

МЕТОДЫ ОЧИСТКИ ДАННЫХ: СВЕРКА ПО СПРАВОЧНИКАМ



bic_list.lgd

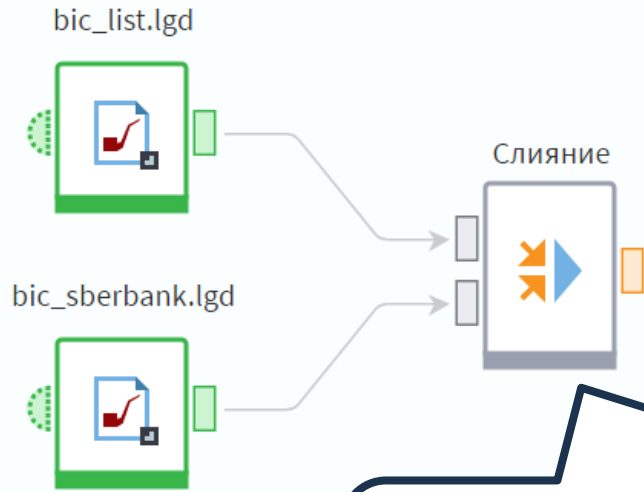


Первый (**bic_list.lgd**) представляет собой загрязненный список из некоторых банковских идентификационных кодов (БИК) отделений Сбербанка: в нем присутствуют опечатки, ошибки, некорректные БИК.

bic_sberbank.
lgd



Второй набор (**bic_sberbank.lgd**) — полный, стандартизированный справочник БИК отделений Сбербанка, в котором не содержится ошибок.



Нам необходимо проверить, какие БИК в первом списке являются корректными. Дополнительно мы добавим к списку наименование отделения. Для этого воспользуемся узлом слияния.

Тип операции

● Левое соединение

Фильтрация

Столбцы основного набора данных

ab БИК

Фильтрация

Столбцы присоединяемого набора данных

ab БИК

ab Отделение

В настройке узла используем **Левое соединение** и добавляем связь между полями **БИК**.

Выходной набор данных

| # | ab БИК | ab Отделение |
|----|-----------|--|
| 1 | 040173604 | АЛТАЙСКОЕ ОТДЕЛЕНИЕ N8644 ПАО СБЕРБАНК |
| 2 | 04049602 | <null> |
| 3 | 040407627 | КРАСНОЯРСКОЕ ОТДЕЛЕНИЕ N 8646 ПАО СБЕРБАНК |
| 4 | 040702615 | СТАВРОПОЛЬСКОЕ ОТДЕЛЕНИЕ N5230 ПАО СБЕРБАНК |
| 5 | 040813608 | ДАЛЬНЕВОСТОЧНЫЙ БАНК ПАО СБЕРБАНК |
| 6 | 041117601 | АРХАНГЕЛЬСКОЕ ОТДЕЛЕНИЕ N 8637 ПАО СБЕРБАНК |
| 7 | 041203602 | АСТРАХАНСКОЕ ОТДЕЛЕНИЕ N8625 ПАО СБЕРБАНК |
| 8 | 41403633 | <null> |
| 9 | 41501601 | <null> |
| 10 | 04178602 | <null> |
| 11 | 041806647 | ВОЛГОГРАДСКОЕ ОТДЕЛЕНИЕ N8621 ПАО СБЕРБАНК |
| 12 | 041909644 | ВОЛОГОДСКОЕ ОТДЕЛЕНИЕ N8638 ПАО СБЕРБАНК |
| 13 | 042202603 | ВОЛГО-ВЯТСКИЙ БАНК ПАО СБЕРБАНК |
| 14 | 042406608 | ИВАНОВСКОЕ ОТДЕЛЕНИЕ N8639 ПАО СБЕРБАНК |
| 15 | 42520607 | <null> |
| 16 | 042748634 | КАЛИНИНГРАДСКОЕ ОТДЕЛЕНИЕ N8626 ПАО СБЕРБАНК |
| 17 | 042809679 | ТВЕРСКОЕ ОТДЕЛЕНИЕ N8607 ПАО СБЕРБАНК |
| 18 | 042908612 | КАЛУЖСКОЕ ОТДЕЛЕНИЕ N8608 ПАО СБЕРБАНК |
| 19 | 043207612 | КЕМЕРОВСКОЕ ОТДЕЛЕНИЕ N8615 ПАО СБЕРБАНК |
| 20 | 04304609 | <null> |
| 54 | 043469623 | КОСТРОМСКОЕ ОТДЕЛЕНИЕ N 8640 ПАО СБЕРБАНК |

Тогда на выходе получаем следующий набор данных. Здесь если информация об отделении банка отсутствует, значит БИК в исходном списке некорректный.

Состояние входа


Вход активирован

Активировано

ab Отделение не пустое ×



В настройках фильтрации зададим
Отделение не пустое.

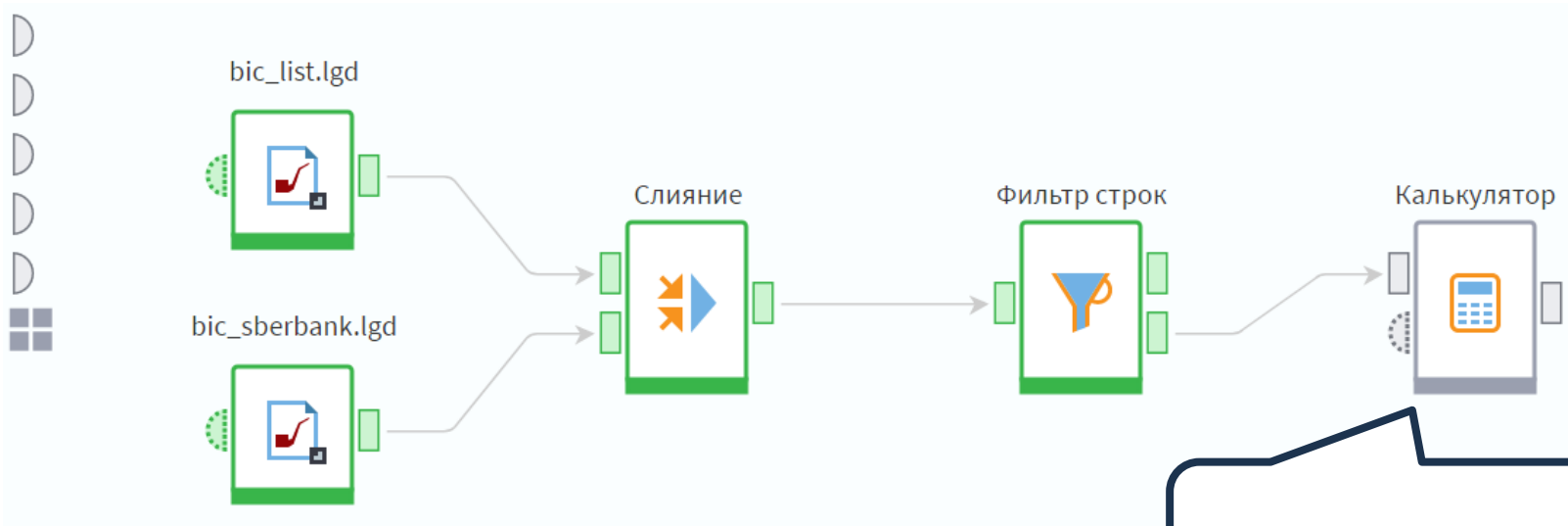
 Удалить все условия

Теперь на первом выходе мы получили список идентификаторов, которые точно были указаны корректно.

| # | ab БИК | ab Отделение |
|----|-----------|---|
| 1 | 040173604 | АЛТАЙСКОЕ ОТДЕЛЕНИЕ N8644 ПАО СБЕРБАНК |
| 2 | 040407627 | КРАСНОЯРСКОЕ ОТДЕЛЕНИЕ N 8646 ПАО СБЕРБАНК |
| 3 | 040702615 | СТАВРОПОЛЬСКОЕ ОТДЕЛЕНИЕ N5230 ПАО СБЕРБАНК |
| 4 | 040813608 | ДАЛЬНЕВОСТОЧНЫЙ БАНК ПАО СБЕРБАНК |
| 5 | 041117601 | АРХАНГЕЛЬСКОЕ ОТДЕЛЕНИЕ N 8637 ПАО СБЕРБАНК |
| 6 | 041203602 | АСТРАХАНСКОЕ ОТДЕЛЕНИЕ N8625 ПАО СБЕРБАНК |
| 7 | 041806647 | ВОЛГОГРАДСКОЕ ОТДЕЛЕНИЕ N8621 ПАО СБЕРБАНК |
| 8 | 041909644 | ВОЛОГОДСКОЕ ОТДЕЛЕНИЕ N8638 ПАО СБЕРБАНК |
| 9 | 042202603 | ВОЛГО-ВЯТСКИЙ БАНК ПАО СБЕРБАНК |
| 10 | 042406608 | ИВАНОВСКОЕ ОТДЕЛЕНИЕ N8639 ПАО СБЕРБАНК |
| 11 | 042748634 | КАЛИНИНГРАДСКОЕ ОТДЕЛЕНИЕ N8626 ПАО СБЕРБАНК |
| 12 | 042809679 | ТВЕРСКОЕ ОТДЕЛЕНИЕ N8607 ПАО СБЕРБАНК |
| 13 | 042908612 | КАЛУЖСКОЕ ОТДЕЛЕНИЕ N8608 ПАО СБЕРБАНК |
| 14 | 043207612 | КЕМЕРОВСКОЕ ОТДЕЛЕНИЕ N8615 ПАО СБЕРБАНК |
| 15 | 043469623 | КОСТРОМСКОЕ ОТДЕЛЕНИЕ N 8640 ПАО СБЕРБАНК |
| 16 | 043601607 | ПОВОЛЖСКИЙ БАНК ПАО СБЕРБАНК |
| 17 | 043735650 | КУРГАНСКОЕ ОТДЕЛЕНИЕ N8599 ПАО СБЕРБАНК |
| 18 | 043807606 | КУРСКОЕ ОТДЕЛЕНИЕ N8596 ПАО СБЕРБАНК |
| 19 | 044206604 | ЛИПЕЦКОЕ ОТДЕЛЕНИЕ N8593 ПАО СБЕРБАНК |
| 20 | 044442607 | СЕВЕРО-ВОСТОЧНОЕ ОТДЕЛЕНИЕ N8645 ПАО СБЕРБАНК |
| 42 | 044525225 | ПАО СБЕРБАНК |

На втором выходе — 12 записей, которые содержат ошибки. Как известно, БИК должен начинаться с 04 и содержать 9 знаков. Легко заметить, что некоторые коды начинаются с 4. Можно предположить, что в таких случаях в БИК отсутствует первая цифра — 0 — добавить ее и провести повторную сверку со справочником. Сделаем это.

| # | ab БИК | ab Отделение |
|----|----------|--------------|
| 1 | 04049602 | <null> |
| 2 | 41403633 | <null> |
| 3 | 41501601 | <null> |
| 4 | 04178602 | <null> |
| 5 | 42520607 | <null> |
| 6 | 04304609 | <null> |
| 7 | 04403653 | <null> |
| 8 | 04680649 | <null> |
| 9 | 47601637 | <null> |
| 10 | 49805609 | <null> |
| 11 | 04120302 | <null> |
| 12 | 41501601 | <null> |



Воспользуемся калькулятором.

Выражения *f(x)* | ^ v

| Имя | Метка |
|------------|-----------|
| ab BIC_new | БИК новый |

Поля/Переменные ▾

| Имя | Метка |
|---------|-----------|
| Поля | ^ |
| ab BIC | БИК |
| ab Name | Отделение |

Предпросмотр... | AND OR NOT XOR | = <> < > <= >= | 9.0 "" 31 ≡

```
If(Left(BIC,1)="4",Concat("0",BIC),BIC)
```

Создадим строковое поле **БИК
новый** и напишем для него
выражение, как показано на
слайде.

Список функций ▾ Категории ▾

- 9.0 Abs (Аргумент)
- 9.0 AbsErr (Аргумент1, Аргумент2)
- 31 AddDay (Дата, Количество)
- 31 AddMonth (Дата, Количество)
- 31 AddQuarter (Дата, Количество)
- 31 AddWeek (Дата, Количество)
- 31 AddYear (Дата, Количество)
- 9.0 AMGD (Стоимость, Остаточная_стоимость, Время...

| Имя | Метка |
|------------|-----------|
| ab BIC_new | БИК новый |

```
If(Left(BIC,1)="4",Concat("0",BIC),BIC)
```

Разберем созданное выражение. Буквально оно означает следующее: **Если первый левый символ в поле БИК – это 4, добавить впереди 0, иначе оставить БИК неизменным.** Для получения такого выражения мы используем функции по работе со строками:

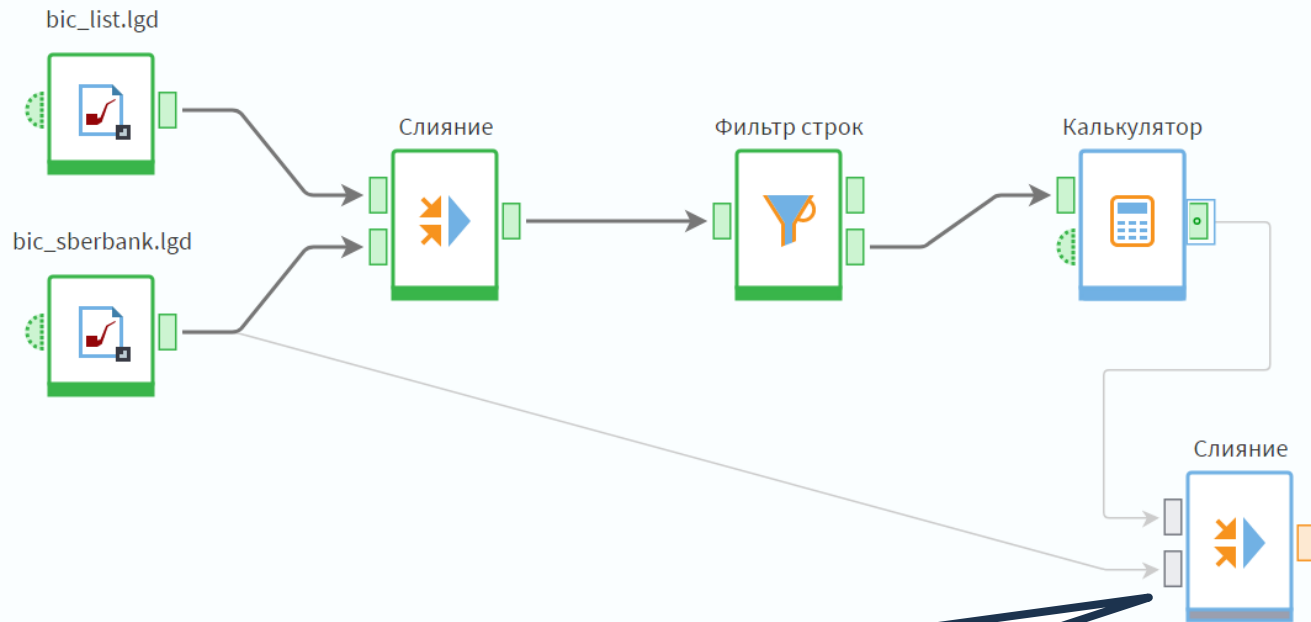
- If()** – позволяет задать условия;
- Left()** – возвращает заданное количество символов левой части строки;
- Concat()** – позволяет объединить несколько строк в одну.

| Поля/Переменные | Фильтр |
|-----------------|--------|
| Имя | |
| Поля | |
| ab BIC | |
| ab Name | |

- 31 AddQuarter (Дата, Количество)
- 31 AddWeek (Дата, Количество)
- 31 AddYear (Дата, Количество)
- 9.0 AMGD (Стоимость, Остаточная_стоимость, Время...

На выходе калькулятора получаем набор данных, который представлен на слайде. Поле **Отделение** исключили, так как оно содержит пустые записи и не несет полезной информации.

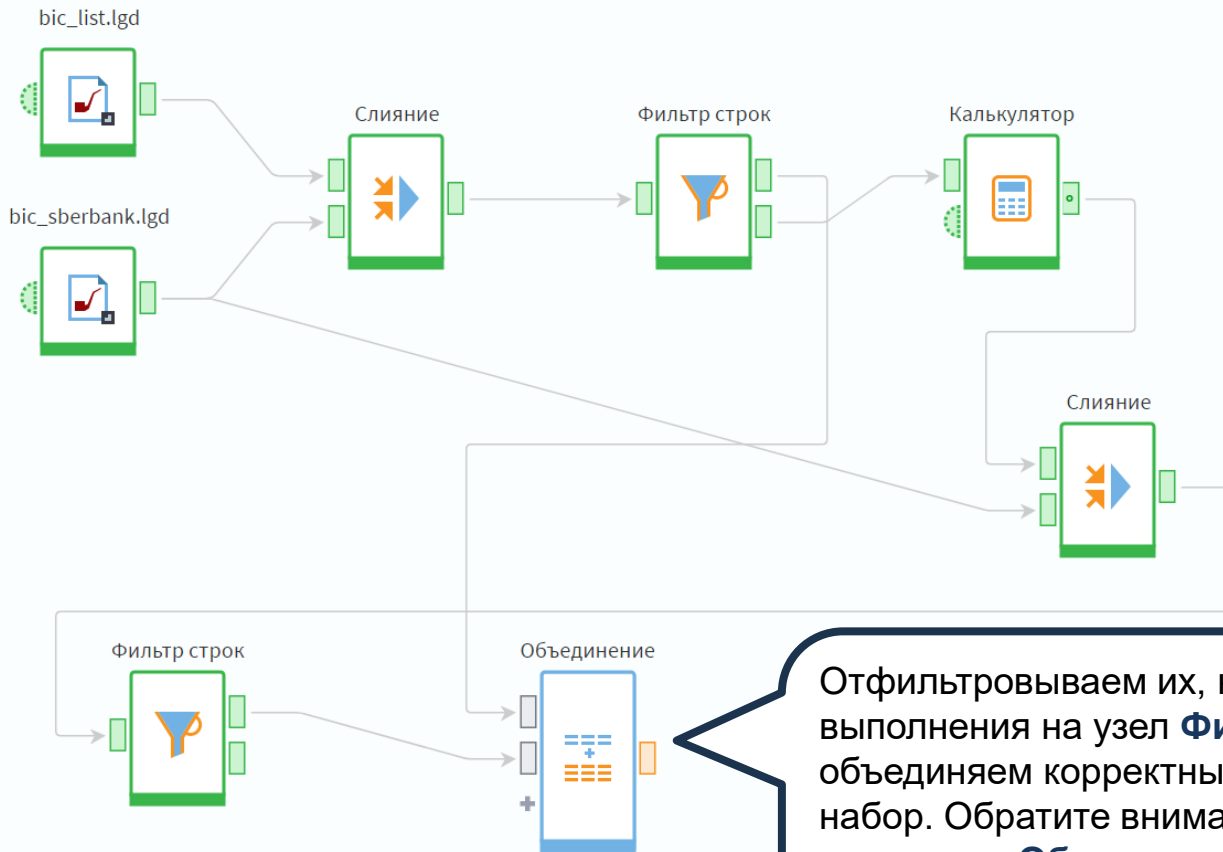
| # | ab БИК | ab БИК нов... |
|----|----------|---------------|
| 1 | 04049602 | 04049602 |
| 2 | 41403633 | 041403633 |
| 3 | 41501601 | 041501601 |
| 4 | 04178602 | 04178602 |
| 5 | 42520607 | 042520607 |
| 6 | 04304609 | 04304609 |
| 7 | 04403653 | 04403653 |
| 8 | 04680649 | 04680649 |
| 9 | 47601637 | 047601637 |
| 10 | 49805609 | 049805609 |
| 11 | 04120302 | 04120302 |
| 12 | 41501601 | 041501601 |



Снова воспользуемся слиянием со справочником, в настройках на этот раз создаем связь между полями **БИК новый** и **БИК**.

В результате нам удалось подтвердить корректность еще нескольких записей.

| # | ab БИК | ab БИК нов... | ab Отделение |
|----|----------|---------------|---|
| 1 | 04049602 | 04049602 | <null> |
| 2 | 41403633 | 041403633 | БЕЛГОРОДСКОЕ ОТДЕЛЕНИЕ N8592 ПАО СБЕРБАНК |
| 3 | 41501601 | 041501601 | БРЯНСКОЕ ОТДЕЛЕНИЕ N8605 ПАО СБЕРБАНК |
| 4 | 04178602 | 04178602 | <null> |
| 5 | 42520607 | 042520607 | БАЙКАЛЬСКИЙ БАНК ПАО СБЕРБАНК |
| 6 | 04304609 | 04304609 | <null> |
| 7 | 04403653 | 04403653 | <null> |
| 8 | 04680649 | 04680649 | <null> |
| 9 | 47601637 | 047601637 | ЧИТИНСКОЕ ОТДЕЛЕНИЕ N8600 ПАО СБЕРБАНК |
| 10 | 49805609 | 049805609 | ЯКУТСКОЕ ОТДЕЛЕНИЕ N8603 ПАО СБЕРБАНК |
| 11 | 04120302 | 04120302 | <null> |
| 12 | 41501601 | 041501601 | БРЯНСКОЕ ОТДЕЛЕНИЕ N8605 ПАО СБЕРБАНК |



Отфильтровываем их, настроив узел выполнения на узел **Фильтр строк**, и объединяем корректные записи в один набор. Обратите внимание: на втором входе узла **Объединение** стоит исключить из набора поле **БИК**, оставив только поля **БИК новый** и **Отделение**.

Получаем очищенный набор данных.

| # | ab БИК | ab Отделение |
|----|-----------|---|
| 1 | 040173604 | АЛТАЙСКОЕ ОТДЕЛЕНИЕ N8644 ПАО СБЕРБАНК |
| 2 | 040407627 | КРАСНОЯРСКОЕ ОТДЕЛЕНИЕ N 8646 ПАО СБЕРБАНК |
| 3 | 040702615 | СТАВРОПОЛЬСКОЕ ОТДЕЛЕНИЕ N5230 ПАО СБЕРБАНК |
| 4 | 040813608 | ДАЛЬНЕВОСТОЧНЫЙ БАНК ПАО СБЕРБАНК |
| 5 | 041117601 | АРХАНГЕЛЬСКОЕ ОТДЕЛЕНИЕ N 8637 ПАО СБЕРБАНК |
| 6 | 041203602 | АСТРАХАНСКОЕ ОТДЕЛЕНИЕ N8625 ПАО СБЕРБАНК |
| 7 | 041806647 | ВОЛГОГРАДСКОЕ ОТДЕЛЕНИЕ N8621 ПАО СБЕРБАНК |
| 8 | 041909644 | ВОЛОГОДСКОЕ ОТДЕЛЕНИЕ N8638 ПАО СБЕРБАНК |
| 9 | 042202603 | ВОЛГО-ВЯТСКИЙ БАНК ПАО СБЕРБАНК |
| 10 | 042406608 | ИВАНОВСКОЕ ОТДЕЛЕНИЕ N8639 ПАО СБЕРБАНК |
| 11 | 042748634 | КАЛИНИНГРАДСКОЕ ОТДЕЛЕНИЕ N8626 ПАО СБЕРБАНК |
| 12 | 042809679 | ТВЕРСКОЕ ОТДЕЛЕНИЕ N8607 ПАО СБЕРБАНК |
| 13 | 042908612 | КАЛУЖСКОЕ ОТДЕЛЕНИЕ N8608 ПАО СБЕРБАНК |
| 14 | 043207612 | КЕМЕРОВСКОЕ ОТДЕЛЕНИЕ N8615 ПАО СБЕРБАНК |
| 15 | 043469623 | КОСТРОМСКОЕ ОТДЕЛЕНИЕ N 8640 ПАО СБЕРБАНК |
| 16 | 043601607 | ПОВОЛЖСКИЙ БАНК ПАО СБЕРБАНК |
| 17 | 043735650 | КУРГАНСКОЕ ОТДЕЛЕНИЕ N8599 ПАО СБЕРБАНК |
| 18 | 043807606 | КУРСКОЕ ОТДЕЛЕНИЕ N8596 ПАО СБЕРБАНК |
| 19 | 044206604 | ЛИПЕЦКОЕ ОТДЕЛЕНИЕ N8593 ПАО СБЕРБАНК |
| 20 | 044442607 | СЕВЕРО-ВОСТОЧНОЕ ОТДЕЛЕНИЕ N8645 ПАО СБЕРБАНК |
| 48 | 044525225 | ПАО СБЕРБАНК |