

Если договоримся, то первым заданием будет такое:

1. Парсим данные с нескольких сайтов (список ниже).
2. Сводим данные в одну БД
3. Ранжируем компании по заданным показателям
4. Выводим рейтинг компаний на свой сайт
5. Ежедневно парсим сайты на изменения
6. Обновляем рейтинг компаний
7. Выводим изменения на свой сайт и в тг канал
8. Через платный тг бот даём пользователям расширенную информацию по компаниям
9. Через платный аккаунт на сайте даём пользователям расширенную информацию

Сайты для парсинга:

<https://zakupki.gov.ru/epz/dishonestsupplier/search/results.html> - здесь данные о недобросовестных поставщиках. Формирует негативный рейтинг.

<https://reestr.nopriz.ru/sro/list> - здесь есть сведения о нарушениях компаний при соблюдении лицензий. Чем больше нарушений, и чем они регулярные - тем ниже рейтинг. Отсутствие нарушений формирует положительный рейтинг.

<http://rating.nopriz.ru/> - здесь берём готовый рейтинг

<https://egrz.ru/> - здесь сведения о прохождении компаний госэкспертиз проектов. Отрицательные заключения формируют негативный рейтинг, положительные - положительный. Если у компании хотя бы один негативный фактор, она отправляется в антирейтинг.

Требования к сайту - разбит по регионам/городам (в каждом свой рейтинг) - у каждой компании должна быть страница с информацией, которая формирует её рейтинг - в бесплатном доступе будет негативный

рейтинг компаний - в платном - положительный рейтинг + связь компаний по директорам

Задание на сейчас:

Предложите стек для парсинга каждого из указанных сайтов

Предложите принципиальную схему БД

Предложите стек для сайта (фронт, бэк, сервер)

Предложите реализацию по отправке данных в тг канал,

Дайте предложения по созданию бота

Предложите сроки исполнения проекта, с учётом предложенного Вами выше затрат времени.

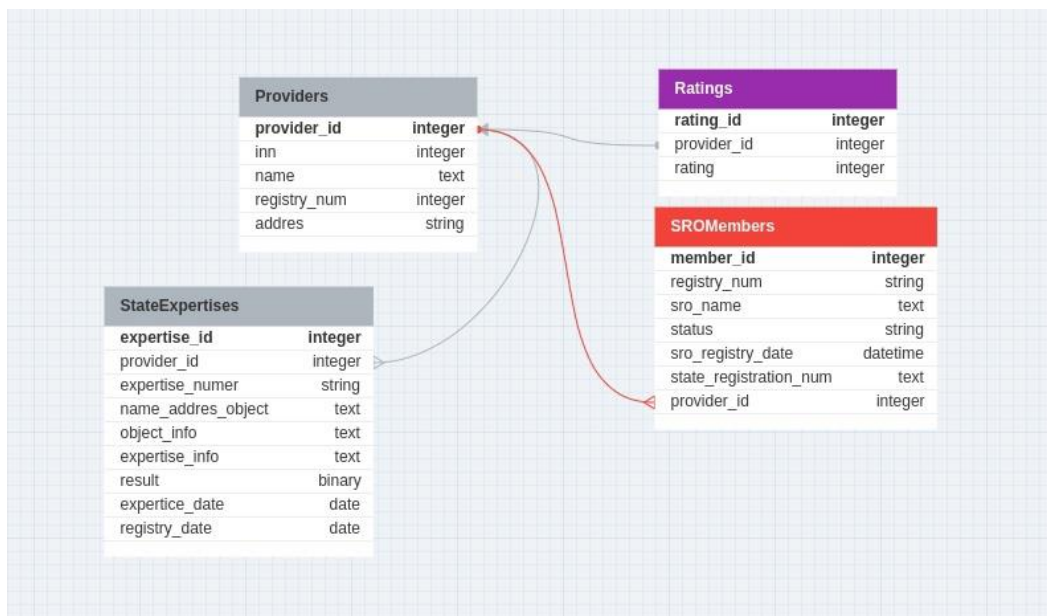
Прошу сначала задать уточняющие вопросы, а затем предоставить развёрнутый ответ на задание.

Стек парсинга:

№	Сайт	Какой он	Формат получаемых данных	Чем парсить
1	https://zakupki.gov.ru/epz/dishones	Статический, пагинация есть, в ссылке инфо о странице. Можно изменить	html	requests Beautifulsoup 4 или scrapy

	tsupplier/se arch/results .html	количество элементов до 500 для уменьшения количества запросов		
2	https://reestr.nopriz.ru/sro/list	Статический, пагинация есть, в ссылке инфы о странице нет, данные получаются путем отправки POST - запроса на сервер (https://reestr.nopriz.ru/api/sro/list) В теле запроса: {page: <номер страницы>, pageCount: "100"} Можно изменить количество элементов для уменьшения количества запросов	json	Requests или scrapy (требуется отправлять только запросы)
3	http://rating.nopriz.ru/ -	Статический, пагинация есть, в ссылке инфы о странице нет, данные получаются путем отправки POST - запроса на сервер (https://reestr.nopriz.ru/api/sro/list) В теле запроса form-data объект Поля: (PageNumber, ActivityDirectionId)	json	Requests или scrapy (требуется отправлять только запросы)
4	https://egrz.ru/	Статический, пагинации нет, в ссылке инфы о странице нет, однако можно производить поиск информации по ссылке которая используется в запросах при поиске (https://open-api.egrz.ru/api/PublicRegistrationBook/?\$filter=(contains(tolower(ExpertiseOrganizationInfo),tolower(%273442112404%27))%20or%20contains(tolower(ExpertiseNumber),tolower(%273442112404%27))%20or%20contains(tolower(WorkType),tolower(%273442112404%27))%20or%20contains(tolower(ExpertiseResultType),tolower(%273442112404%27))%20or%20contains(tolower(ExpertiseType),tolower(%273442112404%27))%20or%20contains(tolower(ExpertiseDocumentType),tolower(%273442112404%27))%20or%20contains(tolower(ExpertiseObjectNameAndAddress),tolower(%273442112404%27))%20or%20contains(tolower(EconomyEfficiencyInfo),tolower(%273442112404%27))%20or%20contains(tolower(PreviousExpertiseResults),tolower(%273442112404%27))%20or%20contains(tolower(DeveloperAndTechnicalCustomerOrganizationInfo),tolower(%273442112404%27))%20or%20contains(tolower(PlannerOrganizationInfo),tolower(%273442112404%27))%20or%20contains(tolower(SubjectRf),tolower(%273442112404%27)))&\$orderby=ExpertiseNumber%20asc%20&\$count=true&\$top=7&\$skip=0) top - количество строк данных, число 27344.... - инн, однако к инн добавляется 27 и запрос работает корректно, без 27 не работает	csv	Requests или scrapy (требуется отправлять только запросы)

Схема БД:



Стэк-сайта:

1. Фронтэнд: (я не фронтэндер) React
2. Бэкэнд: Flask
3. БД: PostgreSQL

Реализация взаимодействия с ТГ:

- Сделать общий API сайта (фронт) и ТГ-бота на Flask (бэк) RestAPI

ТГ бот:

- aiogram3