

Фильтрация данных

Практическое занятие

Проанализируем статистику успеваемости, и выбрать студентов, которым поставите зачет автоматом.

Содержание

Импорт данных.....	1
Фильтрация данных 1.....	1
Фильтрация данных 2.....	2

Импорт данных

В файле `students.xlsx` хранятся оценки студентов по самостоятельным работам. Считайте их виде таблицы (`table`) в переменную `G` (подсказка: `readtable()`)

```
G = readtable("1.Basics\1.2.DataAnalysis\students.xlsx");
```

Заголовки столбцов в файле подписаны по-русски, поэтому они считались некорректно. Измените в таблице `G` подписи столбцов на:

```
Name W1 W2 W3
```

```
G.Properties.VariableNames = {'Name', 'W1', 'W2', 'W3'}
```

Фильтрация данных 1

Вы решили поставить автомат тем студентам, которые получили 5 по самой сложной первой самостоятельной работе.

В переменную `i` запишите логические индексы студентов, которые за первую работу (`W1`) получили оценку 5 (подсказка: `==`)

```
i = (G.W1 == 5)
```

Найдите порядковые номера этих студентов в таблице `G` (подсказка: `find()`)

```
idx = find(i)
```

Выведите фамилии этих студентов

```
G.Name(idx)
```

Обратите внимание, что студенты, прогулявшие первую самостоятельную работу, автоматически не попали в итоговую выборку, потому что сравнение `NaN` с любым числом всегда возвращает логический `0` (`false`).

Итак, вы получили список фамилий, но по нему трудно сказать, заслужили ли эти студенты автомат в полной мере. Для этого нужно обратиться к исходной таблице, и посмотреть полную статистику, что

неудобно. Поэтому вы решили помимо фамилий вывести также полную статистику кандидатов на автомат.

Создайте таблицу G15, которая содержит полную статистику по тем студентам, что получили 5 за первую работу (подсказка: ==)

```
G(idx,G.Properties.VariableNames)
```

Фильтрация данных 2

Как видите в кандидаты на автомат попал Дубинин, который получил 2 за вторую работу, поэтому поставить ему зачет, не спрашивая, вы не можете. Нужно пересмотреть критерий автомата.

В таблице G создайте столбец Auto, в который запишите логические индексы студентов, которые по всем работам получили оценки не ниже 4 (подсказка: all(..., 2), >=)

```
G.Auto = all(G{:[ "W1", "W2", "W3" ]} >= 4, 2)
```

Создайте таблицу G45, которая содержит статистику только по студентам, которые заслужили автомат по последнему критерию (в столбце Auto стоит true)

```
G45 = G(find(G.Auto),G.Properties.VariableNames)
```