

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования**

**Национальный исследовательский университет
“Высшая школа экономики”**

**Факультет экономических наук
Образовательная программа Экономика и статистика**

Проектная работа

На тему "Кластерный анализ регионов Российской Федерации по уровню
инновационной активности с использованием методов машинного обучения"

Выполнил студент группы БСТ213
Курченко М. А.

Руководитель проекта:
Ст. преподаватель Балычева Ю. Е.

Москва, 2024 г.

Содержание

Введение	3
Обзор литературы и публикаций.....	4
Сбор, обработка и визуализация данных	6
Кластерный анализ регионов	10
Алгоритм k-means и метод локтя	10
Кластеризация регионов с учетом объемов выпуска	12
Кластеризация регионов в разрезе статей затрат	17
Список литературы	22

Введение

Инновации играют ключевую роль в развитии общества, определяя его прогресс и улучшая качество жизни граждан. Они меняют ландшафт производства и торговли, стимулируют экономический рост. В современном мире, где технологии стремительно развиваются, инновации становятся все более значимыми. Их интеграция во все сферы деятельности необходима для устойчивого развития и повышения конкурентоспособности. Актуальность работы обусловлена неравномерным развитием региональных систем – современный этап пространственного развития России характеризуется значительным дисбалансом между регионами по уровню их социально-экономического развития, что во многом обусловлено различиями в ресурсной базе. Кластерный анализ может помочь в разработке целевых программ финансирования и поддержки отстающих регионов. Целью работы ставлю изучение инновационной активности регионов Российской Федерации, выделение групп субъектов и их сравнительный анализ. Для этого необходимо провести кластеризацию регионов на основе затрат на инновации, визуализировать результаты, дать качественную оценку получившимся кластерам. В работе был использован алгоритм k-means, а оптимальное число кластеров определялось исходя из суммы квадратов расстояний от точек до центроидов кластеров по методу «локтя» и «силуэта». Вся работа по предобработке данных, визуализации и кластеризации была реализована в Python с применением необходимых библиотек: Pandas, Scikit-learn, statsmodels, SciPy, Plotly, Matplotlib, Folium и др.

Обзор литературы и публикаций

Кластерный анализ с использованием методов машинного обучения нашел широкое применение в исследованиях и статьях, посвящённых изучению инновационной активности и социально-экономического развития регионов РФ. Данный подход позволяет определить отличительные черты и выявить ключевые характеристики для регионов внутри образовавшихся кластеров. Полученные результаты могут быть применены на практике – учитываться при разработке программ по развитию регионов. Например, в работе Суховой А.Ф., Голова И.М. [1] авторы приходят к выводу, что понимание различий в инновационной активности регионов и учет их специфики позволяют повысить эффективность реализации социально-экономической политики. В статье была изучена динамика развития инновационного потенциала регионов РФ за последние 10 лет и проанализирован опыт зарубежных стран. В качестве показателей рассматривались затраты на НИОКР, затраты на технологические инновации, количество выданных патентов, число студентов вузов, а также выпуск продукции обрабатывающих производств и ряд других. Отмечается, что многие из показателей имеют большую дисперсию – максимальное значение превышает минимальное в десятки и сотни раз. Это в очередной раз указывает на неоднородность субъектов РФ. Авторы подчеркивают, что природные, географические и культурно-исторические традиции предопределяют архетипы территорий и ее восприимчивость к инновационным импульсам. При этом, чем ниже технический уровень преобладающих производств, тем более инертна территория в плане инновационных преобразований. Методы сравнительного и экономико-статистического анализа позволили авторам выделить 4 главных критерия дифференциации инновационных стратегий. Для проведения расчетов использовались официальные данные Росстата и сведения НИАЦ МИИРИС. Как результат был предложен и обоснован альтернативный подход к решению задач стратегического планирования регионального инновационного развития.

Кластеризации как метод машинного обучения применяется и в исследовании уровня социально-экономического развития регионов РФ [2]. Целью работы ставилось выделение однородных региональных кластеров для разработки платформы принятия правильных форм поддержки субъектов. Отличительная особенность данной работы заключается в использовании относительно большого массива данных. В качестве анализируемого периода были выбраны 2015-2019 гг. В число показателей вошли, например, численность населения, оборот розничной торговли в регионе, соотношение мужчин и женщин, коэффициенты рождаемости и смертности, уровень безработицы, объем инвестиций в основной капитал, численность врачей и другие. Был также осуществлен переход к удельным величинам некоторых показателей и их стандартизация. При помощи метода главных компонент (РСА) из общего числа показателей были выделены 15 индикаторов. С целью проведения кластеризации был выбран алгоритм k-means. Оптимальное число кластеров определено на основе анализа суммарного квадрата расстояний от предполагаемых центров до регионов в кластере. В результате было выделено 5 региональных кластеров. Авторы подчеркивают, что подход с учетом отраслевой структуры экономики регионов может использоваться при реализации кластерно-ориентированной государственной политики с целью поддержки ускоренного развития субъектов.

Помимо уже рассмотренных подходов к определению однородных регионов России по уровню их инновационной деятельности, также используются и эконометрические уравнения для оценки и прогнозирования показателей [3]. В других исследованиях в контексте кластерного анализа субъектов РФ по уровню инновационной активности помимо алгоритма k-средних применяются методы иерархического кластерного анализа (метод одиночной связи («ближайшего соседа»), парной связи («дальнего соседа»), метод Уорда) [4]. Авторы рассматривали данные за 2016 г. по 83 субъектам РФ. Выбранная система индикаторов во многом похожа на те, которые использовались в ранее упомянутых работах. Разные подходы закономерно давали различное итоговое

число кластеров. Так, метод Уорда выделил 7 кластеров, а метод «дальнего соседа» – 8. Однако, на общие выводы и итоговую картину это не повлияло.

Сбор, обработка и визуализация данных

На первом этапе работ планировалось получить общее представление о масштабах и динамике затрат регионов на инновационную деятельность. Такие данные предоставляет Федеральная служба государственной статистики¹. Немного поработав с данными, было обнаружено, что в ряде строк имеются пропущенные значения. Природа их разная, однако для дальнейшей работы хотелось бы по имеющимся в ряду данным восстановить пропущенные значения. Однако классические методы интерполяции, реализованные в библиотеках SciPy и Pandas, не умеют работать с пропусками на границах рядов. По этой причине 2010-2013гг. были заполнены по первому известному значению. Для последних лет был выбран более аккуратный подход: прогнозирование временного ряда с помощью экспоненциальное сглаживание из библиотеки statsmodels.

Теперь можем ранжировать регионы по объемам затрат на инновации. Возьмем 2021г. и посмотрим, кто входит в десятку лидеров. Достаточно закономерно было увидеть здесь Москву, Санкт-Петербург, а также Республику Татарстан, Нижегородскую область и Ханты-Мансийский автономный округ. Но интересно другое: если взглянуть на график валовой добавленной стоимости регионов РФ², то картина окажется довольно схожей – 6 из 10 регионов останутся в первой десятке (*рис. 1*).

¹ Наука, инновации и технологии. Затраты на инновационную деятельность организаций (с 2010 г.)

URL: <https://rosstat.gov.ru/storage/mediabank/Innov-5.xls>

² Национальные счета. ВРП ОКВЭД 2 (с 2016 г.)

URL: https://rosstat.gov.ru/storage/mediabank/VRP_OKVED2_s2016.xlsx

Можем предположить, что между показателями имеется взаимосвязь. Однако куда более правильно было бы сравнивать разные года – ведь отдача от инвестиций в инновации имеет временной лаг! То есть потенциально значимые изменения (за счёт роста объемов затрат на инновации и ряда иных факторов, которые не учитываются нами в анализе) в валовых добавленных стоимостях по регионам должны быть видны на горизонте 3–5 лет.

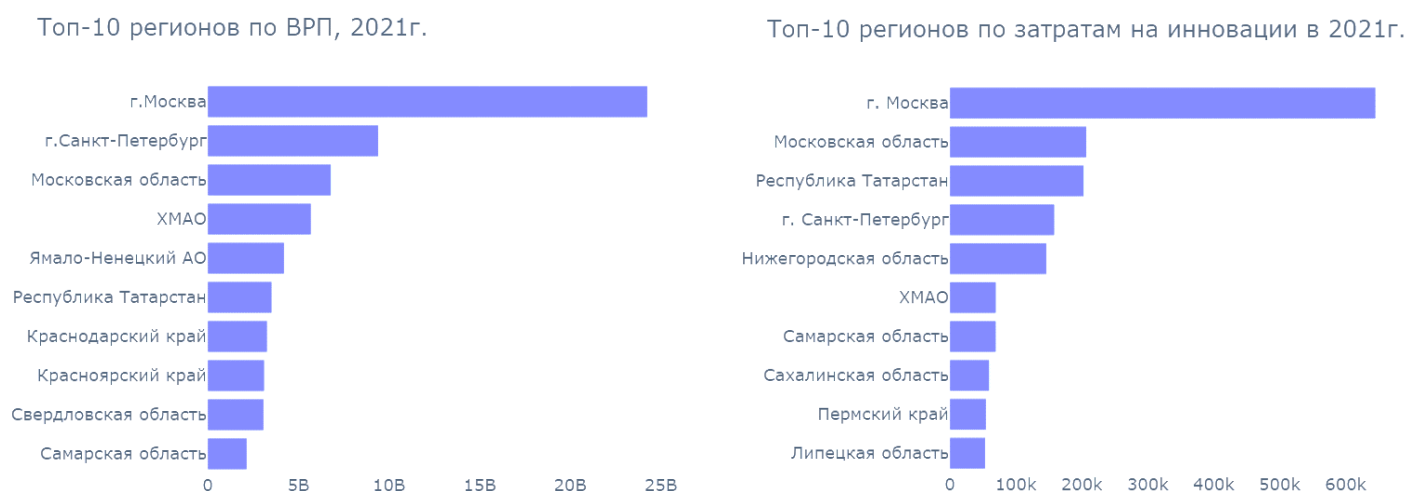


Рис. 1. Топ-10 регионов по ВРП и по затратам на инновации за 2021г.

Возьмем некоторые регионы из тех, что вошли в топ-10 по ВРП в 2021г. и посмотрим на динамику затрат на инновационную деятельность (рис. 2). По графику видим, что данная пятерка регионов из года в год старалась увеличивать объемы инвестиций на инновации. Возможно, именно это и позволило им остаться на лидирующих позициях спустя годы. Также отмечу, что топ-30 регионов по доли в валовой добавленной стоимости практически не изменился за десятилетие, что осложняет сравнительный анализ в динамике.

Динамика затрат на инновации 2010-2022

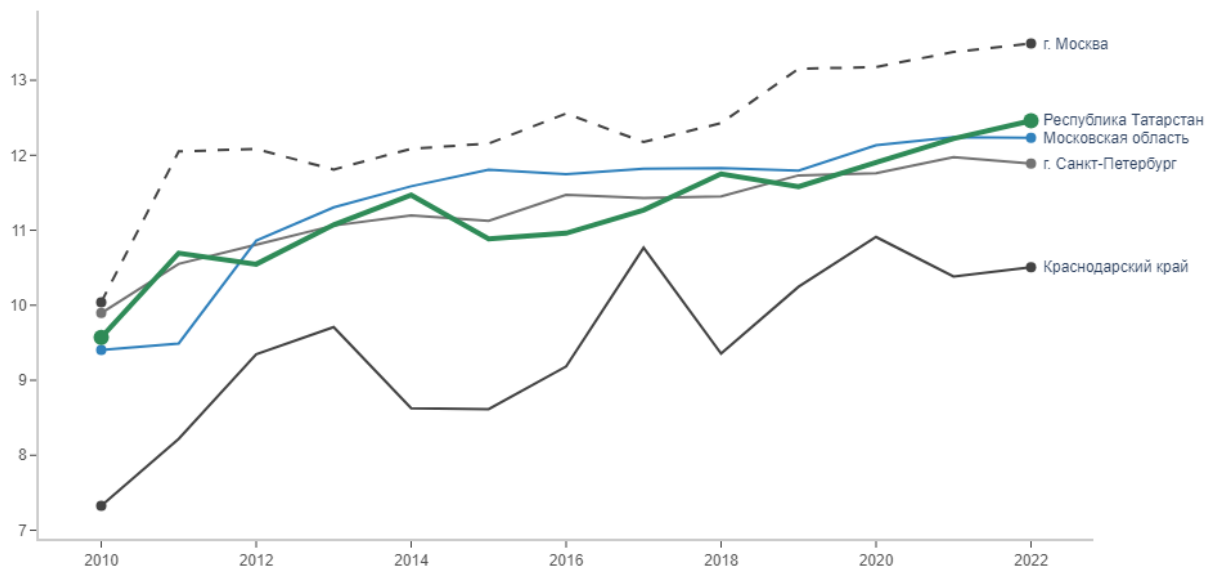


Рис. 2. Динамика затрат регионов на инновации, 2010-2022г.

Для дальнейшего анализа нам потребуется информация по затратам регионов на инновации в разрезе статей расходов по годам. Потому хочется автоматизировать *preprocessing* данных. Для этого вынес соответствующие функции в отдельный модуль: функция *prepare_data* позволяет привести каждый из листов книги Excel к нужному виду, оставляя лишь релевантные строки, осуществляет проверку числа регионов (85), а функция *rename_columns*, как следует из названия, придает колонкам единообразный формат – всё дело в том, что наименования столбцов содержат большое число символов, что визуально излишне нагружает датасет. Более того, в разные годы названия столбцов немного варьируются. С этим справились достаточно простые регулярные выражения, преобразовав огромные предложения в лаконичные словосочетания, в полной мере отражающие суть. Как итог получим аккуратную таблицу с данными (рис. 3)

	region	Исследование / разработка	Дизайн	Машины и оборудование	Обучение персонала	Прочее	Инжиниринг	Маркетинг	Приобретение ПО	Патенты / лицензии
0	Белгородская область	3703.01	4.20	20504.65	11.09	2462.86	343.45	8.80	166.32	6.95
1	Брянская область	1034.44	0.00	719.02	0.00	106.84	364.95	3.15	5.81	6.18
2	Владимирская область	5360.82	0.00	1421.33	68.00	156.17	228.57	7.31	169.79	276.76
3	Воронежская область	8013.89	0.00	10948.62	4.51	190.43	264.18	1.82	211.47	15.93
4	Ивановская область	245.44	0.00	367.07	0.00	9.72	0.00	0.00	26.66	1.34
...
80	Амурская область	688.46	0.00	1649.23	9.77	350.41	0.00	0.00	8.11	0.62
81	Магаданская область	0.00	0.00	0.00	2.59	81.63	0.00	0.00	0.00	0.68
82	Сахалинская область	113.45	0.00	49570.35	3.10	94.72	0.00	0.00	0.00	0.00
83	Еврейская автономная область	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
84	Чукотский автономный округ	0.00	0.00	781.26	0.00	0.00	0.00	0.00	0.00	0.00

Рис. 3. Подготовленный массив данных

В качестве признаков я взял 9 из 10 показателей, по которым были данные Росстата:

1. Исследование / разработка (затраты на исследования и разработку новых продуктов, услуг и методов их производства, новых производственных процессов)
2. Дизайн (деятельность по изменению формы, внешнего вида или удобства использования продуктов или услуг)
3. Машины и оборудование (приобретение машин, оборудования, прочих основных средств, связанных с инновационной деятельностью)
4. Обучение персонала (обучение и подготовка персонала, связанные с инновационной деятельностью)
5. Инжиниринг (включает подготовку технико-экономических обоснований, производственное проектирование и конструкторскую проработку объектов техники и технологий на стадии внедрения инноваций)
6. Маркетинг (маркетинг и создание бренда)
7. Приобретение ПО (разработка и приобретение программ для ЭВМ и баз данных, связанных с инновационной деятельностью)
8. Патенты (приобретение прав на патенты (отчуждение), лицензий на использование изобретений, промышленных образцов, полезных моделей, патентование (регистрация) результатов интеллектуальной деятельности)
9. Прочее (прочие затраты, связанные с осуществлением инновационной деятельности)

Таким образом, вся дальнейшая работа будет построена вокруг данных признаков. Посмотрим на некоторые агрегирующие статистики (рис. 4). Мы можем заметить, насколько сильный разброс внутри показателей по регионам. Это вновь отражает тот факт, как сильно отличаются регионы друг от друга из-за разных масштабов экономик, территорий.

	Исследование / разработка	Дизайн	Машины и оборудование	Обучение персонала
Ср. знач.	12889.97	69.56	11722.96	40.21
Ст. откл.	45079.54	310.23	27648.70	193.46
Медиана	1145.00	0.00	2339.29	4.42
Минимум	0.00	0.00	0.00	0.00
Максимум	375733.59	2324.07	190156.98	1746.09

Рис. 4. Статистики по выборочным показателям, 2022г

Кластерный анализ регионов

Алгоритм k-means и метод локтя

Одним из наиболее распространенных методов машинного обучения в задачах кластеризации является алгоритм k-means. Вспомним как он работает в общих чертах (рис. 5).

Algorithm 1 k-means algorithm

- 1: Specify the number k of clusters to assign.
 - 2: Randomly initialize k centroids.
 - 3: **repeat**
 - 4: **expectation:** Assign each point to its closest centroid.
 - 5: **maximization:** Compute the new centroid (mean) of each cluster.
 - 6: **until** The centroid positions do not change.
-

Рис. 5. Алгоритм k-means

Существенным недостатком k-means является случайная инициализация центров кластеров, что может привести к разным результатам кластеризации и усложняет воспроизводимость экспериментов. Главным же достоинством алгоритма является его скорость. На каждой итерации требуется пересчет только расстояний до центров кластеров. Как и большинство алгоритмов кластеризации k-means

зависит от заданного числа кластеров, т.е. значение данного параметра можно подбирать. Однако в случае кластеризации сложнее выбрать критерий, с помощью которого можно было бы записать постановку оптимизационной задачи. Один из возможных критериев – сумма квадратов расстояний от точек до центроидов кластеров, к которым они относятся:

$$J(c) = \sum_{k=1}^K \sum_{i \in C_k} \|x_i - \mu_k\|^2 \rightarrow \min C$$

здесь C – множество кластеров мощности K , а μ_k – центроид кластера C_k

Потому для выбора числа кластеров часто прибегают к такой эвристике: выбирают то число кластеров, начиная с которого описанный функционал $J(C)$ падает "уже не так быстро". Более формально запишем так:

$$Q(k) = \frac{|J(C_k) - J(C_{k+1})|}{|J(C_{k-1}) - J(C_k)|} \rightarrow \min_k$$

При этом не всегда очевидно, какое оптимальное число стоит выбрать по графику.

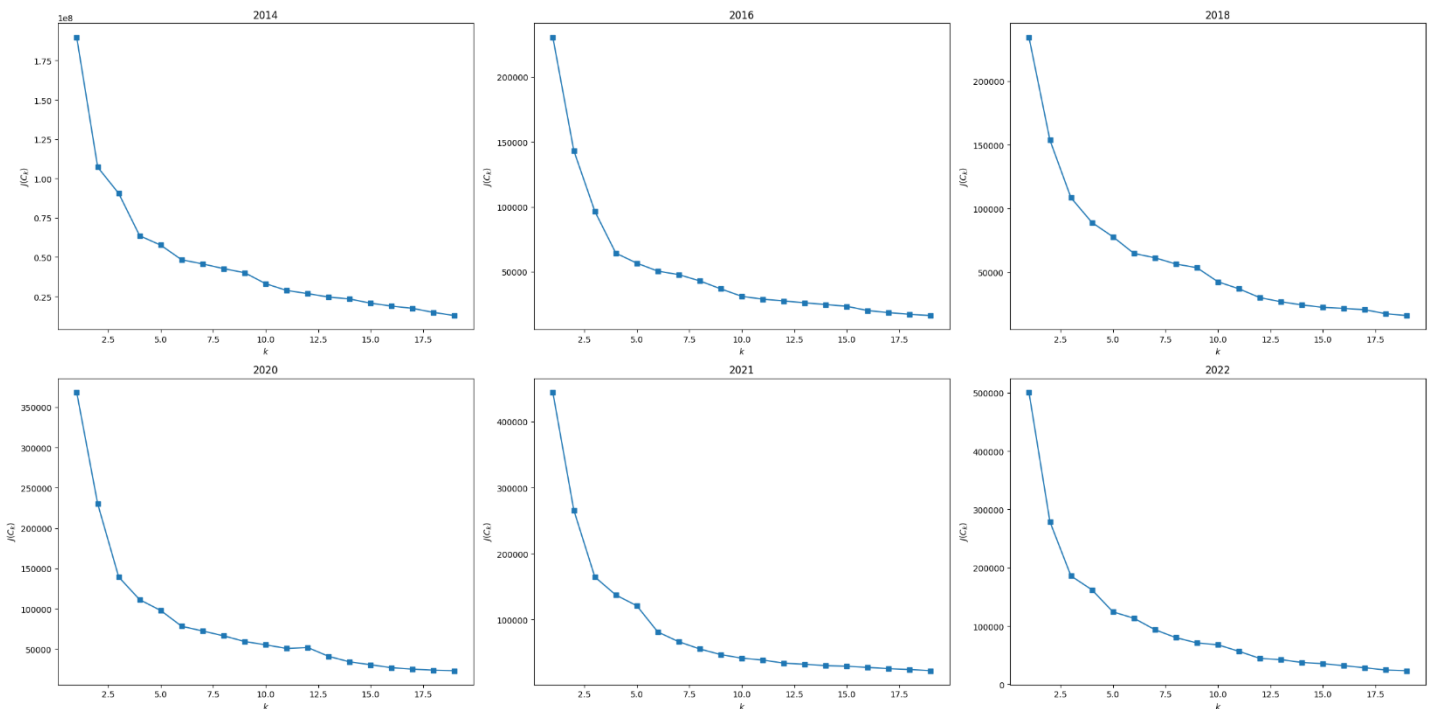


Рис.6. Оценка числа кластеров по методу локтя на ненормированных затратах, 2014-2022гг.

Если запустить алгоритм на ненормированных значениях общих затрат, то по графикам 2014г., 2021г. и 2022г. однозначно сложно сказать, какое число кластеров задать. В таком случае посмотрим на коэффициент «силуэта», который также может использоваться как метрика для оценки количества кластеров. Помимо упомянутых, применяется и байесовский информационный критерий (BIC) с моделью смеси нормальных распределений.

Кластеризация регионов с учетом объемов выпуска

Если производить кластеризацию регионов по объемам инвестиций в инновации AS-IS, при этом никак не нормировать затраты, то мы не учитываем множество других факторов, которые в той или иной степени оказывают влияние на суммарные инвестиции в инновации: численность населения, площадь, наличие/отсутствие градообразующих предприятий, масштабных комплексов по переработке сырья, производств по добыче полезных ископаемых на территории региона (т.е. сама структура экономики региона играет роль). Потому скорректируем данные на величину экономик – для этого необходимо нормировать затраты на объем отгруженной инновационной продукции. Такую информацию нам также предоставляет Росстат³.

Посмотрим на кластеры, которые будут выделены после корректировки суммарных затрат на общий объем выпуска инноваций. Начнем с 2014 года, а затем проследим динамику вплоть до 2022г., тем самым проанализировав устойчивость кластеризации. Для определения оптимального числа кластеров воспользуемся методом локтя и коэффициентом силуэта (*рис. 7–9*).

³ Наука, инновации и технологии. Объем инновационных товаров, работ, услуг (с 2010 г.)
URL: <https://rosstat.gov.ru/storage/mediabank/Innov-3.xls>

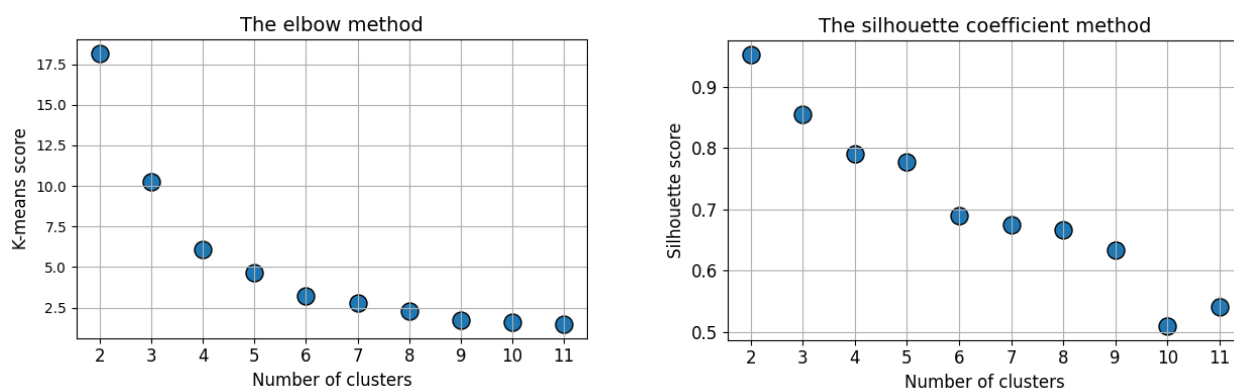


Рис. 7. Выбор оптимального числа кластеров по методу локтя и силуэта (2014г.)

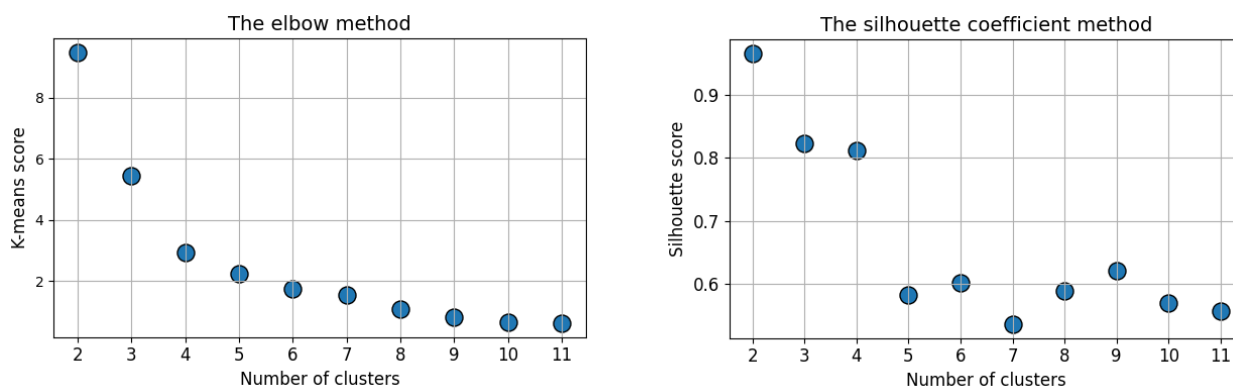


Рис. 8. Выбор оптимального числа кластеров по методу локтя и силуэта (2019г.)

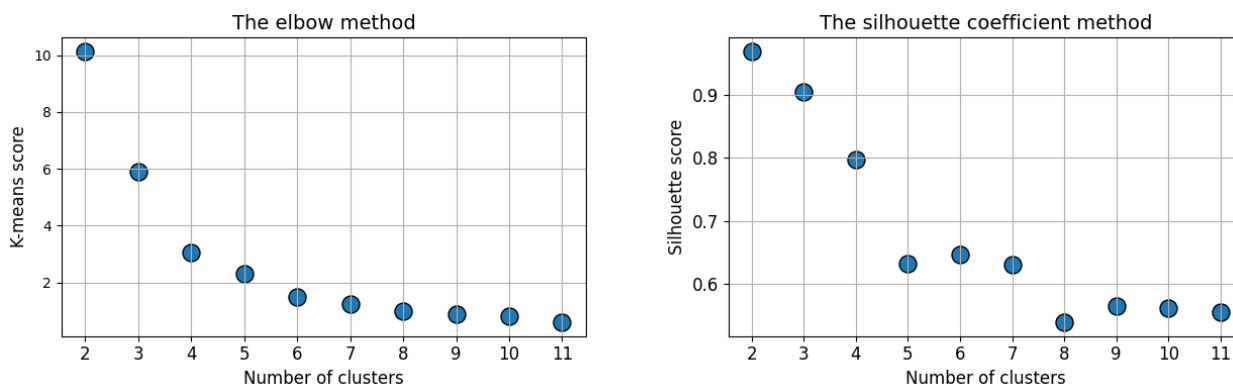


Рис. 9. Выбор оптимального числа кластеров по методу локтя и силуэта (2022г.)

В целом динамика от года к году похожа. По методу локтя разумно выбирать 4 кластера, однако по графику метода силуэта пик достигается при числе кластеров равным 2. Но два кластера — слишком мало, такое разбиение будет не

очень информативно и не интерпретируемо, получатся слишком большие группы, всё же у нас 85 регионов.

Распределение внутри кластеров не очень радует – на первые 2 кластера приходится 3 региона, на другие 2 – 82 (рис. 10)

<i>Кластер</i>	<i>Регионы</i>	<i>Описание</i>
1)	Республика Алтай	Республика Алтай уникальна тем, что в 2014г. коэффициент, отражающий отношение инвестиций в инновации к их производству, превысил 45!
2)	Республика Ингушетия, Республика Северная Осетия – Алания	Эти два региона также имеют сильное превышение инвестиций в инновации (в 12 и 9 раз соответственно), нежели их выпуск, характерный признак отстающих и развивающихся регионов.
3)	Ненецкий автономный округ, Ханты-Мансийский АО, Ямало-Ненецкий АО, Республика Калмыкия, Волгоградская область, Республика Дагестан, Карачаево-Черкесская Республика, Республика Тыва, Приморский край, Магаданская область	Для данных десяти регионов характерно небольшое кратное (~2.3-4.5) превышение инновационных затрат над объемом производимых инновационных товаров и оказанных услуг / работ. В эту группу попали как сильные в плане добывающих производств регионы (Ненецкий автономный округ, Ханты-Мансийский АО, Ямало-Ненецкий АО), так и отстающие регионы – Республика Тыва, Дагестан, Карачаево-Черкесская Республика – имеют одни из наиболее низких показателей ВРП на душу населения ⁴ .
4)	Все остальные регионы	В последний кластер попали регионы, у которых годовые объемы отгруженных инновационных товаров и оказанных услуг больше или сопоставим с затратами на инновации. Средний коэффициент составил 0.49, а медианное значение равно 0.32.

Рис. 10. Результат кластеризации методом *k-means* (2014г.)

Теперь сравним результаты 2014г. с кластерами, которые выделил алгоритм *k-means* для 2019г (рис. 11). Заметим, что распределение внутри кластеров поменялось, однако общий паттерн сохранился – всё также присутствует один

⁴ Национальные счета. ВРП ОКВЭД 2007 (с 2004 г.)

URL: https://rosstat.gov.ru/storage/mediabank/VRP_OKVED2007.xlsx

регион с очень большим значением коэффициента (Республика Тыва, затраты на инновации в 37.4 раза превышают объем отгруженных товаров собственного производства)

<i>Кластер</i>	<i>Регионы</i>	<i>Описание</i>
1)	Республика Тыва	Ранее (2014г.) Республика Тыва входила в число регионов, в которых затраты на инновации превышали производство не более чем в 4-5 раз. По прошествии лет объем инвестицийкратно возрос (как в абсолютном, так и в относительном значении).
2)	Сахалинская область	В прошлый раз соответствующий коэффициент для региона составил 0.1 при немалых затратах на инновации. Спустя 5 лет затраты только возросли, а вот объем отгруженных товаров упал в десятки раз.
3)	Ненецкий автономный округ, Республика Крым, Калининградская область, Астраханская область, Иркутская область, Республика Бурятия, Забайкальский край, Еврейская автономная область	Значительные изменения произошли в данной категории – число регионов, в которых затраты на инновации превышают выпуск от 2 до 4-х раз, сократилось с 10 до 8. При этом только Ненецкий автономный округ вновь попал в эту группу, остальные 9 распределились в другие кластеры.
4)	Все остальные регионы	В последний кластер вновь попали регионы, у которых годовые объемы отгруженных инновационных товаров и оказанных услуг больше или сопоставим с затратами на инновации. Средний коэффициент снизился до 0.37, а медианное - 0.31.

Рис. 11. Результат кластеризации методом k-means (2019г.)

Проделав кластеризацию на данных по регионам за 2022г., можно отметить следующую тенденцию – снижение числа объектов во всех кластерах кроме наибольшего. Это можно интерпретировать следующим образом: всё больше субъектов выходят на уровень, когда объемы собственного инновационного производства превалируют над инвестициями в него. Так, в последнюю

категорию было отнесено 77 регионов из 85, по одному региону в два других кластера и 6 субъектов – в оставшийся.

<i>Кластер</i>	<i>Регионы</i>	<i>Описание</i>
1)	Сахалинская область	Тремя годами ранее Сахалинская область уже была вынесена в отдельный кластер, т.к. имела кратное превышение инвестиций в инновации над самим выпуском. Однако в 2022г. коэффициент для региона составил 39.3, что говорит о разнице на порядок, при этом сами инвестиции в абсолютном выражении снизились, сильно просел выпуск.
2)	Иркутская область	В 2022г. регион увеличил объемы инвестиций в несколько раз, ввиду чего Иркутская область и переместилась из наибольшего кластера в отдельный, соотношение затрат к выпуску стало > 8
3)	Костромская область, Ханты-Мансийский АО, Томская область, Республика Саха (Якутия), Амурская область, Чукотский автономный округ	В данный кластер вернулся ХМАО, а также его дополнили новые регионы, которые ранее в нем не оказывались. Для всех них коэффициент лежит в диапазоне 1.85 – 3.51
4)	Все остальные регионы	В последнем кластере без сильных изменений: средний коэффициент снизился до 0.34, а медианное значение составило 0.26.

Рис. 12. Результат кластеризации методом k-means (2022г.)

Таким образом, кластеризация оказалась слабо устойчивой. От года к году регионы в небольшом количестве переходили между группами, однако общая картина сохранялась. Отмечу, что для последнего кластера характерно снижение среднего значения коэффициента, что может свидетельствовать об отдаче инвестиций в инновации – в силу слабой эластичности объема производимых инновационных товаров необходимы годы, чтобы результат от затрат стал заметен.

Кластеризация регионов в разрезе статей затрат

Теперь проанализируем сами затраты, вернее, их относительные величины – в долях от общих инвестиций в инновации. Такой подход позволит выделить кластеры регионов, ориентированных на определенные отрасли. Для этого вновь определим оптимальное число кластеров (рис. 13-15)

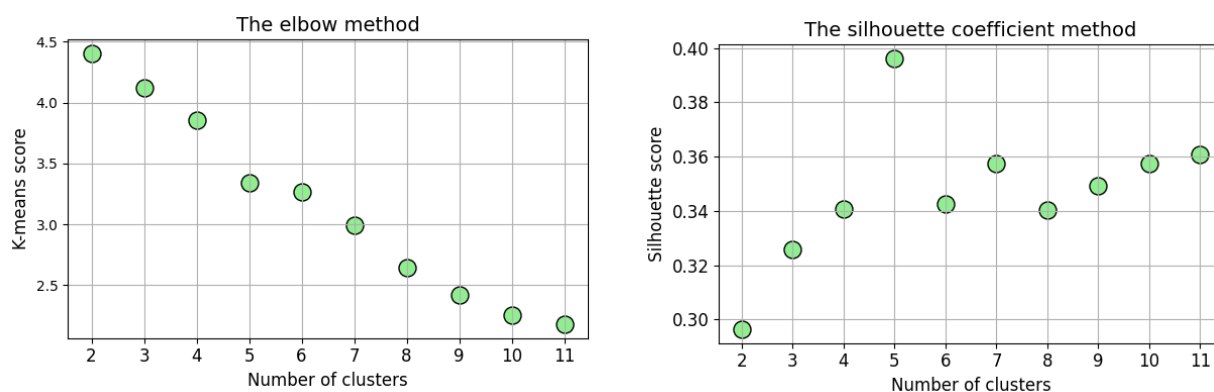


Рис. 13. Выбор оптимального числа кластеров по методу локтя и силуэта (2014г.)

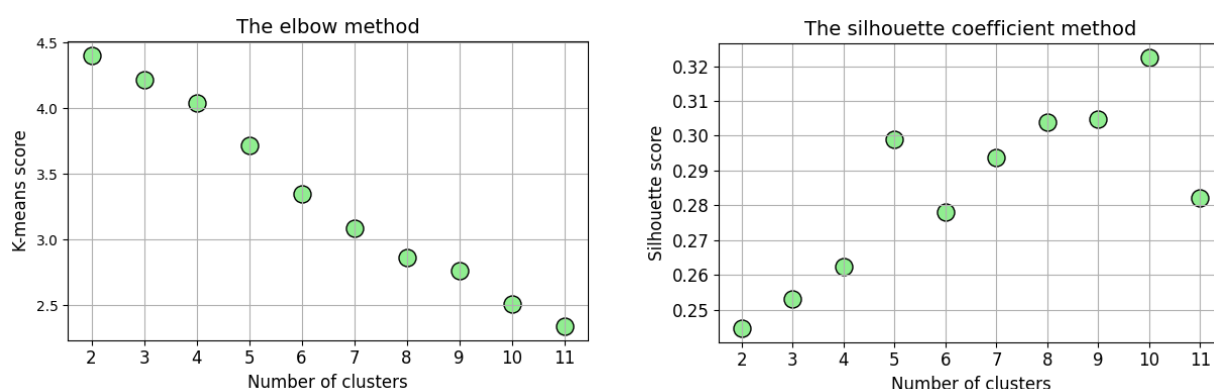


Рис. 14. Выбор оптимального числа кластеров по методу локтя и силуэта (2019г.)

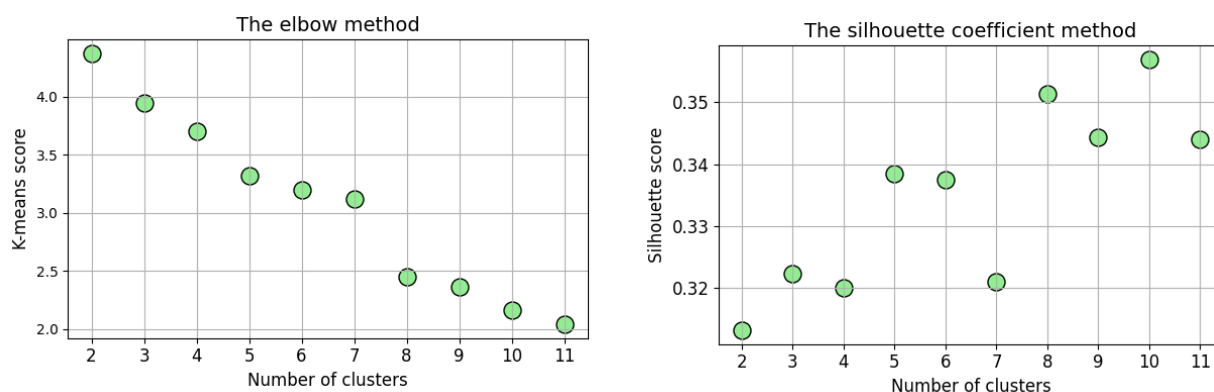


Рис. 15. Выбор оптимального числа кластеров по методу локтя и силуэта (2022г.)

Сразу заметим, что с годами меняется и оптимальное число кластеров по методу силуэта. Для анализа структуры затрат 2014г. выберем число кластеров равное 5 – именно на этой отметке наблюдается пиковое значение по методу силуэта. В этот раз регионы распределились по кластерам более равномерно (*рис. 16*). Тем не менее, всё равно присутствует 2 преобладающих группы.

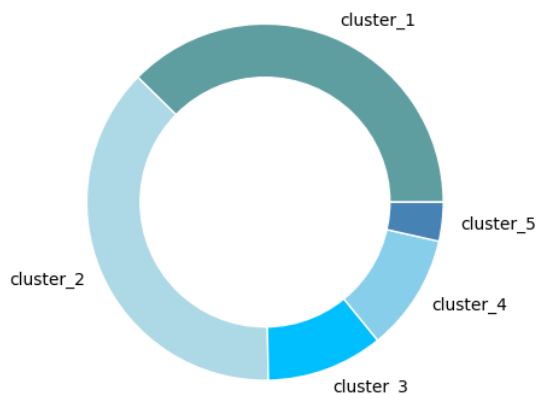


Рис. 16. Удельный вес каждого из кластеров (2014г.)

В контексте изучения инновационных затрат нужно понимать, как они коррелируют с экономико-географической спецификой региона. Для этого будет удобно визуализировать карту России, на которой каждый регион перекрашен в соответствующий цвет. С помощью библиотек Folium и geopandas перенесем полученные результаты кластеризации данных за 2014 год на карту (*рис. 17*).

<i>Кластер</i>	<i>Регионы</i>	<i>Описание</i>
Красный	Владимирская область, Воронежская область, Ивановская область, Калужская область, Московская область, Смоленская область, Тверская область, Тульская область, г. Москва, Республика Карелия, Ненецкий автономный округ, Калининградская область, Мурманская область, Новгородская область, г. Санкт-Петербург, Астраханская область, Ростовская область, Республика Татарстан, Чувашская Республика, Кировская область, Нижегородская область, Самарская область, Курганская область, Ямало-Ненецкий АО, Тюменская область, Алтайский край, Красноярский край, Новосибирская область, Томская область, Камчатский край, Амурская область, Еврейская автономная область	Наибольший процент инвестиций приходится на категорию «Исследование / разработка». В среднем по кластеру он составляет ~55%. Также около четверти (27%) расходов идет на «Машины и оборудование». Наименьший вес имеют статьи «Обучение персонала» и «Маркетинг»

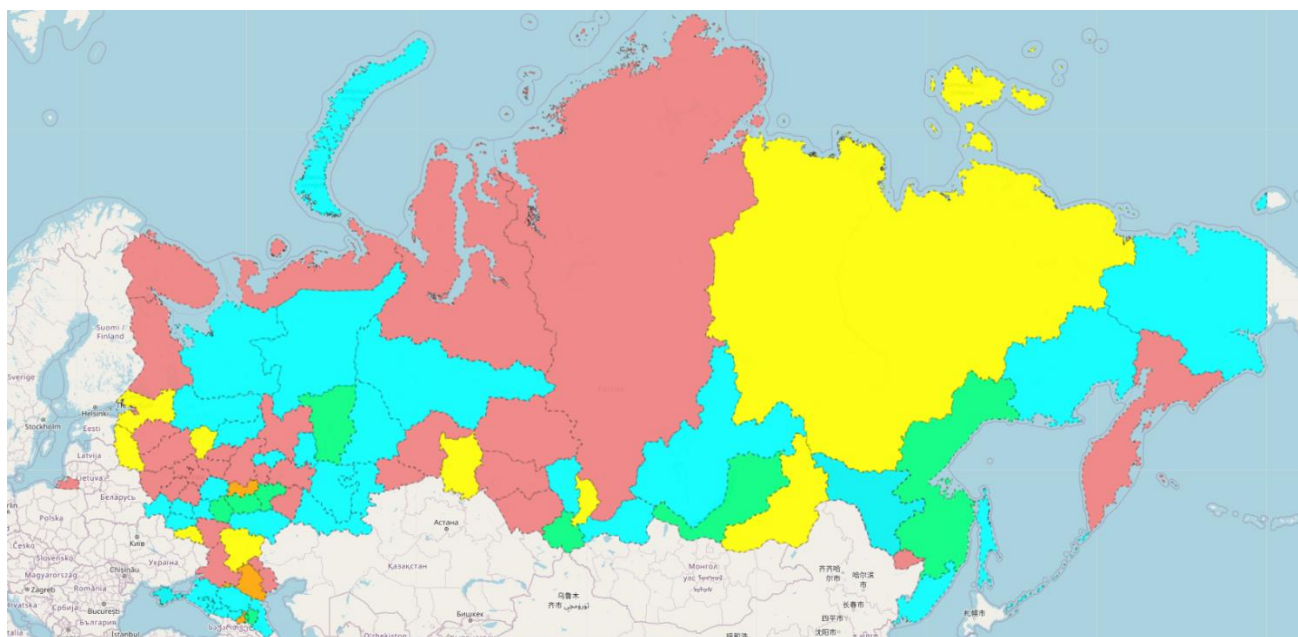


Рис. 17. Регионы в цвет кластеров (2014г.)

Кластер	Регионы	Описание
Бирюзовый	Брянская область, Костромская область, Курская область, Липецкая область, Орловская область, Рязанская область, Республика Коми, Архангельская область, Вологодская область, Республика Адыгея, Краснодарский край, Республика Дагестан, Республика Ингушетия, Кабардино-Балкарская Республика, Карачаево-Черкесская Республика, Ставропольский край, Республика Башкортостан, Республика Марий Эл, Удмуртская Республика, Оренбургская область, Саратовская область, Свердловская область, Ханты-Мансийский АО, Челябинская область, Республика Тыва, Иркутская область, Кемеровская область, Приморский край, Магаданская область, Сахалинская область, Чукотский автономный округ, Республика Крым	В структуре затрат регионов лидирует категория «Машины и оборудование», составляет в среднем более 76% от общих расходов регионов на инновации, что является наибольшим значением по всем кластерам.
Оранжевый	Республика Калмыкия, Республика Северная Осетия – Алания, Республика Мордовия	Больше половины расходов регионов идут на статью «Инжиниринг»
Зеленый	Тамбовская область, Чеченская Республика, Пермский край, Пензенская область, Ульяновская область, Республика Алтай, Республика Бурятия, Хабаровский край, г. Севастополь	Регионы стараются равномерно инвестировать во все сферы жизни общества: по 28% приходится на «Дизайн» и «Машины и оборудование», около 15% на «Исследования / разработки»

Желтый	Белгородская область, Ярославская область, Ленинградская область, Псковская область, Волгоградская область, Республика Хакасия, Забайкальский край, Омская область Республика Саха (Якутия)	Около половины затрат на инновации регионов попадает в категорию «Прочее», а также примерно треть составили расходы на «Машины и оборудование»
--------	--	--

Нужно отметить, что ряд субъектов по тем или иным причинам могут выделяться на фоне остальных в своем кластере. Так, например, для Кировской, Курганской, Омской областей достаточно велика доля затрат на инжиниринг, сильно превышая среднее значение по данному показателю в кластере. Полагаю, это связано с наличием крупных предприятий ВПК на территории регионов (Курганмашзавод, Омсктрансмаш, Маяк). В целом результат кластеризации получился неожиданным и слабо интерпретируемым.

Было принято решение не менять число кластеров при работе с данными 2022г. и проанализировать изменения в ранее сформировавшихся групп. Алгоритм k-means разнес регионы по кластерам еще более равномерно – по 12, 25, 24, 22 и 2 в каждый из пяти кластеров соответственно (рис. 18).

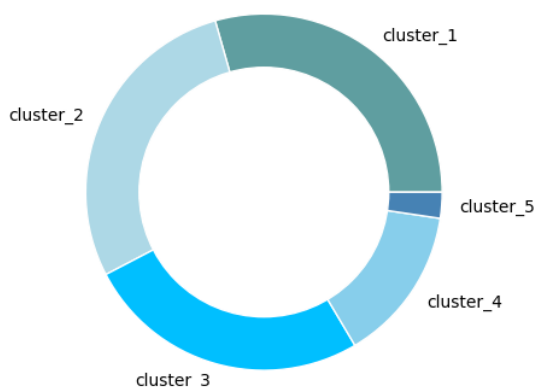


Рис. 18. Удельный вес каждого из кластеров (2022г.)

На географической карте также заметны изменения – отчетливо бросаются в глаза крупные регионы, изменившие цвет (рис. 19)

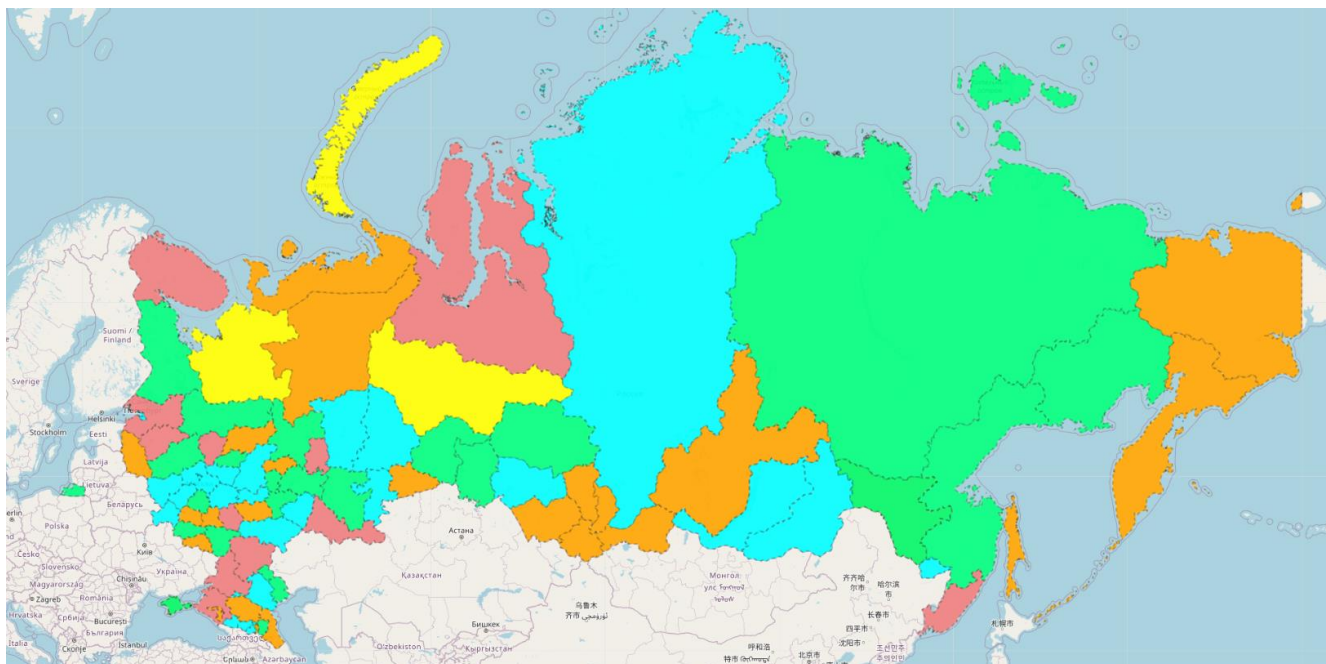


Рис. 19. Регионы в цвет кластеров (2022г.)

Посмотрим чуть детальнее внутри каждого из кластеров. Регионов красного цвета стало не просто меньше, изменились показатели внутри кластера – теперь он состоит из регионов, которые помимо «Машин и оборудования» развивают «Инжиниринг» и вкладываются в «Исследования и разработку». Похожая трансформация произошла и со вторым кластером (бирюзовый цвет). На лидирующую позицию вышли инвестиции в «Исследования и разработку» – в среднем около 63% от общего числа затрат на инновации. Третий (оранжевый) кластер теперь полностью ориентирован на инвестиции в инновационное развитие в сфере «Машины и оборудование» - доля расходов на нее превысила 0.8. Четвертый кластер (зеленый) объединил в себе 22 региона, которые в равной степени вкладываются в «Исследования и разработку» и «Машины и оборудование». В последний кластер (жёлтый) попали субъекты, которые значительную долю (>0.4) инвестируют в «Приобретение ПО».

Таким образом, в ходе работы были опробованы разные методы и подходы к решению задачи кластеризации регионов по уровню инновационного развития.

Полученные результаты в очередной раз доказывают, как разнообразна и неоднородна структура субъектов РФ. Подобные исследования помогают лучше разобраться в особенностях каждого из них, глубже их изучить. Это создает фундамент для выработки дальнейшего плана по улучшению экономических мер и политик, направленных на поддержку и развитие регионов.

Список литературы

1. Суховой А. Ф., Голова И. М. Дифференциация стратегий инновационного развития регионов как условие повышения эффективности социально-экономической политики в РФ // Экономика региона. 2020. Т. 16, вып. 4. С. 1302-1317.
2. 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, прогноз. 2021. Т. 14. № 6. С. 70–85.
3. Деркаченко О.В. Кластеризация регионов России по уровню инновационного развития и построение системы эконометрических уравнений // Российский экономический вестник 2020, Том 3, №2 С. 36-39
4. Шамрай-Курбатова Л. В., Леденева М. В. Кластерный анализ субъектов РФ по уровню инновационной активности // Бизнес. Образование. Право. 2021. № 1 (54). С. 88—97.