



NETFLIX CONTENT ANALYSIS

Team 5: Aryan Kumar, Jenil Shah,
Ruiqi (Freya) Jiang, Sneha Ekka

LET'S START!





CONTENTS

3



ML & ML OPS

4



CHALLENGES

1



PHASE 1 RECAP

5



FUTURE SCOPE

2



IMPROVEMENTS

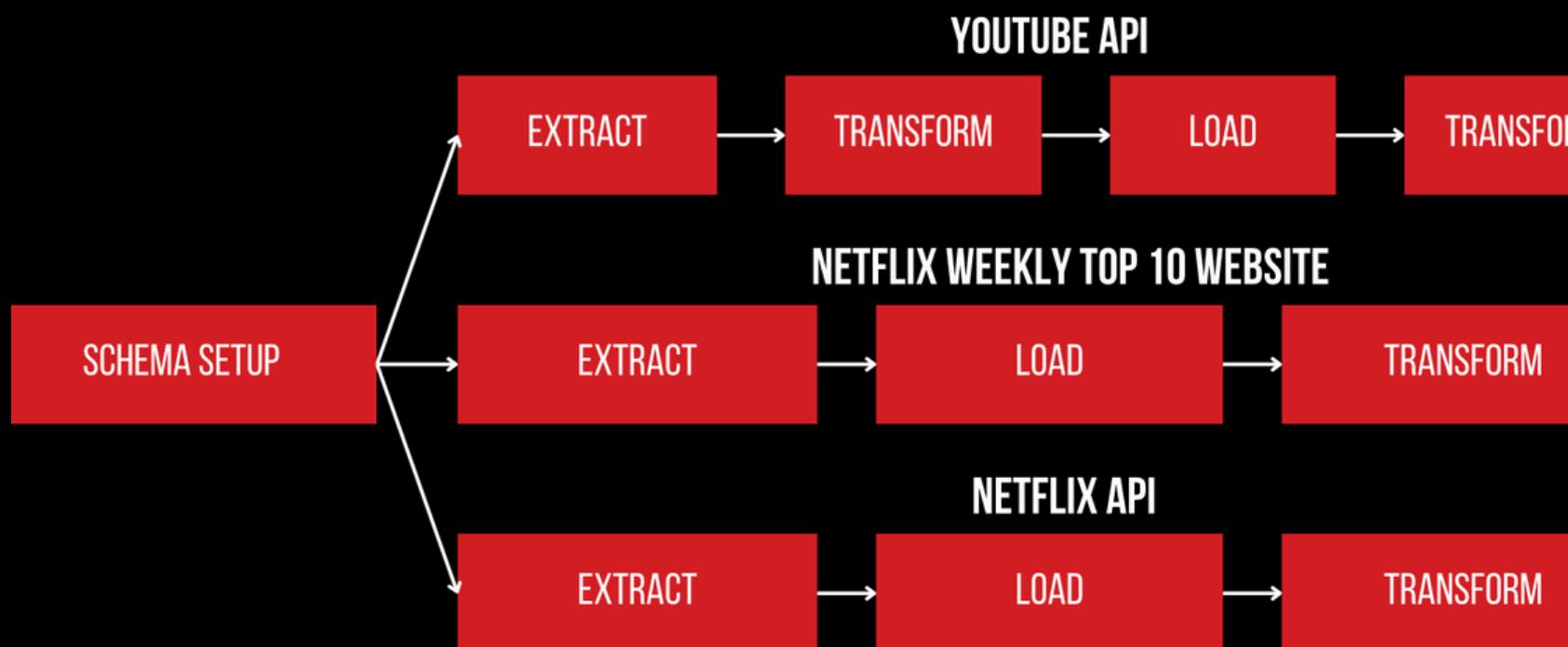
A large, solid red right-pointing triangle, resembling a play button on a media player, is positioned in the center of the image. It overlaps the text "PHASE 1 RECAP".

PHASE 1 RECAP

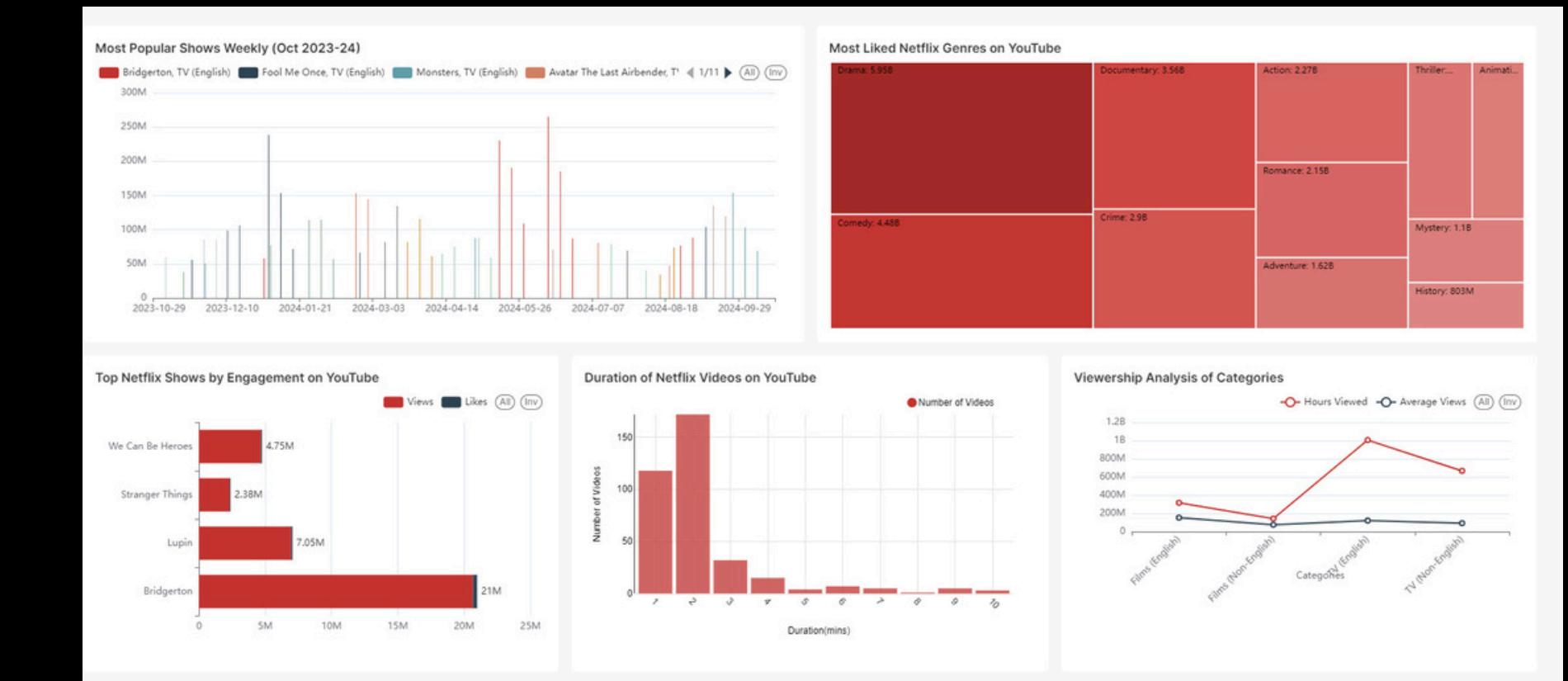


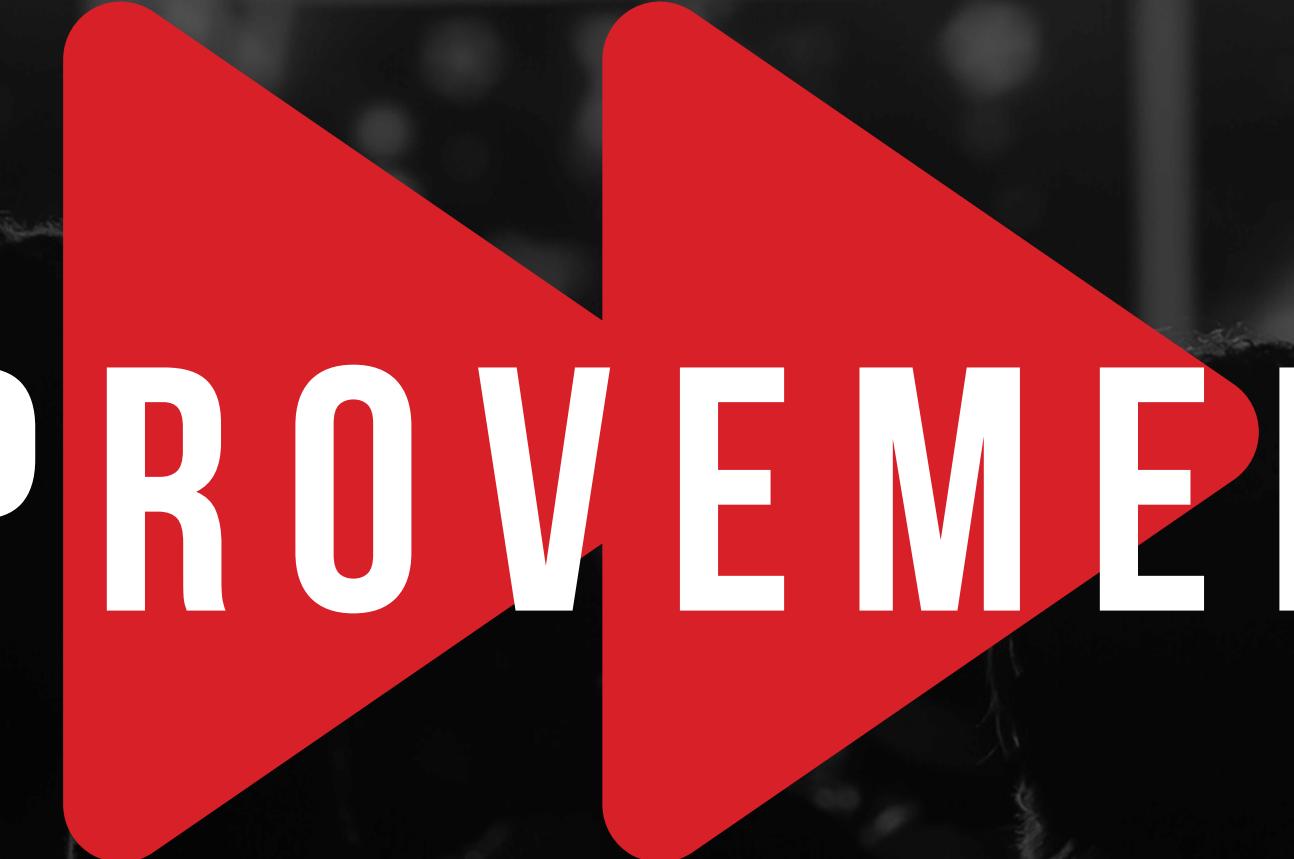
WHAT WE ACHIEVED...

DATA PIPELINE (GCP)



REPORTING (SUPERSET)



A large, semi-transparent red movie play button graphic is positioned in the center of the image. It features a white right-pointing triangle at its base. Overlaid on this triangle is the word "IMPROVEMENTS" in a bold, white, sans-serif font.

IMPROVEMENTS



WHAT WE CHANGED...

1



EtLT to ETL
(based on
changed data
requirements)

2



INSERT Logic
(faulty initial
upsert
process)

3



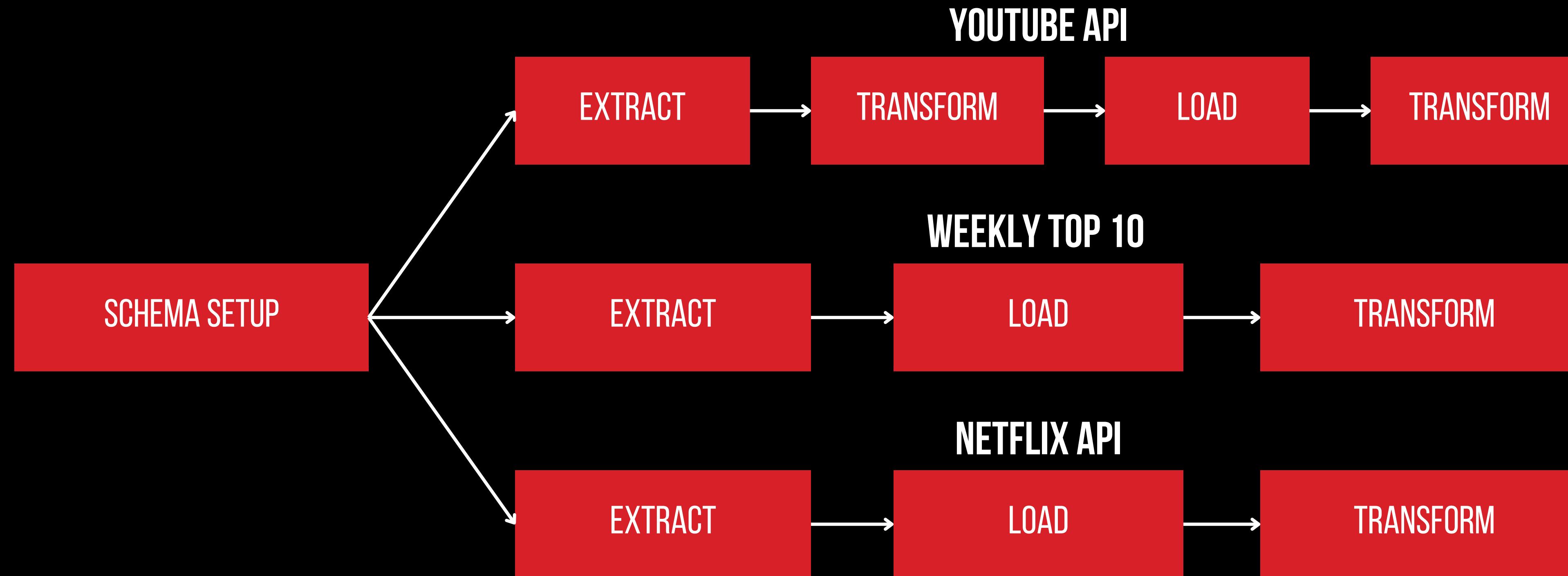
**Prefect
Flow Runs**
(deployed
flow was
crashing)

A large, semi-transparent red movie play button graphic is positioned in the center of the image, partially overlapping the text.

MACHINE LEARNING

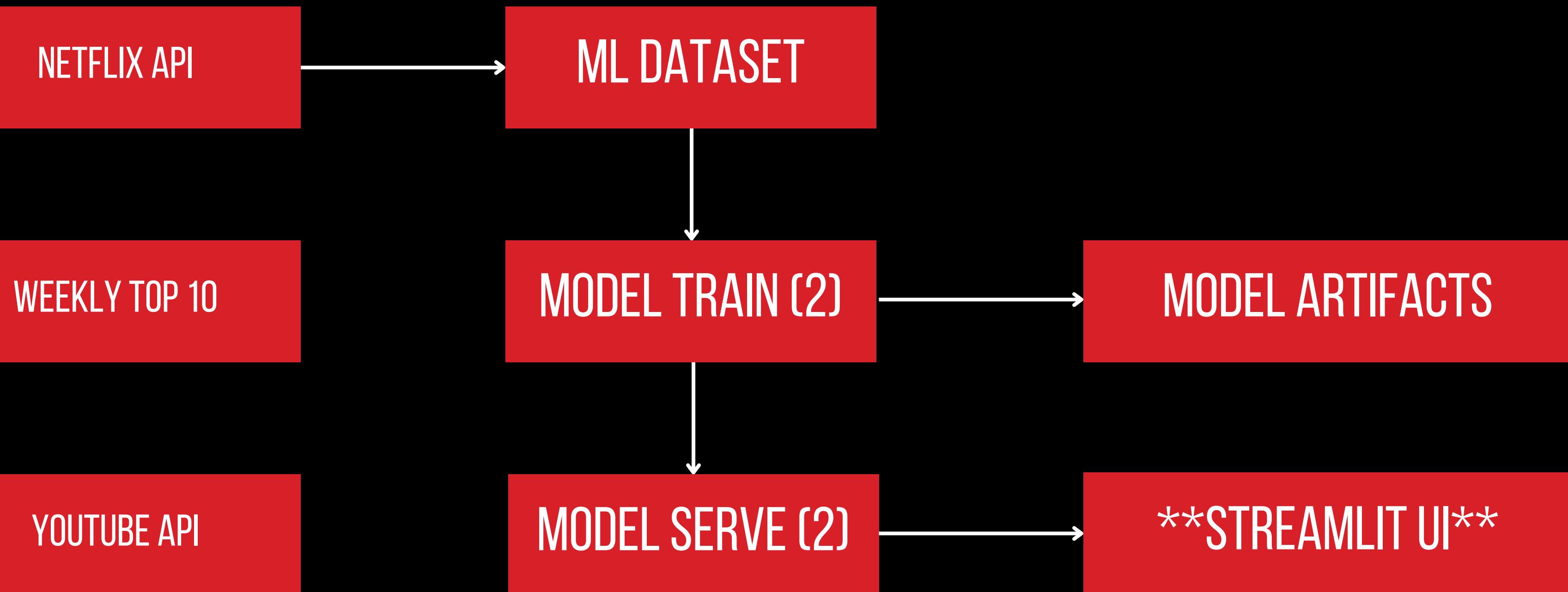


EXTENDING THE DATA PIPELINE ...



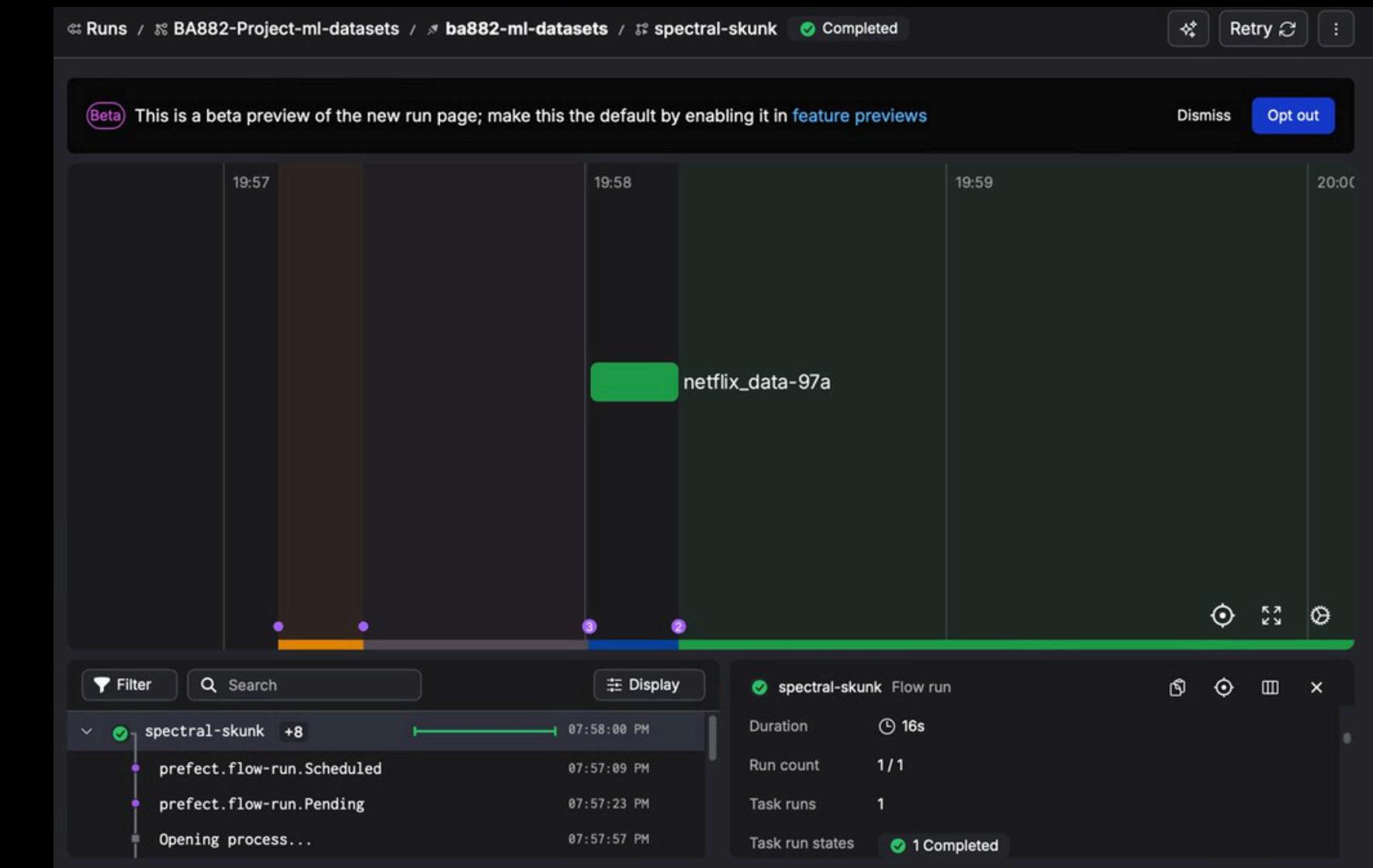
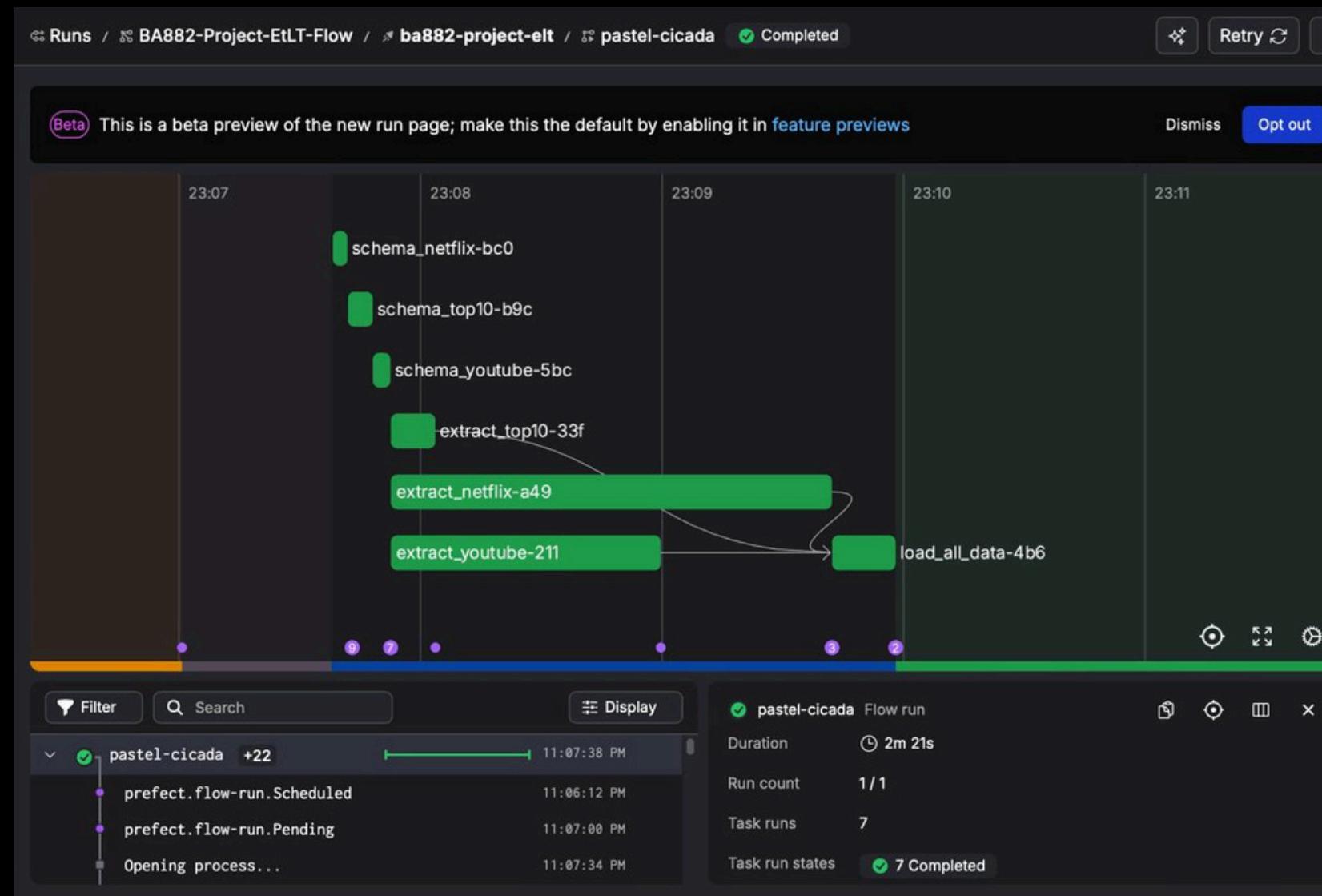


... TO MACHINE LEARNING

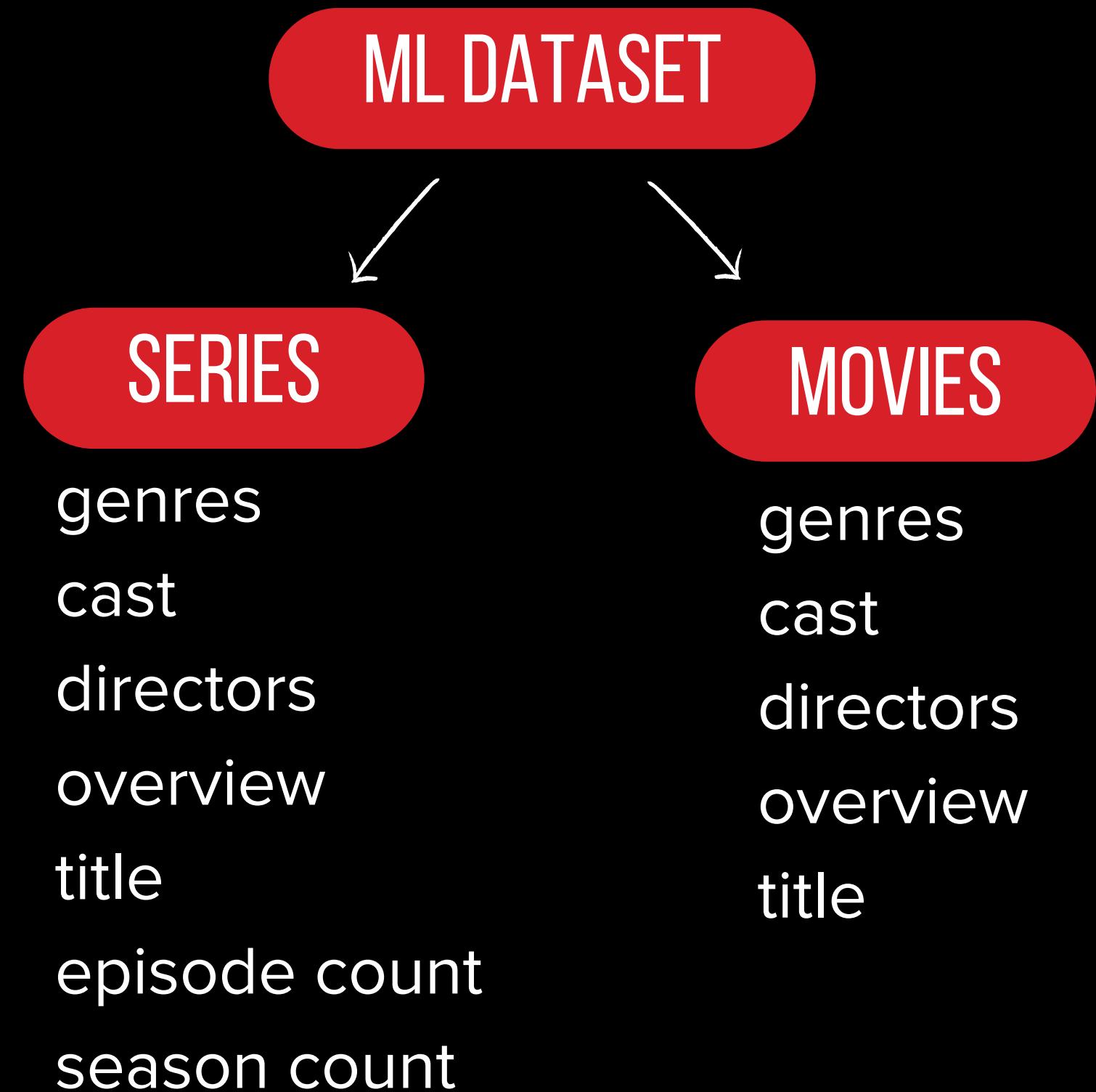




PREFECT FLOW & ORCHESTRATION



NETFLIX RECOMMENDATION MODEL



1

KNN

Used KNN separately for movies and series

2

Dataset Size

Dataset small hence observed overfitting

3

Additional Data

Better for machine learning tasks and deeper insights

NETFLIX RECOMMENDATION MODEL DEMO

The screenshot shows the Google Cloud Platform Cloud Run function details page for a function named "ml-movies-serve". The function was deployed at Nov 12, 2024, 2:05:27AM. The URL is https://us-central1-ba882-inclass-project.cloudfunctions.net/ml-movies-serve. The page includes tabs for METRICS, DETAILS, SOURCE, VARIABLES, TRIGGER, PERMISSIONS, LOGS, and TESTING. The TESTING tab is active, showing a code editor with a JSON input payload and a "TEST THE FUNCTION" button. The Output section displays a JSON response with recommendations. The Logs section shows a list of log entries from November 12, 2024.

ml-movies-serve Cloud Run function (Deployed at Nov 12, 2024, 2:05:27AM) URL: https://us-central1-ba882-inclass-project.cloudfunctions.net/ml-movies-serve

METRICS DETAILS SOURCE VARIABLES TRIGGER PERMISSIONS LOGS TESTING

Configure triggering event

```
Press Option+F1 for Accessibility Options.
1 {
2   "data": [
3     {
4       "genres": "[{"id": "documentary", "name": "Documentary"}]",
5       "cast": "[Robert Downey Sr., Robert Downey Jr., Chris Smith, Alan Arkin, Sean",
6       "directors": "[Chris Smith]",
7       "overview": "Named Best Documentary by the National Board of Review, this fil",
8       "title": "\"Sr.\""
9     },
10    {
11      "genres": "[{"id": "drama", "name": "Drama"}, {"id": "thriller", "name": "Thr",
12      "cast": "[Eyad Nassar, Ghada Adel, Injy Al Moqadem, Omar El Shenawy, Mohamed",
13      "directors": "[Karim Abouzeid]",
14      "overview": "When a troubled lawyer attempts suicide in his car, the stray bu
```

TEST THE FUNCTION

Query parameters

+ ADD QUERY PARAMETER

Headers

+ ADD HEADER

Output

```
$ {"recommendations": [{"\"Sr.\)": [{"title": "The Man from Toronto", "similarity": 0.85}, {"title": "The King's Speech", "similarity": 0.82}, {"title": "The Hurt Locker", "similarity": 0.78}], "status": "Success"}]
```

Logs Fetched (up to 100 entries). View all logs

SEVERITY	TIMESTAMP	SUMMARY
> i	2024-11-12 14:18:43.017 EST	POST 200 3.73 KB 6 s python-requests/2.32.3 https://us-c...
> *	2024-11-12 14:18:48.627 EST	Loaded the KNN model from GCS
> *	2024-11-12 14:18:48.867 EST	Loaded the vectorizer from GCS
> *	2024-11-12 14:18:49.011 EST	Loaded movie metadata from GCS
> i	2024-11-12 14:18:49.109 EST	Default STARTUP TCP probe succeeded after 1 attempt for conta...
> *	2024-11-12 14:18:49.200 EST	Received data: [{"title": "A Fall from Grace"}]
> i	2024-11-12 14:25:55.483 EST	POST 200 3.73 KB 3.7 s python-requests/2.32.3 https://us...
> *	2024-11-12 14:25:58.755 EST	Loaded the KNN model from GCS
> *	2024-11-12 14:25:58.941 EST	Loaded the vectorizer from GCS
> *	2024-11-12 14:25:59.067 EST	Loaded movie metadata from GCS
> i	2024-11-12 14:25:59.115 EST	Default STARTUP TCP probe succeeded after 1 attempt for conta...
> *	2024-11-12 14:25:59.209 EST	Received data: [{"title": "#FriendButMarried 2"}]
> i	2024-11-12 15:04:43.273 EST	POST 200 8.94 KB 5.4 s Google-Cloud-Functions https://us...
> *	2024-11-12 15:04:48.214 EST	Loaded the KNN model from GCS
> *	2024-11-12 15:04:48.471 EST	Loaded the vectorizer from GCS
> *	2024-11-12 15:04:48.633 EST	Loaded movie metadata from GCS
> i	2024-11-12 15:04:48.739 EST	Default STARTUP TCP probe succeeded after 1 attempt for conta...

A large, solid red right-pointing triangle, resembling a play button, is positioned in the center of the image. It contains the text "SEE WHAT'S NEXT" in white, sans-serif capital letters.

SEE WHAT'S NEXT



LIMITATIONS

- Very small dataset for recommendations. Added data for 2020 and 2021 and reached from 2000 to 3500 rows
- Absence of User data. Things such as preferences, user ratings and viewer history could potentially enhance recommendations
- Used only Netflix API and did not use YouTube API for this which limited our recommendations capability





FUTURE SCOPE

- Plan to do sentiment analysis of overview and comments in the YouTube API with either existing or custom build models particularly for our dataset
- Will try to use pre-trained NLP tools such as VADER (great for social media and short texts). We will compare the results with LLMs later
- Make a fully functional UI on Streamlit to recommend movies and tv shows

