

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ

федеральное государственное автономное образовательное учреждение
высшего образования

«Санкт-Петербургский государственный университет
аэрокосмического приборостроения»

Кафедра бизнес-информатики и менеджмента

ОТЧЕТ

ЗАЩИЩЕН С ОЦЕНКОЙ

ПРЕПОДАВАТЕЛЬ

Ассистент

должность, уч. степень,
звание

подпись, дата

А. Д. Державина

инициалы, фамилия

ЛАБОРАТОРНАЯ РАБОТА № 3

по курсу: Интеллектуальные информационные системы

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР.

№

8026



подпись, дата

М. В. Афанасьев

инициалы, фамилия

Санкт-Петербург

2023

Цель: получить навыки в обработке данных в Rapid Miner

В качестве набора данных были выбраны данные о мастерах спорта по шахматам, структура данных с примером представлена на рисунке 1

Row No.	grandmaster	Sex	Age	Region
1	No	Female	25	RUS
2	Yes	Female	29	USA
3	No	Male	11	USA
4	Yes	Male	12	ENG
5	No	Female	17	CHI
6	No	Male	15	CHI
7	Yes	Male	25	RUS
8	Yes	Female	18	RUS
9	No	Male	16	RUS
10	No	Male	30	RUS
11	No	Male	17	RUS
12	No	Male	17	RUS
13	No	Male	10	RUS
14	Yes	Male	27	USA

Рисунок 1 – Набор данных

Решим задачу удаления столбцов, для этого добавим на панель Design оператор Select Attributes, чтобы выбрать атрибуты, которые останутся и убрать те, что мы хотим удалить. На рисунке 2 показана схема, настройки оператора и результат. Со следующими операторами схема показываться не будет, так схемы буду аналогичны.

Select Attributes

type: include attributes

attribute filter type: one attribute

select attribute: Age

☒ Also apply to special attributes (id, label..)

ExampleSet (Retrieve Test — 2) | ExampleSet (Select Attributes)

Open in: Turbo Prep | Auto Model

Row No.	Age
1	25
2	29
3	11
4	12
5	17

Рисунок 2 – Удаление атрибутов

Затем переименуем столбцы на русский язык. Настройки оператора и результат представлен на рисунке 3.

Настройки оператора

Edit Parameter List: **rename attributes**
Use this list to define the renaming of the attributes.

old name	new name
grandmaster	grandmaster
Age	Возраст
Region	Регион
Sex	Пол

ExampleSet (Retrieve Test — 2) | ExampleSet (Rename)

Open in: Turbo Prep | Auto Model

Row No.	grandmaster	Пол	Возраст	Регион
1	No	Female	25	RUS
2	Yes	Female	29	USA
3	No	Male	11	USA
4	Yes	Male	12	ENG
5	No	Female	17	CHI

Рисунок 3 – Переименование атрибутов

Далее, дополним столбцы данных, чтобы появился новый атрибут – зарплата, а также укажем правило, по которому ячейки данного атрибута будут заполняться. Пусть если шахматист-мастер спорта имеет звание «Гранд мастер», то его заработная плата будет равна 10000 единиц, а если нет, то 3000. Реализуем это с помощью оператора **Generate attributes**. На рисунке 4 представлена настройка оператора, а на рисунке 5 результат.

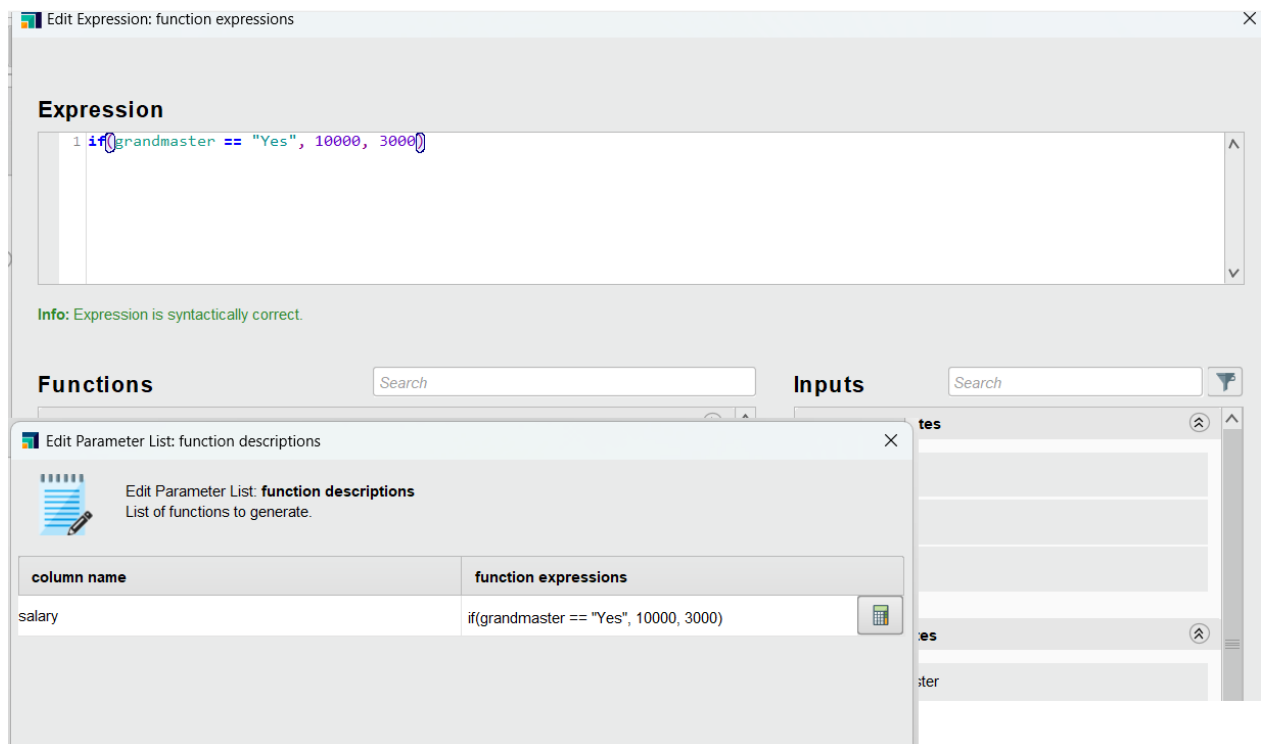


Рисунок 4 – Настройка оператора **Generate Attributes**

Row No.	grandmaster	Sex	Age	Region	salary
1	No	Female	25	RUS	3000
2	Yes	Female	29	USA	10000
3	No	Male	11	USA	3000
4	Yes	Male	12	ENG	10000
5	No	Female	17	CHI	3000
6	No	Male	15	CHI	3000
7	Yes	Male	25	RUS	10000
8	Yes	Female	18	RUS	10000
9	No	Male	16	RUS	3000
10	No	Male	30	RUS	3000
11	No	Male	17	RUS	3000
12	No	Male	17	RUS	3000
13	No	Male	10	RUS	3000
14	Yes	Male	27	USA	10000
15	Yes	Male	29	USA	10000
16	No	Male	10	USA	3000

Рисунок 5 – Дополнение данных

Затем, отсортируем данные о шахматистах: по возрасту, а внутри возраста по зарплате. (оператор, добавляющий данные о зарплате решено было оставить для большей вариативности набора). Для этого был использован оператор Sort. Настройки оператора и результат его работы представлен на рисунке 6.

Edit Parameter List: sort by

Edit Parameter List: **sort by**
This parameter defines how to sort by specifying the attributes to sort by and the associated sorting orders.

attribute name	sorting order
Age	ascending
salary	ascending

Open in

Turbo Prep

Auto Model

Row No.	grandmaster	Sex	Age	Region	salary
19	No	Female	9	USA	3000
20	No	Male	9	RUS	3000
21	No	Female	9	USA	3000
22	No	Female	9	CHI	3000
23	No	Male	9	USA	3000
24	Yes	Male	9	USA	10000
25	Yes	Female	9	ENG	10000
26	Yes	Male	9	RUS	10000
27	Yes	Male	9	RUS	10000
28	Yes	Female	9	CHI	10000

Рисунок 6 – Отсортированные данные

Далее, отфильтруем данные с помощью оператора Filter Examples, чтобы оставить в наборе мужчин гранд мастеров из России. На рисунке 7 представлены настройки оператора, а также результат.

Create Filters: filters

Create Filters: **filters**
Defines the list of filters to apply.

Sex equals Male

grandmaster equals Yes

ExampleSet (Filter Examples)

Open in Turbo Prep Auto Model

Row No.	grandmaster	Sex	Age	Region	salary
1	Yes	Male	25	RUS	10000
2	Yes	Male	16	RUS	10000
3	Yes	Male	14	RUS	10000
4	Yes	Male	17	RUS	10000
5	Yes	Male	20	RUS	10000
6	Yes	Male	19	RUS	10000
7	Yes	Male	20	RUS	10000

Рисунок 7 – Фильтрация набора

Пропущенных значений в наборе нет, так что в качестве примера воспользуемся еще одним оператором – Set Data, чтобы указать заработную плату первому шахматисту 99 000 единиц. Настройки оператора и результат представлен на рисунке 8.

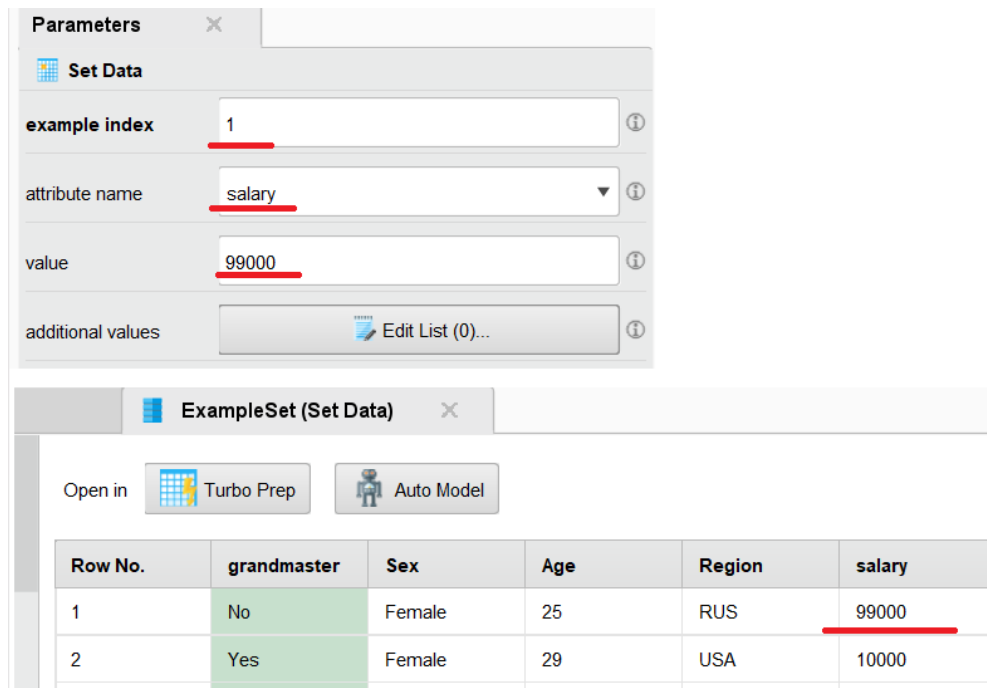
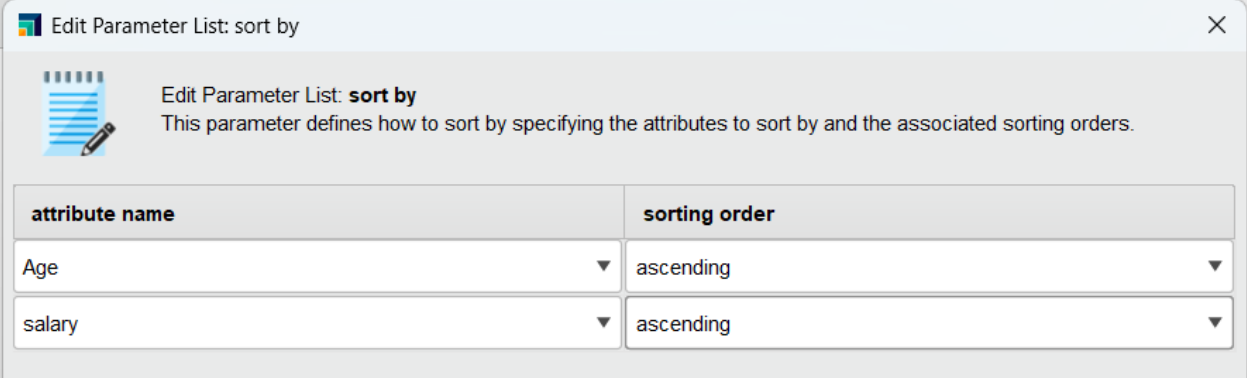


Рисунок 8 – Установка значения

Ответы на контрольные вопросы:

- 1) 5 основных разделов: Get Started, Prepare data, Build a model, Collaborate and Scale, Use Hadoop
- 2) В поиске раздела Operators, в левом нижнем углу рабочего пространства
- 3) В его описании. Правой кнопкой мышь, затем Show operator info. Или нажать на оператор и, затем, F1.
- 4) При нажатии на оператор справа появляется панель его программирования-настройки.
- 5) Графически отобразить результаты работы с данными.
- 6) Множество различных вариантов представления информации, например, около 36 графиков и множество фильтров для них или персонализация каждого графика как тебе это нужно
- 7) С помощью операторов фильтрации, таких как: Filter Examples и Filter examples Range
- 8) С помощью операторов сортировки, таких как: Sort, Sort by Pareto Rank

- 9) С помощью настройки того, какие столбы тебе нужно оставить в операторе Select Attributes
- 10) Оператор Sort в настройках имеет возможность добавить многоуровневую сортировку, как показано на рисунке 9.



Edit Parameter List: sort by

Edit Parameter List: **sort by**
This parameter defines how to sort by specifying the attributes to sort by and the associated sorting orders.

attribute name	sorting order
Age	ascending
salary	ascending

Рисунок 9 – Настройка многоуровневой сортировки

Вывод: в ходе выполнения лабораторной работы были изучены методы обработки данных в системе Rapid Miner, на примере работы с набором данных о шахматистах. Данные были отсортированы, отфильтрованы, изменена их структура и нейминг и так далее.