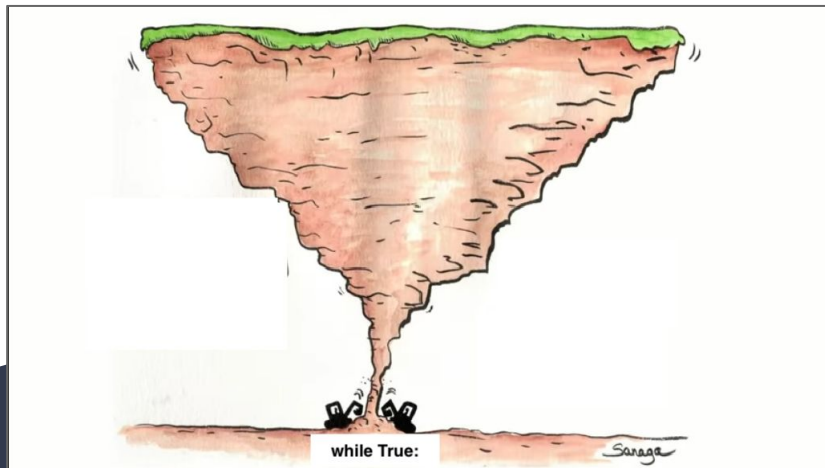


# DEMO PROJET

Système de recommandation de films pour un cinéma de la Creuse



Creuser la data sans se creuser la tête !

# ORDRE DU JOUR :



## INTRODUCTION

→ Contexte et objectifs

### 1. PRÉSENTATION DU PROJET

- L'équipe
- Le schéma de principe
- Fonctionnalités et organisation
- Nos sprints :
  - A) Nettoyage de la donnée
  - B) Tableau de bord
  - C) Système de recommandation
- Retour d'expérience

### 2. DÉMONSTRATION

- Tableau de bord
- Système de recommandation

## CONCLUSION

→ Axes d'amélioration

# INTRODUCTION

→ Générique

Contexte :

- Cinéma en perte de vitesse
- Projet de digitalisation



Objectifs :

- Analyse des données
- Module d'algorithme de recommandation



# 1. L'ÉQUIPE

→ Casting



**Alexandre C.**  
Scrum master



**Marion F.**  
Product owner



**Mohammed B.**  
Developer

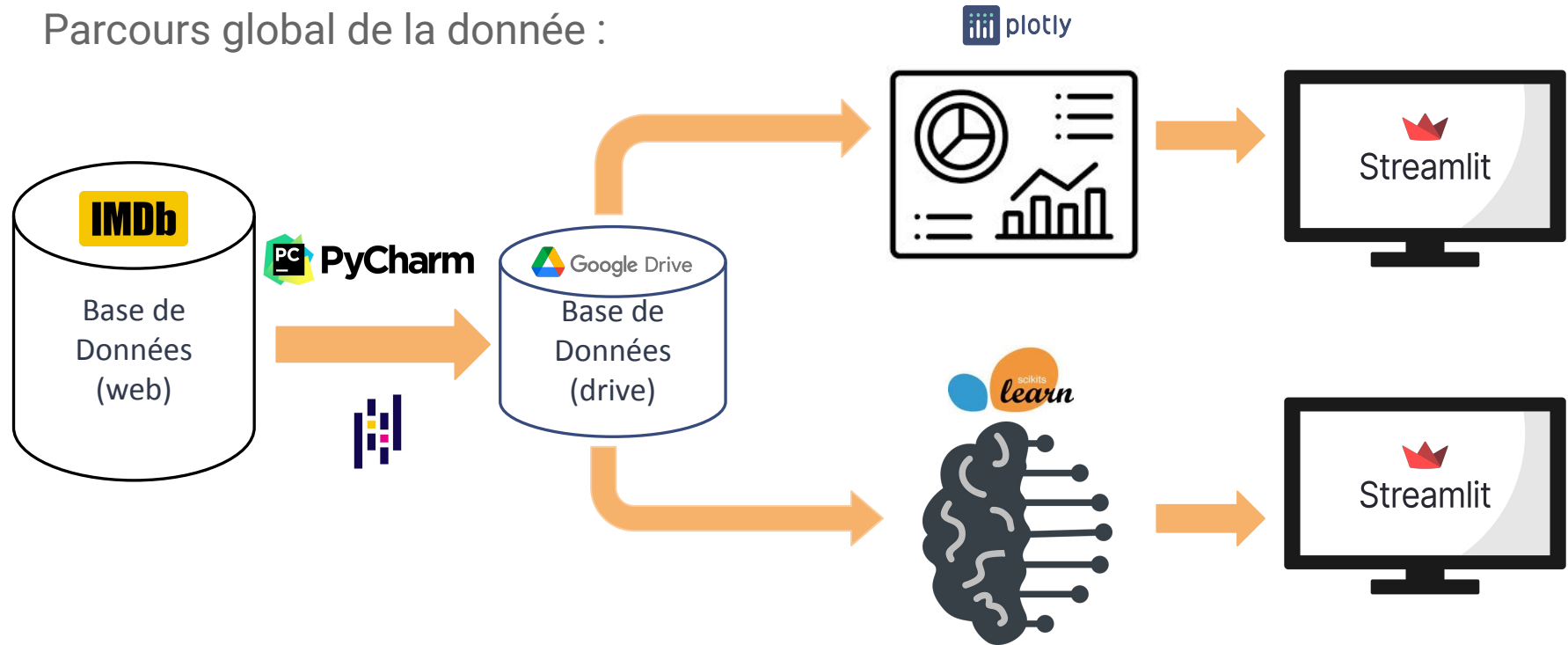


**Aurélien G.**  
Developer

## 2. SCHEMA DE PRINCIPE





→ Storyboard

Parcours global de la donnée :



# 3. FONCTIONNALITÉS

→ Synopsis

	Complexité
	Axes d'analyse pertinents +
	Expérience utilisateur simple et intuitive +
	Exploration de données (data-mining) ++
	Liste de suggestions sur la base d'un film +++

# 4. L'ORGANISATION

→ Chapitres du film



Nettoyage





- 1 module fonctions
- 1 module constantes
- 1 module par table à nettoyer



KPI et  
visualisation





- 1 module tableaux de bord  Streamlit
- 1 module par axe KPI  plotly
- 1 module dataframes nettoyés



Recommandation



- 1 module recommandation  Streamlit
- 1 module machine learning  learn
- 1 module dataframe machine learning

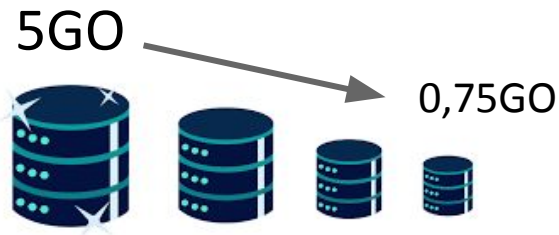


## 5. SPRINT - ACTE 1 : Nettoyage de la donnée

→ Stratégie et résultat :

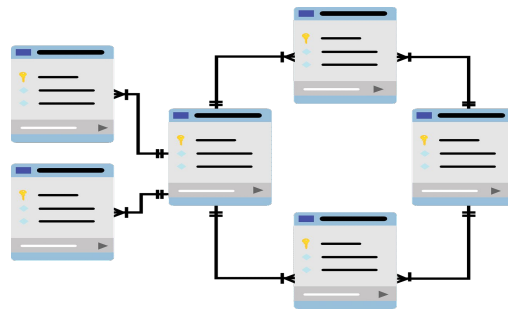
Création d'une table centrale nettoyée (Title basics)

- Sélection du type "film"
- Suppression la catégorie adulte
- Filtre sur la durée (> 40 minutes)
- Filtre sur le troisième quartile (389 votes)



Application sur les autres tables :

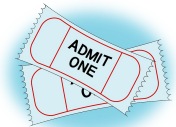

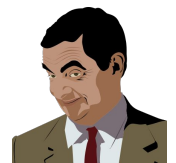

- Fonction de filtrage
- Modification d'architecture générale :
  - Suppression ou création des tables
  - Modélisation de la base de données





## 5. SPRINT - ACTE 2 : Construction d'un tableau de bord

### → Scène 1 : Définition des KPI

	Axe	Analyse
	Nombre / durée	<div data-bbox="937 467 1812 644"><u>KPI généraux</u> Nombre de films, durée moyenne, ...</div>
	Genre	<div data-bbox="981 644 1760 821"><u>KPI techniques</u> Note en fonction du nombre de votes, Note en fonction de la durée, ...</div>
	Casting	<div data-bbox="1025 821 1707 999"><u>Top 10 / Top 20 (dynamiques)</u> Films, acteurs, actrices, ...</div>
	Equipe du film	

## 5. SPRINT - ACTE 3 : Système de recommandation de films

### → Scène 1 : Méthode

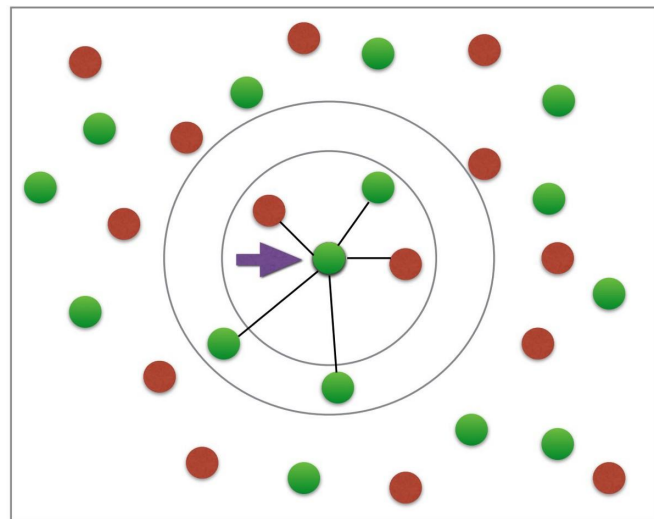
**Objectif :** Proposer des films pertinents et tendances

**Entrée :** 1 film

**Sortie :** 5 films

**Méthode :**

- Machine Learning de classification
- Méthode de recherche des plus proches voisins
- Librairie : Scikit-Learn, NearestNeighbors



# 5. SPRINT - ACTE 3 : Système de recommandation de films

## → Scène 2 : L'algorithme

### Paramètres :

- Année de sortie
- Genres

### Préparation des paramètres :

#### Filtres :

- Nombre de votes
- Notation

#### Standardisation

Entrainement du modèle prédictif

Recherche des plus proches voisins

	startYear	Action	Adventure	Animation	Biography	Comedy	Crime	Documentary	Drama	Family	Fantasy	Film-Noir	History	Horror	Music	Musical	Mystery
0	1920	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1
1	1920	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
2	1921	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1
3	1920	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
0	-2.75086	-0.43415	-0.37570	-0.19990	-0.24293	-0.67539	-0.45322	-0.20568	-1.22400	-0.21372	-0.23680	-0.12685	-0.21533	3.20540	-0.17679	-0.15099	3.49711
1	-2.75086	-0.43415	-0.37570	-0.19990	4.11644	-0.67539	-0.45322	-0.20568	0.81699	-0.21372	-0.23680	-0.12685	-0.21533	-0.31197	-0.17679	-0.15099	-0.28595
2	-2.71226	-0.43415	2.66166	-0.19990	-0.24293	-0.67539	-0.45322	-0.20568	-1.22400	-0.21372	4.22295	-0.12685	-0.21533	-0.31197	-0.17679	-0.15099	3.49711
3	-2.75086	-0.43415	-0.37570	-0.19990	-0.24293	-0.67539	-0.45322	-0.20568	0.81699	-0.21372	-0.23680	-0.12685	-0.21533	-0.31197	-0.17679	-0.15099	-0.28595

## 5. RETOUR D'EXPÉRIENCE

→ Enjeux et stratégies

**BDD énorme**



**Choix des données pertinentes  
Filtrage et nettoyage**

**Temps d'exécution**



**Séparation de la BDD  
Travailler en local  
Export de la BDD cible en csv**

**Organisation du code**



**Mise en place d'une structure  
Google Drive**

# CONCLUSION

→ Et ils vécurent heureux... mais il y aura une suite !

## AXES D'AMÉLIORATIONS :

- **Amélioration du modèle de machine learning**
- Intégration de nouvelles variables
- Regrouper les genres en "clusters de sentiments"
- **Recommandation à partir de plusieurs films**
- "Webscraper" les affiches des films
- Choisir les critères de recommandation
- **Optimisation du code**

