# Revitalizing Rural Cinema: A Data-Driven Recommendation Engine

## Market Analysis & Technical Strategy

Presenter: Anna-Louise A.

Date: December 2025

# The Mission

## Context

**The Client:**
A local cinema in La Creuse, Nouvelle-Aquitaine.

**The Problem:**
Declining attendance and a need for digital transformation.

**The Goal:**
1. Analyze the local market (Cold Start).
2. Build a "testable" movie recommendation engine.
3. Create a dashboard for decision-making.

**Key Insight:** We are not building Netflix; we are building a tool for a specific, aging, rural demographic.

# Market Research - The “Creuse” Context

## Demographics

**Underpopulated** region (115,529 inhabitants).

**Aging population:** Over 45s make up >50% of the populace.

**Primary audience:** Retirees and families living in couples.

## Cinema Landscape


**Small ecosystem:** Only 9 cinemas in the region; absence of large franchises.

**Current Status:** Low attendance indices, though post-COVID numbers are rising.

# User Preferences: The “Why” behind the Data

## Content is King

**French Films**: Perform exceptionally well locally.


**Language Barrier**: **70% preference** for “Version Française” (VF).

**Top Genres:** Strong preference for **Drama** and **Comedy** (specifically French humor).

## Behavior

Cinema is viewed as a **social activity** “good for morale” (9 French out of 10), driven by trailers and internet searches.

# From Insights to Strategy

## The “Cold Start” Strategy

Since we have no user history, we use **Content-Based Filtering** tuned to our Market Research.

## Data Cleaning Rules

**Language Weight:** Prioritize films with **original_language = 'fr'** or major blockbusters likely to be dubbed.

**Genre Focus:** Boost weights for **'Drama'** and **'Comedy'** in the vectorization process.

**Era Sensitivity:** Do not aggressively penalize older films (**1980-2000**) given the 45+ demographic.

# Technical Architecture

## Data Source:

**IMDb**
(7M+ rows)

**TMDB**
(Posters/Budgets)

## ETL Pipeline:

**Chunking**
Chunking massive datasets to manage memory.

**Filtering**
Filtering non-relevant content (shorts, adult, unreleased).

## The Engine:

**Library:** Scikit-Learn
(NearestNeighbors or CosineSimilarity).

**Features:** Genres, Actors, Directors, Keywords.

## The App:

**Streamlit**
(Recommendation, Dashboard)

# Roadmap & Next Steps

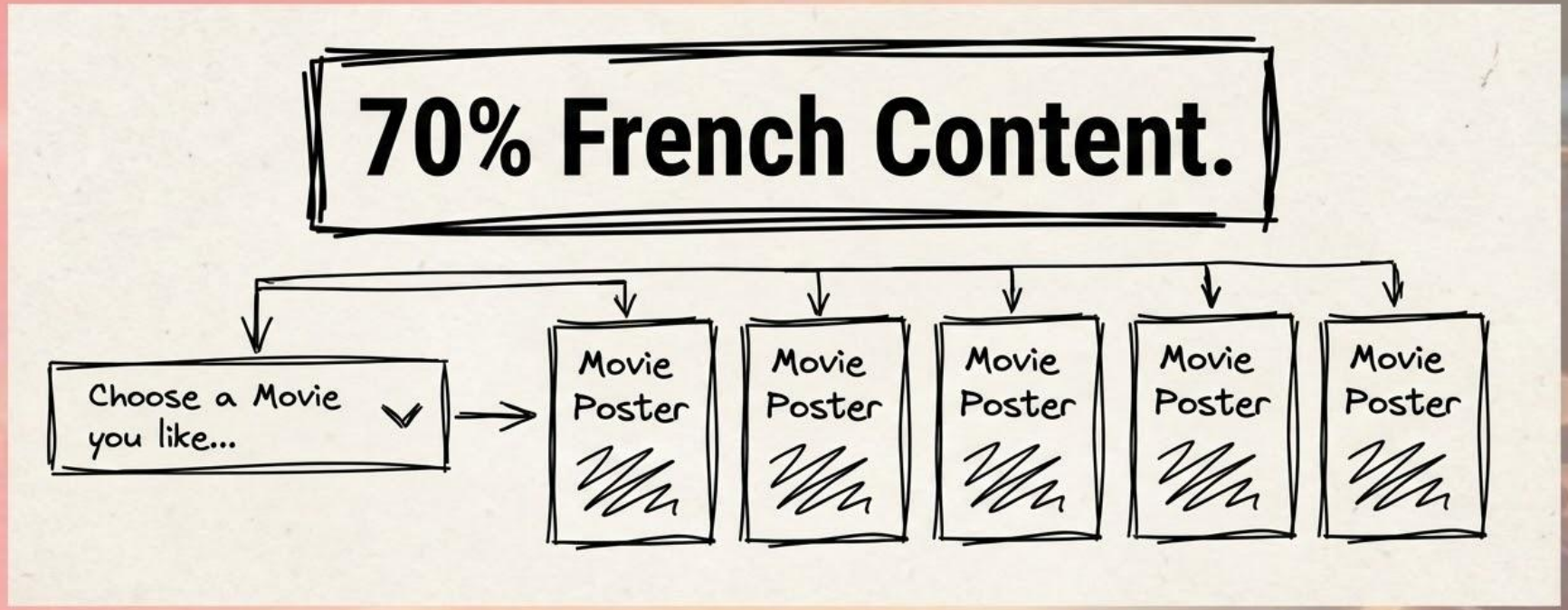
**Phase 1 (Done):** Market Research & User Persona Definition.

**Phase 2 (Current):** Data Cleaning & Feature Engineering (Pandas).

**Phase 3:** Model Training & Streamlit Interface Development.

**Phase 4:** Client Demo (using *Anatomy of a Fall* as the calibration test).

# The Interface Concept



# Thank you