

Практика по курсу распределенных систем обработки данных



Что делать?

- Объединиться в группы по 3-4 человека
- Зарегистрировать бесплатный аккаунт в желаемом облаке
- Выбрать вариант задания
- Выполнить задание с помощью Apache Spark
- Развернуть его в облаке
- Показать мне развернутое и исходный код

Р. S. помнить, что этот семестр
заканчивается в конце марта



Что вам за это будет?

За успешное выполнение задания
вовремя команда получает **автомат.**

Как и когда можно отчитать задание?

- Задание можно отчитать на практических занятиях
- Можно приходить на практику к другой группе

411	451
10:00 - 11:35	13:50 - 15:25
-	15:35 - 17:10

Варианты заданий



1

<https://www.kaggle.com/fivethirtyeight/russian-troll-tweets>

Метрики для задания

- Кол-во:
 - Уникальных авторов
 - Уникальных регионов
 - топ 10 авторов по кол-ву фолловеров

2

<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>

Метрики для задания

- Кол-во медалей для каждого года соревнований
- Топ-10 городов по кол-ву медалей
- Топ-10 стран по кол-ву золотых медалей за каждый год

3

<https://www.kaggle.com/decide-soluciones/air-quality-madrid>

Метрики для задания

- Минимальный и максимальный уровень относительно моря (столбец elevation)
- Средний уровень (elevation)
- Медианный уровень (elevation)

4

<https://www.kaggle.com/kemical/kickstarter-projects>

Метрики для задания

- Кол-во успешных проектов (state)
- Кол-во неуспешных проектов
- Топ-10 стран по успешным проектам
- Топ 10 стран по неуспешным проектам

5

<https://www.kaggle.com/michaelshoemaker/divvy-bike-chicago-2018>

Метрики для задания

- Максимальная/минимальная/средняя/медианная продолжительность поездки
- Топ-10 адресов по аренде велосипедов
- Топ-10 адресов по сдаче велосипеда
- Медианный возраст клиента

6

<https://www.kaggle.com/roccoli/gpx-hike-tracks>

Метрики для задания

- Топ-10 пользователей с самой высокой средней скоростью
- Топ-10 клиентов с самой высокой абсолютной скоростью
- Максимальная/минимальная/средняя/медианная продолжительность восхождения

7

<https://www.kaggle.com/jameslko/gun-violence-data>

Метрики для задания

- Топ-10 стран по кол-ву инцидентов
- Топ-10 штатов США по кол-ву инцидентов
- Кол-во инцидентов в которых оружие было украдено/статус не известен

8

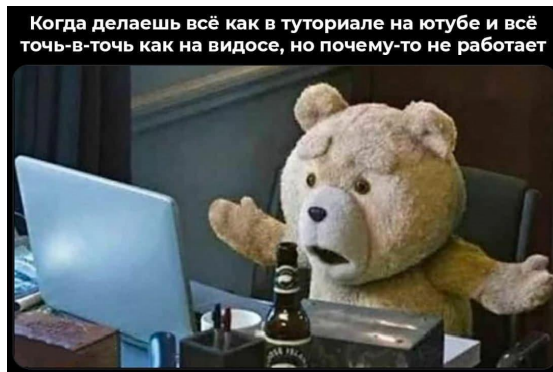
<https://www.kaggle.com/jboysen/spy-plane-finder>

Метрики для задания

- Топ-10 стран по кол-ву полетов
- Топ-3 самых популярных самолетов
- Топ-5 самых популярных самолета относительно года изготовления

Полезные ресурсы

- [Курс по Big Data на Stepik \(лучше брать только теорию, практика там не очень\)](#)
- [Блог Big Data School](#)
- Собственно ваш преподаватель, которому можно задавать вопросы



Вопросы?

Удачи!

The internet will make those bad words go away



Essential

Googling the Error Message

ORLY?

The Practical Developer
@ThePracticalDev