

Анализ данных и машинное обучение в задачах HR

Екатерина Денике

преподаватель факультета компьютерных наук НИУ ВШЭ;
менеджер проектов в Группе управления
проектами контента Яндекс.Маркета

Содержание

- 1 | Яндекс: начало большой дружбы HR и ML
- 2 | Google People Analytics: data for HR decisions
- 3 | Про будущее: области HR в которых применяют ML

Яндекс: начало большой дружбы HR и
машинного обучения



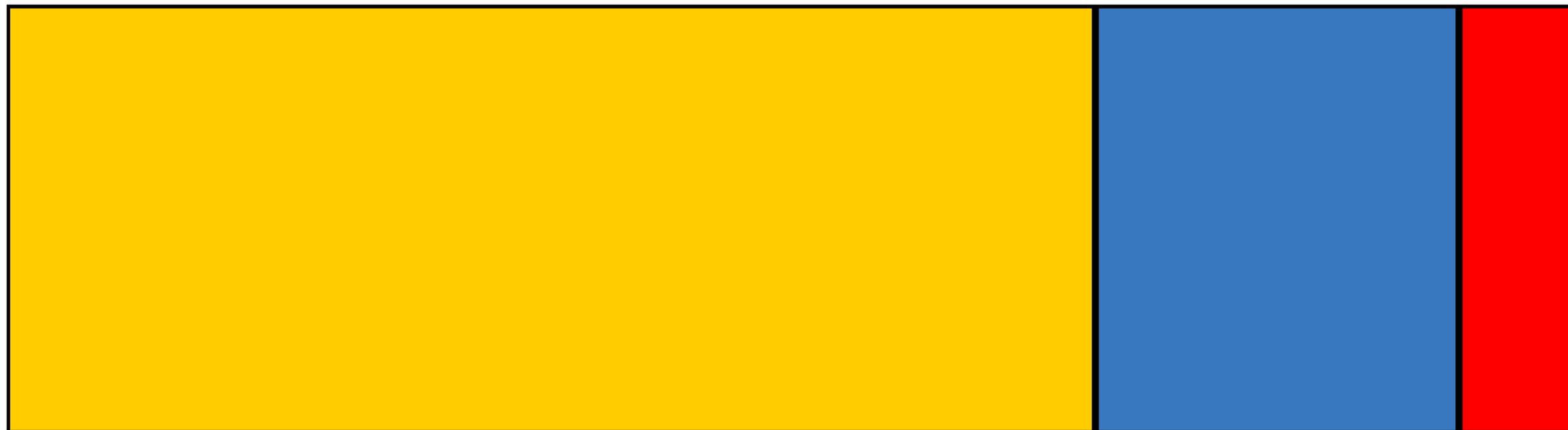
**Изначальная идея: применить интеллект
большого Яндекса в задачах найма**

Что мы делали?

- | Собирали и обрабатывали данные: внутренние, внешние
- | Проводили много экспериментов с данными
- | Обучали модель для предсказания успеха кандидата на собеседовании
- | Обучили модель для показа рекламы потенциально хорошим кандидатам, провели онлайн-эксперимент

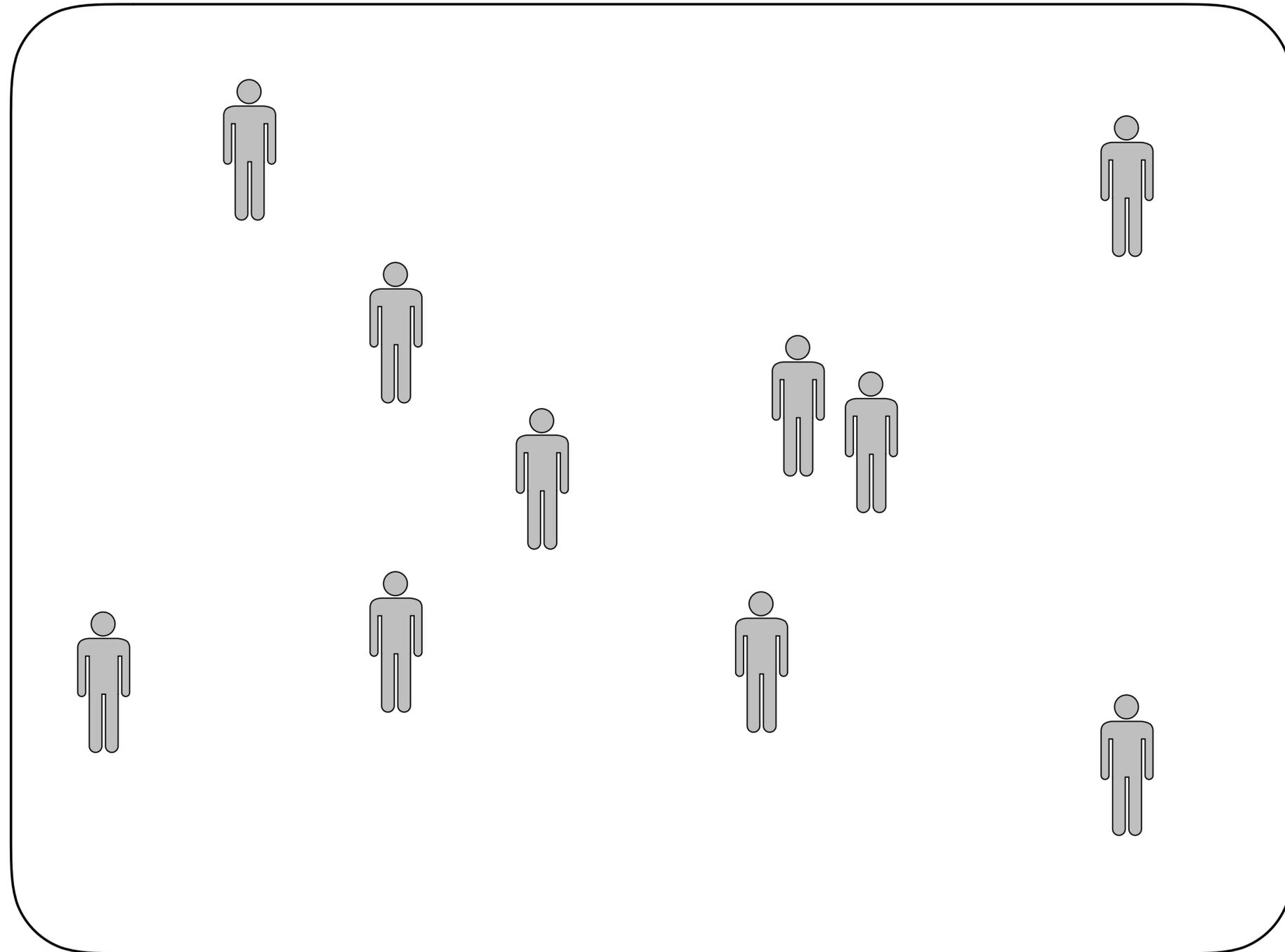
Что мы делали (затраты времени и сил):

■ Сбор данных ■ Обработка данных ■ Обучение модели

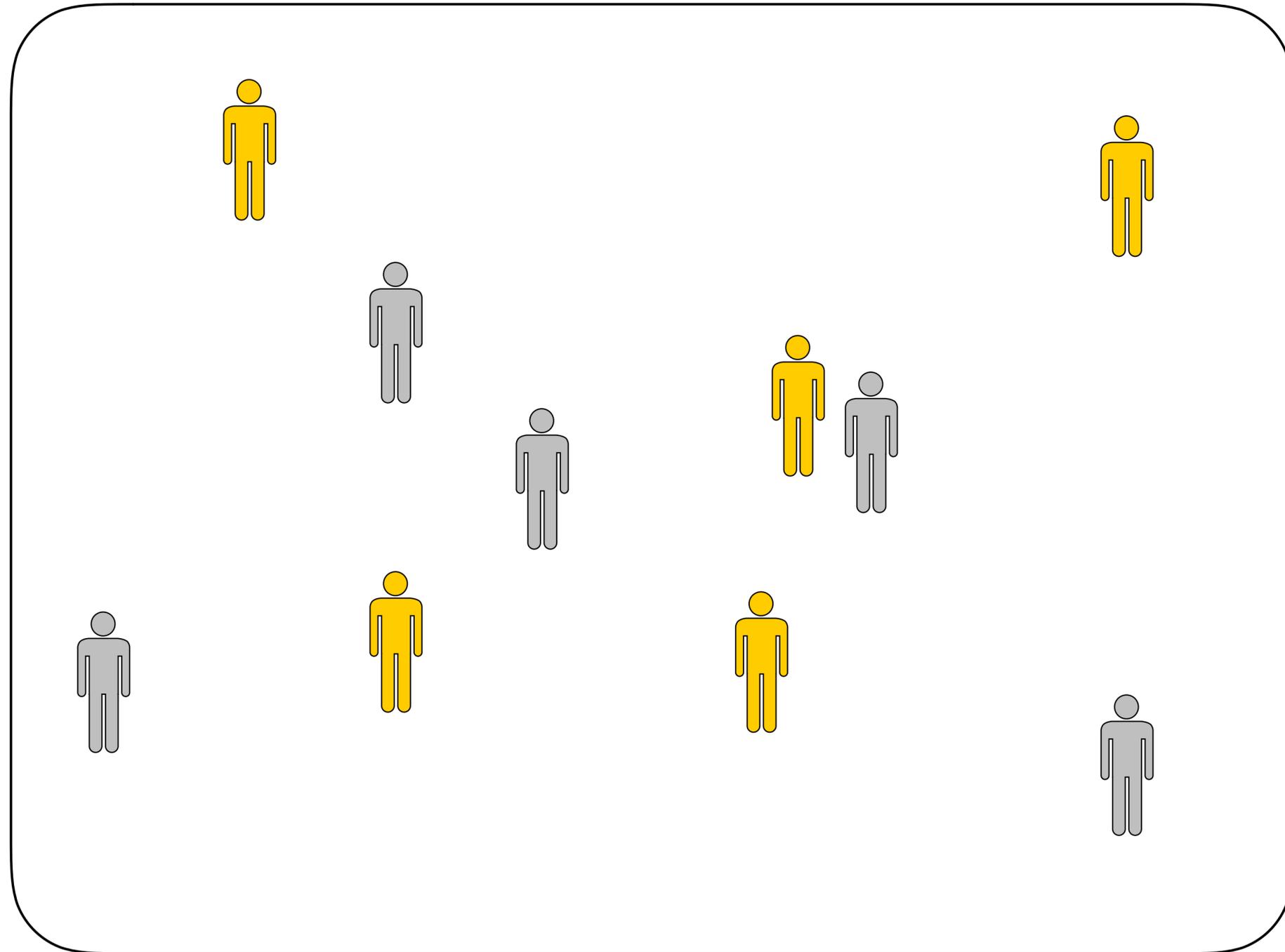


**Задача: по имеющимся данным, определить
вероятность успеха человека на
собеседовании**

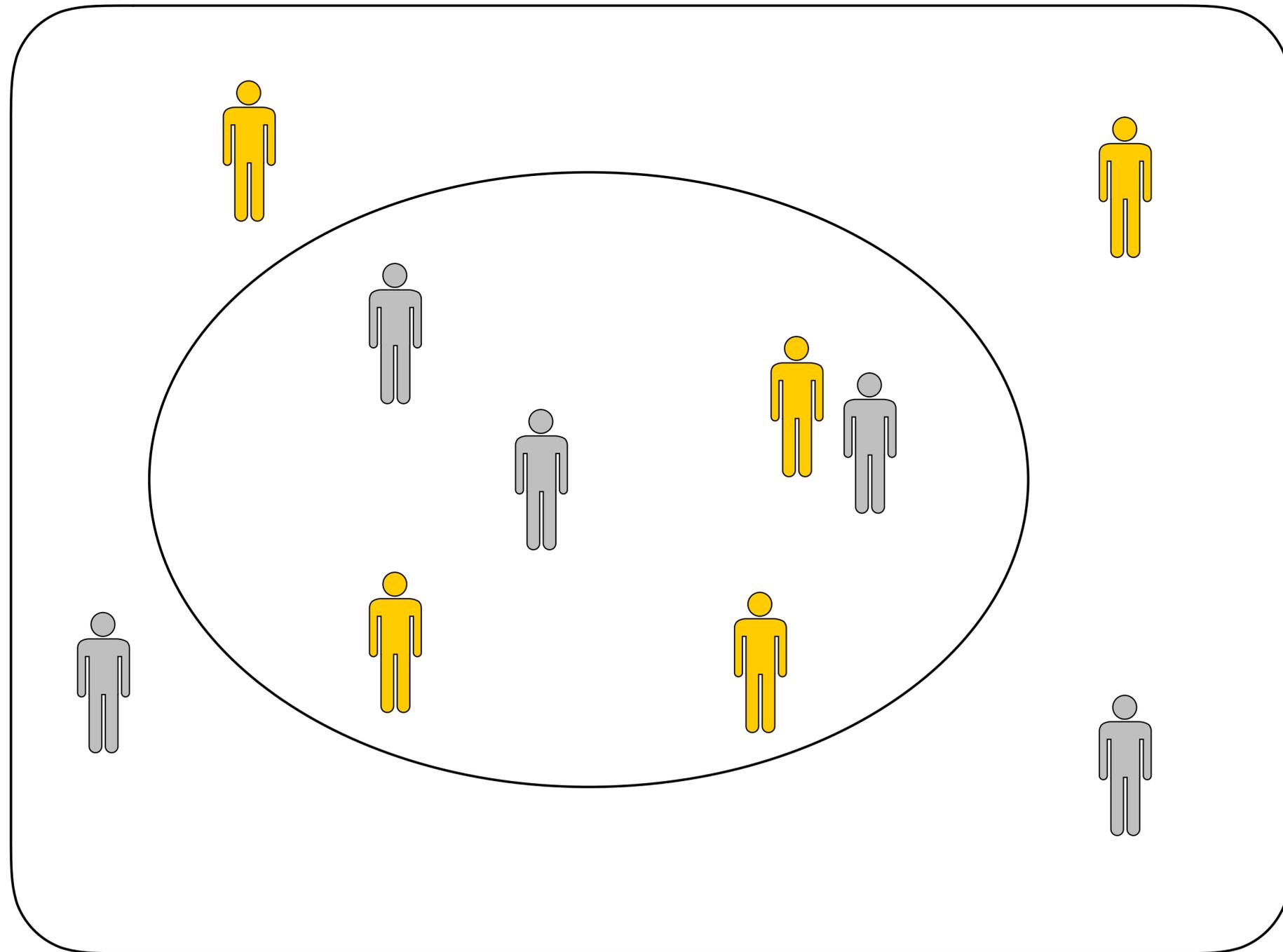
Метрика качества



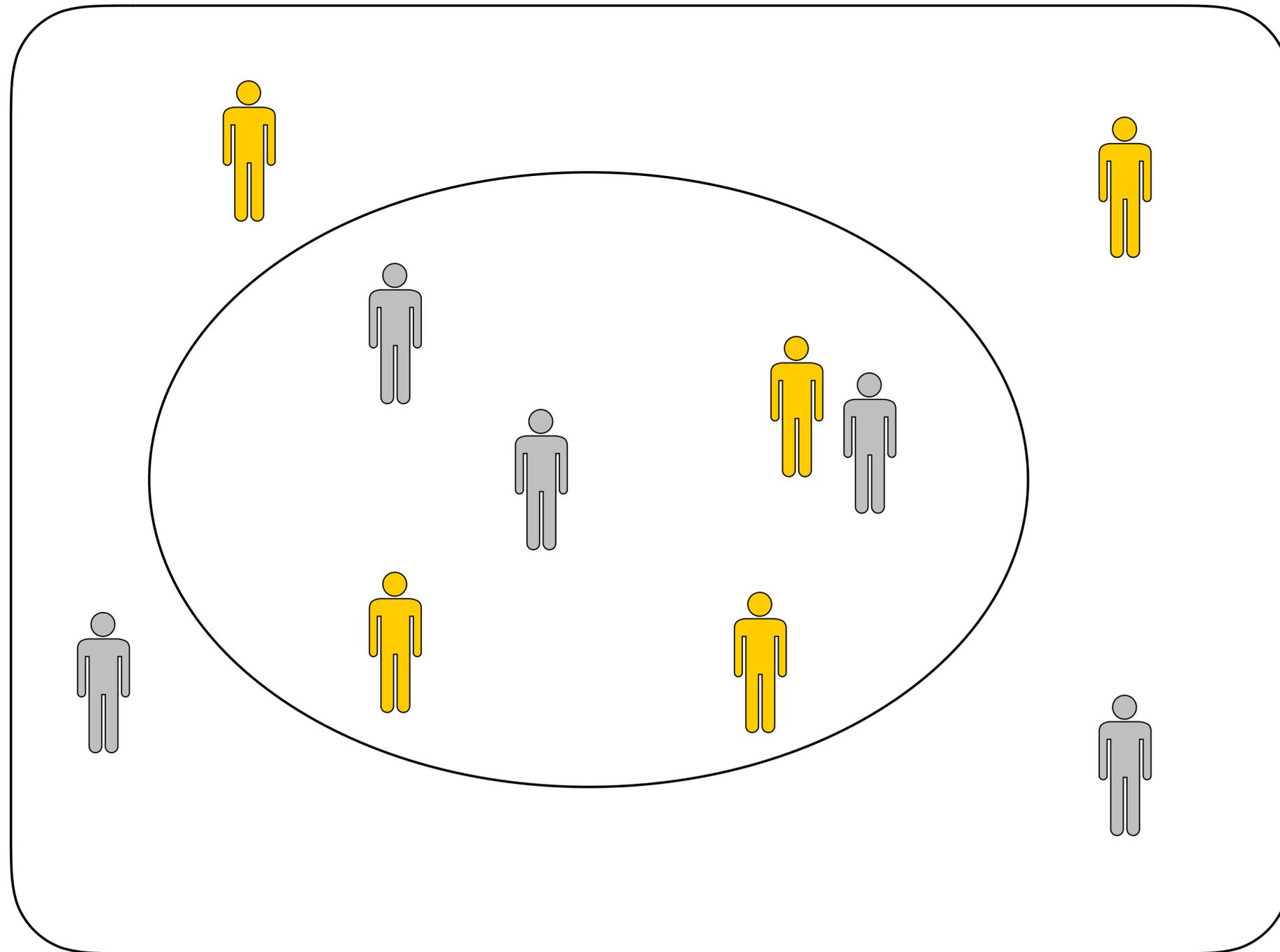
Метрика качества



Метрика качества



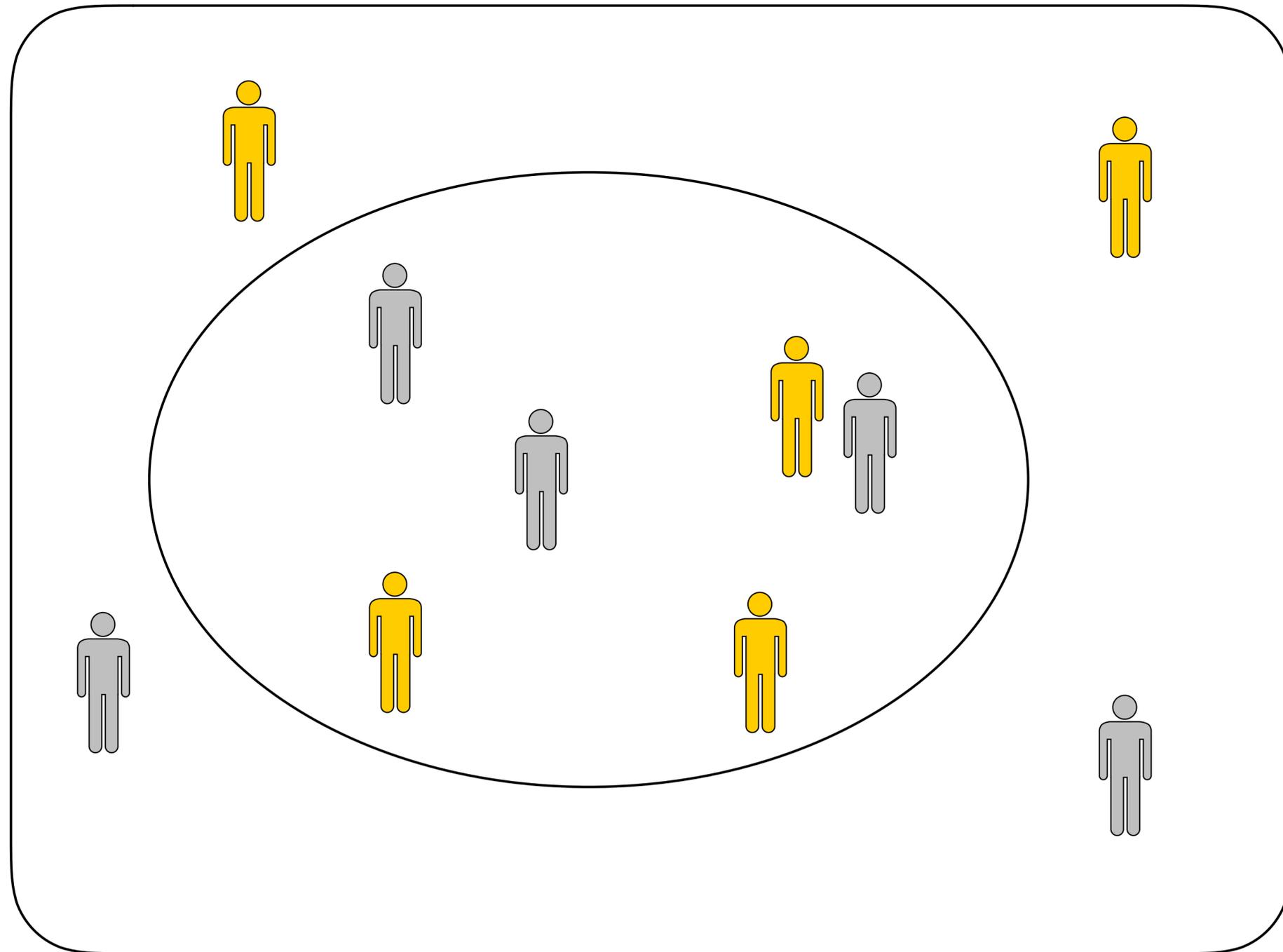
Метрика качества



Точность:

$$\frac{\text{●}}{\text{○}} = \frac{3}{6} = 0,5$$

Метрика качества



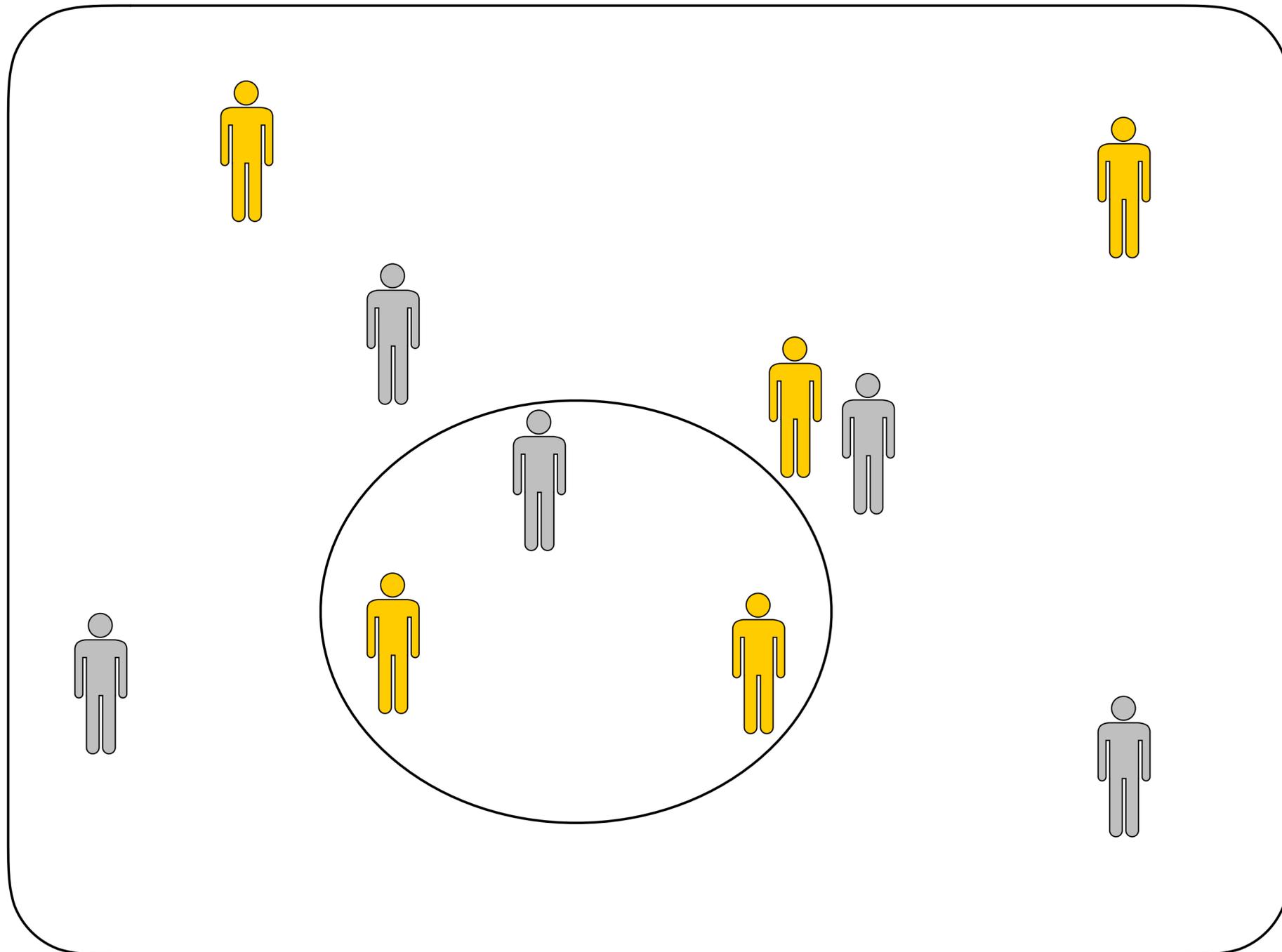
Точность:

$$\frac{\text{●}}{\text{○}} = \frac{3}{6} = 0,5$$

Полнота:

$$\frac{\text{●}}{\text{●}} = \frac{3}{5} = 0,6$$

Метрика качества



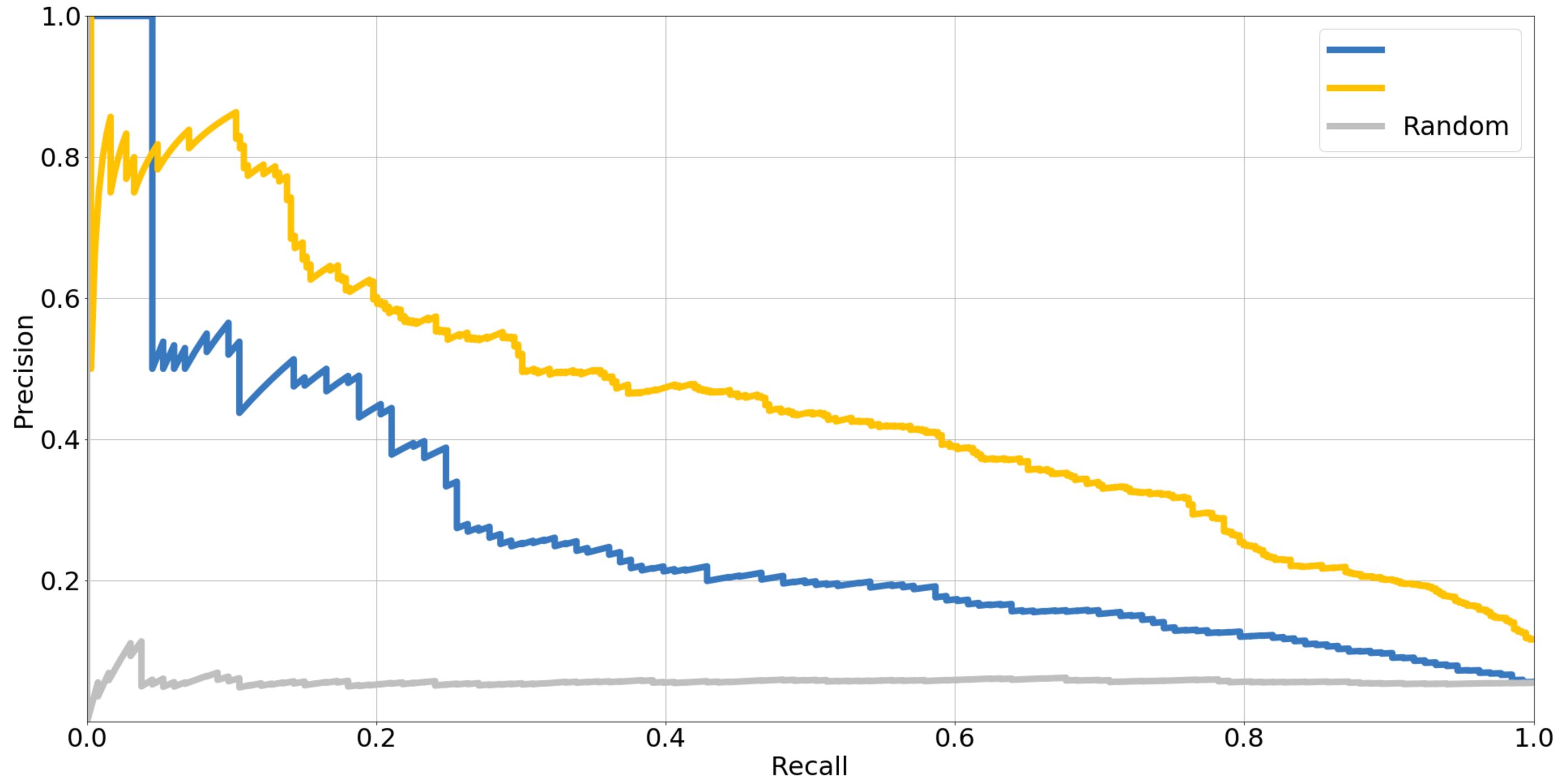
Точность:

$$\frac{\text{●}}{\text{○}} = \frac{2}{3} = 0,7$$

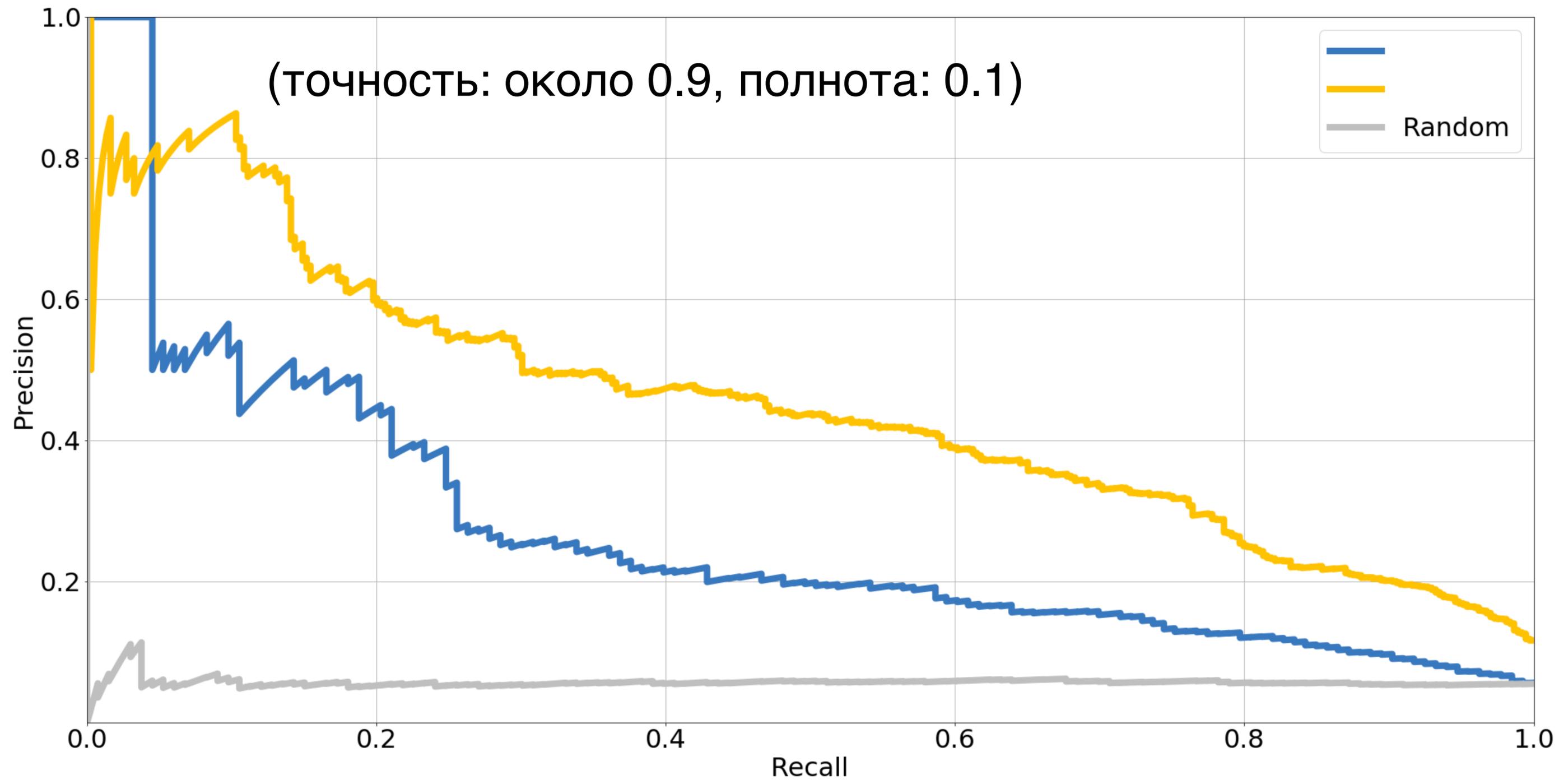
Полнота:

$$\frac{\text{●}}{\text{●}} = \frac{2}{5} = 0,4$$

Метрика качества



Метрика качества



**| Какие еще полезные
результаты
получаются после
обучения модели?**

Важные признаки для модели!

В нашем случае: поведение в интернете,
участие в школах яндекса, победа на топовых
олимпиадах...

**Задача: эксперимент с рекламой, научиться
по показам рекламы приводить успешных
кандидатов в Яндекс**

Проведение эксперимента

Главное в СМИ в Москве 7 июля, пятница 12:04

1. Суд арестовал обвиняемую в резонансном ДТП с «пьяным» ребенком
2. Сенаторы призвали Трампа не возвращать РФ изъятую дипсобственность
3. Siemens начала расследование из-за сообщений о поставках турбин в Крым
4. Путин прибыл в Гамбург на саммит G20
5. СМИ узнали о согласии США оставить Асада у власти

USD ЦБ 60,38 +0,14 EUR ЦБ 68,38 НЕФТЬ 46,85 -2,21% ...



Попробуйте стать разработчиком Яндекса

[Карты](#) [Маркет](#) [Новости](#) [Переводчик](#) [Картинки](#) [Видео](#) [Музыка](#) [ещё](#)

Яндекс

Найти

Найдётся всё. Например, сколько калорий в банане

Проведение эксперимента

Главное в СМИ в Москве 7 июля, пятница 12:04

1. Суд арестовал обвиняемую в резонансном ДТП с «пьяным» ребенком
2. Сенаторы призвали Трампа не возвращать РФ изъятую дипсобственность
3. Siemens начала расследование из-за сообщений о поставках турбин в Крым
4. Путин прибыл в Гамбург на саммит G20
5. СМИ узнали о согласии США оставить Асада у власти

USD ЦБ 60,38 +0,14 EUR ЦБ 68,38 НЕФТЬ 46,85 -2,21% ...



Попробуйте стать разработчиком Яндекса

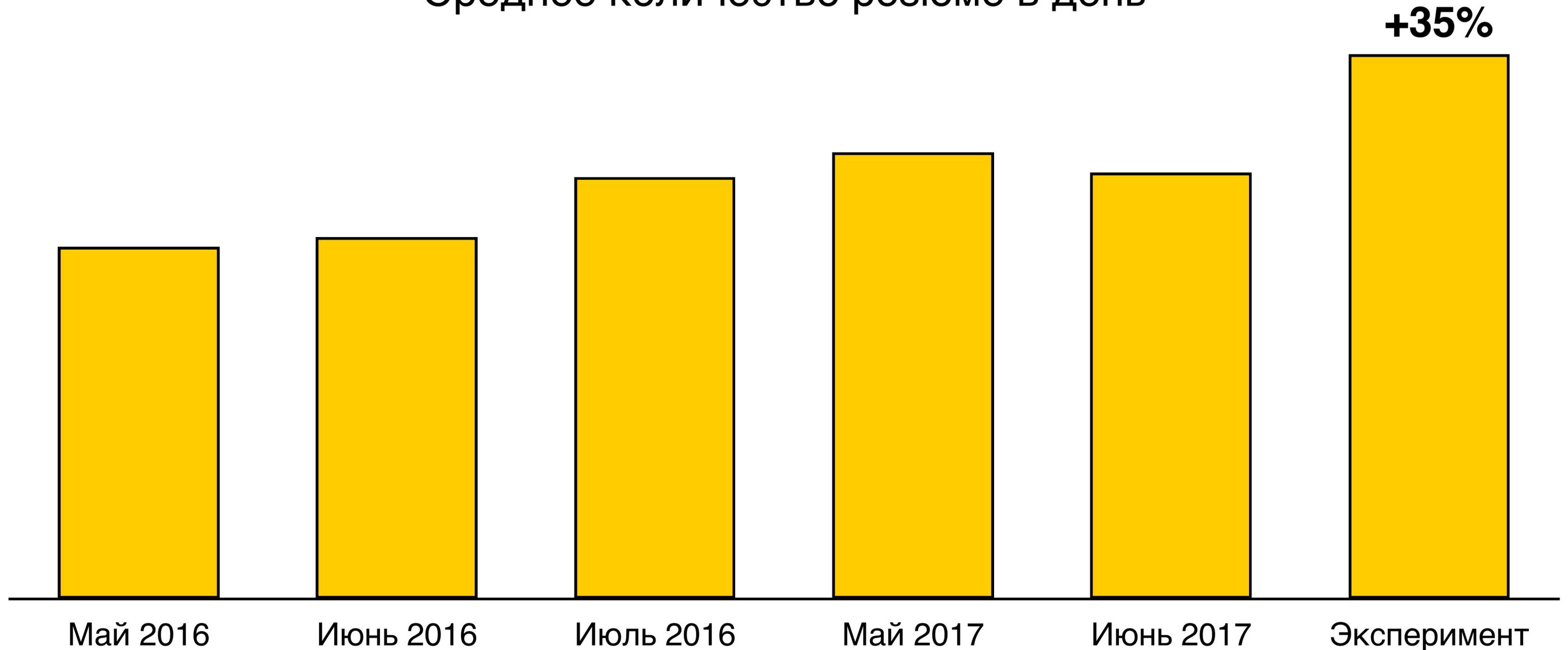
[Карты](#) [Маркет](#) [Новости](#) [Переводчик](#) [Картинки](#) [Видео](#) [Музыка](#) [ещё](#)

Яндекс

Найдётся всё. Например, сколько калорий в банане

Отправка формы

Среднее количество резюме в день



**| Про процесс и самые
интересные моменты**

Наличие и качество данных - залог успеха :)

■ Сбор данных ■ Обработка данных ■ Обучение модели



Хорошие данные для модели

- › долгий систематичный сбор данных, сильно пересекающийся с теми, которые вы хотите анализировать
- › в нашем случае проект, например проект «Люди», данные из школ

Данные, которые не получится использовать для модели

- › плохо структурированные данные (описание, текст)
- › неоднородные данные
- › данные мало пересекающиеся с вашей выборкой
- › в нашем случае, данные про ВУЗы и amazing hiring

**| Отдельно про пример с
вузами**

| Что сложного, давайте посчитаем?

пробегающий мимо аналитик

Не менее 283 различных написаний МГУ:

- > 374 Московский государственный университет имени М. В. Ломоносова
- > 108 МГУ им. Ломоносова
- > 105 МГУ
- > 81 Московский государственный университет им. М.В. Ломоносова, Москва
- > 75 Московский государственный университет им. М.В. Ломоносова
- > ...
- > 1 Lomonosov Moscow State University (MSU) Master's degree, Applied Mathematics and Computer Science, Probability Theory and Mathematical Statistics, graduated with honours
- > 1 МГУ им. М.В. Ломоносова ВМК, математик, системный программист, Специалист, прикладная математика и информатика
- > 1 МГУ им. Ломоносова НИУ ВШЭ ШАД

Не менее 75 различных написаний ИТМО (и каких...):

- › 34 Санкт-Петербургский государственный университет информационных технологий, механики и оптики
- › 14 Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, Санкт-Петербург
- › 12 Университет ИТМО
- › 7 НИУ ИТМО
- › ...

- › 1 Saint-Petersburg State University Information Technologies, Mechanic and Optics (University ITMO) Master of Technology (M.Tech.), Embedded System Design

- › 1 Double degree of University ITMO and University of Rostock

- › 1 Ленинградский ордена Трудового Красного Знамени технологический институт холодильной промышленности

Как с этим справились?

Ручная разметка
практикантами (2000 экз.)

Список министерства
образования

Список в рейтинге
Webometrics

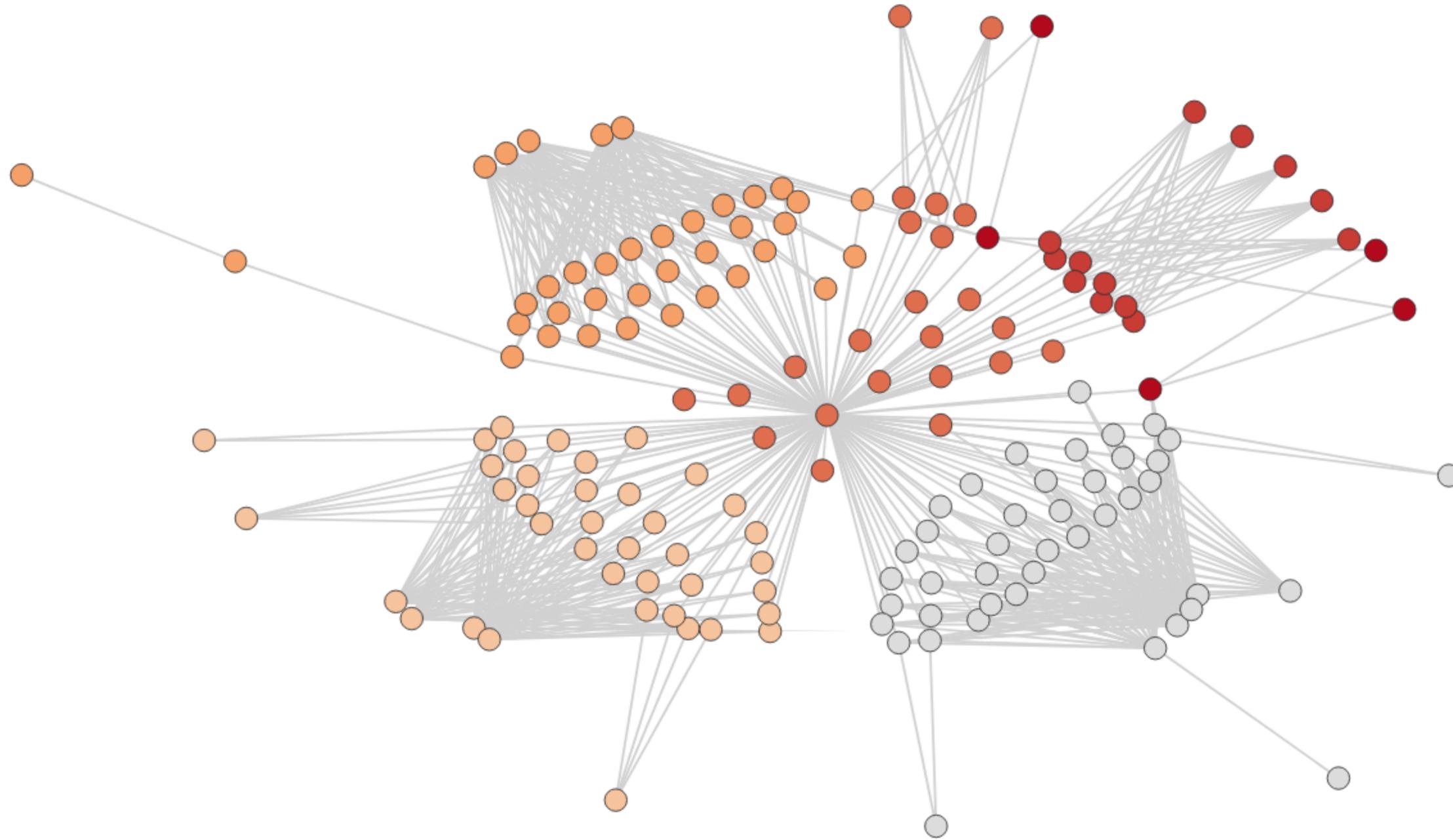
Наш список
со всеми
вариантами
написания

Опечаточник

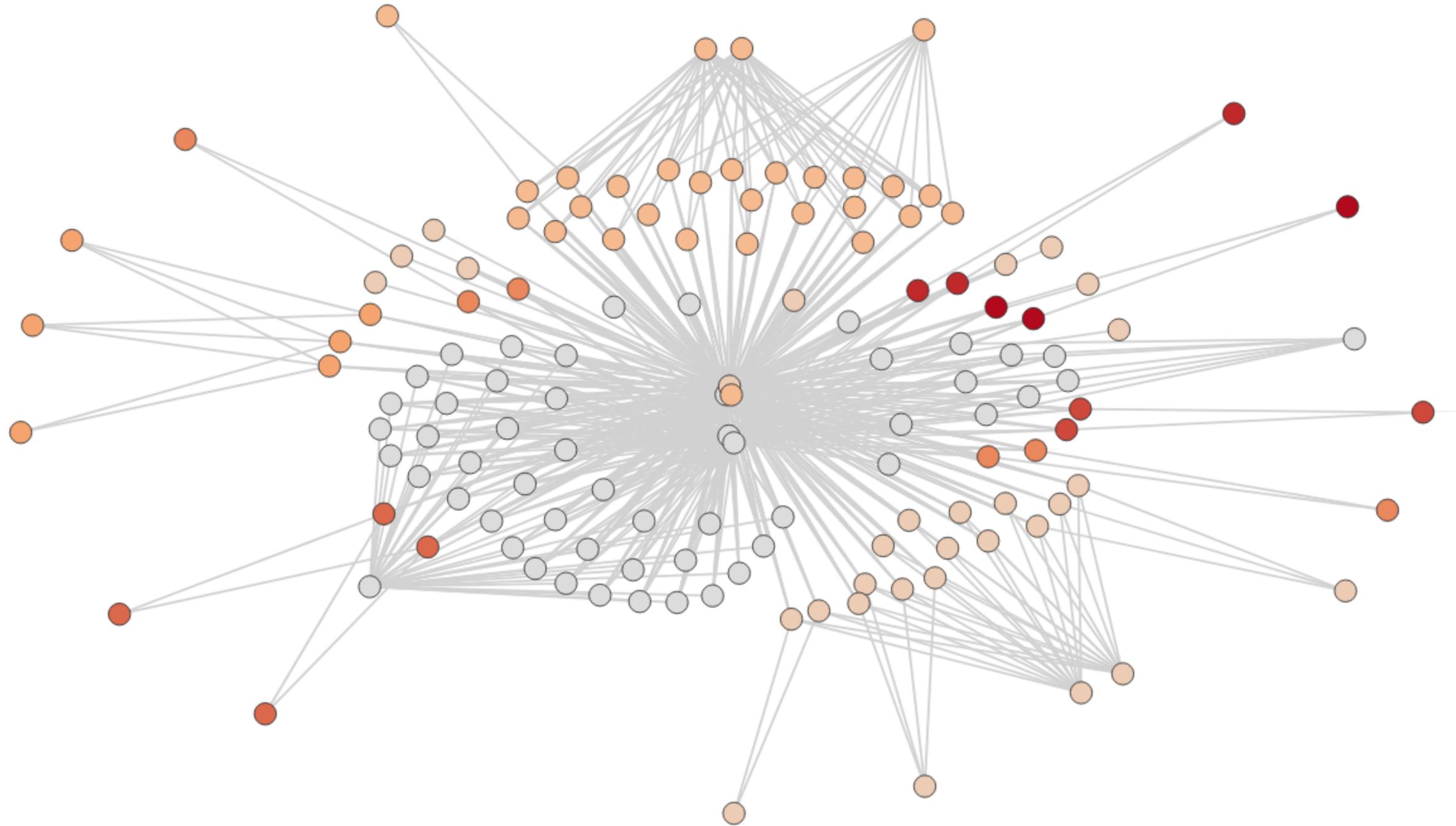
Поисковые расширения

Эвристики, текстовая
близость

В итоге получали такие красивые картинки:



В итоге получали такие красивые картинки:



Προ Google People Analytics

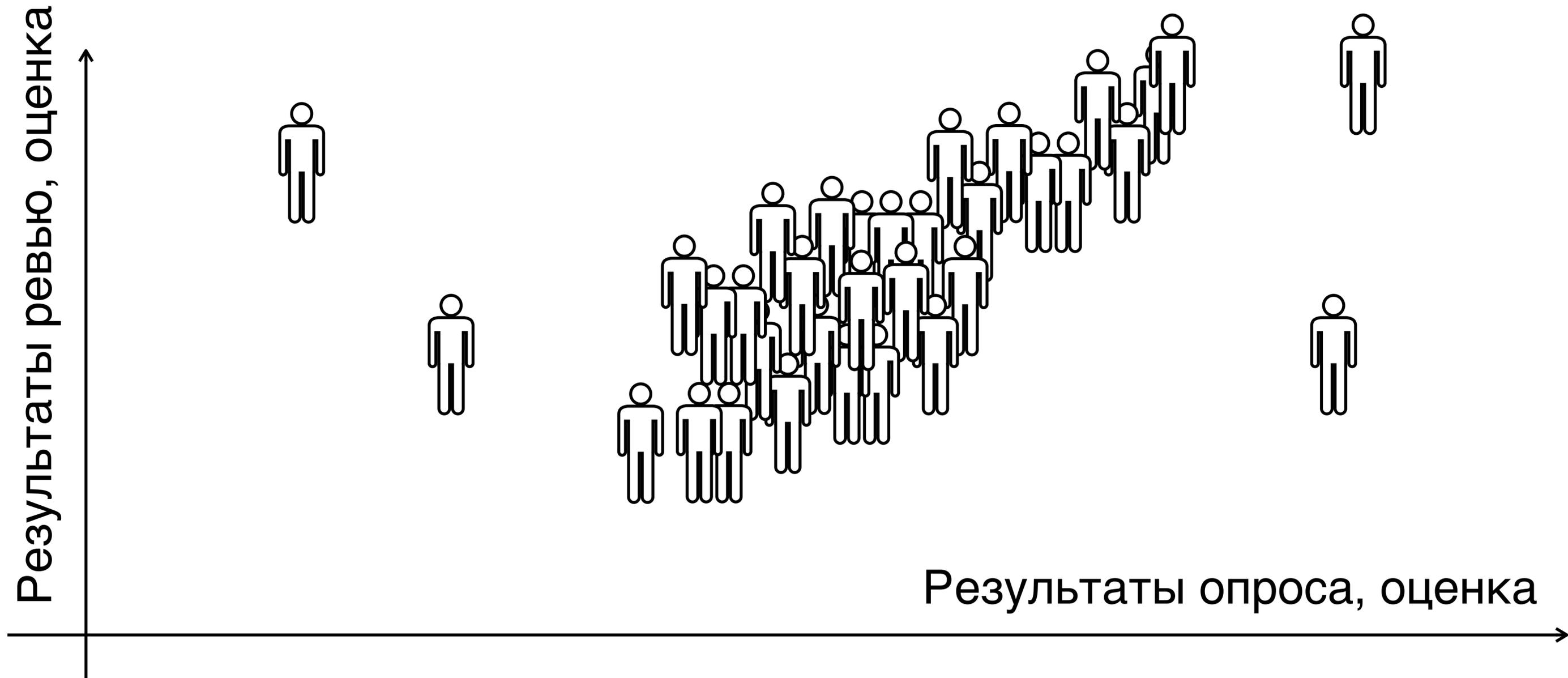


В двух словах про отдел:

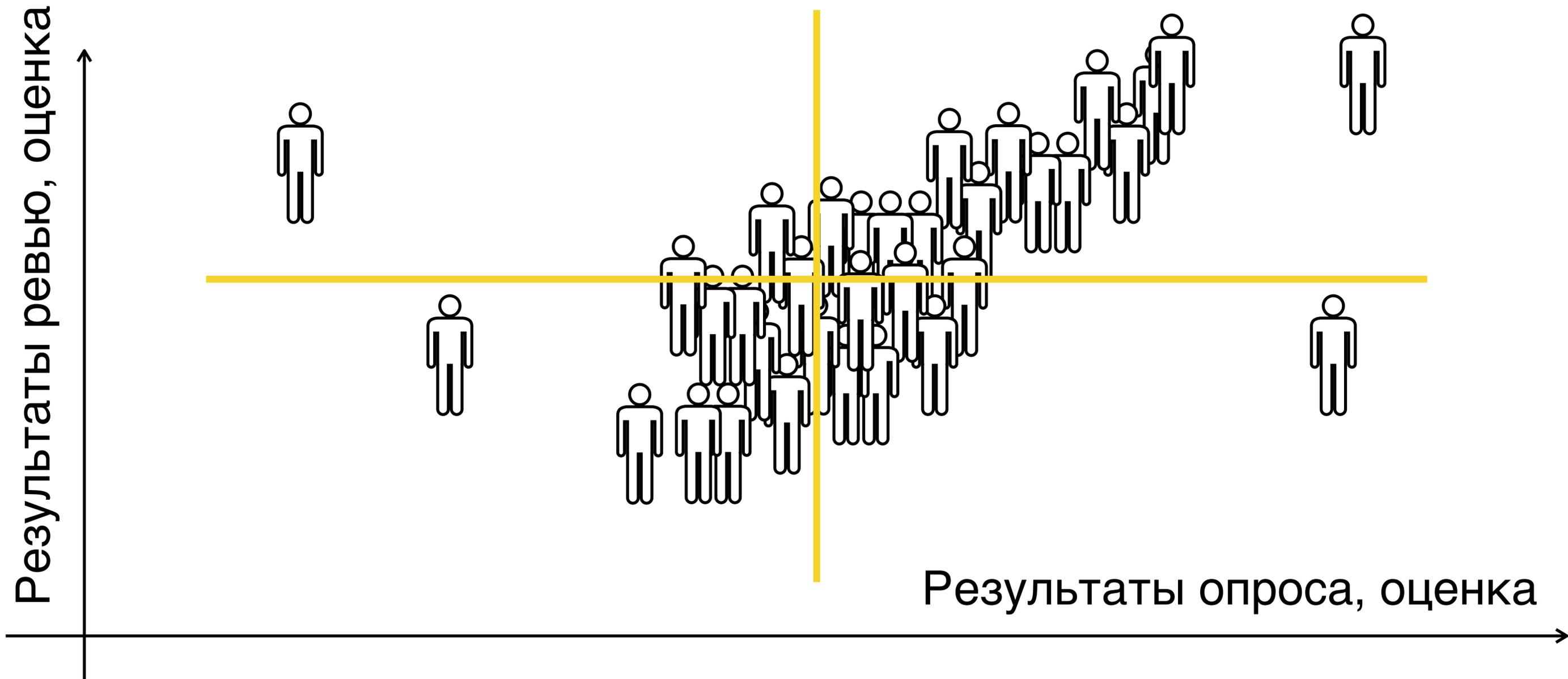
- › все решения про людей основаны на данных
- › около 25% команды специалисты в области анализа данных
- › При принятии решений следующая цепочка действий обязательна: Мнение -> Данные -> Метрики -> Аналитика -> Понимание -> Действие

Проект «Охуген»
«А что делают менеджеры? Полезны ли
они?»

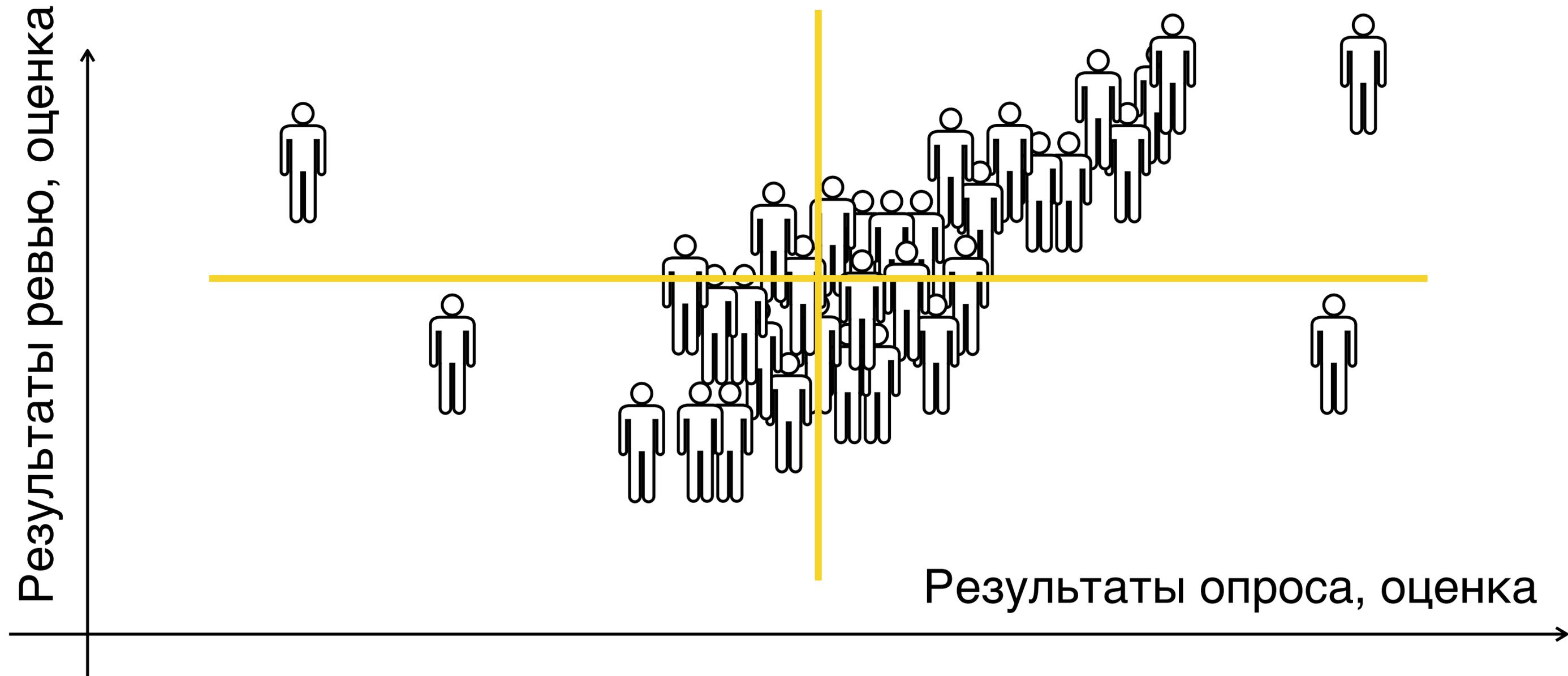
Ревью по результатам/опрос сотрудников



Ревью по результатам/опрос сотрудников



Ревью по результатам/опрос сотрудников



Сравнительный анализ двух групп: лучшие и худшие

1. Анализ текущих показателей: продуктивность команды, увольнения сотрудников, оценки сотрудников, из вовлеченность и т.п.
2. Провели «двойные слепые интервью» и выяснили, в чем разница, между группами

Google's Project Oxygen findings:

1. Is a good coach
2. Empowers the team and does not micromanage
3. Expresses interest in and concern for team members' success and personal well-being
4. Is productive and results-oriented
5. Is a good communicator—listens and shares information
6. Helps with career development
7. Has a clear vision and strategy for the team
8. Has key technical skills that help him or her advise the team

Про цепочку решений:

- › Мнение: «менеджеры не несут ценность в команду»
- › Данные: ревью сотрудников, результаты опросов...
- › Метрики: X% сотрудников ставят хорошую оценку своему менеджеру
- › Аналитика: команды у «хороших» менеджеров имеют продуктивность выше и меньший отток сотрудников
- › Действие: список ключевых компетенций и тренировка менеджеров с его учетом

Ящик пандоры. Развенчивание мифов.

Развенчивание мифов

«Работа над звездными проектами
увеличивает шансы повышения»

«При найме выпускников
смотрят на вуз а не на
возможности!»



«При найме выпускников
смотрят на вуз а не на
ВОЗМОЖНОСТИ»

«Только выпускники
Ivy League получают
повышения!»

«В главном офисе
повышают чаще!»

Магическая формула для повышений

Формула, работающая с 90% точностью
(30% полнотой):

$$\text{Odds} = e^{-22.216 + (5.227 \times \text{AvrgPerf}) + (2.732 \times \text{MgrRecommded}) + (.971 \times \text{SelfRecommended})^*}$$

$$\text{Probability (\%)} = \frac{\text{Odds}}{1 + \text{Odds}}$$

Итог:

- › комитет по найму и повышениям не стал использовать формулу
- › но смог использовать ее для проверки своих решений
- › люди должны принимать решения связанные с людьми (опираясь на данные)

Итог:

- › комитет по найму и повышениям не стал использовать формулу
- › но смог использовать ее для проверки своих решений
- › **люди должны принимать решения связанные с людьми (опираясь на данные)**

Будущее и настоящее



HR становится более развитым и технологичным:

Hr Tech Startups

\$4.9M AVERAGE VALUATION



458 COMPANIES 547 INVESTORS 4,615 FOLLOWERS 239 JOBS

Company	Joined	Followers	Signal	
 ProSky Your Recruiting, Your Training...One ... Mountain View · College Recruiting	May '13	197		
 Captain401 Modern 401(k) provider for startups ... San Francisco · Personal Finance	Sep '15	93		VIEW JOBS
 Pingboard Employee directory & shared org ch... Austin · Human Resources	Mar '14	114		VIEW JOBS
 Crowdworks The Next Generation of crowd sourci... Tokyo · Crowdsourcing	Jan '13	69		
 hackajob				VIEW JOBS

Рынок технологических стартапов HR в США

Что мы научимся предсказывать совсем скоро:

- › хороших кандидатов для найма
- › карьерный путь человека в компании
- › уход сотрудников из компании
- › что нужно сделать, чтобы повысить продуктивность



1. Эффективный найм

- › ранжирование кандидатов
- › соответствие кандидата и позиции
- › автоматизированный и полуавтоматизированный процесс найма

2. Предсказание пути в компании

- › что случилось с сотрудником через 6, 12, 18 месяцев после прихода в компанию
- › предсказание пути, определение типа сотрудника

3. Анализ вовлеченности

- › какие факторы главным образом влияют на вовлеченность
- › что нужно сделать чтобы помочь сотрудникам быть лучше и эффективней

4. Удержание сотрудников

- › предсказание сотрудников, которые захотят перейти
- › удержание сотрудников, находящихся в зоне риска

4. Удержание сотрудников

- › предсказание сотрудников, которые захотят перейти
- › удержание сотрудников, находящихся в зоне риска



Спасибо!