

# Анализ датасета Cycling Metrics

Данный датасет был выбран, поскольку я сам езжу на велосипеде и ранее также записывал свои треки с помощью Strava. К сожалению на данный момент компания принуждает пользователей использовать подписку, но в то же время не улучшает свое приложение. В дальнейшем о проблемах этого приложения будет рассказано подробнее.

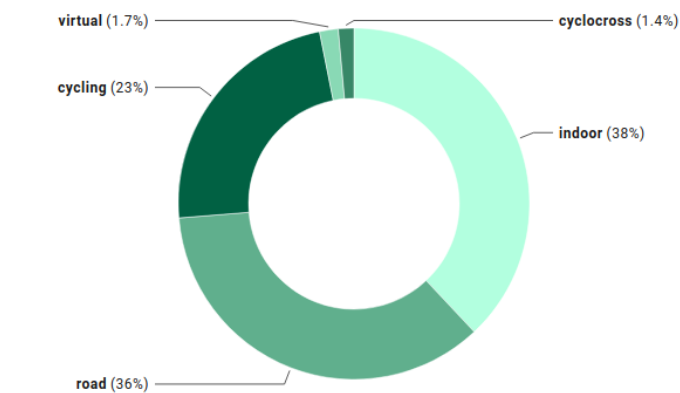
В датасете представлены метрики с 350 велоаездов, предположительно записанных с помощью приложения Strava.

Заезды отличаются по типу:

- cyclocross - пересеченная местность
- cycling - тип неизвестен
- indoor cycling - тренировка на станке
- road biking - шосссе
- virtual ride - тренировка на станке со специальным оборудованием

В данной статье будут представлены диаграммы, позволяющие оценить содержание датасета.

## Распределение заездов по их типу



[Get the data](#) • Created with [Datawrapper](#)

Одними из наиболее важных показателей, позволяющих оценить заезд, являются средняя скорость и расстояние, преодоленное велосипедистом. Стоит отметить, что точки графика, находящиеся на его осях, учитывать не следует, поскольку рассматривать заезды длиной менее 1000 метров не стоит. Аналогично ситуация обстоит и с заездами, средняя скорость которых близка к нулю. Чаще всего такое происходит при сбое датчика gps.

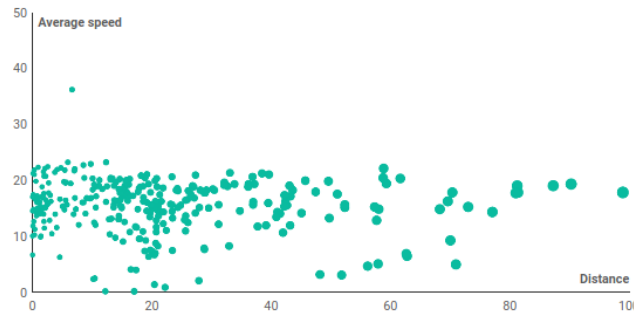
Нагрузку, которую испытывает велосипедист, явно отображает его пульс. Соответственно, чем выше средняя скорость заезда, тем больше трудится спортсмен и тем выше его пульс. Стоит отметить, что существует множество различных факторов, влияющих на среднюю скорость (качество дороги, ее наклон, ветер и тд) и такое же количество факторов, влияющих на пульс (например подготовленность спортсмена). По этим причинам точки на графике не находятся на одной линии. Однако кубическая линия тренда достаточно явно показывает рост пульса при росте средней скорости.

## Распределение заездов по средней скорости и

## Зависимость пульса от средней скорости

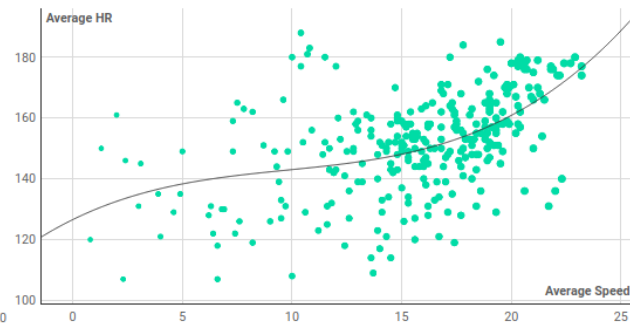
Однако кубическая линия тренда достаточно явно показывает рост пульса при росте средней скорости.

## Распределение заездов по средней скорости и расстоянию.



[Get the data](#) • Created with Datawrapper

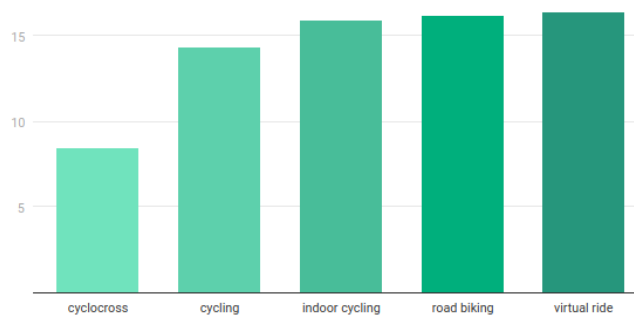
## Зависимость пульса от средней скорости



[Get the data](#) • Created with Datawrapper

Как было сказано ранее, скорость зависит от многих факторов. График ниже наглядно демонстрирует разницу между движением по шоссе и пересеченной местности.

## Средняя скорость заездов каждого типа



[Get the data](#) • Created with Datawrapper

Возникает несколько вопросов: почему для типа cycling средняя скорость настолько мала и почему road biking находится между двумя типами заездов на станках?

Для того, чтобы ответить на них, придется воспользоваться собственным опытом. Тип cycling (что эквивалентно отсутствию типа) чаще всего используется новыми пользователями приложения, многие из которых не являются профессиональными спортсменами, также часть из этих заездов скорее всего проходили по пересеченной местности. Ситуация с тремя последними категориями более интересная. В целом в домашних условиях, на специальных станках занимаются лишь профессиональные спортсмены, поэтому средняя скорость здесь настолько высока. Virtual ride подразумевает использование другого приложения для тренировок, Zwift. Оно платное, поэтому только спортсмены более высокого уровня видят смысл его использовать.

Исходя из собственного опыта и анализа датасета вынужден отметить, что, к сожалению, Strava для записи заездов является далеко не идеальной платформой, поскольку, как было отмечено ранее, происходят ошибки GPS, у некоторых треков отсутствуют важнейшие показатели, такие как максимальная скорость или расстояние, порой треки записываются с ошибками (например в датасете был заезд длиной 10 метров со средней скоростью 111 миль в час).