**Модель данных для анализа одобрения кредитной заявки.**

Структура данных имеет вид таблицы данных по субъектам кредитных заявок вида:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Имя поля* | Фамилия | Имя | Пол | Возраст | Род деятельности | Сфера деятельности | Доход | Залог | Наличие счета |
| *Тип данных* | Текст | Текст | Текст / Бин | Целое | Текст | Текст | Целое | Бин / лист | Бин |

В реальности первичным ключом подобной таблицы идёт номер паспорта плюс проверка на уникальность, так как фамилии (особенно у женщин) могут меняться.

Алгоритм процесса моделирования и анализа данных следующий:

1. Находим список людей в формате .xlsx (Несложно)
2. Парсим имя-фамилию (текст по столбцам) по следующим правилам:
   1. Фамилии:
      1. Если фамилия заканчивается на -ов, - ев, -ин, -ий – муж.
      2. Если на -ова, -ева, -ина, -ая – жен.
      3. Иначе – универсальная
      4. Под массив фамилий целесообразно создать отдельную таблицу
   2. Имена:
      1. Под именами правила достаточно размытые
      2. Если имя заканчивается на -а, -я – женское; исключение – Никита, Илья
      3. Если имя заканчивается на согласную – мужское; исключение - любовь