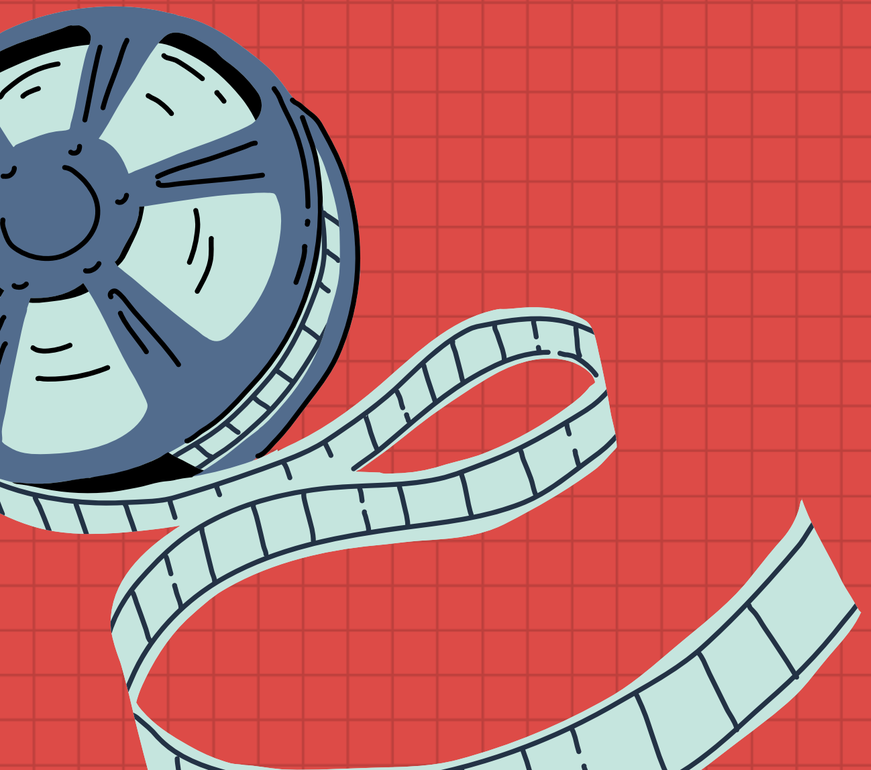
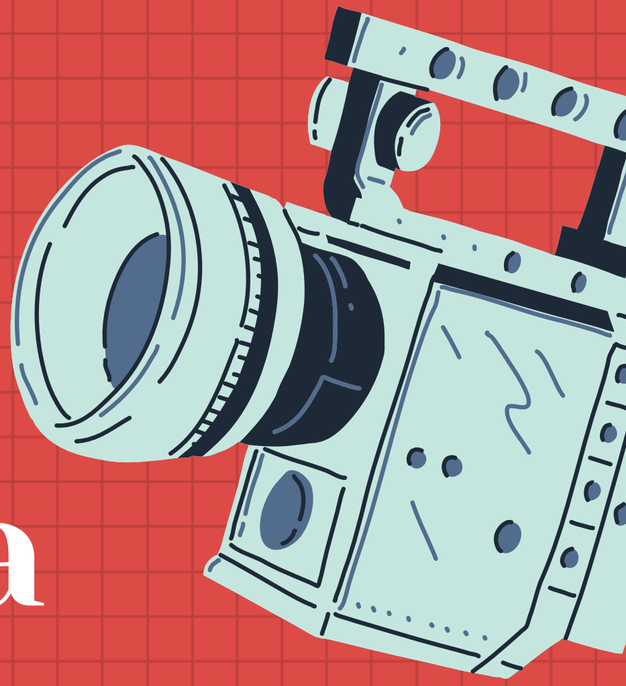
Data Science Capstone

I want to watch a movie...

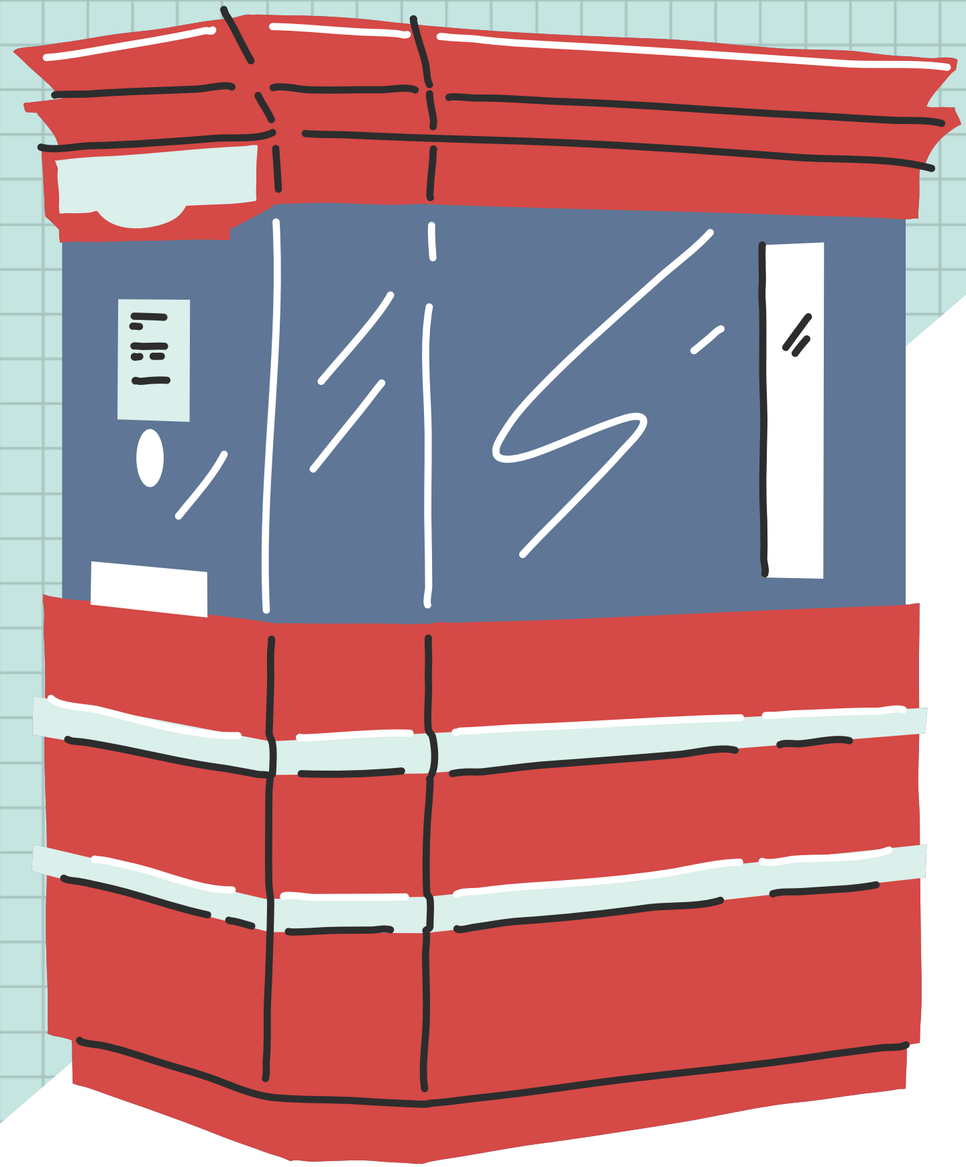
Sprint 1

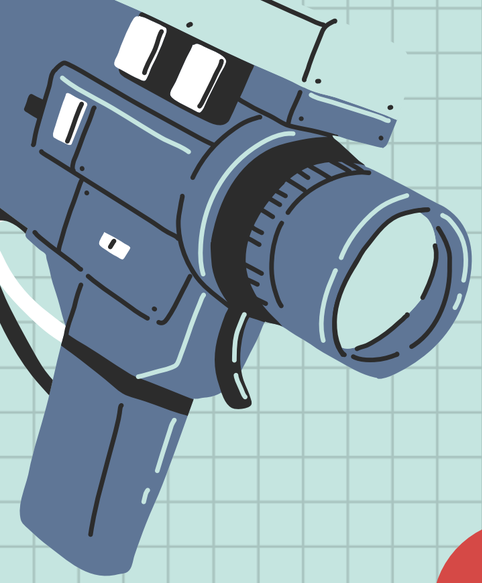


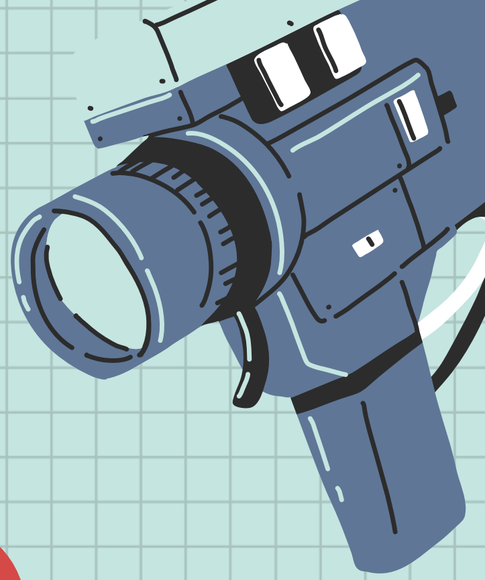


A Movie Reccomender System

1. The Problem
2. The Opportunity
3. The Impact







The Approach

1

The Data

- 3 Tables: Movies, Users, and Ratings
- Over 11 **million ratings**

2

Machine Learning

Use ML models to identify trends and similarities.

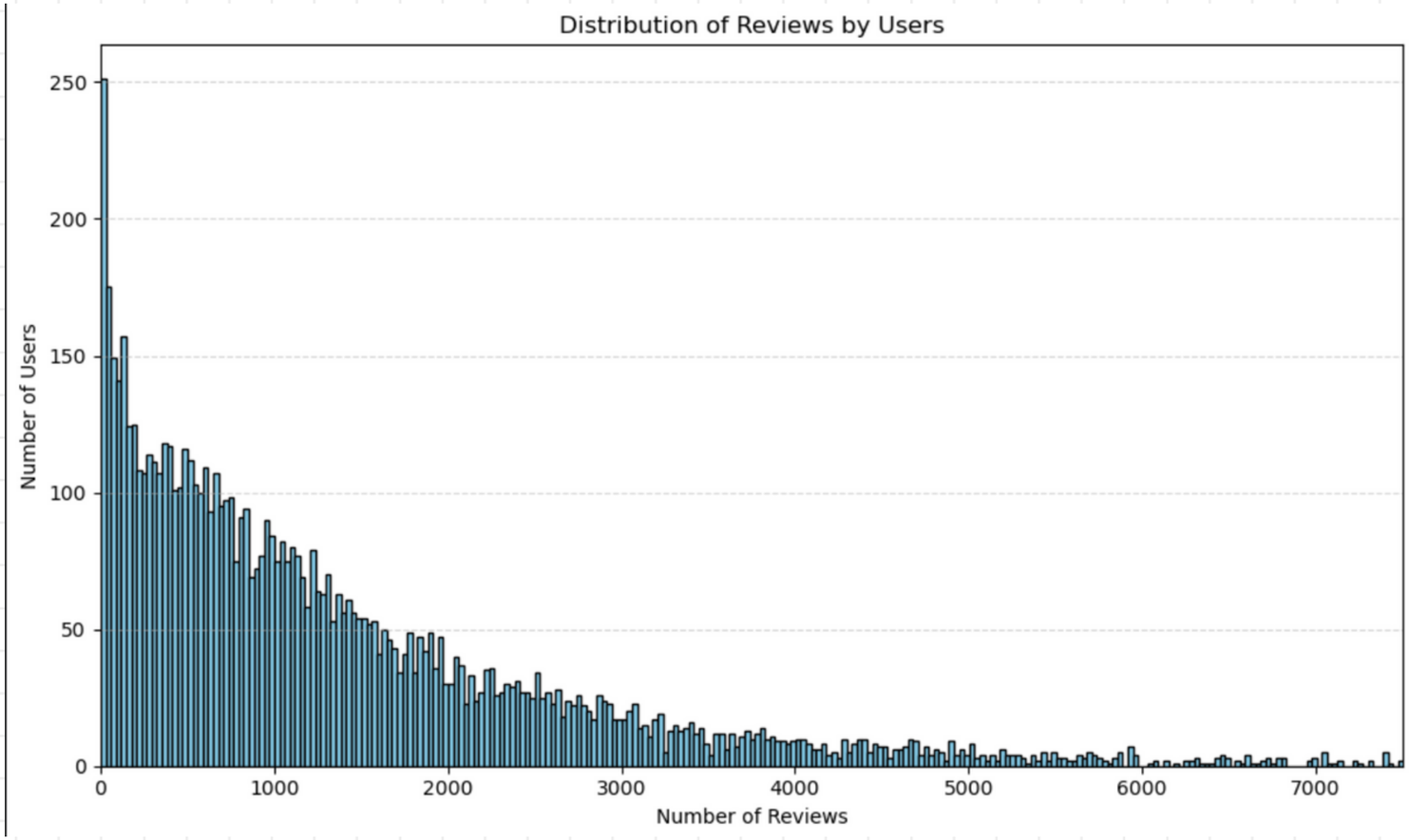
3

Make it Personal

- Scrape Letterboxd data
- Leverage the value add

Deep Dive on Impact

- Leverage Viewership
- Save Time
- Improved enjoyment
- RELAX



Starring: Data

Overview:

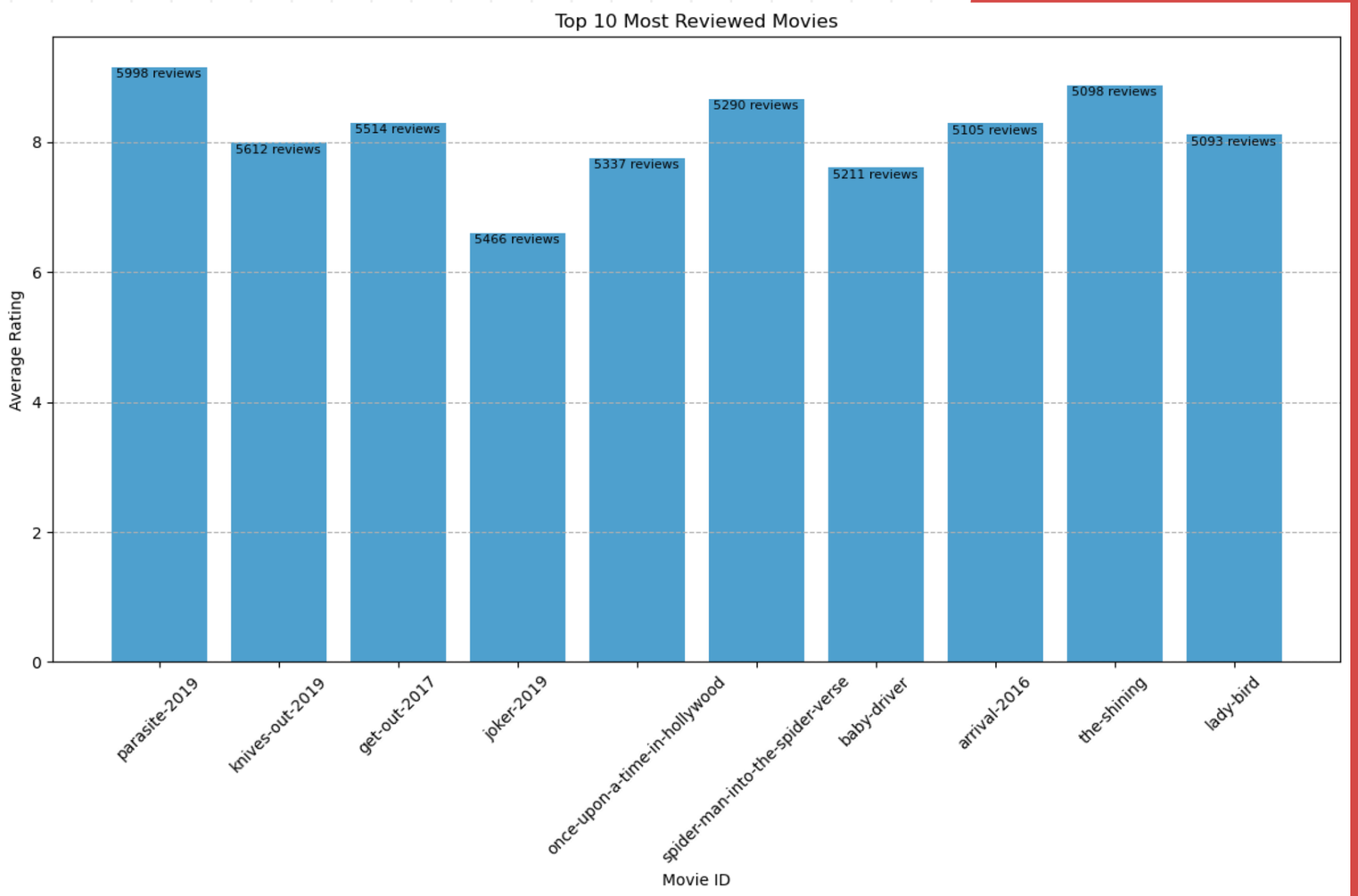
- 11 million plus observations
- 3 files
- 28 Columns

Pros:

- Room to clean
- Personalization opportunity

Cons:

- Do we have characteristics that will be strong indicators?
- Too big?



Next Steps

- Clean & Trim
- Pre-process
- Continually Simplify
- Look ahead
- Scrape data