



BIOINFORMATICS  
INSTITUTE

# Проект 1

# Датасет

- Сервисы проката велосипедов и самокатов продолжают набирать популярность благодаря своему удобству и скорости. Однако, они довольно сильно зависят от погодных условий
- В нашем распоряжении есть дневные данные за 2011-2012 годы по поездкам на велосипедах в Вашингтоне



- 731 день
- 17 колонок со значениями



# Что внутри?

- instant: ID наблюдения
- dteday : дата
- season : сезон (1:весна, 2:лето, 3:осень, 4:зима)
- yr : год (0: 2011, 1:2012)
- mnth : месяц ( 1 to 12)
- holiday : праздничный день
- weekday : день недели
- workingday : если день рабочий 1, иначе 0.
- weathersit :
  - 1: Ясно, немного облаков, частично облачно, частично облачно
  - 2: Туман + облачность, Туман + разорванные облака, Туман + немного облаков, Туман
  - 3: Небольшой снег, небольшой дождь + гроза + рассеянные облака, небольшой дождь + рассеянные облака
  - 4: Проливной дождь + ледяные паллеты + гроза + туман, снег + туман
- temp : Нормированная температура в градусах Цельсия. Значения делятся на 41 (макс.)
- atemp: Нормированная температура ощущений в градусах Цельсия. Значения делятся на 50 (макс.)
- hum: Нормированная влажность. Значения делятся на 100 (макс.)
- windspeed: Нормированная скорость ветра. Значения делятся на 67 (макс.)
- casual: количество случайных пользователей
- registered: количество зарегистрированных пользователей
- cnt: количество велосипедов, взятых напрокат, с учетом случайных и зарегистрированных



# Что мы нужно сделать?

- Проверить данные, сделать EDA
- Проверить 3 гипотезы
- Составить отчёт в формате Rmd
- Проверить по 3 проекта других команд
- Улучшить свой отчёт с учётом пожеланий проверяющих

# Опережая вопросы

- 1) Если вы решили проверить какую-то взаимосвязь, но она оказалась незначимой, то это не повод не писать о ней
- 2) Выбросы дело творческое - можно по-разному работать с ними, главное объяснить
- 3) Лучше ограничиться критериями и не идти в модели/дисперсионный анализ. Они будут в следующем проекте
- 4) Наши данные распределены ненормально. Это может быть связано с тем, что наслаиваются года/сезоны/выходные на один график

# Критерии оценки

7 баллов  
Работа над проектом

5 баллов  
Проверка проектов

3 балла  
Улучшение отчёта



## 1. Проверка и коррекция данных 2 балла

Найдены пропущенные значения 0.5

Объяснено, что делать с пропущенными значениями 0.5

Найдены выбросы 0.5

Объяснено, что с ними делать 0.5

## 2. Проверка гипотез 2.4 балла 2+1 шт

Как в нашем пример, нужно указать почему интересно проверить взаимосвязь и оформить ответ. По алгоритму с пары опишите, гипотезу, выбранный критерий и тд

Каждая гипотеза по 0.8 балла (за каждое по 0.2 балла : описано что мы проверяем, каким образом и почему так, какой ответ получили)

## 3. Оформление отчёта 1.8 балла

Структура, дизайн, разумность скрытия/показа кода и ошибок чанков, описание исходных данных

0.5 наличие структуры документа, организация отчета

0.25 описан датасет

0.65 понятность текста

0.4 работа с ошибками чанков

## 4. Использование пакетов, которые не обсуждались в лекциях

За 2 и более пакетов 0.8 балла

За 1 0.4 балла

Нужно пояснить, в чём смысл использования данного пакета

0.2 использован пакет

0.2 дано объяснение зачем этот пакет

# Критерии оценки

7 баллов  
Работа над проектом

5 баллов  
Проверка проектов

3 балла  
Улучшение отчёта



В ходе проверки вы готовите короткий отчёт, в котором говорите о соответствии / не соответствии критериям, пожелания для улучшений и общее впечатление

5 баллов – рассмотрены все критерии, есть пожелания и мнение по критериям не противоречит мнению команды курса

# Критерии оценки

7 баллов  
Работа над проектом

5 баллов  
Проверка проектов

3 балла  
Улучшение отчёта



На основе полученной обратной связи  
вы улучшаете свой отчёт

3 балла - пожелания учтены и отчёт  
стал лучше 😊

0 баллов - отчёт не изменился



# Сдача отчёта

- Отчёт в формате Rmd – проверьте, что он компилируется и не содержит указаний на авторов (анонимный)
- Консультация – 3 октября
- Срок сдачи – 13 октября
- Проверяем отчёты друг друга – 20 октября
- Сдаём отчёты с доделками – 27 октября

**Вопросы?**