



АНАЛИЗ ДАННЫХ О НАСЕЛЕННЫХ ПУНКТАХ РОССИИ

1/ ВВОДНЫЙ ПРОЕКТ

PYTHON

PANDAS

ОБРАБОТКА ДАННЫХ

РЕГУЛЯРНЫЕ ВЫРАЖЕНИЯ

01

ОБРАБОТКА ДАННЫХ

02

АНАЛИЗ ТИПОВ НАСЕЛЕННЫХ ПУНКТОВ

03

АНАЛИЗ НАЗВАНИЙ НАСЕЛЕННЫХ ПУНКТОВ

Обработан столбец name и дубликаты

Столбец “name” разделен на “type” и “only_name” на основе классификатора ОКАТО и ручного поиска

```
okato_type = [  
    "поселок городского типа",  
    "рабочий поселок",  
    "курортный поселок",  
    "дачный поселок",  
    "сельский поселок",  
    "дачный поселковый совет",  
    "поселок сельского типа",  
    "городской поселок",  
    "поселок при станции",  
    "поселок при железнодорожной станции",  
    "поселок при станции",  
    "поселок станции",  
    "населенный пункт",  
    "железнодорожная станция",  
    "железнодорожная будка",
```

Найдены скрытые дубликаты, причина их возникновения - буквы типа ïïег' и ошибки в словах

35

Общее число
найденных дубликтов

Список уникальных типов населенных пунктов

```
array(['город', 'поселок', 'хутор', 'село', 'поселок городского типа',  
      'поселок при станции', 'деревня', 'поселок станции', 'проселок',  
      'железнодорожная станция', 'населенный пункт',  
      'железнодорожный разъезд', 'разъезд', 'станция', 'рабочий поселок',  
      'местечко', 'починок', 'кордон', 'железнодорожная будка',  
      'слобода', 'железнодорожная площадка', 'остановочный пункт',  
      'сельский поселок', 'погост', 'участок',  
      'поселок при железнодорожной станции', 'слободка',  
      'дачный поселок', 'заимка', 'железнодорожный блокпост', 'казарма',  
      'курортный поселок', 'выселок', 'высел', 'железнодорожная казарма',  
      'остановочная платформа', 'обгонная платформа',  
      'поселок сельского типа', 'аул', 'железнодорожная платформа',  
      'улус', 'аал', 'блок-пост', 'городской поселок', 'обгонный пункт',  
      'железнодорожный путевой пост',  
      'железнодорожный остановочный пункт', 'арбан', 'станица', 'кутан',  
      'отделение', 'мыс', 'остров', 'молочнотоварная ферма совхоза',  
      'экопоселение', 'выселки', 'поселение', 'полустанок', 'колхоз',  
      'городок', 'переезд', 'седо', 'территория'], dtype=object)
```

Доля каждого типа населенного пункта

proportion	
type	
деревня	57.29799%
село	18.73486%
поселок	16.45560%
хутор	3.58062%
город	0.70688%
населенный пункт	0.66817%
поселок городского типа	0.54140%
станция	0.27663%
рабочий поселок	0.22418%

Вывод: Больше всего преобладают такие типы, как деревня, село, поселок что логично - этих типов много и все они разнообразны. Города же более большие объекты. Остальные названия либо редкие, либо используются в определенных местах.

Сумма населения по типам населенных пунктов в относительном и абсолютном выражении

	population	percent
type		
город	102678139.0	67.80209
село	21768502.0	14.37453
поселок	8478293.0	5.59852
деревня	6557748.0	4.33032
поселок городского типа	4923998.0	3.25149
...
мыс	0.0	0.00000
молочнотоварная ферма совхоза	0.0	0.00000
полустанок	0.0	0.00000
седо	0.0	0.00000
территория	0.0	0.00000

Региональная принадлежность для любых трех, неизвестных ранее типов населенных пунктов

type	region_name	
аал	Республика Хакасия	46
арбан	Республика Тыва	12
аул	Карачаево-Черкесская Республика	55
	Краснодарский край	14
	Новосибирская область	12
	Омская область	38
	Оренбургская область	3
	Республика Адыгея (Адыгея)	42
	Республика Ингушетия	2
	Самарская область	2
	Ставропольский край	26
	Тюменская область	12

63
Число уникальных типов населенных пунктов

Самые короткие и длинные
название населенных пунктов

	only_name	len
108703	Ий	2
108301	Яр	2
84551	Юм	2
33548	Ям	2
130439	Ыб	2
...
101243	1-е отделение государственной селекционной ста...	46
155798	Всесоюзного электротехнического института имен...	50
55644	Опытного хозяйства центральной торфо-болотной ...	56
111286	Учебного хозяйства нальчикского сельскохозяйст...	59
149923	Северо-кавказская зональная опытная станция вс...	103

Топ 10 самых популярных
названий населенных пунктов

	count
only_name	
Александровка	331
Ивановка	280
Михайловка	267
Каменка	252
Никольское	240
Николаевка	238
Березовка	233
Сосновка	216
Первомайский	202
Заречье	198

“К”

Буква, на которую чаще всего
начинаются названия. Таких
названий 19542 единиц.

22

Число населенных пунктов на “Ы”

82842

Число уникальных названий
населенных пунктов

23 буквы

Населенный пункт с самым длинным
названием из одного слова без пробелов
и дефисов - “Верхненовокутлумбетьево”

Наиболее популярные населенные пункты, длиной в два символа, и их региональная принадлежность

	region_name	only_name
7238	Калужская область	Ям
23228	Рязанская область	Ур
33548	Тверская область	Ям
37987	Тверская область	Ям
41404	Владимирская область	Ям
...
149439	Краснодарский край	Ея
157093	Республика Северная Осетия — Алания	Ир
158112	Ярославская область	№1
159578	Республика Ингушетия	Гу
160085	Новгородская область	Ям

Наиболее популярные населенные пункты, длиной в три символа, и их региональная принадлежность

	region_name	only_name
2460	Брянская область	Май
2750	Брянская область	Ужа
3371	Брянская область	Ким
3515	Брянская область	Май
3524	Брянская область	Чет
...
159257	Волгоградская область	299
159381	Архангельская область	Бор
159476	Чеченская Республика	Хой
159505	Свердловская область	Лая
159747	Республика Ингушетия	Ний

Некоторые названия населенных пунктов по выделенным категориям

Животные: ['Бобрава' 'Боброво' 'Соколовка' 'Котеневка' 'Котово' 'Быковка' 'Барсук' 'Котов' 'Котляров' 'Волков']

Здания: ['Домановка' 'Хуторище' 'Хлевище' 'Домнино' 'Хуторцы' 'Слобода' 'Теремошка' 'Слобода-селецкая' 'Дом отдыха' 'Доманичи']

Эмоции: ['Надежда' 'Страховка' 'Гневково' 'Любовь' 'Страхово' 'Страхинь' 'Страхиново' 'Тоскаево' 'Радость' 'Любовь труда']

Еда: ['Масловка' 'Кашары' 'Сырцево' 'Чайки' 'Мясоедово' 'Сыроватский' 'Рыбалкин' 'Пирогово' 'Орехово' 'Медвежье']

Особенные_места: ['Борисполье' 'Луговка' 'Гора-подол' 'Степь' 'Вершина' 'Борисов' 'Луганка' 'Боровая' 'Береговой' 'Озеровский']

Жители: ['Кузнецовка' 'Кузнецы' 'Кузнец' 'Земледелец' 'Рабочий' 'Кузнецово' 'Воинка' 'Людиново' 'Кузнецовский' 'Дворяниново']

Столько раз встречаются приведенные прилагательные в названиях населенных пунктов

379

Большой

316

Малый

592

Новый

681

Красный

Работа выполнена Муравьевым Андреем
Все права защищены. 2026©