

Университет ИТМО

Практическая работа №3
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	Проблема	Способы решения	
12 months ending PFA	Закрытие финансового года	date	---	---	
Region	Название района полиции	text	Странные данные частных организаций	Удаление	
Offence	Регион, в котором произошло уголовное преступление	text	Странные данные частных организаций	Удаление	
Rolling year total number of offences	Название уголовного преступления	text	---	---	
	Количество случаев данного правонарушения за последний год	digit	Отрицательные данные. Ненормированность данных	Отрицательные взять по модулю. Нормировать количество преступлений на 1000 жителей	

Предобработка данных

Датасет был расширен дополнительной колонкой с количеством жителей в регионах в прошедшем году и количество преступлений было нормировано на 1000 жителей для более верного отображения опасности регионов и загруженности полицейских районов.

Колонка с преступлениями была взята по модулю, так как отрицательных значений количества быть не может и это скорее всего просто опечатка из-за человеческого фактора.

Удалены строки, в которых в качестве регионов и типов преступлений были указаны кампании, предоставляющие свою статистику преступлений, потому что, во-первых, они представлены с середины исследуемого периода, во-вторых, строчек с ними в 30 раз меньше, чем с нормальными данными, а так же они не распределены по географическим регионам, то есть их почти никак не использовать в исследованиях.

Гипотезы

1. Удаление данных организаций не сильно отразится на общую картину распределения преступлений.
2. Нормировка преступлений гораздо точнее покажет опасность отдельных районов.

3. В зоне ответственности столичной полиции криминальная обстановка не настолько сильно отличается относительно других районов, если сделать поправку на количество жителей, которое они обхватывают.
4. У разных полицейских участков сильно отличается количество подконтрольных им регионов.
5. Возможно, уровень преступности как-то коррелирует с широтами, в которых находится регион, например, чем южнее, тем больше в среднем совершается в год преступлений.