Team 4

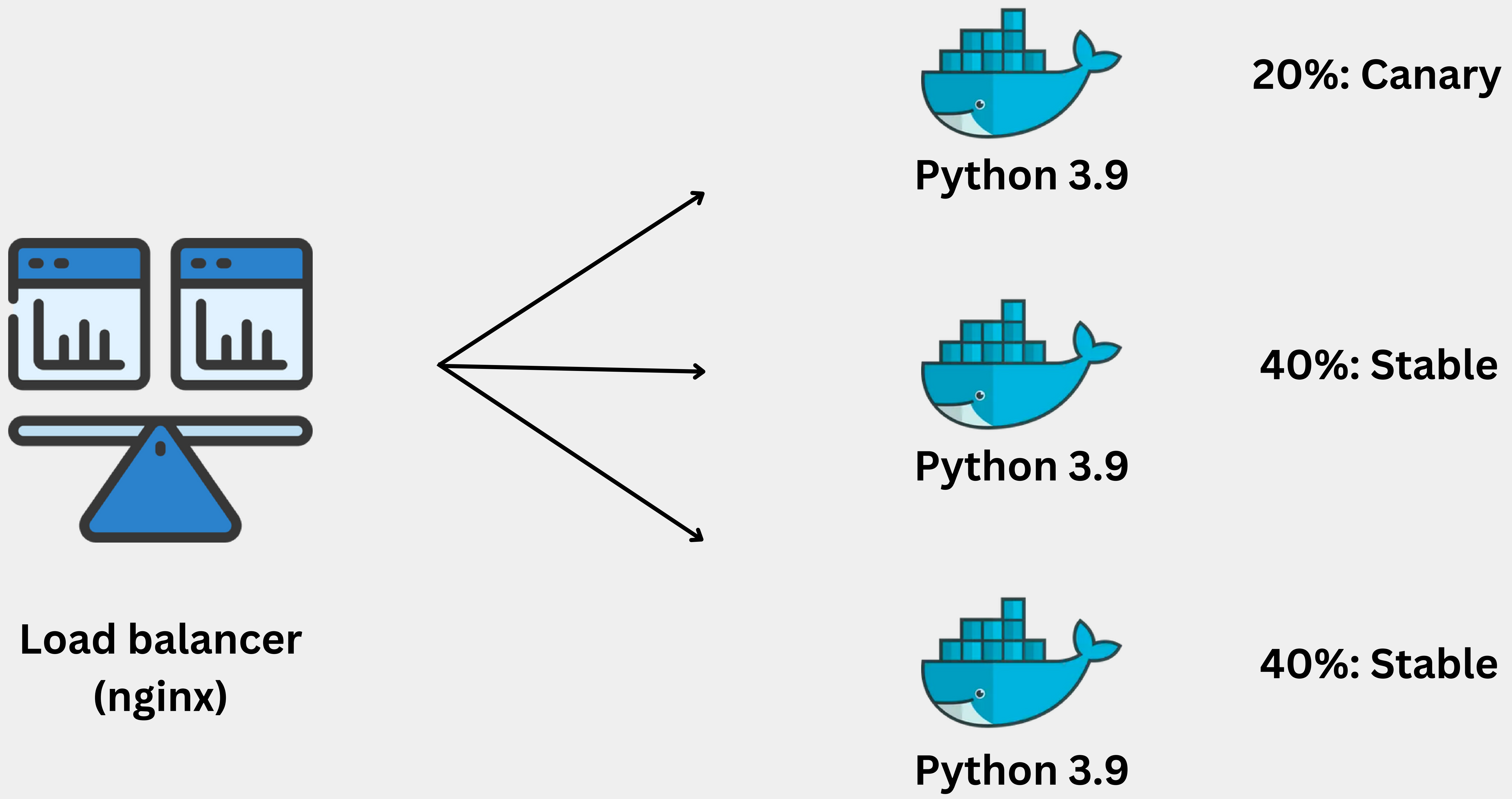
Milestone 3

Aayush, Yaoqiang, Rishabh, Tamara, Varun

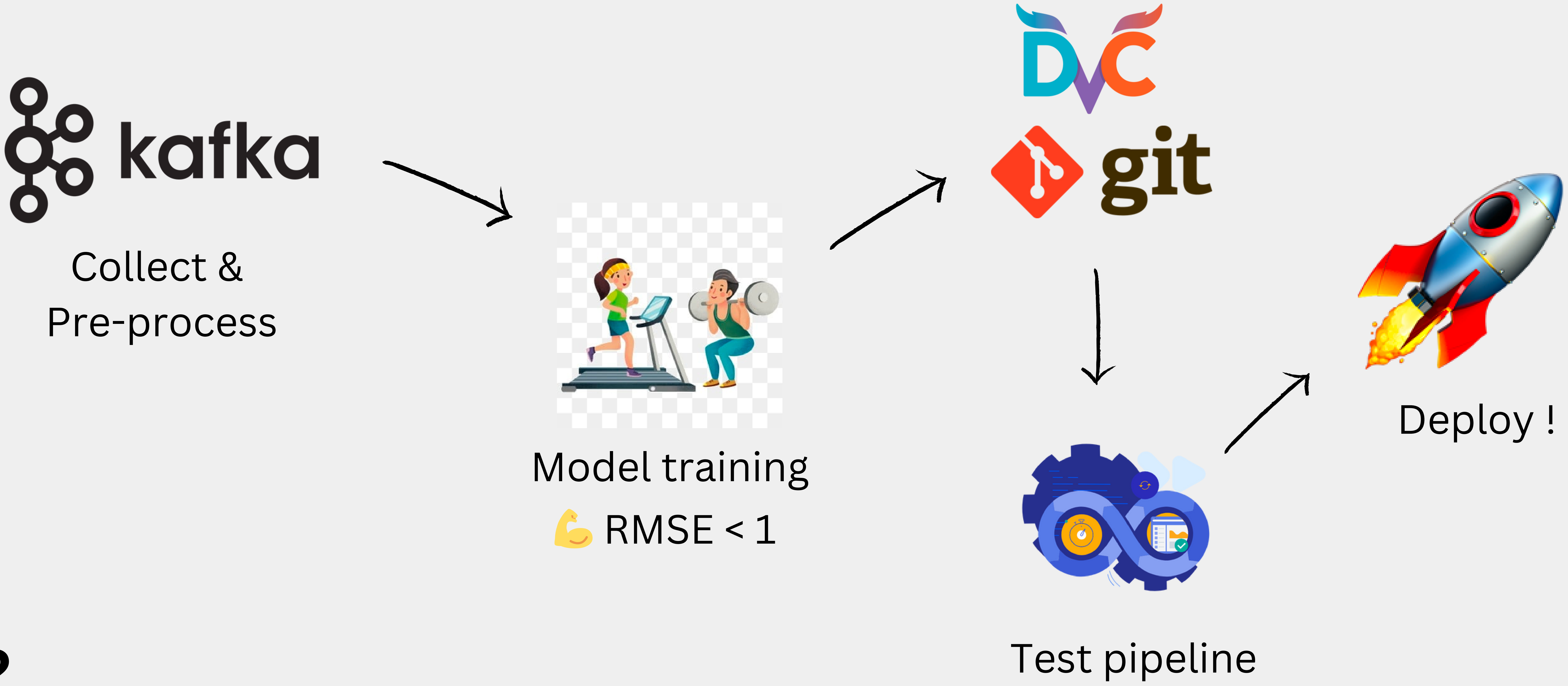
Overview

- ✨ Containerization
- ✨ Auto deployment
- ✨ Canary release
- ✨ Provenance
- ✨ Fairness & Feedback Loops
- ✨ Project Reflections

Containerization



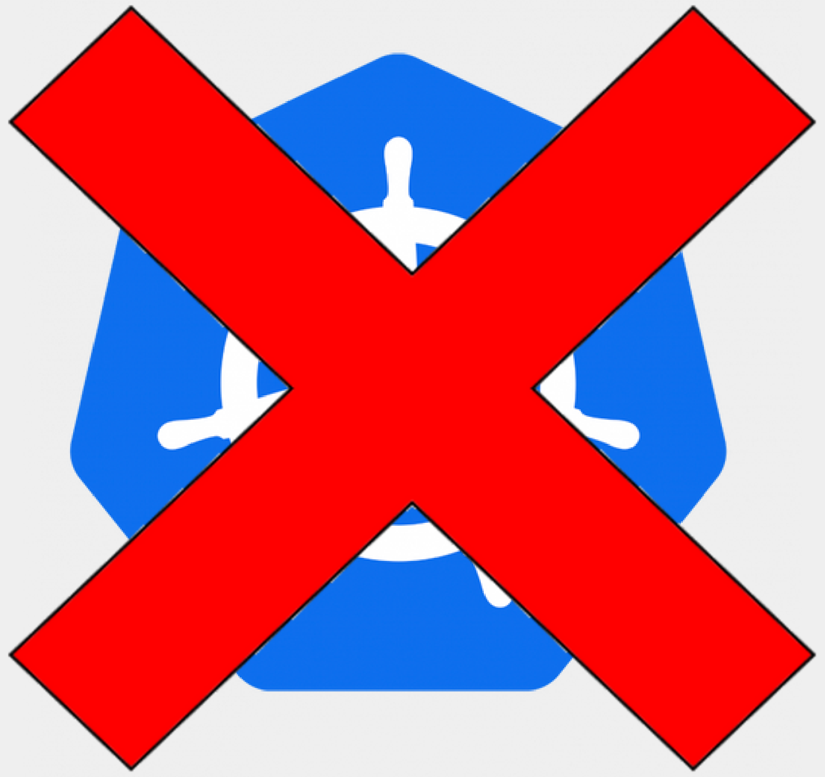
Auto deployment



Canary Release

Kubernetes = Load balancing + Release

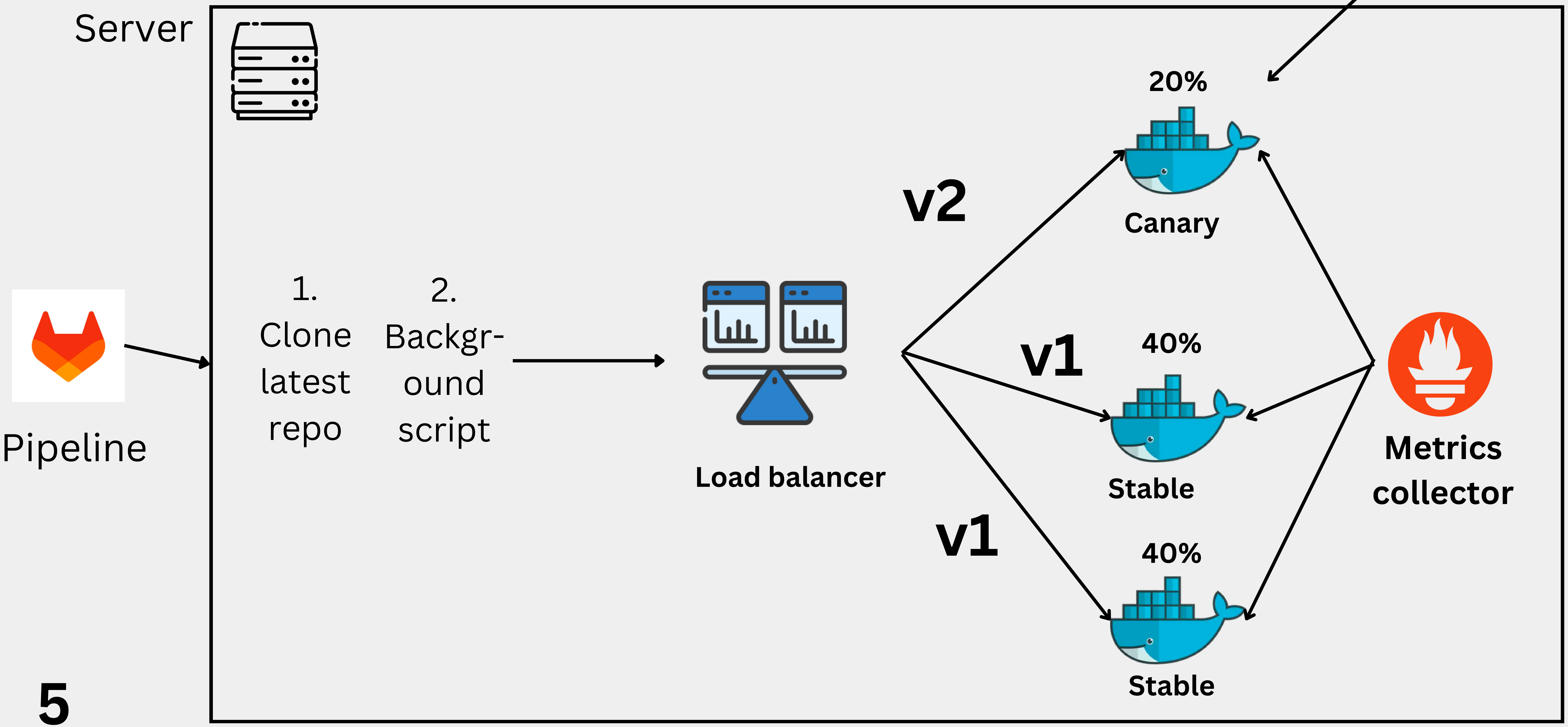
Canary Release



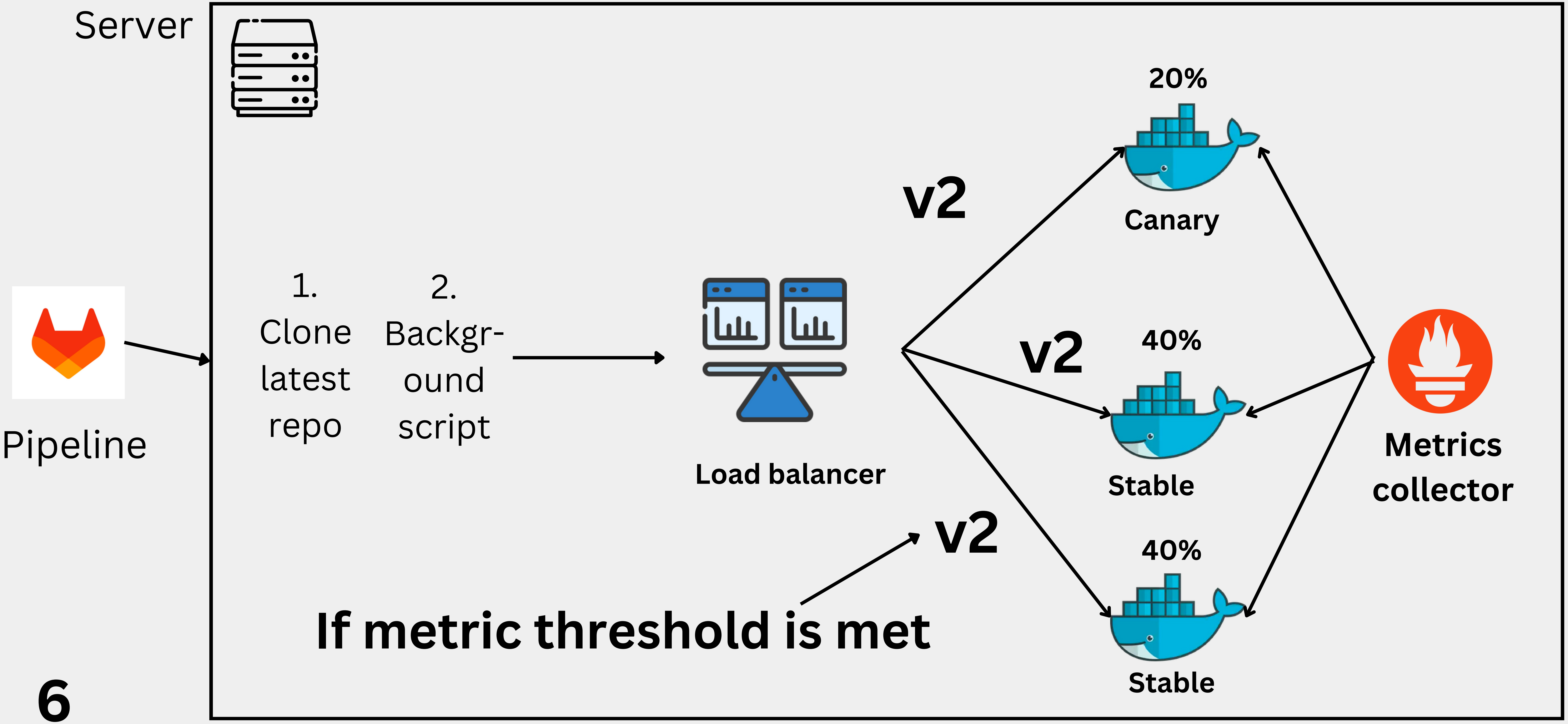
Kubernetes = Not working + Time wasted

http://fall2023-comp585-4.cs.mcgill.ca:80~~82~~

Canary Release



Canary Release



Provenance

Our requirements:

1. **Minimal overhead for tracking.**
2. **A friendly integration to our existing pipeline.**

Use DVC to track provenance..

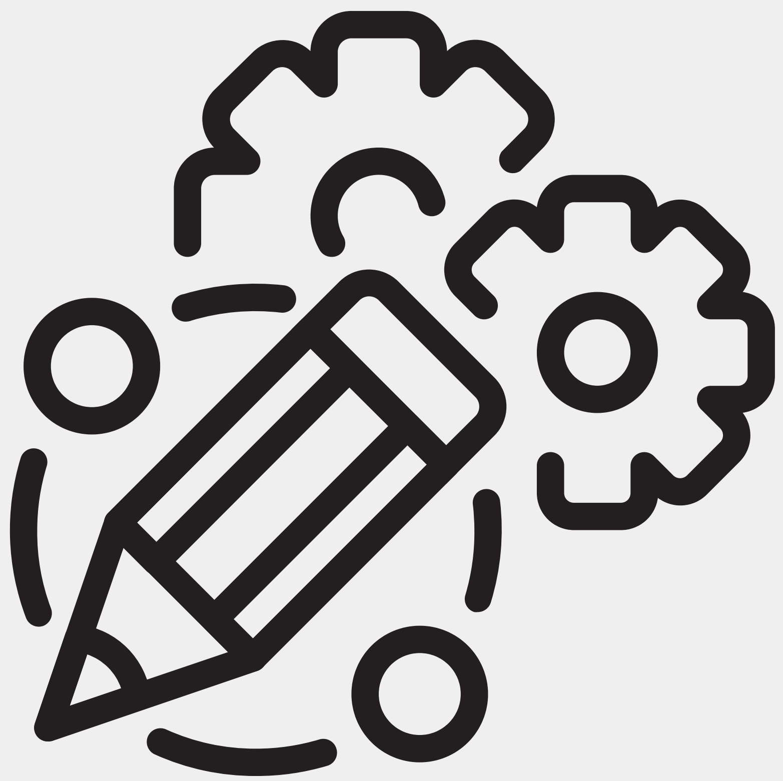
Our stack:

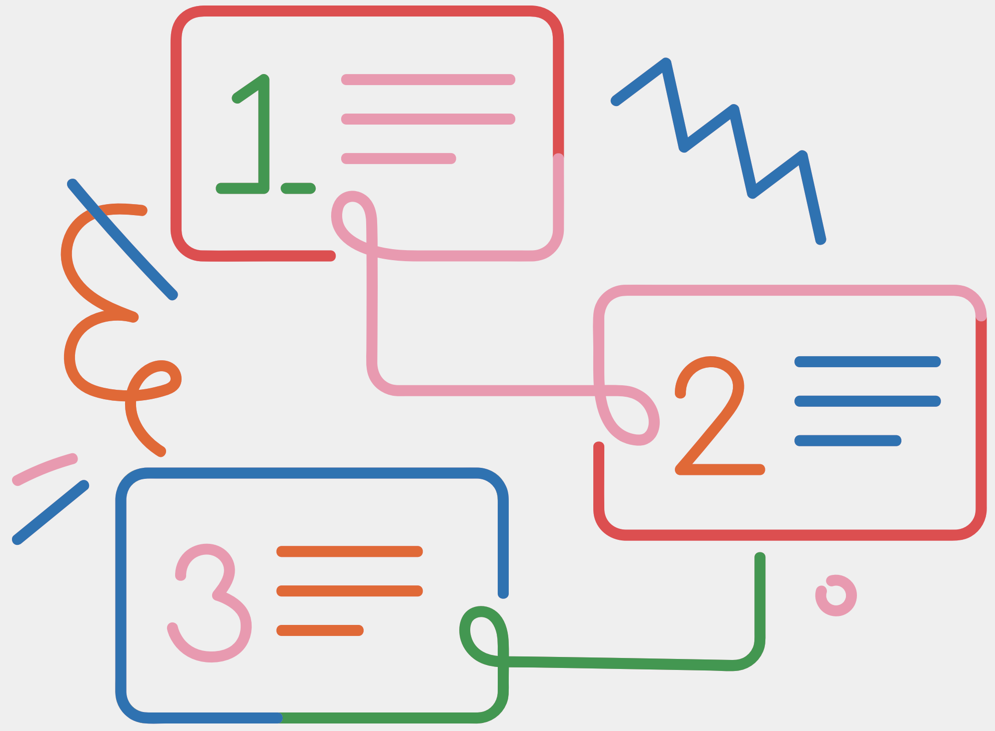
- **Gitlab CI/CD for deployment.**
- **Docker containers for switching releases.**

Provenance

What to track?

- **Pipeline code: (already bound to commit IDs)**
- **Data files: (needs to be tracked)**
- **Model files: (needs to be tracked)**





Where does does dvc fit in?

- After data is captured, cleaned, processed.

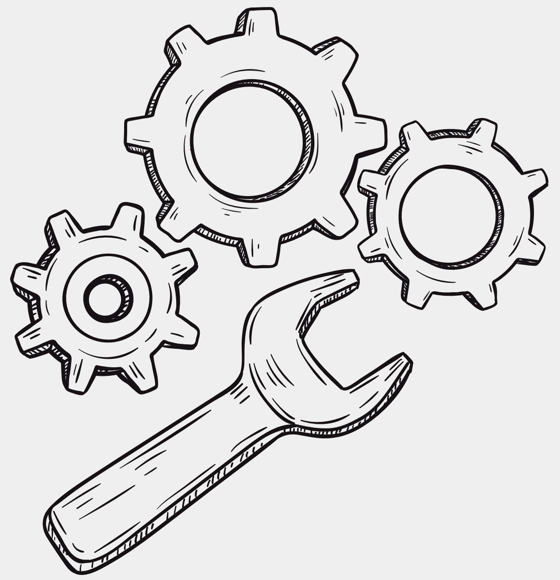
- After models are trained and pkl file created in the container before serving predictions.

Provenance

Under the hood:

- **DvC: acts as an abstraction.**

- **Moves data to cache (offloadable to remote): symbolic links**
    - **Existing code works the same way**



Metadata committed
Associated with commit ID.

Used for tracking

Provenance

- Each release will have the latest commitID.
- Recommendations given by that container linked to that commitID.

Tracks: model, data, pipeline code upto that commit

- No excessive overhead/ data collection apart from logging each recommendation's handler commit.

- This is portable: can even be done in gitlab runner along with pipeline.

Fairness

Similar group

$\longrightarrow$ **Similar result**

There should be no difference between social groups when it is not necessary!

Fairness

Do two people from two different social groups have reason to have different recommendations? Yes!

But there should be no difference in the quality of recommendations between social groups!

Fairness

Gender

Age

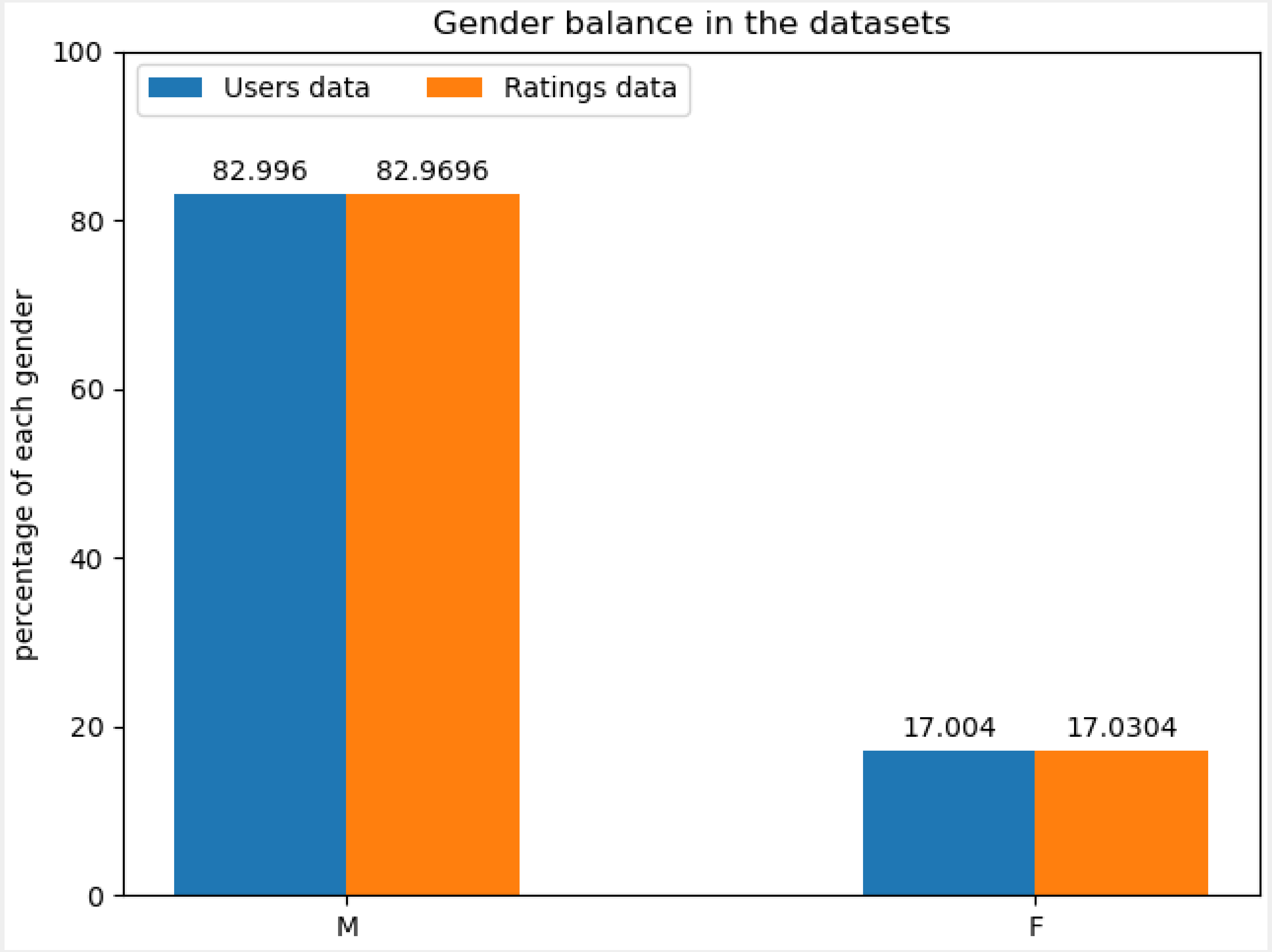
Fairness

- Dataset balance analysis between genders
- Dataset balance analysis between ages
- Dataset balance analysis between ages x genders

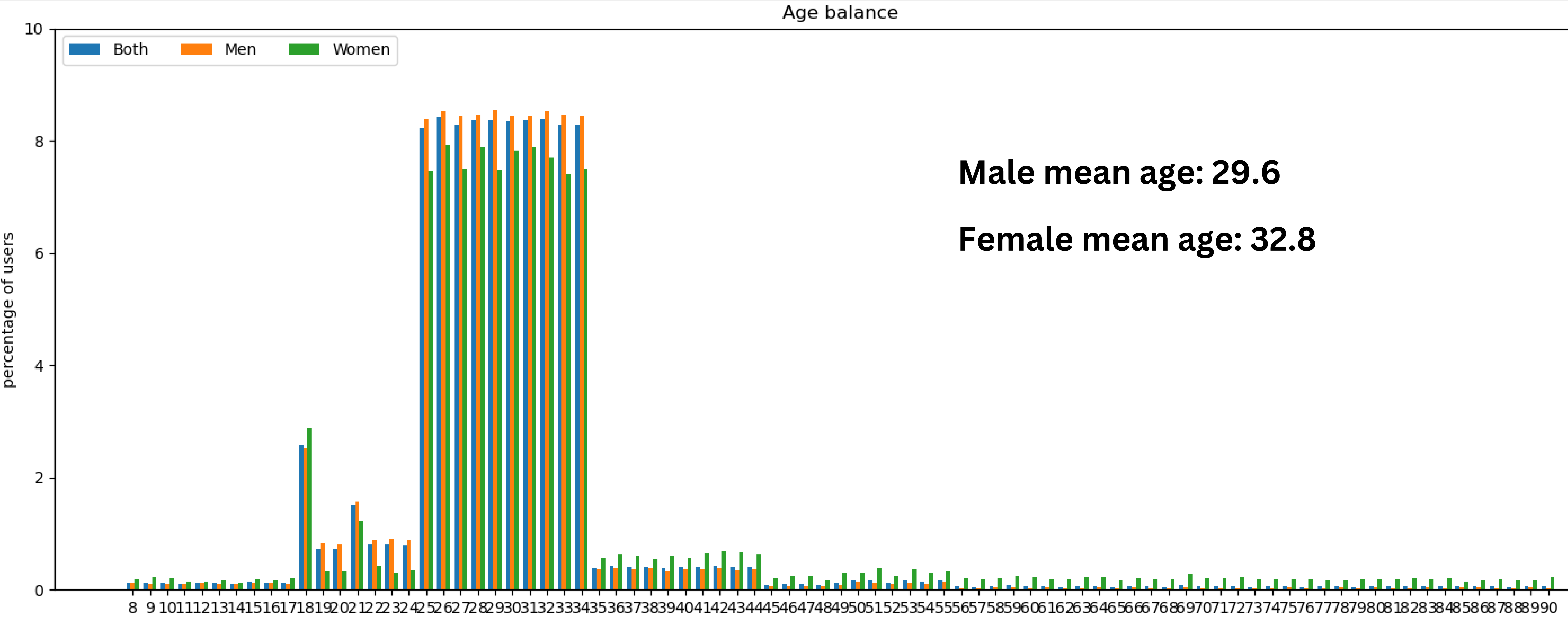
In progress...

- Accuracy analysis between genders
- Accuracy analysis between ages
- Accuracy analysis between ages x genders
- Reduce biases

Fairness


Gender balance in the datasets
Users data
Ratings data
percentage of each gender
82.996
82.9696
17.004
17.0304
0
20
40
60
80
100
M
F

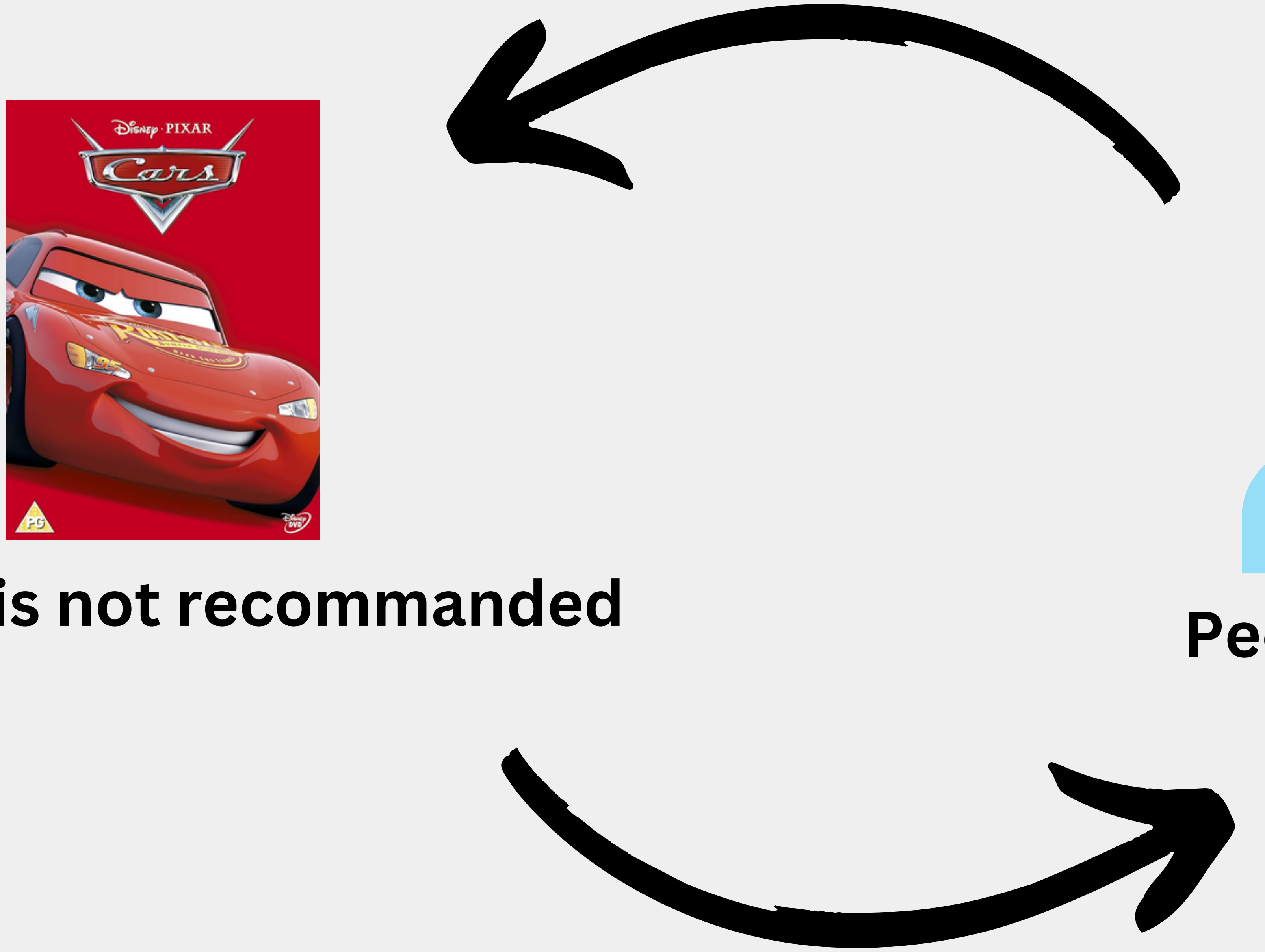
Fairness



Male mean age: 29.6

Female mean age: 32.8

Feedback loop



The movie is not recommanded

People will not watch it

Feedback loop x Fairness

The movie is recommanded a lot to
a particular social group

This social group watch
more this movie

Feedback loop

Mitigate feedback loop: add a little part of randomization in the recommendations!

Project Reflections

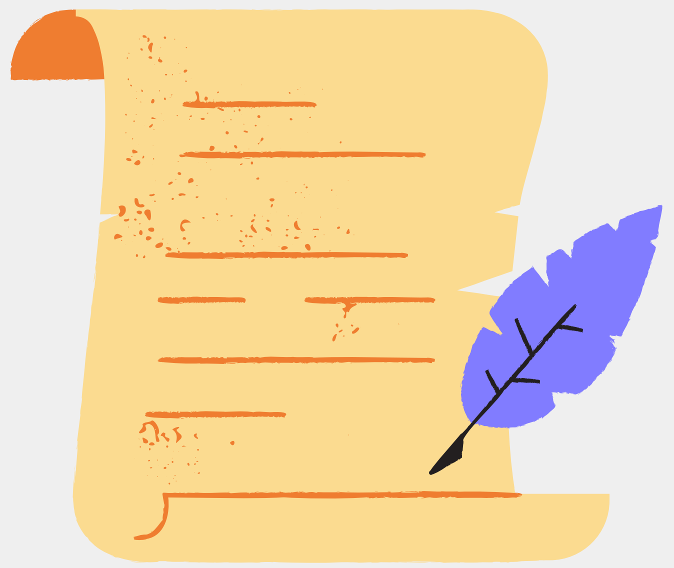
Key Challenges

1. Telemetry System: Transitioned from Loki to CSV files for log management.

2. Kubernetes: Shifted from complex Kubernetes setup to NGINX load balancing.

Conclusion

The project offered crucial lessons in system management and adapting to technical constraints, setting a foundation for future scaling.



End

Thank you

Do you have any questions?