

Контрольное задание

Скачайте файл **Вероятности визитов.txt**, который имеет следующие поля:

- **Клиент.Код** – уникальный идентификатор клиента
- **Отклик** – поле принимает значения **True** или **False**
- **Вероятность визита** – поле принимает значения в диапазоне от 0 до 1
- **Признак тестового множества** – поле принимает значения **True** и **False**
- **Тип клиента** – поле принимает значения **Тип 1** и **Тип 2**

#	ab Клиент.Код	0_1 Отклик	9_0 Вероятность визита	0_1 Признак тестового множества	ab Тип клиента
1	620	false	0,00	false	Тип 2
2	1120	true	0,01	false	Тип 2
3	2771	false	0,01	false	Тип 2
4	5886	true	0,11	false	Тип 1
5	10006	false	0,04	false	Тип 2
6	11207	false	0,01	false	Тип 1
7	11932	false	0,01	false	Тип 1
8	14656	true	0,05	false	Тип 2
9	15350	true	0,08	false	Тип 2
10	17302	false	0,01	false	Тип 1
11	19496	false	0,03	false	Тип 2
12	20478	false	0,00	false	Тип 1
13	20761	false	0,00	false	Тип 2
14	24554	false	0,06	false	Тип 2
15	24880	false	0,00	false	Тип 1
16	29173	false	0,01	false	Тип 2
17	29440	true	0,05	false	Тип 1
18	30600	false	0,05	false	Тип 1
10 946	42439	true	0,36	false	Тип 1

Часть 1

1. Создайте в **Loginom** новый пакет и импортируйте описанный набор данных. Рассчитайте количество клиентов, у которых в идентификаторе **Клиент.Код** присутствует два и более нулей. Обратите внимание: для решения данной задачи не требуется создавать сложный сценарий, достаточно внимательно изучить функции узла **Калькулятор**.

#	ab Клиент.Код	0_1 Отклик	9_0 Вероятность визита	0_1 Признак тестового множества	ab Тип клиента
1	620	false	0,00	false	Тип 2
2	1120	true	0,01	false	Тип 2
3	2771	false	0,01	false	Тип 2
4	5886	true	0,11	false	Тип 1
5	10006	false	0,04	false	Тип 2
6	11207	false	0,01	false	Тип 1
7	11932	false	0,01	false	Тип 1
8	14656	true	0,05	false	Тип 2
9	15350	true	0,08	false	Тип 2
10	17302	false	0,01	false	Тип 1
11	19496	false	0,03	false	Тип 2
12	20478	false	0,00	false	Тип 1
13	20761	false	0,00	false	Тип 2
14	24554	false	0,06	false	Тип 2
15	24880	false	0,00	false	Тип 1
16	29173	false	0,01	false	Тип 2
17	29440	true	0,05	false	Тип 1
18	30600	false	0,05	false	Тип 1
10 946	42439	true	0,36	false	Тип 1

2. Подключите к вашему пакету библиотеку **Loginom Silver Kit**. Отберите всех клиентов, которые входят в тестовое множество. Далее используйте внешний компонент **AUC** из библиотеки **Loginom Silver Kit**. Вам потребуются следующие настройки в мэппинге полей:

- Отклик – Событие
- Вероятность визита – Оценка.

После настройки входного порта достаточно запустить узел на выполнение.

3. Узнайте свой вариант **n** у преподавателя.

4. Ответьте на вопросы:

- 4.1. Отсортируйте клиентов, которые входят в тестовое множество, **по убыванию** вероятности визита. Введите код клиента, который оказался **n-ым** после сортировки.
- 4.2. Сколько клиентов типа **Тип 1** присутствует среди клиентов, у которых в идентификаторе два и более нулей?
- 4.3. Отсортируйте клиентов, которые входят в тестовое множество, **по возрастанию** вероятности визита. Укажите код клиента, который оказался **n-ым** после сортировки.
- 4.4. Сколько клиентов типа **Тип 2** присутствует среди клиентов, у которых в идентификаторе два и более нулей?
- 4.5. Укажите значение переменной **Z-оценка**, полученной при использовании компонента **AUC** на клиентах из тестового множества?
- 4.6. Сколько клиентов входит в тестовое множество?
- 4.7. Сколько клиентов, у которых в идентификаторе **Клиент.Код** присутствует два и более нулей?
- 4.8. Укажите значение переменной **Индекс AUC**, полученной при использовании компонента **AUC** на клиентах из тестового множества.
- 4.9. Сколько клиентов не входит в тестовое множество?
- 4.10. Установите на выходе компонента **AUC** значение переменной **Точность** равным 3. Укажите значение переменной **Z-оценка**, полученной на выходе компонента **AUC** при его использовании на клиентах из текстового множества.

Часть 2

5. Разработайте сценарий, который рассчитывает долгосрочную стоимость клиента **CLV – Customer Lifetime Value** по следующей формуле:

$$CLV = m \left(\frac{r}{1 + i - r} \right),$$

где:

- m – ожидаемый годовой доход с одного клиента в рублях
- r – коэффициент удержания клиента, от 0 до 1
- i – годовая ставка дисконтирования в долях от 0 до 1

6. Рассчитайте для каждого клиента из файла Вероятности визитов.txt значение показателя CLV, если ставка дисконтирования равна 12%, а коэффициент удержания и ожидаемый годовой доход зависят от типа клиента:
- $r = 0,9$ и $m = 7200$ – для клиентов с типом 1
 - $r = 0,7$ и $m = 9500$ – для клиентов с типом 2

7. Отсортируйте по убыванию CLV.

8. **Ответьте на следующие вопросы:**

- 8.1. Введите значение CLV для 10-ого клиента из списка, упорядоченного по убыванию CLV. Укажите целую часть числа.
- 8.2. Введите значение вероятности визита для 3-го клиента из списка, упорядоченного по убыванию CLV. Ответ округлить до 2-ого знака после запятой.
9. Проделайте то же самое, рассчитав CLV по другой формуле, с учетом ежегодного роста прибыли на $g\%$ (при расчетах возьмите $g=8\%$):

$$CLV = m \left(\frac{r}{1 + i - r(1 + g)} \right)$$

10. **Ответьте на следующие вопросы:**

- 10.1. Введите количество клиентов, для которых целая часть значения CLV равна 18269.
- 10.2. Введите значение CLV для 100-го клиента из списка, упорядоченного по убыванию CLV. В ответе укажите целую часть числа.