

# Отчет по кейсу

«Анализ демографических характеристик по регионам в России»

1. Постановка задачи

2. Описание датасета

3. Ход работы

3.1. Гипотеза

3.2. Первичная обработка датасета

3.3. Анализ датасета

4. Выводы

## 1. Постановка задачи

Данный кейс предлагает. Выполнить анализ демографических характеристик по регионам в России с выявлением закономерностей и изменений по числу населения за период с 1990 по 2017-ый год

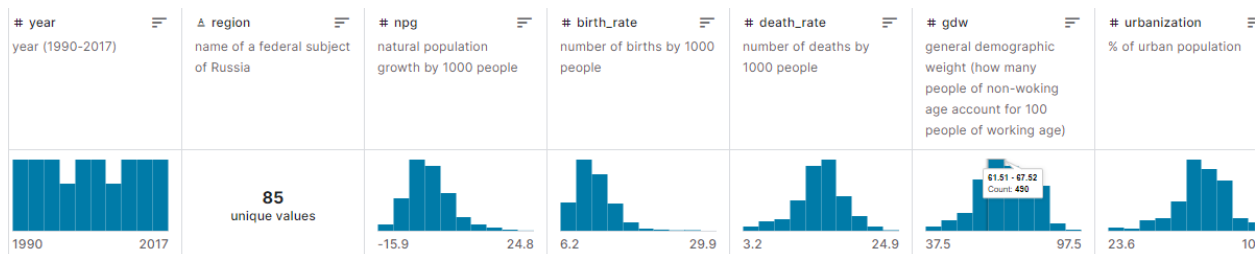
"Исследование демографической ситуации в регионах позволяет выявить основные тенденции и закономерности изменения численности, состава и структуры населения, определить факторы, влияющие на эти процессы, и разработать меры для улучшения демографической ситуации и повышения качества жизни населения.

Исходя из этого, появляется необходимость исследования демографической ситуации в России, и мы выдвинули такие задачи:

1. Выполнить анализ демографических характеристик по регионам в России.
2. Визуализировать полученный результат.
3. Сравнить итоговый результат с аналогичными работами для определения и выявления сходств/различий в найденных закономерностях.

## 2. Описание датасета

Датасета демографические данные России (1990-2017) представляет натуральный набор данных, созданный на основе реальных данных, связанный с мониторингом демографических показателей в различных регионах России. Общая схема приведена на рисунке:



В данный датасет входит набор данных содержит 2380 строк и 7 столбцов. Ключи для столбцов:

- "year" - год (1990–2017)
- "region" - название субъекта Российской Федерации.
- "birth\_rate" - количество рождений на 1000 человек
- "death\_rate" - количество смертей на 1000 человек
- "gdw" - общий демографический вес (нерабочие на 100 рабочих).
- "urbanization" - % городского населения.
- "npg" - естественный прирост населения на 1000 человек.

## 3. Ход работы

### 3.1. Гипотеза

Исходя из описания датасета, наиболее полезной информацией для анализа стала рождаемость и смертность в ходе чего мы сделали гипотезу:

В период с 1990 по 2017 год в регионах России с более высоким уровнем смертности (death\_rate) наблюдалось более заметное снижение коэффициента рождаемости в последующие годы, чем в регионах с более низким уровнем смертности, что может свидетельствовать о долгосрочном демографическом кризисном периоде.

### 3.2. Первичная обработка датасета

Загрузили датасет в ruspark вывели датасет и общую информацию (кол-во строк/столбцов), проверили на наличие NULL значений. На данном изображении показаны результаты:

```
Количество строк: 2380
Количество столбцов: 7

Схема:
root
|-- year: integer (nullable = true)
|-- region: string (nullable = true)
|-- npg: double (nullable = true)
|-- birth_rate: double (nullable = true)
|-- death_rate: double (nullable = true)
|-- gdw: double (nullable = true)
|-- urbanization: double (nullable = true)

ВЫВОДИМ весь датасет

df.show(2380)
```

year	region	npg	birth_rate	death_rate	gdw	urbanization
1990	Republic of Adygea	1.9	14.2	12.3	84.66	52.42
1990	Altai Krai	1.8	12.9	11.1	80.24	58.07
1990	Amur Oblast	7.6	16.2	8.6	69.55	68.37
1990	Arkhangelsk Oblast	3.7	13.5	9.8	73.26	73.63
1990	Astrakhan Oblast	4.7	15.1	10.4	77.05	68.01
1990	Republic of Bashk...	6.5	16.2	9.7	80.53	64.22
1990	Belgorod Oblast	0.0	12.9	12.9	84.17	63.26
1990	Bryansk Oblast	0.1	13.0	12.9	86.48	67.49
1990	Republic of Buryatia	9.2	18.3	9.1	79.47	62.16
1990	Vladimir Oblast	-0.4	12.1	12.5	77.78	79.31
1990	Volgograd Oblast	1.3	13.0	11.7	77.3	75.76
1990	Vologda Oblast	1.4	13.4	12.0	82.16	65.48
1990	Voronezh Oblast	-2.4	11.5	13.9	83.78	60.94
1990	Republic of Dagestan	19.9	26.1	6.2	94.26	43.49
1990	Jewish Autonomous...	8.2	17.8	9.6	76.11	65.01

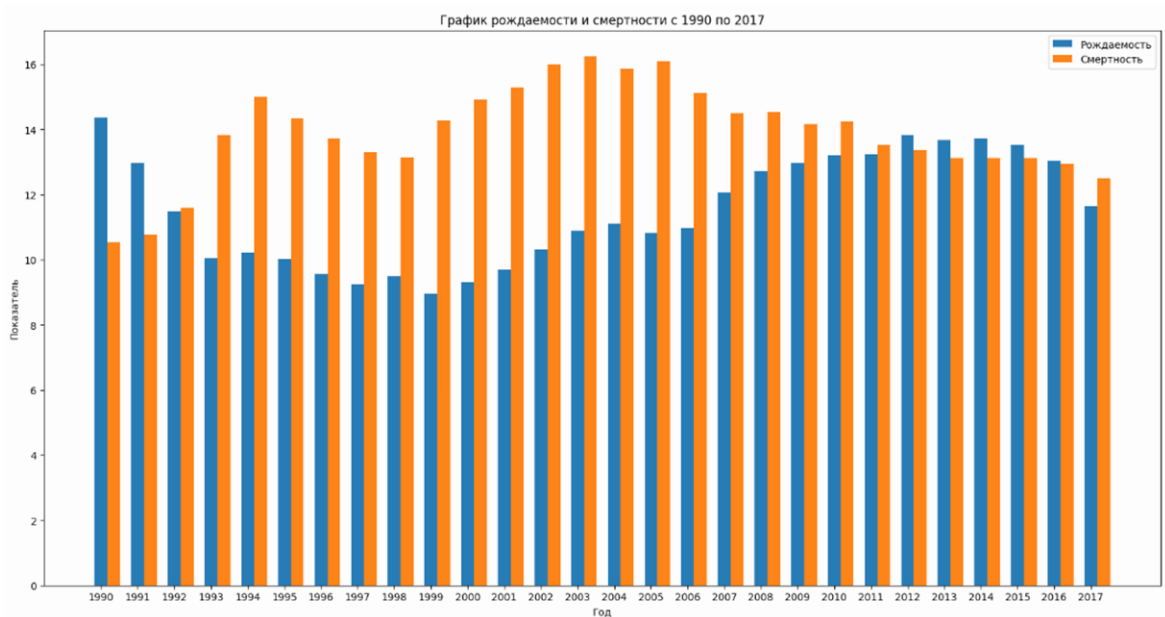
Далее удалили из набора данных NULL строки для дальнейшей обработки данных, вывели результат

```
Количество строк: 2315
Количество столбцов: 7
```

year	region	npg	birth_rate	death_rate	gdw	urbanization
1990	Republic of Adygea	1.9	14.2	12.3	84.66	52.42
1990	Altai Krai	1.8	12.9	11.1	80.24	58.07
1990	Amur Oblast	7.6	16.2	8.6	69.55	68.37
1990	Arkhangelsk Oblast	3.7	13.5	9.8	73.26	73.63
1990	Astrakhan Oblast	4.7	15.1	10.4	77.05	68.01
1990	Republic of Bashk...	6.5	16.2	9.7	80.53	64.22
1990	Belgorod Oblast	0.0	12.9	12.9	84.17	63.26
1990	Bryansk Oblast	0.1	13.0	12.9	86.48	67.49
1990	Republic of Buryatia	9.2	18.3	9.1	79.47	62.16
1990	Vladimir Oblast	-0.4	12.1	12.5	77.78	79.31
1990	Volgograd Oblast	1.3	13.0	11.7	77.3	75.76
1990	Vologda Oblast	1.4	13.4	12.0	82.16	65.48
1990	Voronezh Oblast	-2.4	11.5	13.9	83.78	60.94

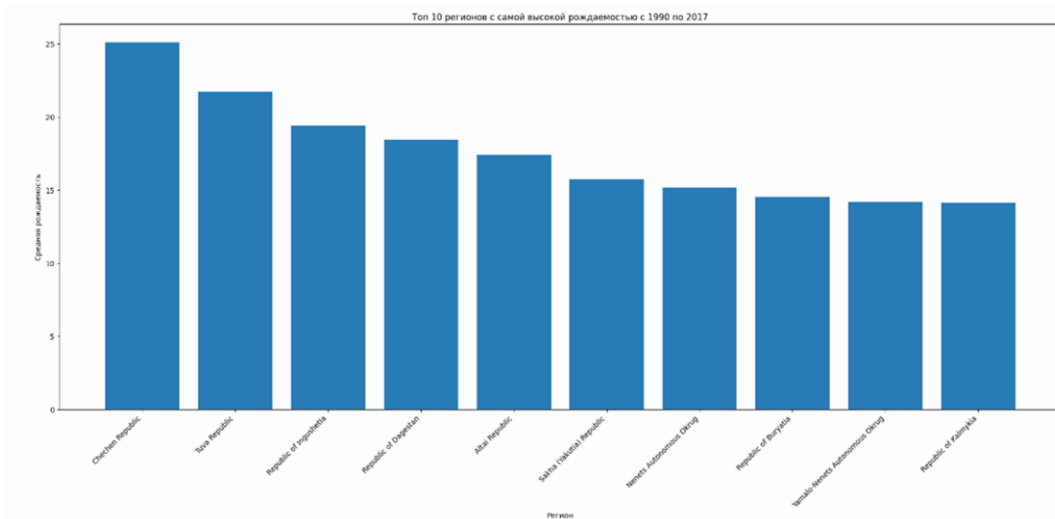
### 3.3. Анализ датасета

Анализируя датасет, мы вывели следующие гистограммы:



В данной гистограммы выполнили группировку по годам и вывели средние показатели по рождаемости/смертности за каждый год

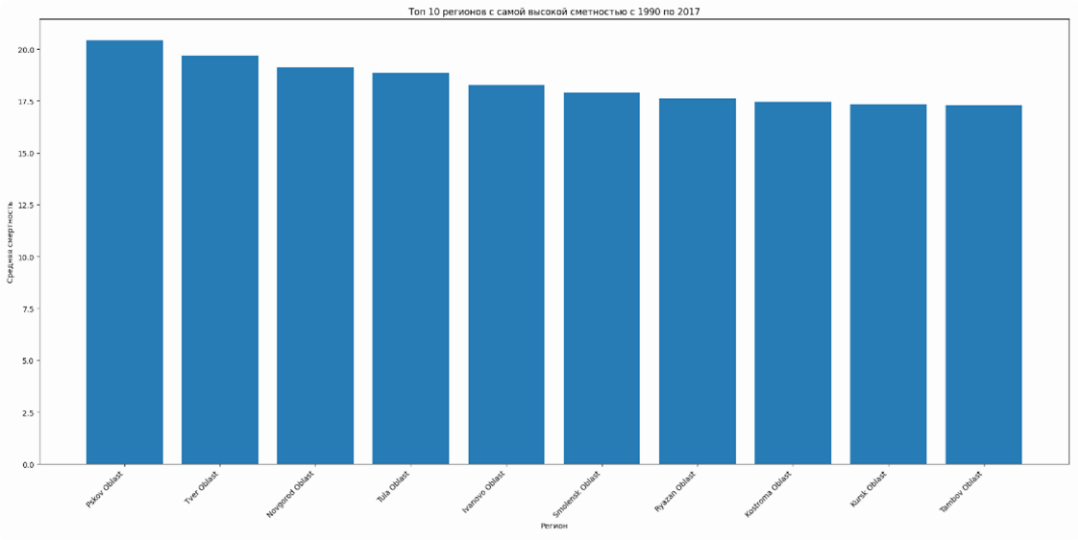
В следующие графиках определили топ 10 регионов по каждой из характеристик: рождаемость и смертность



Рождаемость:

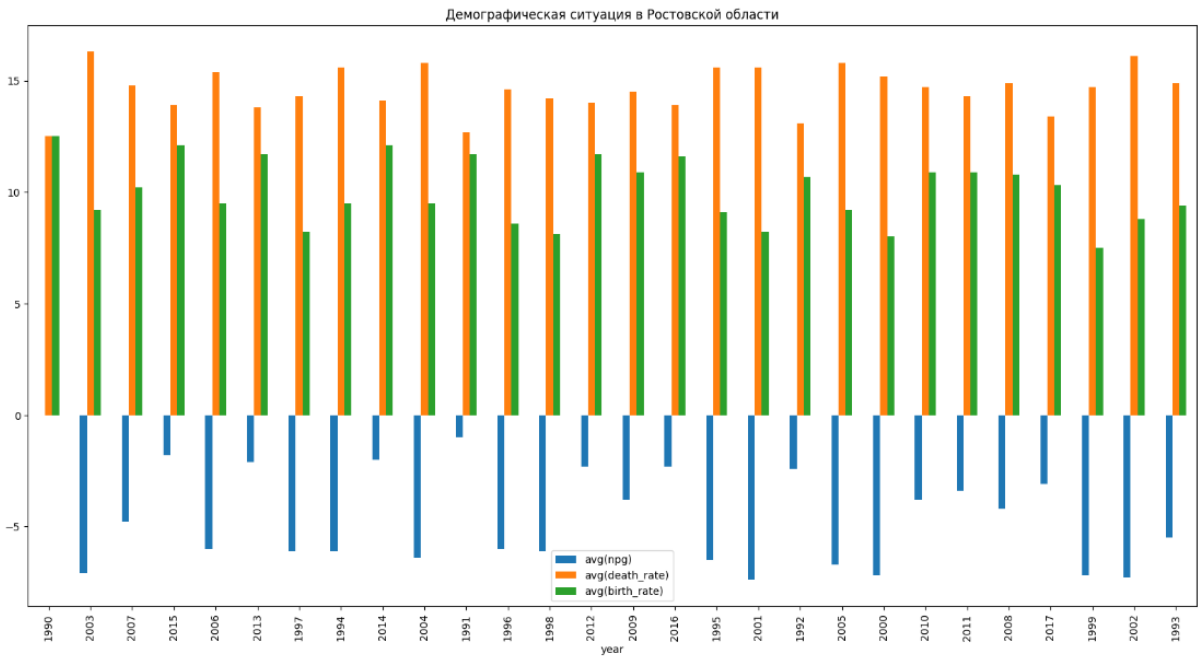
1. Чеченская респ.

- 2. Респ. Тува
- 3. Респ. Ингушетия



- Смертность:
- 1. Псковская обл.
  - 2. Тверская обл.
  - 3. Новгородская обл.

Следующий График: Демографическая ситуация в Ростовской области



Из представленного графика наблюдаем, что демографическая ситуация не в лучшем положении

Далее были созданы 3 таблицы для построения кругового графика показателя смертности

year	region	npg	birth_rate	death_rate	gdw	urbanization
1990	Altai Krai	1.8	12.9	11.1	80.24	58.07
1990	Amur Oblast	7.6	16.2	8.6	69.55	68.37
1990	Arkhangelsk Oblast	3.7	13.5	9.8	73.26	73.63
1990	Astrakhan Oblast	4.7	15.1	10.4	77.05	68.01
1990	Republic of Bashk...	6.5	16.2	9.7	80.53	64.22

Смертность ниже средней

year	region	npg	birth_rate	death_rate	gdw	urbanization
1990	Ivanovo Oblast	-2.4	11.6	14.0	81.82	82.3
1990	Novgorod Oblast	-1.8	12.3	14.1	83.06	69.68
1990	Pskov Oblast	-3.2	11.9	15.1	84.99	63.14
1990	Ryazan Oblast	-2.4	11.6	14.0	80.7	65.5
1990	Tambov Oblast	-3.2	11.7	14.9	85.32	56.28
1990	Tula Oblast	-3.4	11.5	14.0	85.76	71.00

Смертность средняя

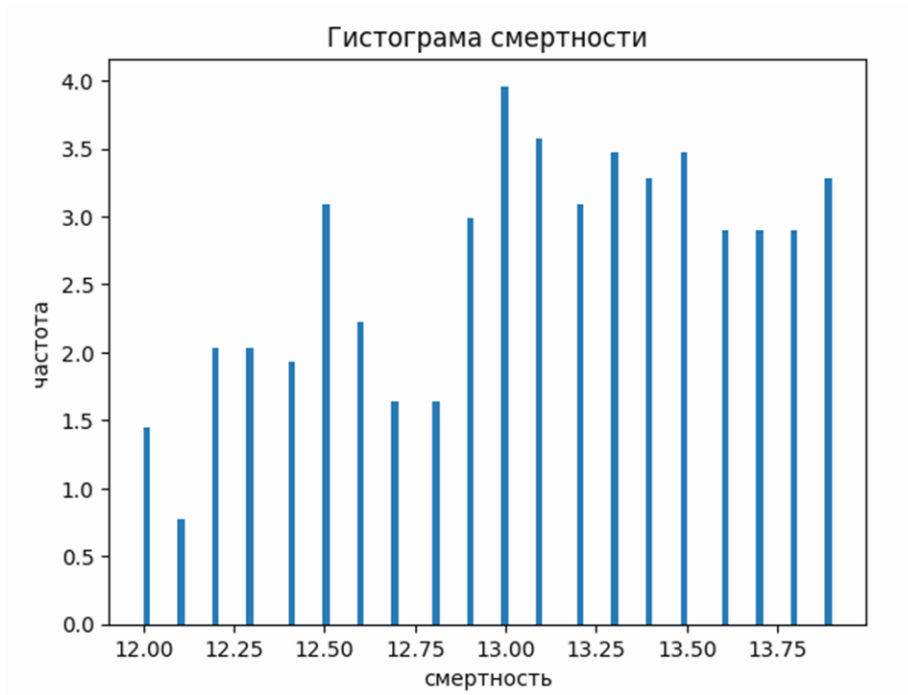
year	region	npg	birth_rate	death_rate	gdw	urbanization
1990	Republic of Adygea	1.9	14.2	12.3	84.66	52.42
1990	Belgorod Oblast	0.0	12.9	12.9	84.17	63.26
1990	Bryansk Oblast	0.1	13.0	12.9	86.48	67.49
1990	Vladimir Oblast	-0.4	12.1	12.5	77.78	79.31
1990	Vologda Oblast	1.4	13.4	12.0	82.16	65.48
1990	Voronezh Oblast	-2.4	11.5	13.9	83.78	60.94

Смертность выше средней



Круговой график смертности и по его результатам видно что в России смертность выше средней

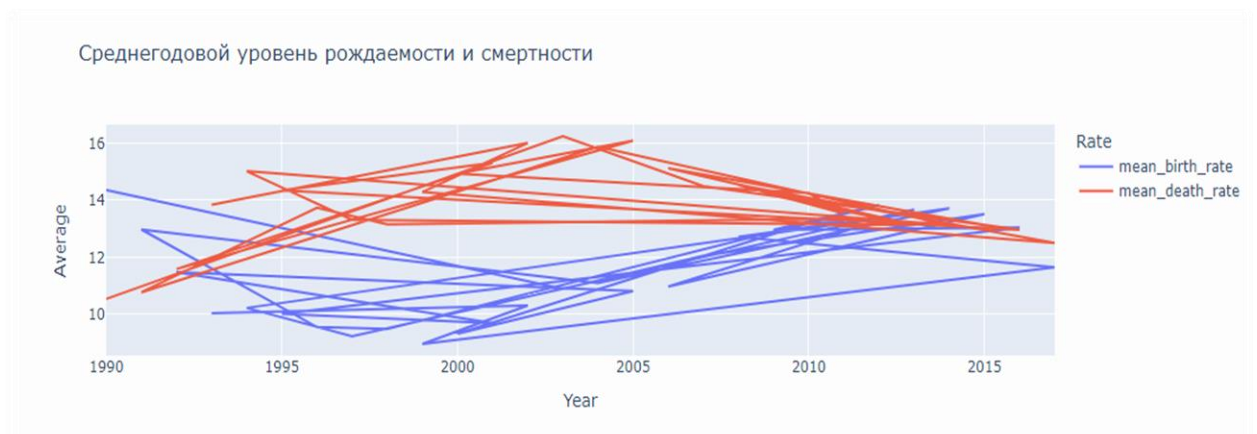
Далее делаем гистограмму смертности



наиболее частое встречаемое значение: 13.0

Выполнили расчет среднего значения рождаемости и смертности, сгруппировав их по каждому году. Полученный результат визуализировали с помощью интерактивного графика.

year	mean_birth_rate	mean_death_rate
1990	14.354878048780485	10.542682926829269
2003	10.89277108433735	16.23855421686747
2007	12.054216867469883	14.48915662650602
2015	13.51529411764706	13.131764705882356
2006	10.974698795180723	15.119277108433733
2013	13.67469879518072	13.126506024096383
1997	9.240243902439024	13.307317073170733
1994	10.228395061728396	15.009876543209884
2014	13.713253012048195	13.131325301204818





Сформировали два графика для визуального представления среднего показателя рождаемости и смертности по регионам, среднего естественного прироста населения на 1000 человек



## 4. Выводы

В ходе решения данного кейса, было сделано:

Анализ демографических данных регионов России в период с 1990 по 2017 год указывает на наличие долгосрочного демографического кризисного периода. Обнаружена связь между повышенной смертностью в определенный год и более заметным снижением коэффициента рождаемости в последующие годы. Этот эффект проявляется сильнее в регионах с изначально высокой смертностью, что может свидетельствовать о более сильном влиянии кризисных явлений на демографическую структуру таких регионов.