



Результаты кластерного анализа школ Невского района

Александр Пронин
ГБУ ДПО «СПБЦОКОиИТ»
socpron@gmail.com

Ключевые идеи исследования

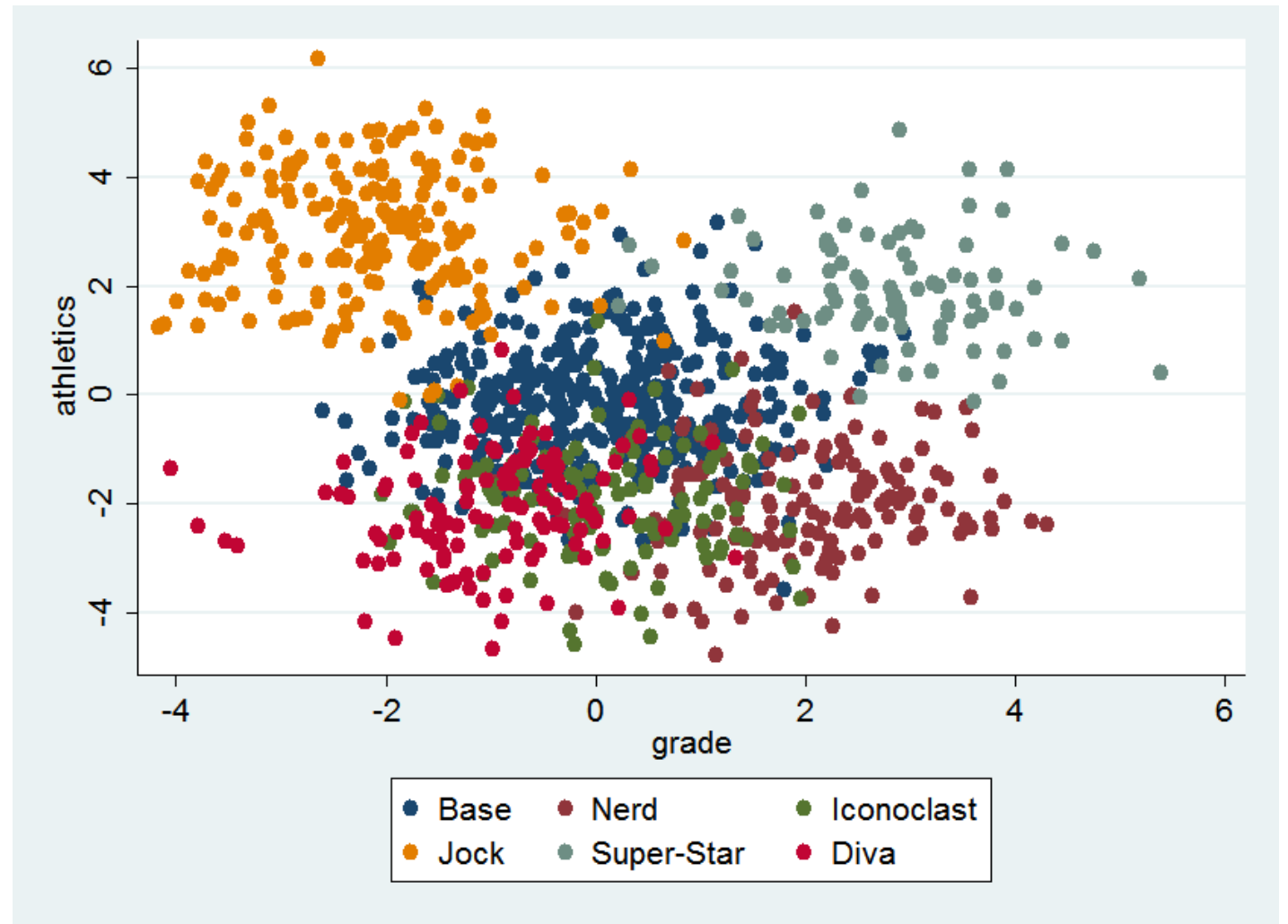
Выявить реальную структуру образовательной среды Невского района: количество кластеров и школ внутри них

Определить основные характеристики каждого кластера и направления адресной помощи

Описать полученные кластеры школ в контексте географии и инфраструктуры Невского района

Пример использования кластерного анализа

Выделение 6 групп
студентов на основе двух
переменных – успеваемости
(grade) и физических
способностей (athletics)



Алгоритм исследования

Для 53 школ Невского района, вошедших в рейтинг 2020, формируется массив данных. Используются рейтинговые баллы и контекстные данные о школе: характеристики контингента, процессы мобильности, география учеников и другие группы переменных.

В статистическом пакете IBM SPSS Statistics 20 проводится процедура кластеризации методом К-средних на основе сформированного массива данных. Каждой школе присваивается номер кластера, в который она попала.

Для каждого кластера рассчитываются средние значения всех переменных, которые попали в исходный массив данных и на основе которых проводился кластерный анализ. Выделяются различия, на основе которых определяются ключевые характеристики кластеров.

Для детального описания кластеров строятся диаграммы рассеивания, на которых с помощью цветовой индикации отражаются инфраструктурные и географические параметры района.

Данные для кластеризации

Рейтинги школ 2020 года

- Сумма баллов школы в рейтинге по результатам массового образования
- Сумма баллов школы в рейтинге по высоким образовательным результатам и достижениям обучающихся
- Сумма баллов школы в рейтинге по качеству условий ведения образовательной деятельности
- Сумма баллов школы в рейтинге по кадровому обеспечению
- Сумма баллов школы в рейтинге по эффективности управления
- Количество попаданий школы в топ-100 по каждому из пяти рейтингов в 2020 году

Данные для кластеризации

Социальный паспорт школы

- Процент обучающихся-инвалидов
- Процент обучающихся на дому (по медицинским показаниям)
- Процент обучающихся, оставленных на повторный курс обучения
- Процент обучающихся, не имеющие гражданства РФ
- Процент обучающихся, не имеющие постоянной регистрации в Санкт-Петербурге
- Процент обучающихся, которые занимаются в кружках/секциях школы
- Процент обучающихся, которые занимаются в кружках/секциях вне школы
- Процент обучающихся, имеющих проблемы в обучении (неуспевающие)
- Процент обучающихся из семей, где родители ненадлежащим образом исполняют свои обязанности
- Процент обучающихся из семей, состоящие на контроле в отделе опеки и попечительства
- Процент обучающихся из семей, имеющие детей, находящихся в социально опасном положении, (безнадзорность, беспризорность)
- Процент обучающихся, родители которых инвалиды (по информации/документам родителей)
- Процент обучающихся из многодетных семей (по документам)
- Процент опекаемых обучающихся (по документам)
- Процент детей из неполных семей вследствие потери кормильца (по документам)
- Процент обучающихся из семей мигрантов (по документам)
- Процент обучающихся из неблагополучных семей
- Процент обучающихся, грубо или неоднократно нарушившие Устав ОУ и Правила поведения

Данные для кластеризации

Мобильность учеников

- Процент учеников, проживающих в другом районе Санкт-Петербурга
- Процент учеников, проживающих в Ленинградской области
- Процент учеников, покинувших школу после первого года обучения
- Процент учеников, поступивших в школу после первого года обучения
- Процент учеников, покинувших школу после четвертого класса
- Процент учеников, поступивших в школу после четвертого класса

Данные для интерпретации кластеров

Контекстные данные

- Плотность населения МО
- Сведения о застройке прилегающей к ОО территории
- Территориальная близость ОО к общественно-деловым зонам
- Территориальная близость ОО к производственным зонам и инженерной инфраструктуре

Результаты кластерного анализа школ на основе исходного массива данных



Кластеризация школ проводилась с помощью метода К-средних



Изначально не задавалось требование к количеству формируемых кластеров, выбор сделан в пользу 3-факторной модели



С помощью критерия Краскела — Уоллиса были определены переменные, по которым между полученными кластерами есть значимые различия. Эти переменные будут использованы как ключевые характеристики для описания состава кластеров



На следующих слайдах представлены данные по составу кластеров, а также их характеристикам

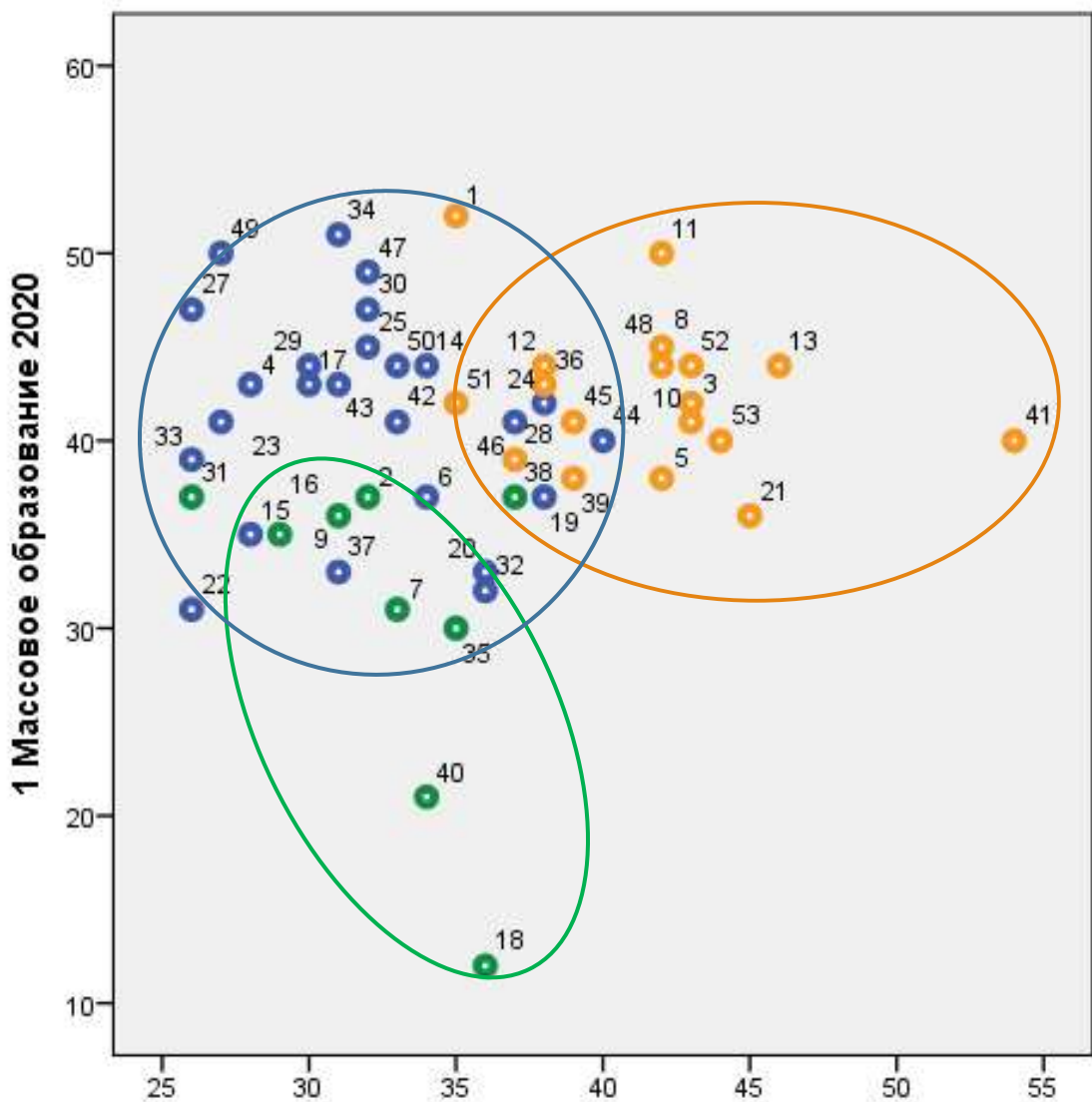
Ключевые характеристики кластеров

Переменные, различия в которых по кластерам статистически значимы	Среднее значение переменных по кластерам		
	1 кластер	2 кластер	3 кластер
Рейтинг 1 Массовое образование 2020	41	31	42
Рейтинг 2 Высокие результаты 2020	12	4	18
Рейтинг 3 Условия 2020	32	33	42
Рейтинг 4 Кадровое обеспечение 2020	25	20	29
Рейтинг 5 Эффективность управления 2020	46	32	44
Попадание в топ 100 2020	0	0	0,8
Занимаются в кружках/секциях вне ОУ	,403	,253	,445
Обучающиеся из семей мигрантов (по документам)	,018	,050	,012
Обучающиеся из неблагополучных семей	,000	,002	,001
Обучающиеся-инвалиды	,009	,005	,008
Обучающиеся, оставленные на повторный курс обучения	,001	,004	,002
Обучающиеся, не имеющие гражданства РФ	,021	,053	,015

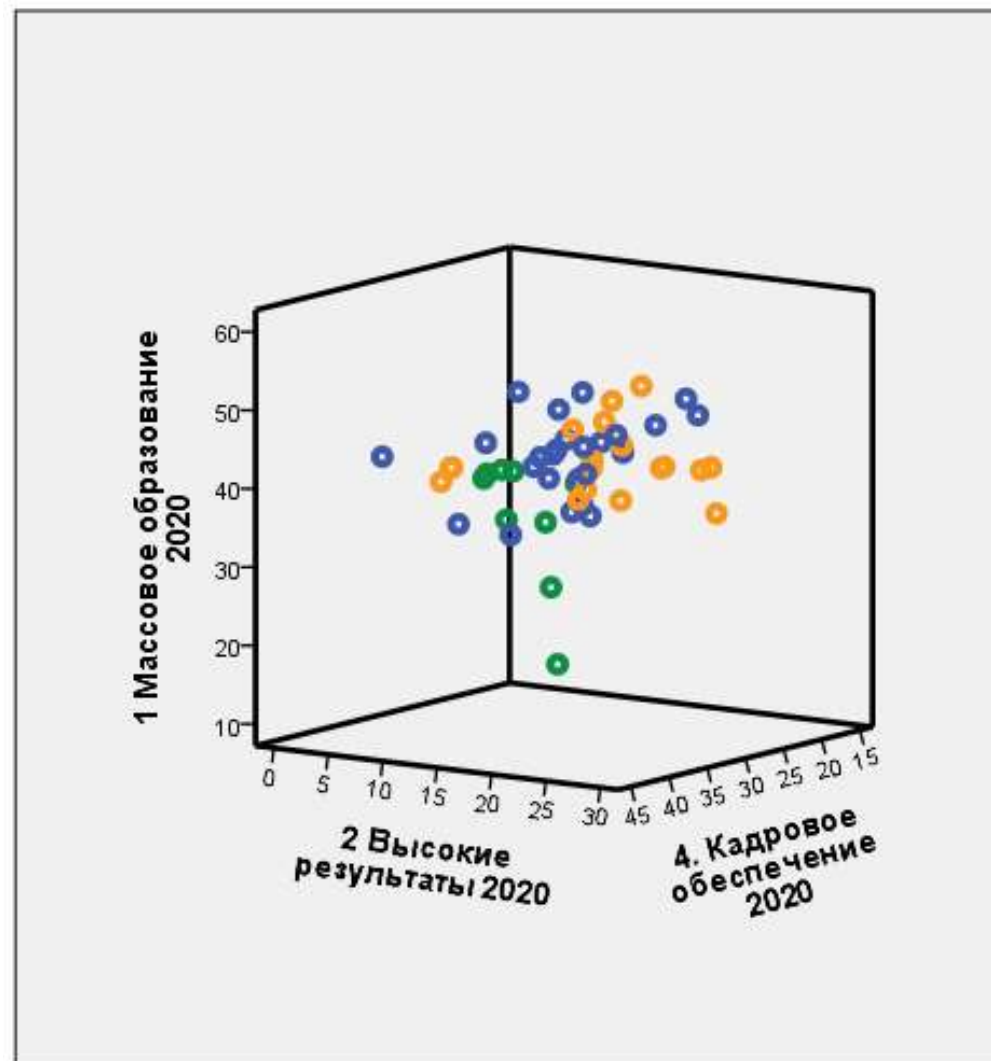
Переменные миграции учеников	Среднее значение переменных по кластерам		
	1 кластер	2 кластер	3 кластер
1 -2 классы убыло	0,021	0,073	0,034
1 -2 классы прибыло	0,055	0,008	0,018
4-5 классы убыло	0,034	0,064	0,051
4-5 классы прибыло	0,044	0,020	0,030

- Кластер 1 (48% школ района): развитие программ сетевого партнерства и наставничества, обмен успешными практиками с образовательными учреждениями из кластера 3, мониторинг мобильности учеников.
- Кластер 2 (17% школ района): комплексная помощь со стороны ИМЦ, повышение квалификации учителей, контроль объективности и низких результатов, мониторинг мобильности учеников.
- Кластер 3 (35 % школ района): мониторинг ключевых показателей.

Визуализация кластерной структуры



3. Условия 2020



Номера школ анонимизированы

Ключевые комментарии

При интерпретации кластеров требуется сопоставление с результатами других процедур анализа.

Результаты кластерного анализа позволяют определять актуальные направления для управленческих решений и адресной поддержки школ.

Следует помнить о главном принципе кластерного анализа: мы ищем неизвестную структуру (группы / кластеры) на основе известных данных.

Спасибо за ваше внимание!

И спасибо моим коллегам за совместные проекты
по оценке качества образования!

<https://monitoring.spbcokoit.ru/>