

Оглавление

Предисловие и входные данные	2
Само исследование	4
Вывод	15
Об авторе.....	16
Приложение 1 – проверка гипотез на нормальное распределение совокупностей.....	17
Приложение 2 – центрировано-нормированные значения показателей по странам.....	18
Приложение 3 – распределение стран по кластерам и исходные значения признаков на момент исследования.....	23
Приложение 4 – проверка гипотез на нормальное распределение совокупностей по кластерам	29
Приложение 5 – корреляционный анализ по однородным кластерам	31

Предисловие и входные данные

Я решил вспомнить, как анализировать данные в Python.

Статистика COVID-19, пожалуй, самая хайповая тема последних полтора лет, поэтому решил исследовать её.

Данные для исследования собраны из открытых источников.

Основной источник данных - Our World in Data. Они, в свою очередь, опираются на данные Университета Хопкинса. Ссылка на набор данных:

<https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/latest/owid-covid-latest.csv>

Дата проведения исследования и актуальности данных – 19 января 2022 года.

Задачи исследования

1. Сравнить текущую ситуацию с COVID-19 в разных странах мира
2. Распределить страны по кластерам (*группам похожих между собой объектов*)
3. Выяснить, какие факторы влияют на смертность и заболеваемость COVID-19.

Данное исследование не имеет задачи проверить данные тех или иных стран на корректность.

Из всего массива доступных показателей для исследования отобраны следующие:

- число зарегистрированных случаев заболевания COVID-19 на 1 миллион населения за всё время пандемии
- число новых случаев заболевания COVID-19 в сутки на 1 миллион населения (*со сглаживанием¹ за последние 7 дней*)
- число смертей от COVID-19 на 1 миллион населения за всё время пандемии
- число новых смертей от COVID-19 на миллион населения в сутки (*со сглаживанием за последние 7 дней*)
- коэффициент распространения COVID-19 (*оценка на последних данных*) – сколько людей успеет заразить 1 больной.
- количество полностью вакцинированных людей на 100 человек населения
- количество людей, получивших первую дозу вакцины, на 100 человек населения
- доля населения с возрастом более 65 лет
- количество больничных коек на 1000 населения
- индекс человеческого развития как показатель общего уровня жизни и здравоохранения в стране

Удалены все страны, по которым недостаёт каких-либо данных. Из исходного массива в 217 стран остаётся только 132.

¹ Для сглаживания, как правило, используется метод «скользящей средней», когда вместо самого значения подставляется среднее за последние X дней (в данном случае – 7 дней). Этот метод используется для сглаживания разовых скачков и аномалий в данных, а также более наглядной иллюстрации долгосрочных тенденций развития рассматриваемого процесса.

Из исследования были удалён ряд показателей, которые могли бы существенно обогатить исследование. Причины удаления этого каждого из показателей приведены в таблице 1.

Таблица 1 – причины удаления показателей из исследования

Показатель	Причина удаления
Оценка избыточной смертности за всё время пандемии	Актуальные данные за последний месяц имеются только по 20 странам, всего хоть за какой период имеются всего по 108 странам
Число новых госпитализаций	Менее 50 стран с доступными данными, даже без проверки на актуальность
Число занятых койкомест в больницах	Менее 50 стран с доступными данными, даже без проверки на актуальность
Число новых пациентов в палатах реанимации и интенсивной терапии	Менее 50 стран с доступными данными, даже без проверки на актуальность
Число занятых койкомест в палатах реанимации и интенсивной терапии	Менее 50 стран с доступными данными, даже без проверки на актуальность
Общее число проведённых тестов на COVID-19 на тысячу человек населения за всё время пандемии	<ol style="list-style-type: none"> 1. Менее 120 стран с данными актуальностью в 1 месяц 2. Отсутствие единой методологии показателя. Одни страны публикуют число проведённых тестов. Другие – число протестированных образцов (<i>в спорных ситуациях один образец может быть протестирован больше одного раза</i>). Третьи – число протестированных людей (<i>у одного человека можно взять более одного образца</i>). По четвёртым нет никакой информации о единицах измерения. Сравнивать такие показатели между собой некорректно.
Число проведённых тестов на COVID-19 за последние сутки на тысячу населения (7-дневное сглаживание)	
Доля положительных тестов в общем числе сданных	
Доля населения старше 70 лет	Заведомо высокая корреляция с долей населения старше 65 лет. Чтобы не плодить мультиколлинеарность ² , убрал эти показатели из рассмотрения.
Медианный возраст	
Численность и плотность населения	<ol style="list-style-type: none"> 1. Численность не показательна в числу разной территории стран 2. Плотность населения не показательна в силу неравномерного распределения населения по территории страны (<i>в качестве примеров – Россия, Китай, Гренландия</i>)
Вакцинация бустерными дозами вакцины на 100 человек населения	Актуальные данные доступны менее чем по 100 странам.

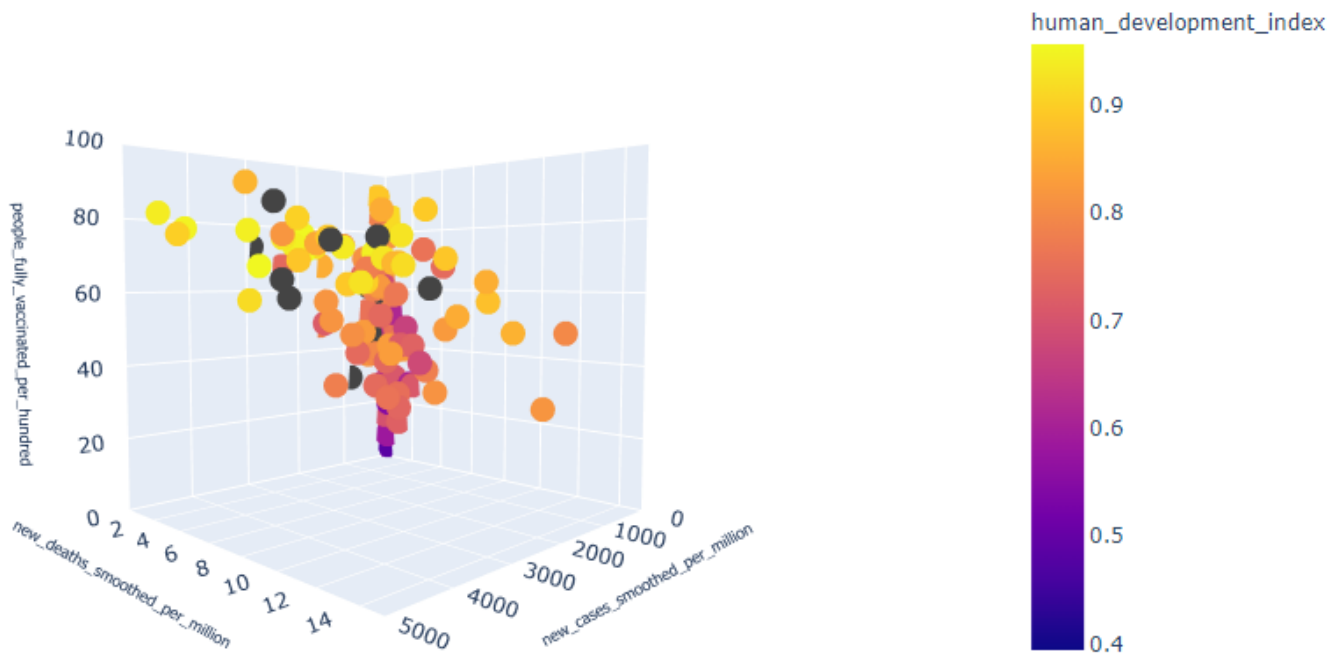
² Мультиколлинеарность – это явление сильной взаимозависимости между рассматриваемыми показателями. Для ряда математических методов анализа данных это нарушает критичные допущений, которые используются при обосновании применения тех или иных инструментов решения поставленных задач. Применение таких методов на наборах данных с мультиколлинеарностью приводит к некорректным статистическим выводам.

Само исследование

Рассмотрим распределение данных графически в разрезе нескольких признаков. Выберем для примера число новых случаев на 1 млн. населения, новые смерти на 1 млн. населения и долю полностью вакцинированного населения.

В качестве четвёртого измерения добавим индекс человеческого развития – его покажем цветом объектов на трёхмерном графике.

Рисунок 1 – распределение стран в выбранной системе координат



В общем случае для того, чтобы проверить взаимосвязи между признаками, нужно построить матрицы парных и частных корреляций.

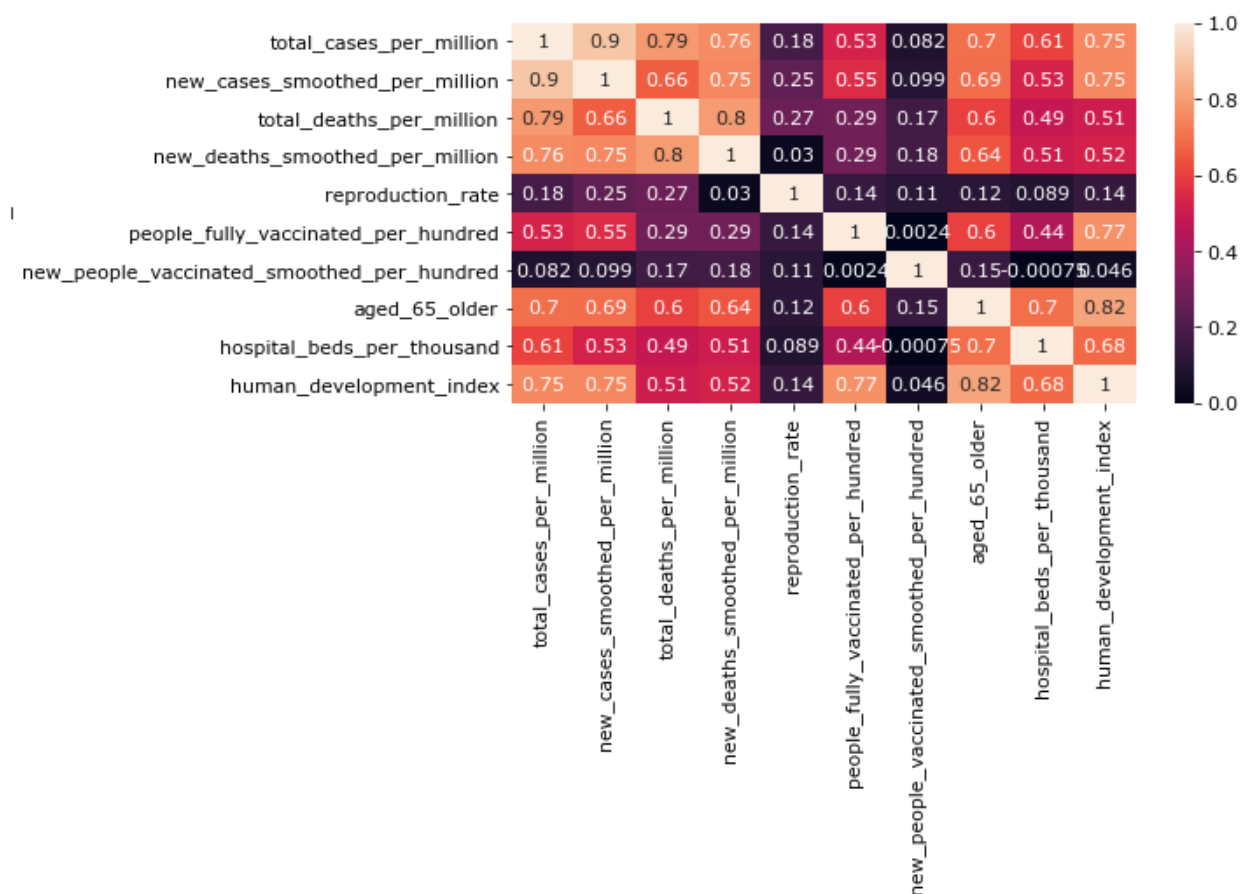
Судя по графику выше, мы имеем несколько разнородных групп объектов, для каждой из которых действуют свои взаимосвязи. Также наблюдаем, что страны

Но прежде нужно проверить распределения параметров нашего массива данных на нормальность. Если распределение хотя бы одного из параметров будет отличаться от нормального (нулевая гипотеза может быть отклонена) – параметрические коэффициенты корреляции строить нельзя.

Проверка на нормальность распределения показана в приложении 1.

Распределение большинства признаков отлично от нормального, поэтому для реального исследования взаимосвязей нужно использовать непараметрические коэффициенты корреляции. Для этого исследования я выбрал метод ранговых корреляций Спирмана.

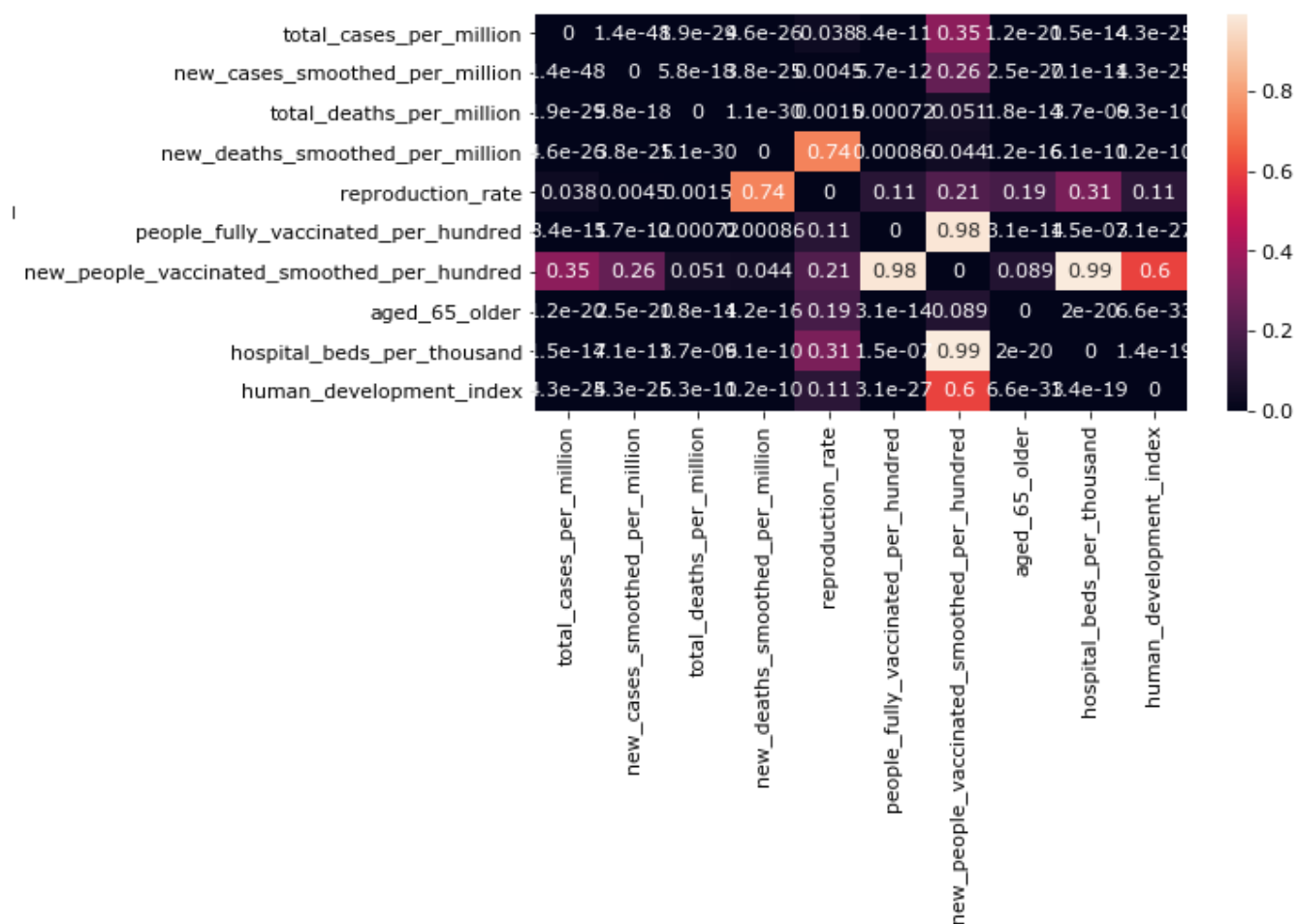
Рисунок 2 – матрица парных коэффициентов корреляции



Уровень значимости³, при котором будем считать гипотезу о наличии взаимосвязи между признаками, примем за 0.05. Проверка значимости этих коэффициентов показана на рисунке 3.

³ Уровень значимости – это вероятность, с которой проверка гипотезы о значимости коэффициента корреляции покажет отсутствие связи, когда она на самом деле есть (т.н. «ошибка первого рода»). Фиксируя уровень значимости, мы устанавливаем желаемый уровень точности вывода. Чем выше цена неправильного вывода об отсутствии связи – тем меньший уровень значимости следует устанавливать. Уровень значимости обычно выставляют из следующего диапазона значений: 0.05, 0.01, 0.001 (5%, 1%, 0.1% вероятности ошибиться соответственно).

Рисунок 3 – уровни значимости парных коэффициентов корреляции



Подавляющее большинство связей значимо на уровне значимости 0.05.

Много значимых связей – не повод для радости и поспешных выводов (к тому же нередко противоречащих здравому смыслу).

Поскольку парные коэффициенты корреляции учитывают взаимосвязи этих двух признаков со всеми прочими, необходимо исключить эти "остальные".

Для такой задачи используются частные коэффициенты корреляции.

Рисунок 4 – матрица частных коэффициентов корреляции

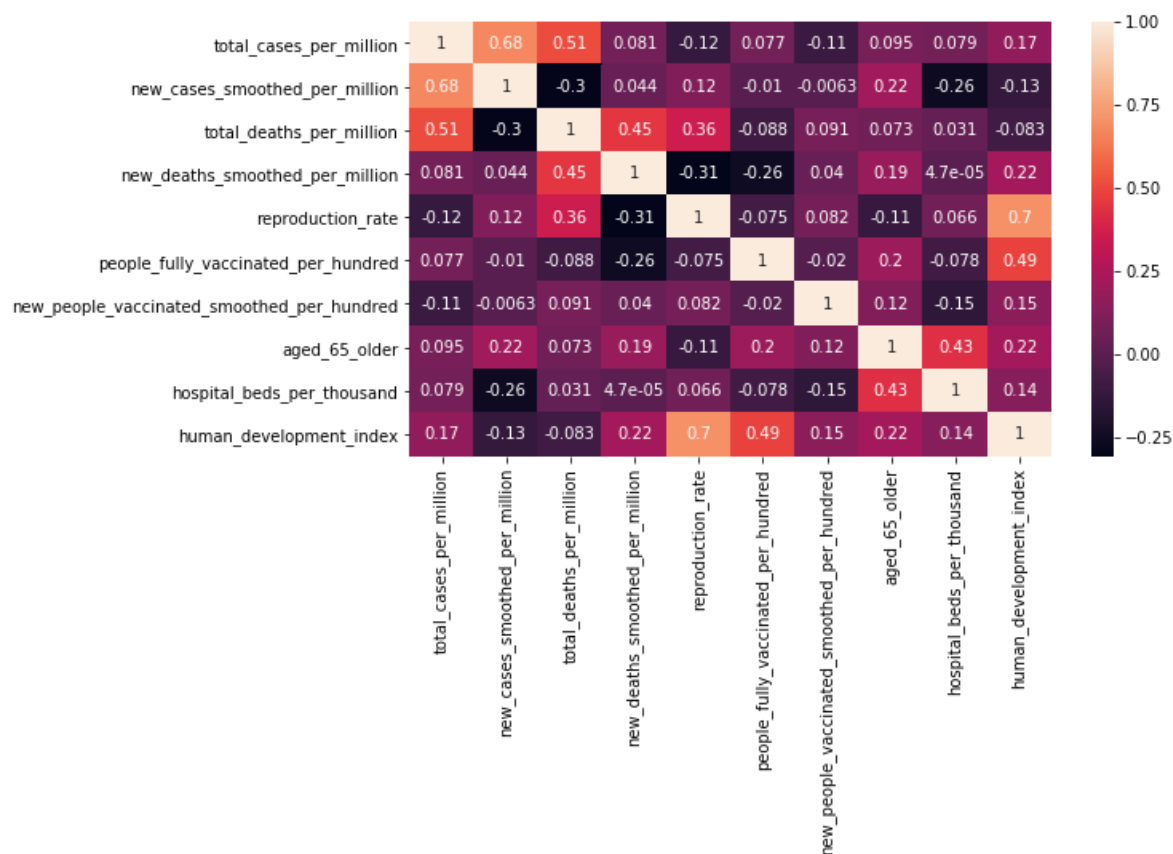
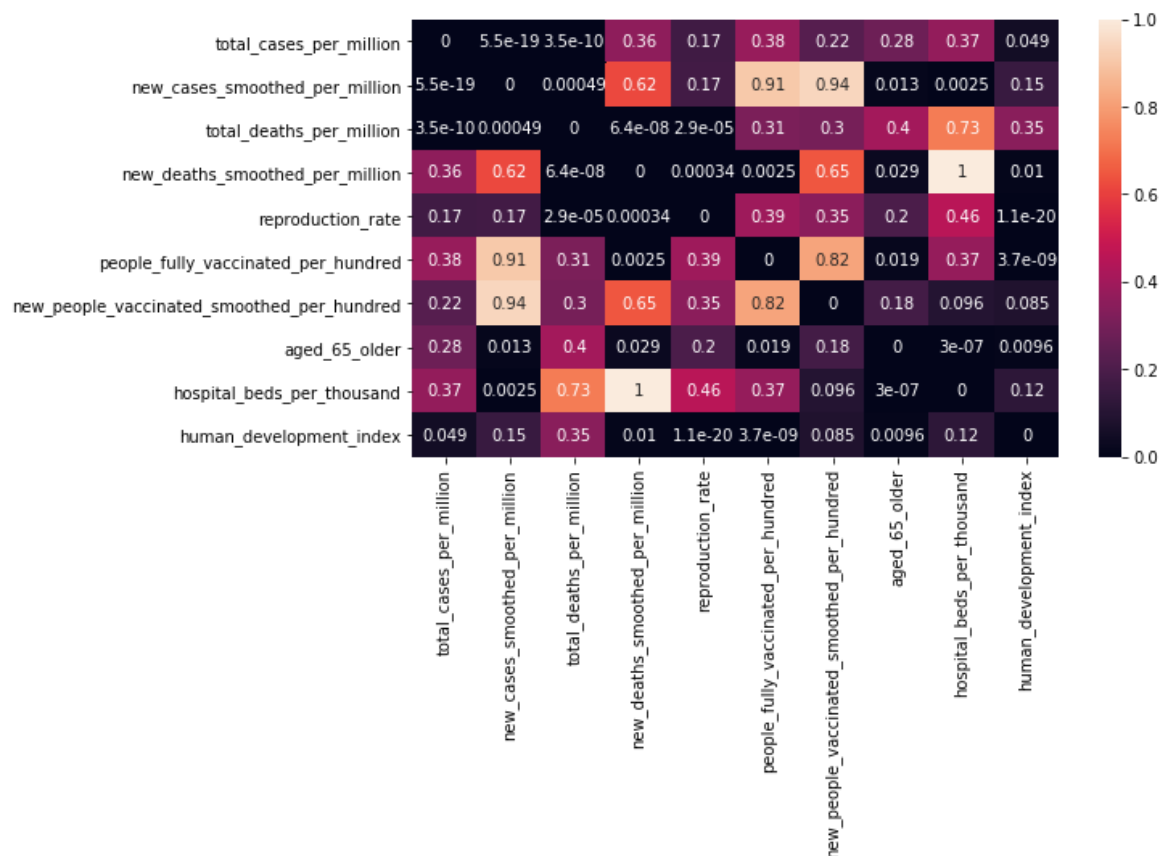


Рисунок 5 – уровни значимости частных коэффициентов корреляции



Среди частных коэффициентов корреляции значима примерно половина, при этом их значения существенно меньше, чем у парных.

Это происходит как между сильной взаимосвязи между признаками (*например, общее число зарегистрированных случаев и общее число зарегистрированных смертей*), так и из-за разнородности совокупности.

Чтобы преодолеть эту разнородность, я соберу похожие между собой объекты в несколько однородных групп. Это можно сделать с помощью методов кластерного анализа.

По принадлежности страны к той или иной группе можно будет сделать выводы о ситуации в ней.

Перед применением методов кластерного анализа сначала нужно привести все показатели примерно к одному масштабу. Иначе на кластеризацию будет влиять только признак с большим абсолютным разбросом значений показателя. Например, если мы проведём кластеризацию только по доле вакцинированного населения и числу выявленных случаев заболеваний на 100 000 населения за последние 7 дней, первый показатель практически не будет влиять на кластеризацию.

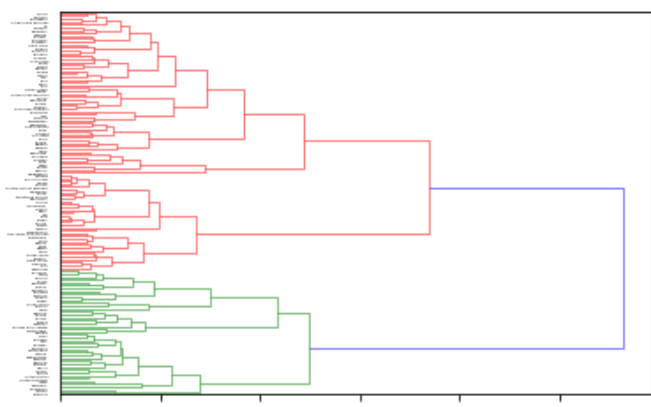
Для этой цели используются процедуры центрирования (вычитания среднего арифметического) и нормирования (деления на среднеквадратическое отклонение). В 99,7% случаев значения преобразованного таким способом признака не выйдут за пределы интервала $[-3; +3]$. Если речь, конечно, идёт о нормально распределённой совокупности.

Так как количество доступных вакцин не влияет на ключевые показатели, выпилим его из кластеризации.

Центрировано-нормированные значения показателей по странам представлены в приложении 2.

Самое сложное в процедурах кластерного анализа – понять, на какое число классов нужно разбить совокупность. Приблизительно оценить это число можно с помощью процедур иерархического кластерного анализа. Я использую метод Уорда.

Рисунок 6 – дендрограмма по результатам кластерного анализа

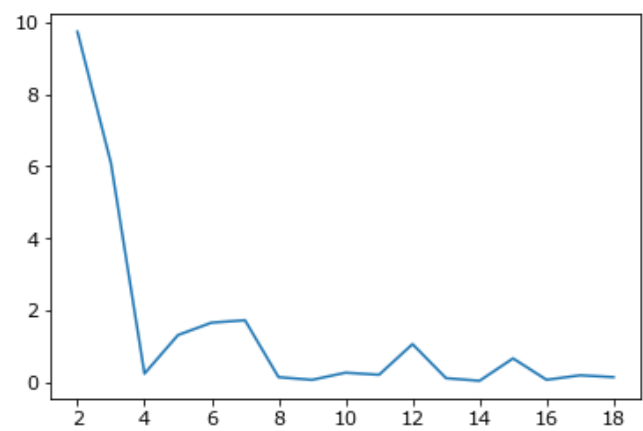


Дендрограмма в более читабельном виде с видимыми названиями стран:

https://github.com/Dimidro/COVID_research/blob/main/hier_covid_clust_v2.png

Оптимальным будет то число классов, при котором расстояние между ними увеличивается больше всего.

Рисунок 7 – изменение расстояния между объединяемыми кластерами по мере роста их количества



Выбор между 6 и 7 классами. Остановимся на 7, так как в нём наибольшее расстояние между классами.

Проведём кластерный анализ методом k-средних. Вычислим координаты центров полученных кластеров в признаковом пространстве, а также среднеквадратическое отклонение значений признаков. Это позволит нам характеризовать как сами классы, так и входящие в них объекты.

Рисунок 8 – центры полученных классов по центрировано-нормированным значениям признаков

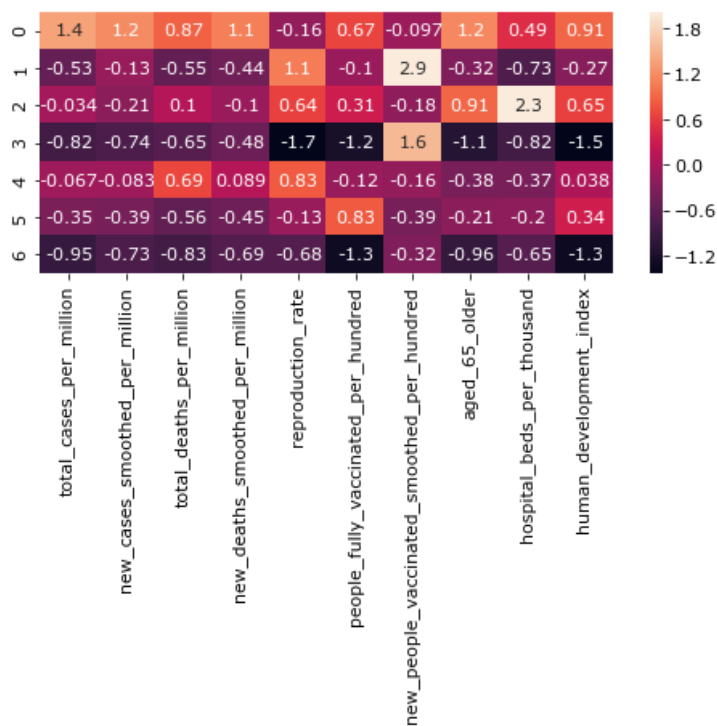
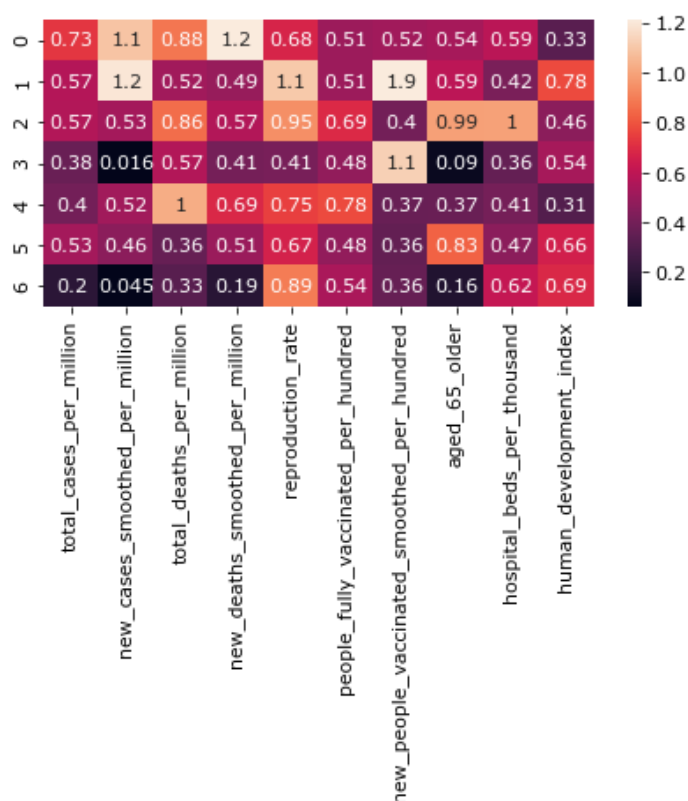


Рисунок 9 – среднеквадратические отклонения полученных классов по центрированно-нормированным значениям признаков



Используя центрированно-нормированные координаты центров классов, можем характеризовать ситуацию в них относительно "средней температуры по больнице" и друг друга. Среднеквадратические отклонения позволяют сделать вывод, насколько объекты в этом кластере однородны по значению данного признака.

Таблица 2 – описание кластеров

№	Характерные признаки	Однородность признаков	Страны в кластере
0	<p>Наивысшие значения:</p> <ul style="list-style-type: none"> • Индекс человеческого развития • Доля населения старше 65 лет • Общее число зарегистрированных случаев COVID-19 на 1 млн. населения за всю пандемию • Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию • Число новых зарегистрированных случаев COVID-19 на 1 млн. населения • Число новых зарегистрированных смертей от COVID-19 на 1 млн. населения <p>Высокие значения:</p> <ul style="list-style-type: none"> • Обеспеченность больничными койками на 1000 населения • Прошедших полный курс вакцинации на 100 населения 	Признаки однородны с точки зрения соотношения значений признаков в центре кластера и среднеквадратических отклонений.	<p>В кластер попали 33 страны.</p> <p>В кластере, в основном, находятся страны Западной Европы и развитые страны Латинской Америки.</p> <p>Примеры:</p> <ul style="list-style-type: none"> • Аргентина • Нидерланды • Словения

1	<p>Наивысшие значения:</p> <ul style="list-style-type: none"> Скорость распространения COVID-19 Новые вакцинации на 100 человек населения <p>Низкие значения:</p> <ul style="list-style-type: none"> Общее число зарегистрированных случаев COVID-19 на 1 млн. населения за всю пандемию Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию Число новых зарегистрированных смертей от COVID-19 на 1 млн. населения Обеспеченность больничными койками на 1000 населения 	<p>Разнородны по всем показателям, кроме:</p> <ul style="list-style-type: none"> Скорость распространения COVID-19 Новые вакцинации на 100 населения 	<p>В кластер попали 8 стран.</p> <p>Большинство стран кластера находится либо на островах, либо в горной местности.</p> <p>Примеры:</p> <ul style="list-style-type: none"> Албания Исландия Непал
2	<p>Наивысшие значения:</p> <ul style="list-style-type: none"> Обеспеченность больничными койками на 1000 населения <p>Высокие значения:</p> <ul style="list-style-type: none"> Доля населения старше 65 лет Индекс человеческого развития Скорость распространения COVID-19 <p>Низкие значения:</p> <ul style="list-style-type: none"> Новые вакцинации на 100 человек населения 	<p>Разнородны по всем показателям, кроме:</p> <ul style="list-style-type: none"> Обеспеченность больничными койками на 1000 населения Новые вакцинации на 100 человек населения 	<p>В кластер попали 11 стран.</p> <p>В этом кластере находятся развитые страны бывшего СССР с добавкой Западной Европы и азиатских соседей России.</p> <p>Примеры:</p> <ul style="list-style-type: none"> Беларусь Румыния Россия
3	<p>Низкие значения:</p> <ul style="list-style-type: none"> Прошедших полный курс вакцинации на 100 населения Общее число зарегистрированных случаев COVID-19 на 1 млн. населения за всю пандемию Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию Число новых зарегистрированных смертей от COVID-19 на 1 млн. населения <p>Наименьшие значения:</p> <ul style="list-style-type: none"> Индекс человеческого развития Доля населения старше 65 лет Скорость распространения COVID-19 Число новых зарегистрированных случаев COVID-19 на 1 млн. населения 	<p>Однородны практически по всем показателям. Это является следствием как схожести стран, так и их малого числа в кластере.</p>	<p>В кластер попали 3 страны «Чёрной Африки».</p> <p>Примеры:</p> <ul style="list-style-type: none"> Эсватини (Свазиленд) Мозамбик Уганда
4	<p>Высокие значения:</p> <ul style="list-style-type: none"> Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию Скорость распространения COVID-19 	<p>Разнородны по всем показателям, кроме:</p> <ul style="list-style-type: none"> Индекс человеческого развития Новые вакцинации на 100 человек населения Доля населения старше 65 лет 	<p>В кластер попали 22 страны.</p> <p>Этот кластер состоит из развивающихся стран. Большинство стран этого кластера</p>

			находится в Латинской Америке. Примеры: <ul style="list-style-type: none"> • Бразилия • Мексика • Кыргызстан
5	<p>Наивысшие значения:</p> <ul style="list-style-type: none"> • Прошедших полный курс вакцинации на 100 населения <p>Низкие значения:</p> <ul style="list-style-type: none"> • Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию • Число новых зарегистрированных смертей от COVID-19 на 1 млн. населения <p>Наименьшие значения:</p> <ul style="list-style-type: none"> • Новые вакцинации на 100 человек населения 	<p>Разнородны по всем показателям, кроме:</p> <ul style="list-style-type: none"> • Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию • Число новых зарегистрированных смертей от COVID-19 на 1 млн. населения • Прошедших полный курс вакцинации на 100 населения 	<p>В кластер попали 25 стран.</p> <p>В этом кластере демократические северные страны удивительным образом сочетаются с авторитарными государствами Латинской Америки и Азии.</p> <p>Примеры:</p> <ul style="list-style-type: none"> • Куба • Норвегия • Саудовская Аравия
6	<p>Низкие значения:</p> <ul style="list-style-type: none"> • Число новых зарегистрированных случаев COVID-19 на 1 млн. населения • Скорость распространения COVID-19 • Обеспеченность больничными койками на 1000 населения • Доля населения старше 65 лет • Индекс человеческого развития <p>Наименьшие значения:</p> <ul style="list-style-type: none"> • Прошедших полный курс вакцинации на 100 населения • Общее число зарегистрированных случаев COVID-19 на 1 млн. населения за всю пандемию • Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию • Число новых зарегистрированных смертей от COVID-19 на 1 млн. населения 	<p>Однородны практически по всем показателям, кроме:</p> <ul style="list-style-type: none"> • Скорость распространения COVID-19 • Прошедших полный курс вакцинации на 100 населения • Обеспеченность больничными койками на 1000 населения • Индекс человеческого развития <p>Это является следствием схожести стран в кластере.</p>	<p>В кластер попали 30 стран.</p> <p>Сюда вошли страны Африки и Азии, большинство из которых пострадало в ходе войн и революций («арабская весна») последних десятилетий.</p> <p>Примеры:</p> <ul style="list-style-type: none"> • Афганистан • Центральная Африканская Республика • Йемен

Для того, чтобы получить содержательно интерпретируемые значения признаков, я проведу операцию обратную той, которую делал перед кластеризацией – умножу значения центров кластеров на среднеквадратические отклонения исходных признаков и прибавлю к ним соответствующие математические ожидания.

Таблица 3 – центры полученных классов по исходным значениям признаков

№	total_cases_per_1mln	new_cases_smoothed_per_1mln	total_deaths_per_1mln	new_deaths_smoothed_per_1mln	reproduction_rate	people_fully_vaccinated_per_100	new_people_vaccinated_smoothed_per_100	aged_65_and_over	hospital_beds_per_1000	human_development_index
0	201 739	2 461	2 138	4,69	1,23	67,1%	0,07%	17,1%	4,25	0,886
1	46 673	795	556	0,90	1,78	46,9%	0,37%	7,7%	1,31	0,715
2	86 789	696	1 278	1,73	1,58	57,5%	0,06%	15,5%	8,63	0,849
3	22 743	29	435	0,78	0,55	18,1%	0,23%	2,8%	1,10	0,537
4	84 108	858	1 937	2,20	1,67	46,4%	0,06%	7,4%	2,17	0,760
5	61 406	471	539	0,87	1,25	71,1%	0,04%	8,4%	2,58	0,803
6	12 595	39	243	0,28	1,00	15,7%	0,04%	3,7%	1,50	0,563

Получены центры классов в исходных значениях признаков. Это поможет более наглядно увидеть разницу между классами.

Видим, что на 19.01.2022 в подавляющем большинстве кластеров эпидемия будет продолжаться в силу скорости распространения COVID-19 более единицы.

Распределение стран по кластерам с исходными значениями признаков представлено в приложении 3.

В начале исследования мы говорили о том, что отсутствие сильной взаимосвязи между признаками может быть вызвано разнородностью объектов.

Сейчас мы получили относительно однородные объекты. По кластерам численностью более 20 объектов мы можем провести корреляционный анализ. Такими кластерами являются кластер 0, кластер 4, кластер 5, кластер 6.

Сначала по каждому из кластеров нужно проверить нормальность распределения признаков, после чего решать вопрос о применении параметрических или непараметрических методов корреляционного анализа.

Проверки нормальности распределения признаков для каждого кластера приведены в приложении 4.

Расчёты частных коэффициентов корреляции и проверки значимости приведены в приложении 5.

Результаты корреляционного анализа по кластерам приведены в таблице 4.

Таблица 4 – результаты корреляционного анализа по кластерам

Кластер	Есть признаки без НР ⁴ ?	Значимые связи с числом зарегистрированных новых случаев и новых смертей от COVID-19	Возможен прогноз числа новых случаев или смертей с помощью КЛММР ⁵ ?
0 (страны первого мира)	Да	<ul style="list-style-type: none"> Прошедших полный курс вакцинации на 100 населения – Новых зарегистрированных случаев COVID-19 на 1 млн. населения, прямая связь ($\hat{r} = 0.38$)⁶. 	Нет
4 (Латинская Америка)	Да	<ul style="list-style-type: none"> Прошедших полный курс вакцинации на 100 населения – Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию, прямая связь ($\hat{r} = 0.5$)⁷. 	Нет
5 (страны с политической волей)	Да	<ul style="list-style-type: none"> Прошедших полный курс вакцинации на 100 населения – Общее число зарегистрированных смертей от COVID-19 на 1 млн. населения за всю пандемию, обратная связь ($\hat{r} = -0.46$)⁸. Прошедших полный курс вакцинации на 100 населения – Новых зарегистрированных случаев COVID-19 на 1 млн. населения за всю пандемию, обратная связь ($\hat{r} = -0.51$)⁹. 	Нет
6 (беднейшие страны)	Да	Значимых взаимосвязей нет	Нет

⁴ Нормальное распределение

⁵ КЛММР – классическая линейная модель регрессии. Она имеет вид $\hat{y}(x) = const + b_1x_1 + \dots + b_nx_n$

⁶ Возможно, дело в том, что после вакцинации некоторые люди считают себя бессмертными и начинают ходить в общественных без масок, а также не соблюдать требования гигиены. Омикрон-штамм коронавируса, доминирующий сейчас в Европе, по моим наблюдениям среди знакомых пробивает как естественную защиту (у переболевших COVID-19), так и вакцинный иммунитет, в том числе при комбинировании Спутник-Pfizer (одна из них – как основная вакцина, вторая – как бустер).

⁷ Возможно, это связано с тем, что если среди чьего-то круга общения от ковида умрёт много людей, то он с большей вероятностью пойдёт делать прививку / это заставить сделать его правительство.

⁸ Возможно, эта взаимосвязь объясняется политической волей этих стран к жёстким действиям (локдауны, массовое тестирование, принудительная вакцинация и т.п.) в том числе в случае угрозы появления нового штамма. Для примера почитайте про [антиковидные меры в Саудовской Аравии](#), относящейся к 5 кластеру по сделанной классификации.

⁹ См. пункт 8

Вывод

В ситуации с распространением Омикрон-штамма COVID-19, игнорирующего приобретённый иммунитет как после вакцинации, как и после самого заболевания, сама по себе вакцинация перестала давать существенное снижение шансов заражения новым штаммом.

Однако, на примере кластера 5 мы видим, что вакцинация может быть эффективна в комплексе с другими мерами, такими как:

- Ношение масок в общественных местах
- Массовое тестирование на COVID-19
- Своевременное выявление контактов заболевших людей и тестированием этих контактов с изоляцией до выяснения ковидного статуса
- Перевод большинства сотрудников на удалённую работу
- Контроль за реализацией антиковидных ограничений со штрафами за их нарушение

Поэтому вопрос защиты населения от коронавируса должен решаться комплексно. Рано или поздно это придётся сделать – вопрос лишь в том, каковы будут потери к моменту, когда соответствующие решения будут приняты.

Об авторе

Меня зовут Дмитрий Евдокимов.

Профессионально Machine Learning не занимаюсь, но развлекаюсь иногда. Благо что-то помню из университетского образования ☺

Контакты:

- Facebook: <https://www.facebook.com/dmitry.evdokimov.12/>
- Telegram: @Dimidr0
- GitHub с исходным кодом исследования:
https://github.com/Dimidro/COVID_research
- Instagram: @dimidr0

Приложение 1 – проверка гипотез на нормальное распределение совокупностей

```
In [10]: alpha = 0.01

for i, col in enumerate(clean_col_names):
    k2, p = stats.normaltest(master_data_clean[col])
    print("p = {:g}".format(p))
    if p < alpha: # null hypothesis: x comes from a normal distribution
        print(col+": The null hypothesis can be rejected")
    else:
        print(col+": The null hypothesis cannot be rejected")

p = 0.000636694
total_cases_per_million: The null hypothesis can be rejected
p = 1.19854e-11
new_cases_smoothed_per_million: The null hypothesis can be rejected
p = 5.85577e-08
total_deaths_per_million: The null hypothesis can be rejected
p = 1.64653e-12
new_deaths_smoothed_per_million: The null hypothesis can be rejected
p = 0.17446
reproduction_rate: The null hypothesis cannot be rejected
p = 2.31182e-06
people_fully_vaccinated_per_hundred: The null hypothesis can be rejected
p = 2.13466e-31
new_people_vaccinated_smoothed_per_hundred: The null hypothesis can be rejected
p = 3.5599e-06
aged_65_older: The null hypothesis can be rejected
p = 4.84143e-11
hospital_beds_per_thousand: The null hypothesis can be rejected
p = 0.00799628
human_development_index: The null hypothesis can be rejected
```

Приложение 2 – центрировано-нормированные значения показателей по странам

location	total_cases_per_1m ln	new_cases_smoothed_per_1 mln	total_deaths_per_1 mln	new_deaths_smoothed_per_1 mln	reproduction_rate	people_fully_vaccinated_per_1 00	new_people_vaccinated_smoothed_per_1 100	aged_65_and_over	hospital_beds_per_10 00	human_development_index
Afghanistan	-1,051	-0,763	-0,878	-0,784	0,312	-1,537	-0,531	-1,127	-1,064	-1,681
Albania	-0,077	-0,190	-0,020	-0,198	0,585	-0,420	2,179	0,537	-0,077	0,280
Algeria	-1,037	-0,751	-0,915	-0,707	0,084	-1,408	-0,521	-0,557	-0,486	-0,044
Antigua and Barbuda	-0,386	-0,159	0,055	-0,216	-0,553	0,410	-0,186	-0,444	0,298	0,163
Argentina	0,906	1,179	1,288	0,410	0,016	0,957	0,110	0,225	0,794	0,626
Armenia	0,344	-0,692	1,380	-0,586	0,494	-0,891	0,140	0,230	0,463	0,149
Australia	-0,136	1,978	-0,944	0,022	-0,098	1,086	1,637	0,900	0,315	1,309
Austria	0,943	0,802	0,340	-0,328	0,517	0,933	-0,245	1,480	1,772	1,157
Azerbaijan	-0,345	-0,706	-0,294	-0,405	-0,235	-0,131	-0,521	-0,588	0,670	0,011
Bahamas	-0,123	-0,245	0,580	-0,509	0,061	-0,450	-0,442	-0,120	-0,073	0,412
Bahrain	1,131	0,457	-0,327	-0,766	0,721	0,712	-0,511	-1,159	-0,445	0,674
Bangladesh	-0,979	-0,736	-0,892	-0,776	2,269	-0,574	7,106	-0,732	-0,940	-0,845
Barbados	0,483	0,714	-0,200	0,001	-0,121	0,047	-0,324	0,814	1,124	0,412
Belarus	-0,164	-0,658	-0,487	-0,135	-0,235	-0,393	0,534	0,790	3,271	0,474
Belgium	1,621	1,459	1,170	-0,007	0,562	1,015	0,051	1,381	1,058	1,219
Belize	0,240	0,690	0,309	-0,231	0,767	0,005	-0,491	-0,927	-0,734	-0,265
Bhutan	-1,049	-0,707	-1,041	-0,799	1,063	0,905	-0,639	-0,765	-0,569	-0,693
Bolivia	-0,296	-0,162	0,500	0,533	0,289	-0,279	2,297	-0,480	-0,816	-0,251
Brazil	0,245	-0,393	1,563	-0,400	1,222	0,748	-0,225	-0,190	-0,362	0,073
Brunei	-0,655	-0,704	-0,845	-0,799	0,289	1,598	-0,639	-0,811	-0,156	0,577
Burkina Faso	-1,089	-0,762	-1,030	-0,745	-1,828	-1,770	-0,432	-1,154	-1,105	-2,088
Burundi	-1,063	-0,755	-1,042	-0,799	-1,213	-1,906	-0,728	-1,130	-0,940	-2,220
Cambodia	-1,013	-0,763	-0,885	-0,799	0,039	1,207	-0,521	-0,840	-0,940	-1,108

Cameroon	-1,051	-0,764	-0,983	-0,799	-1,464	-1,814	-0,728	-1,035	-0,734	-1,322
Canada	-0,181	-0,221	-0,291	0,463	-0,690	1,095	-0,107	1,133	-0,238	1,206
Cape Verde	0,100	-0,328	-0,438	0,430	-0,986	-0,073	0,337	-0,832	-0,404	-0,617
Central African Republic	-1,067	-0,744	-1,025	-0,775	-2,169	-1,524	-0,077	-0,958	-0,858	-2,468
Chile	0,117	-0,408	0,796	-0,377	1,199	1,457	-0,255	0,208	-0,399	0,667
China	-1,099	-0,764	-1,042	-0,799	-0,963	1,344	-0,599	0,138	0,521	0,046
Colombia	0,248	-0,292	1,255	0,220	0,039	0,353	0,938	-0,332	-0,565	0,087
Costa Rica	0,410	-0,073	0,253	-0,351	0,608	0,809	0,130	-0,046	-0,804	0,384
Croatia	1,419	0,674	1,868	3,063	-0,326	0,156	-0,373	1,562	1,017	0,667
Cuba	-0,005	-0,532	-0,383	-0,707	0,539	1,416	-0,294	0,780	0,876	0,197
Cyprus	2,097	1,489	-0,355	0,485	-0,872	0,758	-0,107	0,573	0,133	0,916
Czechia	1,940	0,192	2,044	0,124	0,403	0,512	-0,412	1,453	1,467	1,005
Denmark	1,505	3,167	-0,499	0,211	-0,075	1,202	-0,511	1,555	-0,238	1,282
Djibouti	-0,915	-0,697	-0,875	-0,799	0,562	-1,587	-0,097	-0,871	-0,693	-1,591
Dominican Republic	-0,519	-0,339	-0,694	-0,720	-0,030	0,112	-0,540	-0,437	-0,610	0,011
Ecuador	-0,668	-0,324	0,672	0,916	-0,439	0,912	0,475	-0,417	-0,651	0,032
Egypt	-1,053	-0,755	-0,853	-0,690	-0,189	-0,991	0,603	-0,722	-0,610	-0,327
Equatorial Guinea	-0,968	-0,711	-0,934	-0,719	-1,191	-1,366	-0,629	-1,085	-0,404	-1,122
Estonia	1,447	0,664	0,302	0,373	0,152	0,486	-0,412	1,520	0,666	0,950
Eswatini	-0,388	-0,725	0,000	-0,013	-2,192	-0,853	0,573	-1,035	-0,404	-0,990
Ethiopia	-1,052	-0,756	-0,990	-0,750	-1,555	-1,856	-0,728	-0,978	-1,147	-1,860
Fiji	-0,277	-0,511	-0,297	1,176	0,175	0,701	-0,678	-0,555	-0,321	-0,079
Finland	-0,278	0,576	-0,766	0,093	-0,690	0,932	-0,659	1,798	0,084	1,268
France	1,732	3,003	0,654	0,350	-0,530	0,996	-0,353	1,561	1,198	1,012
Gabon	-0,855	-0,715	-0,927	-0,748	-0,644	-1,533	0,002	-0,834	1,330	-0,355
Georgia	2,036	0,271	2,236	2,035	0,243	-0,748	0,356	0,800	-0,197	0,398
Germany	0,125	0,025	0,199	0,069	0,130	0,878	-0,176	1,834	2,032	1,330
Ghana	-1,040	-0,757	-1,006	-0,773	-0,257	-1,554	-0,274	-1,001	-0,899	-0,990
Greece	0,942	0,668	0,883	2,796	-0,849	0,764	0,386	1,668	0,468	0,923
Grenada	0,034	1,067	0,574	0,728	0,835	-0,671	-0,383	-0,386	0,257	0,170

Guatemala	-0,656	-0,664	-0,248	-0,562	0,699	-0,842	0,071	-0,795	-1,023	-0,631
Guinea	-1,068	-0,759	-1,017	-0,769	-0,348	-1,538	-0,678	-1,040	-1,147	-1,916
Guyana	-0,262	0,228	0,213	1,603	0,995	-0,431	-0,166	-0,699	-0,610	-0,500
Honduras	-0,629	-0,725	-0,112	-0,753	-1,691	-0,232	-0,471	-0,802	-0,982	-0,832
Hungary	0,647	-0,045	2,743	2,802	0,721	0,516	-0,363	1,382	1,628	0,688
Iceland	0,617	2,770	-0,937	-0,019	0,175	1,070	4,081	0,732	-0,069	1,344
India	-0,763	-0,610	-0,731	-0,689	1,677	-0,065	1,322	-0,592	-1,052	-0,756
Indonesia	-0,910	-0,761	-0,576	-0,790	0,699	-0,222	2,376	-0,697	-0,841	-0,251
Iran	-0,199	-0,738	0,349	-0,690	0,471	0,506	-0,402	-0,678	-0,651	0,197
Iraq	-0,464	-0,685	-0,516	-0,740	2,770	-1,335	-0,048	-1,032	-0,693	-0,555
Ireland	1,668	1,061	0,051	-0,198	-1,282	1,070	0,593	0,653	-0,049	1,385
Israel	1,592	3,699	-0,237	-0,353	-0,394	0,595	-0,402	0,309	-0,036	1,137
Italy	0,777	1,580	1,068	1,425	-0,280	0,993	0,396	2,080	0,042	0,950
Jamaica	-0,620	-0,394	-0,274	0,149	0,653	-1,130	-0,412	-0,013	-0,569	-0,141
Japan	-0,908	-0,597	-0,913	-0,776	2,702	1,133	-0,619	2,711	4,117	1,137
Kazakhstan	-0,333	-0,240	-0,179	-0,647	1,882	-0,157	-0,314	-0,435	1,496	0,487
Kenya	-1,029	-0,755	-0,954	-0,745	-1,760	-1,583	0,031	-1,110	-0,693	-1,059
Kuwait	0,262	0,102	-0,531	-0,706	0,426	0,973	-0,540	-1,164	-0,445	0,356
Kyrgyzstan	-0,744	-0,687	-0,660	-0,694	1,677	-1,319	-0,196	-0,827	0,587	-0,396
Laos	-0,888	-0,674	-0,983	-0,464	-0,940	0,092	-0,176	-0,900	-0,651	-0,977
Latvia	0,968	0,495	1,242	1,327	0,357	0,730	-0,324	1,567	1,029	0,771
Lebanon	0,426	0,002	0,205	0,085	-0,030	-0,805	-0,363	-0,196	-0,073	-0,072
Libya	-0,394	-0,682	-0,287	-0,295	-0,348	-1,384	-0,008	-0,838	0,257	-0,210
Lithuania	1,558	0,325	1,522	1,190	0,198	0,742	-0,452	1,449	1,438	0,881
Malawi	-1,048	-0,754	-0,931	-0,640	-1,327	-1,759	-0,728	-1,064	-0,734	-1,874
Malaysia	-0,044	-0,689	-0,173	-0,636	-0,735	1,108	-0,688	-0,544	-0,486	0,384
Mali	-1,083	-0,753	-1,014	-0,733	-0,143	-1,807	-0,363	-1,136	-1,229	-2,213
Malta	0,441	-0,151	-0,151	1,207	-1,418	1,386	0,790	1,516	0,581	0,971
Mexico	-0,676	-0,520	1,036	-0,204	0,858	0,337	-0,255	-0,456	-0,701	0,170
Mongolia	0,441	-0,181	-0,481	-0,574	1,108	0,598	-0,688	-0,899	1,619	-0,120
Montenegro	2,973	1,317	2,516	2,499	-0,986	-0,215	-0,550	0,784	0,323	0,515
Morocco	-0,748	-0,608	-0,684	-0,660	0,175	0,474	-0,659	-0,470	-0,816	-0,472

Mozambique	-1,016	-0,743	-0,985	-0,736	-1,464	-1,024	2,790	-1,036	-0,982	-2,061
Myanmar	-0,981	-0,762	-0,728	-0,789	-0,917	-0,675	0,149	-0,632	-0,899	-1,184
Nepal	-0,737	-0,614	-0,693	-0,762	2,998	-0,335	2,593	-0,620	-1,147	-1,052
Netherlands	1,540	0,887	0,065	-0,585	-0,735	0,846	-0,649	1,414	0,100	1,309
New Zealand	-1,063	-0,755	-1,035	-0,799	-0,667	1,025	-0,521	0,872	-0,193	1,219
Nicaragua	-1,068	-0,764	-1,015	-0,790	-0,052	-0,007	1,292	-0,677	-0,899	-0,652
Niger	-1,096	-0,763	-1,034	-0,781	-0,394	-1,753	0,100	-1,131	-1,147	-2,489
Norway	0,147	0,945	-0,813	-0,378	0,243	0,883	-0,550	1,107	0,216	1,399
Oman	-0,356	-0,615	-0,336	-0,744	2,019	0,255	-0,511	-1,162	-0,610	0,405
Pakistan	-1,027	-0,748	-0,929	-0,786	1,495	-0,569	0,445	-0,826	-1,023	-1,363
Panama	0,590	0,854	0,502	0,094	1,359	0,278	0,396	-0,290	-0,321	0,418
Paraguay	-0,232	-0,229	1,051	0,763	-0,348	-0,285	-0,324	-0,531	-0,734	-0,182
Peru	-0,097	0,197	4,433	0,162	0,949	0,679	-0,294	-0,410	-0,610	0,156
Poland	0,322	-0,389	1,401	2,434	-0,121	0,279	-0,127	1,098	1,463	0,867
Portugal	1,321	2,233	0,668	0,514	-0,007	1,564	-0,481	1,841	0,129	0,757
Romania	0,149	-0,297	1,740	0,016	0,926	-0,313	0,524	1,268	1,575	0,508
Russia	-0,198	-0,613	0,903	1,110	0,107	-0,090	-0,028	0,692	2,053	0,481
Saint Lucia	0,091	0,425	0,492	1,073	0,813	-0,852	-0,432	-0,007	-0,734	0,032
Sao Tome and Principe	-0,789	-0,553	-0,779	-0,541	0,312	-0,850	-0,412	-1,079	-0,073	-0,894
Saudi Arabia	-0,880	-0,639	-0,818	-0,777	1,017	0,653	-0,294	-1,015	-0,156	0,688
Serbia	1,555	0,859	0,672	0,825	1,199	-0,096	-0,599	1,192	1,045	0,356
Seychelles	2,981	3,614	0,225	2,691	0,152	1,149	-0,225	-0,182	0,216	0,287
Singapore	-0,433	-0,599	-0,906	-0,746	0,198	1,431	0,504	0,495	-0,280	1,268
Slovakia	2,122	-0,226	1,823	3,184	-0,986	-0,056	-0,393	0,832	1,132	0,729
Slovenia	2,206	2,147	1,424	0,419	0,608	0,321	-0,501	1,459	0,587	1,123
South Africa	-0,371	-0,707	0,353	0,085	-1,395	-0,869	-0,570	-0,693	-0,313	-0,314
South Korea	-0,930	-0,690	-0,931	-0,496	-0,280	1,367	-0,491	0,651	3,795	1,116
Spain	1,181	1,432	0,710	0,345	-0,235	1,236	0,317	1,517	-0,044	1,033
Sri Lanka	-0,758	-0,739	-0,408	-0,585	-0,576	0,576	0,494	0,048	0,216	0,191

Sudan	-1,086	-0,754	-0,977	-0,758	-0,667	-1,787	-0,560	-0,975	-0,940	-1,688
Suriname	0,307	0,566	0,807	1,243	0,539	-0,404	-0,560	-0,444	0,009	-0,113
Sweden	0,956	1,368	0,335	0,124	1,063	0,932	-0,383	1,603	-0,354	1,316
Switzerland	1,440	1,800	0,246	-0,026	-0,030	0,693	-0,264	1,360	0,600	1,385
Thailand	-0,687	-0,679	-0,763	-0,702	-1,623	0,709	-0,146	0,252	-0,404	0,156
Timor	-0,919	-0,764	-0,963	-0,799	-1,282	-0,312	-0,550	-0,974	1,165	-1,025
Togo	-1,048	-0,746	-1,017	-0,778	-0,667	-1,428	-0,629	-1,086	-0,982	-1,653
Trinidad and Tobago	-0,199	-0,403	1,025	4,080	-0,599	-0,041	-0,255	0,039	-0,032	0,287
Tunisia	-0,272	-0,307	0,898	-0,245	1,131	0,069	-0,284	-0,277	-0,321	-0,100
Turkey	0,441	-0,140	-0,145	-0,036	-0,394	0,454	-0,540	-0,253	-0,110	0,453
Uganda	-1,059	-0,757	-0,979	-0,703	-1,486	-1,760	1,312	-1,191	-1,064	-1,453
Ukraine	0,021	-0,593	1,127	0,627	0,312	-0,634	-0,186	1,051	2,363	0,170
United Arab Emirates	-0,099	-0,530	-0,847	-0,689	-0,348	1,642	-0,737	-1,352	-0,775	0,936
United Kingdom	1,701	0,311	0,967	0,776	-1,236	0,801	-0,452	1,373	-0,222	1,226
United States	1,434	1,056	1,266	1,340	-0,576	0,511	-0,284	0,886	-0,127	1,185
Uruguay	0,815	1,512	0,566	0,076	0,835	1,059	1,036	0,767	-0,115	0,432
Yemen	-1,096	-0,764	-0,986	-0,785	0,312	-1,870	-0,649	-1,073	-0,982	-1,964
Zimbabwe	-0,915	-0,746	-0,731	-0,604	-1,851	-1,079	-0,373	-1,089	-0,569	-1,267

Приложение 3 – распределение стран по кластерам и исходные значения признаков на момент исследования

cluster	location	total_cases_per_1mln	new_cases_smoothed_per_1mln	total_deaths_per_1mln	new_deaths_smoothed_per_1mln	reproduction_rate	people_fully_vaccinated_per_100	new_people_vaccinated_smoothed_per_100	aged_65_and_over	hospital_beds_per_1000	human_development_index
0	Argentina	163282,35	2449,277	2601,159	3,001	1,31	74,45	0,086	11,198	5	0,845
0	Australia	78430,432	3455,648	112,067	2,039	1,26	77,81	0,241	15,504	3,84	0,944
0	Belgium	221392,63	2801,907	2469,496	1,965	1,55	75,98	0,08	18,571	5,64	0,931
0	Croatia	205016,49	1812,079	3247,946	9,59	1,16	53,63	0,037	19,724	5,54	0,851
0	Cyprus	260134,71	2839,748	768,969	3,189	0,92	69,28	0,064	13,416	3,4	0,887
0	Czechia	247346,91	1204,845	3444,153	2,291	1,48	62,89	0,033	19,027	6,63	0,9
0	Denmark	211968,69	4954,155	608,088	2,507	1,27	80,82	0,023	19,677	2,5	0,94
0	Estonia	207259,65	1800,284	1501,674	2,911	1,37	62,23	0,033	19,452	4,69	0,892
0	France	230484,47	4747,247	1894,278	2,852	1,07	75,48	0,039	19,718	5,98	0,901
0	Georgia	255188,42	1305,064	3657,244	7,036	1,41	30,15	0,111	14,864	2,6	0,812
0	Greece	166188,22	1805,022	2148,833	8,926	0,93	69,45	0,114	20,396	4,21	0,888
0	Hungary	142199,6	905,958	4223,097	8,941	1,62	63	0,038	18,577	7,02	0,854
0	Ireland	225255,8	2299,749	1221,577	1,491	0,74	77,41	0,135	13,928	2,96	0,955
0	Israel	219074,91	5624,537	900,011	1,107	1,13	65,06	0,034	11,733	2,99	0,919
0	Italy	152721,17	2954,105	2355,656	5,521	1,18	75,41	0,115	23,021	3,18	0,892
0	Latvia	168266,26	1587,553	2549,099	5,28	1,46	68,55	0,042	19,754	5,57	0,866
0	Lithuania	216302,18	1372,986	2861,113	4,939	1,39	68,87	0,029	19,002	6,56	0,882
0	Malta	125400,12	772,552	995,931	4,982	0,68	85,62	0,155	19,426	4,485	0,895
0	Montenegro	331455,57	2622,853	3969,423	8,189	0,87	43,99	0,019	14,762	3,861	0,829
0	Netherlands	214853,19	2081,254	1237,575	0,532	0,98	71,58	0,009	18,779	3,32	0,944
0	Poland	115716,01	473,05	2726,724	8,028	1,25	56,84	0,062	16,763	6,62	0,88

0	Portugal	197008,67	3776,948	1909,239	3,26	1,3	90,23	0,026	21,502	3,39	0,864
0	Serbia	216038,11	2045,744	1913,979	4,033	1,83	47,09	0,014	17,366	5,609	0,806
0	Seychelles	332120,11	5517,281	1415,428	8,666	1,37	79,46	0,052	8,606	3,6	0,796
0	Slovakia	262194,08	677,984	3197,194	9,889	0,87	48,13	0,035	15,07	5,82	0,86
0	Slovenia	269065,19	3668,667	2752,652	3,024	1,57	57,94	0,024	19,062	4,5	0,917
0	Spain	185621,5	2767,431	1956,072	2,839	1,2	81,72	0,107	19,436	2,97	0,904
0	Sweden	167353,29	2687,05	1537,87	2,292	1,77	73,81	0,036	19,985	2,22	0,945
0	Switzerland	206703,37	3231,536	1439,276	1,918	1,29	67,61	0,048	18,436	4,53	0,955
0	Trinidad and Tobago	73353,219	454,822	2307,297	12,114	1,04	48,53	0,049	10,014	3	0,796
0	United Kingdom	227899,44	1355,324	2243,109	3,91	0,76	70,4	0,029	18,517	2,54	0,932
0	United States	206248,69	2293,663	2576,537	5,311	1,05	62,88	0,046	15,413	2,77	0,926
0	Uruguay	155851,45	2868,741	1795,618	2,172	1,67	77,1	0,18	14,655	2,8	0,817
1	Albania	83235,118	723,999	1142,734	1,492	1,56	38,66	0,296	13,188	2,89	0,795
1	Bangladesh	9875,283	35,21	169,425	0,056	2,3	34,68	0,796	5,098	0,8	0,632
1	Bolivia	65429,831	759,72	1722,058	3,308	1,43	42,33	0,308	6,704	1,1	0,718
1	Iceland	139740,02	4453,924	119,308	1,937	1,38	77,4	0,489	14,431	2,91	0,949
1	India	27428,251	194,881	350	0,273	2,04	47,91	0,209	5,989	0,53	0,645
1	Indonesia	15470,764	3,841	521,751	0,022	1,61	43,81	0,316	5,319	1,04	0,718
1	Nepal	29577,738	189,395	391,846	0,091	2,62	40,87	0,338	5,809	0,3	0,602
1	Nicaragua	2626,53	0,874	32,675	0,021	1,28	49,4	0,206	5,445	0,9	0,66
2	Austria	166278,45	1973,429	1543,281	1,169	1,53	73,84	0,05	19,202	7,37	0,922
2	Barbados	128821,58	1863	941,927	1,986	1,25	50,8	0,042	14,952	5,8	0,814
2	Belarus	76193,491	133,933	621,21	1,649	1,2	39,38	0,129	14,799	11	0,823
2	Germany	99656,914	995,053	1386,417	2,154	1,36	72,41	0,057	21,453	8	0,947
2	Japan	15674,324	210,728	146,433	0,057	2,49	79,03	0,012	27,049	13,05	0,919

2	Kazakhstan	62433,831	660,837	965,414	0,376	2,13	45,5	0,043	6,991	6,7	0,825
2	Mongolia	125419,54	734,736	628,364	0,558	1,79	65,14	0,005	4,031	7	0,737
2	Romania	101680,22	588,628	3104,648	2,024	1,71	41,44	0,128	17,85	6,892	0,828
2	Russia	73444,236	191,406	2171,528	4,741	1,35	47,26	0,072	14,178	8,05	0,824
2	South Korea	13887,544	93,204	126,303	0,752	1,18	85,13	0,025	13,914	12,27	0,916
2	Ukraine	91194,084	215,853	2421,065	3,54	1,44	33,11	0,056	16,462	8,8	0,779
3	Eswatini	57980,892	50,082	1164,309	1,95	0,34	27,41	0,133	3,163	2,1	0,611
3	Mozambique	6868,38	27,099	66,536	0,155	0,66	22,97	0,358	3,158	0,7	0,456
3	Uganda	3379,84	9,571	73,191	0,239	0,65	3,84	0,208	2,168	0,5	0,544
4	Armenia	117484,15	91,303	2703,724	0,529	1,52	26,44	0,089	11,232	4,2	0,776
4	Bahamas	79485,732	654,694	1811,476	0,72	1,33	37,89	0,03	8,996	2,9	0,814
4	Belize	109022,88	1832,836	1508,959	1,411	1,64	49,73	0,025	3,853	1,3	0,716
4	Brazil	109467,8	468,078	2907,215	0,991	1,84	69,04	0,052	8,552	2,2	0,765
4	Chile	99032,956	449,406	2052,376	1,048	1,83	87,46	0,049	11,087	2,11	0,851
4	Colombia	109712,82	595,835	2563,832	2,53	1,32	58,76	0,17	7,646	1,71	0,767
4	Costa Rica	122845,79	871,367	1446,57	1,112	1,57	70,61	0,088	9,468	1,13	0,81
4	Ecuador	35190,648	555,481	1913,634	4,257	1,11	73,3	0,123	7,104	1,5	0,759
4	Grenada	92262,089	2308,164	1805,07	3,792	1,67	32,15	0,036	7,304	3,7	0,779
4	Guyana	68226,017	1250,112	1401,948	5,965	1,74	38,39	0,058	5,305	1,6	0,682
4	Iraq	51761,379	100,414	589,3	0,146	2,52	14,9	0,07	3,186	1,4	0,674
4	Jamaica	39040,015	466,267	859,268	2,354	1,59	20,21	0,033	9,684	1,7	0,734
4	Kyrgyzstan	28990,335	97,805	428,161	0,259	2,04	15,31	0,055	4,489	4,5	0,697
4	Lebanon	124168,3	965,937	1392,937	2,195	1,29	28,67	0,038	8,514	2,9	0,744
4	Mexico	34509,699	308,232	2319,26	1,478	1,68	58,35	0,049	6,857	1,38	0,779
4	Oman	60587,635	188,275	789,528	0,137	2,19	56,22	0,023	2,355	1,6	0,813

4	Panama	137531,57	2038,859	1724,034	2,217	1,9	56,82	0,115	7,918	2,3	0,815
4	Paraguay	70612,237	674,925	2336,127	3,878	1,15	42,17	0,042	6,378	1,3	0,728
4	Peru	81631,108	1211,125	6107,721	2,385	1,72	67,24	0,045	7,151	1,6	0,777
4	Saint Lucia	96913,791	1498,288	1713,657	4,648	1,66	27,43	0,031	9,721	1,3	0,759
4	Suriname	114512,05	1676,489	2064,894	5,069	1,54	39,09	0,018	6,933	3,1	0,738
4	Tunisia	67390,156	576,874	2165,425	1,376	1,8	51,38	0,046	8,001	2,3	0,74
5	Antigua and Barbuda	58149,664	762,557	1225,589	1,447	1,06	60,24	0,056	6,933	3,8	0,778
5	Azerbaijan	61444,279	73,851	836,321	0,978	1,2	46,18	0,022	6,018	4,7	0,756
5	Bahrain	181536,87	1539,296	799,636	0,082	1,62	68,1	0,023	2,372	2	0,852
5	Bhutan	4140,274	71,804	3,847	0	1,77	73,11	0,01	4,885	1,7	0,654
5	Brunei	36205,756	76,358	221,954	0	1,43	91,13	0,01	4,591	2,7	0,838
5	Cambodia	7132,882	1,34	177,913	0	1,32	80,96	0,022	4,412	0,8	0,594
5	Canada	74760,941	684,341	840,13	3,133	1	78,06	0,064	16,984	2,5	0,929
5	Cape Verde	97688,027	549,665	676,276	3,051	0,87	47,7	0,109	4,46	2,1	0,665
5	China	72,988	0,102	3,21	0	0,88	84,52	0,014	10,641	4,34	0,761
5	Cuba	89120,051	293,023	737,619	0,227	1,54	86,39	0,045	14,738	5,2	0,783
5	Dominican Republic	47254,383	536,113	390,187	0,196	1,29	52,51	0,02	6,981	1,6	0,756
5	Fiji	67016,355	319,289	832,873	4,905	1,38	67,81	0,006	6,224	2,3	0,743
5	Finland	66890,925	1689,379	310,722	2,214	1	73,81	0,008	21,228	3,28	0,938
5	Iran	73291,778	32,799	1553,968	0,269	1,51	62,73	0,034	5,44	1,5	0,783
5	Kuwait	110808,39	1092,183	572,94	0,231	1,49	74,87	0,02	2,345	2	0,806
5	Malaysia	85951,496	95,583	971,162	0,405	0,98	78,38	0,005	6,293	1,9	0,81
5	Morocco	28623,567	196,551	402,332	0,344	1,38	61,91	0,008	6,769	1,1	0,686
5	New Zealand	2993,402	12,466	10,151	0	1,01	76,22	0,022	15,322	2,61	0,931
5	Norway	101503,05	2154,32	258,342	1,045	1,41	72,53	0,019	16,821	3,6	0,957

5	Saudi Arabia	17903,9	158,421	252,174	0,053	1,75	66,55	0,045	3,295	2,7	0,854
5	Singapore	54290,194	208,303	154,76	0,131	1,39	86,77	0,126	12,922	2,4	0,938
5	Sri Lanka	27842,372	31,971	709,066	0,532	1,05	64,55	0,125	10,069	3,6	0,782
5	Thailand	33638,794	107,433	314,321	0,241	0,59	68,02	0,06	11,373	2,1	0,777
5	Turkey	125422,85	787,329	1002,472	1,895	1,13	61,38	0,02	8,153	2,81	0,82
5	United Arab Emirates	81465,743	294,791	220,196	0,272	1,15	92,26	0	1,144	1,2	0,89
6	Afghanistan	3993,179	2,005	185,413	0,036	1,44	9,63	0,021	2,581	0,5	0,511
6	Algeria	5130,778	16,368	144,408	0,227	1,34	12,99	0,022	6,211	1,9	0,748
6	Burkina Faso	949,617	3,509	16,421	0,133	0,5	3,59	0,031	2,409	0,4	0,452
6	Burundi	3001,527	12,006	3,101	0	0,77	0,05	0,001	2,562	0,8	0,433
6	Cameroon	4028,245	0	68,064	0	0,66	2,44	0,001	3,165	1,3	0,563
6	Central African Republic	2707,121	25,116	21,951	0,058	0,35	9,98	0,067	3,655	1	0,397
6	Djibouti	15100,823	84,529	188,586	0	1,55	8,35	0,065	4,213	1,4	0,524
6	Egypt	3874,894	11,376	213,297	0,269	1,22	23,83	0,136	5,159	1,6	0,707
6	Equatorial Guinea	10730,462	67,69	123,458	0,197	0,78	14,08	0,011	2,846	2,1	0,592
6	Ethiopia	3902,051	10,861	60,996	0,121	0,62	1,35	0,001	3,526	0,3	0,485
6	Gabon	19924,707	62,438	131,208	0,125	1,02	9,73	0,075	4,45	6,3	0,703
6	Ghana	4859,113	8,765	42,544	0,063	1,19	9,2	0,047	3,385	0,9	0,611
6	Guatemala	36145,741	126,874	887,842	0,587	1,61	27,7	0,082	4,694	0,6	0,663
6	Guinea	2646,393	7,261	30,377	0,074	1,15	9,6	0,006	3,135	0,3	0,477
6	Honduras	38354,788	50,028	1039,154	0,114	0,56	43,55	0,027	4,652	0,7	0,634
6	Kenya	5801,708	11,746	100,39	0,133	0,53	8,45	0,078	2,686	1,4	0,601
6	Laos	17257,192	114,431	68,705	0,832	0,89	51,97	0,057	4,029	1,5	0,613

6	Libya	57499,578	103,552	844,143	1,252	1,15	13,61	0,074	4,424	3,7	0,724
6	Malawi	4235,563	13,124	126,783	0,393	0,72	3,87	0,001	2,979	1,3	0,483
6	Mali	1402,828	14,439	33,612	0,164	1,24	2,62	0,038	2,519	0,1	0,434
6	Myanmar	9736,231	2,74	352,261	0,023	0,9	32,04	0,09	5,732	0,9	0,583
6	Niger	337,116	1,205	11,699	0,045	1,13	4,02	0,085	2,553	0,3	0,394
6	Pakistan	5976,028	21,273	128,961	0,032	1,96	34,81	0,12	4,495	0,6	0,557
6	Sao Tome and Principe	25299,511	266,062	295,482	0,64	1,44	27,49	0,033	2,886	2,9	0,625
6	South Africa	59368,08	72,442	1558,426	2,194	0,69	27	0,017	5,344	2,32	0,709
6	Sudan	1181,936	12,673	75,485	0,102	1,01	3,15	0,018	3,548	0,8	0,51
6	Timor	14782,625	0,638	90,782	0	0,74	41,48	0,019	3,556	5,9	0,606
6	Togo	4252,768	23,556	31,021	0,051	1,01	12,46	0,011	2,839	0,7	0,515
6	Yemen	342,695	1,073	65,397	0,033	1,44	0,98	0,009	2,922	0,7	0,47
6	Zimbabwe	15033,424	23,229	348,923	0,483	0,49	21,53	0,037	2,822	1,7	0,571

Приложение 4 – проверка гипотез на нормальное распределение совокупностей по кластерам

Кластер 0

```
In [48]: for i, col in enumerate(clean_col_names):
         testing_cluster=clusters_initial[0]
         k2, p = stats.normaltest(testing_cluster[col])
         print("p = {:g}".format(p))
         if p < alpha: # null hypothesis: x comes from a normal distribution
             print(col+": The null hypothesis can be rejected")
         else:
             print(col+": The null hypothesis cannot be rejected")

p = 0.686158
total_cases_per_million: The null hypothesis cannot be rejected
p = 0.206571
new_cases_smoothed_per_million: The null hypothesis cannot be rejected
p = 0.901769
total_deaths_per_million: The null hypothesis cannot be rejected
p = 0.135521
new_deaths_smoothed_per_million: The null hypothesis cannot be rejected
p = 0.754766
reproduction_rate: The null hypothesis cannot be rejected
p = 0.0941101
people_fully_vaccinated_per_hundred: The null hypothesis cannot be rejected
p = 0.000161037
new_people_vaccinated_smoothed_per_hundred: The null hypothesis can be rejected
p = 0.127186
aged_65_older: The null hypothesis cannot be rejected
p = 0.0505696
hospital_beds_per_thousand: The null hypothesis cannot be rejected
p = 0.267464
human_development_index: The null hypothesis cannot be rejected
```

Кластер 4

```
In [52]: for i, col in enumerate(clean_col_names):
         testing_cluster=clusters_initial[4]
         k2, p = stats.normaltest(testing_cluster[col])
         print("p = {:g}".format(p))
         if p < alpha: # null hypothesis: x comes from a normal distribution
             print(col+": The null hypothesis can be rejected")
         else:
             print(col+": The null hypothesis cannot be rejected")

p = 0.223321
total_cases_per_million: The null hypothesis cannot be rejected
p = 0.238822
new_cases_smoothed_per_million: The null hypothesis cannot be rejected
p = 1.36233e-06
total_deaths_per_million: The null hypothesis can be rejected
p = 0.295665
new_deaths_smoothed_per_million: The null hypothesis cannot be rejected
p = 0.236021
reproduction_rate: The null hypothesis cannot be rejected
p = 0.589046
people_fully_vaccinated_per_hundred: The null hypothesis cannot be rejected
p = 0.00145083
new_people_vaccinated_smoothed_per_hundred: The null hypothesis can be rejected
p = 0.62083
aged_65_older: The null hypothesis cannot be rejected
p = 0.0595824
hospital_beds_per_thousand: The null hypothesis cannot be rejected
p = 0.984338
human_development_index: The null hypothesis cannot be rejected
```

Кластер 5

```
In [56]: for i, col in enumerate(clean_col_names):
         testing_cluster=clusters_initial[5]
         k2, p = stats.normaltest(testing_cluster[col])
         print("p = {:.g}".format(p))
         if p < alpha: # null hypothesis: x comes from a normal distribution
             print(col+": The null hypothesis can be rejected")
         else:
             print(col+": The null hypothesis cannot be rejected")

p = 0.150398
total_cases_per_million: The null hypothesis cannot be rejected
p = 0.00125703
new_cases_smoothed_per_million: The null hypothesis can be rejected
p = 0.354101
total_deaths_per_million: The null hypothesis cannot be rejected
p = 9.35302e-05
new_deaths_smoothed_per_million: The null hypothesis can be rejected
p = 0.925475
reproduction_rate: The null hypothesis cannot be rejected
p = 0.847934
people_fully_vaccinated_per_hundred: The null hypothesis cannot be rejected
p = 0.00253349
new_people_vaccinated_smoothed_per_hundred: The null hypothesis can be rejected
p = 0.191913
aged_65_older: The null hypothesis cannot be rejected
p = 0.309956
hospital_beds_per_thousand: The null hypothesis cannot be rejected
p = 0.894642
human_development_index: The null hypothesis cannot be rejected
```

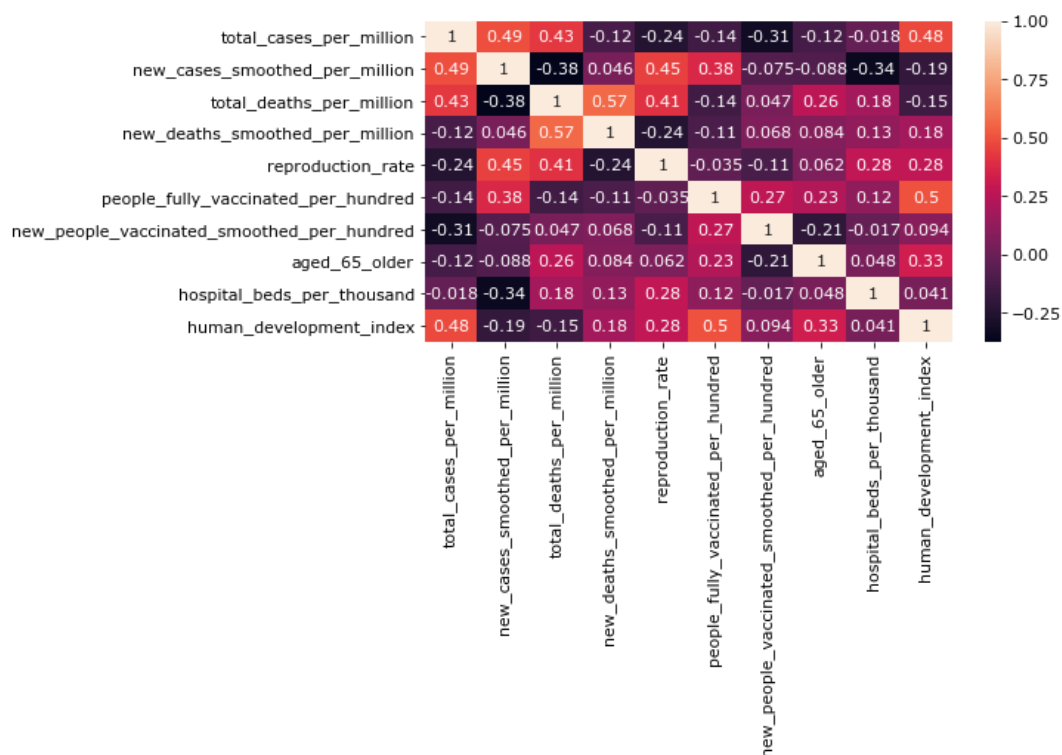
Кластер 6

```
In [61]: for i, col in enumerate(clean_col_names):
         testing_cluster=clusters_initial[6]
         k2, p = stats.normaltest(testing_cluster[col])
         print("p = {:.g}".format(p))
         if p < alpha: # null hypothesis: x comes from a normal distribution
             print(col+": The null hypothesis can be rejected")
         else:
             print(col+": The null hypothesis cannot be rejected")

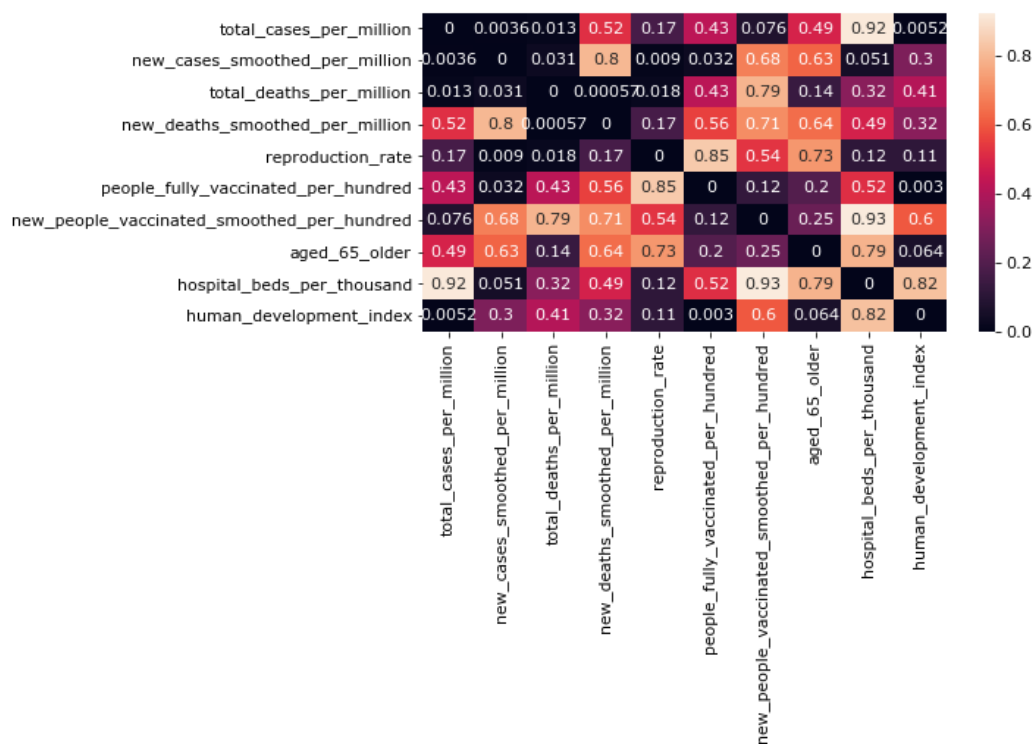
p = 4.59603e-05
total_cases_per_million: The null hypothesis can be rejected
p = 3.0075e-08
new_cases_smoothed_per_million: The null hypothesis can be rejected
p = 4.47436e-07
total_deaths_per_million: The null hypothesis can be rejected
p = 1.88939e-09
new_deaths_smoothed_per_million: The null hypothesis can be rejected
p = 0.586964
reproduction_rate: The null hypothesis cannot be rejected
p = 0.0601578
people_fully_vaccinated_per_hundred: The null hypothesis cannot be rejected
p = 0.1486
new_people_vaccinated_smoothed_per_hundred: The null hypothesis cannot be rejected
p = 0.167966
aged_65_older: The null hypothesis cannot be rejected
p = 3.12341e-06
hospital_beds_per_thousand: The null hypothesis can be rejected
p = 0.379192
human_development_index: The null hypothesis cannot be rejected
```

Приложение 5 – корреляционный анализ по однородным кластерам

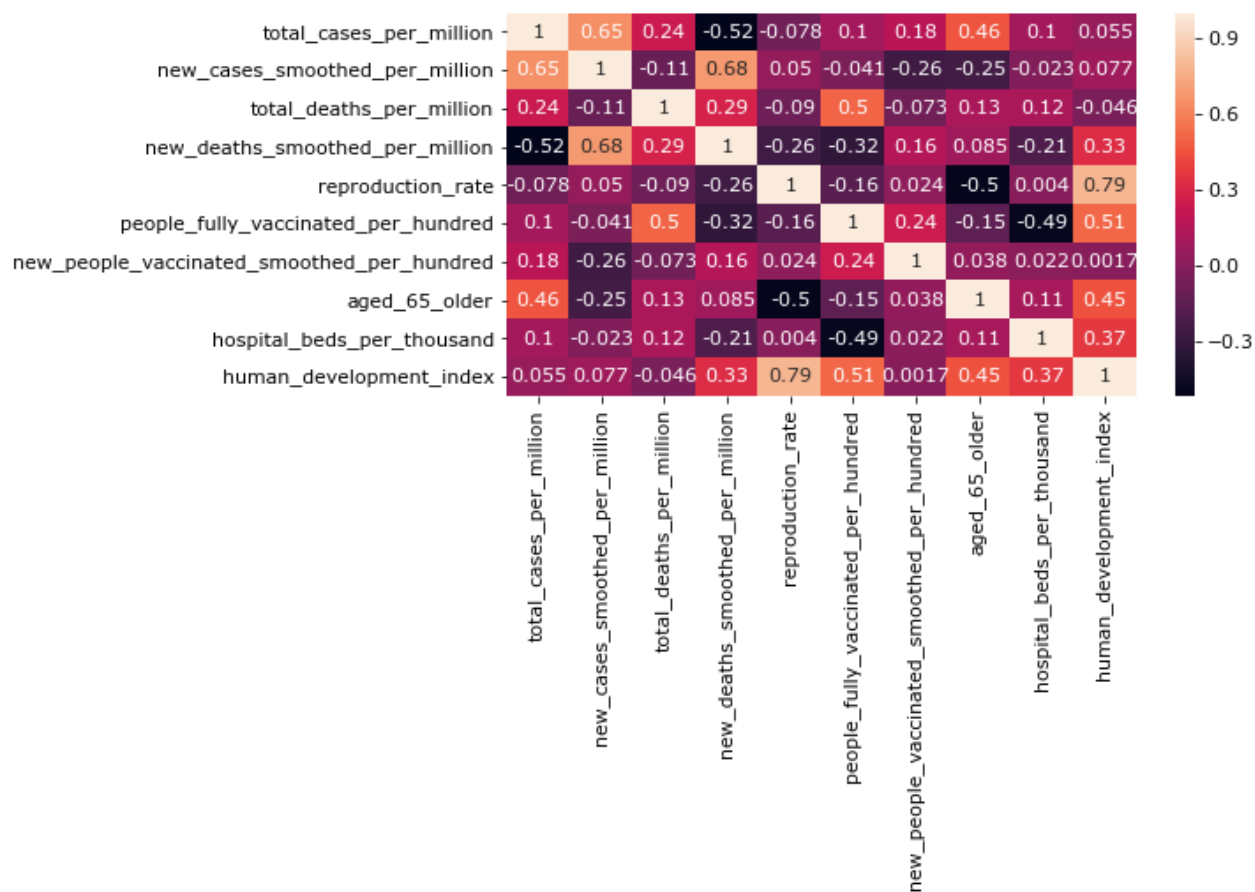
Кластер 0 – оценка частных коэффициентов корреляции



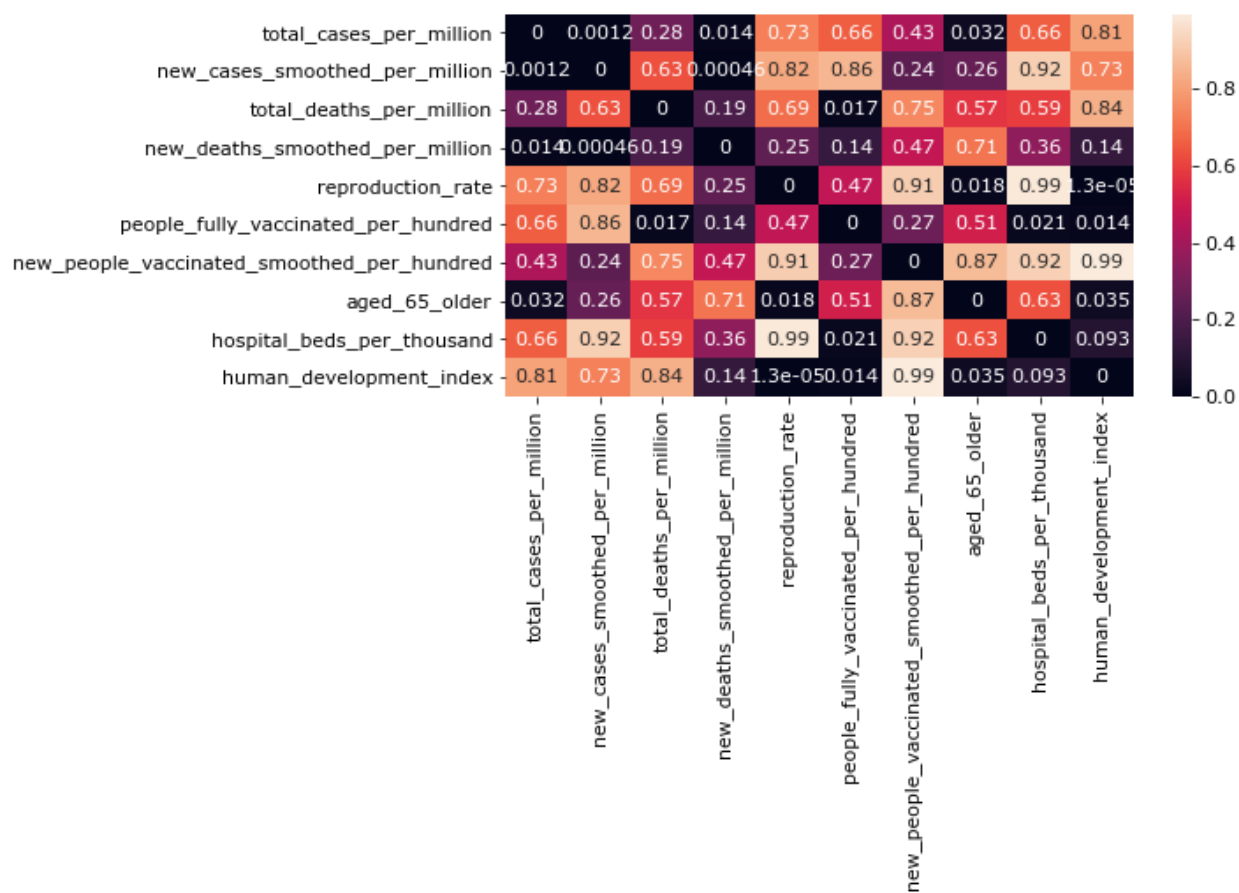
Кластер 0 – оценка значимости частных коэффициентов корреляции



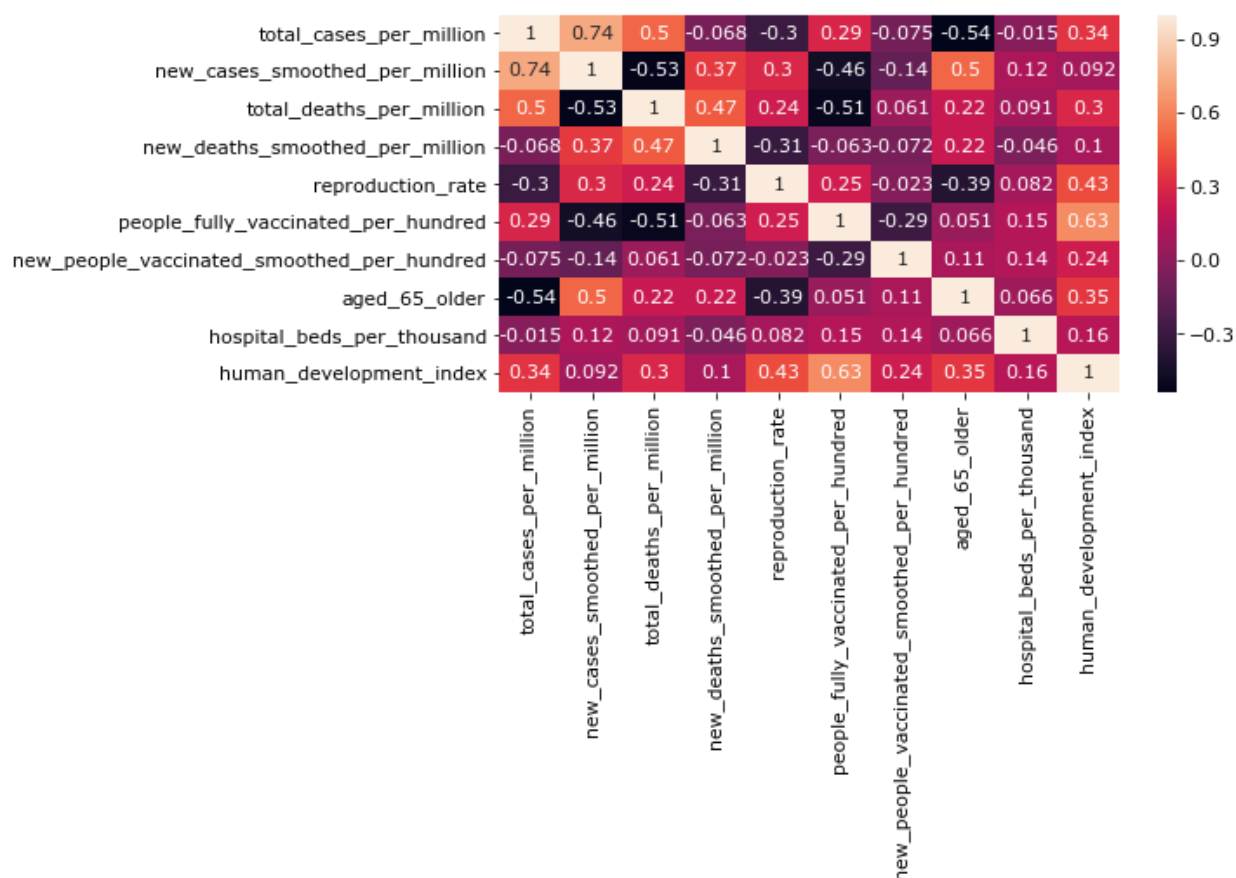
Кластер 4 – оценка частных коэффициентов корреляции



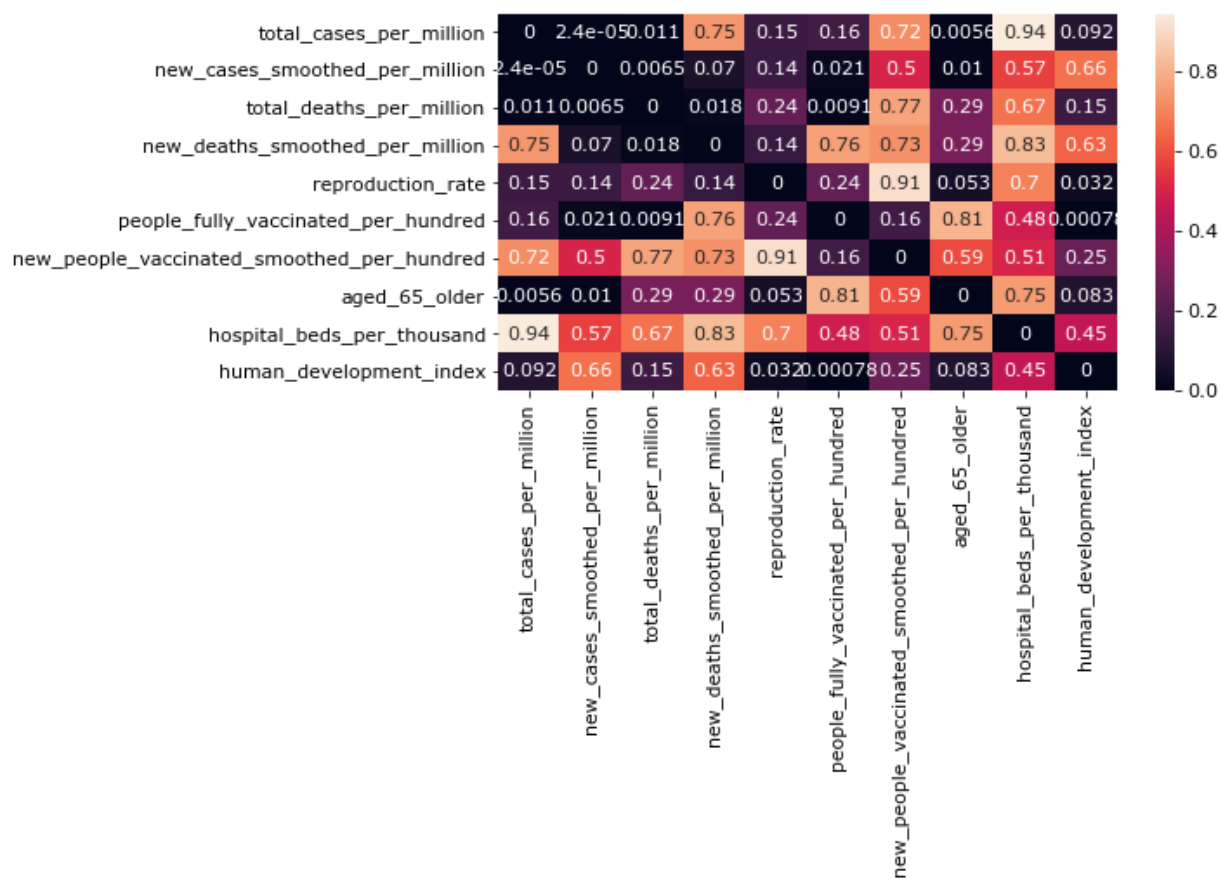
Кластер 4 – оценка значимости частных коэффициентов корреляции



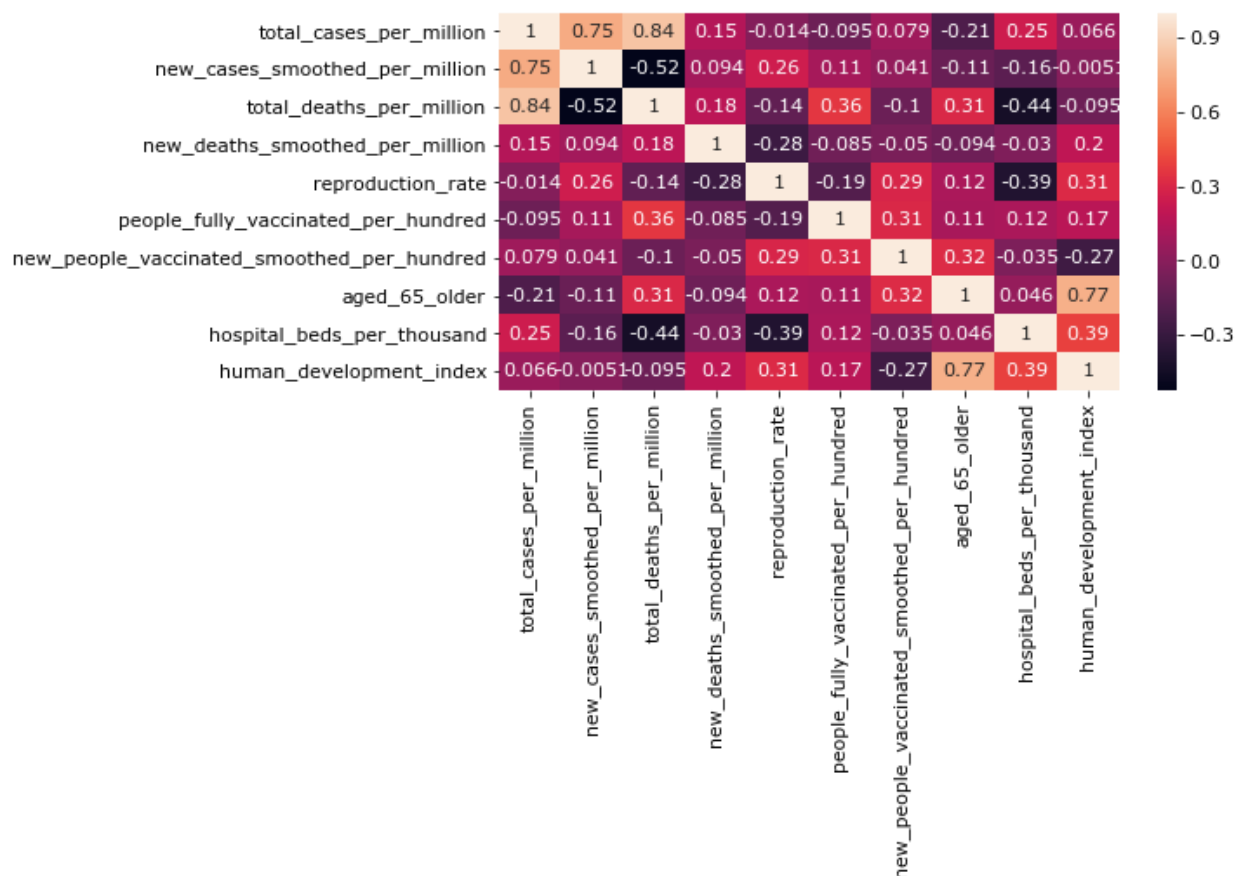
Кластер 5 – оценка частных коэффициентов корреляции



Кластер 5 – оценка значимости частных коэффициентов корреляции



Кластер 6 – оценка частных коэффициентов корреляции



Кластер 6 – оценка значимости частных коэффициентов корреляции

