

DS-поток.

Продвинутое практическое задание.

Анализ данных системы велопроката.



Описание задачи.

Система велопроката Pronto Cycle Share в Сиэтле состоит из 54 станций и 500 велосипедов. На Kaggle компания Pronto предоставляет открытые данные об отдельных поездках, станциях и погоде. Имеются следующие данные:

1. `station.csv` — информация о станциях проката. Для каждой станции файл содержит `id` станции, ее имя, координаты, количество мест для велосипедов, а так же информацию об изменениях на станции.
2. `trip.csv` — информация о поездках. Для каждой поездки файл содержит `id` поездки, `id` используемого велосипеда, время поездки, станции отправления и прибытия, а также информацию о пользователях — тип абонемента, пол, возраст.
3. `weather.csv` — информация о погоде за каждый день — температура, точка росы, влажность, давление, видимость, скорость ветра.

<https://www.kaggle.com/pronto/cycle-share-dataset>

Формат выполнения задания

Каждая исследовательская задача по заданию должна состоять из следующих шагов.

1. Формулировка задачи на простом "пользовательском" языке.
2. Формулировка задачи на математическом языке.
3. Исследование на Питоне или на R. Весь необходимый код должен быть включен в решение задания.
4. Выводы как на математическом, так и на "пользовательском" языке.

Идеи исследования

Для выполнения задания можно выбрать некоторые идеи из приведенного ниже списка или же предложить свои идеи.

1. Простая визуализация данных. В том числе визуализация на карте города поездок, станций проката.
2. Исследование сезонности поездок.
3. Влияние погодных условий на количество поездок.
4. Выявление самых популярных маршрутов. Маршруты могут зависеть от дня недели (рабочий, выходной) и от времени суток.
5. Поиск станций, на которых наблюдается нехватка велосипедов или же, наоборот, их переизбыток. Рекомендации для сервиса по балансировке количества велосипедов.
6. Поиск поездок, совершаемых компанией из нескольких человек. В какое время суток и в какие дни совершаются такие поездки? Какие популярные маршруты для путешествия компанией? Есть ли зависимость от полового состава компании?
7. Как пользуются велопрокатом обладатели различных абонементов?
8. Где стоит открыть еще несколько станций?
9. Предсказание количества совершаемых поездок на неделю вперед в зависимости от погодных условий. Какие факторы оказывают наибольшее влияние на прогноз?
10. Предсказание поездок на 1-2 часа вперед. На какие станции стоит скорее подвезти велосипеды? На каких станциях стоит освободить места для велосипедов?

Некоторые поездки могут отображаться как перескок велосипеда с одной станции на другую. Скорее всего это не поездки, а переброски велосипедов сотрудниками службы проката с одной станции на другую.

При анализе разрешается пользоваться доступными ноутбуками на Kaggle.