

Исходные данные и гипотеза

В работе дополнительно используются следующие данные.

1. Количество учеников разрезе муниципальных образований. Данные получены из государственной информационной системы о государственных и муниципальных учреждения [1](#), так же из реестра образовательных учреждений Рособрназдора [2](#);
2. Доходы муниципальных образований. Источником данных по доходам является портал "Электронный бюджет" [3](#);
3. Зона покрытия 4g, загруженная с инвесткарты РФ [4](#).

С учетом этого собран следующий датасет [5](#).

Показатель	Название	Описание
territory_id	Код муниципального образования	
c114_sum	Ученики 1-4 классов	Количество учеников 1 - 4 классов в соответствии с планом финансово-хозяйственной деятельности учреждения. В данном показателей, как и в последующих двух использован ряд допущений и упрощений: отсекались очевидно нереалистичные цифры, не делалось различий между общим числом обучающихся и числом обучающихся по специальным программам.
c159_sum	Ученики 5-9 классов	
c11011_sum	Ученики 10-11 классов	
quality_ratio	Коэффициент качества данных	Доля учреждений, по которым данные загружены в ГИС ГМУ.
full_area	Площадь территории	Площадь территории, кв. км.
cell_area_ratio	Доля покрытия территории 4G	
tax	Налоговые доходы	Общая сумма налоговых доходов местного бюджета, млн. руб.
tax_pers	В том числе НДФЛ	В том числе НДФЛ (15% по бюджетному кодексу), млн. руб.
transfert	Трансферты	Сумма трансфертов в местный бюджет, млн. руб.
tax_ratio	Доля налогов	Доля налогов в общем бюджете.
pers_tax_ratio	Доля НДФЛ	Доля НДФЛ в общем бюджете.
pop_total	Общее население	Общее население, чел
pop_work_ratio	Доля трудоспособного населения	В интервале 18 - 60 лет.
migration_total	Общая миграция	Общая миграция, чел.
total_cons_val	Общие траты	Общие траты по всем категориям (этот и последующие три показателя по данным sber index)
rest_cons_val	Траты на общественное питание	
rest_ratio	Доля трат на общественное питание	
rest_inc_ratio	Изменение трат на общественное питание	С января 2023 по декабрь 2024 с учетом 17% накопленной инфляции

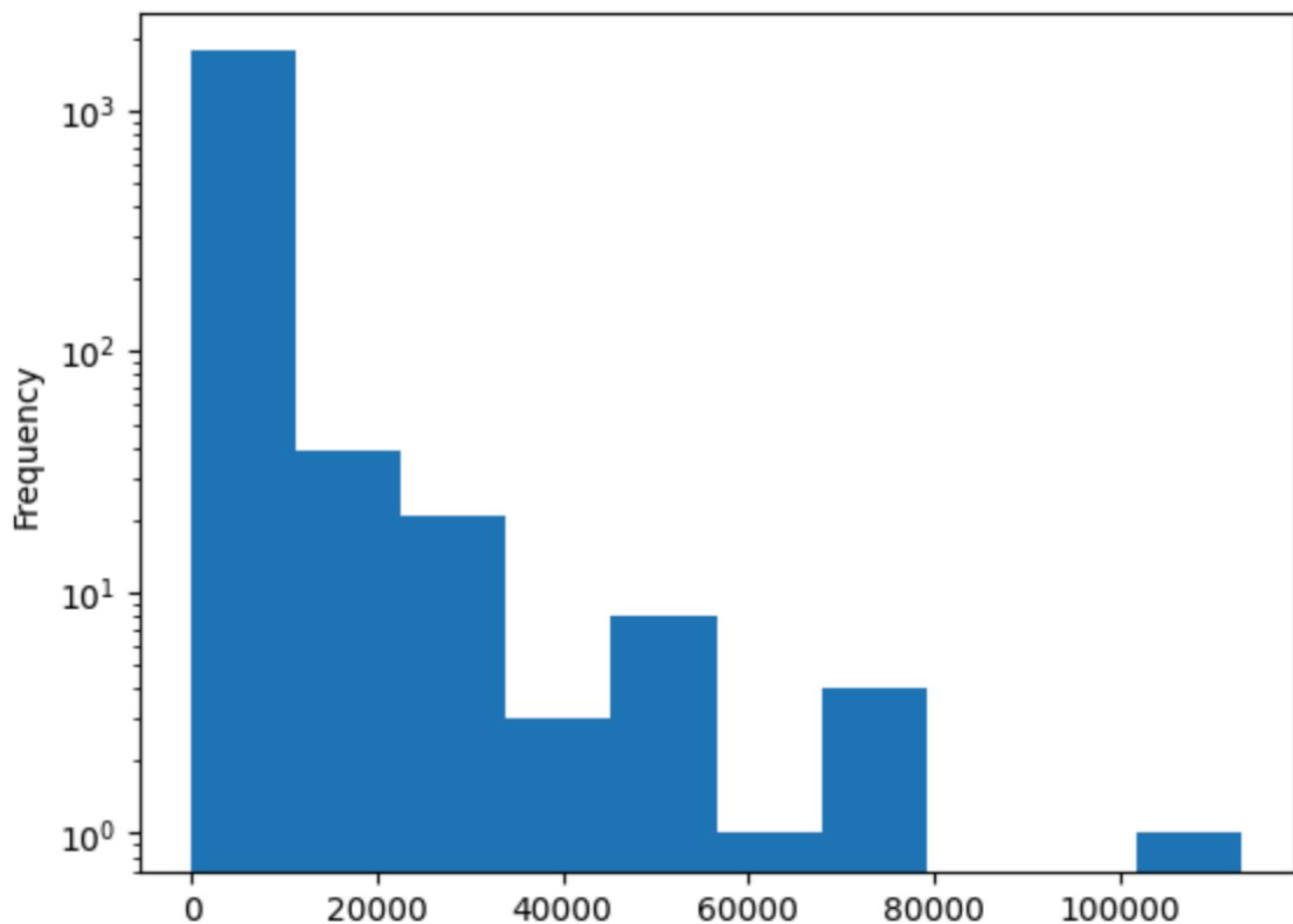
Динамику в работе рассматривать не будем, так как это требует дополнительных усилий по обеспечению сопоставимости данных.

Выдвинем следующую гипотезу, подтверждение которой постараемся найти в собранных данных. Будем считать регион привлекательным с точки зрения миграции, если:

- в регионе много школьников 1 - 4 классов (а значит и их молодых родителей);
- в регионе высокая доля налоговых доходов, а особенно НДФЛ (так как оплата идет по месту регистрации работодателя);
- большая площадь территории с 4g покрытием (города, дороги), то есть большая площадь урбанизированных территорий;
- высокая доля трат на общественное питание, а также высокий индекс роста по этому показателю. Считаем, что эти показатели характеризуют степень урбанизации.

Анализ исходных данных

Предварительно посмотрим на отдельные показатели. Начнем с учеником 1 - 4 классов. Вот так выглядит гистограмма.



Видим, что в основной массе муниципалитетов не более 10 000 учащихся 1 - 4 классов. Шкала логарифмическая, и все что больше 10 000 даже не набирает сотни муниципалитетов. В цифрах получается следующая картина: 87 муниципалитетов, где больше 10 000 учащихся, против 2 507 где менее 10 000 учащихся 1 - 4 классов.

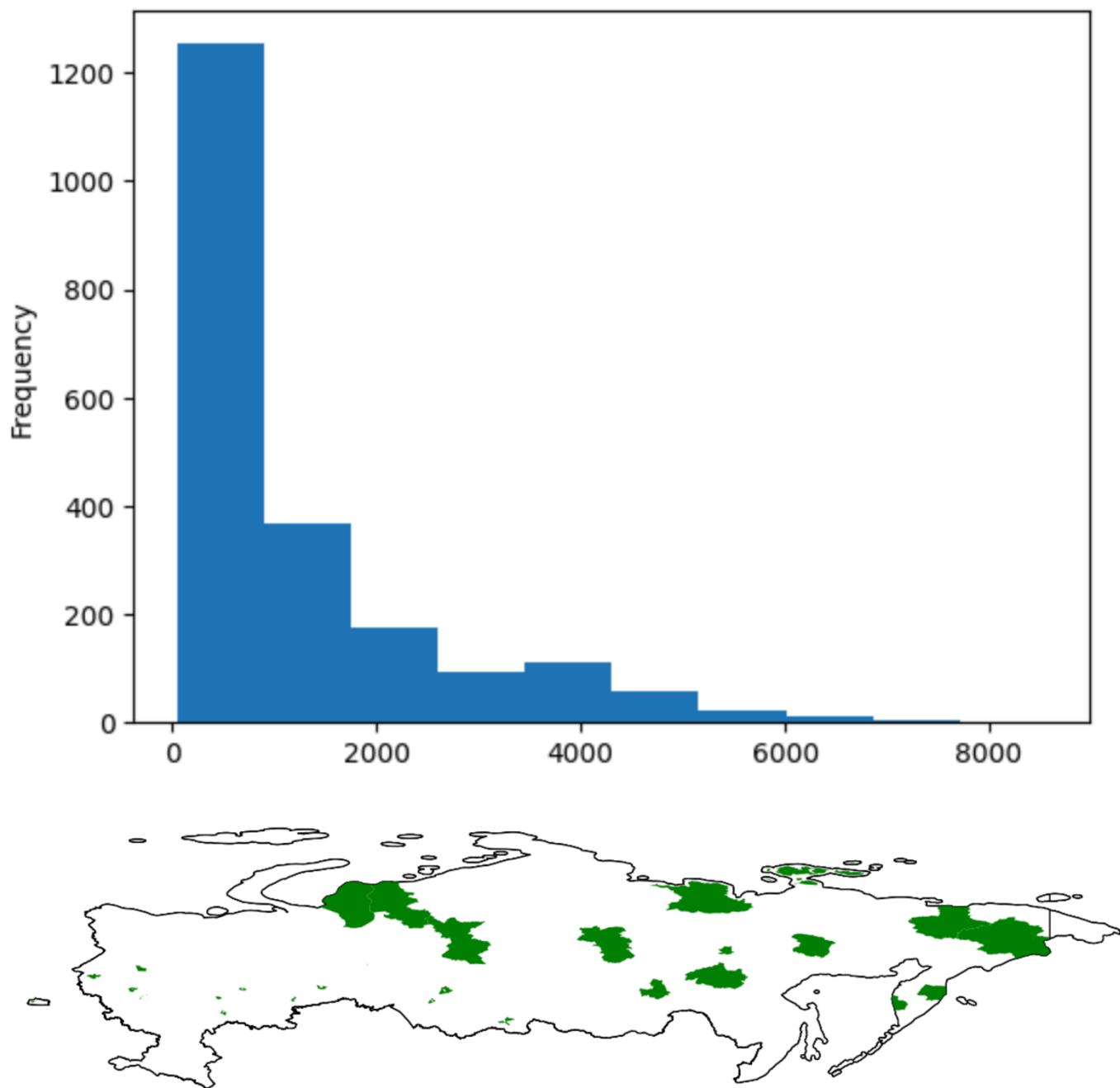
Для наглядности отобразим эти МО на карте, а топ-10 выведем в таблицу. Заметим, что все это преимущественно крупные города, расположенные в центральной и южной части РФ.



municipal_district_name	cl14_sum	full_area	cell_area_ratio	tax	tax_ratio	pop_total	migration_total	rest_ratio	rest_inc_ratio
городской округ Пермский	113107	2852,23	0,72	33491,11	0,56	1176541	2485	0,06	0,28
городской округ город Краснодар	78795,62	1682,91	0,98	38006,79	0,4	1408263	16054		
городской округ город Казань	77414	2018,39	0,98	33215,63	0,57	1509742	2679	0,06	0,25
городской округ город Екатеринбург	71045,31	3814,62	0,74	41993,98	0,47	1827616	-2807	0,06	0,19
городской округ город Новосибирск	69582	1524,27	0,99	49902,24	0,5	1882661	2040	0,06	0,33
городской округ город Уфа	61752	2144,93	0,92	23692,05	0,47	1362766	6864	0,06	0,21
городской округ город Красноярск	54950,05	1211,81	0,96	36140,23	0,57	1366057	9325	0,05	0,24
городской округ город Омск	53687	1763,15	0,99	19888,17	0,44	1279744	-1645	0,05	0,14
городской округ город Тюмень	52946	2366,48	0,87	24698,41	0,5	978772	2894	0,06	0,21
городской округ город Воронеж	51087	1556,24	0,67	16823,4	0,34	1216022	-992		

Аналогичным образом выведем на карту топ-50 МО по росту расходов на общественное питание, так как это единственный показатель в нашем датасете, который хоть как-то отражает динамику. Ну и заметим, что по тратам нет данных по Краснодару и Воронежу, так как, видимо не все терминалы привязанные к этим регионам физически там же и находятся.

Так как данные сильно скошены, будем брать только те МО где абсолютное значение трат на общественное питание больше чем 0.8 квантиль, равный 1 955, чтобы избавиться от эффекта низкой базы.

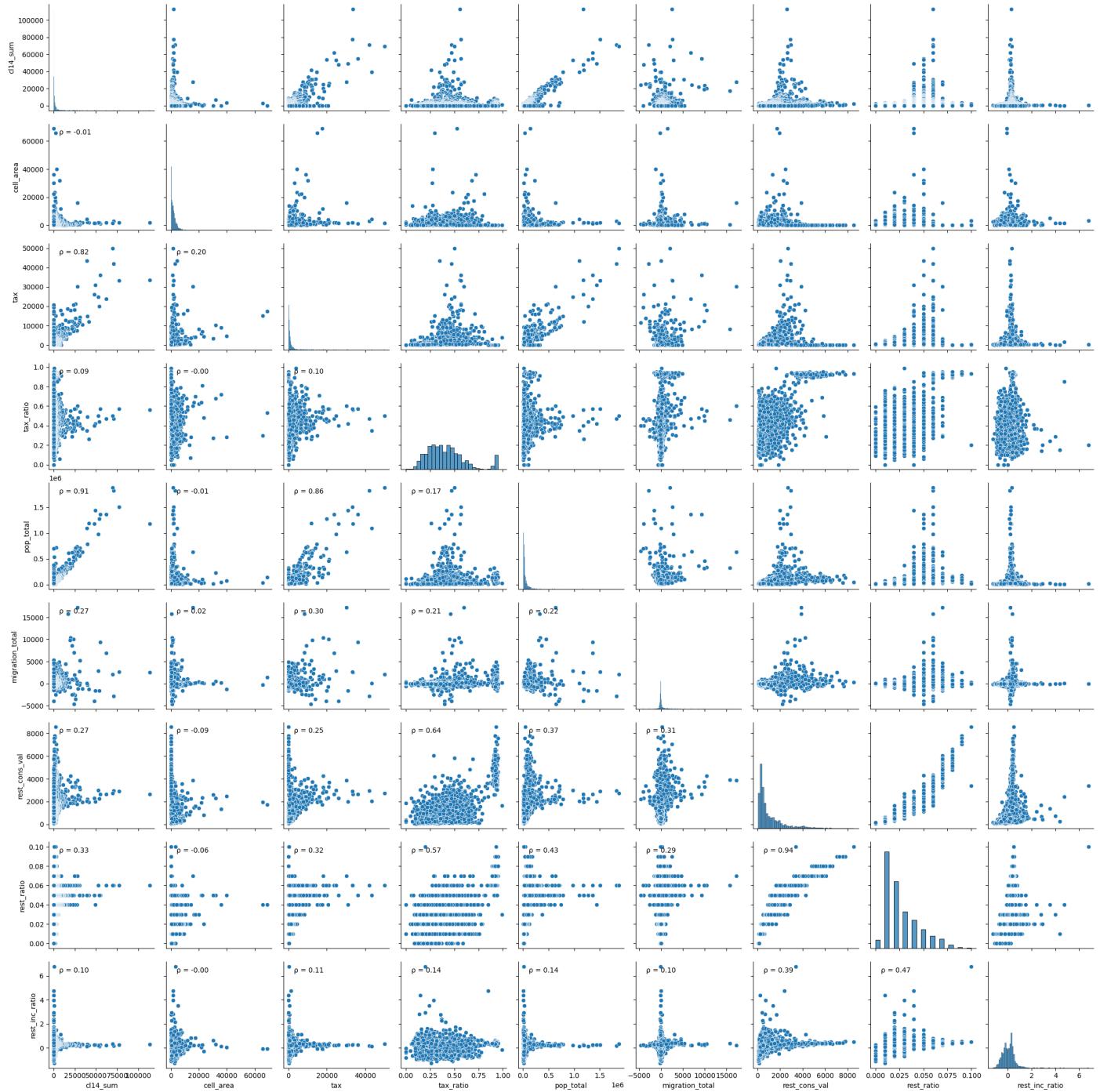


В основном на карте покрасились северные и восточные МО, что, все таки, скорее объясняется эффектом низкой базы, так как если анализировать регионы с абсолютным значениями больше 0.9 квантиля, то там будут в основном Москва и Санкт-Петербург.

Проверка гипотезы

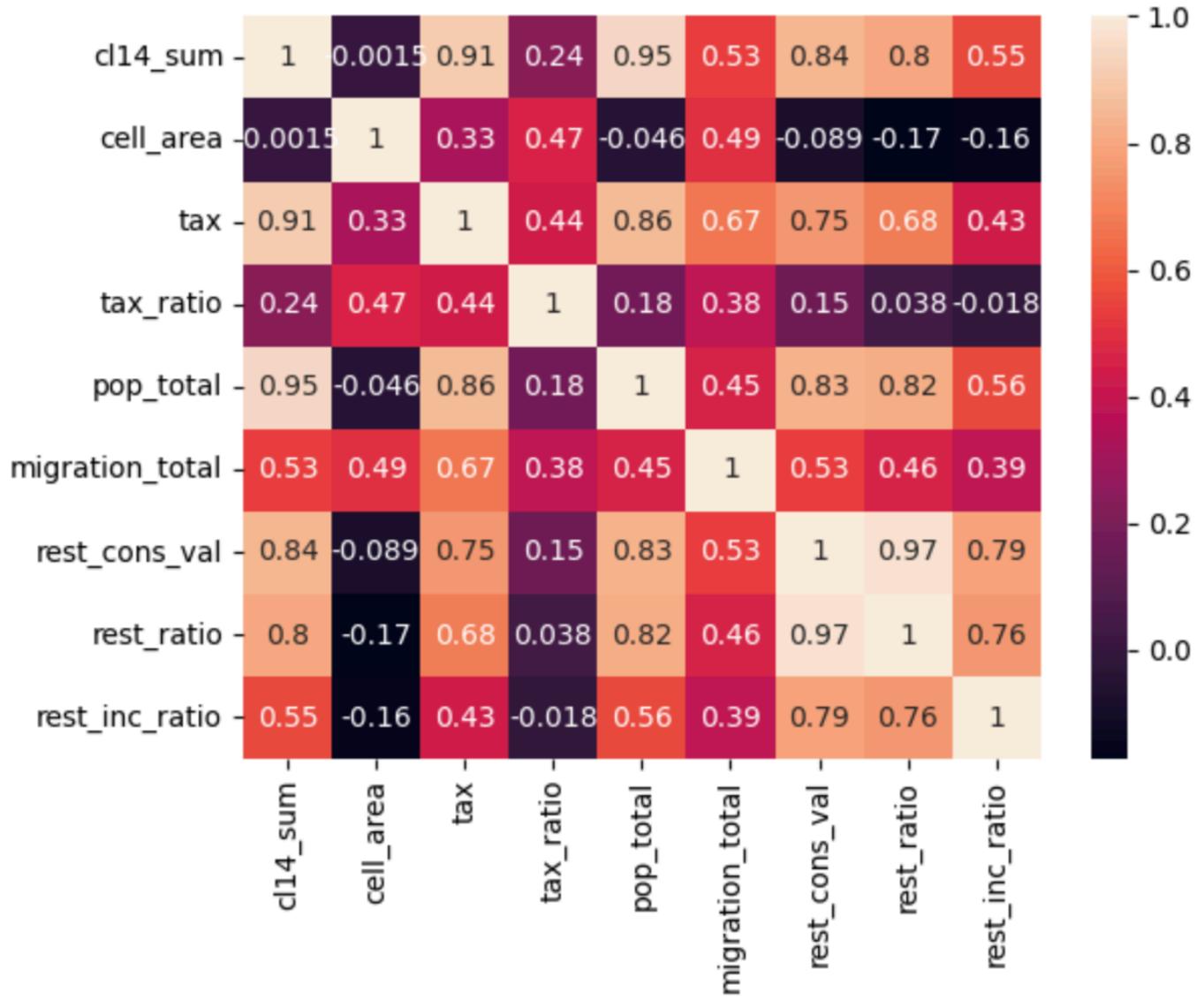
Вернемся к нашей исходной гипотезе. Построим попарные графики показателей.

На масштабе все страны значимые корреляции обнаружены только между населением и НДФЛ, что вполне ожидаемо. МО слишком разные по географическим, социально-экономическим признакам. Попробуем посмотреть корреляции на примере одного более-менее однородного субъекта, например Самарской области (без учета Самары). Результат опишем в заключении.



Заключение

При анализе Самарской области (без учета Самары) видны уже более явные корреляции. Наиболее интересная для нас – между количеством учеников 1 – 4 классов и общими тратами на общественное питание = 0.84. Интерпретировать эту связь можно следующим образом: чем больше детей – тем больше молодых родителей, более склонных тратить на общественное питание.



Рассчитаем линейную регрессию и посмотрим какие МО лучше конвертируют демографический потенциал в траты на общественное питание (на графике подпишем названия этих МО).
Зафиксируем, что это преимущественно городские округа.

Вместе с тем наша исходная гипотеза о том, что миграция определяется некоторыми из собранных нами показателей не нашла своего подтверждения. Уровень корреляции по Росстата миграции крайне слабый. Можем предположить, что это связано с несовершенством регистрационной системы учета МВД.

