

# Бухгалтерская отчетность 1,6 млн российских компаний *Как получить и как использовать?*

Евгений Погребняк

Version 0.0.1, January 29, 2019

Мы сделали интерфейс для удобного доступа к открытым данным корпоративной отчетности российских компаний (РСБУ) за 2012-2017 год. Пользователи могут получить эти данные в виде CSV файлов.

Для анализа с помощью [pandas](#) или [R](#) в CSV файлах записаны заголовки с названиями столбцов (переменных). Размер файлов, составляющий 0,7-2,3 Гб за один год, уменьшается примерно в два раза за счет удаления пустых и неиспользуемых столбцов.

Код для доступа к данным и примеры их использования находятся в репозитории <https://github.com/ru-corporate/sandbox>.

## Как получить отчетность?

1. Скачать файлы Росстата
2. Ввести удобные названия колонок
3. Уменьшить размер файлов за счет пустых и неиспользуемых столбцов
4. Привести все строки к одинаковым единицам измерения (тыс. руб.)
5. Преобразовать отдельные колонки в более удобные
  - короткое название компании
  - код ОКВЭД разбить на три уровня
6. Устранить повторы в данных
7. Указать тип переменных по колонкам, чтобы ускорить чтение данных

## Как использовать?

Сферы применения отчетности:

- рейтинги крупнейших компаний и отраслевые рейтинги
- финансовая устойчивость и эффективность компаний
- оценка стоимости компаний, кредитный риск, банкротства
- оценка экономической концентрации в отрасли
- анализ налоговых поступлений и налоговой нагрузки
- инвестиционный процесс (кто инвестирует в расширение мощностей)
- региональная экономика и планирование (городской транспорт, ЖКХ, ТБО)

## Что хотим показать?

- краткий код доступа к данным
- полноту данных
- разметку по отраслям

- крупнейшие холдинги
- красивые графики

```
from booby import read_dataframe
df = read_dataframe(2017)
```

## Возможные разделы

- Карта экономики
- Использование в обучении финансам

## Учим asciidoc

### Расширения:

- Можем ли вставить таблицы из `pd.DataFrame`?
- [asciidoctor-new-noteworthy-beyond](#)

### Базовые знания:

- [asciidoc-syntax-quick-reference.adoc](#)
- [asciidoc\\_example.adoc](#)