

## ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ  
ИССЛЕДОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО  
РАЗВИТИЯ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ  
по программе профессиональной переподготовки:  
«Анализ данных на языке Python»

Выполнила: Карапетова Елена Александровна  
Руководитель: к.т.н. Семендяев Родион Юрьевич

Санкт-Петербург, 2023



# Актуальность исследования

**Проблема:** неравномерность регионального развития РФ.

**Последствия:** региональные диспропорции.

Средний размер пенсий (данные 2021 года)

Самые высокие		Самые низкие	
Чукотский АО	26381	Дагестан респ.	11923
Ненецкий АО	23854	Кабардино-Балкарская респ.	12216
Камчатский АО	23317	Ингушетия респ.	13282

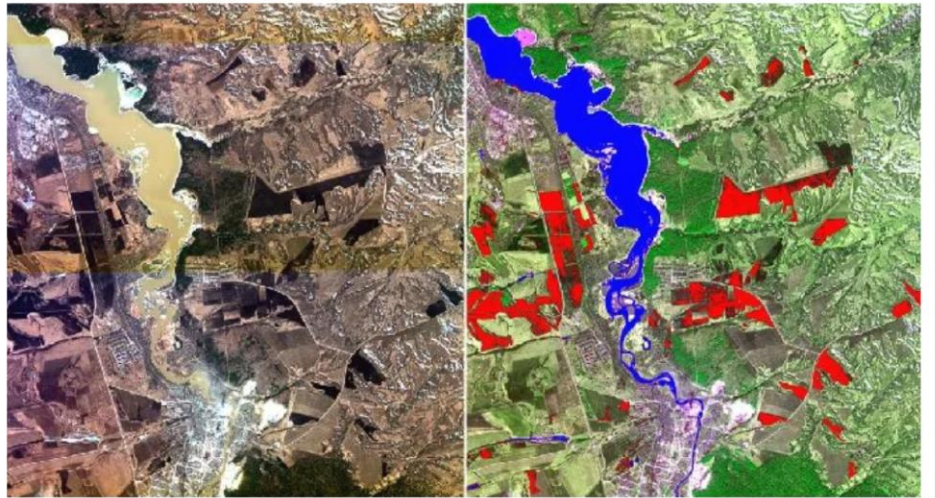


## Цель работы

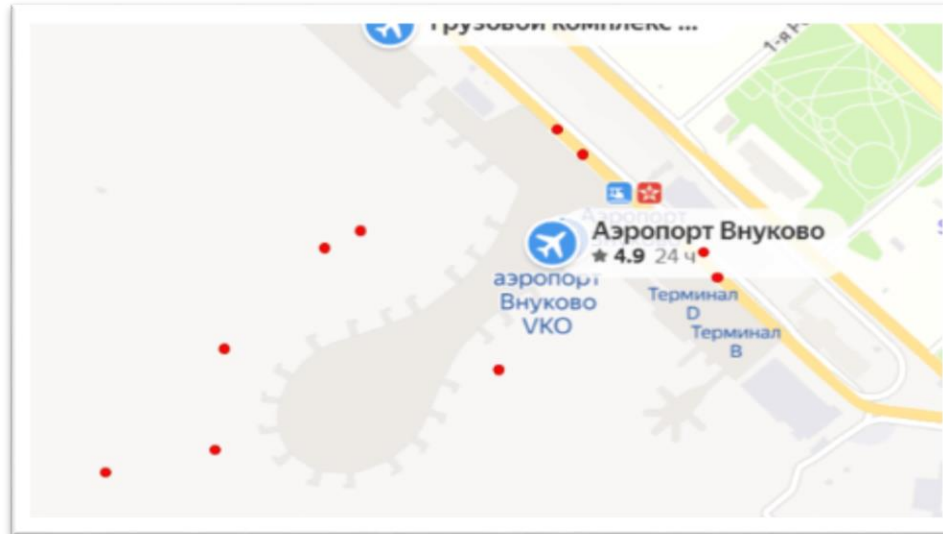
Цель данного исследования заключается в выявлении групп регионов со схожими социально-экономическими характеристиками и изучении изменений, происходящих в этих группах за последние годы.

- выбрать и собрать основные показатели для сравнения социально-экономического развития регионов РФ;
- проверить качество набора исходных данных, провести предобработку данных;
- выделить группы регионов со схожими социально-экономическими показателями путем применения методов машинного обучения;
- провести анализ распределения регионов по отдельным кластерам.

# Область применения кластерного анализа



Сегментация изображений



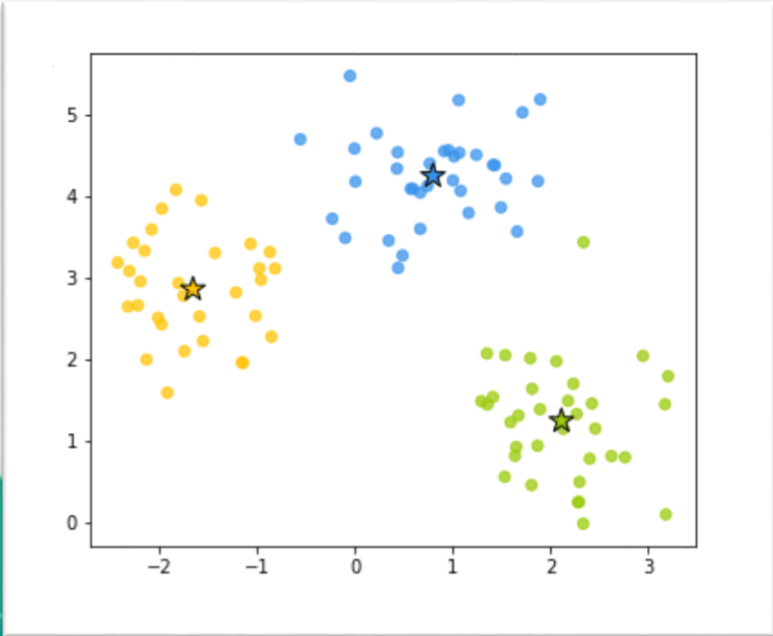
Анализ геоданных  
(используя GPS- координаты)

- биология
- психология
- медицина
- химия
- экономика
- экономическая география

# Методы машинного обучения для задачи кластеризации

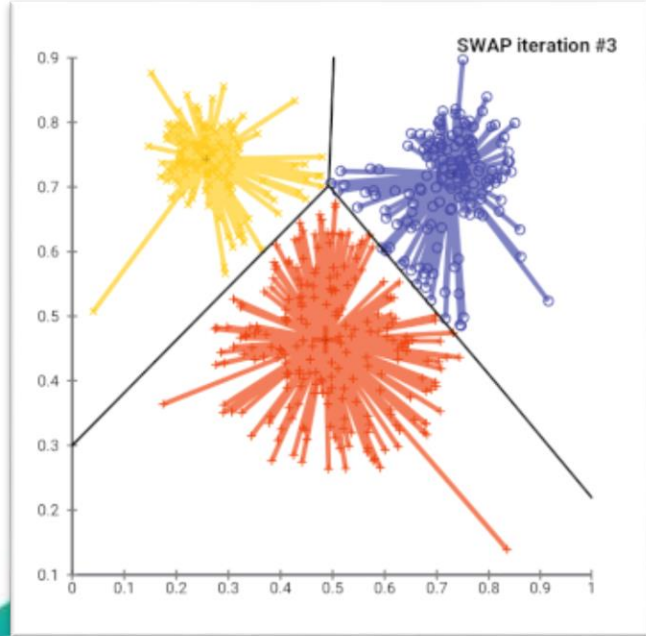
Метод  $k$ -средних

Результат кластеризации  
методом  $k$ -средних



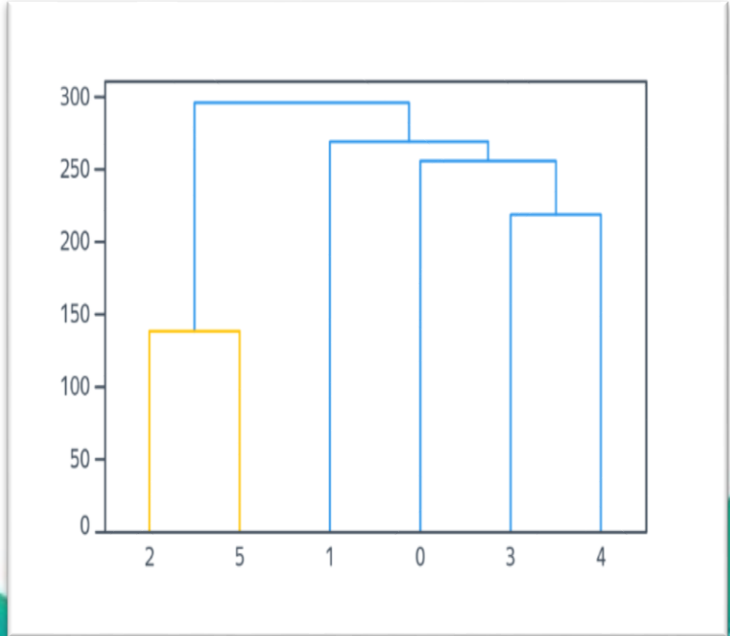
Метод  $k$ -medoids

Результат кластеризации  
методом  $k$ -medoids



Агломеративная  
иерархическая  
кластеризация

Результат кластеризации –  
дендрограмма



# Источники данных



Федеральная служба государственной статистики



Генеральная прокуратура РФ

Федеральная служба государственной статистики

О Росстате    **Статистика**    Публикации

Официальная статистика    Переписи и обследования

Методология и нормативно-справочная информация    Понятная статистика

**Новости Росстата**

## Региональная статистика

- WEB    СОЦИАЛЬНО-ЭКОНОМИЧЕСКОЕ ПОЛОЖЕНИЕ субъектов Российской Федерации
- WEB    ПОКАЗАТЕЛИ ДЛЯ МОНИТОРИНГА ОЦЕНКИ ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ с/х предприятий и организаций в субъектах Российской Федерации

...е к ним местности

ГЕНЕРАЛЬНАЯ ПРОКУРАТУРА РОССИЙСКОЙ ФЕДЕРАЦИИ

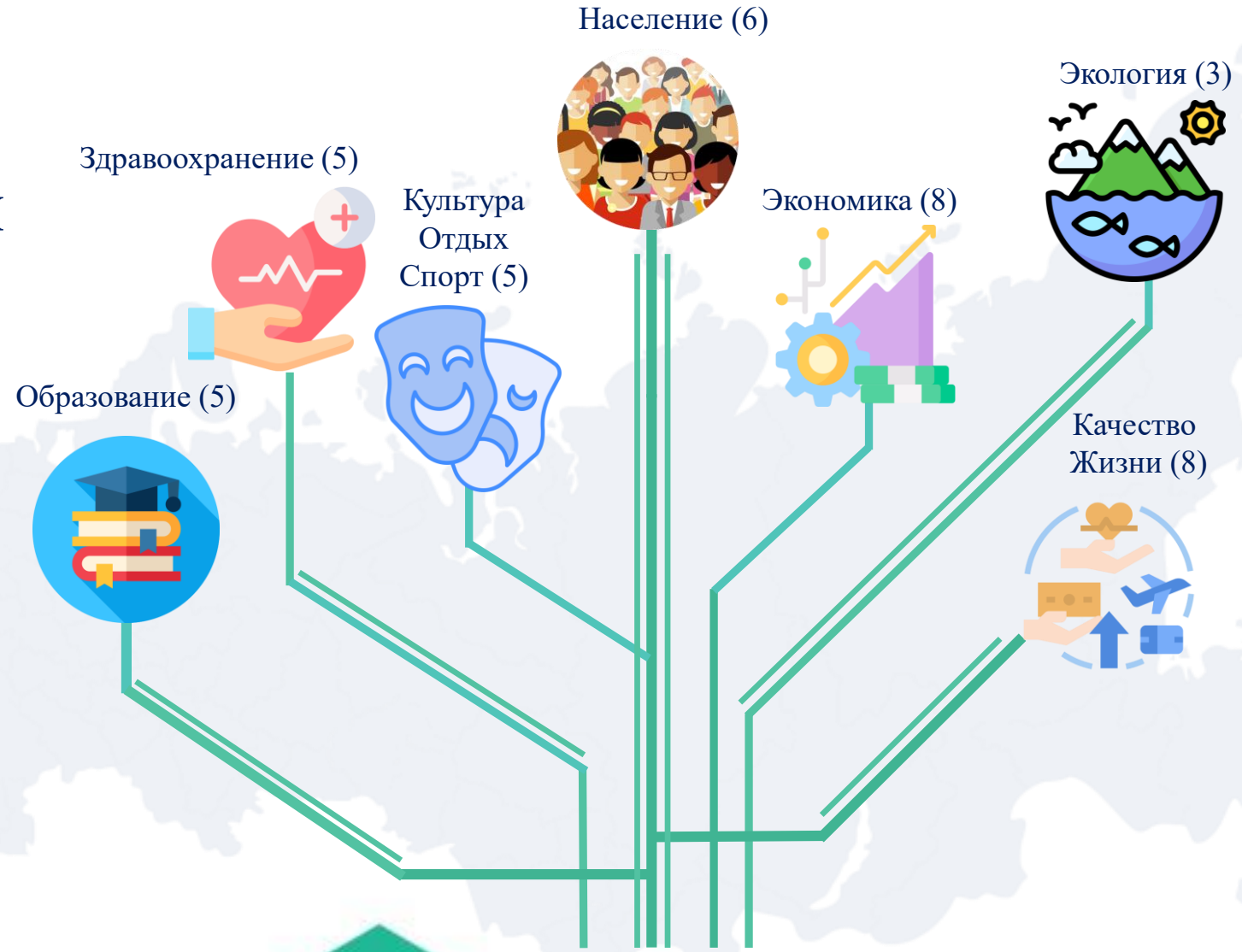
### ПОРТАЛ ПРАВОВОЙ СТАТИСТИКИ

В связи с доработкой функциональности портала правовой статистики размещение статистических данных по субъектам Российской Федерации

Показатели преступности России    Преступность в регионах    Социальный портрет преступности

Показатели преступности России

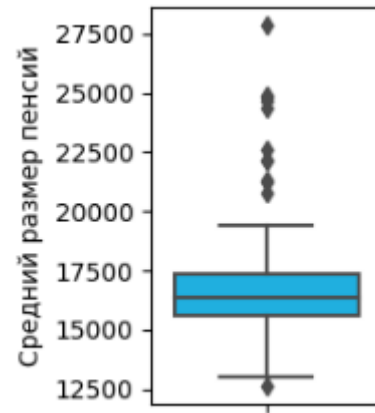
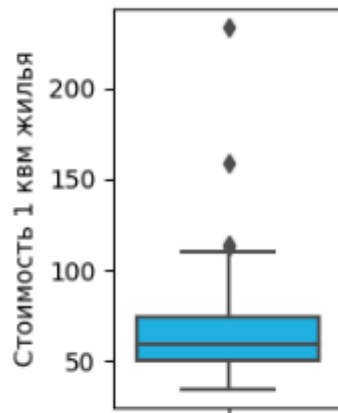
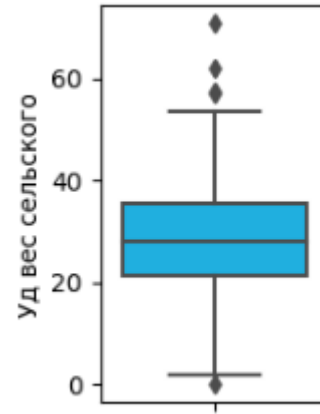
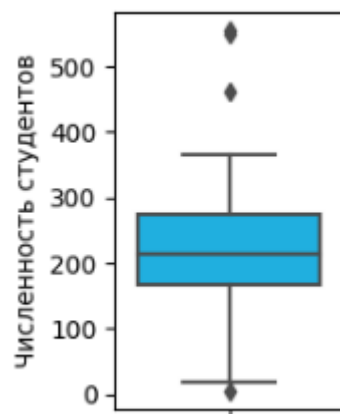
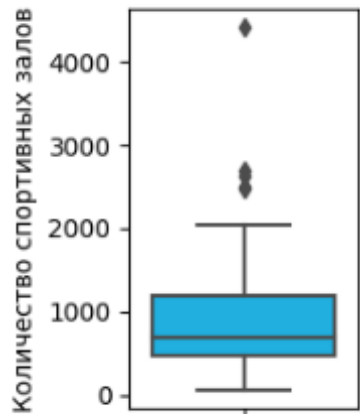
# Группы показателей развития регионов



Площадь территории	1.0	-0.1	0.3	0.2	0.0	-0.1	0.1	0.2	-0.2	0.1	0.5	-0.1	-0.2	-0.1	-0.2	0.2	-0.0	-0.2	-0.1	-0.1	0.2	0.0	-0.1	-0.1	-0.0	-0.1	0.0	0.1	-0.3	0.5	-0.1	0.1	0.0	-0.0	-0.1	0.5	0.4	0.0		
Численность населения	-0.1	1.0	0.3	0.6	0.9	-0.3	0.3	0.0	-0.0	-0.4	-0.1	-0.2	0.3	0.1	0.0	0.2	1.0	-0.1	-0.1	0.1	0.9	0.7	0.2	0.8	0.2	0.1	0.3	0.4	1.0	0.1	-0.2	-0.2	0.1	-0.2	0.9	0.9	0.3	0.2	0.6	0.5
Доходы	0.3	0.3	1.0	0.8	0.5	-0.5	0.5	0.0	-0.3	-0.6	0.8	-0.4	-0.1	0.0	-0.1	0.6	0.3	-0.2	-0.2	-0.1	0.3	0.6	0.0	0.4	0.5	0.1	0.1	-0.2	0.2	0.2	-0.3	0.6	-0.2	0.1	0.2	0.3	0.2	0.2	0.5	0.1
Расходы	-0.2	0.6	0.8	1.0	0.7	-0.6	0.6	-0.1	-0.1	-0.7	0.6	-0.5	-0.0	0.3	0.1	0.5	0.5	-0.0	-0.1	0.1	0.6	0.8	0.1	0.6	0.3	0.2	0.2	0.1	0.5	0.3	-0.4	0.3	-0.0	-0.1	0.5	0.6	0.3	0.1	0.6	0.3
ВРП	0.0	0.9	0.5	0.7	1.0	-0.4	0.4	0.0	-0.1	-0.4	0.1	-0.2	0.3	0.0	-0.1	0.4	0.8	-0.1	-0.1	0.0	0.7	0.8	0.1	1.0	0.3	0.1	0.3	0.3	0.7	0.1	-0.3	0.0	0.0	-0.1	0.8	0.8	0.5	0.2	0.6	0.3
Уд вес сельского	-0.1	-0.3	-0.5	-0.6	-0.4	1.0	-1.0	0.4	0.4	-0.3	0.6	-0.6	0.2	-0.4	-0.3	-0.4	-0.3	0.0	-0.3	-0.4	-0.2	-0.5	-0.3	0.4	0.0	0.4	-0.4	-0.1	-0.2	-0.1	0.5	-0.3	-0.3	-0.0	-0.3	-0.4	-0.4	-0.3	-0.4	-0.1
Уд вес городского	-0.1	0.3	0.5	0.6	0.4	-1.0	1.0	-0.4	0.3	-0.6	0.6	-0.6	-0.2	0.3	0.3	0.4	0.3	0.0	0.3	0.4	0.2	0.5	0.3	0.4	-0.0	0.4	0.4	0.1	0.2	0.1	-0.5	0.3	0.3	0.0	0.3	0.4	0.3	0.3	0.4	0.1
Козф рождаемости	-0.2	0.0	0.0	-0.1	0.0	0.4	-0.4	1.0	-0.8	0.5	-0.0	0.7	0.4	-0.5	-0.7	0.1	-0.1	-0.1	-0.7	-0.7	-0.0	0.0	-0.2	0.0	0.4	-0.2	0.1	-0.1	0.1	-0.1	0.2	-0.5	-0.2	-0.1	-0.1	-0.3	0.0	0.1	0.0	
Козф смертности	-0.2	-0.0	-0.3	-0.1	-0.1	-0.3	0.3	-0.8	1.0	-0.3	-0.2	-0.6	-0.6	0.4	0.8	-0.2	0.1	0.1	0.9	0.8	0.0	-0.1	0.2	-0.1	-0.5	0.2	-0.0	0.2	-0.1	-0.1	0.1	-0.4	0.6	0.1	0.1	0.0	0.3	-0.1	-0.2	-0.1
Уровень бедности	-0.1	-0.4	-0.6	-0.7	-0.4	0.6	-0.6	0.5	-0.3	1.0	-0.4	0.7	0.1	-0.4	-0.5	-0.3	-0.4	-0.0	-0.2	-0.3	-0.4	-0.4	-0.3	-0.3	-0.1	-0.3	-0.1	-0.1	-0.3	-0.3	0.2	-0.2	-0.3	0.0	-0.4	-0.5	-0.3	-0.1	-0.4	-0.3
Средний размер пенсий	0.5	-0.1	0.8	0.6	0.1	-0.6	0.6	-0.0	-0.2	-0.4	1.0	-0.4	-0.3	0.1	-0.0	0.5	-0.0	-0.2	-0.1	-0.0	-0.0	0.4	0.1	0.1	0.3	0.0	0.0	-0.3	-0.1	0.1	-0.4	0.7	-0.1	0.2	-0.1	0.1	0.1	0.2	0.3	0.0
Уровень безработицы	-0.1	-0.2	-0.4	-0.5	-0.2	0.6	-0.6	0.7	-0.6	0.7	-0.4	1.0	0.5	-0.4	-0.6	-0.2	-0.3	-0.0	-0.5	-0.6	-0.2	-0.3	-0.4	-0.2	0.1	-0.2	-0.2	-0.1	-0.1	-0.2	0.2	-0.1	-0.4	-0.1	-0.3	-0.4	-0.3	-0.2	-0.3	-0.1
Продолжит жизни	-0.2	0.3	-0.1	-0.0	0.3	0.2	-0.2	0.4	0.6	0.1	-0.3	0.5	1.0	-0.2	-0.3	0.1	0.1	0.2	-0.5	-0.6	0.2	0.1	-0.1	0.3	0.2	0.1	0.0	0.2	0.2	-0.2	-0.2	-0.3	0.2	0.2	-0.2	-0.1	0.1	0.1	0.2	
Колво авто	-0.1	0.1	0.0	0.3	0.0	-0.4	0.3	-0.5	0.4	-0.4	0.1	-0.4	-0.2	1.0	0.5	-0.1	0.1	0.1	0.4	0.4	0.1	0.0	0.1	0.0	-0.1	0.0	0.1	0.0	0.0	0.1	-0.1	0.3	-0.2	0.2	0.1	0.1	0.1	0.1	0.0	
Общая площадь	-0.2	0.0	-0.1	0.1	-0.1	-0.3	0.3	-0.7	0.8	-0.5	-0.0	-0.6	-0.3	0.5	1.0	-0.1	0.0	0.1	0.6	0.6	0.2	-0.1	0.3	-0.0	-0.4	0.2	-0.1	0.1	-0.0	0.1	0.2	-0.2	0.6	0.0	0.1	0.1	0.2	-0.1	-0.1	-0.1
Численность врачей	-0.2	0.2	0.6	0.5	0.4	-0.4	0.4	0.1	-0.2	-0.3	0.5	-0.2	0.1	-0.1	-0.1	1.0	0.2	-0.2	-0.2	-0.1	0.1	0.6	0.1	0.4	0.3	0.2	0.4	0.3	0.1	0.0	-0.3	0.4	-0.3	0.0	0.1	0.3	0.2	0.0	0.2	-0.1
Колво преступлений	-0.0	1.0	0.3	0.5	0.8	-0.3	0.3	-0.1	0.1	-0.4	-0.0	-0.3	0.1	0.1	0.0	0.2	1.0	-0.1	0.0	0.2	0.8	0.6	0.2	0.8	0.1	0.0	0.3	0.4	0.9	0.1	-0.3	-0.1	0.2	-0.1	0.9	0.9	0.3	0.3	0.6	0.5
Доля граждан ЗОЖ	-0.2	-0.1	-0.2	-0.0	-0.1	-0.0	0.0	-0.1	0.1	-0.0	-0.2	-0.0	0.2	0.1	0.1	-0.2	-0.1	1.0	0.2	0.0	-0.0	-0.1	0.1	-0.1	-0.1	0.4	0.1	0.0	-0.1	0.3	0.1	-0.3	0.1	-0.1	-0.1	-0.1	0.0	-0.2	-0.2	-0.1
Смертность от ССЗ	-0.2	-0.1	-0.2	-0.1	-0.1	-0.3	0.3	-0.7	0.9	-0.2	-0.1	-0.5	-0.5	0.4	0.6	-0.2	0.0	0.2	1.0	0.7	-0.1	-0.1	0.2	-0.1	-0.5	0.2	-0.0	0.2	-0.1	-0.1	0.1	-0.3	0.5	0.1	0.0	0.0	0.3	-0.0	-0.2	-0.2
Смертность от онкологии	-0.1	0.1	-0.1	0.1	0.0	-0.4	0.4	-0.7	0.8	-0.3	-0.0	-0.6	-0.6	0.4	0.6	-0.1	0.2	0.0	0.7	1.0	0.0	0.1	0.1	0.1	-0.4	0.3	0.2	0.2	-0.0	-0.0	-0.2	0.5	0.1	0.1	0.1	0.3	0.1	0.1	-0.1	-0.1
Объем жил строительства	-0.1	0.9	0.3	0.6	0.7	-0.2	0.2	-0.0	0.0	-0.4	-0.0	-0.2	0.2	0.1	0.2	0.1	0.8	-0.0	-0.1	0.0	1.0	0.6	0.2	0.7	0.2	0.1	0.2	0.2	0.9	0.3	-0.2	-0.2	0.1	-0.1	0.8	0.9	0.2	0.1	0.6	0.5
Стоимость 1 квм жилья	-0.2	0.7	0.6	0.8	0.8	-0.5	0.5	0.0	-0.1	-0.4	0.4	-0.3	0.1	0.0	-0.1	0.6	0.6	-0.1	-0.1	0.1	0.6	1.0	0.1	0.8	0.3	0.3	0.4	0.3	0.5	0.2	-0.4	0.2	-0.1	-0.0	0.6	0.7	0.4	0.1	0.5	0.1
Доля молодых ученых	0.0	0.2	0.0	0.1	0.1	-0.3	0.3	-0.2	0.2	-0.3	0.1	-0.4	-0.1	0.1	0.3	0.1	0.2	0.1	0.2	0.1	0.2	0.1	1.0	0.1	-0.1	0.3	0.1	0.2	0.2	0.0	-0.3	0.1	0.3	-0.2	0.3	0.3	0.2	0.3	0.3	-0.0
Затраты на исследования и разработки	-0.1	0.8	0.4	0.6	1.0	-0.4	0.4	0.0	-0.1	-0.3	0.1	-0.2	0.3	0.0	-0.0	0.4	0.8	-0.1	-0.1	0.1	0.7	0.8	0.1	1.0	0.2	0.2	0.4	0.4	0.7	0.1	-0.3	-0.1	0.1	-0.1	0.7	0.8	0.5	-0.0	0.5	0.2
Доля домохозяйств с сетью Интернет	-0.1	0.2	0.5	0.3	0.3	0.0	-0.0	0.4	-0.5	-0.1	0.3	0.1	0.2	-0.1	-0.4	0.3	0.1	-0.1	-0.5	-0.4	0.2	0.3	-0.1	0.2	1.0	-0.1	-0.0	-0.1	0.2	0.1	-0.0	0.4	-0.5	-0.1	0.1	0.1	-0.1	0.0	0.2	0.1
Посещения музеев	-0.1	0.1	0.1	0.2	0.1	-0.4	0.4	-0.2	0.2	-0.3	0.0	-0.2	0.1	0.0	0.2	0.2	0.0	0.4	0.2	0.3	0.1	0.3	0.3	0.2	-0.1	1.0	0.4	0.2	0.0	0.4	-0.2	-0.2	0.2	0.0	-0.0	0.2	0.3	-0.1	-0.0	-0.1
Посещения театров	-0.0	0.3	0.1	0.2	0.3	-0.4	0.4	-0.1	-0.0	-0.1	0.0	-0.2	0.0	0.1	-0.1	0.4	0.3	0.1	-0.0	0.2	0.2	0.4	0.1	0.4	-0.0	0.4	1.0	0.5	0.3	0.0	-0.3	-0.1	0.1	-0.1	0.3	0.4	0.2	0.0	0.1	-0.1
Численность студентов	-0.1	0.4	-0.2	0.1	0.3	-0.1	0.1	-0.1	0.2	-0.1	-0.3	-0.1	0.2	0.1	0.1	0.3	0.4	0.0	0.2	0.2	0.2	0.3	0.2	0.4	-0.1	0.2	0.5	1.0	0.3	-0.1	0.1	-0.3	0.0	-0.2	0.4	0.4	0.3	-0.1	-0.0	-0.1
Численность учителей	0.0	1.0	0.2	0.5	0.7	-0.2	0.2	0.1	-0.1	-0.3	-0.1	-0.1	0.3	0.0	-0.0	0.1	0.9	-0.1	-0.1	-0.0	0.9	0.5	0.2	0.7	0.2	0.0	0.3	0.3	1.0	0.1	-0.2	-0.2	0.1	-0.2	0.9	0.9	0.2	0.2	0.6	0.5
Козф миграц прироста	-0.1	0.1	0.2	0.3	0.1	-0.1	0.1	-0.1	-0.1	-0.3	0.1	-0.2	0.2	0.0	0.1	0.0	0.1	0.3	-0.1	-0.0	0.3	0.2	0.0	0.1	0.1	0.4	0.0	-0.1	0.1	1.0	-0.1	-0.1	0.1	0.1	0.0	0.1	0.0	0.0	0.1	0.2
Сельское хозяйство	-0.3	-0.2	-0.3	-0.4	-0.3	0.5	-0.5	-0.1	0.1	0.2	-0.4	0.2	0.2	0.1	0.2	-0.3	-0.3	0.1	0.1	-0.2	-0.2	-0.4	-0.3	-0.3	-0.0	-0.2	-0.3	0.1	-0.2	-0.1	1.0	-0.4	-0.1	-0.2	-0.2	-0.3	-0.2	-0.3	-0.4	-0.1
Добыча полезных ископаемых	0.5	-0.2	0.6	0.3	0.0	-0.3	0.3	0.2	-0.4	-0.2	0.7	-0.1	-0.2	-0.1	-0.2	0.4	-0.1	-0.3	-0.3	-0.2	-0.2	0.2	0.1	-0.1	0.4	-0.2	-0.1	-0.3	-0.2	-0.1	-0.4	1.0	-0.4	0.1	-0.2	-0.1	-0.1	0.4	0.3	-0.0
Обрабатывающие производства	-0.1	0.1	-0.2	-0.0	0.0	-0.3	0.3	0.5	0.6	-0.3	-0.1	-0.4	-0.2	0.3	0.6	-0.3	0.2	0.1	0.5	0.5	0.1	-0.1	0.3	0.1	-0.5	0.2	0.1	0.0	0.1	0.1	-0.1	-0.4	1.0	-0.0	0.2	0.2	0.3	0.2	0.2	0.0
Электро- и газоснабжение	-0.1	-0.2	0.1	-0.1	-0.1	-0.0	0.0	-0.2	0.1	0.0	0.2	-0.1	-0.3	-0.2	0.0	0.0	-0.1	-0.1	0.1	0.1	-0.1	-0.0	-0.2	-0.1	-0.1	0.0	-0.1	-0.2	-0.2	0.1	-0.2	0.1	-0.0	1.0	-0.2	-0.1	0.2	-0.0	-0.2	0.0
Количество спортивных залов	0.0	0.9	0.2	0.5	0.8	-0.3	0.3	-0.1	0.1	-0.4	-0.1	-0.3	0.2	0.2	0.1	0.1	0.9	-0.1	0.0	0.1	0.8	0.6	0.3	0.7	0.1	-0.0	0.3	0.4	0.9	0.0	-0.2	-0.2	0.2	-0.2	1.0	0.9	0.3	0.3	0.6	0.4
Количество плавательных бассейнов	-0.0	0.9	0.3	0.6	0.8	-0.4	0.4	-0.1	0.0	-0.5	0.1	-0.4	0.2	0.1	0.1	0.3	0.9	-0.1	0.0	0.1	0.9	0.7	0.3	0.8	0.1	0.2	0.4	0.4	0.9	0.1	-0.3	-0.1	0.2	-0.1	0.9	1.0	0.4	0.2	0.6	0.4
Выпуск газет	-0.1	0.3	0.2	0.3	0.5	-0.4	0.3	-0.3	0.3	-0.3	0.1	-0.3	-0.0	0.1	0.2	0.2	0.3	0.0	0.3	0.3	0.2	0.4	0.2	0.5	-0.1	0.3	0.2	0.3	0.2	0.0	-0.2	-0.1	0.3	0.2						



# Стандартизация



$$x_{i,\text{станд.}} = \frac{x_i - \mu}{\sigma}$$

$x_{i,\text{станд.}}$  - стандартизованный элемент признака,  
 $x_i$  - исходный элемент,  
 $\mu$  - среднее арифметическое,  
 $\sigma$  - стандартное отклонение

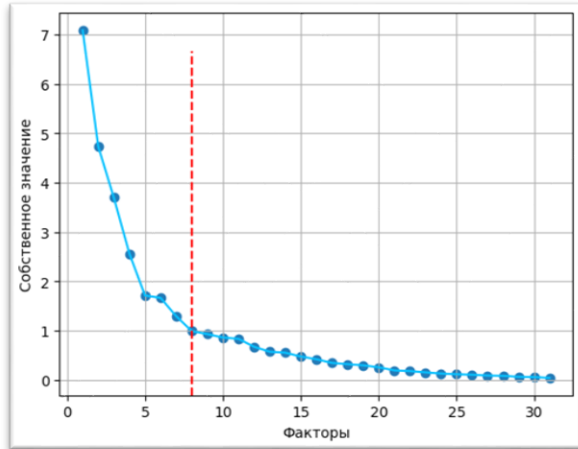
## Фрагмент кода

```
# Стандартизация
import numpy as np
from sklearn.preprocessing import StandardScaler
x = np.array(df)
scaler = StandardScaler()
x_scaler = scaler.fit_transform(x)
```

# Факторный анализ

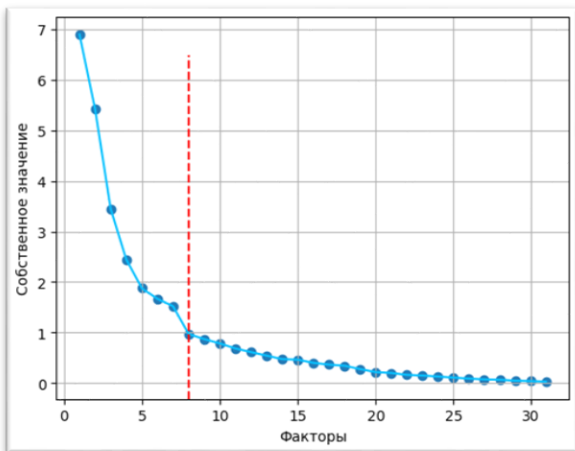
КМО = 0.692

График “Каменистая осыпь” (2016 год)

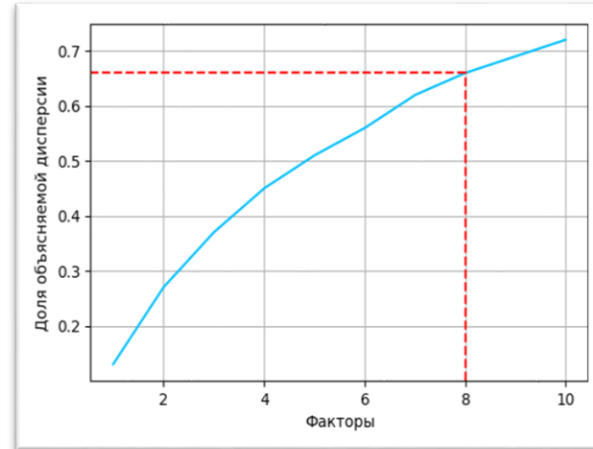


КМО = 0.693

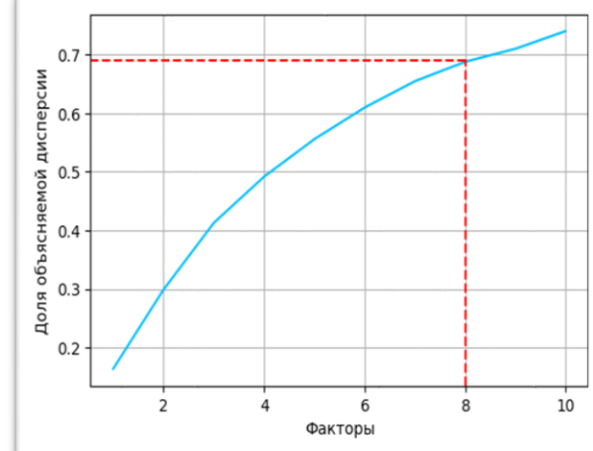
График “Каменистая осыпь” (2021 год)



Зависимость доли объясненной дисперсии от числа факторов (2016 год)



Зависимость доли объясненной дисперсии от числа факторов (2021 год)



Факторы (данные 2016 года)

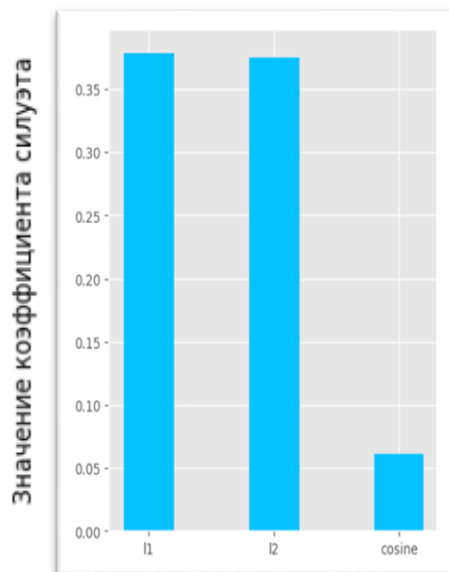
- Фактор 1:** Фактор благосостояния населения (2)
- Фактор 2:** Фактор физического здоровья, условий жизни и занятости населения (6)
- Фактор 3:** Фактор инновационного развития (2)
- Фактор 4:** Фактор экологии (3)
- Фактор 5:** Фактор качества жизни (3)
- Фактор 6:** Фактор культурной жизни (2)
- Фактор 7:** Фактор образования (2)
- Фактор 8:** Фактор популяризации науки (1)

Факторы (данные 2021 года)

- Фактор 1:** Фактор физического здоровья, условий жизни и занятости населения (5)
- Фактор 2:** Фактор благосостояния населения (2)
- Фактор 3:** Фактор благоустройства жилищной сферы и социального строительства (2)
- Фактор 4:** Фактор культурной жизни (2)
- Фактор 5:** Фактор экономики и охраны окружающей среды (5)
- Фактор 6:** Фактор качества жизни (4)
- Фактор 7:** Фактор сельского хозяйства и развития энерго- и газоснабжения (2)
- Фактор 8:** Фактор популяризации науки (2)

# Агломеративная иерархическая кластеризация

Подбор метрик

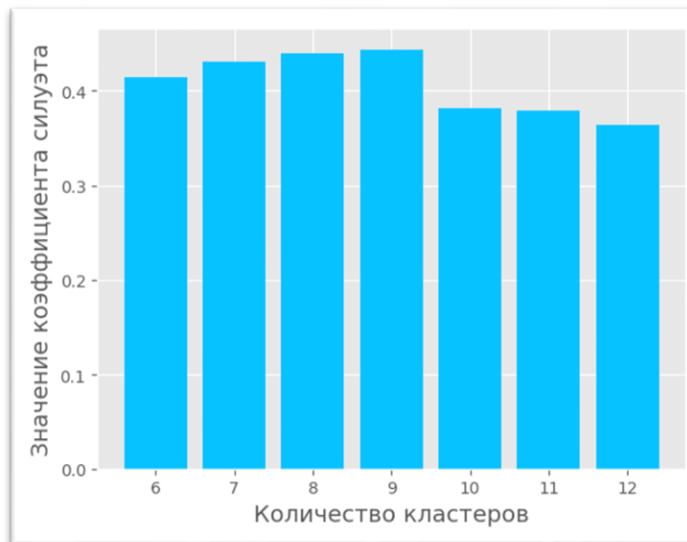


Фрагмент кода

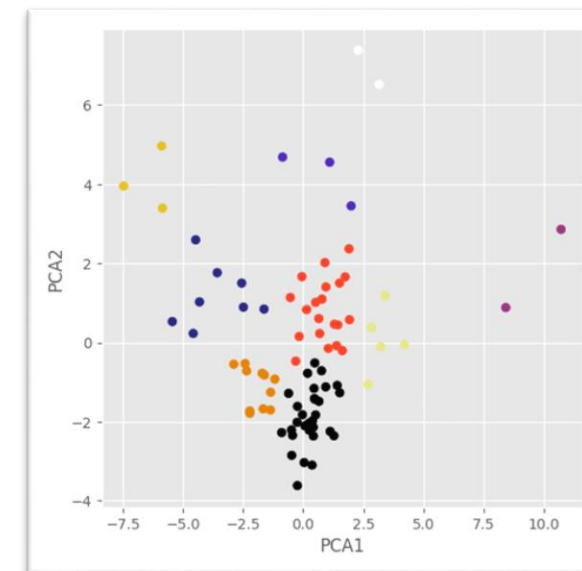
```

model_agg = AgglomerativeClustering(n_clusters = 9,
                                     affinity='euclidean',
                                     linkage='ward')
model_agg.fit(x_fa)
df_agg = df.copy()
df_agg['cluster'] = model_agg.labels_
    
```

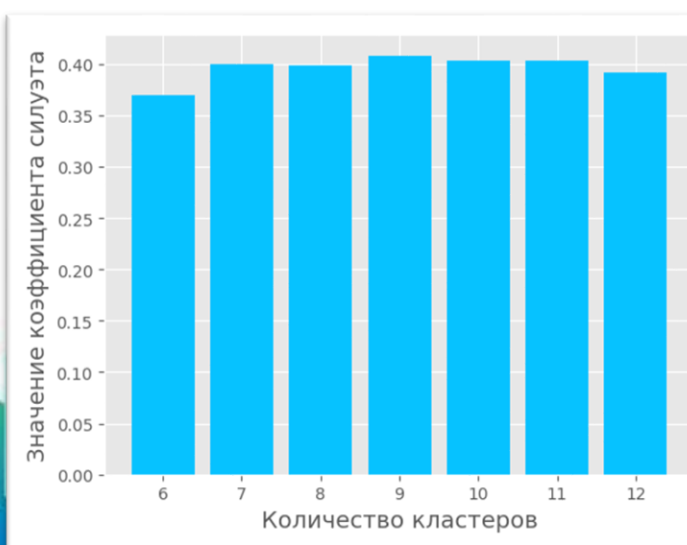
Выбор количества кластеров (2016 год)



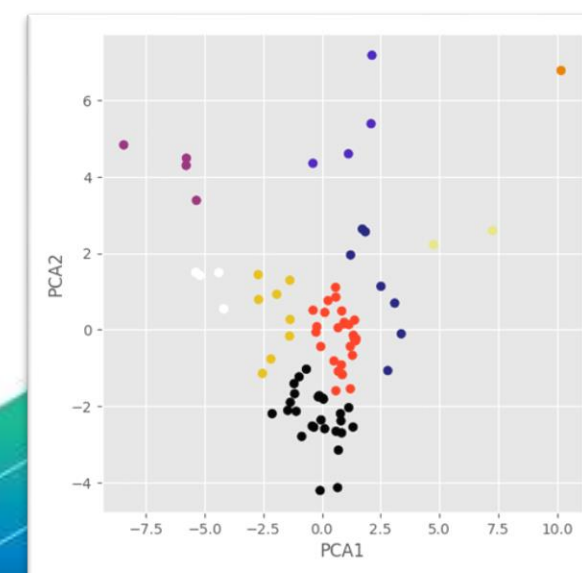
Визуализация кластеризации (2016 год)



Выбор количества кластеров (2021 год)

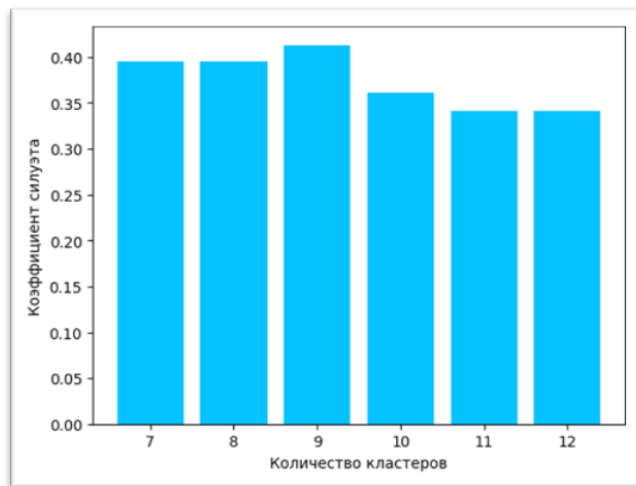


Визуализация кластеризации (2021 год)



# Кластеризация *k*-medoids

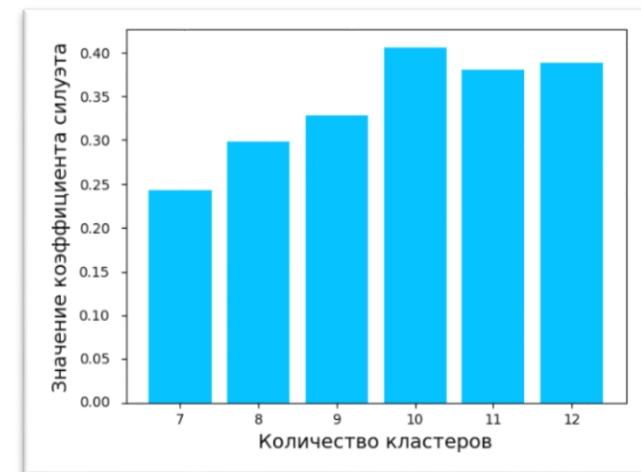
Выбор количества кластеров (данные 2016 года)



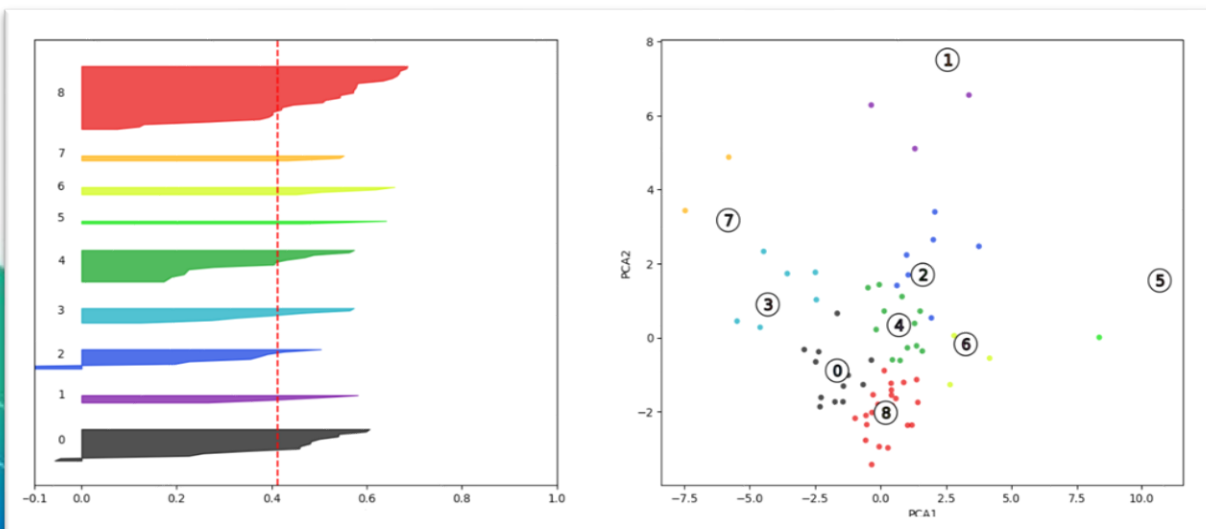
Фрагмент кода

```
model_KMedoids = KMedoids(n_clusters = 10,  
                           random_state = 0,  
                           metric = 'euclidean',  
                           method = 'pam',  
                           init = 'k-medoids++').fit(x_fa)  
model_KMedoids.fit_predict(x_fa)  
df_KMedoids = df.copy()  
df_KMedoids['cluster'] = model_KMedoids.labels_
```

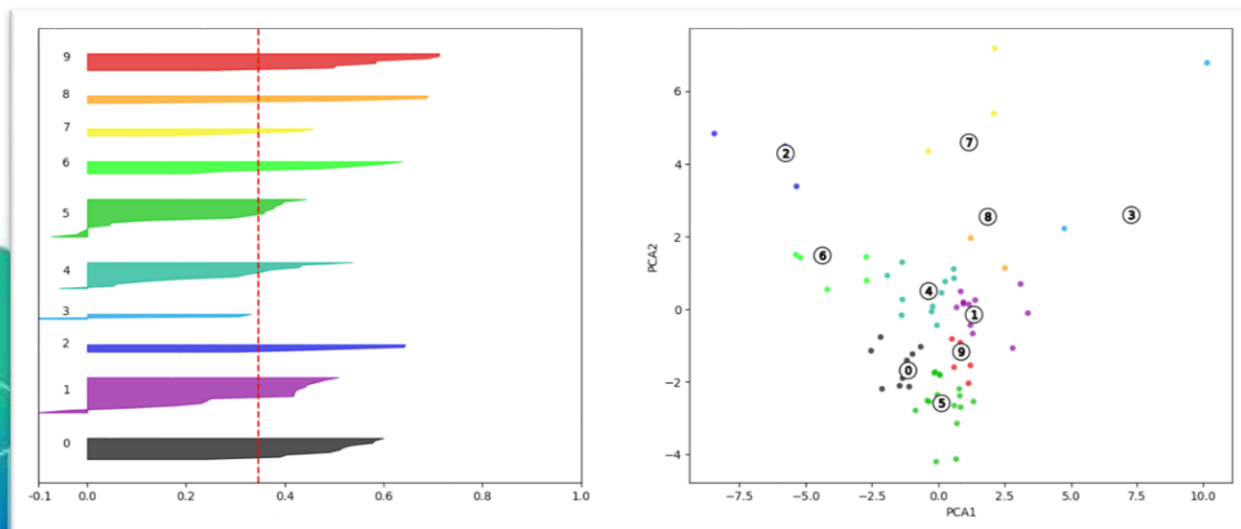
Выбор количества кластеров (данные 2021 года)



Значение коэффициента силуэта и визуализация кластеров (данные 2016 года)



Значение коэффициента силуэта и визуализация кластеров (данные 2021 года)



# Кластеризация $k$ -средних

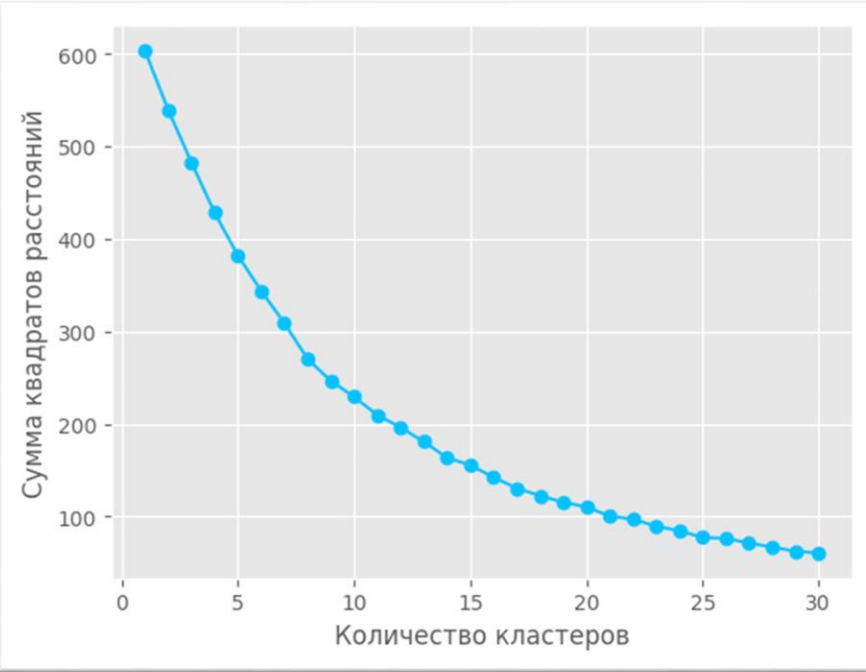


График “локтя” (данные 2016 года)

Фрагмент кода

```

kmeans_kwargs = {
    'init': 'k-means++',
    'n_init': 20,
    'random_state': 42,
}
model_fa = KMeans(n_clusters=10, **kmeans_kwargs)
model_fa.fit_predict(x_fa)
df_means = df.copy()
df_means['cluster'] = model_fa.labels_
    
```

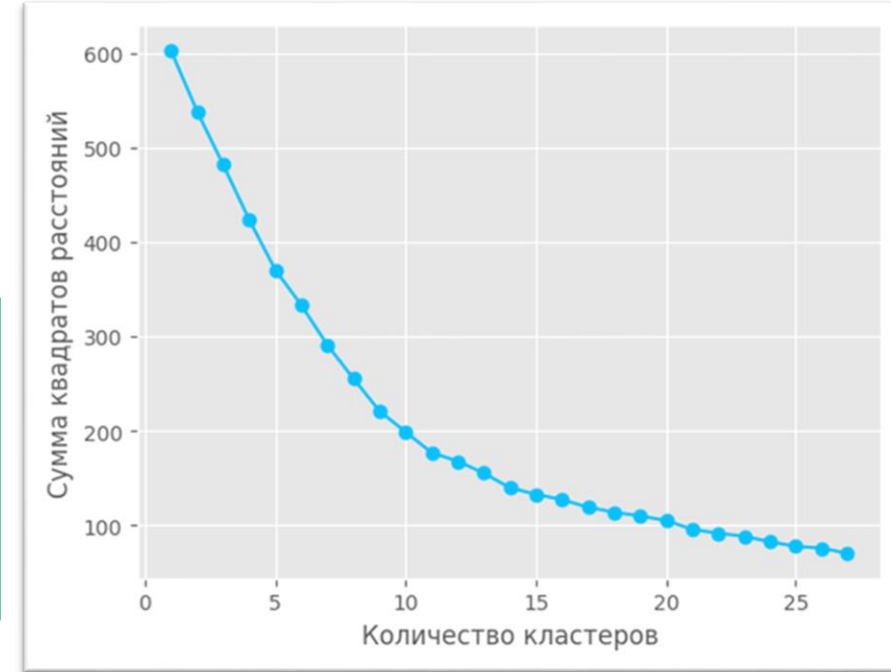


График “локтя” (данные 2021 года)

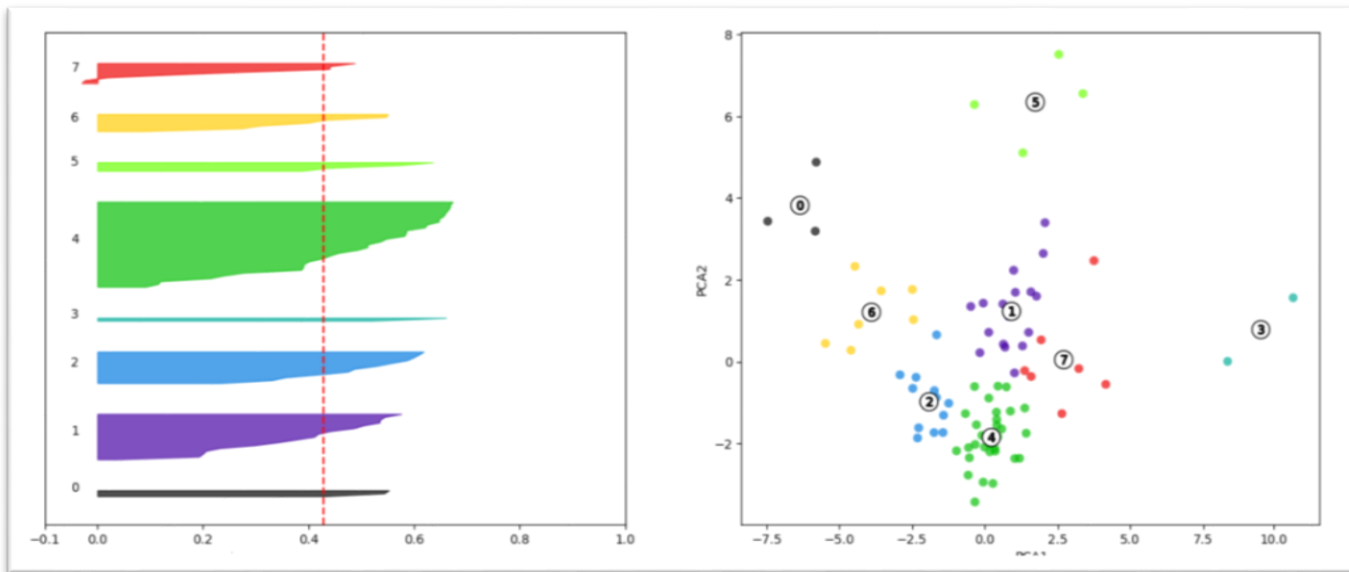


# Кластеризация $k$ -средних

Визуализация кластеризации К-средних  
(при 8 кластерах (данные 2016 года))

Значение коэффициента силуэта

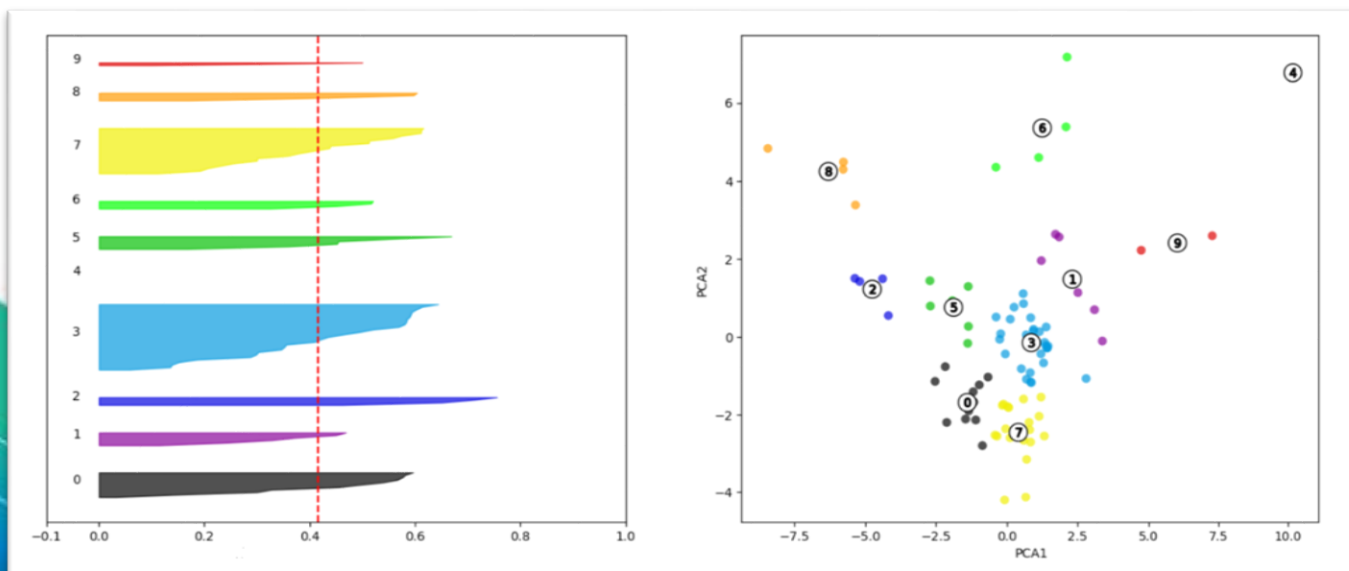
```
For n_clusters = 7 The average silhouette_score is : 0.4364
For n_clusters = 8 The average silhouette_score is : 0.4449
For n_clusters = 9 The average silhouette_score is : 0.4466
For n_clusters = 10 The average silhouette_score is : 0.4092
For n_clusters = 11 The average silhouette_score is : 0.4094
For n_clusters = 12 The average silhouette_score is : 0.3902
```



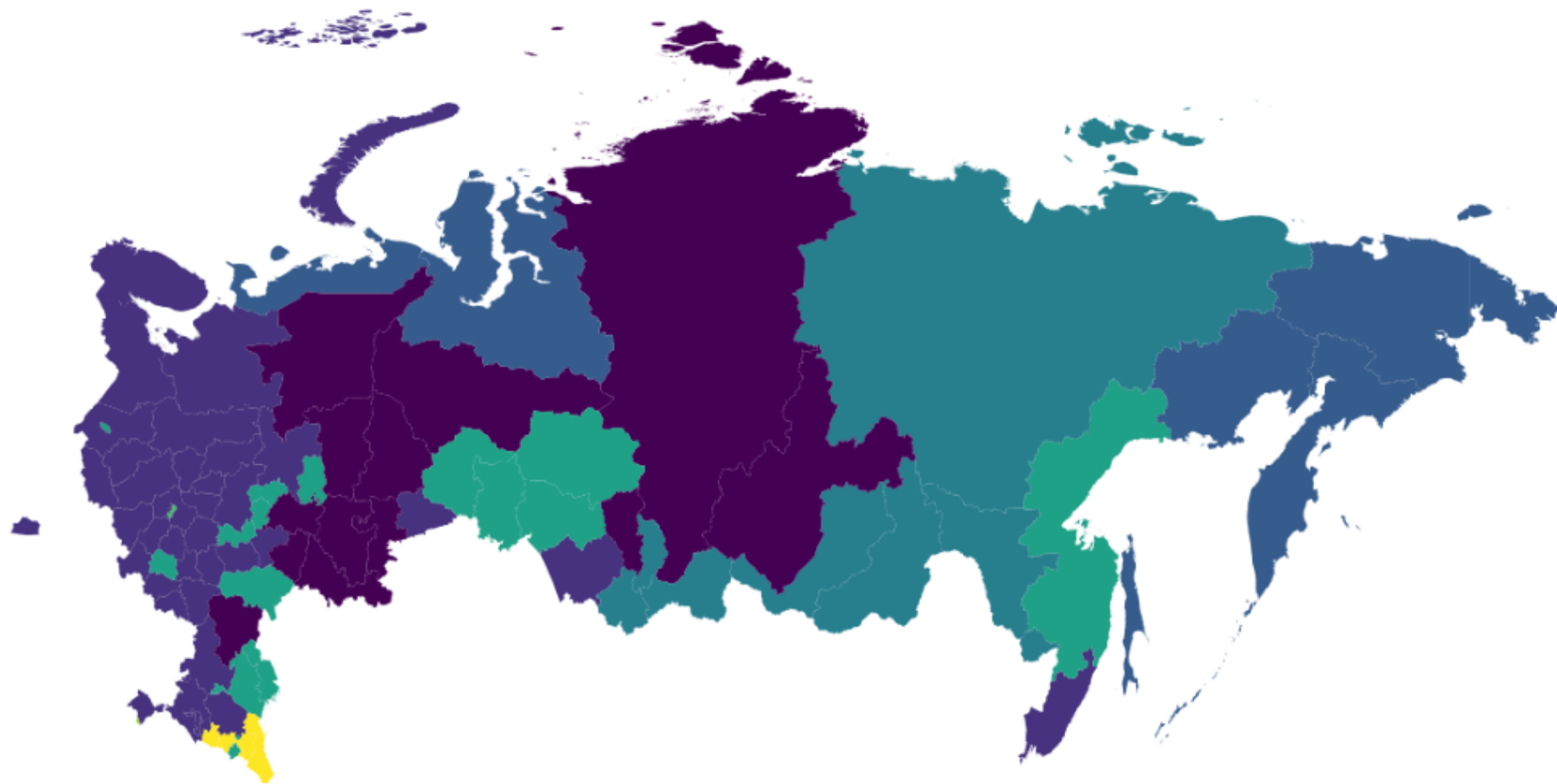
Визуализация кластеризации К-средних  
(при 8 кластерах (данные 2021 года))

Значение коэффициента силуэта

```
For n_clusters = 7 The average silhouette_score is : 0.4157
For n_clusters = 8 The average silhouette_score is : 0.415
For n_clusters = 9 The average silhouette_score is : 0.4014
For n_clusters = 10 The average silhouette_score is : 0.4162
For n_clusters = 11 The average silhouette_score is : 0.3872
For n_clusters = 12 The average silhouette_score is : 0.3764
```



# Картограмма кластерного распределения субъектов РФ (*k*-средних, 2016 год)





# Результаты кластеризации методом $k$ -средних (2016 год)

Название кластера	Состав кластера
Индустриальный кластер	Республика Коми, Волгоградская обл., Республика Башкортостан, Республика Татарстан, Пермский край, Оренбургская обл., Самарская обл., Свердловская обл., Ханты-Мансийский АО, Челябинская обл., Красноярский край, Иркутская обл., Кемеровская обл.
Смешанный кластер 1	Белгородская обл., Брянская обл., Владимирская обл., Воронежская обл., Ивановская обл., Калужская обл., Костромская обл., Курская обл., Липецкая обл., Московская обл., Рязанская обл., Смоленская обл., Тамбовская обл., Тверская обл., Тульская обл., Ярославская обл., Республика Карелия, Архангельская обл., Вологодская обл., Калининградская обл., Ленинградская обл., Мурманская обл., Новгородская обл., Псковская обл., Республика Адыгея, Республика Крым, Краснодарский край, Ростовская обл., Ставропольский край, Кировская обл., Нижегородская обл., Пензенская обл., Ульяновская обл., Курганская обл., Алтайский край, Приморский край
Сырьевой кластер	Ямало-Ненецкий АО, Камчатский край, Магаданская обл., Сахалинская обл., Чукотский АО, Ненецкий АО
Кризисный кластер 1 (дотационный)	Республика Алтай, Республика Тыва, Республика Хакасия, Республика Бурятия, Республика Саха (Якутия), Забайкальский край, Амурская обл., Еврейская АО
Смешанный кластер 2	Орловская обл., г. Санкт-Петербург, Республика Калмыкия, Астраханская обл., Республика Северная Осетия-Алания, Республика Марий Эл, Республика Мордовия, Удмуртская республика, Чувашская республика, Саратовская обл., Тюменская обл., Новосибирская обл., Омская обл., Томская обл., Хабаровский край
Кластер-Лидер	г. Москва
Кластер- "Новая территория"	г. Севастополь
Кризисный кластер 2 (дотационный)	Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская республика, Карачаево-Черкесская республика, Чеченская республика

## Картограмма кластерного распределения субъектов РФ (*k*-средних, 2021 год)

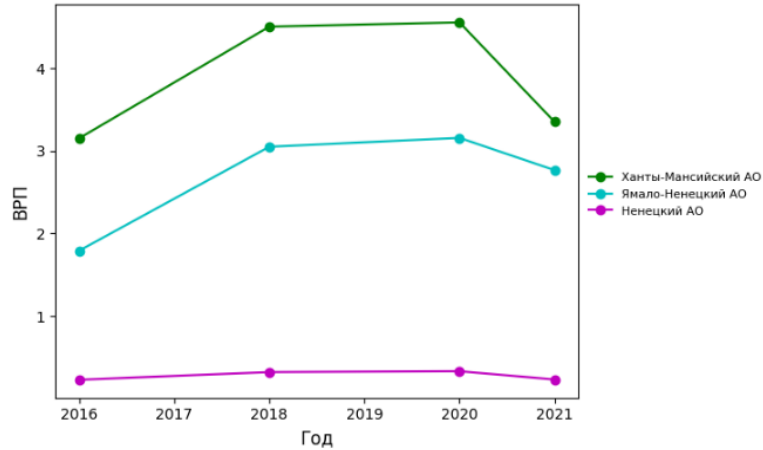


# Результаты кластеризации методом $k$ -средних (2021 год)

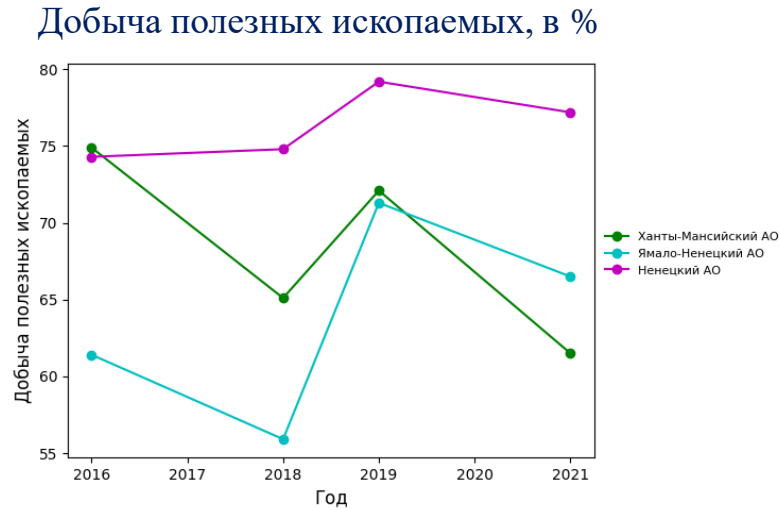
Название кластера	Состав кластера
Промышленный кластер	Брянская обл., Владимирская обл., Ивановская обл., Калужская обл., Костромская обл., Липецкая обл., Рязанская обл., Смоленская обл., Тверская обл., Тульская обл., Ярославская обл., Республика Карелия, Республика Коми, Архангельская обл., Вологодская обл., Мурманская обл., Новгородская обл., Псковская обл., Республика Крым, Волгоградская обл., Республика Марий Эл, Кировская обл., Нижегородская обл., Ульяновская обл., Курганская обл., Республика Хакасия
Индустриальный кластер	Республика Башкортостан, Пермский край, Самарская обл., Свердловская обл., Челябинская обл., Красноярский край, Иркутская обл., Кемеровская обл.
Кризисный кластер (дотационный)	Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Чеченская Республика
Сырьевой кластер 1	Камчатский край, Магаданская обл., Сахалинская обл., Чукотский АО
Кластер-Лидер	г. Москва
Кластер Прогрессивный 1	Московская обл., Ленинградская обл., Краснодарский край
Сырьевой кластер 2	Ханты-Мансийский АО, Ямало-Ненецкий АО, Ненецкий АО
Кластер-Прогрессивный 2	Калининградская обл., г. Санкт-Петербург, г. Севастополь
Депрессивный кластер	Республика Алтай, Республика Тыва, Республика Бурятия, Забайкальский край, Амурская обл., Еврейская АО
Смешанный кластер	Белгородская обл., Воронежская обл., Курская обл., Орловская обл., Тамбовская обл., Республика Адыгея, Республика Калмыкия, Астраханская обл., Ростовская обл., Северная Осетия-Алания, Ставропольский край, Республика Мордовия, Удмуртская республика, Республика Чувашия, Оренбургская обл., Пензенская обл., Саратовская обл., Тюменская обл., Алтайский край, Новосибирская обл., Омская обл., Томская обл., Республика Саха (Якутия), Приморский край, Хабаровский край, республика Татарстан, Тюменская обл.

# Анализ кластера “Сырьевой” (2021 год)

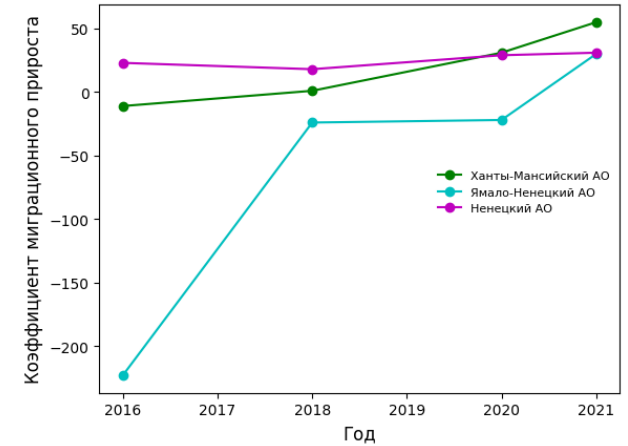
Ханты-Мансийский АО, Ямало-Ненецкий АО, Ненецкий АО



ВРП, млрд. руб.

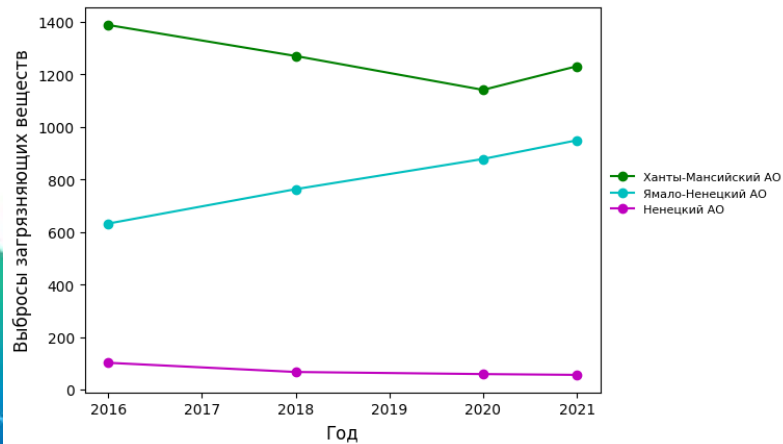


Добыча полезных ископаемых, в %

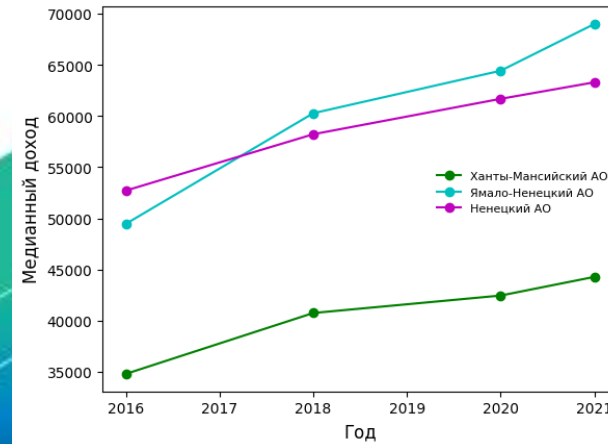


Коэффициент миграционного прироста, На 10000 чел. населения

Выбросы загрязняющих веществ, млн.м<sup>3</sup>



Медианный доход, руб.

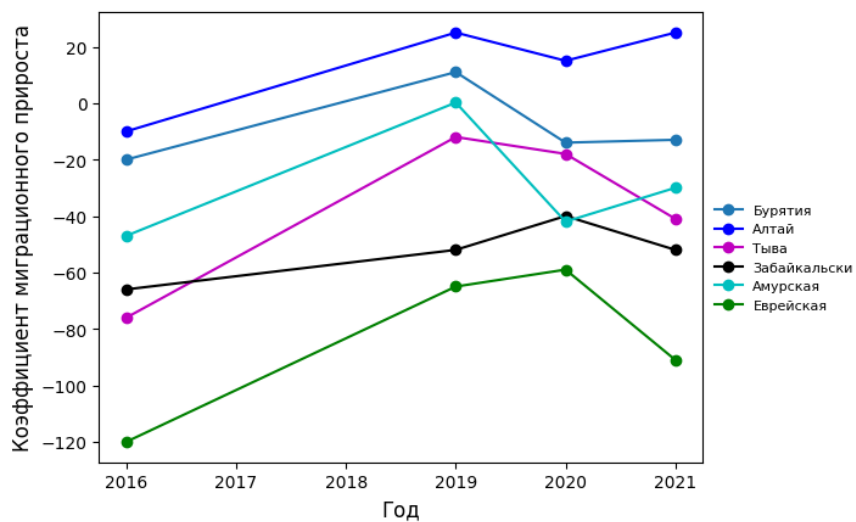


# Анализ кластера Кризисный 1 (2016 год) – Депрессивный (2021 год)

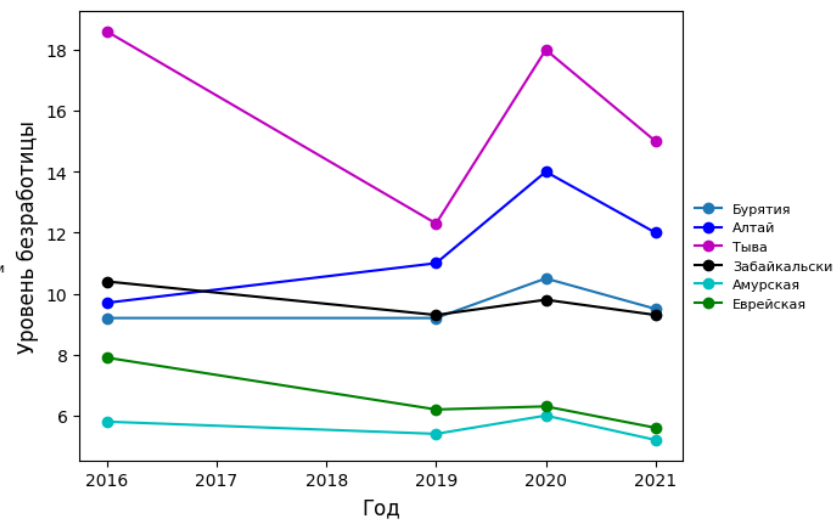
Республика Алтай, Республика Тыва,  
Республика Хакасия, Республика Бурятия,  
Республика Саха (Якутия), Забайкальский край,  
Амурская обл., Еврейская АО



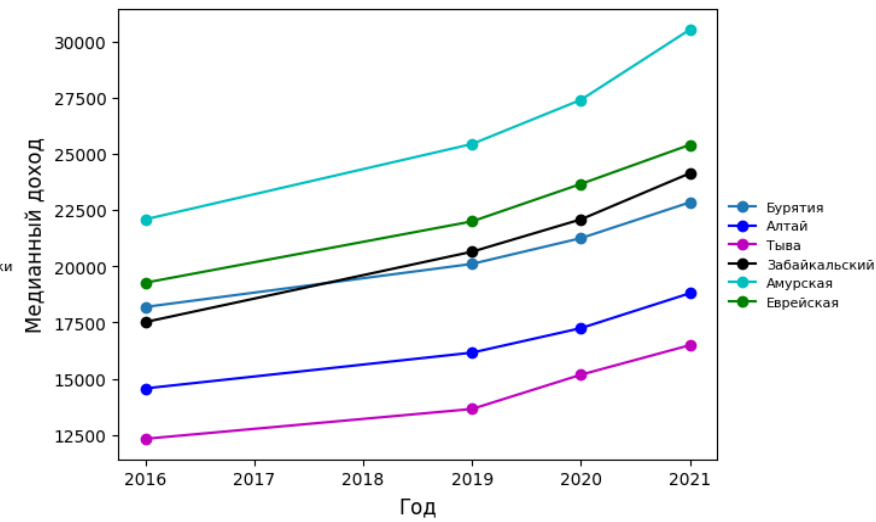
Республика Алтай, Республика Тыва,  
Республика Бурятия, Забайкальский край,  
Амурская обл., Еврейская АО



Коэффициент миграционного прироста, на 10000 чел. населения



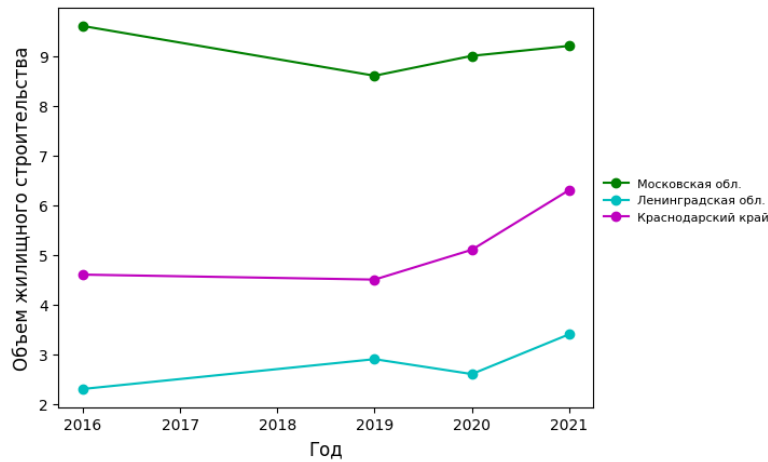
Уровень безработицы, в %



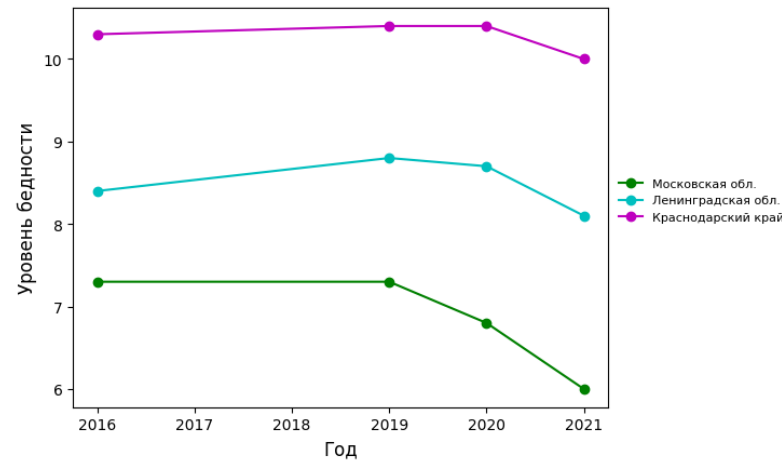
Медианный доход, в руб.

# Анализ кластера “Прогрессивный” (2021 год)

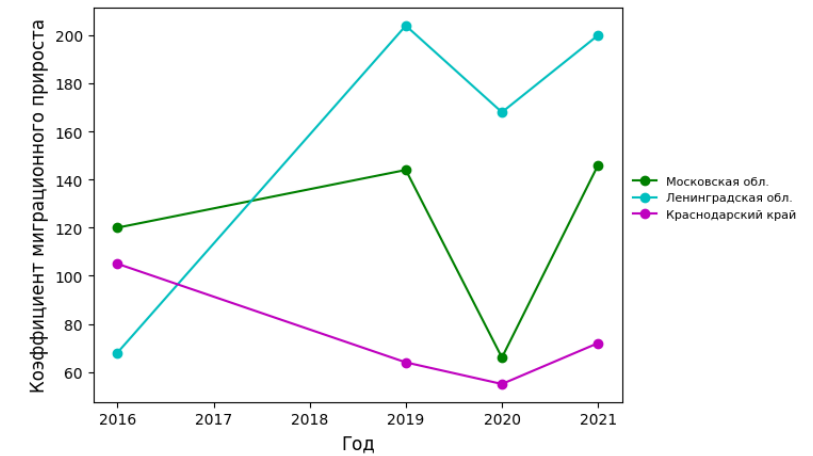
Московская обл., Ленинградская обл., Краснодарский край



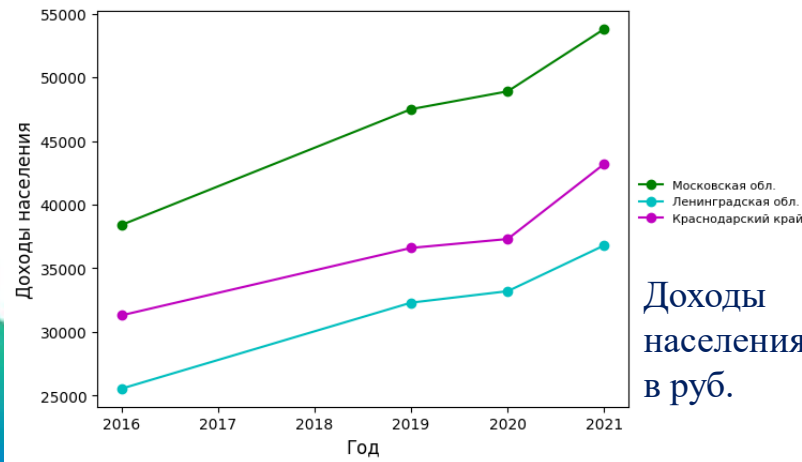
Объем жилищного строительства, млн. м<sup>2</sup>



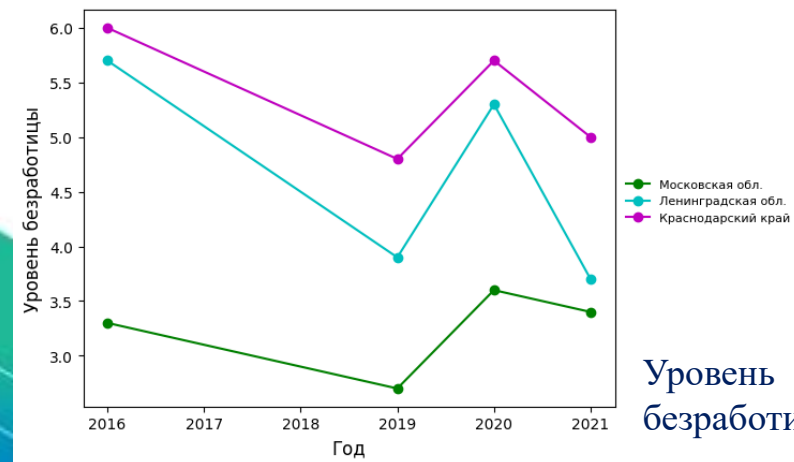
Уровень бедности, в %



Коэффициент миграционного прироста, на 10000 чел. населения



Доходы населения, в руб.



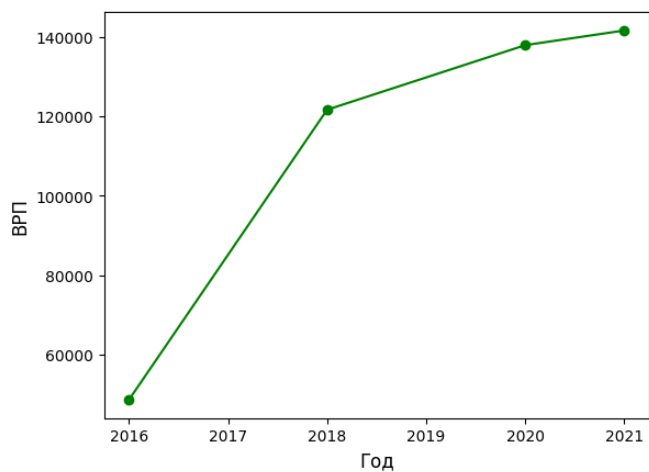
Уровень безработицы, в %

# Анализ кластера “Новая территория” (2016 г.) – “Прогрессивный” (2021 г.)

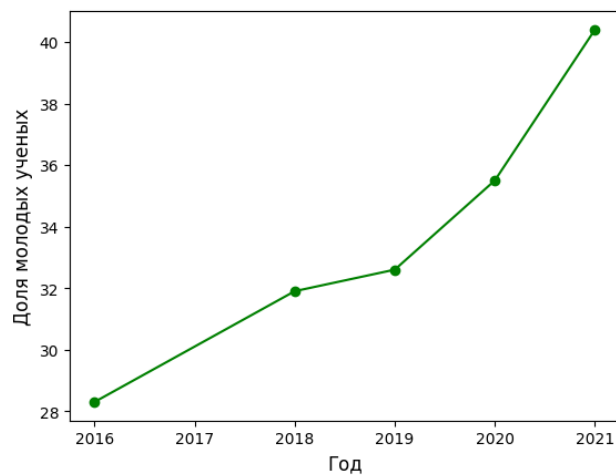
г. Севастополь



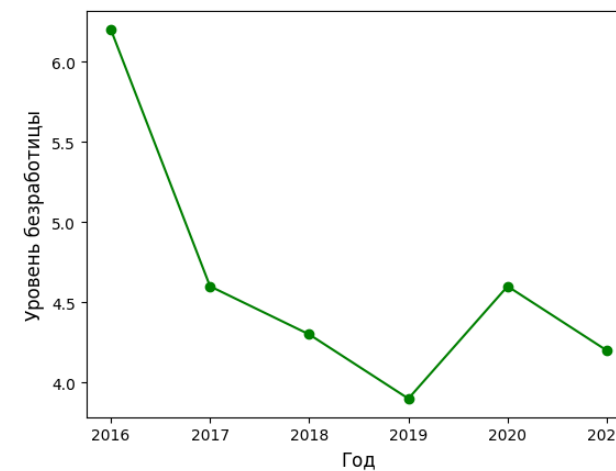
Калининградская обл.,  
г. Санкт-Петербург,  
г. Севастополь



Валовой региональный продукт, млн. руб.  
(по г. Севастополь)



Доля молодых ученых, в %  
(по г. Севастополь)



Уровень безработицы, в %  
(по г. Севастополь)

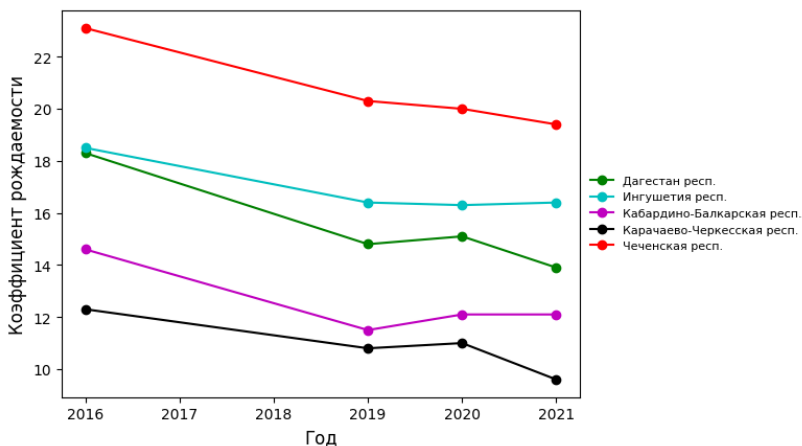


# Анализ кластера “Кризисный” (2016-2021)

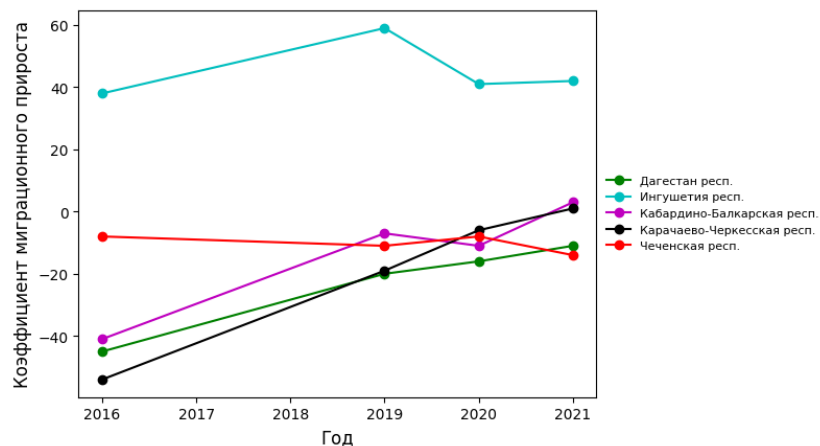
Республика Дагестан, Республика Ингушетия,  
Кабардино-Балкарская республика,  
Карачаево-Черкесская республика, Чеченская республика



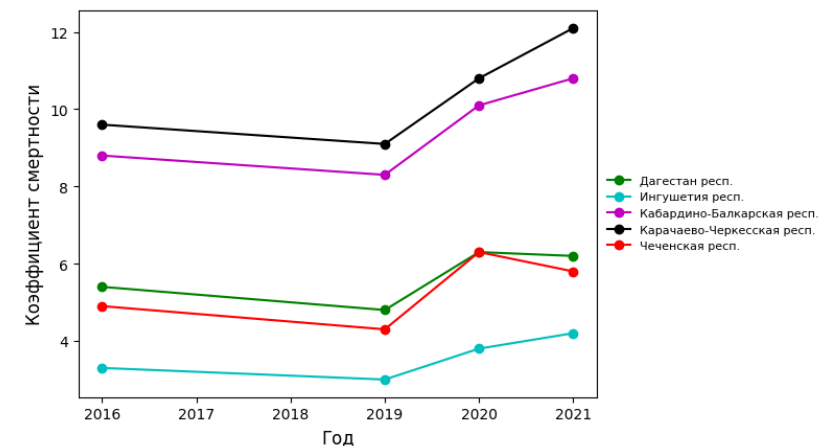
Республика Дагестан, Республика Ингушетия,  
Кабардино-Балкарская Республика,  
Карачаево-Черкесская республика, Чеченская Республика



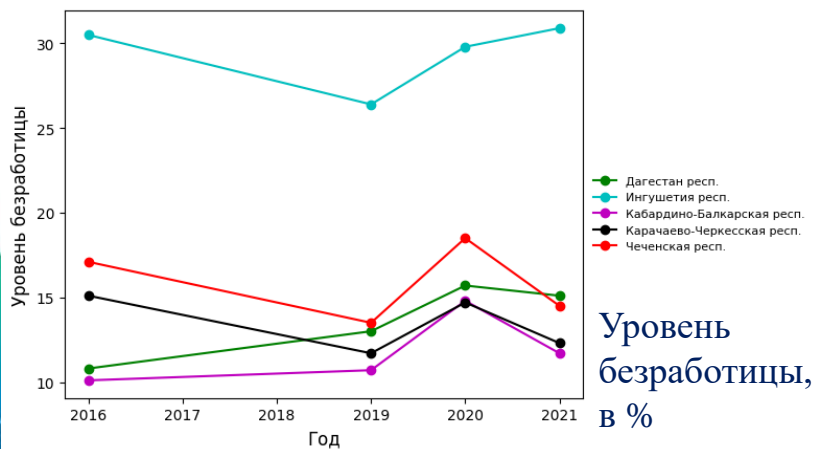
Кoeffициент рождаемости,  
на 1000 чел. населения



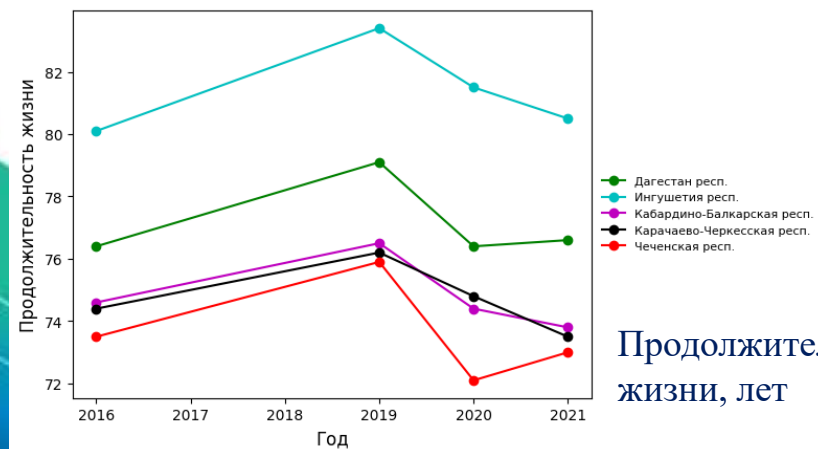
Кoeffициент миграционного прироста  
на 10000 чел. населения



Кoeffициент смертности,  
на 1000 чел. населения



Уровень безработицы,  
в %



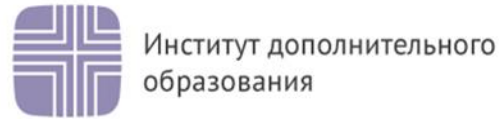
Продолжительность  
жизни, лет

## Выводы

В ходе исследования выполнена кластеризация регионов РФ по уровню социально-экономического развития. На основе имеющихся статистических данных выявлено 8 кластеров (2016 год) и 10 кластеров (2021 год).

Проведен анализ кластеров для определения существенных различий в их социально-экономическом положении.

Выполненная кластеризация может использоваться при реализации кластерно-ориентированной государственной политики с целью поддержки ускоренного развития субъектов.



## ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА

ПРИМЕНЕНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ  
ИССЛЕДОВАНИЯ СОЦИАЛЬНО-ЭКОНОМИЧЕСКОГО  
РАЗВИТИЯ РЕГИОНОВ РОССИЙСКОЙ ФЕДЕРАЦИИ  
по программе профессиональной переподготовки:  
«Анализ данных на языке Python»

Выполнила: Карапетова Елена Александровна  
Руководитель: к.т.н. Семендяев Родион Юрьевич

Санкт-Петербург, 2023