

A wide-angle, nighttime photograph of a city skyline, likely Chicago, featuring the Willis Tower (formerly Sears Tower) and other modern skyscrapers. The buildings are illuminated from within, and their reflections are clearly visible in the dark water of a frozen river or lake in the foreground. Bare trees line the riverbank on both sides.

# База стартапов Data Shamans

# Команда



Надежда  
Лосева

**Капитан**



Константин  
Некрасов



Анна  
Симонова



Анатолий  
Кайда



Анна  
Зверева



Дамир  
Уталиев

# Проблема

- Не существует единой удобной базы стартапов с красивым визуалом
- Имеющиеся решения – промежуточные, в лучшем случае покрывают лишь какой-то аспект деятельности
- Данные быстро устаревают

# Решение

- Готовая база данных с максимально подробной информации о стартапах и их признаках в различных географических зонах
- Удобное визуальное решение для комфортной работы с базой
- Использование технологических решений: SQL, Python, Requests, Gradio, ChatGPT и другие

# Что сделано на Хакатоне

## Результаты:

- ✓ Очищенные и обогащенные данные о 226 тысячах стартапов со всего мира
- ✓ Единая база данных SQLite
- ✓ Веб-интерфейс Gradio с удобным набором параметров для вывода данных

## Что предстоит сделать:

- ✓ Дополнять базу данных новыми данными
- ✓ Улучшить производительность базы данных, вынеся ее на коммерческий сервер
- ✓ Доработать веб-интерфейс, добавить визуализацию

**~50 баз  
исследовано  
~450 000  
обработанных  
записей**

# Собранный dataset

Собрали, очистили и обогатили базу данных:

**226 175 компаний**

**60 признаков**

**+ 9 таблиц доп  
данных**

- ✓ Данные очищены от дубликатов и обогащены из различных баз данных
- ✓ Все объекты имеют информацию по статусу и по раундам финансирования
- ✓ По большинству объектов есть данные по общему объему привлеченного финансирования
- ✓ Связанные таблицы. люди, события, фонды, слияния и поглощения, ipo, инвестиции

# Признаки 1

|                                |   |
|--------------------------------|---|
| <code>id</code>                | Тип объекта и код   |
| <code>entity_id</code>         | Код компании  |
| <code>name</code>              | Название компании   |
| <code>normalized_name</code>   | Название, приведенное в нижний регистр  |
| <code>category_code</code>     | Основная категория  |
| <code>status</code>            | Статус компании (1-работает, 2-проинвестирована, 3-вышла на IPO, 4-закрылась) |
| <code>founded_at</code>        | Открыта в году  |
| <code>closed_at</code>         | Закрыта в году  |
| <code>domain</code>            | URL домен   |
| <code>homepage_url</code>      | URL официального сайта  |
| <code>twitter_username</code>  | Официальный Twitter аккаунт   |
| <code>logo_url</code>          | URL логотипа  |
| <code>logo_width</code>        | Ширина логотипа   |
| <code>logo_height</code>       | Высота логотипа   |
| <code>short_description</code> | Краткое описание компании   |
| <code>overview</code>          | Подробное описание деятельности компании                                      |
| <code>tag_list</code>          | Список тэгов  |
| <code>country_code</code>      | Код страны  |
| <code>state_code</code>        | Код штата/региона   |
| <code>city</code>              | Город   |
| <code>region</code>            | Регион  |

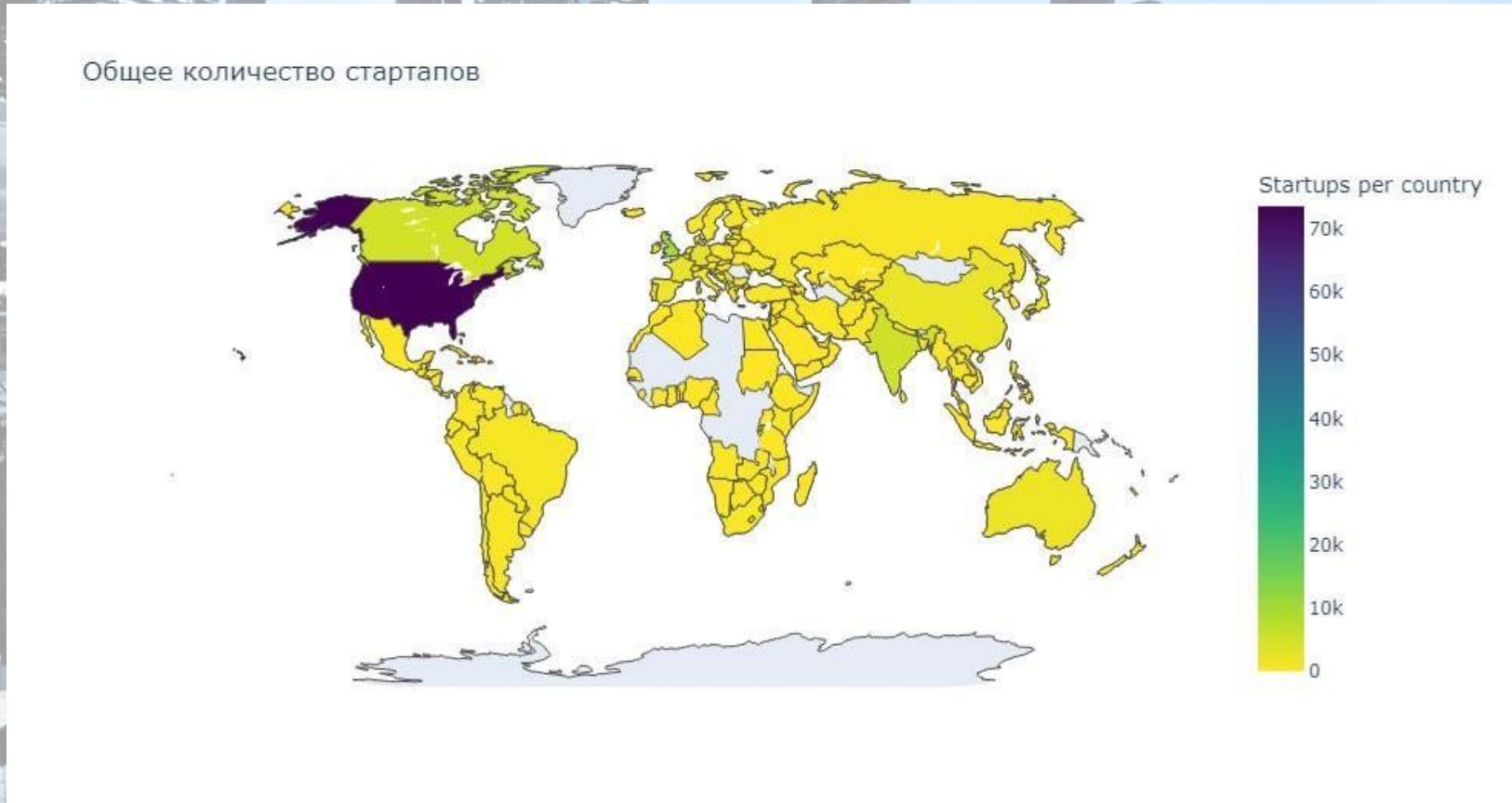
# Признаки 2

|                     |   |
|---------------------|---|
| first_investment_at | Дата первой инвестиции (агрегированные данные Crunch и Y-combinator)  |
| last_investment_at  | Дата последней зафиксированной инвестиции   |
| investment_rounds   | Количество пройденных раундов инвестиций  |
| invested_companies  | В какое кол-во компаний проинвестировано  |
| first_funding_at    | Дата первого раунда инвестирования фондом   |
| last_funding_at     | Дата последнего зафиксированного раунда инвестирования фондом   |
| funding_rounds      | Кол-во раундов фондового инвестирования   |
| funding_total_usd   | Всего поднято фондов в USD (если на конце присутствует 1 - то проинвестировано на дополнительную неизвестную сумму) |
| first_milestone_at  | Дата первого зафиксированного события, связанного с компанией   |
| last_milestone_at   | Дата последнего зафиксированного события, связанного с компанией  |
| milestones          | Кол-во связанных с компанией событий  |
| relationships       | Кол-во связанных персон   |
| updated_at          | Дата последнего обновления записи   |
| category_list       | Список категорий  |
| crunchbase_profile  | Ссылка на страницу в базе crunch  |
| country             | Страна дислокации   |
| founders            | Имена основателей   |
| investors           | Названия инвесторов   |
| address             | Адрес головного офиса   |
| state               | Штат/провинция  |

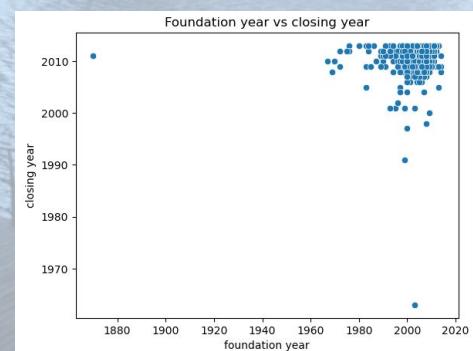
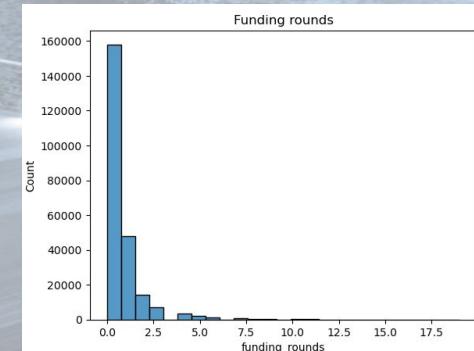
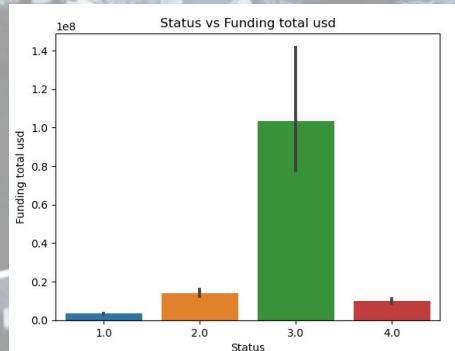
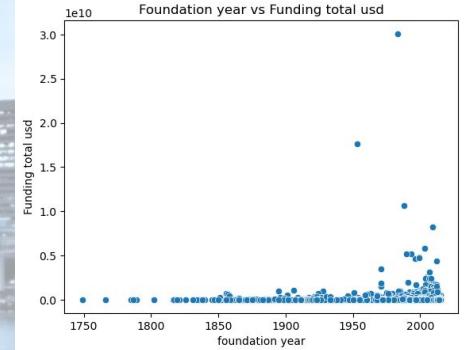
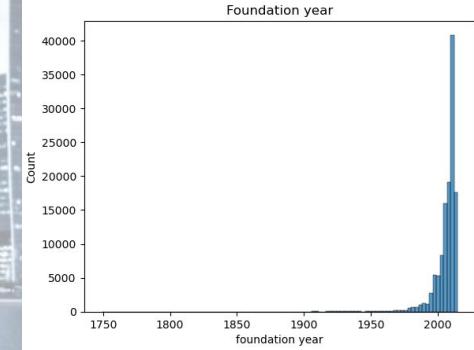
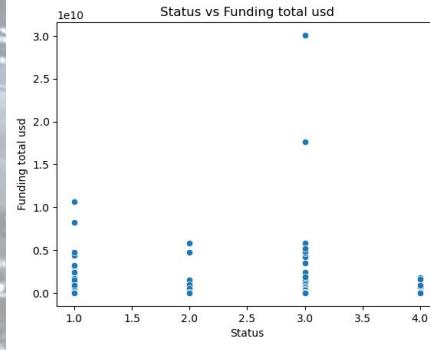
# Признаки 3

|                    |   |
|--------------------|---|
| seed_profile       | страница компании на Y-combinator   |
| y_combinator_year  | Год участия в Y-combinator  |
| city_2023          | Город 2023 (оставлен в датасете для последующего уточнения, была ли релокация или открыт новый региональный офис) |
| country_2023       | Страна 2023 (аналогично)  |
| homepage_url_2023  | URL 2023 года (аналогично)  |
| Industries         | Отрасли промышленности  |
| Employees          | Кол-во сотрудников  |
| Company_values     | Ценности компании   |
| Org_structure      | Организационная структура   |
| Teams and qty      | Топовые команды компании  |
| acquired_at_first  | Первое осуществленное поглощение  |
| acquired_at_last   | Последнее поглощение  |
| price_amount       | Заплаченная за поглощения сумма   |
| term_code          | Условия поглощений  |
| acquired_object_id | Код поглощенной компании  |
| acquired_by        | Кем поглощена компания (код поглотившего)   |
| acquired_at        | Дата продажи  |
| final_sell_price   | Цена продажи компании   |
| acquired_on_terms  | На каких условиях продана   |

# Интерактивная карта



# Собранный dataset



# Преимущества



**Быстрый справочник** по стартапам и их инвестиционной истории



**Инвесторы** могут использовать базу данных стартапов для выявления **потенциальных инвестиционных возможностей** (отбора и фильтрации стартапов на основе конкретных критериев)



База данных может дать ценную информацию о **возникающих тенденциях и изменениях на рынке**. Это может быть полезно для компаний, которые стремятся **опередить конкурентов** и выявить новые возможности в своей отрасли



**Инкубаторы и акселераторы** могут использовать базу данных стартапов для выявления потенциальных **кандидатов** на участие в их программах. Это может помочь им найти стартапы, которые хорошо подходят для их программ

# Модель монетизации

## На первом этапе:

Определим целевой рынок: инвесторы, исследователи, аналитики, инкубаторы и акселераторы, а также предприятия, желающие расширить свою сеть

Определим ролевую модель реализации продукта и его стоимость

Можно использовать онлайн-платформы LinkedIn, AngelList и Crunchbase и др, чтобы рекламировать свою базу данных потенциальным покупателям

## На последующих этапах:

Предложим бесплатную пробную версию доступа

Далее – доступ по подписке. Обеспечиваем постоянный мониторинг актуальности, техническую поддержку пользователям

Предлагаем ассоциированные услуги – продажа части базы данных, аналитические услуги, собственный образовательный контент и т.д.

# Ссылка на Gradio

**<https://2ec669cb325455cea3.gradio.live>**

# Ссылка на GitHub

**[https://github.com/loseff-n/hakaton\\_2t](https://github.com/loseff-n/hakaton_2t)**

# Ссылка на GDrive

<https://drive.google.com/drive/folders/1SMVCifbNe-kT4cweFQLDnAOfhCq7hpy?usp=sharing>