

Yandex

Crowdsourcing Practice for Efficient Data Labeling: Aggregation, Incremental Relabeling, and Pricing

Alexey Drutsa, Valentina Fedorova, Dmitry Ustalov, Olga Megorskaya, Evfrosiniya Zerminova, Daria Baidakova

Part VI

Setting up and running label collection projects cont.

Daria Baidakova,
Project Manager

Yandex.Toloka is a service of Swiss company Yandex Services AG

Tutorial schedule

Introduction: 15 min

Part I: 20 min
Main Components

Part II: 10 min
Introduction to
Crowd Platform

Part III: 15 min
Brainstorming
pipeline

Part IV: 60 min
Set & Run Projects

Part V: 25 min
Theory on
Aggregation

**Break:
30 min**

Part VI: 20 min
Set & Run Projects
cont.

Part VII: 10 min
Results &
Conclusions

In this part we continue configuring projects:

- You configure crowd projects

- You connect them in a pipeline

- You run them on real performers

- You have:

- > The instruction with all details

https://tlk.s3.yandex.net/re substitutes/SIGMOD_instruction.pdf

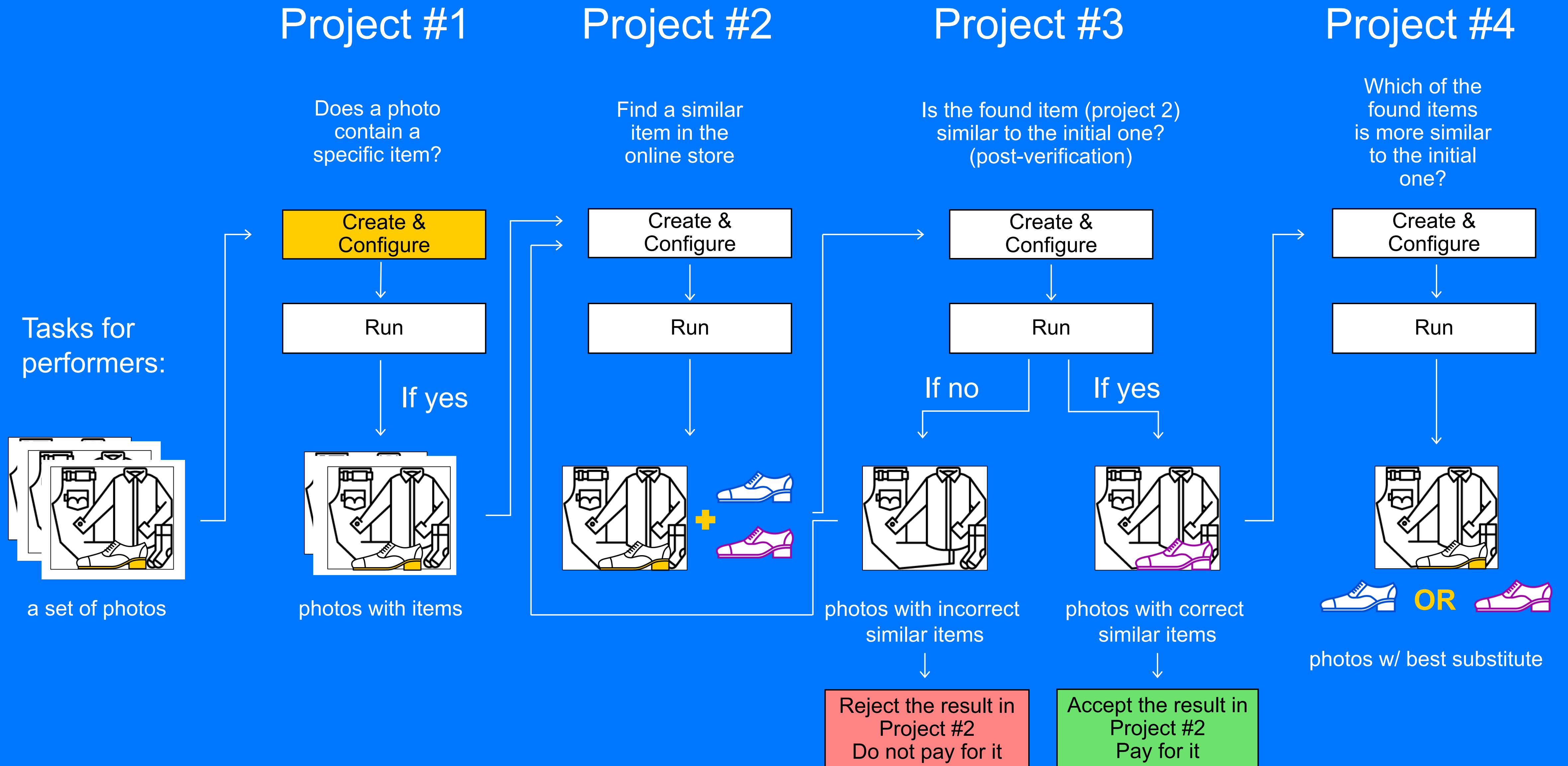
- > The dataset with photos

[http://tlk.s3.yandex.net/re substitutes/dataset Y.tsv](http://tlk.s3.yandex.net/re substitutes/dataset_Y.tsv)

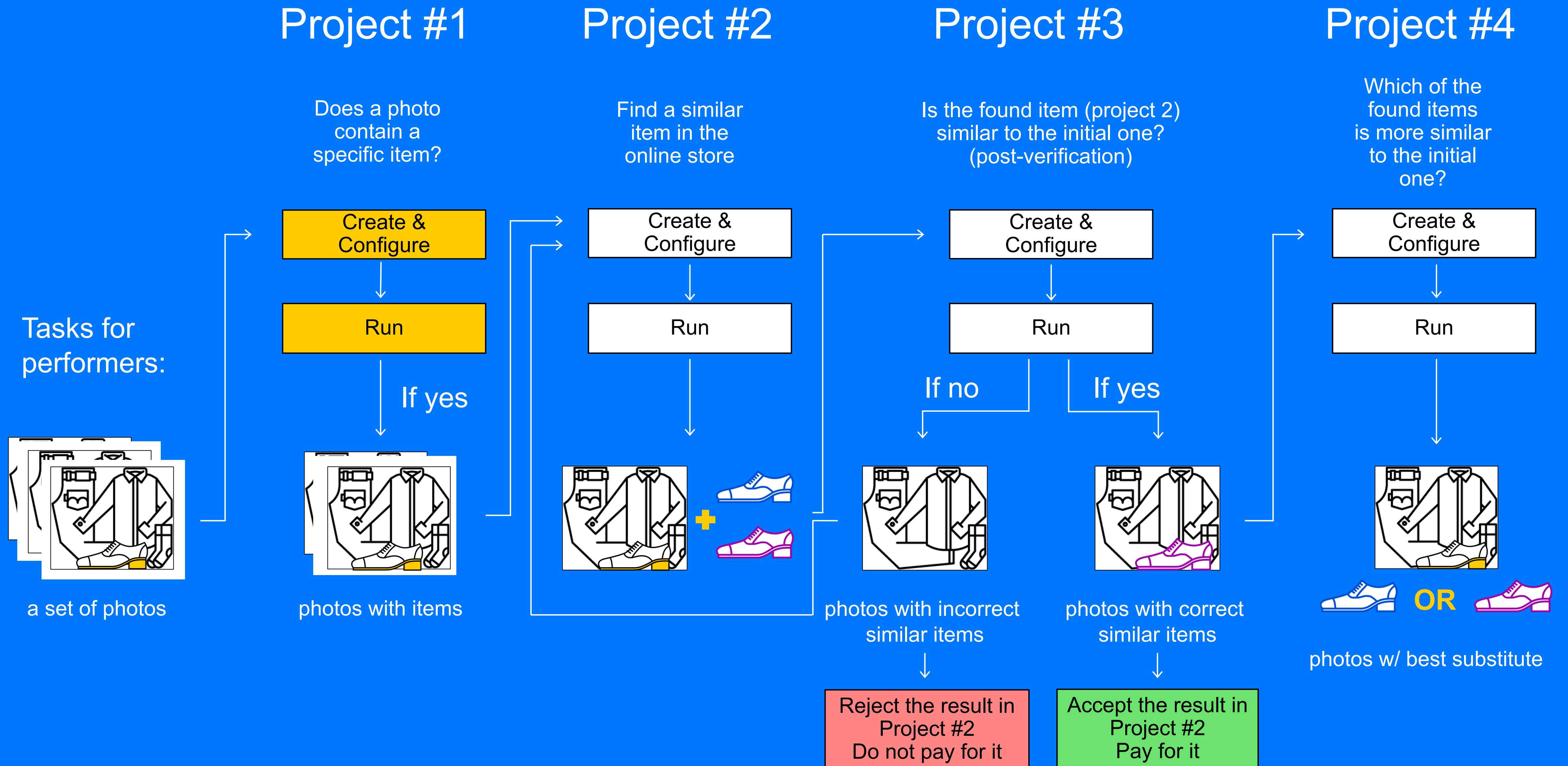
where **Y** is any number from 1 to 10



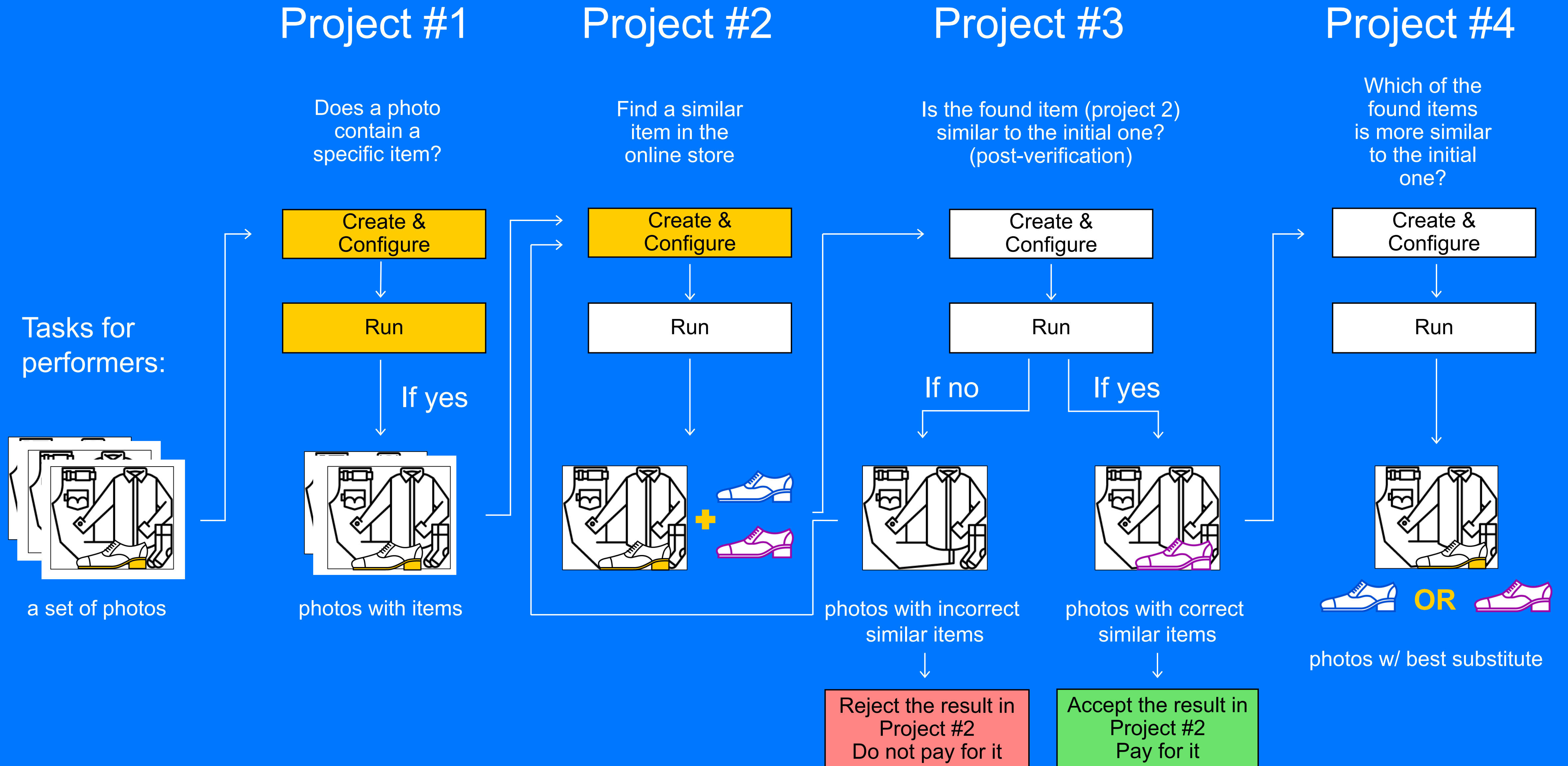
Most of us are at this step



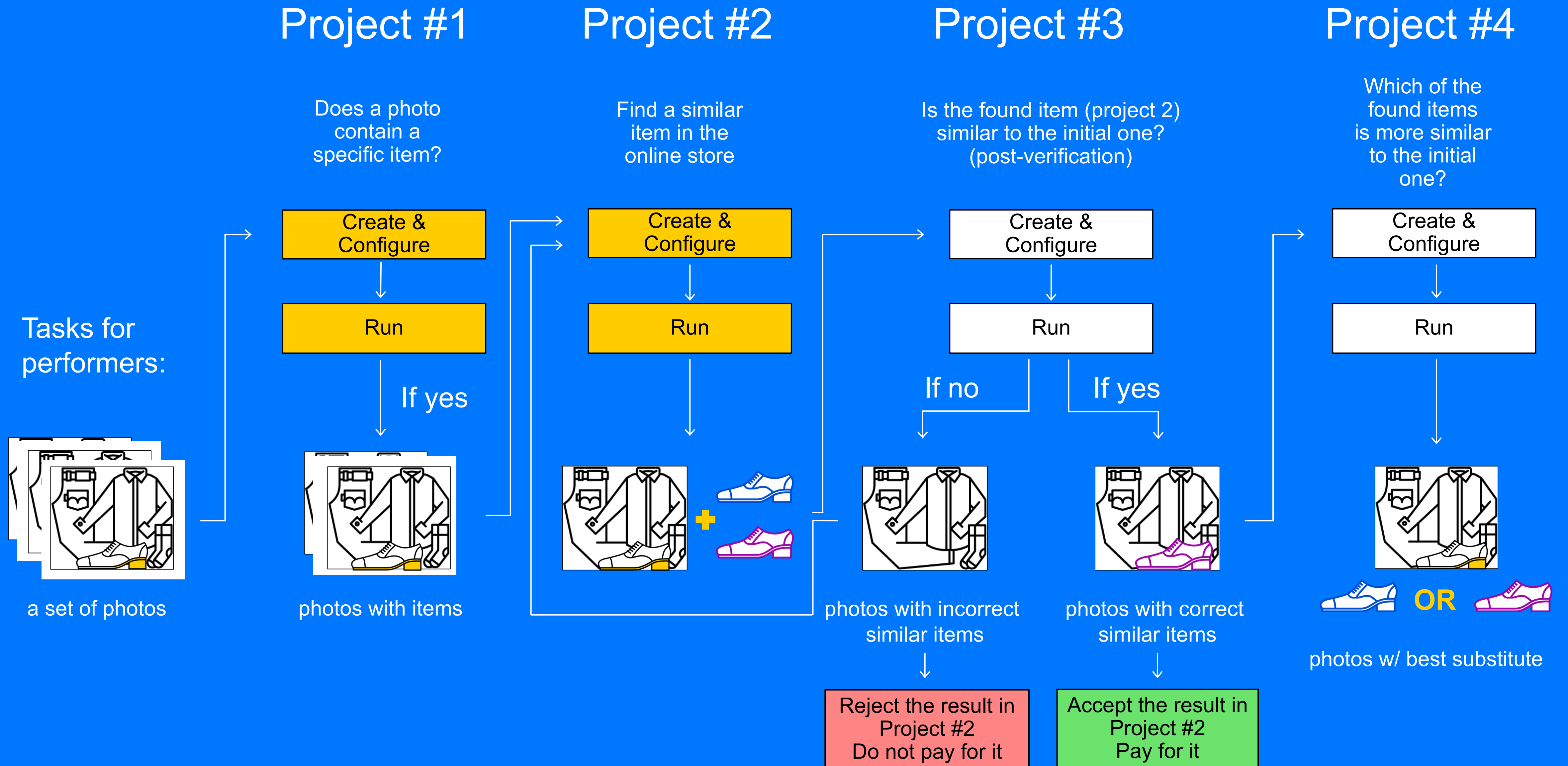
Most of us are at this step



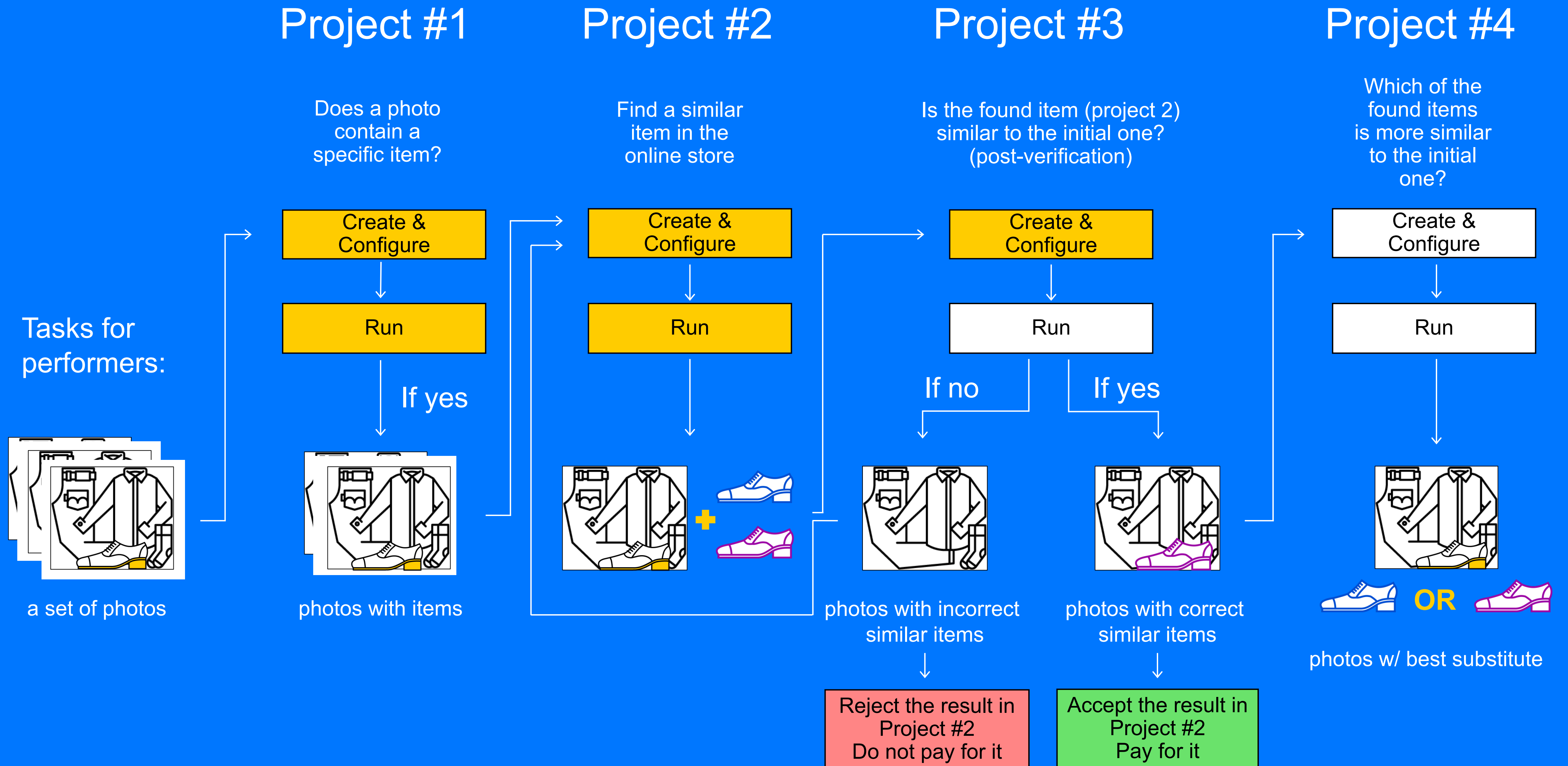
Most of us are at this step



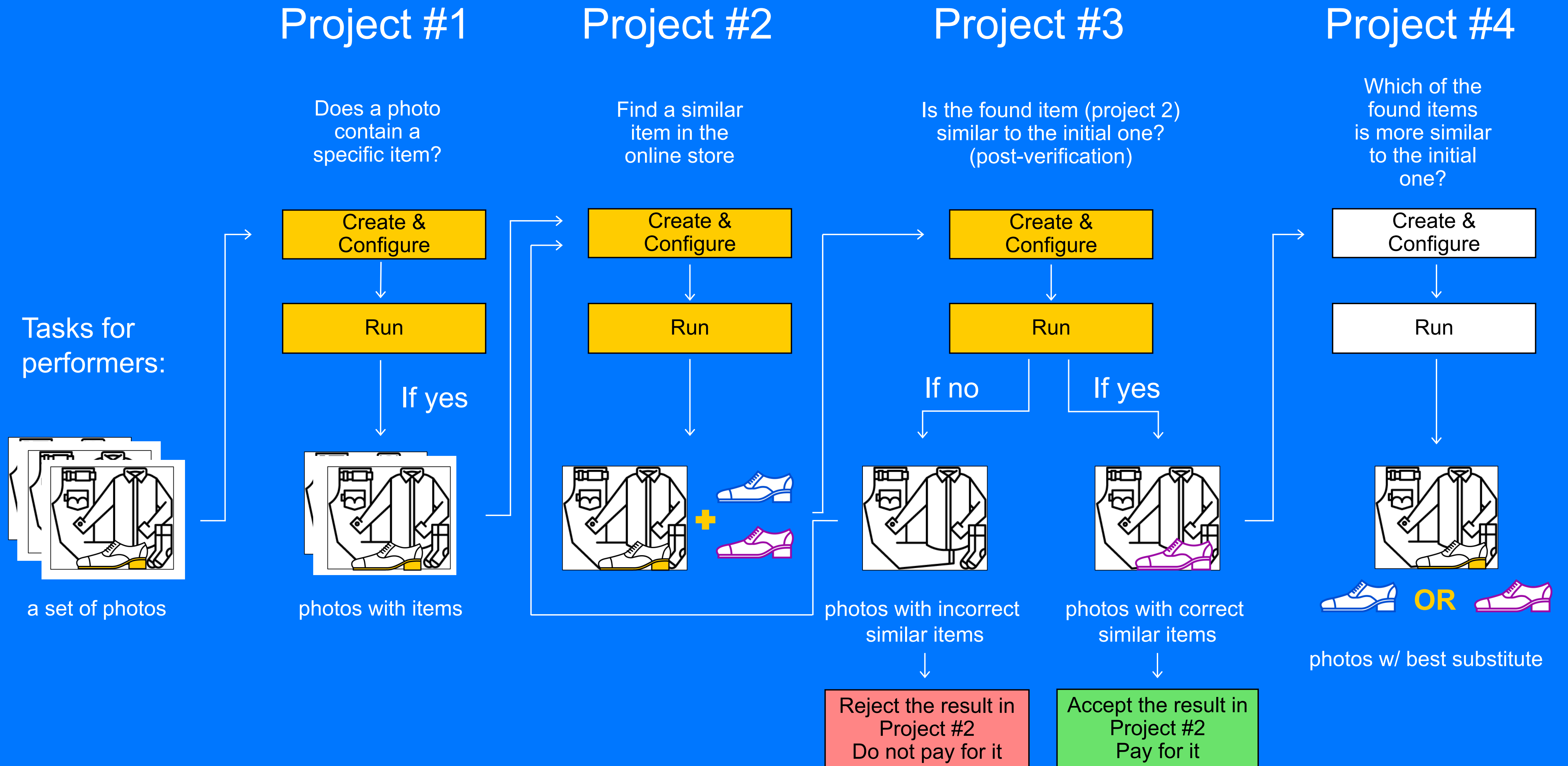
Most of us are at this step



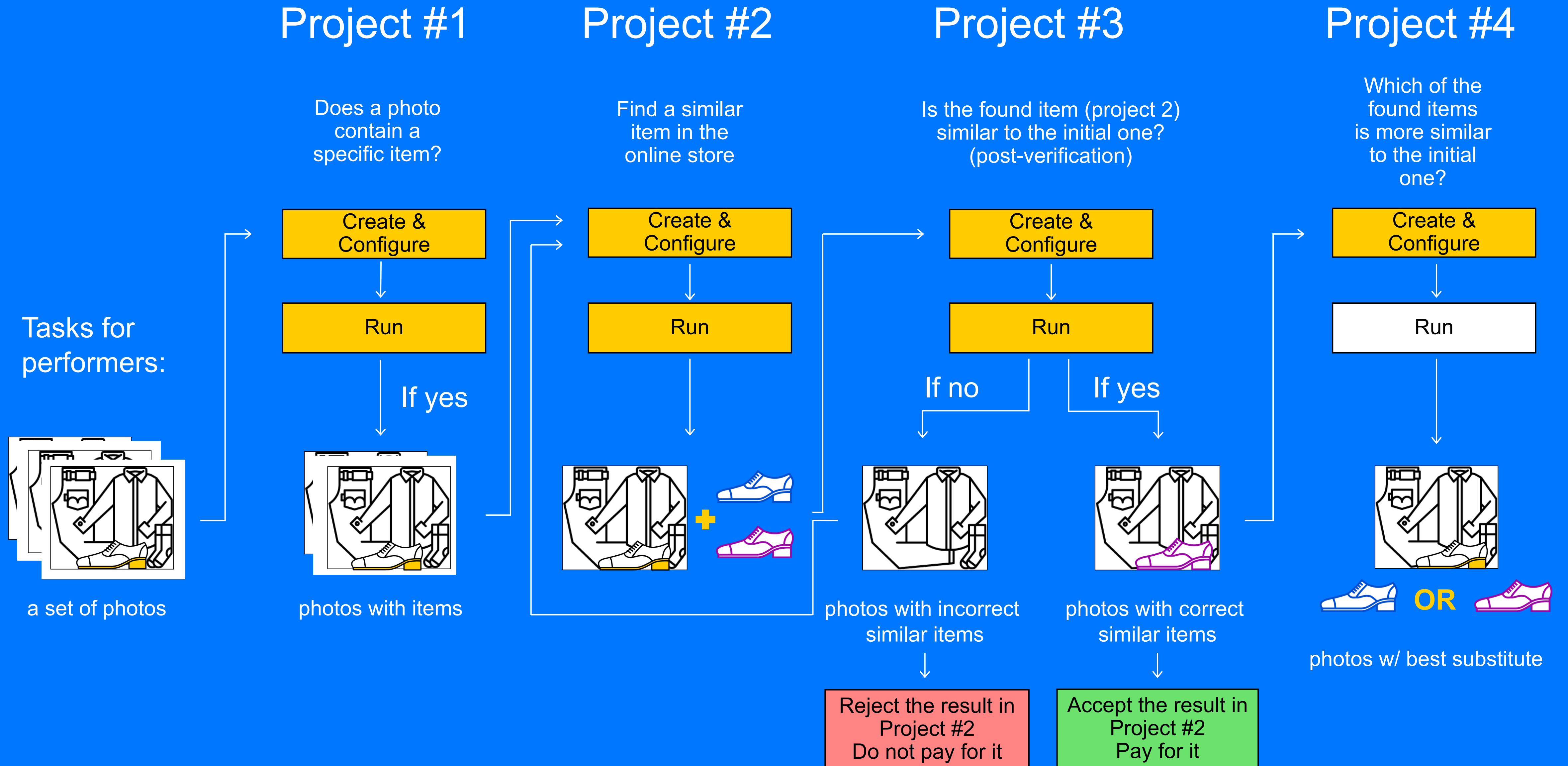
Most of us are at this step



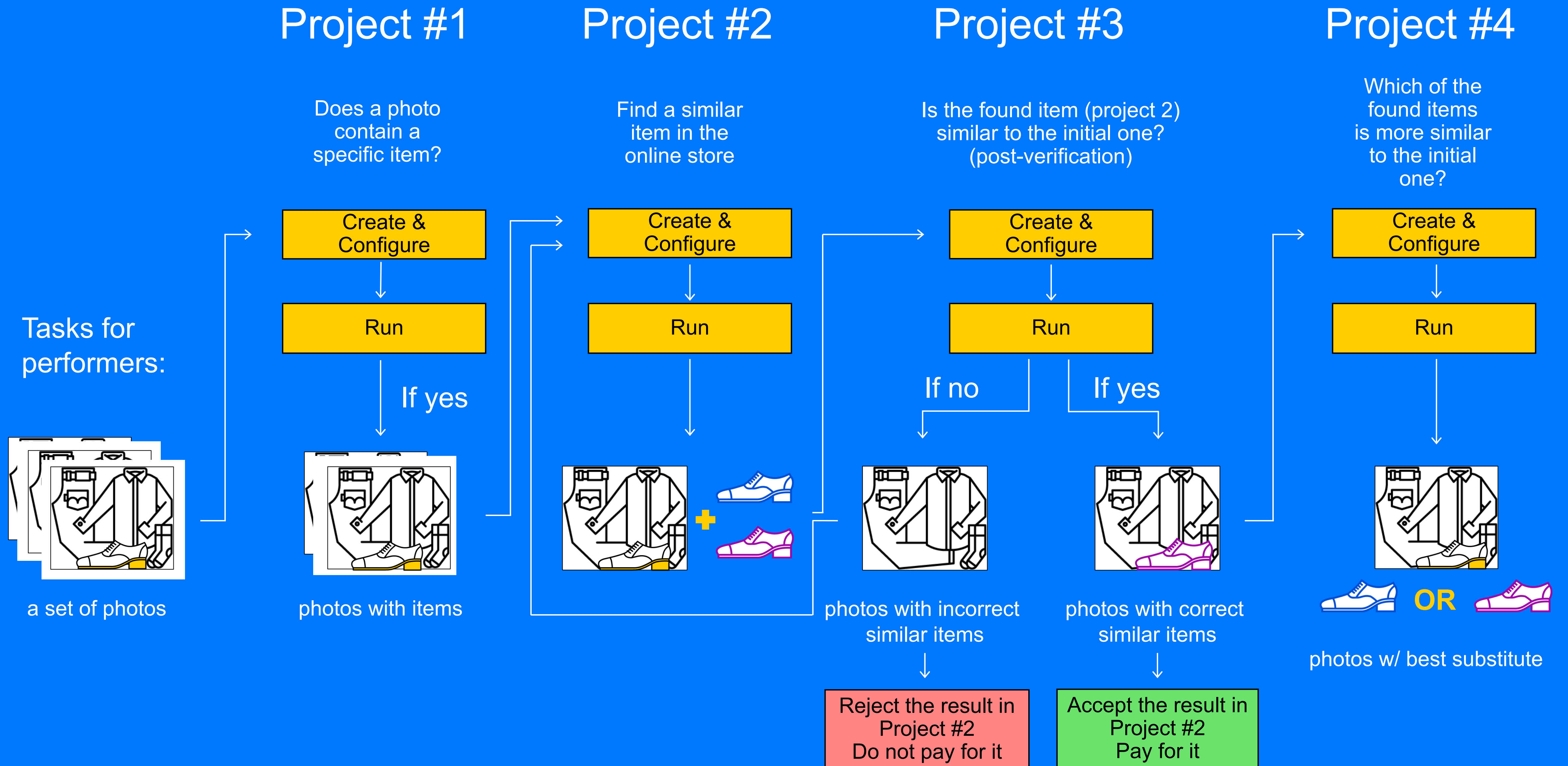
Most of us are at this step



Most of us are at this step



Most of us are at this step



Yandex

Thank you!
Questions?

Daria Baidakova

Project Manager



dbaidakova@yandex-team.ru



<https://research.yandex.com/tutorials/crowd/sigmod-2020>

