**ТЕХНИЧЕСКОЕ ЗАДАНИЕ**

*Анализ столкновений птиц с воздушными судами*

Описание: Изучение влияния погоды на столкновения птиц с самолетами. Проанализировать данные для проведения дальнейшего исследования и выявить закономерности в данных. На основе исследования будет построена модель предсказания столкновений птиц с самолётами.

Стек технологий DA: Обработка сырых данных, Python (pandas, seaborn, plotly и другие библиотеки на ваш выбор).

Цель проекта: проанализировать данные + передать данные для дальнейших исследований ваших коллег из DS-направления (будет отдельная Мастерская).

Что нужно передать команде DS:

– проанализировать какие факторы коррелируют со столкновениями

- погода?

- аэропорт, трафик в аэропорте?

- на вылете или прилете?

- есть ли тренды?

– как можно агрегировать по году-месяцу эти данные чтобы создать признаки для прогнозирования количества столкновений?

На выходе нужны данные в виде таблицы csv и описание инсайтов

год, месяц

количество полетов

аэропорт, вылет, прилет

найденные признаки по колонкам

**Этапы и задачи, ТЗ проекта:**

1. Предобработать данные, определить типы данных в таблице.
2. Обработать данные о погоде (WND, CIG, VIS, TMP, DEW, SLP), преобразовать их в отдельные столбцы.
3. Сагрегировать таблицу по месяцам, найти общие суммы полетов по аэропортам, найти среднемесячную температуру, сила ветра, возможно направление ветра, температуру точки росы и т.д. По направлению ветра возможно сделать отдельные колонки. Найти закономерности, зависимости одних факторов от других.
4. Проанализировать данные, найти интересные инсайты, построить визуализации, которые в дальнейшем помогут построить модель предсказания страйков для коллег из направления DS.

Источники данных: Коллеги из направления ДЕ собрали следующие данные о погоде: скорость и направление ветра, состояние неба, видимость, температура воздуха, температура точки росы, атмосферное давление. Эти показатели находятся в полях WND, CIG, VIS, TMP, DEW, SLP соответственно.

Полное описание полей по ссылке: <https://www.ncei.noaa.gov/data/global-hourly/doc/isd-format-document.pdf>

Данные из FAA и NCEI объединены в один датасет в формате CSV.   
Данные из BTS объединены в один датасет в формате CSV.

**Ограничения:**

FAA - все записи с 1/1/2018 по 12/31/2022

NCEI - все записи для строчек из датасета FAA, у которых указаны координаты. Если не указано время, то данные за 12:00.

BTS - данные для первых 10 аэропортов по количеству столкновений.

**Рекомендации по оформлению работы:**1. Описание проекта, название, цели, исходные данные

Краткие комментарии в коде

Гипотеза-исследование-выводы

Чистый код, PEP8

Меняете данные – должно быть подтверждение ДО и ПОСЛЕ

По итогу исследования можно подготовить короткий, но содержательный отчет.

2. Для сдачи проекта необходимо разместить работу в любом репозитории, например, в Git (опционально), яндекс/гугл диске и открыть доступ. Или прислать тетрадку Jupyter;

Ссылку на работу вносим в специальную форму (не забывайте оставлять контактную информацию);

Не забывайте размещать ФИО и Telegram-аккаунт в “шапке проекта” – это необходимо для ускорения отправки обратной связи;

После проверки ссылку на работу с комментариями пришлем в личные сообщения (ТГ указанный в форме);

При наличии критических ошибок в работе, вы получите обратную связь с комментариями и предложением доработать ваши проекты;