

Университет ИТМО

Практическая работа №4
по дисциплине «Визуализация и моделирование»

Автор: Михайлов Иван Юрьевич
Поток: Вим 1.2
Группа: К3221
Факультет: ИКТ
Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

Датасет - Recorded Crime Data at the Police Force Area Level.

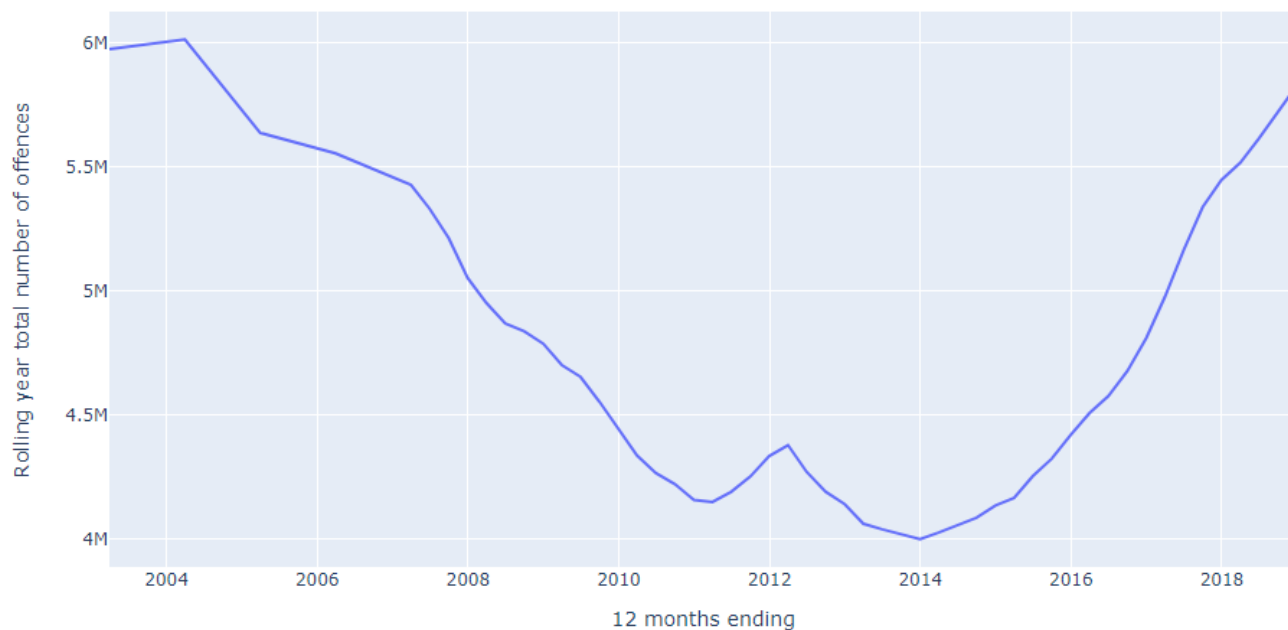
Датасет представляет из себя список преступлений в Англии и Уэльсе за период с 2003 по 2018 гг.

Описание данных		
Название столбца	Данные	Тип данных
12 months ending PFA Region	Заккрытие финансового года	date
Offence	Название района полиции Регион, в котором произошло уголовное преступление	text
Rolling year total number of offences	Название уголовного преступления	text
	Количество случаев данного правонарушения за последний год	digit

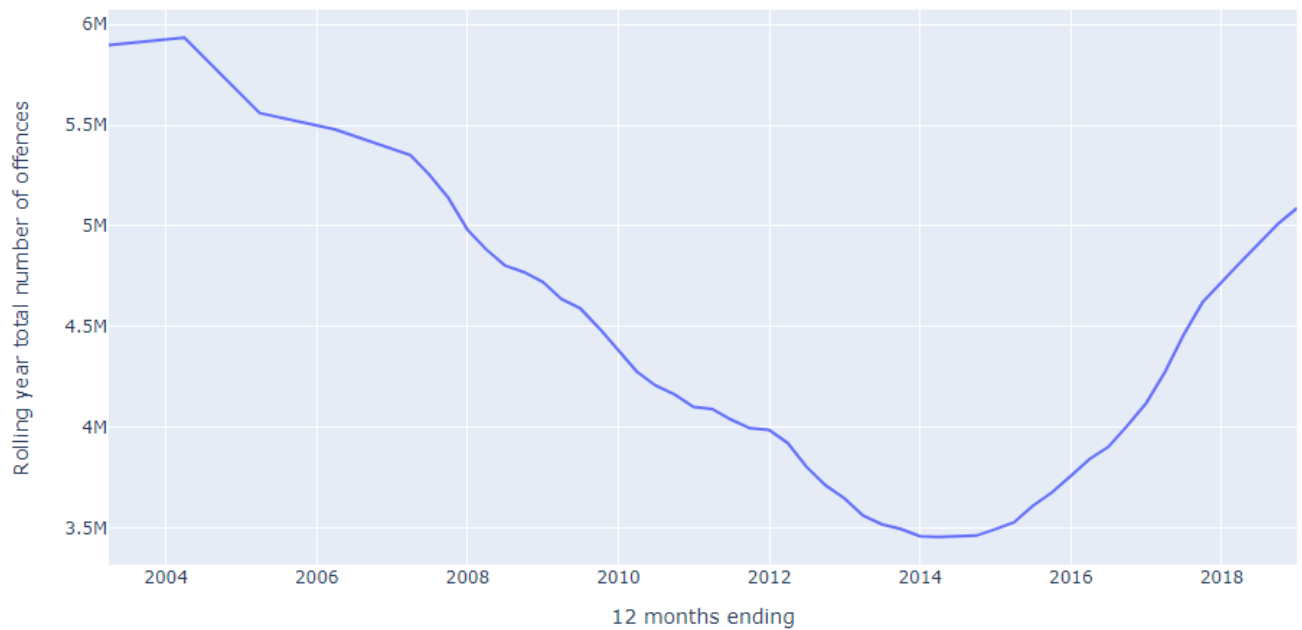
Обработка гипотез

1. Гипотеза: Удаление данных организаций не сильно отразится на общую картину распределения преступлений.

Сначала покажем общее распределение преступлений с течением времени до обработки.



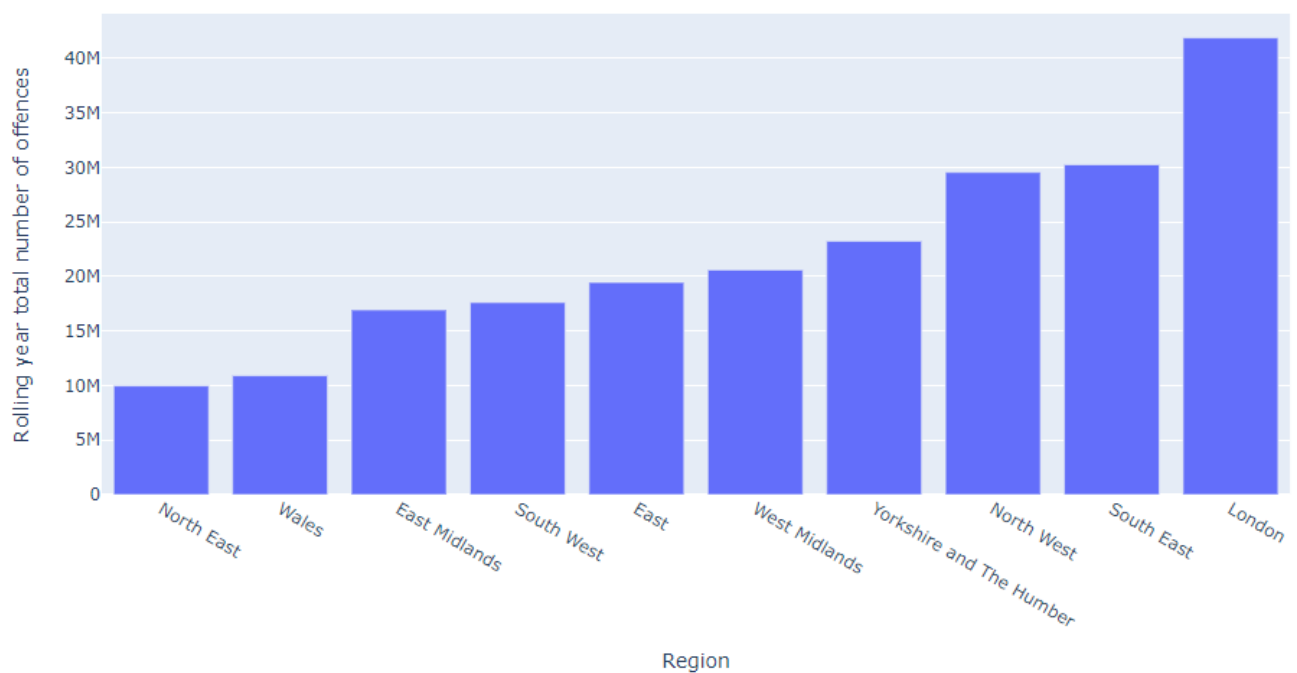
После удаления неудобных данных распределение приняло следующий вид:



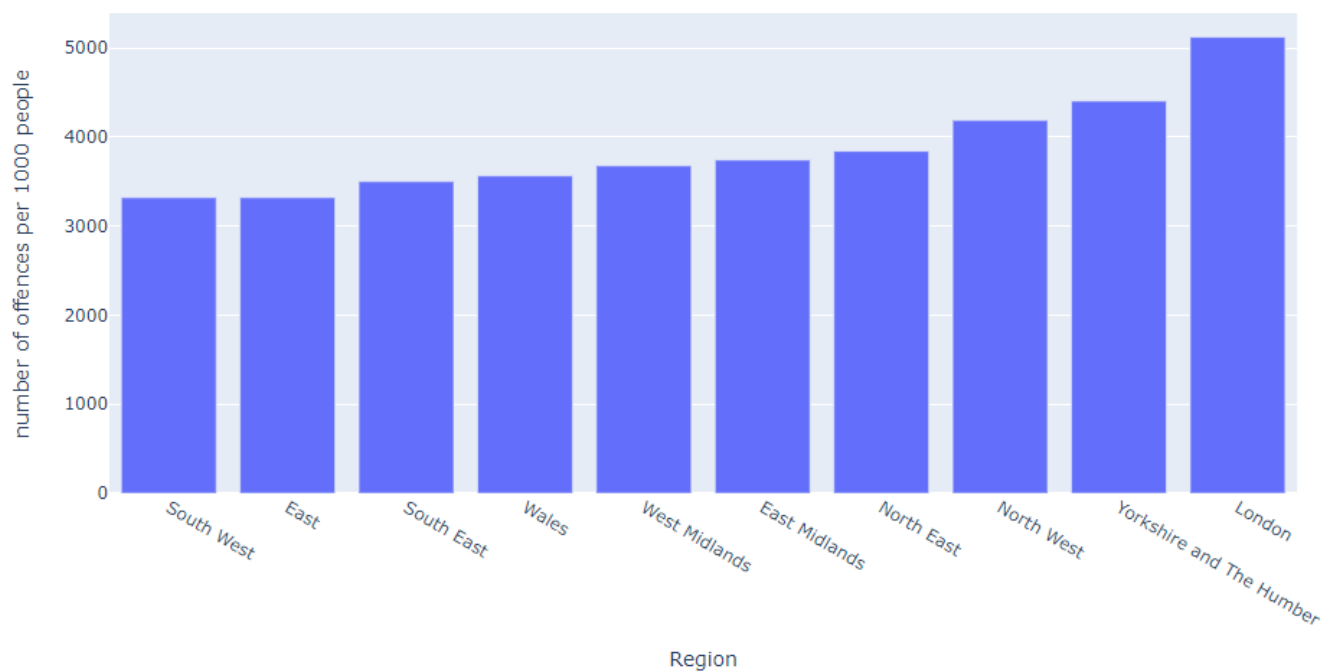
Вывод: По всей видимости, мы нашли объяснение двойному минимуму распределения преступлений. Он появлялся только из-за внезапного появления в середине 2011 года новых строк с данными от частных организаций. Однако возрастание преступности, начиная с 2014 года никуда не делся. В целом можно заключить, что гипотеза почти подтвердилась.

2. Гипотеза: Нормировка преступлений гораздо точнее покажет опасность отдельных районов.

Сначала покажем общее распределение преступлений по регионам до обработки.



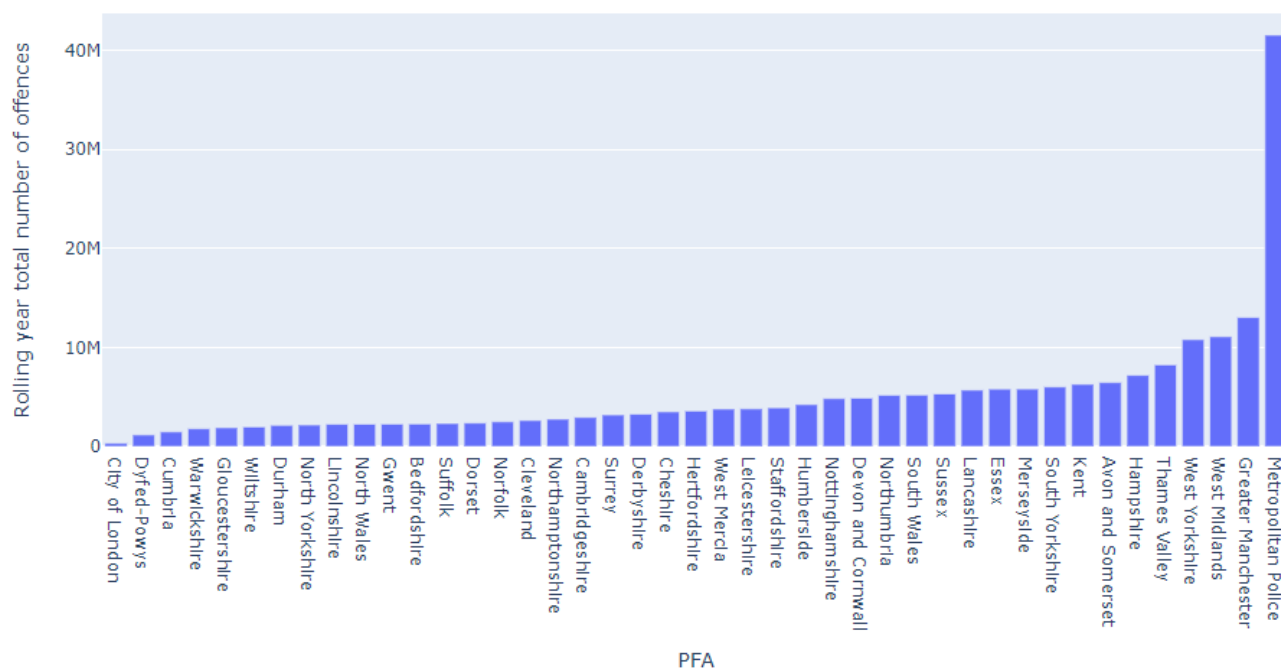
Теперь покажем распределение преступлений с учётом количества жителей в регионах:



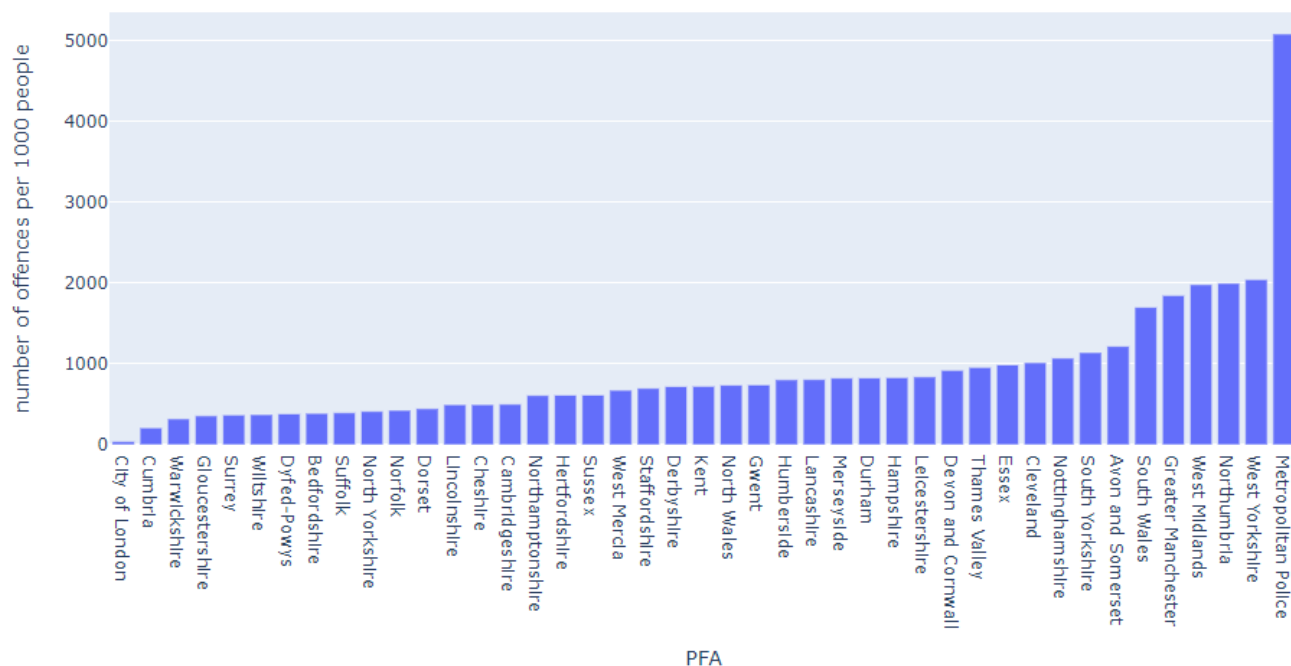
Вывод: Хотя Лондон и оставил за собой криминальное лидерство, можно заметить существенные изменения в распределении преступлений. Так Северо-Восточный округ с последнего 10 места переместился аж на 4, а Юго-Восточный со 2 на 8 место. Но, что самое главное, теперь распределение приблизилось к равномерному, то есть на самом деле в стране нет такого сильного криминогенного перекоса. Гипотеза полностью подтвердилась.

3. Гипотеза: В зоне ответственности столичной полиции криминальная обстановка не настолько сильно отличается относительно других районов, если сделать поправку на количество жителей, которое они обхватывают.

Покажем, какое распределение мы видели во 2 задании.



С поправкой на население распределение принимает следующий вид:



Вывод: Видим, что некоторые районы сместились, например Northumbria с 15 места поднялась на 3. Но общая картина распределения практически не изменилась и столичная полиция лидирует с большим отрывом. Следовательно гипотеза не подтвердилась.

- Гипотеза. У разных полицейских участков сильно отличается количество подконтрольных им регионов. Так как полицейских участков намного больше, чем регионов, корректнее было бы сформулировать гипотезу наоборот: в разных регионах количество полицейских отделов сильно разнится.

Списки полицейских участков по регионам:

South West - ['Avon and Somerset', 'Devon and Cornwall', 'Dorset', 'Gloucestershire', 'Wiltshire']

East - ['Bedfordshire', 'Cambridgeshire', 'Essex', 'Hertfordshire', 'Norfolk', 'Suffolk']

North West - ['Cheshire', 'Cumbria', 'Greater Manchester', 'Lancashire', 'Merseyside']

London - ['City of London', 'Metropolitan Police']

North East - ['Cleveland', 'Durham', 'Northumbria']

East Midlands - ['Derbyshire', 'Leicestershire', 'Lincolnshire', 'Northamptonshire', 'Nottinghamshire']

Wales - ['Dyfed-Powys', 'Gwent', 'North Wales', 'South Wales']

South East - ['Hampshire', 'Kent', 'Surrey', 'Sussex', 'Thames Valley']

Yorkshire and The Humber - ['Humberside', 'North Yorkshire', 'South Yorkshire', 'West Yorkshire']

West Midlands - ['Staffordshire', 'Warwickshire', 'West Mercia', 'West Midlands']

Вывод: Видим, что количество полицейских отделов разнится от 2 в Лондоне до 6 в Восточном округе. Из этой картины мы понимаем, почему на графике преступлений по районам столичная полиция имеет такой выброс. В Лондоне работают всего 2 участка, 1 из которых вообще почти не имеет (или не выкладывает) преступлений. Получается всего один участок на самый крупный регион страны, логично, что в его юрисдикции совершается большое количество преступлений. Можно заключить, что гипотеза скорее подтвердилась.