**Проект 1. Определение сентимента и предсказание доходности акций**

**Проект представляет собой несколько самостоятельных модулей (их не требуется объединять в один модуль)**

1. **Парсер сообщений Пульс**

- в папку с кодом должен быть помещен файл типа блокнот с указанием тикеров компаний, по которым требуется собрать информацию (пример: SBER MAGN – указываем последовательно через ENTER)

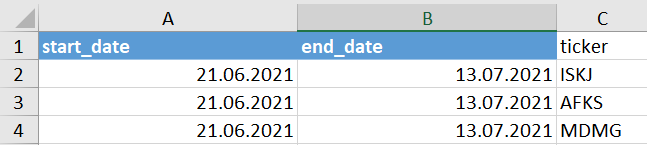
- запуск кода файла **«exаmple.py»**

- сохранение итоговых данных происходит в файл **all\_posts.xlsx**

1. **Модуль сбора торговых показателей компаний**

- в папку с кодом должен быть помещен файл **tickerlist.xlsx** с указанием тикеров компаний, по которым необходимо собрать данные, а также дата начала и окончания периода сбора данных

*Пример*

**

- в коде должна быть указана переменная со значением от 1 до 7, которая указывает таймфрейм сбора данных: periods = {'tick': 1, 'min': 2, '5min': 3, '10min': 4, '15min': 5, '30min': 6, 'hour': 7, 'daily': 8, 'week': 9, 'month': 10}. По умолчанию стоит часовой таймфрейм = 7.

- запуск кода файла **«Сбор торговых данных.ipynb»**

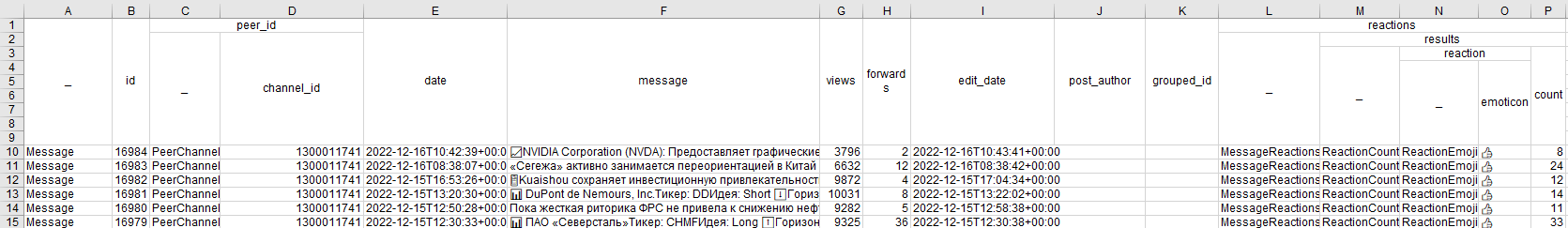
- после завершения работы кода собранные данные сохраняются в файл «**all\_companies\_stock\_prices.xlsx»**

1. **Модуль построения индекса сентимента по сообщениям инвесторов**

- определение сентимента производится с помощью обученной авторами модели, сохраненной в файле **«model\_cb.cbm»**, который должен находится в папке с запускаемым кодом

- после запуска кода **«TG AND PULSE\_SENTIMENT.ipynb»** последовательно происходят следующие шаги:

* модель дообучается на тренировочных данных из файла **«2\_Labeled sentiment data.xlsx - Sheet1»** (должен быть в папке с кодом)
* выполняем предобработку текста
* пробуем разные алгоритмы для классификации
* получаем тикеры сообщений (для получения тикеров сообщений из Телеграмма и последующего определения сентимента на вход должен быть подан файл **«Telegram.xlsx»**, который должен находиться в папке с кодом)



* анализируем сентимент ТГ
* анализируем сообщения из Пульса (файл с сообщениями должен находиться в папке с кодом и называться **«4\_tcs\_pulse\_posts.csv»**)

1. **Модуль построения индикаторов технического анализа и сигналов на покупку/продажу акций на основе торговых показателей акций компании**

**-** расчет технических индикаторов с помощью кода «Расчет технических индикаторов и сигналов.ipynb» на основе данных торгов по акциям из файла **«Технические\_инидкаторы\_input.xlsx»** (данный файл должен быть заранее подготовлен и находиться в папке с кодом)

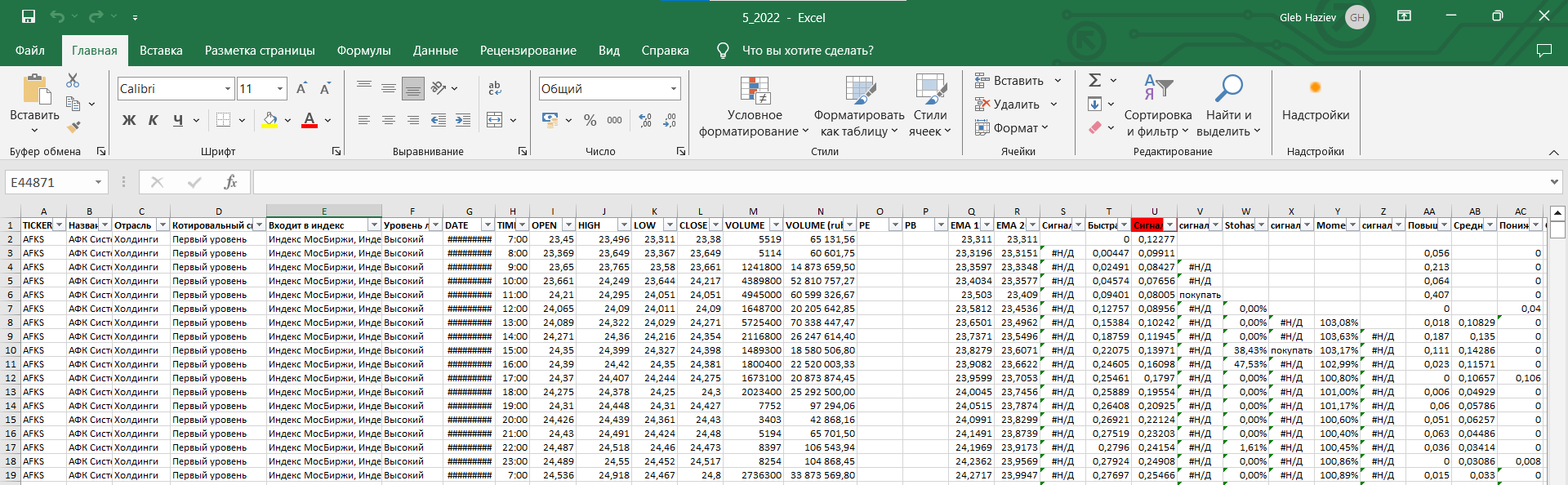
**-** после расчета технических индикаторов и сигналов на покупку/продажу акций на выходе получаем файл **«signals\_output.xlsx»**

1. **Модуль прогнозирования цен акций на часовых данных с учетом индексов сентимента и индикаторов технического анализа с применением моделей искусственного интеллекта**

**-** перед построением предсказательной модели изменения доходности происходит дообработка данных по сообщениям из социальных сетей и технических данных.

Для проведения нужных операция необходимо чтобы в папке с кодом были файлы, полученные на предыдущих шагах:

**«tg\_2022\_label.csv», «pulse\_2022\_label.csv», «5\_2022.xlsx».** Файл 5\_2022.xlsx – содержит в себе информацию по торгам акциями и сигналы на покупку/продажу акций



**-** запускаем код **«MERGE DATA AND PREDICT RETURN.ipynb»,** который ссылаясь на файлы, указанные в предыдущем пункте, произведет необходимые вычисления и построит итоговую модель предсказания доходности акций и сохранит результаты в excel файлы «results\_with\_RSMI.csv», «results\_no\_RSMI.csv»

results\_with\_RSMI.csv – показатели надежности моделей с использованием авторского индекса сентимента

results\_no\_RSMI.csv **-** показатели надежности моделей без использования авторского индекса сентимента

**6. Модуль выявления манипуляций**

- перед построением предсказательной модели выявления манипуляций необходимо загрузить в папку с кодом заранее подготовленный файл «train.xlsx»

- запускаем код «Классические модели.ipynb», который ссылаясь на файлы, указанные в предыдущем пункте, произведет необходимые вычисления и построит итоговую модель предсказания манипуляций с использованием методов логистической регрессии, а также стандартной модели Catboost

- запускаем код «Catboostcustom.ipynb», который ссылаясь на файлы, указанные в предыдущем пункте, произведет необходимые вычисления и построит итоговую модель предсказания манипуляций с использованием авторской кастомизированной модели Catboost