ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ

«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ

«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

*Факультет экономических наук*

Исследовательская работа

Кластеризация регионов по типам инновационного процесса

по направлению подготовки Экономика

образовательная программа «Экономика и статистика»

Выполнил:

Студент группы БСТ213

Еськов Михаил Михайлович

Руководитель:

Балычева Юлия Евгеньевна

Москва 2023

Оглавление

[Введение 3](#_Toc170010628)

[Формулировка темы и актуальности исследования 3](#_Toc170010629)

[Введение в концепцию кластеризации и ее значение для анализа инновационных процессов в регионах 4](#_Toc170010630)

[Цели и задачи исследования 4](#_Toc170010631)

[Обзор литературы 5](#_Toc170010632)

[Методы кластерного анализа 5](#_Toc170010633)

[Иерархические методы кластеризации 5](#_Toc170010634)

[Чёткие (Partitioning) методы 6](#_Toc170010635)

[Мягкие (Fuzzy) методы 6](#_Toc170010636)

[Алгоритмы плотностной кластеризации 7](#_Toc170010637)

[Другие методы 7](#_Toc170010638)

[Примеры кластеризации регионов 7](#_Toc170010639)

[Обзор работы: "Сравнительный анализ региональных различий в сферах здоровья населения, экологии и здравоохранения" 7](#_Toc170010640)

[Обзор работы: "Кластерный анализ регионов Центрального федерального округа по уровню развития человеческого капитала" 8](#_Toc170010641)

[Обзор работы: "Кластеризация регионов Российской Федерации по уровню социально-экономического развития с использованием методов машинного обучения" 10](#_Toc170010642)

[Обзор работы: "Кластеризация регионов России по показателям качества жизни и качества населения" 11](#_Toc170010643)

[Данные и методы исследования 12](#_Toc170010644)

[Исследуемые показатели 12](#_Toc170010645)

[Численность персонала, занятого научными исследованиями и разработками 12](#_Toc170010646)

[Численность исследователей, имеющих ученую степень, по субъектам Российской Федерации 12](#_Toc170010647)

[Внутренние затраты на научные исследования и разработки по субъектам Российской Федерации 13](#_Toc170010648)

[Уровень инновационной активности организаций по субъектам Российской Федерации 13](#_Toc170010649)

[Удельный вес инновационных товаров, работ, услуг в общем объеме отгруженных товаров, выполненных работ, услуг 13](#_Toc170010650)

[Затраты на инновационную деятельность организаций по субъектам Российской Федерации 13](#_Toc170010651)

[Разработанные передовые производственные технологии по субъектам Российской Федерации 14](#_Toc170010652)

[Важность исследования этих показателей 14](#_Toc170010653)

[Сбор и предобработка данных: 15](#_Toc170010654)

[Кластеризация регионов 15](#_Toc170010655)

[Выводы 18](#_Toc170010656)

[Источники 18](#_Toc170010657)

# Введение

## Формулировка темы и актуальности исследования

В условиях стремительных технологических изменений и глобализации инновации становятся одним из ключевых факторов устойчивого экономического роста и повышения конкурентоспособности регионов. Инновации способствуют созданию новых отраслей, модернизации существующих предприятий, улучшению качества жизни населения, а также увеличению доходов. Однако инновационные процессы распределены по регионам неравномерно, что приводит к различиям в уровнях экономического развития, инвестиционной привлекательности и социальной стабильности.

Тема "Кластеризация регионов по типам инновационного процесса" становится всё более актуальной в современных экономических реалиях. Кластеризация позволяет выявить регионы с схожими характеристиками инновационной активности и определить специфические факторы, влияющие на их развитие. Это, в свою очередь, предоставляет ценную информацию для государственных органов и региональных властей, позволяя разрабатывать и адаптировать экономические политики и стратегии, направленные на поддержание и ускорение инновационного роста.

**Актуальность данного исследования подтверждается следующими факторами:**

1. **Неравномерность инновационного развития.** Различия в инновационной активности между регионами порождают дисбалансы в их экономическом развитии, что требует детального анализа и разработки мер по поддержке отстающих регионов.

2. **Эффективное распределение ресурсов.** В условиях ограниченности ресурсов, их направленность на наиболее перспективные и нуждающиеся регионы требует знания особенностей и потребностей каждого из них.

3. **Усиление глобальной конкуренции.** В условиях глобализации регионы должны повышать свою конкурентоспособность на международной арене, что невозможно без активного внедрения и поддержки инноваций.

4. **Реализация национальных стратегий и программ.** Для успешного применения государственных программ поддержки инновационной активности необходимо детально понимать региональные особенности.

## Введение в концепцию кластеризации и ее значение для анализа инновационных процессов в регионах

Кластеризация является мощным аналитическим инструментом, который позволяет группировать объекты (в данном случае регионы) в однородные группы по определённым характеристикам. Метод кластеризации широко используется в разных областях науки и техники для структурирования и упрощения анализа данных. В контексте инновационных процессов кластеризация помогает выявить регионы с похожими характеристиками, что существенно упрощает анализ и позволяет сосредоточиться на специфических потребностях каждой группы регионов.

Значение кластеризации для анализа инновационных процессов в регионах:

1. **Выявление региональных особенностей.** Кластеризация позволяет выявить группы регионов с подобными инновационными характеристиками, что помогает лучше понять их сильные и слабые стороны.

2. **Оптимизация экономической политики.** Разбиение регионов на кластеры способствует разработке более эффективных и целенаправленных мер государственной поддержки, адаптированных к конкретным потребностям каждой группы.

3. **Распределение ресурсов.** Кластеризация позволяет рационально распределять ограниченные ресурсы, направляя их в наиболее перспективные направления.

4. **Идентификация успешных моделей.** Анализ кластеров помогает выявить успешные модели инновационного развития, которые могут быть адаптированы и применены в других регионах.

5. **Мониторинг и оценка динамики.** Кластеризация облегчает отслеживание изменений и динамики инновационной активности в регионах, оценивая эффективность применяемых стратегий и программ поддержки.

6. **Форсайт и прогнозирование.** Кластерный анализ позволяет строить прогнозы и сценарии развития для различных групп регионов, что помогает в стратегическом планировании.

## Цели и задачи исследования

Основной целью данного исследования является проведение кластерного анализа регионов по типам их инновационного процесса для выделения однородных групп и детального понимания их особенностей. Для достижения этой цели необходимо решить следующие задачи:

1. **Выбор основного экономического принципа для разделения регионов на однородные группы.**

2. **Обзор научной литературы по теме исследования** с целью выявления наиболее эффективных методов и подходов к кластеризации.

3. **Выбор и обоснование показателей**, которые могут быть использованы для обучения моделей кластерного анализа.

4. **Сбор и предварительная обработка данных**, включающая работу с выбросами и пропусками для обеспечения надежности анализа.

5. **Подробный предварительный анализ данных** для выявления основных тенденций и характеристик.

6. **Выбор метода кластеризации и проведение анализа** на основе отобранных данных.

7. **Исследование устойчивости полученных кластеров, анализ динамики и переходов** между кластерами во времени.

8. **Обучение дополнительных моделей по кластерам** для более детального анализа выделенных групп регионов и проведение сравнительного анализа.

9. **Оценка основных параметров моделей и метрик качества** для подтверждения достоверности полученных результатов.

10. **Подробная экономическая интерпретация полученных результатов**, включающая сравнительный анализ построенных моделей и сравнение с результатами других исследований.

Проведение данного исследования предоставит теоретическую и практическую базу для анализа и понимания инновационных процессов в регионах, что, в свою очередь, позволит разрабатывать и реализовывать более эффективные стратегии экономической политики.

# Обзор литературы

## Методы кластерного анализа

Кластеризация — это метод анализа данных, который позволяет группировать объекты в однородные группы (кластеры) на основе выбранных характеристик, чтобы объекты в одном кластере были более схожи между собой, чем с объектами из других кластеров. Рассмотрим подробнее различные методы кластеризации.

### Иерархические методы кластеризации

#### Агломеративные методы (bottom-up approach)

1. **Метод одиночной связи (Single Linkage):**

- **Описание:** Этот метод объединяет кластеры, минимизируя минимальное расстояние между точками из разных кластеров. Объекты сначала образуют свои собственные кластеры, затем они последовательно объединяются в большее количество кластеров до тех пор, пока не останется один кластер.

- **Преимущества:** Простота реализации и вычислительная эффективность при малом объеме данных.

- **Недостатки:** Склонен к созданию "цепеобразных" кластеров, что может быть нежелательным.

2. **Метод полной связи (Complete Linkage):**

- **Описание:** объединяет кластеры, минимизируя максимальное расстояние между точками из разных кластеров. Это заставляет кластеры быть более компактными.

- **Преимущества:** создает плотные, компактные кластеры.

- **Недостатки:** Меньшая устойчивость к шуму по сравнению с методом одиночной связи.

3. **Метод средней связи (Average Linkage):**

- **Описание:** объединяет кластеры, минимизируя среднее расстояние между точками из разных кластеров. Это метод является компромиссом между методами одиночной и полной связи.

- **Преимущества:** более сбалансированный подход, чем методы одиночной и полной связи.

- **Недостатки:** Вычислительная сложность выше, чем у методов одиночной и полной связи.

4. **Метод Уорда (Ward's Method):**

- **Описание:** объединяет кластеры, минимизируя увеличение общей внутрикластерной дисперсии. Этот метод использует анализ разброса для оценки расстояния между кластерами.

- **Преимущества:** часто создает кластеры равномерного размера и формы.

- **Недостатки:** Высокая вычислительная сложность, особенно на больших наборах данных.

#### Дивизионные методы (top-down approach)

1. **Метод DIANA (Divisive Analysis):**

- **Описание:** Этот подход начинается с одного крупного кластера, включающего все объекты, и последовательно делит его на более мелкие кластеры, пока не будут достигнуты заданные критерии разделения.

- **Преимущества:** Удобен для определения крупномасштабных кластерных структур.

- **Недостатки:** Вычислительно более сложен и менее популярен, чем агломеративные методы.

### Чёткие (Partitioning) методы

1. **K-средних (K-means):**

- **Описание:** делит множество объектов на K кластеров, минимизируя сумму квадратов расстояний объектов до центроидов кластеров. Алгоритм итеративно улучшает разбиение, чтобы минимизировать внутренние вариации классов.

- **Преимущества:** Простота и быстрота выполнения, особенно при больших объемах данных.

- **Недостатки:** требует заранее заданного числа кластеров и чувствителен к выбросам.

2. **K-medoids (PAM, Partitioning Around Medoids):**

- **Описание:** Аналогичен методу K-средних, но центроиды (медоиды) являются реальными объектами данных, что делает метод менее чувствительным к выбросам.

- **Преимущества:** Устойчивость к выбросам.

- **Недостатки:** более высокая вычислительная сложность по сравнению с K-средними.

### Мягкие (Fuzzy) методы

1. **Fuzzy C-means (FCM):**

- **Описание:** позволяет каждому объекту принадлежать к нескольким кластерам одновременно с различными степенями принадлежности (weights). Метод итеративно обновляет степени принадлежности и центроиды кластеров для минимизации внутрикластерной дисперсии.

- **Преимущества:** Гибкость и возможность моделировать данные с нечеткими границами кластеров.

- **Недостатки:** требует больше вычислительной мощности и точной настройки параметров.

### Алгоритмы плотностной кластеризации

1. **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):**

- **Описание:** определяет кластеры как области с высокой плотностью объектов, разделенные областями с низкой плотностью (шум). Алгоритм не требует заранее задавать число кластеров.

- **Преимущества:** Способен выявлять кластеры произвольной формы и устойчив к шуму.

- **Недостатки:** Выбор параметров (minPts и ε) может быть сложным и критически влияет на результаты.

### Другие методы

1. **Спектральная кластеризация (Spectral Clustering):**

- **Описание:** использует спектр (собственные значения) матрицы смежности графа данных для определения кластеров. Методы спектральной кластеризации основаны на теории графов и линейной алгебре.

- **Преимущества:** Эффективен для данных с сильно неконвексными формами кластеров.

- **Недостатки:** Высокая вычислительная сложность и зависимость от параметров.

2. **Самоорганизующиеся карты (Self-Organizing Maps, SOM):**

- **Описание:** Нейросетевая модель, которая преобразует многомерное пространство данных в двумерную решетку, сохраняя топологическую структуру данных. SOM менее подвержены локальным минимумам по сравнению с методами, основанными на случайной инициализации.

- **Преимущества:** Визуализация высокоразмерных данных и выявление сложных связей в данных.

- **Недостатки:** Сложность настройки и интерпретации результатов.

## Примеры кластеризации регионов

### Обзор работы: "Сравнительный анализ региональных различий в сферах здоровья населения, экологии и здравоохранения"

Аннотация:

Статья посвящена сравнительному анализу региональных различий в России по ключевым аспектам: здоровье населения, экология и система здравоохранения. Авторы использовали эконометрические методы, в частности кластерный анализ, для выявления и анализа различий по выбранным параметрам.

Методология:

- **Кластерный анализ:** Метод K-средних для классификации регионов.

- **Исходные данные:** Официальная статистика Росстата.

- **Параметры анализа:**

- "Здоровье населения": Средняя продолжительность жизни и общая заболеваемость населения.

- "Экология": Выбросы вредных веществ в атмосферу и грязных вод в сточные воды.

- "Здравоохранение": Количество коек на 100 000 населения, количество больниц, количество врачей на 10 000 населения и мощность больниц.

Основные выводы:

- В течение 2000-х годов наблюдается значительное улучшение по показателям здоровья населения и экологии.

- К 2010 году количество регионов с "плохим" и "низким" уровнем здоровья населения значительно снизилось.

- Большинство регионов демонстрируют высокий уровень экологической безопасности.

- Динамика изменений в сфере здравоохранения носит биполярный характер. Наблюдается уменьшение числа регионов с "хорошим" и "плохим" уровнем здравоохранения, одновременно увеличилось количество регионов с "высоким" и "низким" уровнем.

Основные данные и результаты:

1. **Здоровье населения:**

- Средняя продолжительность жизни увеличилась почти на 4 года за период с 1999 по 2011 год.

- Наибольший рост показателя наблюдается в Северо-Западном ФО.

- Заболеваемость по регионам увеличилась на 16% за 16 лет.

2. **Экология:**

- В целом снижение выбросов вредных веществ на 10%, выбросов грязных вод — на 35%.

- Стабильная экологическая ситуация в большинстве регионов.

3. **Здравоохранение:**

- Снижение количества коек на 25%.

- Увеличение числа врачей на 10%.

- Уменьшение числа больничных учреждений на 49%.

- Мощность амбулаторно-поликлинических учреждений выросла на 12%.

### Обзор работы: "Кластерный анализ регионов Центрального федерального округа по уровню развития человеческого капитала"

Аннотация:

В статье предложен расширенный набор показателей для оценки уровня развития человеческого капитала в регионе. Представлены результаты кластерного анализа регионов Центрального Федерального округа (ЦФО) по этим показателям, проведенного на основе расширенной системы индикаторов. Были выявлены ключевые направления развития человеческого капитала для различных кластеров регионов ЦФО.

Ключевые слова:

Человеческий капитал, показатели уровня развития человеческого капитала, кластерный анализ, кластерный анализ регионов по уровню развития человеческого капитала, направления развития человеческого капитала.

Методология исследования:

- **Кластерный анализ:** Метод K-средних для классификации регионов.

- **Исходные данные:** Использованы данные Федеральной службы государственной статистики.

- **Параметры для анализа:**

- X1 — Ожидаемая продолжительность жизни, лет.

- X2 — Обеспеченность врачами на 10 000 населения, чел.

- X3 — Уровень занятости, %.

- X4 — Доля безработных с высшим образованием, %.

- X5 — Численность учащихся профессиональных учебных заведений всех видов на 10 000 населения, чел.

- X6 — Отношение среднедушевых денежных доходов к прожиточному минимуму, раз.

- X7 — Обеспеченность жильем, м²/чел.

Все переменные были стандартизованы методом линейного масштабирования, который используется при расчете индекса развития человеческого потенциала (ИРЧП) по методике ПРООН.

Основные результаты:

Кластеры 2007 года:

1. **Кластер 1 (лидеры):** г. Москва.

2. **Кластер 2 (образовательный рост):** Воронежская, Ивановская, Курская, Рязанская, Смоленская, Тверская, Ярославская области.

3. **Кластер 3 (профессиональный рост и благосостояние):** Белгородская, Орловская, Тамбовская области.

4. **Кластер 4 (средний уровень):** Липецкая, Московская области.

5. **Кластер 5 (аутсайдеры):** Брянская, Владимирская, Калужская, Костромская, Тульская области.

Кластеры 2009 года:

1. **Кластер 1 (лидеры):** г. Москва.

2. **Кластер 2 (образовательный рост):** Воронежская, Ивановская, Курская, Рязанская, Смоленская, Тверская, Ярославская области.

3. **Кластер 3 (профессиональный рост и благосостояние):** Белгородская, Калужская, Липецкая, Московская, Тамбовская, Тульская области.

4. **Кластер 4 (средний уровень):** Брянская, Орловская области.

5. **Кластер 5 (аутсайдеры):** Владимирская, Костромская области.

Кластеры 2011 года:

1. **Кластер 1 (лидеры):** г. Москва.

2. **Кластер 2 (образовательный рост):** Воронежская, Ивановская, Курская, Рязанская, Смоленская, Тверская, Ярославская области.

3. **Кластер 3 (профессиональный рост и благосостояние):** Московская область.

4. **Кластер 4 (средний уровень):** Белгородская, Калужская, Липецкая, Орловская, Тамбовская области.

5. **Кластер 5 (аутсайдеры):** Брянская, Владимирская, Костромская, Тульская области.

Основные выводы и рекомендации:

1. **Регионы-лидеры:** Основные проблемы связаны с высоким уровнем безработицы среди людей с высшим образованием и низкой обеспеченностью жильем. Рекомендации включают создание программ доступного жилья и меры по повышению мобильности рабочей силы.

2. **Образовательные регионы:** Проблемы связаны с низким уровнем материального благополучия и высокой безработицей. Рекомендовано создание эффективных рабочих мест, стимулирование малого бизнеса и улучшение качества медицинского обслуживания.

3. **Регионы с профессиональным ростом:** требуется поддержка малого бизнеса и создание новых рабочих мест. Необходимо развивать систему среднего специального образования и совершенствовать качество медицинской помощи.

4. **Средние уровни:** Регионы сохраняют достаточный уровень благосостояния, но испытывают трудности в сфере здравоохранения и профессионального образования. Требуются меры по привлечению врачей и улучшению медицинских услуг.

5. **Аутсайдеры:** Низкий уровень всех показателей развития человеческого капитала. Необходим комплексный подход, включающий поддержку здравоохранения, образование и создание высокооплачиваемых рабочих мест.

### Обзор работы: "Кластеризация регионов Российской Федерации по уровню социально-экономического развития с использованием методов машинного обучения"

Аннотация:

В данной работе решается задача кластеризации регионов Российской Федерации по уровню социально-экономического развития с учетом отраслевой структуры валового регионального продукта (ВРП) с использованием классических методов машинного обучения.

Основные цели и задачи:

- **Цель исследования:** Выделение однородных кластеров регионов с целью разработки эффективных форм поддержки регионов и стимулирования экономического роста.

- **Задачи:**

1. Определить структуру показателей социально-экономического развития регионов.

2. Собрать и проверить качество большого набора данных для проведения кластерного анализа.

3. Выполнить предобработку данных: заполнение пропусков, преобразование данных, стандартизация, выделение основных индикаторов методом главных компонент.

4. Выделить однородные региональные кластеры с использованием методов машинного обучения.

5. Проанализировать дифференциацию средних показателей развития региональных кластеров.

Основные методы:

1. **Метод главных компонент (PCA):** используется для выявления 15 базовых индикаторов социально-экономического развития региона.

2. **Метод K-средних (K-means):** Применен для кластеризации регионов на основании 15 выделенных индикаторов.

Предобработка данных:

- **Исходные данные:** Статистическая информация из Федеральной службы государственной статистики за 2019 год, включающая 5525 записей по 65 показателям социально-экономического развития 85 регионов РФ.

- **Преобразование данных:** Переход к удельным величинам, стандартизация показателей.

- **Метод главных компонент:** Снижение размерности до 15 индикаторов, объясняющих около 85% дисперсии.

Результаты кластерного анализа:

- **Выделено 5 кластеров:**

1. **Кластер 1 (г. Москва, г. Санкт-Петербург):** Высокая доля оптовой и розничной торговли, операций с недвижимым имуществом, профессиональной и научной деятельности, высокая занятость, низкий уровень безработицы.

2. **Кластер 2 (регионы ЦФО):** Специализация на обрабатывающем производстве, сельском хозяйстве, торговле.

3. **Кластер 3 (смешанная экономика):** Средние значения по основным социально-экономическим показателям.

4. **Кластер 4 (СКФО, ЮФО, СФО, ДФО):** Высокая безработица, большая доля государственной деятельности, низкие значения социально-экономических показателей.

5. **Кластер 5 (ДФО, УФО, СЗФО):** Специализация на добыче полезных ископаемых, максимальные среднедушевые доходы.

Заключение:

- **Научная новизна:** Исследование позволяет выделить устойчивые региональные кластеры на основе большого объема статистических данных.

- **Практическая значимость:** Полученные результаты могут использоваться при формировании государственной политики, направленной на поддержание и ускорение социально-экономического развития регионов.

### Обзор работы: "Кластеризация регионов России по показателям качества жизни и качества населения"

Аннотация:

В статье рассматривается задача кластеризации регионов Российской Федерации по показателям качества жизни и качества населения. Цель исследования — сформировать группы регионов с схожими показателями и разработать стратегии для улучшения качественных характеристик населения.

Основные цели и задачи:

- **Цель исследования:** Кластеризация регионов России для последующего поиска эффективных мер воздействия на качество жизни с целью улучшения качества населения.

- **Задачи:**

1. Анализ корреляций между показателями качества жизни и качества населения.

2. Выбор значимых показателей для кластеризации.

3. Проведение кластерного анализа регионов по выбранным показателям.

4. Разработка типологии регионов на основе результатов кластерного анализа.

Основные методы:

1. **Корреляционный анализ:** позволил выявить взаимосвязь между показателями качества жизни и качества населения.

2. **Кластерный анализ:** Проведен для группировки регионов по совокупности семи показателей.

Показатели для анализа:

- Показатели качества жизни: среднедушевые денежные доходы, уровень бедности, уровень безработицы.

- Показатели качества населения: естественный прирост населения, продолжительность жизни, уровень образования, численность больных алкоголизмом и наркоманией.

Результаты кластерного анализа:

1. **Кластер 1 (6 регионов):** Основные характеристики включают высокие денежные доходы, низкий уровень безработицы, высокий уровень образования и высокий уровень обеспеченности медицинскими услугами.

2. **Кластер 2 (16 регионов):** характеризуется низкими денежными доходами и высоким уровнем бедности.

3. **Кластер 3 (3 региона):** Низкие уровни доходов и высокая безработица, региональный лидер по естественному приросту и продолжительности жизни.

4. **Кластер 4 (15 регионов):** Самый благополучный кластер, лидирует по уровню безработицы и обеспеченности жильем.

5. **Кластер 5 (13 регионов):** Отрицательный естественный прирост населения и высокий уровень бедности.

6. **Кластер 6 (28 регионов):** Показатели, близкие к среднероссийским, большинство из которых занимают средние позиции по основным показателям.

7. **Кластер 7 (4 региона):** Высокие доходы, низкая безработица, но наибольший уровень алкоголизма и наркомании, а также наименьшая продолжительность жизни.

Заключение:

Проведенная кластеризация регионов позволила выделить 7 групп с различными проблемами и достижениями. Полученные результаты могут служить основой для разработки целевых программ и мер поддержки, направленных на улучшение качественных характеристик населения в каждом кластере.

# Данные и методы исследования

## Исследуемые показатели

Исследование показателей научной и инновационной активности регионов является критически важным для понимания и стимулирования экономического роста, повышения конкурентоспособности и устойчивого развития. Рассмотрим каждый из упомянутых показателей более детально:

### Численность персонала, занятого научными исследованиями и разработками

Этот показатель отражает количество людей, которые работают в сфере научных исследований и разработок (НИОКР) в конкретном регионе. Это могут быть как исследователи, так и технический персонал и административные работники, занимающиеся поддержкой НИОКР задач.

**Важность:**

- Даёт представление о масштабе научно-исследовательской деятельности.

- Помогает оценить уровень инвестиционного потенциала региона в наукоёмкие отрасли.

- Может служить показателем перспективности региона для научных проектов и инновационных инициатив.

### Численность исследователей, имеющих ученую степень, по субъектам Российской Федерации

Этот показатель указывает на количество исследователей с научными степенями (кандидаты и доктора наук) в различных регионах.

**Важность:**

- Является индикатором квалификационного уровня научного сообщества региона.

- Высокая численность высококвалифицированных исследователей может свидетельствовать о развитой научной инфраструктуре и высоком уровне научных исследований.

- Привлекателен для инвесторов и компаний, занимающихся передовыми технологическими разработками.

### Внутренние затраты на научные исследования и разработки по субъектам Российской Федерации

Этот показатель показывает объем финансовых ресурсов, направляемых на НИОКР в конкретном регионе.

**Важность:**

- Отражает финансирование научной и инновационной деятельности.

- Позволяет оценить государственную и частную поддержку научных проектов.

- Связан с возможностями выполнения крупных проектов и программ.

### Уровень инновационной активности организаций по субъектам Российской Федерации

Этот показатель демонстрирует процент организаций в регионе, участвующих в инновационной деятельности.

**Важность:**

- Индикатор динамики и потенциала развития новых технологий и продуктов.

- Способствует пониманию степени вовлеченности бизнеса в инновационные процессы.

- Связан с общей конкурентоспособностью и экономическим развитием региона.

### Удельный вес инновационных товаров, работ, услуг в общем объеме отгруженных товаров, выполненных работ, услуг

Этот показатель измеряет долю инновационной продукции и услуг в общем объеме производственной и сервисной деятельности.

**Важность:**

- Показывает влияние инноваций на экономику региона.

- Важен для оценки коммерциализации научных разработок.

- Указывает на зрелость и востребованность инновационных продуктов на рынке.

### Затраты на инновационную деятельность организаций по субъектам Российской Федерации

Этот показатель отражает объем средств, направляемых организациями региона на инновационную деятельность (НИОКР, внедрение новых технологий и т.д.).

**Важность:**

- Подчеркивает инвестиции в развитие нововведений.

- Связан с потенциалом масштабирования и распространения инноваций.

- Отражает ориентированность компаний и государственного сектора на модернизацию и инновации.

### Разработанные передовые производственные технологии по субъектам Российской Федерации

Этот показатель демонстрирует количество новых технологий, разработанных в регионе, готовых к внедрению в производство.

**Важность:**

- Индицирует уровень технологического прогресса и изобретательства.

- Служит основой для улучшения производственной эффективности и конкурентоспособности.

- Связан с возможностями трансфера технологий и улучшением производственных процессов.

## Важность исследования этих показателей

Исследование и анализ этих показателей важно для нескольких ключевых аспектов:

1. **Обоснованность региональной политики:** Понимание текущего состояния и динамики научной и инновационной активности позволяет разрабатывать и корректировать стратегии регионального развития, направленные на увеличение конкурентоспособности и устойчивого экономического роста.

2. **Инвестиционная привлекательность:** Эти данные помогают инвесторам принимать информированные решения, оценивая потенциал и риски инвестиций в конкретных регионах, особенно в наукоемкие и технологичные сектора экономики.

3. **Разработка программ поддержки и финансирования:** Указанные показатели служат ориентиром для распределения государственных и частных ресурсов на поддержку и развитие ключевых инновационных направлений, инфраструктуры и кадрового потенциала.

4. **Целевая помощь и стимулы:** Анализ позволяет выявить регионы с недостаточной активностью, требующие дополнительных стимулов и поддержки для развития НИОКР и инноваций.

5. **Мониторинг и оценка прогресса:** Регулярный сбор и анализ этих показателей позволяет отслеживать прогресс в области науки и технологий, оценивать эффективность внедряемых мер и программ, а также сравнивать результаты как в динамике, так и относительно других регионов и стран.

Исследование данных показателей и их сотрудничество с научными и инновационными процессами в регионах способствует формированию комплексного подхода к управлению и развитию научно-технического и инновационного потенциала, что является ключевым фактором для обеспечения устойчивого и долгосрочного развития регионов и страны в целом.

## Сбор и предобработка данных:

**Источники данных:**

- Статистические данные с официального сайта Росстата.

- Базы данных, предоставляемые государственными органами и научными организациями.

- Отчеты и публикации по результатам научных исследований и инновационной деятельности.

- Собраны данные за 2015-2021гг

**Методы сбора данных:**

- Скачивание данных с официальных сайтов и порталов.

- Использование встроенных функций обработки табличных данных, например, pandas для чтения CSV, Excel файлов и баз данных.

**Предобработка данных**

- Исходные данные имели много пропусков, которые были заполнены средними. Этот вариант подходит лучше всего, так как большинство методов кластеризации основаны на расстоянии между объектами, а средняя меньше всего влияет на это.

- Были удалены регионы, для которых очень мало данных: Еврейская автономная область и Чукотский автономный округ.

- Данные приведены к одному масштабу с помощью нормализации, чтобы каждый регион вносил равный «вклад» во время алгоритма определения кластера принадлежности.

В результате был получен очищенный, нормализованный набор данных, готовый для последующего анализа и моделирования. Приведение данных к единому формату и масштабу обеспечивает точность и надежность выводов, полученных в ходе исследования, и повышает качество решений, основанных на этих данных.

# Кластеризация регионов

В качестве основных методов кластеризации были выбраны методы k-средних и метод Варда. Эти методы является самыми устойчивыми к аномалиям в данных и достаточно просты в применении. Например, для использования метода DBS-Scan необходимо было бы выбирать гиперпараметры, от которых завесили бы выводы исследования. Нейросетевой подход имел бы большое смещение, поскольку модель достаточно сложная. Остальные методы либо слишком простые, либо достаточно схожи с выбранными, но имеют больше недостатков. Метод k-средних и метод Варда получили сильно схожие результаты. 81 регион из 86 получили одинаковые методы кластеров, что говорит и высоком качестве выполненной предобработке и выборе показателей для исследования. Остановимся на методе Варда, поскольку он имеет статистическое обоснование и не имеет элемента случайности.

В первый кластер попали регионы: Брянская область, Владимирская область, Воронежская область, Ивановская область, Калужская область, Костромская область, Курская область, Орловская область, Рязанская область, Смоленская область, Тамбовская область, Тверская область, Ярославская область, Республика Карелия, Республика Коми, Архангельская область, в том числе Ненецкий автономный округ, Архангельская область без АО, Вологодская область, Калинингpадская область, Ленинградская область, Мурманская область, Новгородская область, Псковская область, Республика Адыгея, Республика Калмыкия, Республика Крым2) , Краснодарский край, Астраханская область, Волгоградская область, г. Севастополь 2) , Республика Дагестан, Республика Ингушетия , Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Республика Северная Осетия - Алания, Чеченская Республика , Ставропольский край, Республика Марий Эл, Удмуртская Республика, Кировская область, Оренбургская область, Саратовская область, Ульяновская область, Курганская область, в том числе:, Ямало-Ненецкий автономный округ, Республика Алтай , Республика Тыва, Республика Хакасия, Алтайский край, Иркутская область, Кемеровская область, Омская область, Томская область, Республика Бурятия, Республика Саха (Якутия), Забайкальский край, Камчатский край, Приморский край, Амурская область, Магаданская область.

Во второй кластер попали: Белгородская область, Липецкая область, Московская область1), Тульская область, г. Москва1), Ростовская область, Республика Башкортостан, Республика Мордовия, Республика Татарстан, Чувашская Республика, Пермский край, Нижегородская область, Пензенская область, Самарская область, Свердловская область, Тюменская область, Ханты-Мансийский автономный округ - Югра, Тюменская область без АО, Челябинская область, Красноярский край, Новосибирская область, Хабаровский край, Сахалинская область.

В третий кластер попал только Санкт-Петербург.

Третий кластер является самым развитым в контексте инновационного процесса. Он имеет больше всего технологий, больше технологий развивает и больше тратит ресурсов на развитие. Вторым по развитости является второй кластер, и первый кластер на последнем месте. Посмотрим на графики некоторых показателей по полученным кластерам и убедимся в сделанных выводах.

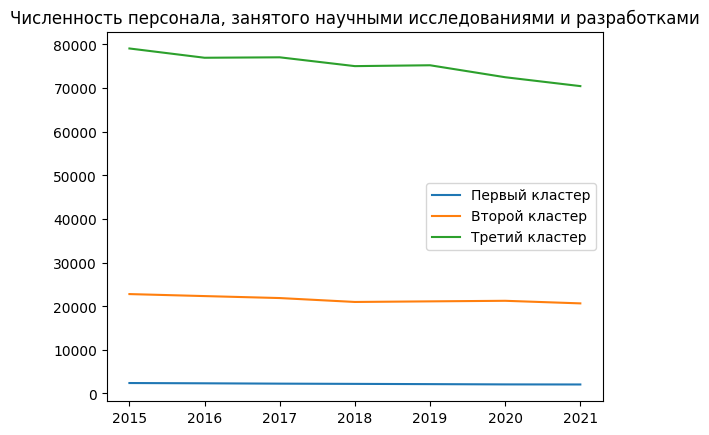


Рис. 1

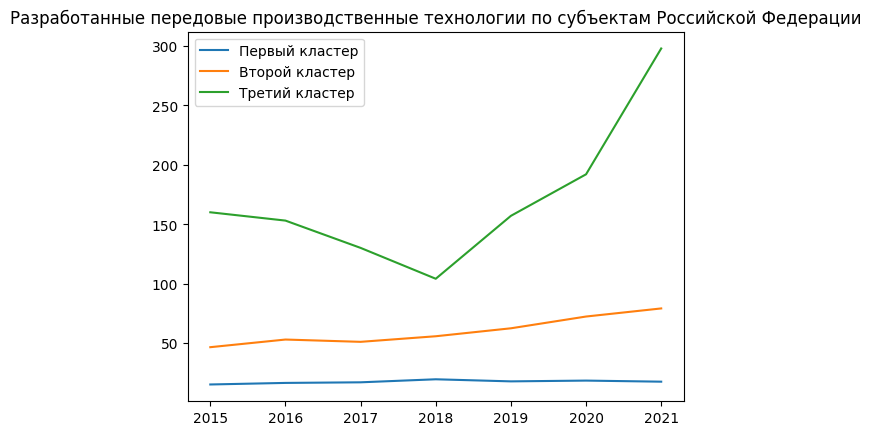


Рис. 2

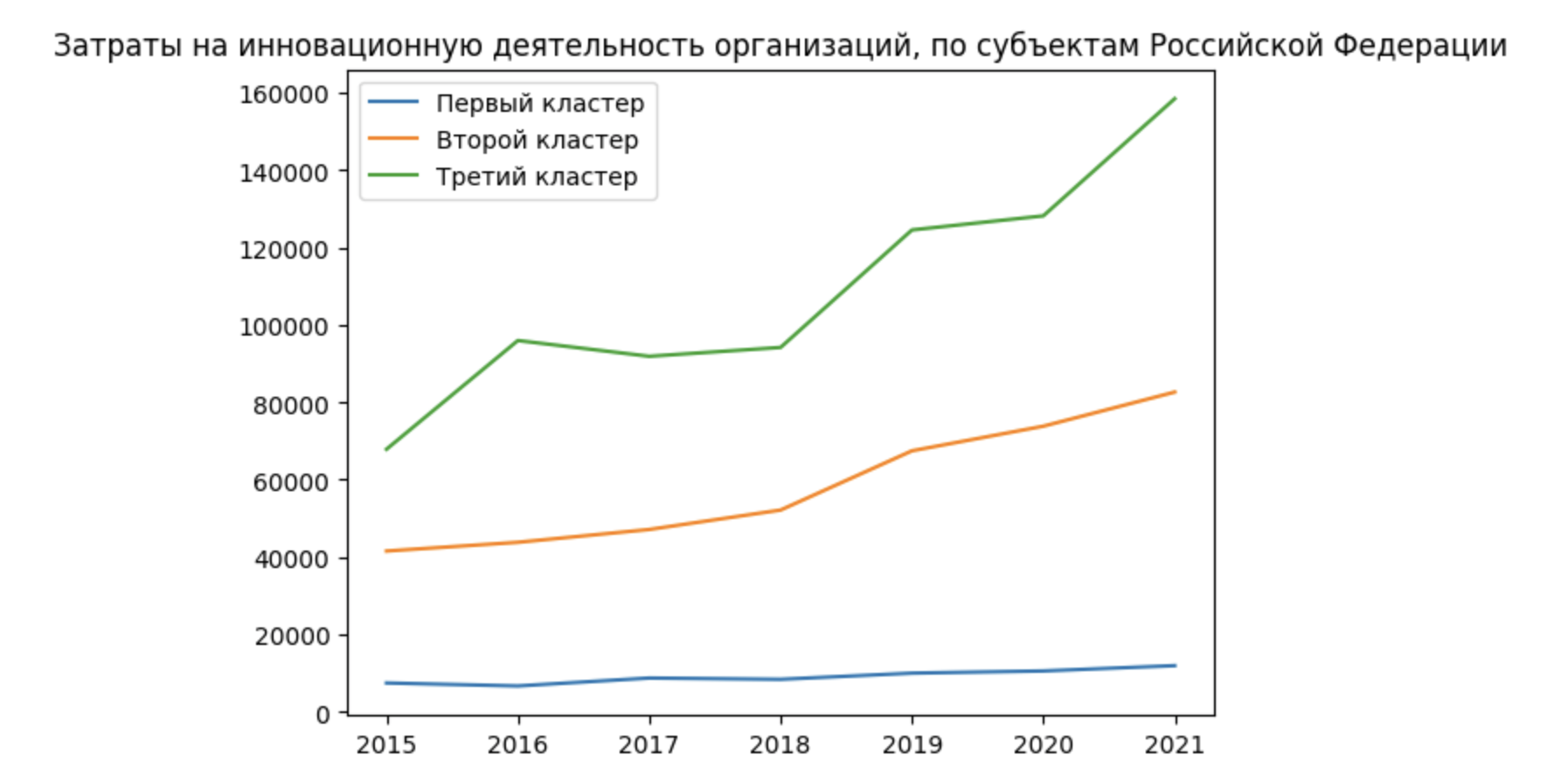


Рис. 3



Рис. 4

Как можно видеть, алгоритм кластеризации получил очень хорошие и интерпретируемые результаты. Графики подтверждают выводы, сделанные выше. Основная причина отставания остальных регионов из-за нехватки затрат на инновационную деятельность. По графикам хорошо видно корреляция этих величин.

# Выводы

Исследование, направленное на кластеризацию регионов по типам инновационного процесса, позволило сделать несколько значимых выводов, которые могут служить основой для дальнейших исследований и разработки стратегий регионального развития.

1. **Разделение регионов на кластеры:**

- Кластеризация регионов Российской Федерации по типам инновационного процесса выявила несколько четко различимых групп, каждая из которых характеризуется своими особенностями и уровнями развития научной и инновационной деятельности.

- Этот подход позволяет лучше понять неоднородность регионов и их сравнительные преимущества в разных аспектах инновационного процесса.

2. **Выявление ведущих и отстающих регионов:**

- Анализ показателей, таких как численность персонала, занятая научными исследованиями, расходы на НИОКР, уровень инновационной активности и другие, позволил выделить лидирующие регионы, которые демонстрируют высокую инновационную активность и значительные внутренние затраты на НИОКР.

- Одновременно были идентифицированы регионы с низкой инновационной активностью, где наблюдается недостаток ресурсов и структурных условий для стимулирования научно-технического прогресса.

3. **Характеристика кластеров:**

- **Кластер 1:** Регионы с высокой концентрацией научных кадров и значительными внутренними расходами на НИОКР. Эти регионы характеризуются высоким уровнем инновационной активности и успешной коммерциализацией научных разработок.

- **Кластер 2:** Регионы со средним уровнем научно-технической активности, умеренными расходами на НИОКР и потенциальными возможностями для роста. В этих регионах имеются условия для развития, но требуется дополнительная поддержка.

- **Кластер 3:** Регионы с низкой инновационной активностью, ограниченными внутренними расходами на НИОКР и минимальным числом высококвалифицированных исследователей. Здесь требуется значительная государственная и частная поддержка для стимулирования инновационных процессов.

4. **Рекомендации для региональной политики:**

- Ведение целенаправленных стратегий для каждого кластера может содействовать более эффективному распределению ресурсов и стимулированию инновационного процесса.

- Для регионов-лидеров рекомендуется поддерживать и усиливать существующие научно-технические и образовательные инфраструктуры.

- Для регионов со средним уровнем активности целесообразно внедрение программ, направленных на повышение доступности финансирования и развитие межрегиональных коопераций.

- Для отстающих регионов необходимо создание и реализация комплексных программ поддержки, включая меры по привлечению высококвалифицированных специалистов, развитие инфраструктуры и улучшение условий для инновационной деятельности.

5. **Важность нормализации и предварительной обработки данных:**

- Проведенная нормализация и предварительная обработка данных, включая работу с выбросами и заполнение пропусков методом средних, обеспечили более точные и надежные результаты кластеризации.

- Приведение данных к единому масштабу позволило учесть различия в единицах измерения и масштабах, что существенно повысило качество кластерного анализа.

Заключение

Кластеризация регионов Российской Федерации по типам инновационного процесса предоставила ценные знания о текущем состоянии и потенциале инновационного развития в разных частях страны. Эти выводы могут стать основой для разработки эффективной региональной политики, направленной на стимулирование научно-технического прогресса и улучшение конкурентоспособности регионов. Методологический подход и результаты анализа могут быть применены для подобных исследований в других странах и регионах, что подчеркивает универсальность и значимость проведенного исследования.

# Список литературы

Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). "Cluster Analysis". Wiley.

Gan, G., Ma, C., & Wu, J. (2007). "Data Clustering: Theory, Algorithms, and Applications". SIAM.

Bishop, C. M. (2006). "Pattern Recognition and Machine Learning". Springer.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise". KDD.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). "An Introduction to Statistical Learning with Applications in R". Springer.

Лавриненко П.А., Рыбакова Д.А. Сравнительный анализ региональных различий в сферах здоровья населения, экологии и здравоохранения // Экономические и социальные перемены: факты, тенденции, прогноз. 2015. No 5 (41). С. 198–210.

Орлова И.В., Филонова Е.С. Кластерный анализ регионов Центрального федерального округа по социально-экономическим и демографическим показателям // Статистика и экономика. 2015. No 5. С. 111–115. DOI: 10.21686/2500-3925-2015-5-136-142

Кетова К.В., Касаткина Е.В., Вавилова Д.Д. Кластеризация регионов Российской Федерации по уровню социально-экономического развития с использованием методов машинного обучения. // Экономические и социальные перемены: факты, тенденции, прогнозы DOI: 10.15838/esc.2021.6.78.4