



От чего зависят результаты участников забега?

Бег — это самый популярный вид физической активности в России. По данным ВЦИОМа, в 2021 году бегом, легкой атлетикой или спортивной ходьбой занимались более четверти россиян. Самым популярным в России форматом массовых пробежек является «5 верст» — еженедельные дружеские забеги по 5 километров, проходящие по утрам (обычно в 9:00) в городских парках в десятках городов по всей России. Несмотря на то, что «5 верст» ориентированы на любителей и не являются спортивными соревнованиями, время всех зарегистрированных участников фиксируется и отражается в протоколе забега. Мероприятия «5 верст» проходят в одних и тех же парках в одно и то же время круглых год, при самой разной погоде, от летней жары, до зимних холодов.

В базе данных, которую вы получили, присутствуют все участники забегов «5 верст» зарегистрировавшиеся в 2023 году. Мероприятия «5 верст» проходят с 2013 год (под названием parkrun), но для данного задания мы исключили всех участников, которые зарегистрировались ранее 2023 года. Единицей наблюдения в базе является одна пробежка одного человека, представлены все пробежки за 2023 год участников, зарегистрировавшихся в этом же году.

Описание переменных представлено ниже.

Переменная	Объяснение
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
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]
Характеристики бегуна	
athlete_gender	Пол бегуна (мужской, женский)
athlete_age	Возраст бегуна, лет
best_time	Лучшее время на всех пробежках «5 верст» с момента регистрации и до данной пробежки (включая данную пробежку)
worst_time	Худшее время на всех пробежках «5 верст» с момента регистрации и до данной пробежки (включая данную пробежку)
mean_time	Среднее время на всех пробежках «5 верст» с момента регистрации и до данной пробежки (включая данную пробежку)
median_time	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»)

Переменная**Объяснение****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

Характеристики локации

track_description_surface	Покрытие (напр., «грунт», «асфальт», «смешанное»)
---------------------------	---

Используя полученные данные, вы должны ответить на вопрос: «От чего зависят результаты участников забега?».

Переменной, доступной в базе, которую мы можем использовать в качестве результата участника забега, 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.

Очевидно, что, степень подготовленности участника зависит от того, как давно он занимается бегом, а также от объема его тренировочных нагрузок. Эти данные нам не доступны, мы видим лишь активность участников в рамках пробежек «5 верст» и не видим их активность за пределами данных пробежек.

Первая групп факторов, влияющих на результаты участников, — это социально-демографические характеристики бегунов. Насколько быстрее в среднем бегают мужчины, чем женщины? Как возраст влияет на результаты? В каком возрасте участники показывают лучшие результаты?

Вторая группа факторов — активность в рамках «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?

Третья группа факторов — погодные условия. 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?

Наконец, нам доступна такая характеристика локации, как тип покрытия. Насколько сильно покрытие влияет на результаты участников? Как этот эффект связан с погодными условиями?

Учитывая все вышеперечисленное, проведите анализ данных и выявите ключевые факторы, влияющие на результаты участников.

Обязательно подведите итоги проведенного исследования. Что именно вы хотели выяснить? Какие результаты вы получили? Как эти результаты могут быть использованы на практике? Какие ограничения в рамках 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?

Примечание. Пункты выше примерные и отражают программу- максимум. Успешная работа не обязательно требует выполнения всех из них — вы можете сфокусироваться на отдельных аспектах. Более того, работа, где глубоко исследуют какой-то один важный аспект, представляет большую ценность, чем работа, где исследуют все, но поверхностно.

Желаем вам плодотворной работы на хакатоне!