Отчет по кейсу

«Анализ демографических характеристик по регионам в России»

[1. Постановка задачи](https://wandb.ai/valentinkar/telecom100k/reports/---Vmlldzo3NTI3OTMx?accessToken=hm01nr3axgk95guu3phlliugexscvc0m175m1rw4ucmdg06shztx1hbjs5ecb7f5#1.-%D0%BF%D0%BE%D1%81%D1%82%D0%B0%D0%BD%D0%BE%D0%B2%D0%BA%D0%B0-%D0%B7%D0%B0%D0%B4%D0%B0%D1%87%D0%B8)

[2. Описание датасета](https://wandb.ai/valentinkar/telecom100k/reports/---Vmlldzo3NTI3OTMx?accessToken=hm01nr3axgk95guu3phlliugexscvc0m175m1rw4ucmdg06shztx1hbjs5ecb7f5#2.-%D0%BE%D0%BF%D0%B8%D1%81%D0%B0%D0%BD%D0%B8%D0%B5-%D0%B4%D0%B0%D1%82%D0%B0%D1%81%D0%B5%D1%82%D0%B0)

[3. Ход работы](https://wandb.ai/valentinkar/telecom100k/reports/---Vmlldzo3NTI3OTMx?accessToken=hm01nr3axgk95guu3phlliugexscvc0m175m1rw4ucmdg06shztx1hbjs5ecb7f5#3.-%D1%85%D0%BE%D0%B4-%D1%80%D0%B0%D0%B1%D0%BE%D1%82%D1%8B)

[3.1. Гипотеза](https://wandb.ai/valentinkar/telecom100k/reports/---Vmlldzo3NTI3OTMx?accessToken=hm01nr3axgk95guu3phlliugexscvc0m175m1rw4ucmdg06shztx1hbjs5ecb7f5#3.1.-%D0%B3%D0%B8%D0%BF%D0%BE%D1%82%D0%B5%D0%B7%D0%B0)

3.2. Первичная обработка датасета

[3.3. Анализ датасета](https://wandb.ai/valentinkar/telecom100k/reports/---Vmlldzo3NTI3OTMx?accessToken=hm01nr3axgk95guu3phlliugexscvc0m175m1rw4ucmdg06shztx1hbjs5ecb7f5#3.3.-%D1%81%D1%82%D0%B0%D1%82%D0%B8%D1%81%D1%82%D0%B8%D1%87%D0%B5%D1%81%D0%BA%D0%B8%D0%B5-%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5)

[4. Выводы](https://wandb.ai/valentinkar/telecom100k/reports/---Vmlldzo3NTI3OTMx?accessToken=hm01nr3axgk95guu3phlliugexscvc0m175m1rw4ucmdg06shztx1hbjs5ecb7f5#4.-%D0%B2%D1%8B%D0%B2%D0%BE%D0%B4%D1%8B)

﻿

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 рразных регионах России. Общая схема приведена на рисунке:

Изображение выглядит как текст, линия, График, Шрифт

Автоматически созданное описание

﻿

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

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

﻿

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

﻿

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

﻿

﻿

3.3. Анализ датасета

Анализируя датасет, мы вывели следующие гистограммы:

Изображение выглядит как снимок экрана, График, линия, текст

Автоматически созданное описание

﻿

﻿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

Изображение выглядит как текст, снимок экрана, линия, График

Автоматически созданное описание

Рождаемость:  
1. Чеченская респ.  
2. Респ. Тува  
3. Респ. Ингушетия

Изображение выглядит как снимок экрана, текст, линия, Параллельный

Автоматически созданное описание

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

Следующий График: Демографическая ситуация в Ростовской области

Изображение выглядит как снимок экрана, текст, линия, График

Автоматически созданное описание

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

Далее были созданы 3 таблице для построения кругового графика показателя смертности

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Смертность ниже средней

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Смертность средняя

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Смертность выше средней

Изображение выглядит как текст, снимок экрана, диаграмма, Шрифт

Автоматически созданное описание

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

Далее делаем гистограмму смертности

Изображение выглядит как текст, снимок экрана, линия, График

Автоматически созданное описание  
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.

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Изображение выглядит как линия, снимок экрана, График, диаграмма

Автоматически созданное описание

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

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

Изображение выглядит как текст, снимок экрана, Шрифт, линия

Автоматически созданное описание

4. Выводы

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

В ходе решения данного кейса, было сделан:

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.﻿