

DANC

От чего зависят результаты участников забега?

Спикер: Кислицын Дмитрий, ведущий научный сотрудник Междисциплинарного центра исследования общественного здоровья Сеченовского университета, доцент НИУ ВШЭ



Пробежки «5 верст»



«5 верст» - самый популярный в России формат массовых пробежек



Пробежки «5 верст» проходят каждую субботу в 9 утра (за редкими исключениями) в городских парках в десятках городов по всей России круглый год



«5 верст» ориентированы на любителей и не являются спортивными соревнованиями, но время всех зарегистрированных участников фиксируется и отражается в протоколе забега



База данных



Единица наблюдения – одна пробежка одного человека



Данные за 2023 год



Только участники, зарегистрировавшиеся в 2023 году



Всего в базе 74984 наблюдений



Описание переменных

Идентификаторы

record_id

Уникальный идентификатор забега, напр., 13138

athlete_id

Уникальный идентификатор человека, напр., 790097123

event_name

Название локации, напр., “Сосновка” [парк в Санкт-Петербурге]

event_date

Дата пробежки, напр., 2023-01-01

Описание переменных

Информация о старте

n_runners

Количество участников забега в этот день в этом парке [включая участников, исключенных из базы данных]

Результат

finish_time

Время, за которое бегун пробежал эту конкретную пробежку 5 км (в секундах)

position

Каким бегун был на финише данной пробежки (напр., 1, если бегун прибежал самым первым из всех участвующих, включая бегунов, исключенных из базы данных]

Описание переменных

Характеристики бегуна

athlete_gender

Пол бегуна (мужской, женский)

athlete_age

Возраст бегуна, лет

**best_time, worst_time, mean_time,
median_time**

Лучшее, худшее, среднее и медианное время на всех пробежках “5 верст” с момента регистрации и до данной пробежки (включая данную пробежку)

times

Количество участия в пробежках “5 верст” с момента регистрации и до данной пробежки (включая данную пробежку)

num_locations

Количество локаций, в которых бегун участвовал в пробежках “5 верст” с момента регистрации и до данной пробежки (включая данную пробежку)

home

Домашняя локация (в которой пробежал больше всего “5 верст”, напр., “Сосновка”)

Описание переменных

Характеристики погоды в день старта

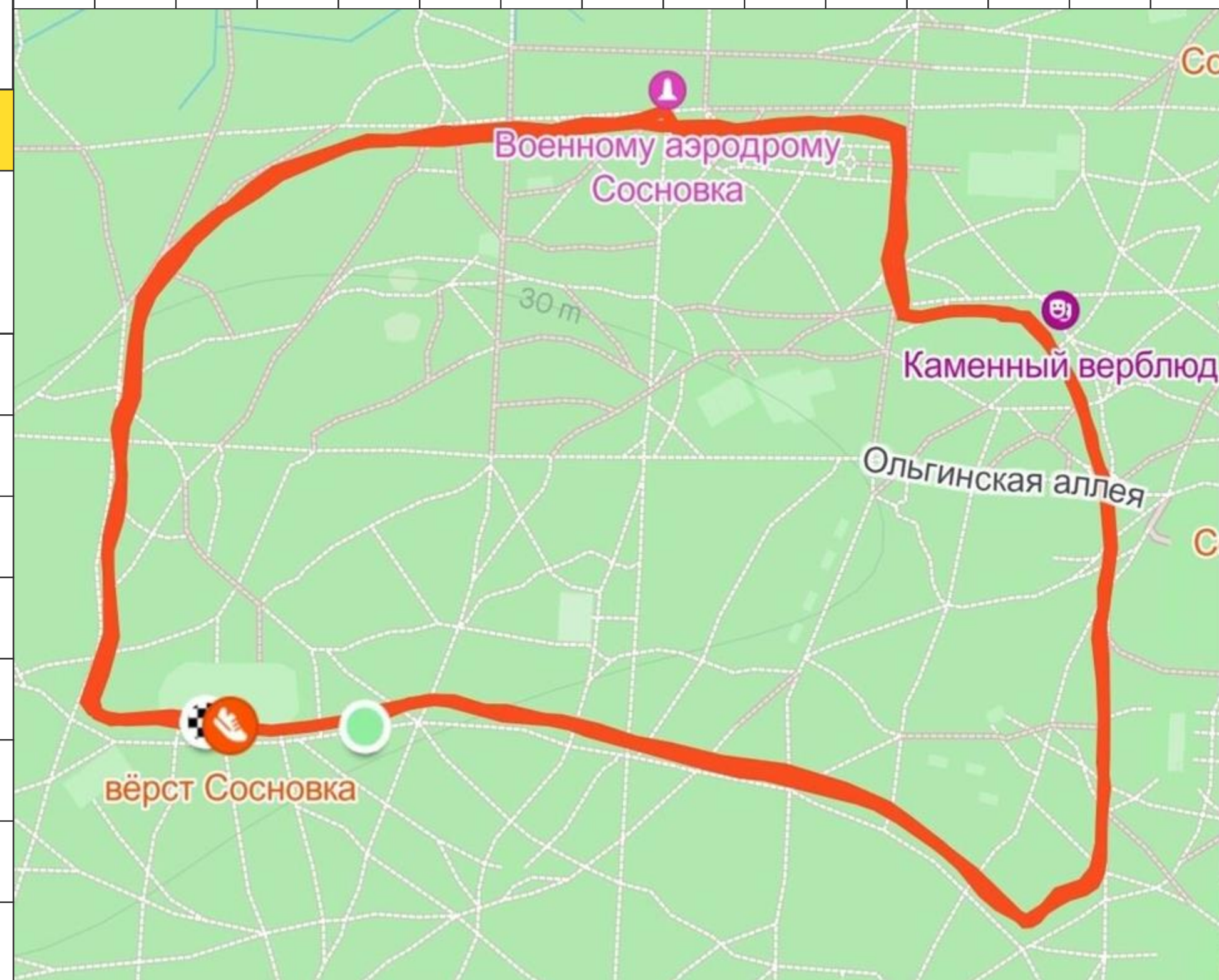
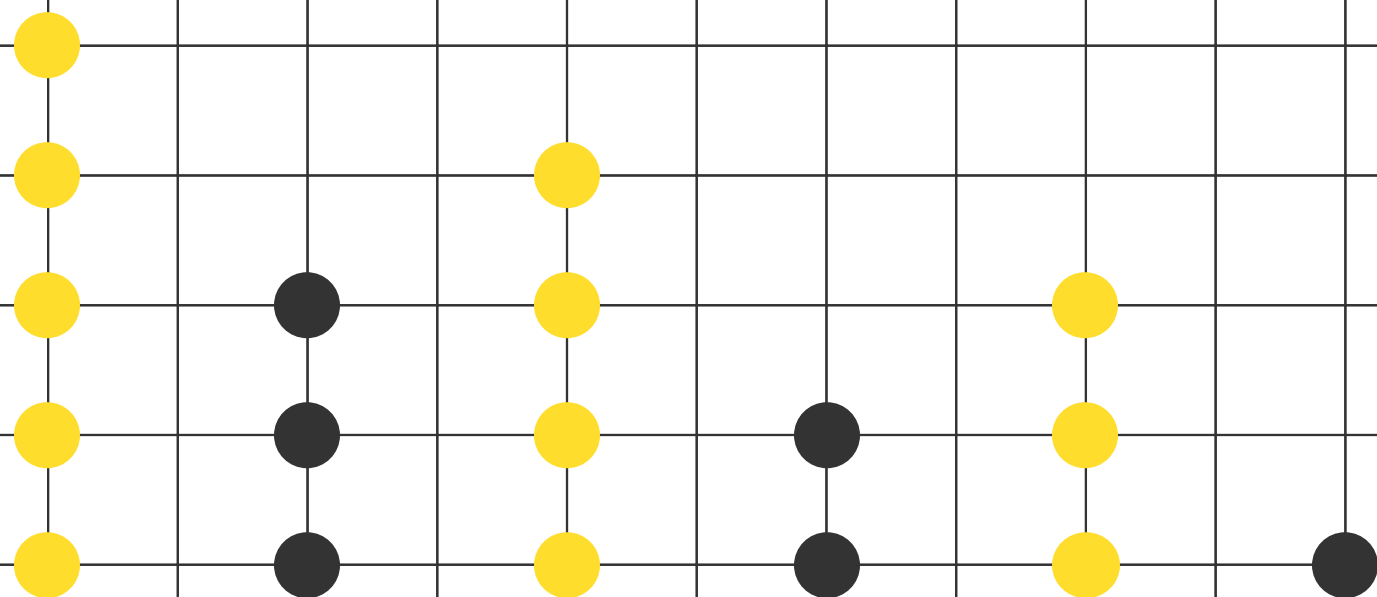
temp	Температура в данной локации в день старта в градусах Цельсия, напр., 15.22
feels_like	Как ощущается температура в данной локации в день старта, напр., 14.42
pressure	Атмосферное давление в гектапаскалях (hPa)
humidity	Влажность, %
dew_point	Точка росы в градусах Цельсия
clouds	Облачность, %
wind_speed	Скорость ветра, м/с
wind_deg	Метеорологическое направление ветра (азимут ОТКУДА дует)
weather.main	Погодные параметры (например, "Clouds" — облачно)
weather.description	Уточнённые погодные параметры (например, "broken clouds")
rain.1h	Интенсивность осадков в мм/ч: дождь
snow.1h	Интенсивность осадков в мм/ч: снег

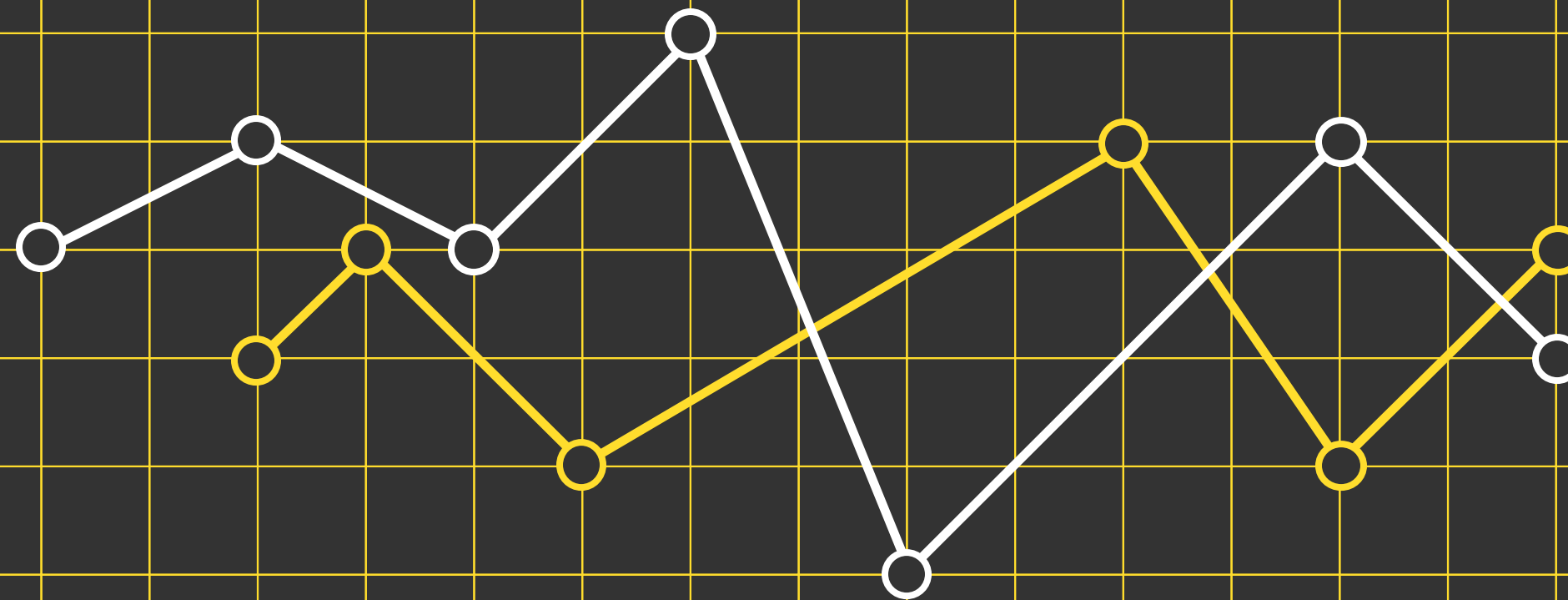
Описание переменных

Характеристики локации

track_description_surface

Покрытие (напр., "грунт", "асфальт", "смешанное")





Исследовательский вопрос

От чего зависят результаты участников забега?

Результат участника



Переменной, доступной в базе, которую мы можем использовать в качестве меры скорости бегуна, является `finish_time` – время, за которое участник пробежал данную пробежку



Для того, чтобы правильно интерпретировать результаты, помните, что пробежка «5 верст» — это не легкоатлетическое соревнование, и участник может бежать не в полную силу



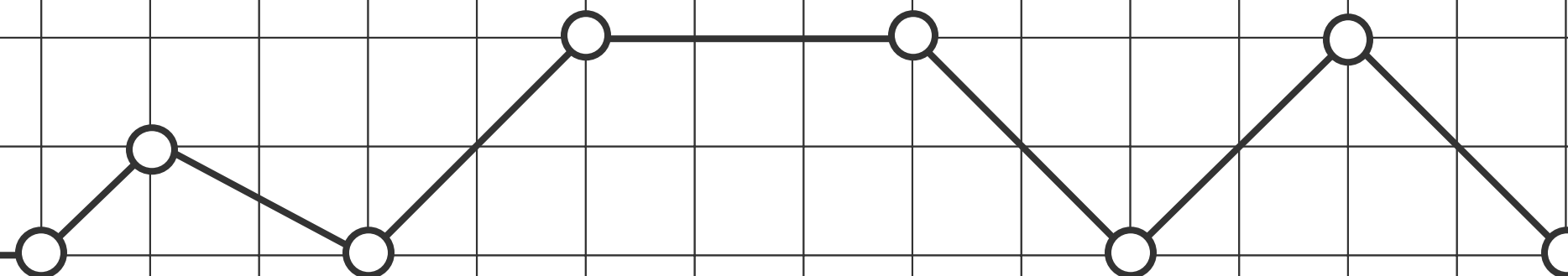
Тем не менее, часто участники испытывают азартное желание побить свой личный рекорд или финишировать раньше, чем другие участники, поэтому выкладываются изо всех сил

**Что мы
ожидаем?**



**Какие доступные в базе данных переменные
могут быть связаны с результатами участников?
Какие зависимости вы ожидаете?**

- Демографические характеристики (пол и возраст)
- Активность на мероприятиях «5 верст»
- Погода
- Покрытие



Демографические характеристики



Насколько быстрее в среднем бегают мужчины, чем женщины?



Как возраст влияет на результаты?



В каком возрасте участники показывают лучшие результаты?

Активность на мероприятиях «5 верст»



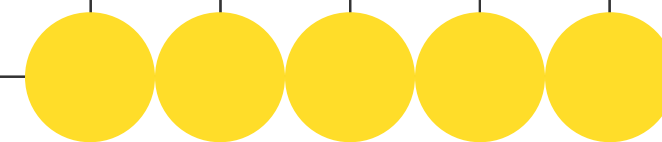
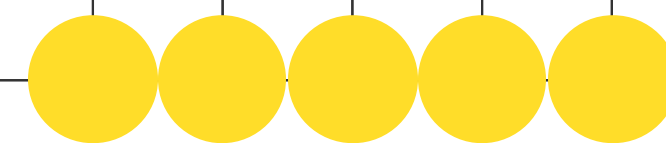
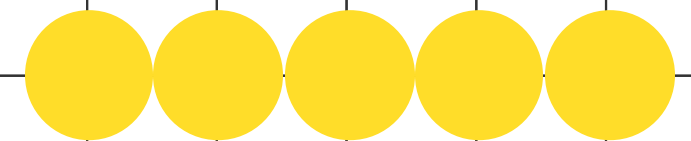
Правда ли, что люди, регулярно участвующие в забегах, улучшают свои результаты? Насколько?



Правда ли, что участники, «коллекционирующие» локации (бегающие в разных парках), показывают лучшие результаты, чем те, кто бегают только в одном парке?



Правда ли, что чем более массовым является забег, тем лучшие результаты показывают участники?



Погода



Из-за того, что мероприятия «5 верст» проходят в одних и тех же локациях в одно и то же время, мы смогли собрать детальную информацию о погодных условиях во время пробежек



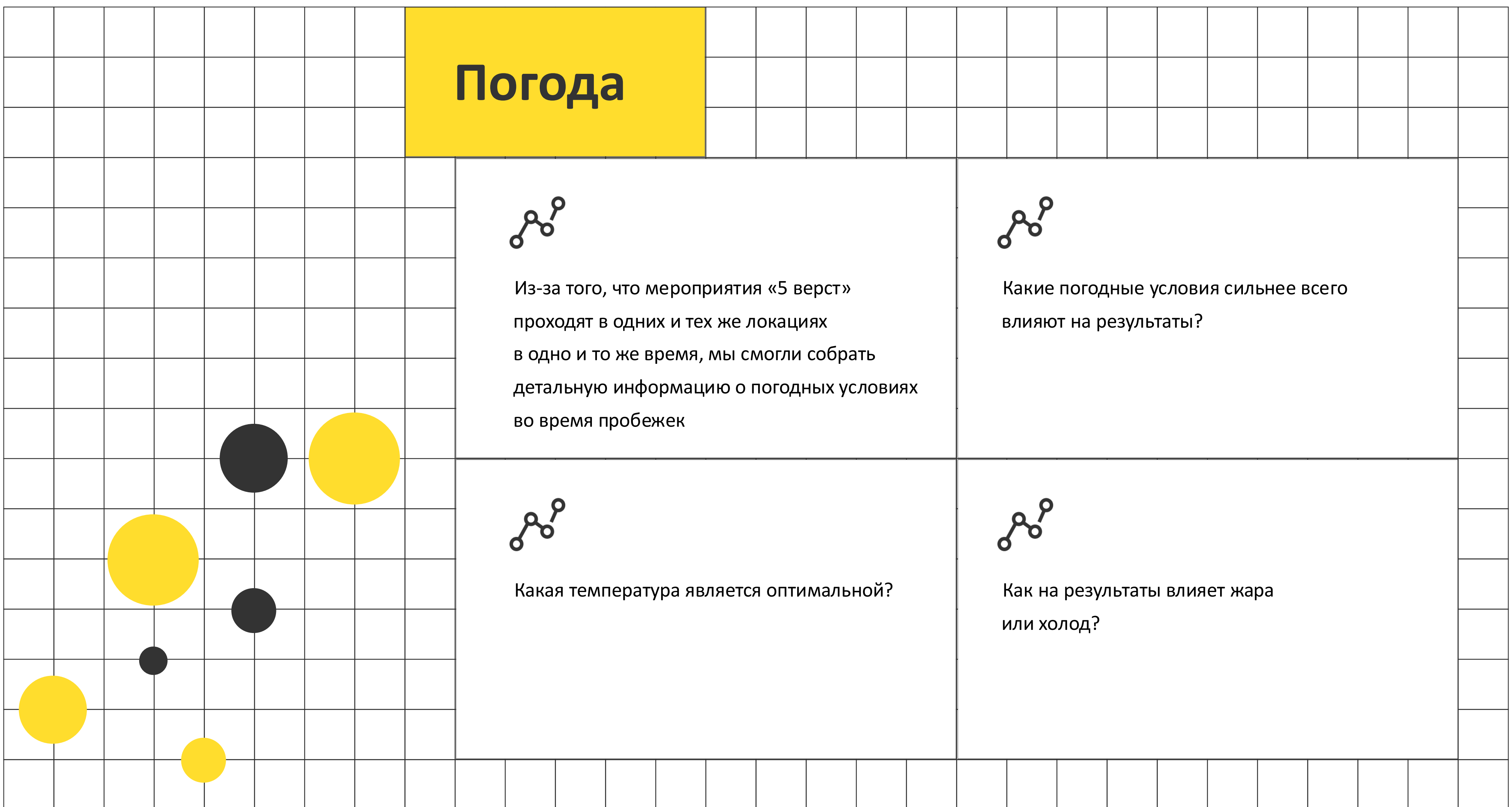
Какие погодные условия сильнее всего влияют на результаты?



Какая температура является оптимальной?



Как на результаты влияет жара или холод?



Покры́тие



Наконец, нам доступна такая характеристика локации, как тип покрытия



Насколько сильно покрытие влияет на результаты участников?



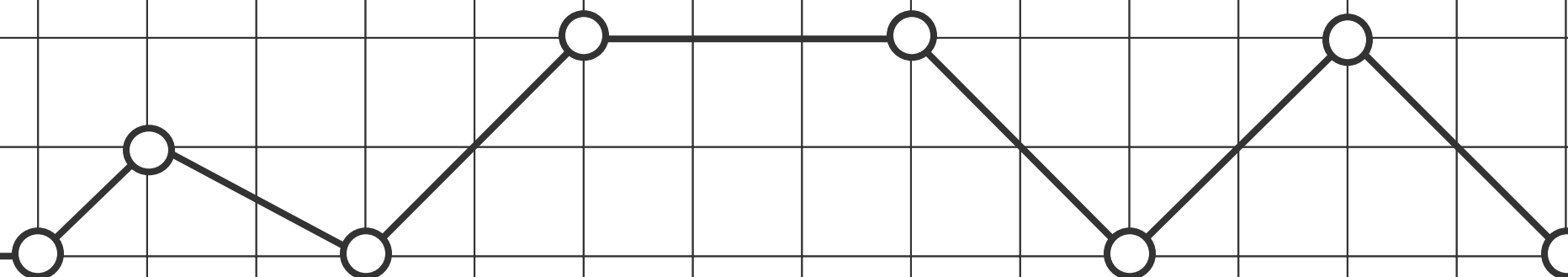
Как этот эффект связан с погодными условиями?

Агрегирование данных



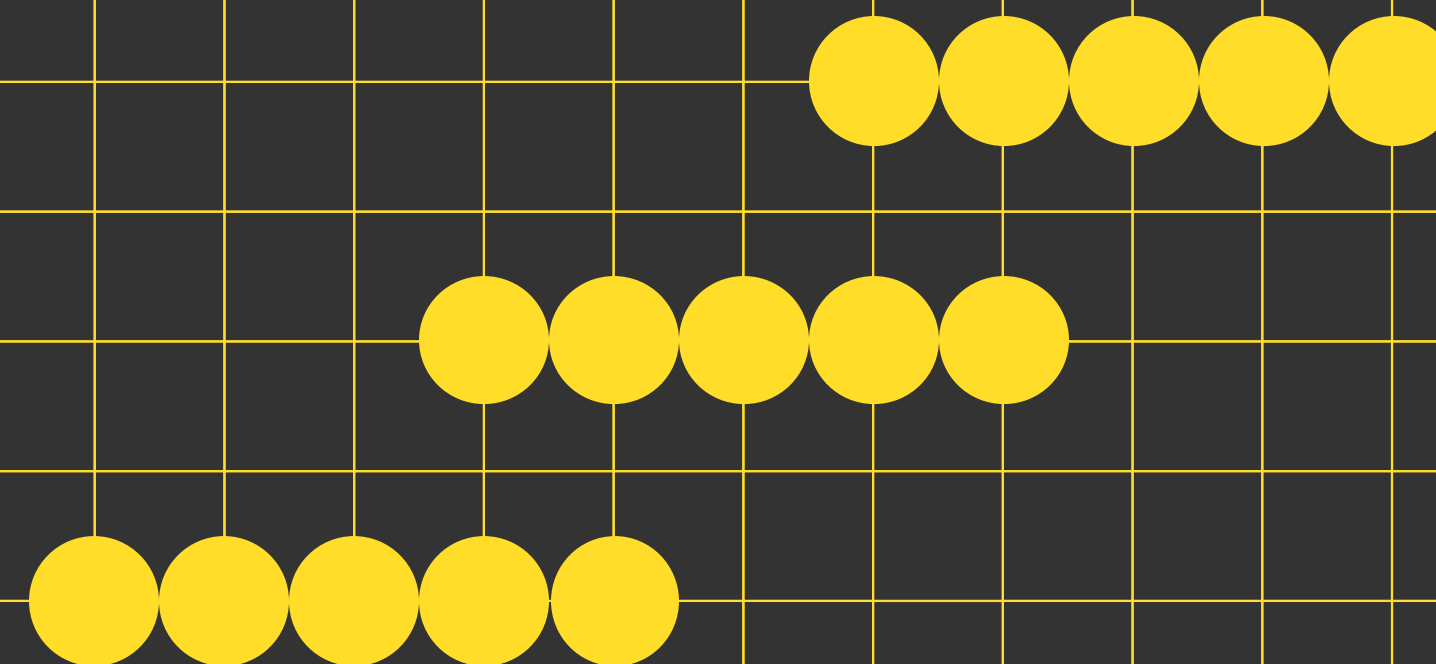
**Есть ли необходимость агрегировать данные?
(на уровне участников, забегов, локаций...)**

Это зависит от того, влияние каких именно факторов вы исследуете. В целом, это не является необходимым



Анализ

Учитывая все вышеперечисленное, проведите анализ данных и выявите ключевые факторы, влияющие на результаты участников пробежек «5 верст»



Помните



Представленные ранее пункты являются примерными и отражают “программу максимум”



Успешная работа не обязательно требует выполнения всех из них – вы можете сфокусироваться на отдельных аспектах



Более того – работа, глубоко исследующая какой-то один важный аспект представляет большую ценность, чем работа, исследующая все, но поверхностно

Структура работы

Введение

- Мотивация
- Исследовательский вопрос
- Гипотеза
- Механизм

Данные

- Описание переменных
- Разведывательный анализ

Методы

- Проверка гипотезы
- Визуализация

Результаты

- Основные результаты
- Проверка устойчивости результатов

Обсуждение

- Основные выводы
- Применение результатов
- Ограничения исследования

**Обсуждение:
выводы
и практические
рекомендации**



Подведите итог проведенному исследованию



Что именно вы хотели выяснить?



Какие результаты были получены? Как эти результаты могут быть использованы на практике?



Какие ограничения своего исследования вы видите?



Если бы у вас было больше времени и не было ограничений, связанных с доступностью данных, как вы могли бы дополнить и улучшить свое исследование?

DANG

**Желаю вам хорошо
поработать на Хакатоне!**

