# Кейс-1. Кредитный скоринг юридических лиц

* + ~~Все данные~~ ***~~train~~*** ~~имеют одинаковое количество столбцов~~
  + ~~Все данные~~ ***~~train~~*** ~~совпадают с теми, которые были даны~~
  + ~~Все данные~~ ***~~test~~*** ~~имеют одинаковое количество столбцов~~
  + ~~Все данные~~ ***~~test~~*** ~~совпадают с теми, которые были даны (за исключением target) - проверка через объединение двух разных тестовых данных и сравнения~~
  + ~~Все данные~~ ***~~test~~*** ~~не содержат столбца target~~
  + ***~~Baseline~~*** ~~имеет четкую структуру~~
  + ***Baseline*** содержит фото мероприятия в начале файла
  + ***~~Baseline~~*** ~~содержит текст про правообладателей в конце файла~~
  + ***~~Baseline~~*** ~~содержит установку библиотек для работы~~
  + ***~~Baseline~~*** ~~содержит отсортированный в правильном порядке список файлов train - от train\_1 до train\_10~~
  + ***~~Baseline~~*** ~~содержит отсортированный в правильном порядке список файлов test - от test\_1 до test\_10~~
  + ***~~Baseline~~*** ~~не содержит отсылок на тестовые данные с таргетами~~
  + ***~~Baseline~~*** ~~содержит только 2 пути к папкам - train и test (path\_train и path\_test соответственно)~~
  + ***~~Baseline~~*** ~~содержит объединение файлов train и проверку того, что все файлы - только train~~
  + ***~~Baseline~~*** ~~содержит объединение файлов test и проверку того, что все файлы - только test~~
  + ***~~Baseline~~*** ~~содержит разделение данных на тренировочную и тестовую часть~~
  + ***~~Baseline~~*** ~~содержит разделение данных на тренировочные и валидационные~~
  + ***~~Baseline~~*** ~~содержит обучение модели на тренировочной части тренировочных данных и предсказании на валидационных~~
  + ***~~Baseline~~*** ~~содержит подсчет метрики roc-auc для валидационной части данных~~
  + ***~~Baseline~~*** ~~содержит комментарии только с большой буквы~~
  + ***~~Baseline~~*** ~~содержит идентичные комментарии с другими бейзлайнами~~
  + ***~~Baseline~~*** ~~содержит обучение новой модели на всех тренировочных данных~~
  + ***~~Baseline~~*** ~~содержит предсказание для тестовых данных~~
  + ***~~Baseline~~*** ~~содержит получение из предсказания только вероятностей перехода в кластер 1~~
  + ***~~Baseline~~*** ~~содержит объединение предсказания с основной частью данных~~
  + ***~~Baseline~~*** ~~содержит сохранение id и предсказаний в CSV-формат без индексов~~ **~~(baseline\_submission\_case1.csv)~~**
  + ~~В отдельном файле произведен расчет метрики roc-auc для тестовых данных (~~**~~baseline\_submission\_case1.csv~~**~~)~~
  + ~~Подтверждено, что в файле~~ **~~baseline\_submission\_case1.csv~~** ~~правильный порядок столбцов - сначала~~ **~~id~~**~~, затем~~ **~~target~~**
  + ~~Подтверждено, что~~ **~~baseline\_submission\_case1.csv~~** ~~имеет верную размерность и что его можно использовать для предсказания~~

# Кейс-2. Склонность физических лиц к инвестициям

* + ~~Все данные~~ ***~~train~~*** ~~имеют одинаковое количество столбцов~~
  + ~~Все данные~~ ***~~train~~*** ~~совпадают с теми, которые были даны~~
  + ~~Все данные~~ ***~~test~~*** ~~имеют одинаковое количество столбцов~~
  + ~~Все данные~~ ***~~test~~*** ~~совпадают с теми, которые были даны (за исключением target) - проверка через объединение двух разных тестовых данных и сравнения~~
  + ~~Все данные~~ ***~~test~~*** ~~не содержат столбца target~~
  + ***~~Baseline~~*** ~~имеет четкую структуру~~
  + ***Baseline*** содержит фото мероприятия в начале файла
  + ***~~Baseline~~*** ~~содержит текст про правообладателей в конце файла~~
  + ***~~Baseline~~*** ~~содержит установку библиотек для работы~~
  + ***~~Baseline~~*** ~~содержит отсортированный в правильном порядке список файлов train - от train\_1 до train\_10~~
  + ***~~Baseline~~*** ~~содержит отсортированный в правильном порядке список файлов test - от test\_1 до test\_10~~
  + ***~~Baseline~~*** ~~не содержит отсылок на тестовые данные с таргетами~~
  + ***~~Baseline~~*** ~~содержит только 2 пути к папкам - train и test (path\_train и path\_test соответственно)~~
  + ***~~Baseline~~*** ~~содержит объединение файлов train и проверку того, что все файлы - только train~~
  + ***~~Baseline~~*** ~~содержит объединение файлов test и проверку того, что все файлы - только test~~
  + ***~~Baseline~~*** ~~содержит разделение данных на тренировочную и тестовую часть~~
  + ***~~Baseline~~*** ~~содержит разделение данных на тренировочные и валидационные~~
  + ***~~Baseline~~*** ~~содержит обучение модели на тренировочной части тренировочных данных и предсказании на валидационных~~
  + ***~~Baseline~~*** ~~содержит подсчет метрики roc-auc для валидационной части данных~~
  + ***~~Baseline~~*** ~~содержит комментарии только с большой буквы~~
  + ***~~Baseline~~*** ~~содержит идентичные комментарии с другими бейзлайнами~~
  + ***~~Baseline~~*** ~~содержит обучение новой модели на всех тренировочных данных~~
  + ***~~Baseline~~*** ~~содержит предсказание для тестовых данных~~
  + ***~~Baseline~~*** ~~содержит получение из предсказания только вероятностей перехода в кластер 1~~
  + ***~~Baseline~~*** ~~содержит объединение предсказания с основной частью данных~~
  + ***~~Baseline~~*** ~~содержит сохранение id и предсказаний в CSV-формат без индексов (baseline\_submission\_case2.csv)~~
  + ~~В отдельном файле произведен расчет метрики roc-auc для тестовых данных~~ **~~(baseline\_submission\_case2.csv)~~**
  + ~~Подтверждено, что в файле~~ **~~baseline\_submission\_case2.csv~~** ~~правильный порядок столбцов - сначала~~ **~~id~~**~~, затем~~ **~~target~~**
  + ~~Подтверждено, что~~ **~~baseline\_submission\_case2.csv~~** ~~имеет верную размерность и что его можно использовать для предсказания~~

# Кейс-3. Отток юридических лиц из расчетно-кассового обслуживания

* + ~~Все данные~~ ***~~train~~*** ~~имеют одинаковое количество столбцов~~
  + ~~Все данные~~ ***~~train~~*** ~~совпадают с теми, которые были даны~~
  + ~~Все данные~~ ***~~test~~*** ~~имеют одинаковое количество столбцов~~
  + ~~Все данные~~ ***~~test~~*** ~~совпадают с теми, которые были даны (за исключением target) - проверка через объединение двух разных тестовых данных и сравнения~~
  + ~~Все данные~~ ***~~test~~*** ~~не содержат столбца target~~
  + ***~~Baseline~~*** ~~имеет четкую структуру~~
  + ***Baseline*** содержит фото мероприятия в начале файла
  + ***~~Baseline~~*** ~~содержит текст про правообладателей в конце файла~~
  + ***~~Baseline~~*** ~~содержит установку библиотек для работы~~
  + ***~~Baseline~~*** ~~содержит отсортированный в правильном порядке список файлов train - от train\_1 до train\_10~~
  + ***~~Baseline~~*** ~~содержит отсортированный в правильном порядке список файлов test - от test\_1 до test\_10~~
  + ***~~Baseline~~*** ~~не содержит отсылок на тестовые данные с таргетами~~
  + ***~~Baseline~~*** ~~содержит только 2 пути к папкам - train и test (path\_train и path\_test соответственно)~~
  + ***~~Baseline~~*** ~~содержит объединение файлов train и проверку того, что все файлы - только train~~
  + ***~~Baseline~~*** ~~содержит объединение файлов test и проверку того, что все файлы - только test~~
  + ***~~Baseline~~*** ~~содержит разделение данных на тренировочную и тестовую часть~~
  + ***~~Baseline~~*** ~~содержит разделение данных на тренировочные и валидационные~~
  + ***~~Baseline~~*** ~~содержит обучение модели на тренировочной части тренировочных данных и предсказании на валидационных~~
  + ***~~Baseline~~*** ~~содержит подсчет метрики roc-auc для валидационной части данных~~
  + ***~~Baseline~~*** ~~содержит комментарии только с большой буквы~~
  + ***~~Baseline~~*** ~~содержит идентичные комментарии с другими бейзлайнами~~
  + ***~~Baseline~~*** ~~содержит обучение новой модели на всех тренировочных данных~~
  + ***~~Baseline~~*** ~~содержит предсказание для тестовых данных~~
  + ***~~Baseline~~*** ~~содержит получение из предсказания только вероятностей перехода в кластер 1~~
  + ***~~Baseline~~*** ~~содержит объединение предсказания с основной частью данных~~
  + ***~~Baseline~~*** ~~содержит сохранение id и предсказаний в CSV-формат без индексов (baseline\_submission\_case3.csv)~~
  + ~~В отдельном файле произведен расчет метрики roc-auc для тестовых данных~~ **~~(baseline\_submission\_case3.csv)~~**
  + ~~Подтверждено, что в файле~~ **~~baseline\_submission\_case3.csv~~** ~~правильный порядок столбцов - сначала~~ **~~id~~**~~, затем~~ **~~target~~**
  + ~~Подтверждено, что~~ **~~baseline\_submission\_case3.csv~~** ~~имеет верную размерность и что его можно использовать для предсказания~~