

Исследование данных об инвестициях венчурных фондов в компании-стартапы

Статус проекта: *завершен*

Использованные библиотеки: pandas, sqlalchemy, numpy, sys, datetime, warnings

Описание проекта:

Мы будем работать с базой данных, которая хранит информацию о венчурных фондах и инвестициях в компании-стартапы. Эта база данных основана на датасете [Startup Investments](#), опубликованном на популярной платформе для соревнований по исследованию данных Kaggle.

Анализировать рынок инвестиций без подготовки может быть непросто. Поэтому сначала познакомимся с важными понятиями, которые нам встретятся в работе с базой данных:

Венчурные фонды — это финансовые организации, которые могут позволить себе высокий риск и инвестировать в компании с инновационной бизнес-идеей или разработанной новой технологией, то есть в стартапы.

Цель венчурных фондов — в будущем получить значительную прибыль, которая в разы превысит размер их трат на инвестиции в компанию. Если стартап подорожает, венчурный фонд может получить долю в компании или фиксированный процент от её выручки.

Чтобы процесс финансирования стал менее рискованным, его делят на стадии — *раунды*. Тот или иной раунд зависит от того, какого уровня развития достигла компания.

Первые этапы — *предпосевной и посевной раунды*. Предпосевной раунд предполагает, что компания как таковая ещё не создана и находится в стадии замысла. Следующий — посевной — раунд знаменует рост проекта: создатели компании разрабатывают бизнес-модель и привлекают инвесторов.

Если компании требуется ментор или наставник — она привлекает *бизнес-ангела*. Бизнес-ангелы — инвесторы, которые помимо финансовой поддержки предлагают экспертную помощь. Такой раунд называют *ангельским*.

Когда стартап становится компанией с проверенной бизнес-моделью и начинает зарабатывать самостоятельно, предложений инвесторов становится больше. Это *раунд А*, а за ним следуют и другие: *В, С, D* — на этих этапах компания активно развивается и готовится к IPO.

Иногда выделяют *венчурный раунд* — финансирование, которое могло поступить от венчурного фонда на любом этапе: начальном или более позднем.

В данных об инвестициях нам встретятся упоминания раундов, но самостоятельный проект не предполагает, что мы должны разбираться в их специфике лучше любого инвестора. Главное — понимать, как устроены данные.

Мы будем работать с базой данных StackOverflow — сервиса вопросов и ответов о программировании. StackOverflow похож на социальную сеть — пользователи сервиса задают вопросы, отвечают на посты, оставляют комментарии и ставят оценки другим ответам. Работа будет проходить с версией базы, где хранятся данные о постах за 2008 год, но в таблицах будет информация и о более поздних оценках, которые эти посты получили.

Наша задача произвести различные выгрузки данных с помощью SQL-запросов и проанализировать результаты.

Исследование пройдёт в три этапа:

1. Изучение схемы (ER-диаграммы).
2. Обзор данных.
3. Выгрузки данных с помощью SQL-запросов и анализ.

В данных, которые мы будем использовать в проекте, упоминаются сервисы и компании, запрещённые в РФ в настоящий момент.