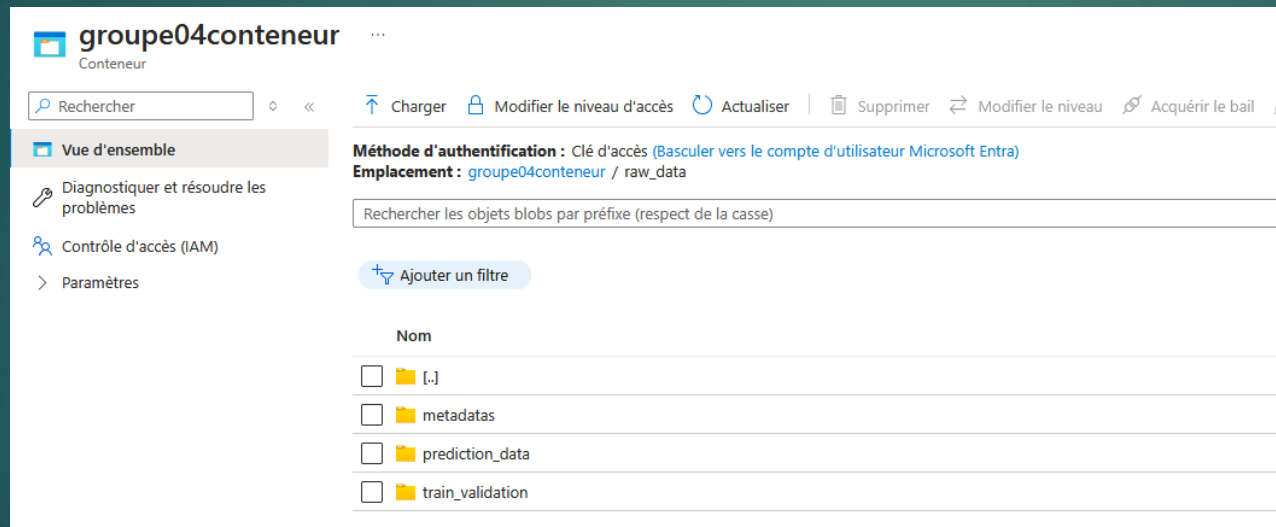


- Modele Custom



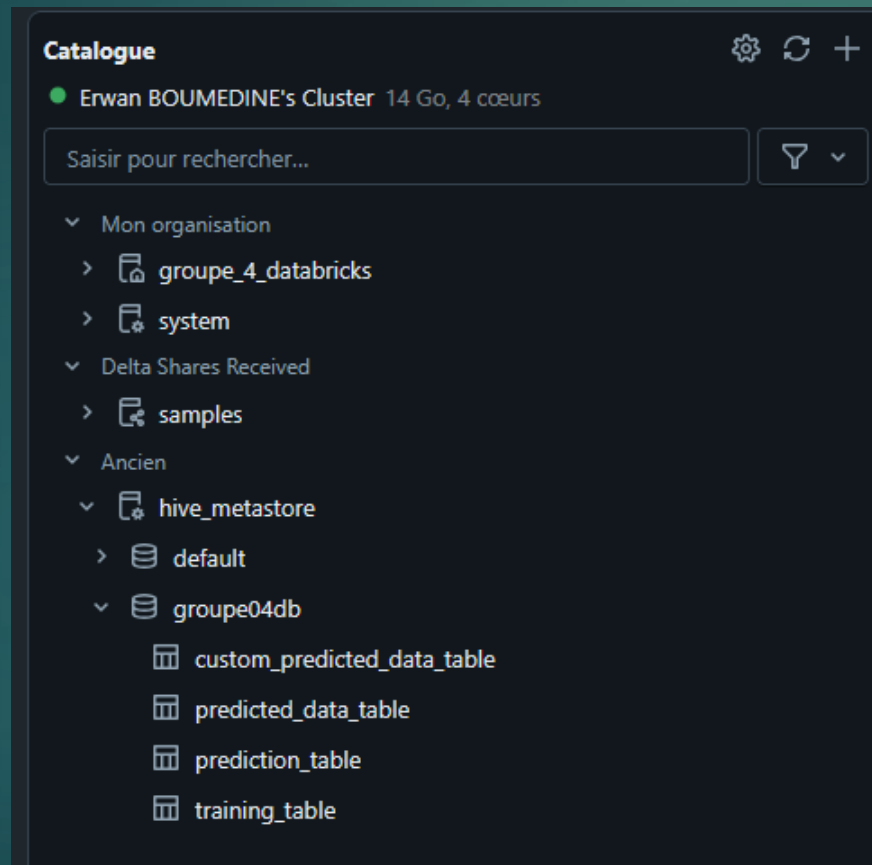
# Mise en production (3)

- Données brutes sauvegardées sur Azure Data Blob Storage
  - Données d'entrainement
  - Metadonnées
  - Données à prédire



# Mise en production (4)

- Données agrégées sauvegardées dans des tables SQL



# Résultats

- ▶ Résultats présentés directement dans Databricks

# Coûts du projet

- Estimation faite :
  - Fourchette basse : 8€
  - Fourchette haute : 31€ si besoin de GPUs

# Coûts du projet

- Coûts totaux :

COÛT RÉEL (EUR UNIQUEMENT) ⓘ

€14.39 ▾

- Dans le détail

Filtrer les éléments		20 lignes				
Ressource	Resource type	Location	Resource group name	Tags		Cost ↑↓
▼ groupe_4_databricks	Service Azure Databricks	fr central	esgi_1_groupe_04	clusterid:0109-115315-xqo33v8o	clustername:erwan bou...	€12.71
Service name		Meter		Cost ↑↓		
Azure Databricks		Premium Serverless SQL DBU		€6.67		
Azure Databricks		Premium All-purpose Compute DBU		€6.02		
Azure Databricks		Premium Jobs Compute DBU		€0.02		
> groupe4comptedestockage	Compte de stockage	fr central	esgi_1_groupe_04	--		€0.37
> groupe04keyvault	Coffre de clés	fr central	esgi_1_groupe_04	--		

Merci pour votre écoute