Лабораторная работа 2. Кластерный анализ в Orange.

Кластерный анализ– задача разбиения заданной выборки объектов на непересекающиеся подмножества, называемые кластерами так, чтобы каждый кластер состоял из схожих объектов, а объекты разных кластеров существенно отличались.

1.Цель исследования. Это может быть или определение кластерной структуры данных (проверка наличия кластеров в данных) или разбиение объектов на группы по заранее определенному принципу.

2. Преобразование имеющихся данных в объекты для их кластеризации, например, определение набора слов, которые характеризуют тексты, определение переменных, характеризующих объекты.

3.Преобразование выбранных объектов в значимый вид, т.е. сопоставление значениям переменных числовые значения.

4.Задание метрики пространства. (Формула расчета расстояния между кластерами)

5.Выбор алгоритма кластеризации.

6.Расчет и анализ результатов.

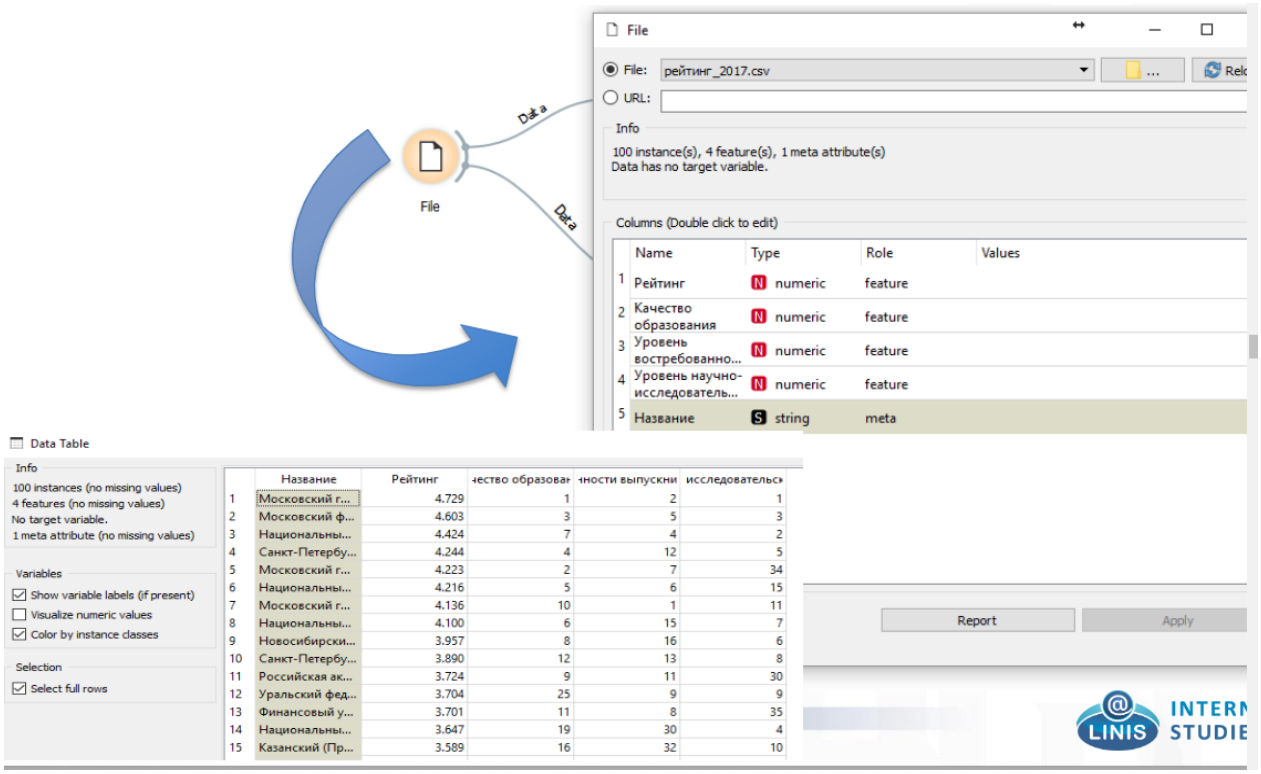
Для определения сходства используем расстояния между объектами (см. конспект).

Задание: провести кластерный анализ вузов РФ.

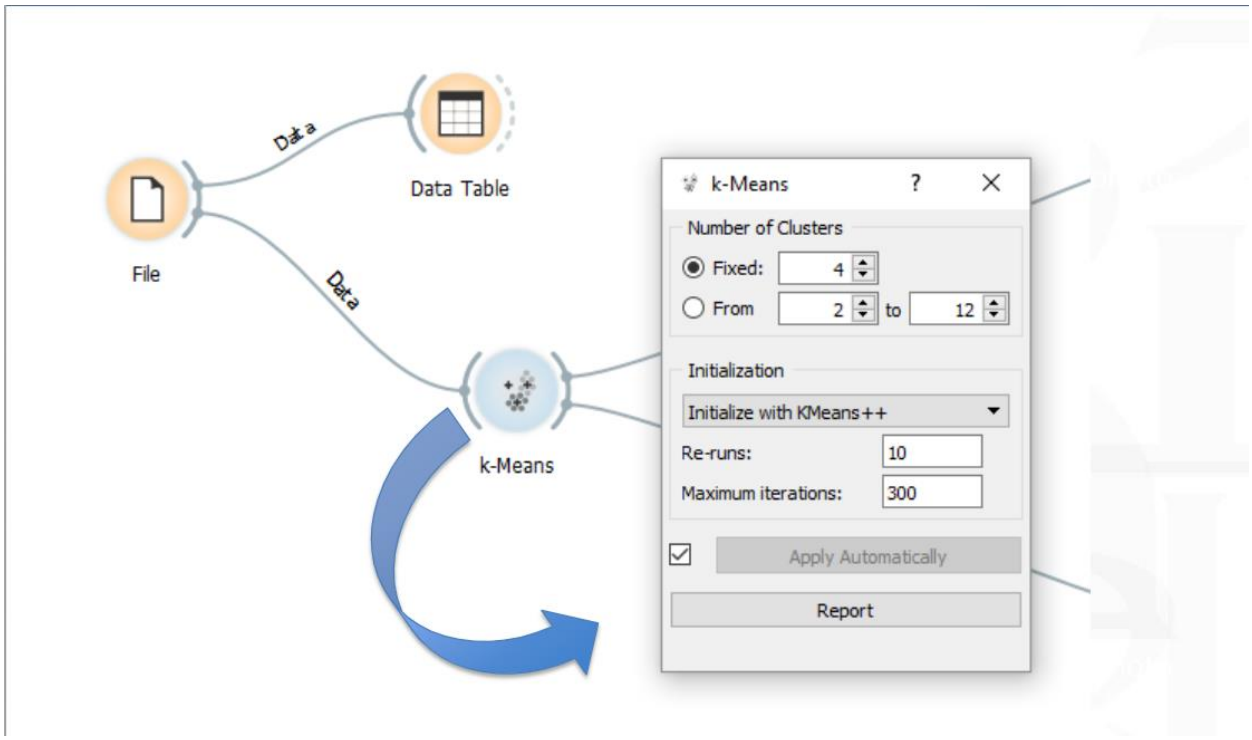
Исходные данные: https://raexpert.ru/rankingtable/university/2017/main

A	B	C	D	E
Название	Рейтинг	Качество образования	Уровень востребованности выпускников р	Уровень научно-исследовательской деятельности
Московский государственный университет им. М.В. Ломоносова	4.729	1	2	1
Московский физико-технический институт (государственный университет)	4.603	3	5	3
Национальный исследовательский ядерный университет «МИФИ»	4.424	7	4	2
Санкт-Петербургский государственный университет	4.244	4	12	5
Московский государственный институт международных отношений (университет) МИД РФ	4.223	2	7	34
Национальный исследовательский университет "Высшая школа экономики"	4.216	5	6	15
Московский государственный технический университет имени Н.Э. Баумана	4.136	10	1	11
Национальный исследовательский Томский политехнический университет	4.1	6	15	7
Новосибирский национальный исследовательский государственный университет	3.957	8	16	6
Санкт-Петербургский политехнический университет Петра Великого	3.89	12	13	8
Российская академия народного хозяйства и государственной службы при Президенте РФ	3.724	9	11	30
Уральский федеральный университет имени первого Президента России Б.Н. Ельцина	3.704	25	9	9
Финансовый университет при Правительстве РФ	3.701	11	8	35
Национальный исследовательский Томский государственный университет	3.647	19	30	4
Казанский (Приволжский) федеральный университет	3.589	16	32	10
Сибирский федеральный университет	3.565	26	10	14
Национальный исследовательский технологический университет «МИСиС»	3.56	13	31	16
Российский государственный университет нефти и газа (национальный исследовательский университет	3.553	24	3	37
Университет ИТМО	3.461	15	47	12
Российский университет дружбы народов	3.412	14	29	23
Первый Московский государственный медицинский университет имени И.М. Сеченова Министерства з	3.338	21	20	21
Первый Санкт-Петербургский государственный медицинский университет имени академика И.П. Павло	3.32	18	25	33
Российский экономический университет имени Г.В. Плеханова	3.253	23	14	41
Российский национальный исследовательский медицинский университет имени Н.И. Пирогова Минист	3.175	22	23	31
Национальный исследовательский университет "МЭИ"	3.088	28	26	27
Новосибирский государственный технический университет	3.06	32	21	19
Московский государственный лингвистический университет	3.018	17	52	75
Национальный исследовательский Нижегородский государственный университет имени Н.И. Лобачевс	2.957	29	58	18

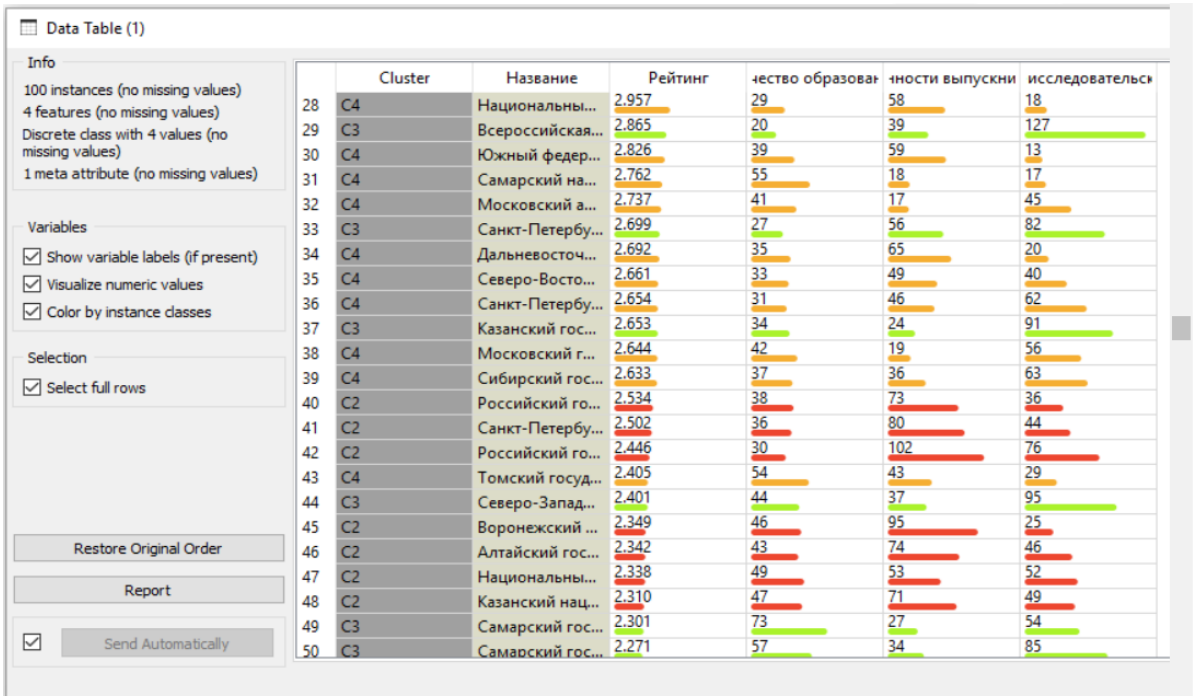
Загрузить данные *.csv в Orange (источник данных - файл):



Настроить число кластеров:

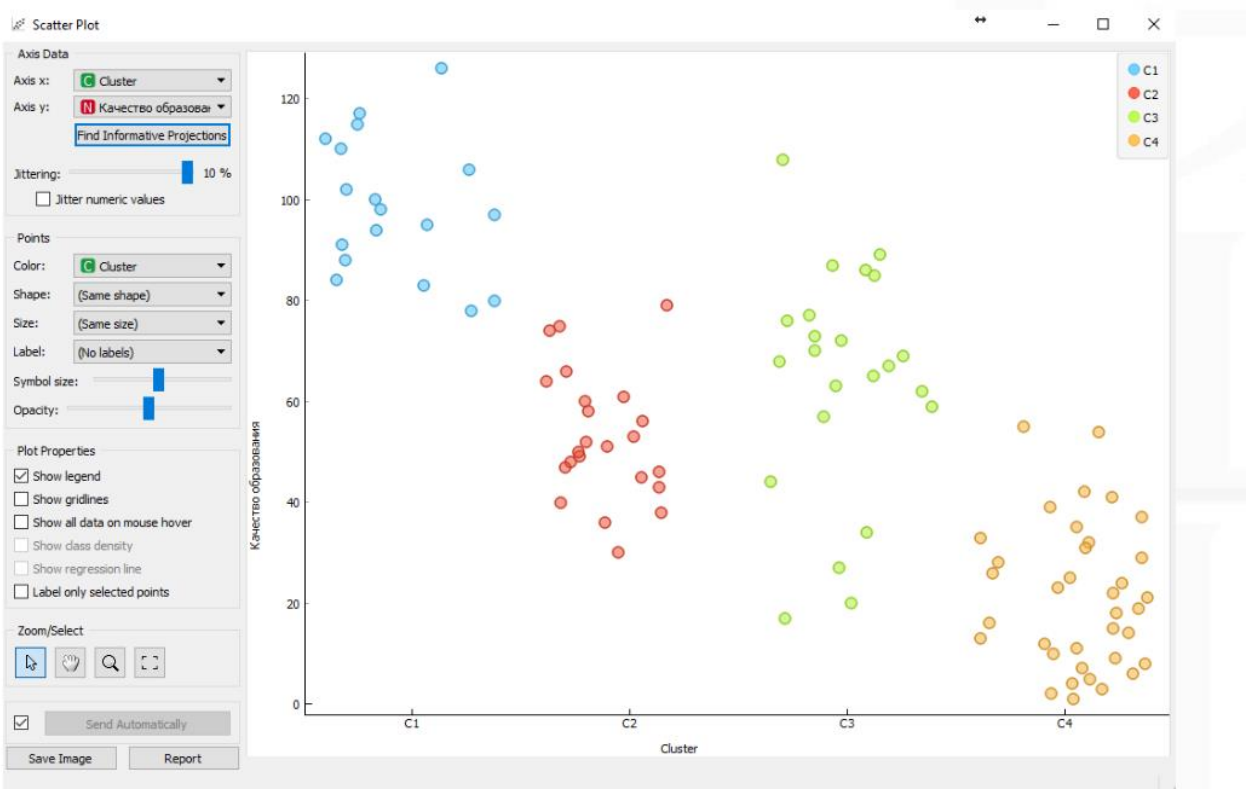


Результат кластеризации в виде Data Table:



	Cluster	Название	Рейтинг	чество образовани	чности выпускни	исследовательск
28	C4	Национальны...	2.957	29	58	18
29	C3	Всероссийская...	2.865	20	39	127
30	C4	Южный федер...	2.826	39	59	13
31	C4	Самарский на...	2.762	55	18	17
32	C4	Московский а...	2.737	41	17	45
33	C3	Санкт-Петербу...	2.699	27	56	82
34	C4	Дальневосточ...	2.692	35	65	20
35	C4	Северо-Восто...	2.661	33	49	40
36	C4	Санкт-Петербу...	2.654	31	46	62
37	C3	Казанский гос...	2.653	34	24	91
38	C4	Московский г...	2.644	42	19	56
39	C4	Сибирский гос...	2.633	37	36	63
40	C2	Российский го...	2.534	38	73	36
41	C2	Санкт-Петербу...	2.502	36	80	44
42	C2	Российский го...	2.446	30	102	76
43	C4	Томский госуд...	2.405	54	43	29
44	C3	Северо-Запад...	2.401	44	37	95
45	C2	Воронежский ...	2.349	46	95	25
46	C2	Алтайский гос...	2.342	43	74	46
47	C2	Национальны...	2.338	49	53	52
48	C2	Казанский нац...	2.310	47	71	49
49	C3	Самарский гос...	2.301	73	27	54
50	C3	Самарский гос...	2.271	57	34	85

Распределение университетов по качеству образования:



Также представить распределение университетов по востребованности среди работодателей. Отобразить университеты первой группы С4 в виде таблицы.

Полученные результаты представить в виде текстового отчета со скриншотами и исходного файла Orange.