

Study of data on investments of venture capital funds in start-up companies

Project status: *completed*

Libraries used: pandas, sqlalchemy, numpy, sys, datetime, warnings

Description of the project:

We will work with a database that stores information about venture funds and investments in start-up companies. This database is based on the [Startup Investments](#) dataset published on the popular data mining competition platform Kaggle. **Analyzing the investment market without preparation can be difficult. Therefore, first, let's get acquainted with the important concepts that we will meet in working with the database: **

Venture funds are financial institutions that can afford high risk and invest in companies with an innovative business idea or new technology developed, i.e. start-ups.

The goal of venture funds is to get a significant profit in the future, which will be several times higher than the amount of their spending on investments in the company. If a startup rises in price, a venture capital fund can receive a stake in the company or a fixed percentage of its revenue.

To make the financing process less risky, it is divided into stages - * rounds *. This or that round depends on what level of development the company has reached.

The first stages are *pre-seed and seed rounds*. The pre-seed round assumes that the company as such has not yet been created and is in the concept stage. The next - seed - round marks the growth of the project: the founders of the company develop a business model and attract investors.

If a company needs a mentor or mentor, it attracts a *business angel*. Business angels are investors who, in addition to financial support, offer expert assistance. Such a round is called *angelic*.

When a startup becomes a company with a proven business model and starts making money on its own, there are more investor offers. This is *round A*, and others follow: *B, C, D* - at these stages the company is actively developing and preparing for an IPO.

Sometimes a *venture round* is singled out - financing that could come from a venture fund at any stage: initial or later.

We'll see rounds mentioned in the investment data, but being a standalone project doesn't mean we have to understand them better than any investor. The main thing is to understand how the data is arranged.

We will work with the StackOverflow database - a service for questions and answers about programming. StackOverflow is like a social network - users of the service ask questions, reply to posts, leave comments, and rate other answers. The work will be carried out with the version of the database where data on posts for 2008 is stored, but the tables will also contain information about later ratings that these posts received.

Our task is to perform various data uploads using SQL queries and analyze the results.

Research will take place in three stages:

1. Studying the scheme (ER-diagrams).
2. Data review.
3. Unloading data using SQL queries and analysis.

The data that we will use in the project mentions services and companies that are currently banned in the Russian Federation.