

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
**«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИТМО»**

Отчёт

по лабораторной работе №1 «Анализ данных и построение графиков в
Python»

по дисциплине «**Математические модели исторических процессов**»

Автор: Малаев Степан Геннадьевич
Факультет инфокоммуникационных технологий
Группа: K33422
Преподаватель: Екатерина Ивановна

Санкт-Петербург


2024

Выбор данных

Для анализа я выбрал датасет с платформы Kaggle.

Этот датасет предоставляет обширное представление о динамике мировой демографии на протяжении последнего века, выделяя ключевые аспекты роста, изменений и старения населения по различным странам и регионам.


Ссылка на датасет: "[World Population Data: Insights and Trends](#)".

 MOHAMEDFAHIM · UPDATED 9 DAYS AGO

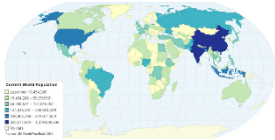
13

New Notebook

Download (9 kB)



Exploring Global Population Patterns Over Time



Data Card

Code (1)

Discussion (0)

Suggestions (0)

About Dataset

Welcome to the World Population!

This dataset offers a comprehensive view of global population trends over the past century, providing insights into how populations have grown, shifted, and aged across countries and regions. The dataset includes population figures from various sources, allowing for detailed analyses and visualizations of population dynamics.

Explore the data to uncover fascinating trends, such as population growth rates, age distributions, and urbanization patterns. Whether you're a data enthusiast, researcher, or student, this dataset offers a rich resource for understanding the complex and ever-changing landscape of global populations.

Usability
10.00

License
[Apache 2.0](#)

Expected update frequency
Annually

Tags
Social Science

Обзор данных

Проведя краткий обзор данных можно вынести следующие заметки:

- Все атрибуты в датафрейме за исключением названия страны "Country (or dependency)" представляют собой числовые форматы, однако не все из них представлены в корректных типах данных.
- Размер датасета относительно небольшой, всего 234 записи.
- Отсутствуют пустые данные
- Формат всех данных унифицирован

display(df.sample(5))
df.info()
Executed at 2024.03.06 04:35:17 in 16ms

<< 5 rows 5 rows x 12 columns Static output

#	Country (or dependency)	Population (2023)	Yearly Change	Net Change	Density (P/Km ²)
129	130 Croatia	4008617	-0.54 %	-21741	72
161	162 Fiji	936375	0.71 %	6609	51
189	190 Curaçao	192077	0.48 %	914	433
12	13 Philippines	117337368	1.54 %	1778359	394
139	140 Armenia	2777970	-0.09 %	-2499	98

0	#	234 non-null	int64
1	Country (or dependency)	234 non-null	object
2	Population (2023)	234 non-null	int64
3	Yearly Change	234 non-null	object
4	Net Change	234 non-null	int64
5	Density (P/Km ²)	234 non-null	int64
6	Land Area (Km ²)	234 non-null	int64
7	Migrants (net)	234 non-null	int64
8	Fert. Rate	234 non-null	object
9	Med. Age	233 non-null	float64
10	Urban Pop %	234 non-null	object
11	World Share	234 non-null	object

dtypes: float64(1), int64(6), object(5)
memory usage: 22.1+ KB

Предобработка данных

```
df['Yearly Change'] = df['Yearly Change'].str.rstrip(' %').astype(float)
```

```
df['Fert. Rate'] = df['Fert. Rate'].replace({'...': None})
```

```
df['Fert. Rate'] = df['Fert. Rate'].astype(float)
```

```
df['Urban Pop %'] = df['Urban Pop %'].replace({'N.A.': None})
```

```
df['Urban Pop %'] = df['Urban Pop %'].str.rstrip(' %').astype(float)
```

```
df['World Share'] = df['World Share'].str.rstrip(' %').astype(float)
```

Executed at 2024.03.06 05:08:42 in 15ms

```
df.info()
```

Executed at 2024.03.06 05:08:42 in 7ms

```
<class 'pandas.core.frame.DataFrame'>
```

```
RangeIndex: 234 entries, 0 to 233
```

```
Data columns (total 12 columns):
```

#	Column	Non-Null Count	Dtype
0	#	234 non-null	int64
1	Country (or dependency)	234 non-null	object
2	Population (2023)	234 non-null	int64
3	Yearly Change	234 non-null	float64
4	Net Change	234 non-null	int64
5	Density (P/Km ²)	234 non-null	int64
6	Land Area (Km ²)	234 non-null	int64
7	Migrants (net)	234 non-null	int64
8	Fert. Rate	233 non-null	float64
9	Med. Age	233 non-null	float64
10	Urban Pop %	216 non-null	float64
11	World Share	234 non-null	float64

```
dtypes: float64(5), int64(6), object(1)
```

```
memory usage: 22.1+ KB
```

После обработки у нас появились пустые значения.

Базовый анализ

```
mean_population = df['Population (2023)'].mean()
median_population = df['Population (2023)'].median()
std_population = df['Population (2023)'].std()

print(
    "Анализ населения по всем странам:",
    '-' * 55,
    f"Среднее значение населения: {mean_population:.0f}",
    f"Медиана населения: {median_population:.0f}",
    f"Стандартное отклонение населения: {std_population:.0f}",
    '-' * 55,
    sep='\n'
)
```

Executed at 2024.03.06 05:15:57 in 4ms

Анализ населения по всем странам:

Среднее значение населения: 34375648

Медиана населения: 5643895

Стандартное отклонение населения: 137386102

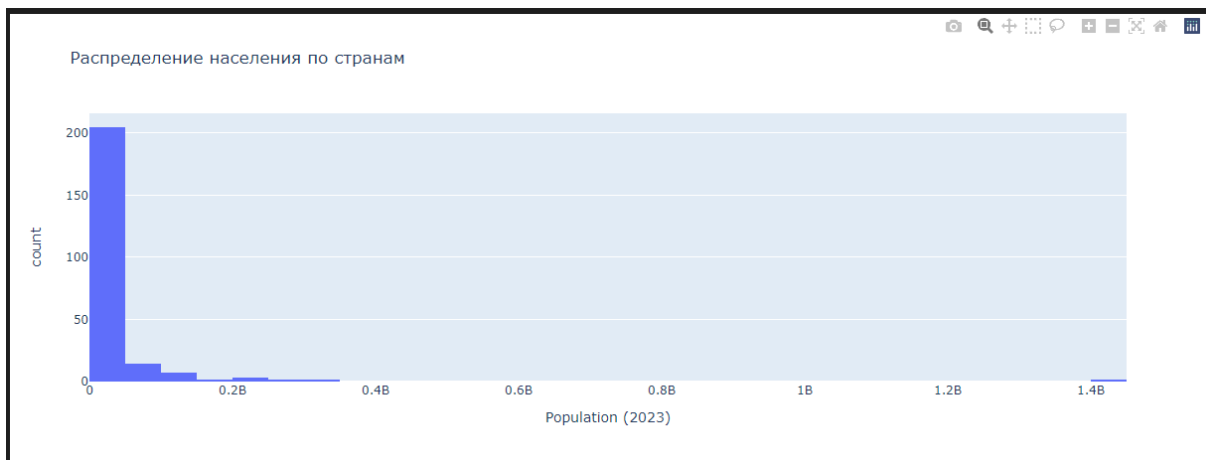
Выводы, которые можно сделать из этого анализа:

1. Большое различие между средним значением и медианой. Среднее значение населения значительно выше медианы показывает что в наборе данных есть страны с очень большим населением, которые сдвигают среднее значение вверх. Медиана менее чувствительна к экстремальным значениям, поэтому она дает более реалистичное представление о размере населения стран.
2. Очень большое стандартное отклонение говорит о том, что размеры населения стран сильно различаются. Страны могут варьироваться от очень маленьких например, островных государств до стран с миллиардным населением например, Китай или Индия
3. Распределение населения по странам является асимметричным с длинным правым хвостом. Тем самым, меньшее количество стран с очень большим населением сильно влияет на среднее значение, тогда как большинство стран имеют относительно небольшое население.

Базовая визуализация

Для начала построим гистограмму распределения населения стран. Этот график покажет, как население распределено между странами, и поможет выявить, сколько стран имеют похожее население.

```
fig1 = px.histogram(df, x='Population (2023)', nbins=50, title="Распределение населения по странам")  
  
fig1.show()  
Executed at 2024.03.06 06:01:35 in 451ms
```



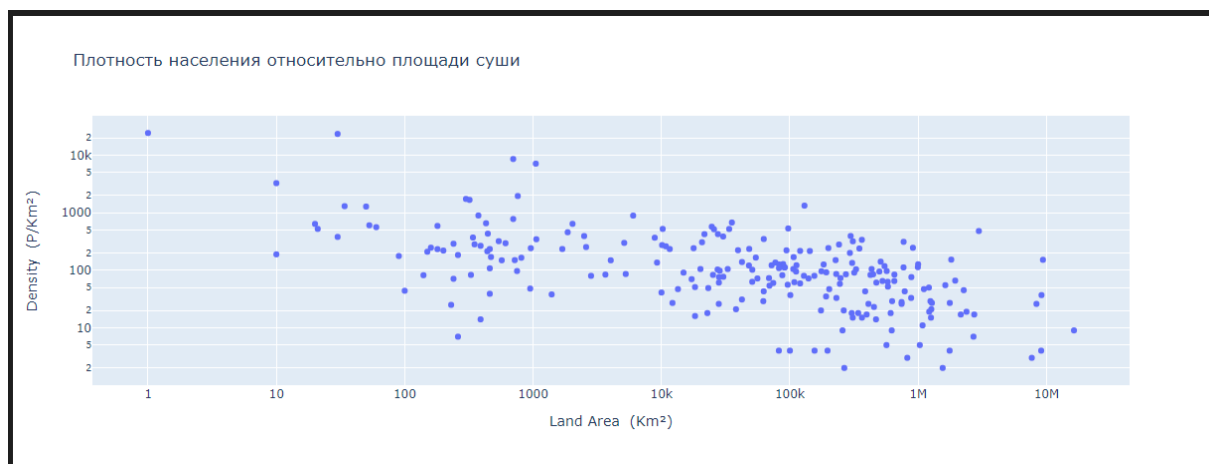
Анализ демографических данных показывает, что большинство стран мира имеют население в диапазоне от 0 до 50 миллионов человек. Это указывает на то, что небольшие и средние по размеру страны преобладают на мировой арене. Однако стоит отметить, что две страны выделяются существенным образом, имея население более 1.4 миллиарда человек каждая

Кроме того, интересно было бы посмотреть на график рассеяния плотности населения относительно площади суши. Это даст представление о том, как плотность населения связана с размером территории страны.

```
fig2 = px.scatter(  
    df, x='Land Area (Km2)', y='Density (P/Km2)', hover_name='Country (or dependency)',  
    title="Плотность населения относительно площади суши", log_x=True, log_y=True  
)
```

```
fig2.show()
```

Executed at 2024.03.06 06:01:36 in 44ms



Что касается территориального распределения и плотности населения, большинство стран занимают площадь от 10 тысяч до 1 миллиона квадратных километров, с плотностью населения варьирующейся от 10 до 1000 человек на квадратный километр. Россия выделяется как страна с самой большой территорией, в то время как Монголия и Западная Сахара имеют наименьшую плотность населения. С другой стороны, [Макао](#) обладает самой высокой плотностью населения. Монако является государством с наименьшей площадью территории.

Последним шагом построим график изменения численности населения по странам. Это поможет выявить страны с наибольшим приростом и убытком населения.

```
fig3 = px.bar(
    df[:45], x='Country (or dependency)', y='Net Change',
    title="Изменение численности населения по странам",
)

fig3.update_layout(
    xaxis_tickangle=-45,
    title_font_size=22,
    font_size=14,
    autosize=False,
    width=1200,
    height=800,
)

fig3.show()
```

Executed at 2024.03.06 06:01:36 in 46ms



Индия показывает замечательный годовой рост населения, превышающий 12 миллионов человек, что выделяет ее среди других стран мира. Это подчеркивает ее растущее демографическое значение и потенциальное влияние на мировые экономические и социальные процессы. В то же время, наблюдается тенденция сокращения населения в странах, которые в настоящее время находятся в состоянии военных конфликтов или испытывали их в недавнем прошлом, на 2023 год.

Вывод

В ходе данной лабораторной работы был проведен анализ датасета "[World Population Data: Insights and Trends](#)", выбранного с платформы Kaggle. Этот датасет предоставил обширный набор данных для изучения динамики мировой демографии за последний век, акцентируя внимание на ключевых аспектах роста и изменений населения по различным странам и регионам.

Базовая визуализация включала гистограмму распределения населения и график рассеяния плотности населения относительно площади суши, что позволило лучше понять взаимосвязь между размером территории страны и плотностью ее населения.

Особое внимание было уделено анализу изменения численности населения.