



# Product Data Science

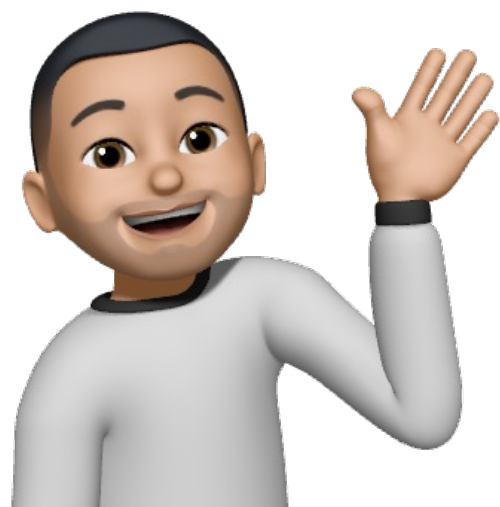
How to use machine learning to  
discover problems worth solving

**Azamat Omuraliev** | Senior Data Scientist @ ING  
Datafest, 16 September 2023

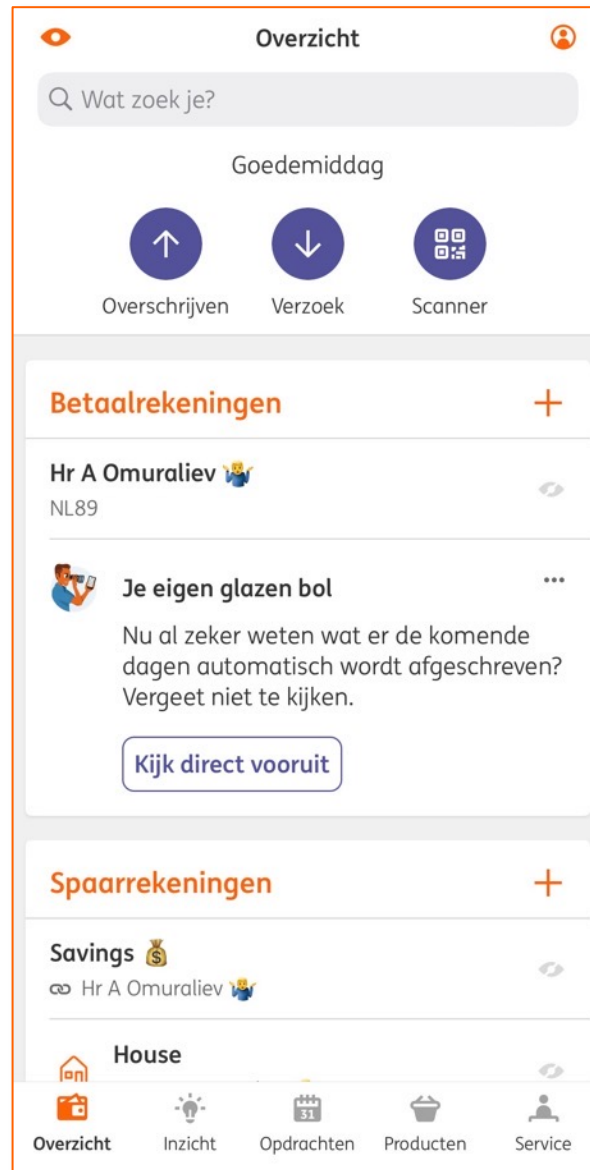
**DataFest**  
Tbilisi 2023

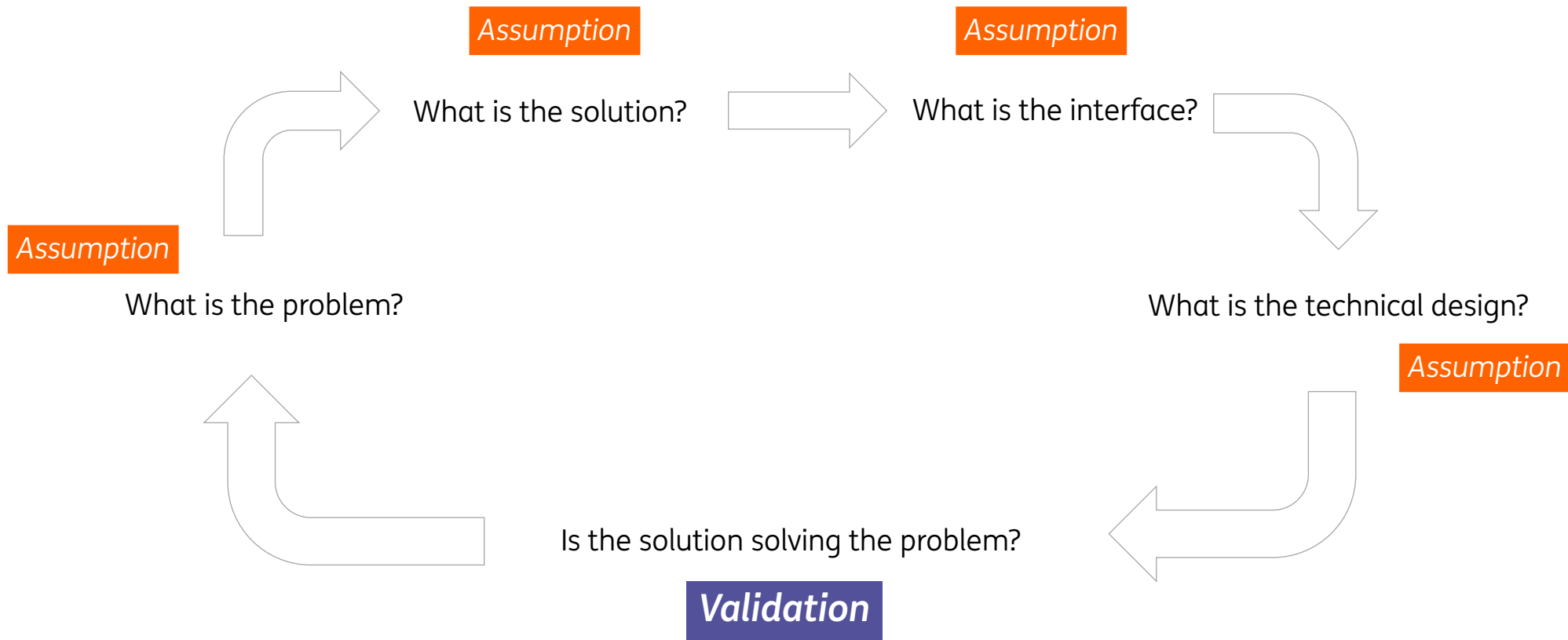


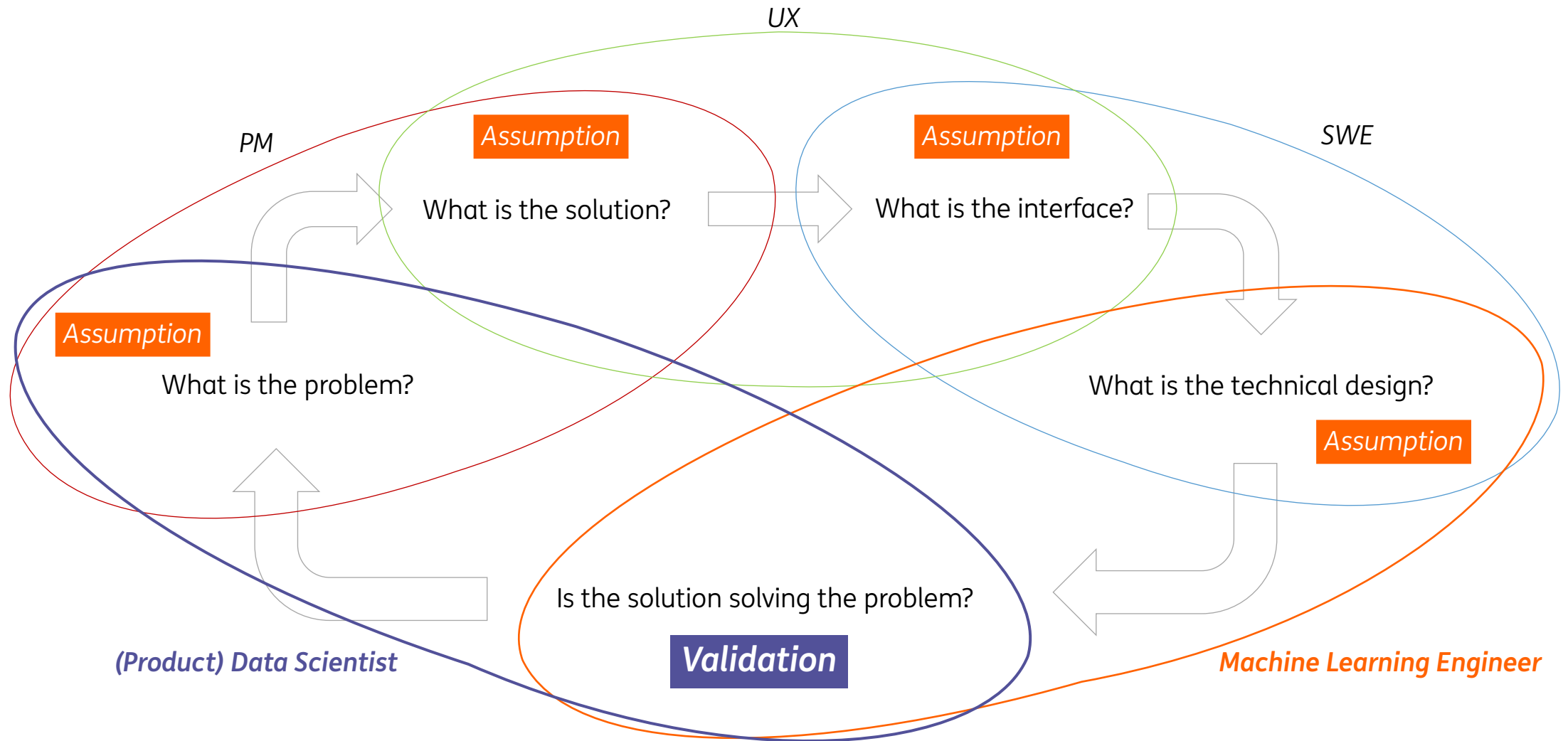
do your thing











Learning 1: We make assumptions on every step. **Validate them**

How to improve your chances of building something that **users want** or **solves the business need**?

Talk to customers

Market research & expert advice

Cycle through faster (early MVPs)

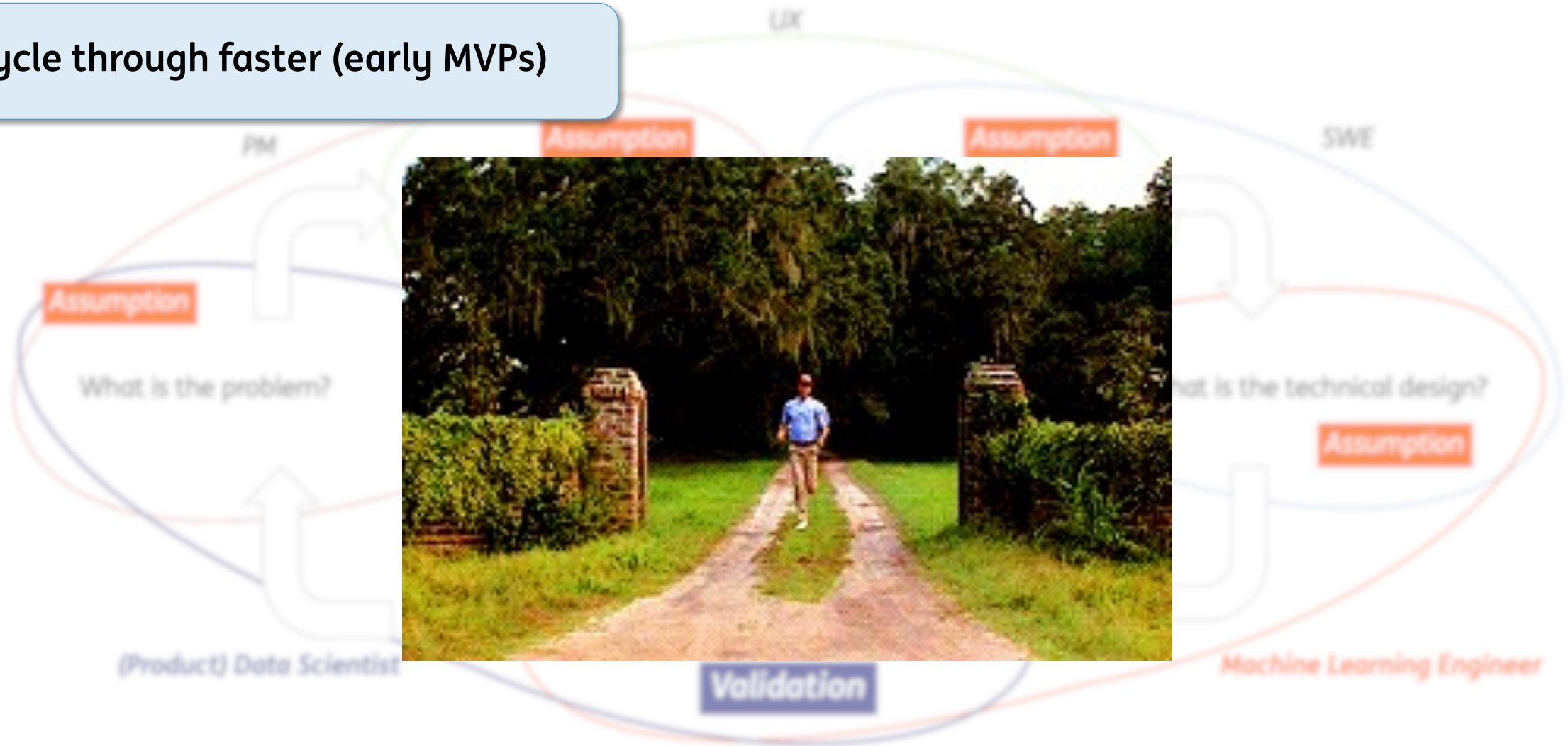
Make a hypothesis, then experiment

Find validation evidence in data

*You should do this  
as a data scientist!*



Cycle through faster (early MVPs)





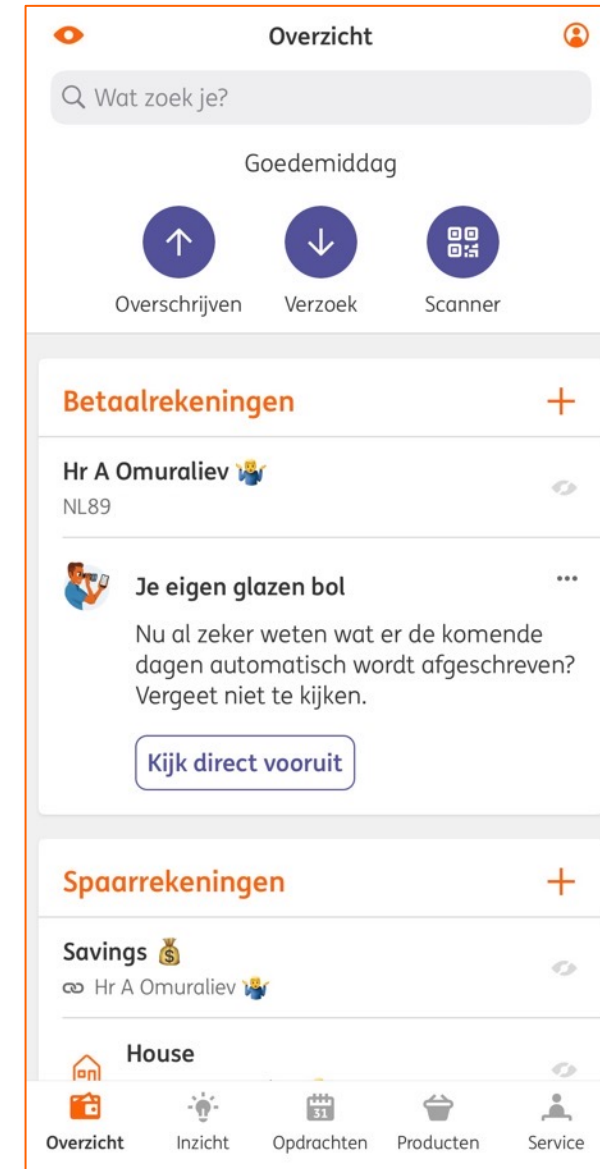
## Case 1: Social ads optimization

MVP cycle: 2 months

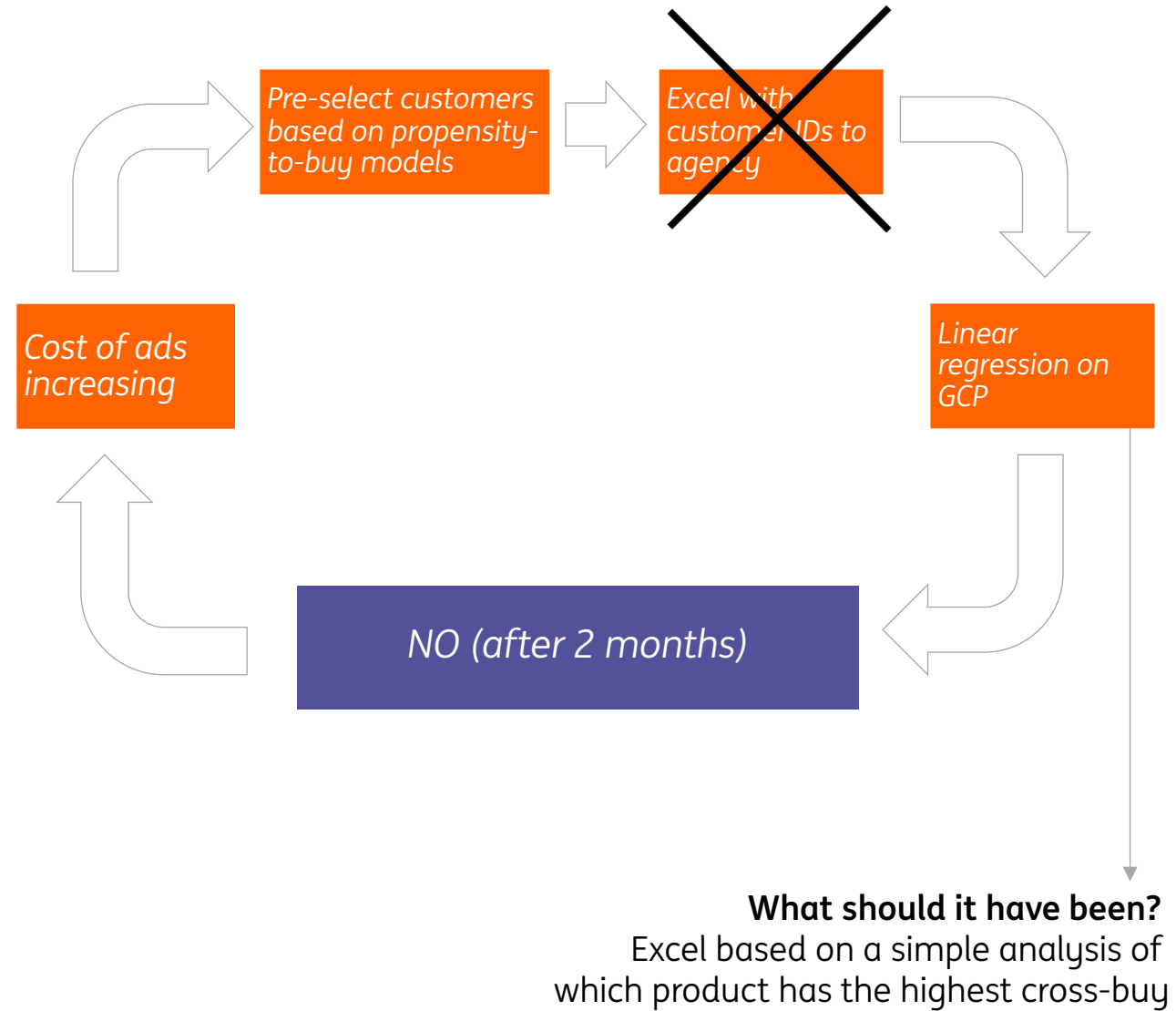


## Case 2: Login intent prediction model

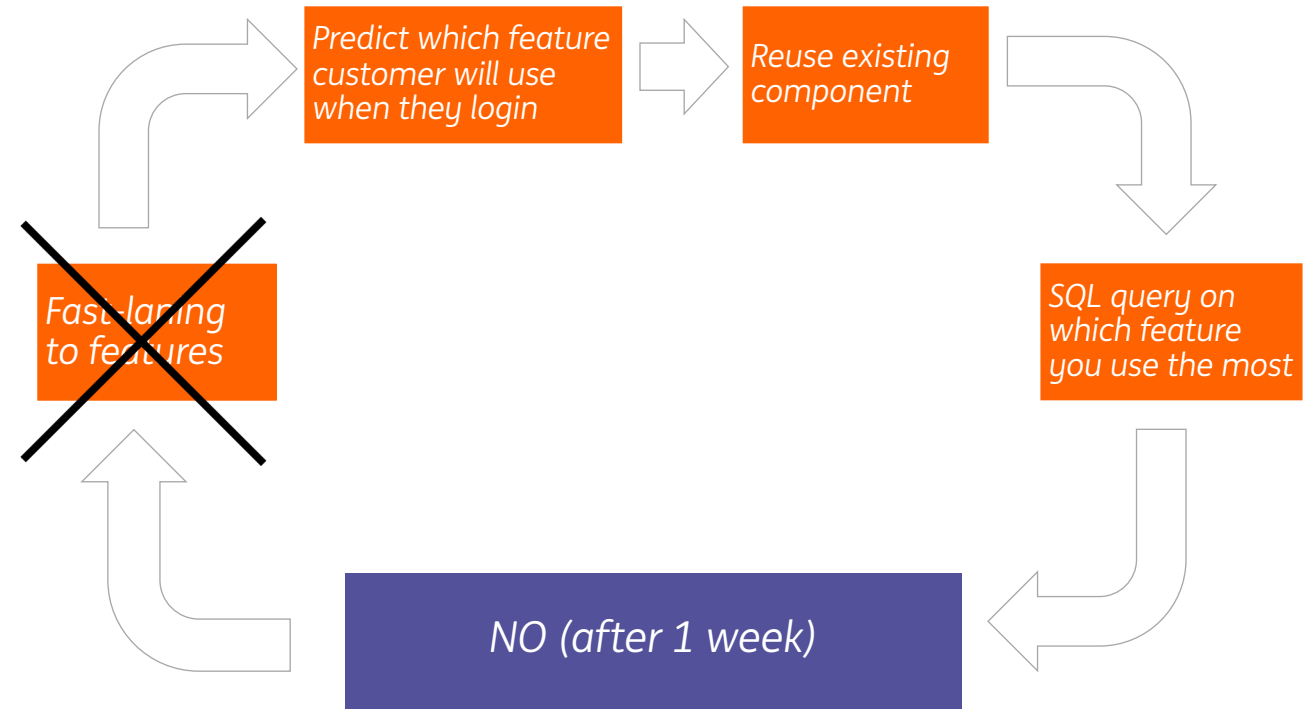
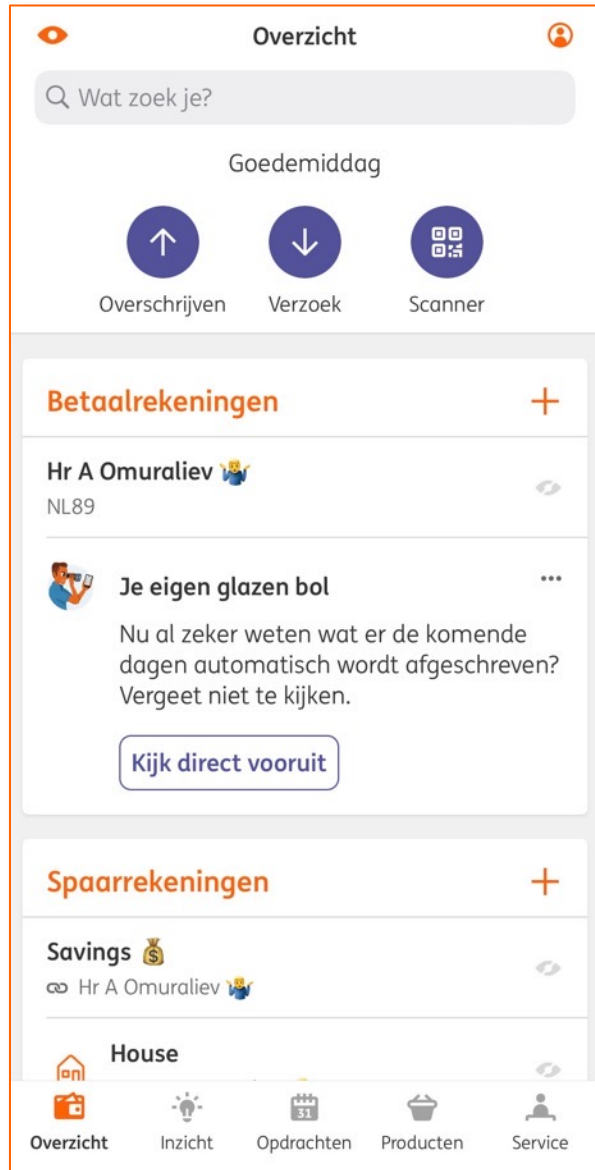
MVP cycle: 1 week



## Case 1: Social ads optimization

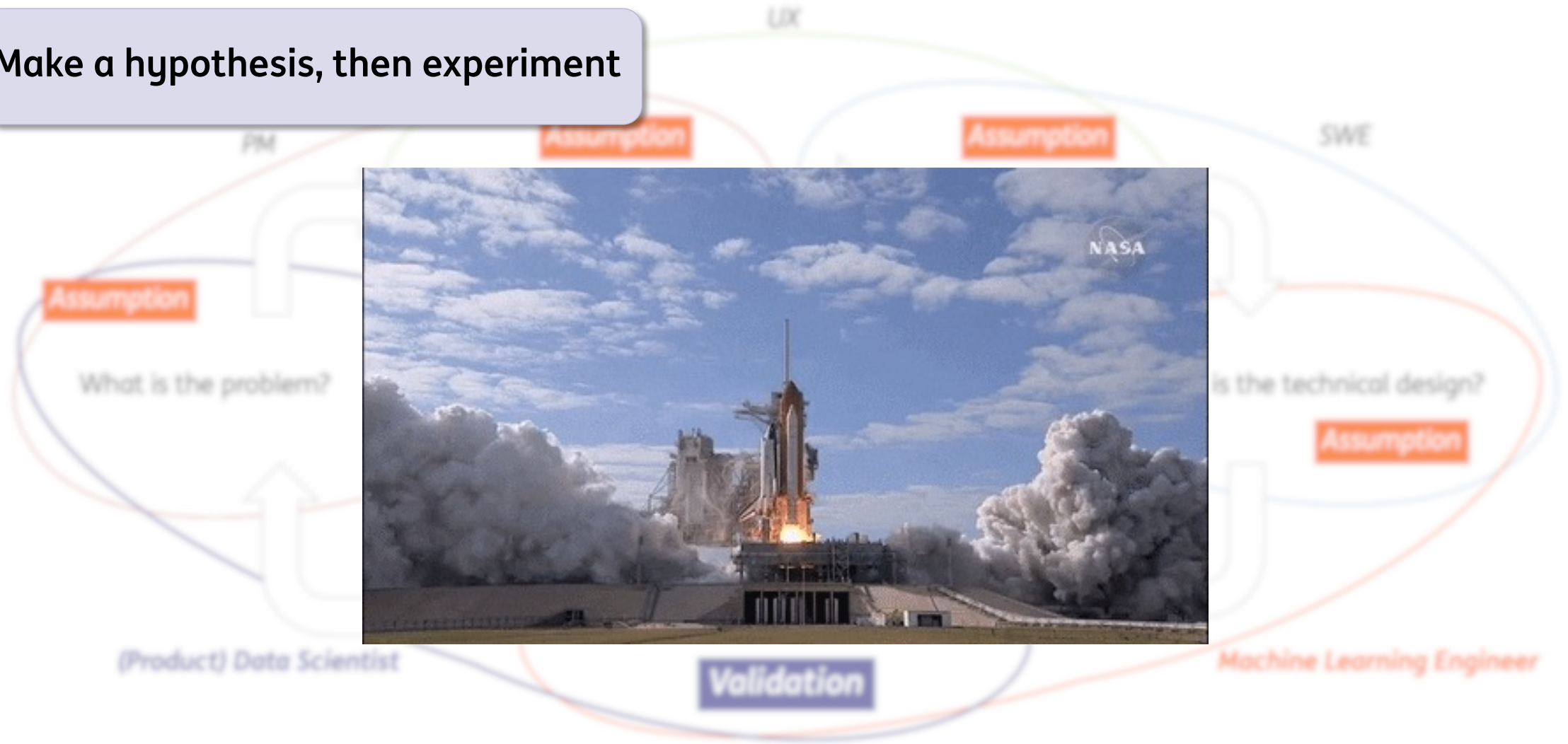


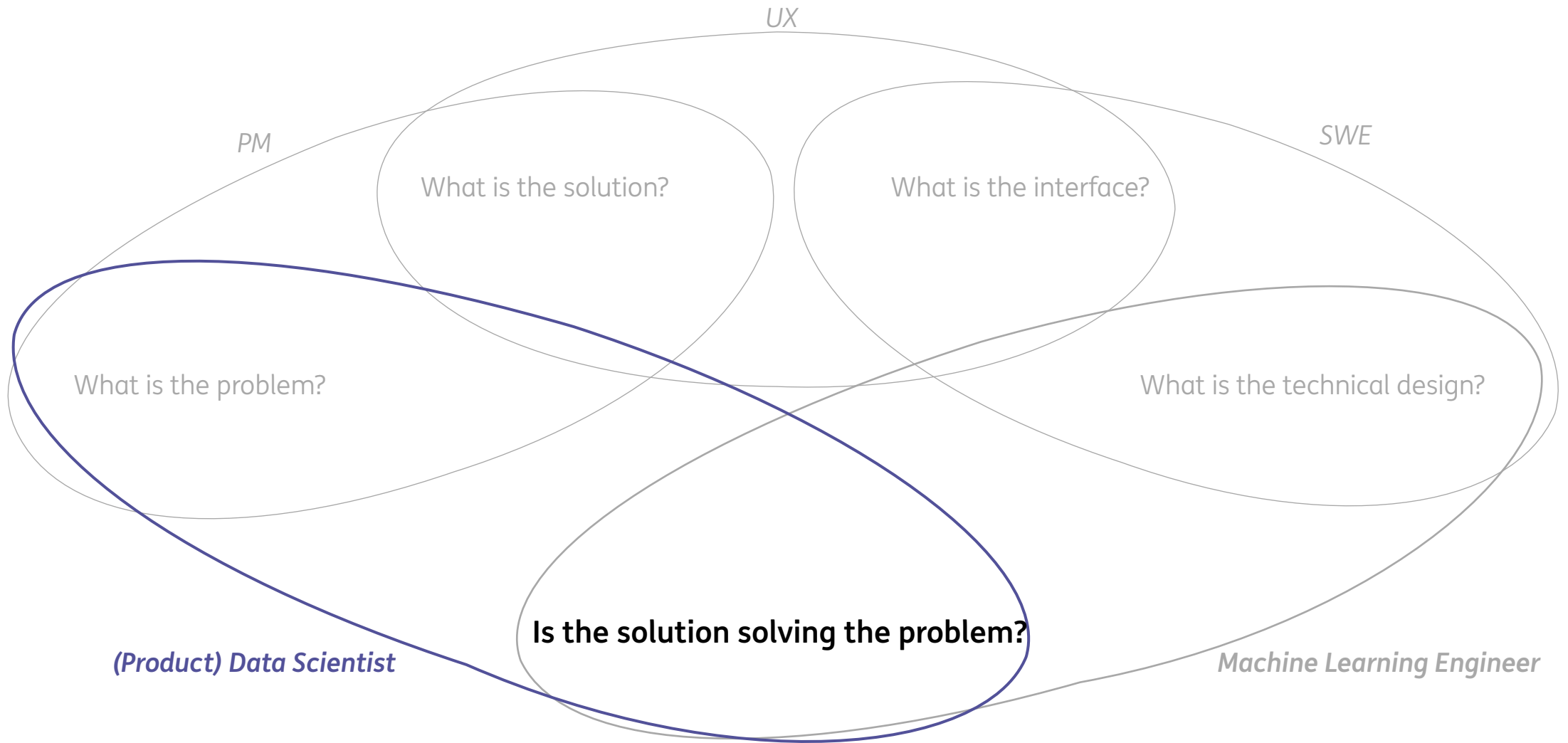
## Case 2: Login intent prediction model




Learning 2: Done imperfectly is **way** better than done late

Make a hypothesis, then experiment





# How to find the next application for your model?

**Start nu met Eenvoudig Beleggen** ...

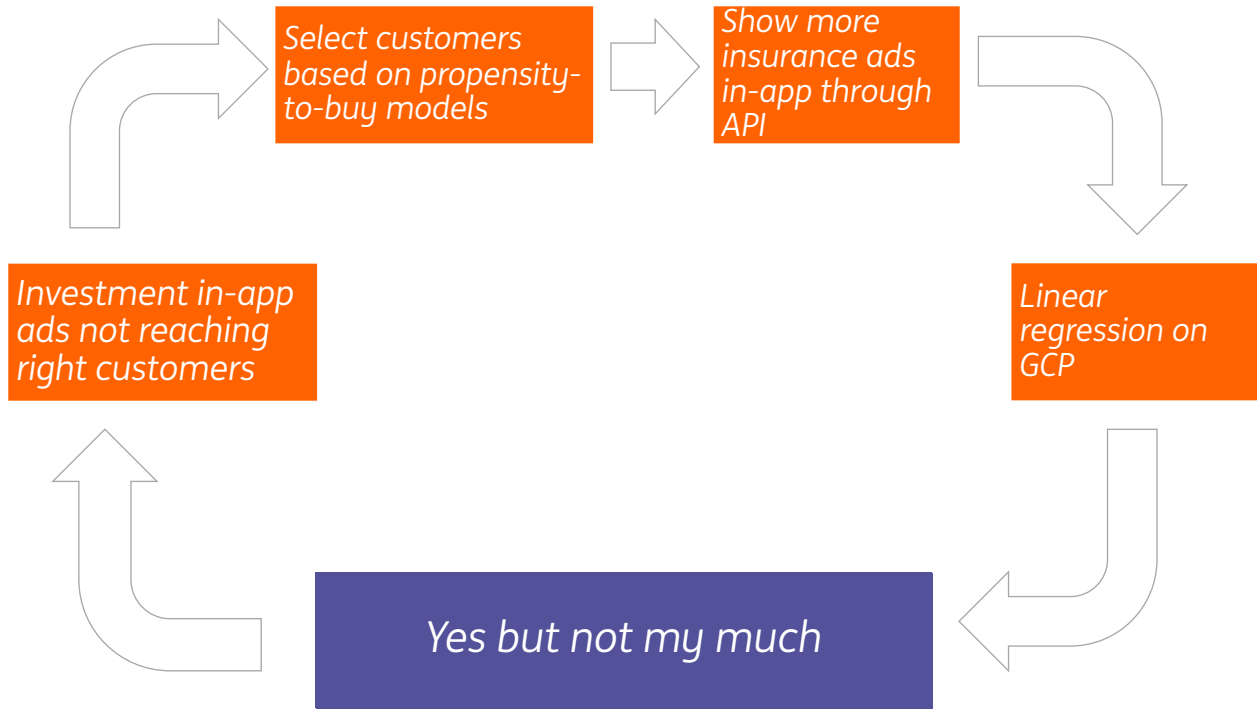
En krijg tijdelijk tot € 50 startgeld cadeau.  
Let op: Beleggen kent risico's en kosten.

[Lees meer over deze actie](#)

**Overweeg je om te beleggen?** ...

Misschien is Eenvoudig Beleggen iets voor jou. Lees er meer over.

[Lees meer](#)

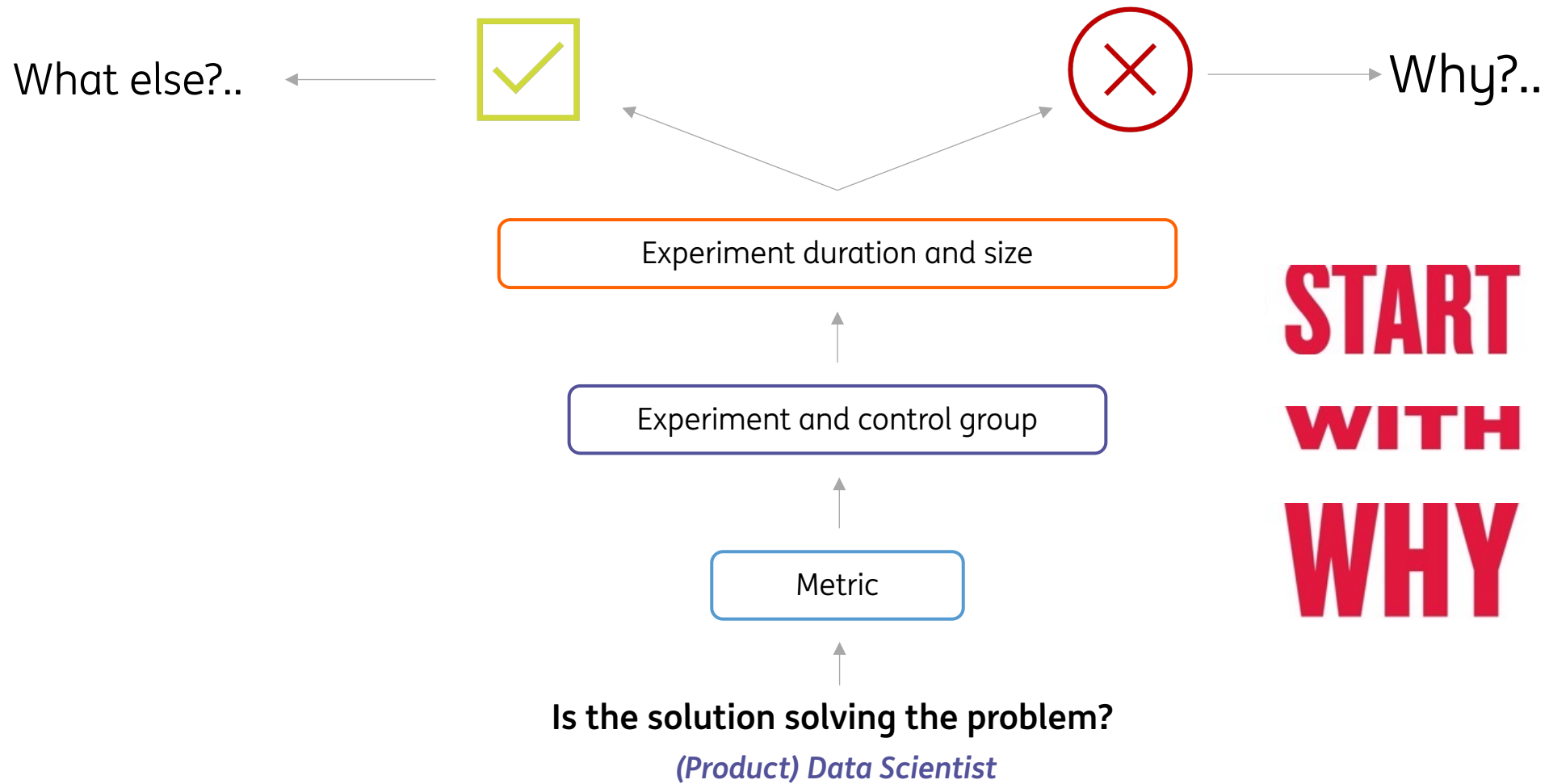


Who converted recently?

Who did not convert?

Cross-sell opportunities?





You know  
what to do next



You know what to check



Start nu met Eenvoudig Beleggen

En krijg tijdelijk tot € 50 startgeld cadeau.  
Let op: Beleggen kent risico's en kosten.

[Lees meer over deze actie](#)

Sales uplift by activating  
customers who had an  
interest in investing

**Next:** analyze customers  
with high propensity who  
did not convert and did  
not show interest in  
investing

Experiment duration and size

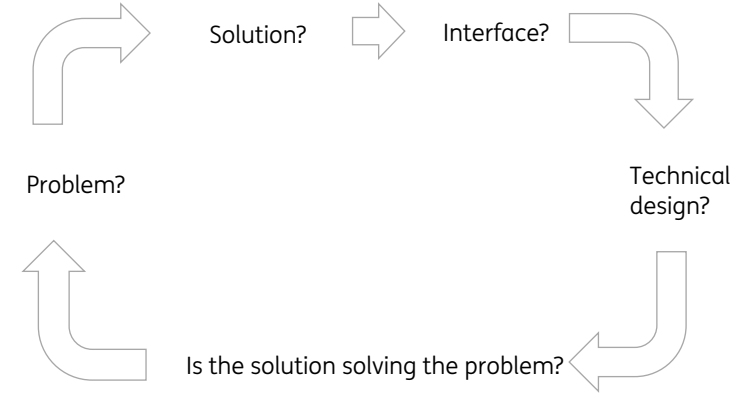
Experiment and control group

Metric

Is the solution solving the problem?

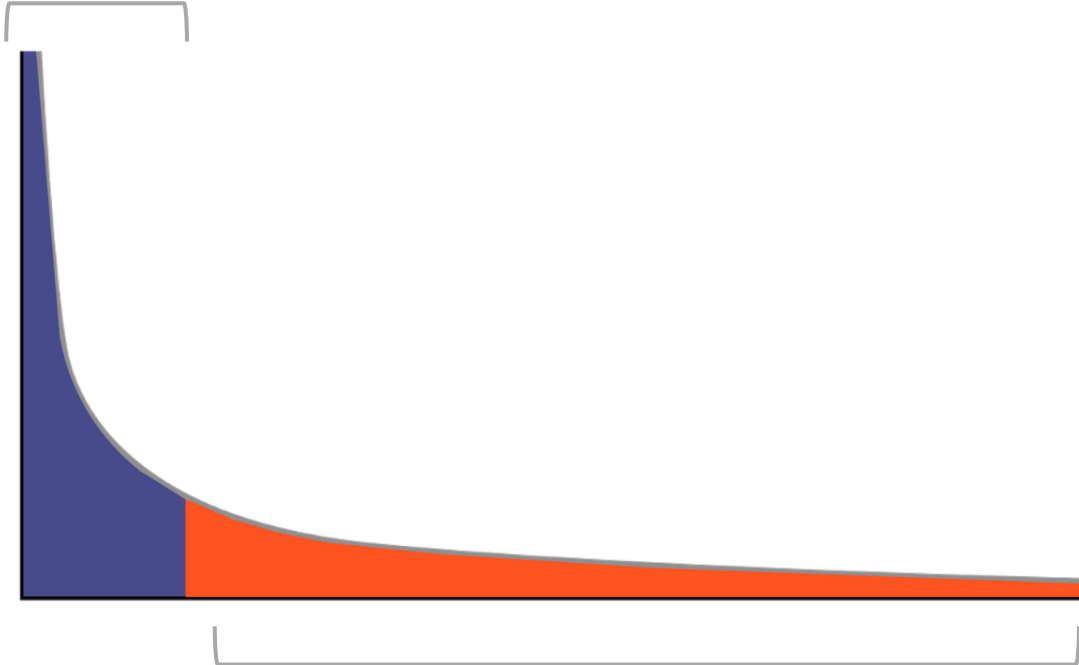
*(Product) Data Scientist*

**Why? (Hypothesis)**



## Power users

Your processes, models and intuition is biased towards them

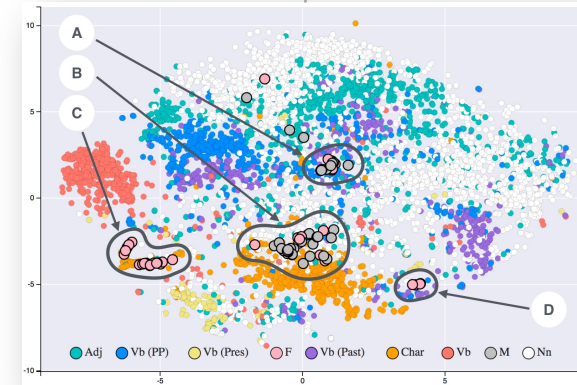
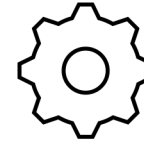


## Long tail

A lot of opportunities lie here as well

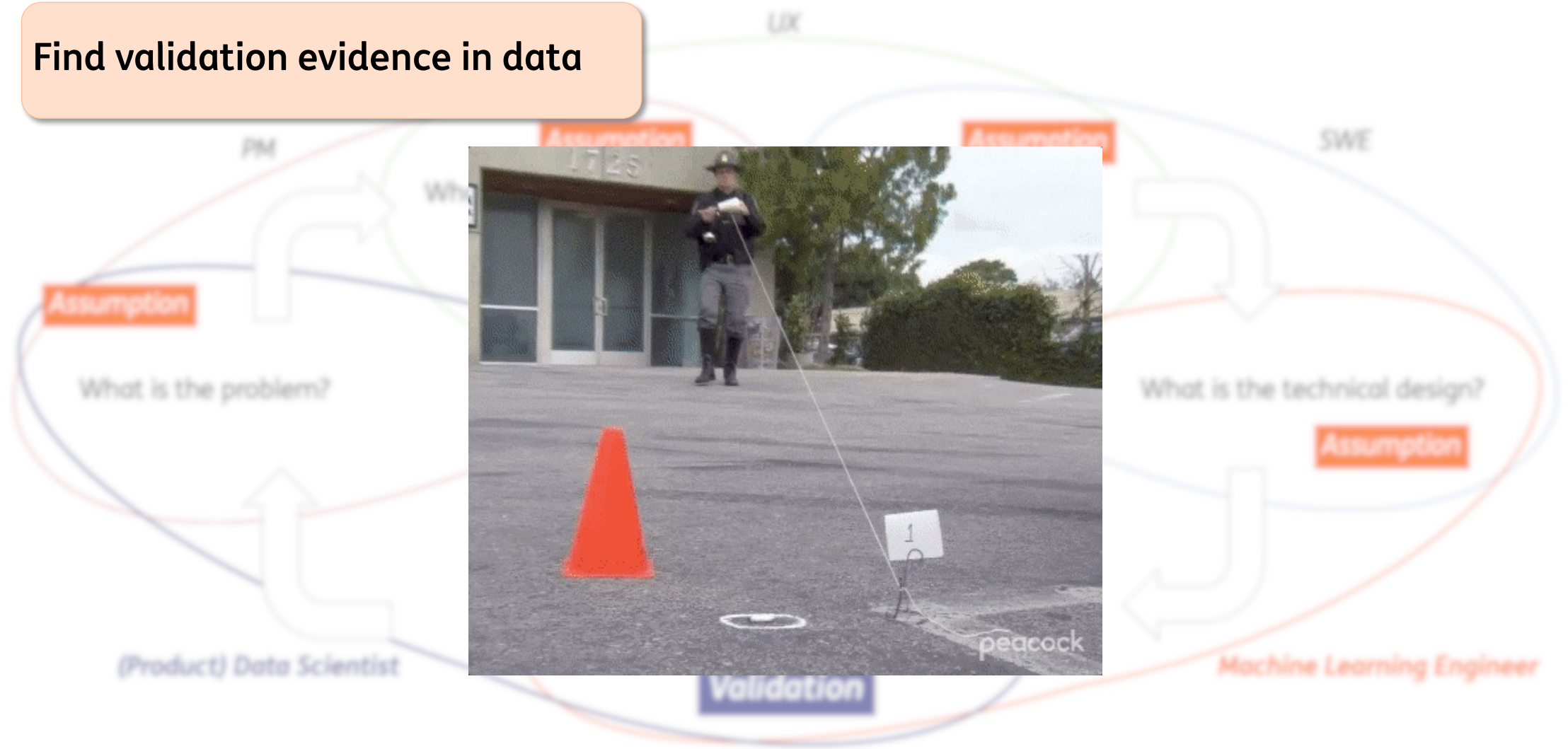
Learning 3: Use your toolset for things **other than** model building

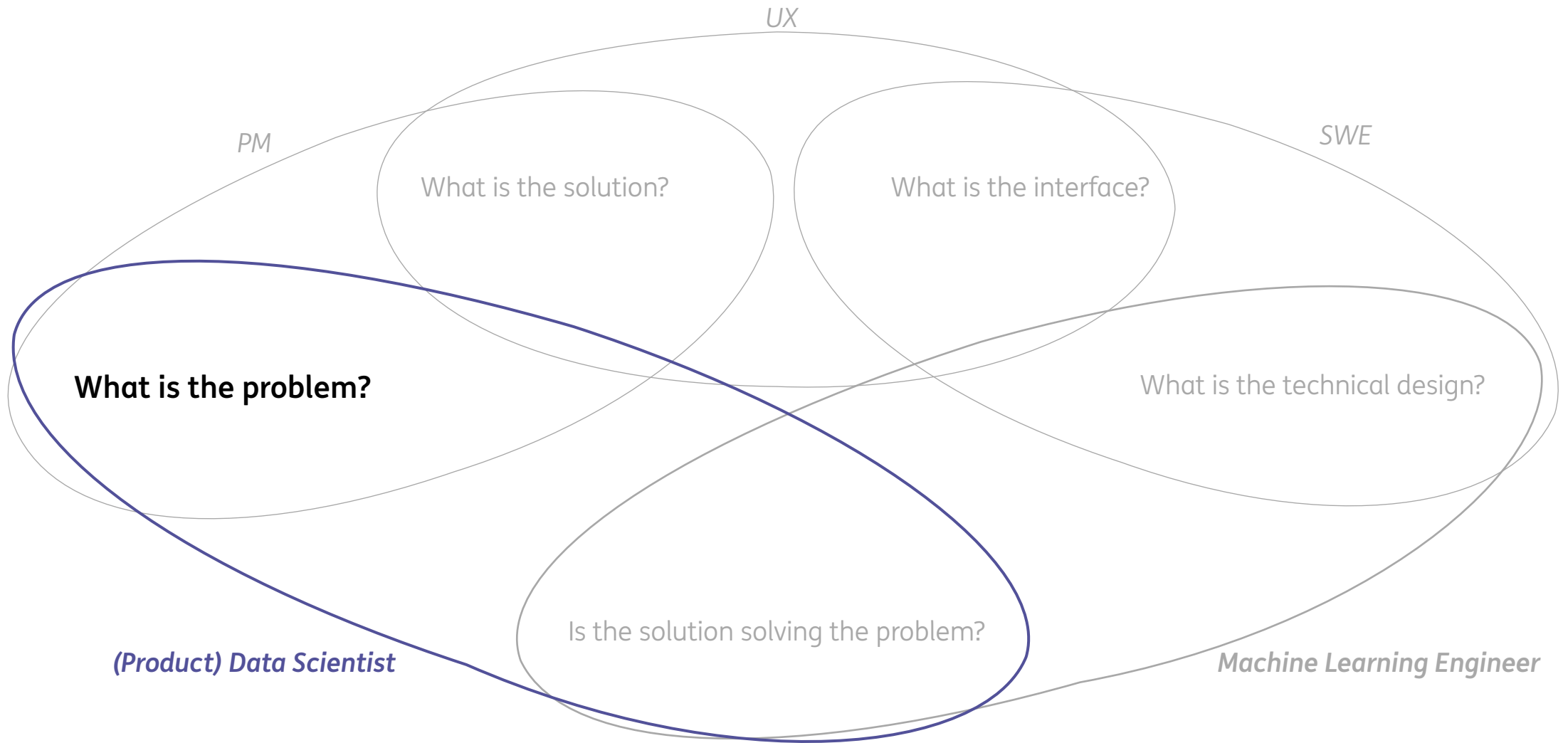
Create software



Analyze problems

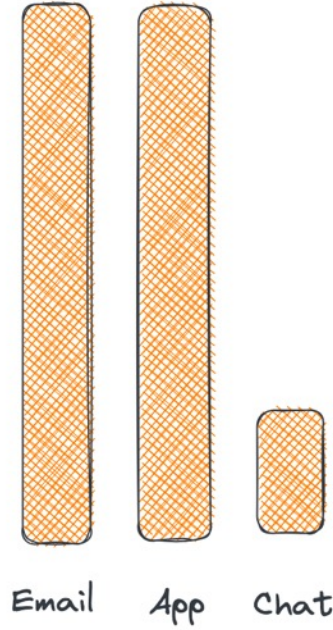
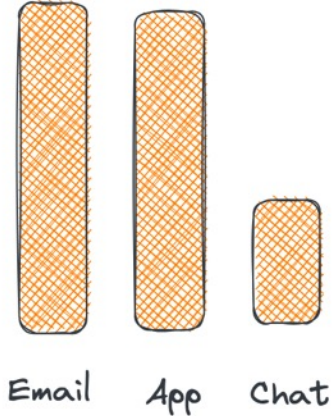
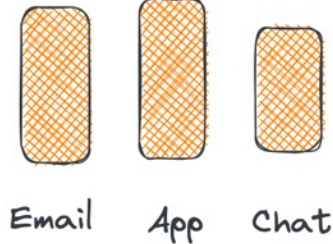
## Find validation evidence in data





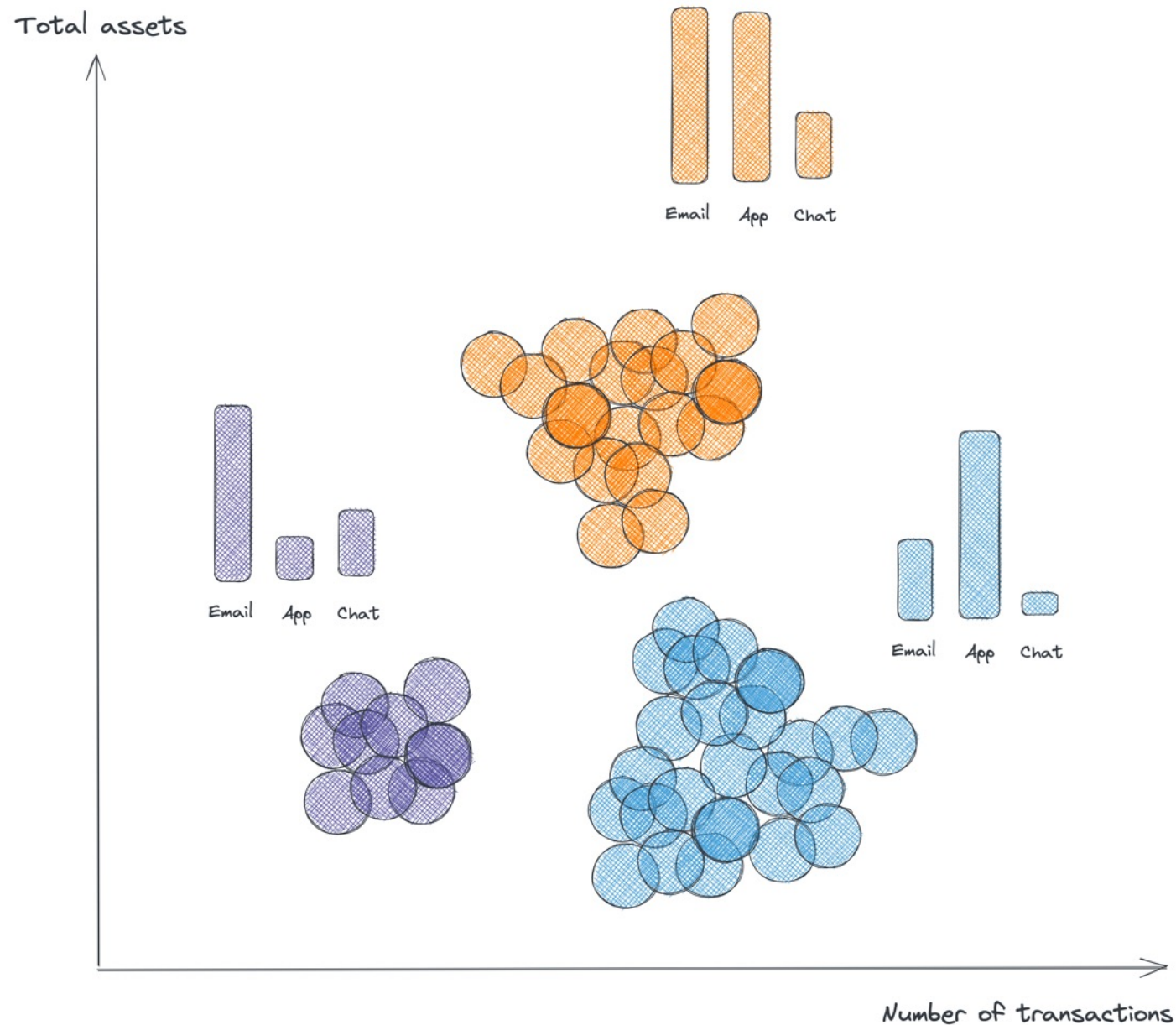
Case: Finding most suitable channel

Engagement rate





# Case: Finding most suitable channel



Do customers have a channel they prefer to be contacted via?

YES (if you control for other factors)

- 1<sup>st</sup> step: K-means clustering on non-channel information
- 2<sup>nd</sup> step: Analyze channel usage per cluster

# Key takeaways

We make assumptions on every step.  
**Validate them**

Done imperfectly is **way** better  
than done late

Use your toolset for things  
**other than** model building

**Slides**



**do your thing**

**LinkedIn:** Azamat Omuraliev

**Twitter:** @azamatomu