

Yandex

Practice of Efficient Data Collection via Crowdsourcing: Aggregation, Incremental Relabelling, and Pricing

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

Part IV

Setting up and running label collection projects cont.

Daria Baidakova,
Project Manager

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

Tutorial outline

Introduction: 20 min

Part I: 40 min
Main Components

Coffee break: 30 min

Part II: 25 min
Brainstorming pipeline

Part III: 10 min
Introduction to Crowd Platform

Part IV: 85 min
Set & Run Projects

Lunch break: 90 min

Part V: 35 min
Interface & Quality control

Part VI: 25 min
Theory on Aggregation

Coffee break: 30 min

Part VI: 60 min
Set & Run Projects cont.

Part VII: 20 min
Incremental relabeling and pricing

Part VIII: 10 min
Results & Conclusions

Practice starts: you have 60 min left

You create 4 crowd projects

You connect them in a pipeline

You run them on real performers

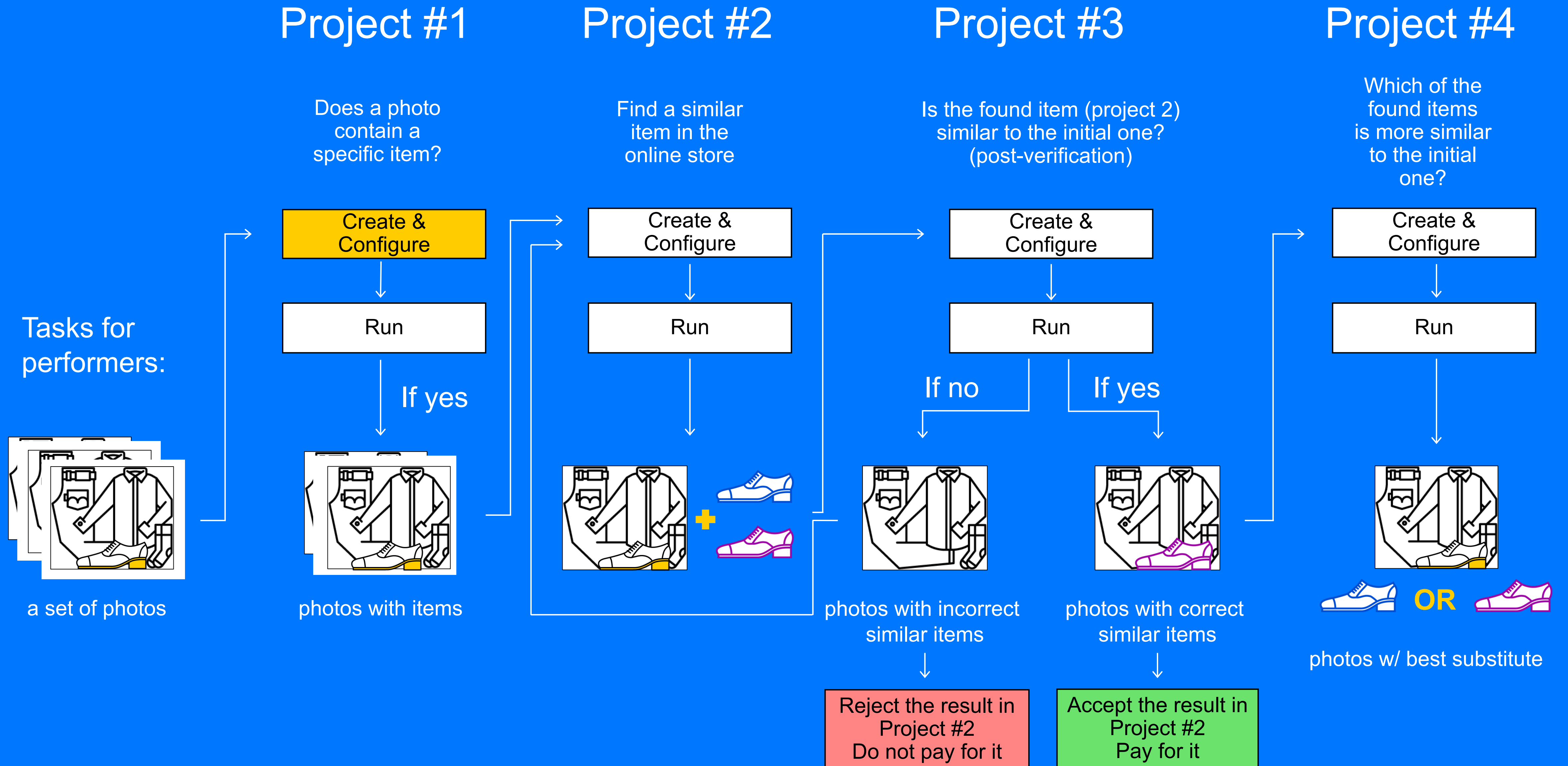
You have:

- › The instruction with all details
- › The dataset with photos

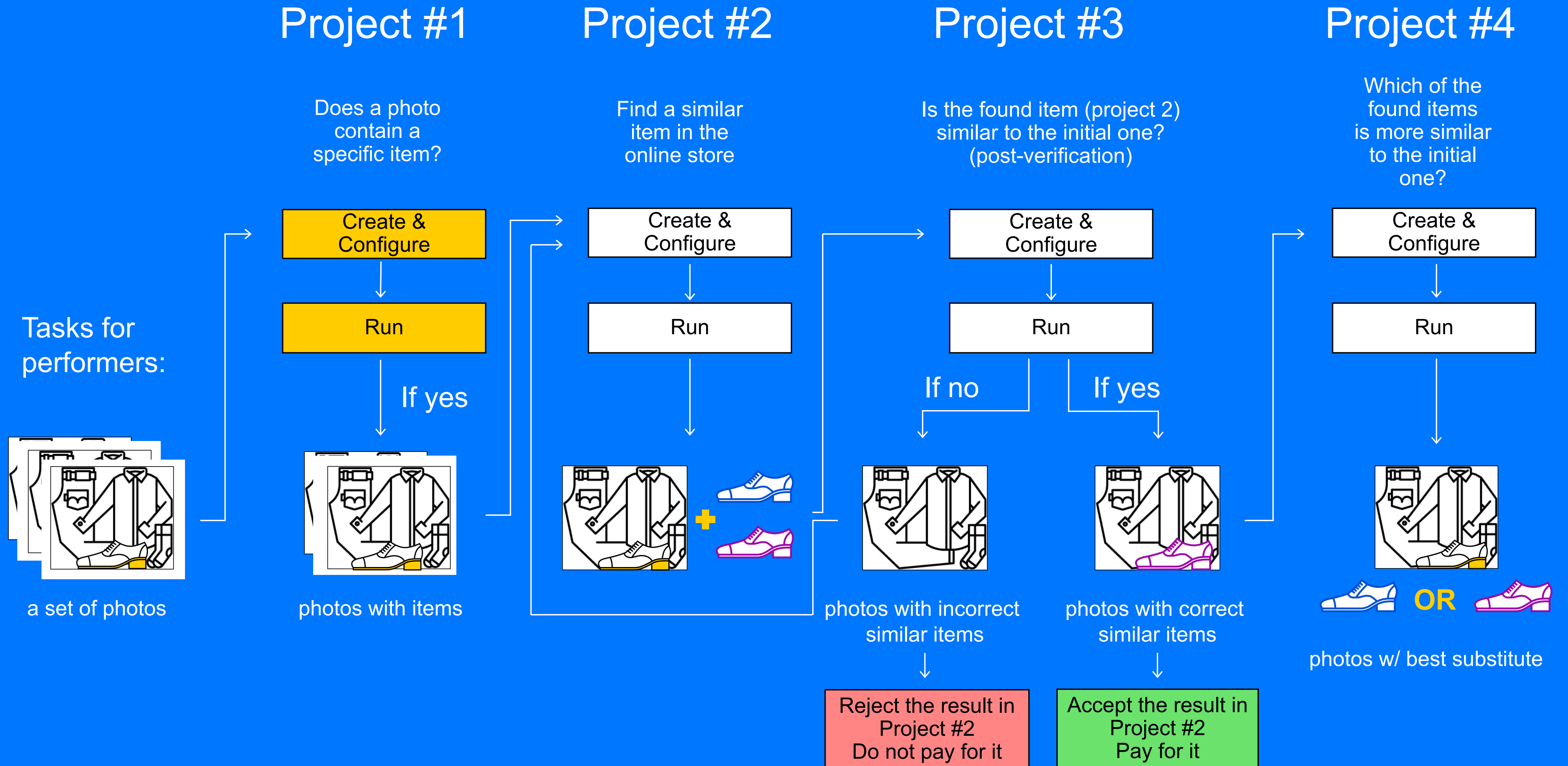
Tutorial chat with materials:



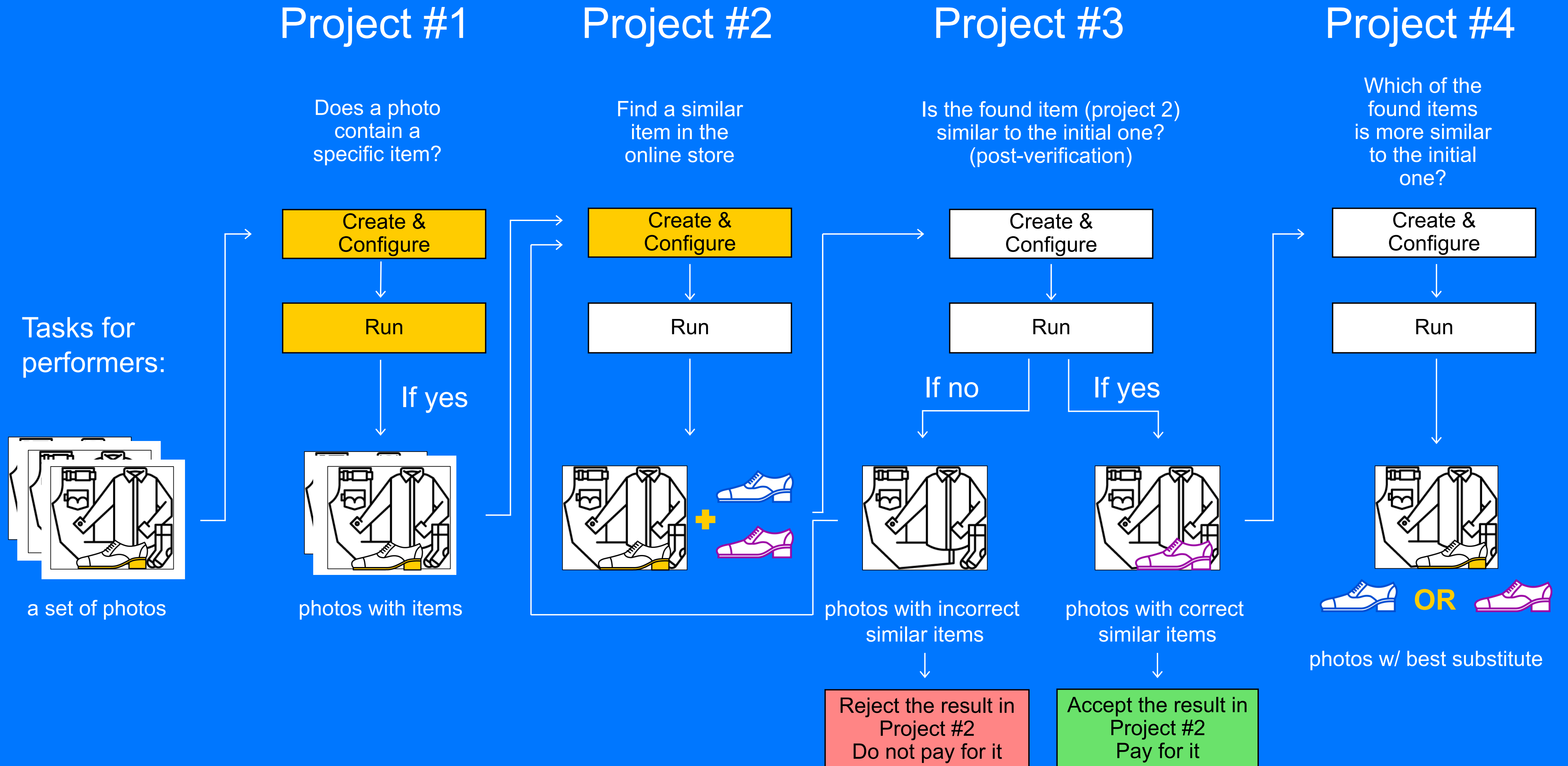
Most of us are at this step



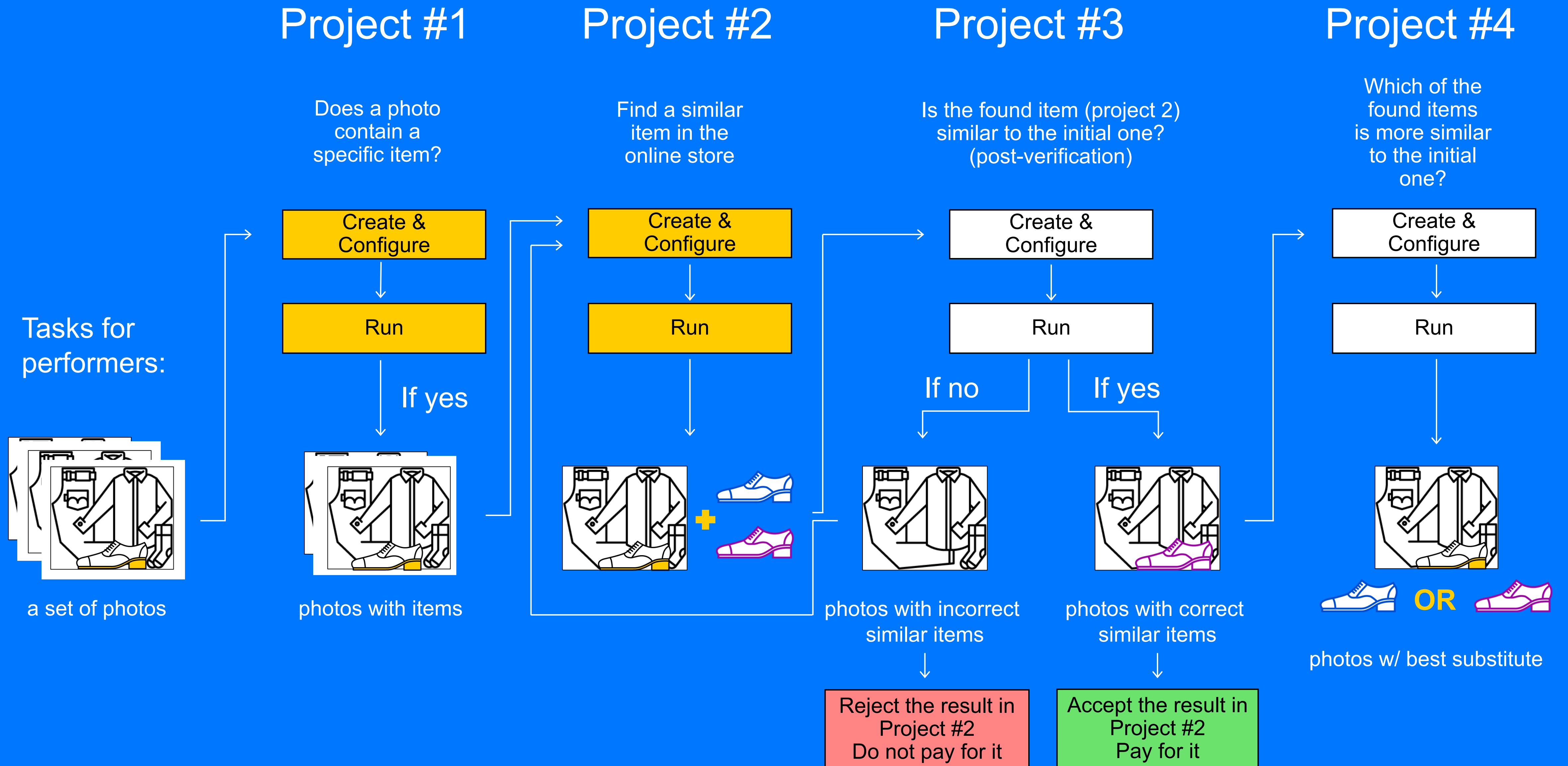
Most of us are at this step



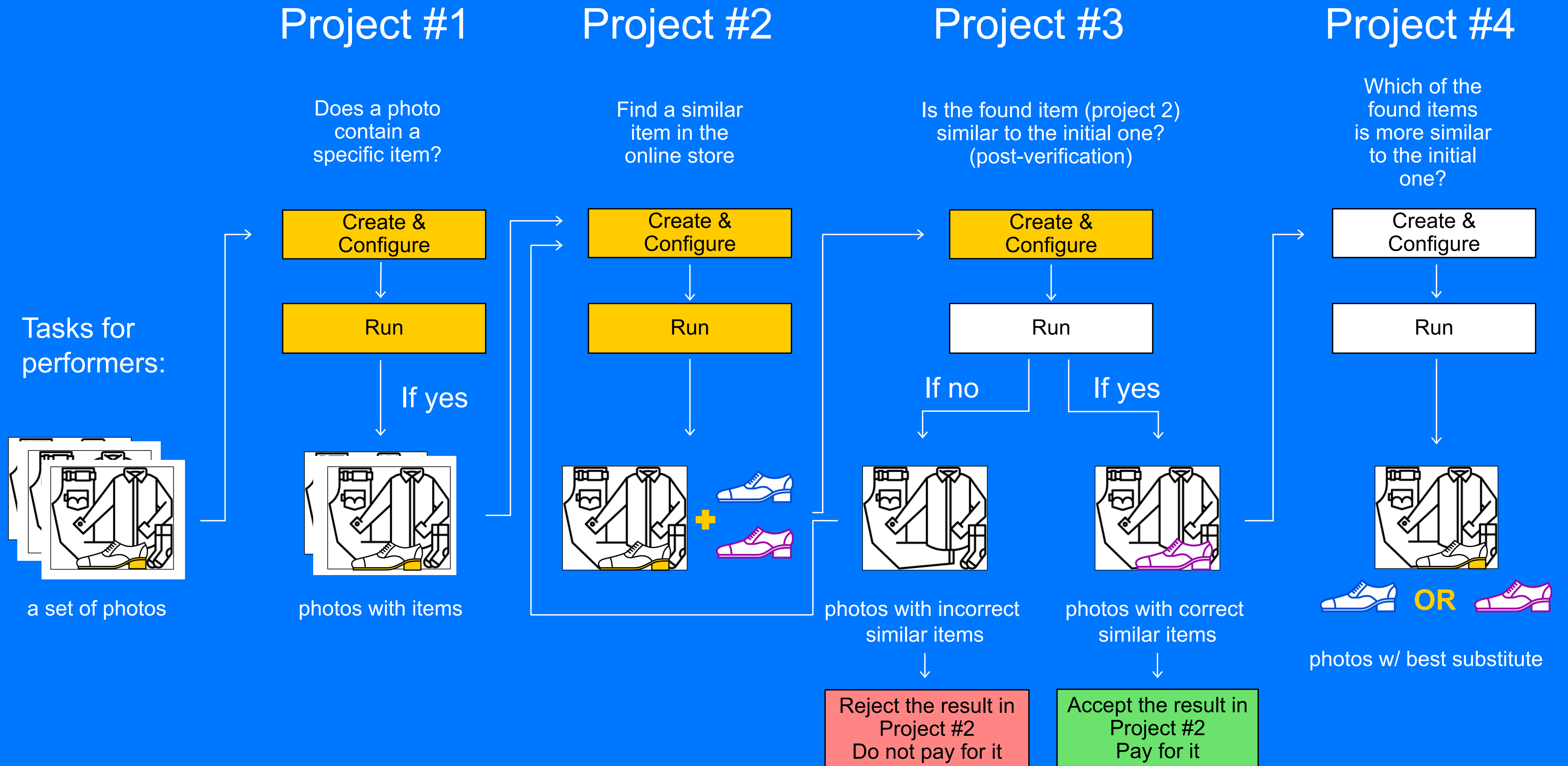
Most of us are at this step



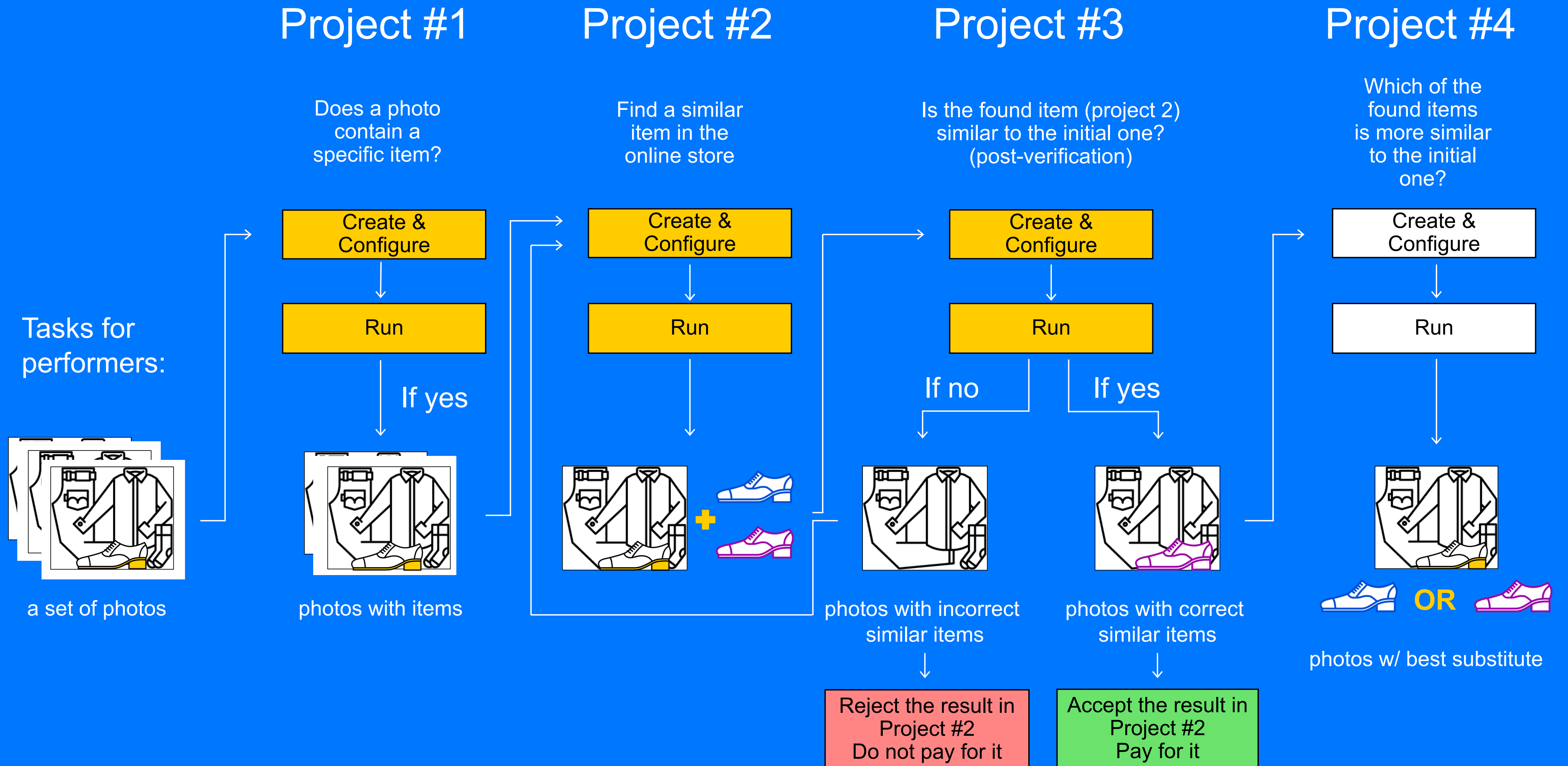
Most of us are at this step



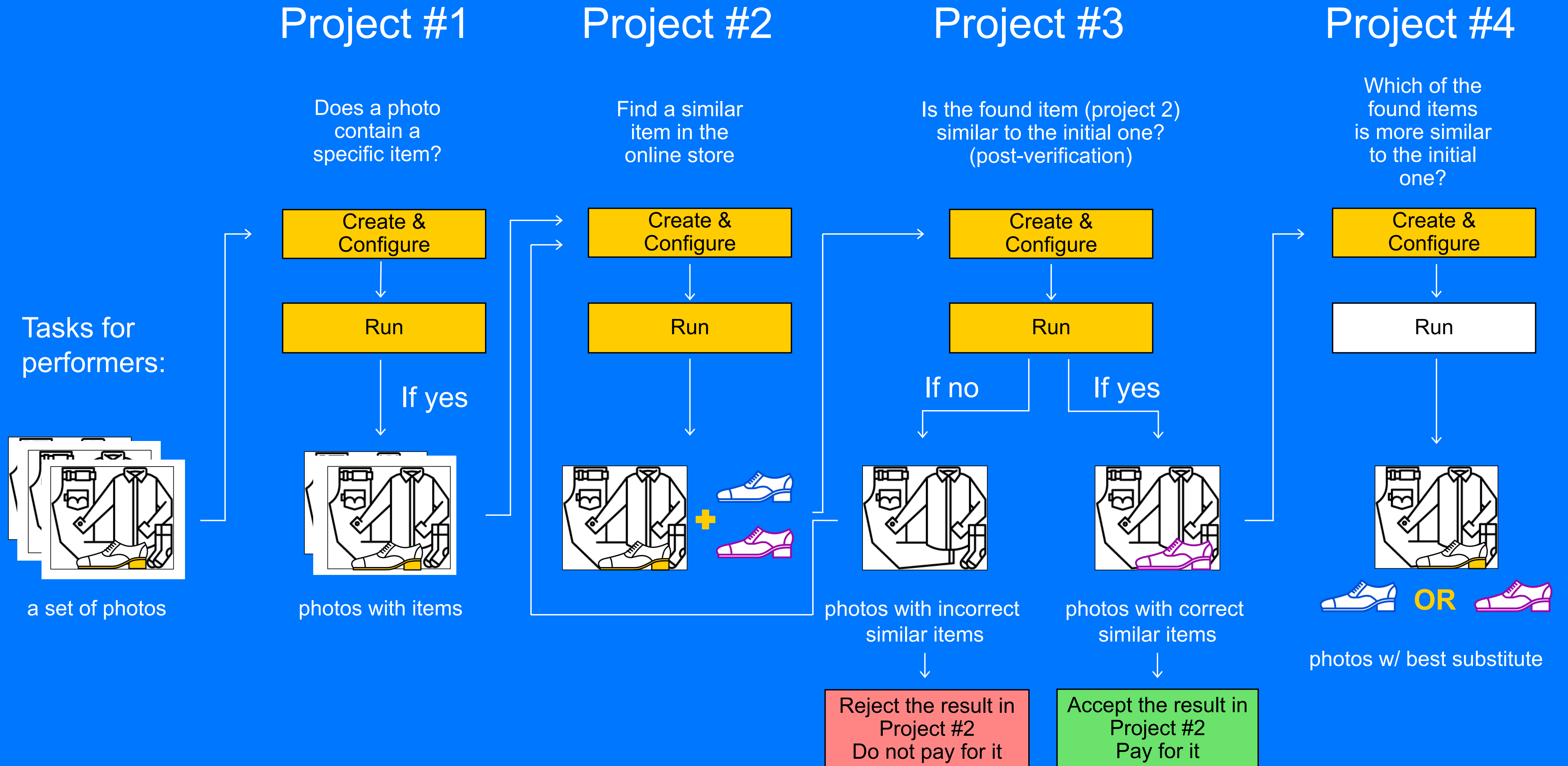
Most of us are at this step



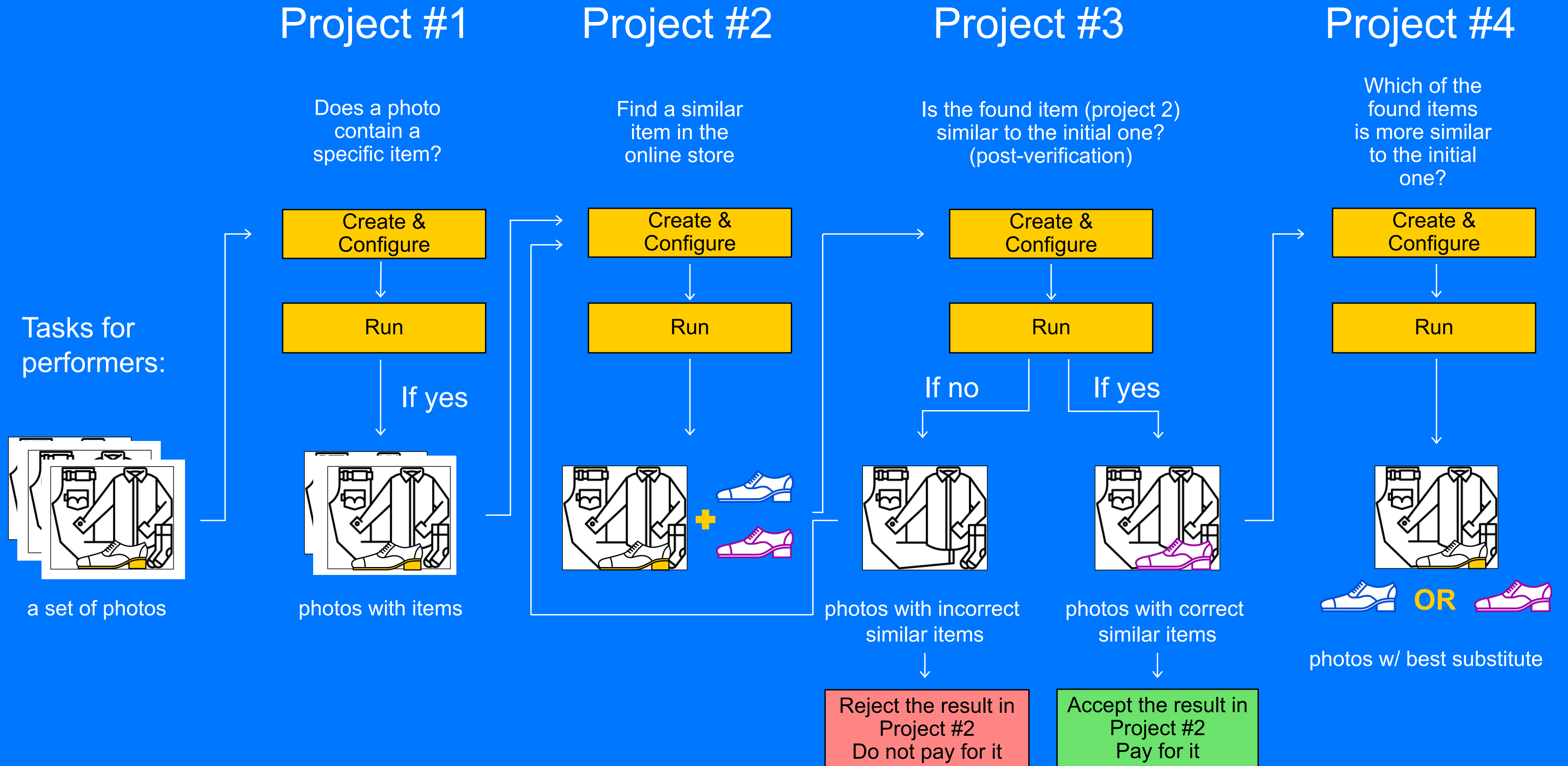
Most of us are at this step



Most of us are at this step



Most of us are at this step



Yandex

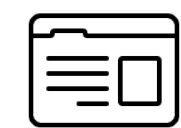
**Thank you!
Questions?**

Daria Baidakova

Project Manager



dbaidakova@yandex-team.ru



<https://research.yandex.com/tutorials/crowd/kdd-2019>