

УДК 004.421

<https://doi.org/10.33619/2414-2948/114/20>

АНАЛИЗ ДАННЫХ С ИСПОЛЬЗОВАНИЕМ ЯЗЫКА ПРОГРАММИРОВАНИЯ PYTHON

©*Тайырбекова Р., Ошский государственный университет,
г. Ош, Кыргызстан, rahimar589@gmail.com*

©*Кочконбаева Б. О., SPIN-код: 6521-0121, Ошский государственный университет,
г. Ош, Кыргызстан, buajar@mail.ru*

©*Калбаева Д. И., Ошский технологический университет им. М. М. Адышева,
г. Ош, Кыргызстан, d.kalbaeva@list.ru*

DATA ANALYSIS USING PYTHON PROGRAMMING LANGUAGE

©*Taiirbekova R., Osh State University, Osh city, Kyrgyzstan, rahimar589@gmail.com*

©*Kochkonbaeva B., SPIN-code: 6521-0121, Osh State University,
Osh, Kyrgyzstan, buajar@mail.ru*

©*Kalbaeva D., Osh Technological University after named M.M. Adyshev,
Osh, Kyrgyzstan, d.kalbaeva@list.ru*

Аннотация. Сегодня язык Python является самым мощным языком программирования для обработки и анализа больших данных. Это достигается за счет многих полезных библиотек языка, который каждый день обновляются с добавлением новых возможностей. В данной статье рассматривается анализ демографического состояния населения Кыргызстана за 2019–2023 годы с использованием языка программирования Python и его библиотек для обработки и визуализации данных. Использование Python для анализа демографических данных позволяет эффективно обработать большие массивы информации, выявить закономерности и представить результаты в наглядной форме. Такой подход может быть полезен для государственных органов и исследователей при планировании демографической политики.

Abstract. Today, Python is the most powerful programming language for processing and analyzing big data. This is achieved through many useful language libraries, which are updated every day with new features. This article discusses the analysis of the demographic state of the population of Kyrgyzstan for 2019-2023 using the Python programming language and its libraries for processing and visualizing data. Using Python to analyze demographic data allows you to effectively process large amounts of information, identify patterns and present the results in a visual form. This approach can be useful for government agencies and researchers when planning demographic policy.

Ключевые слова: python, pandas, matplotlib, анализ данных, обработка данных, визуализация данных.

Keywords: python, pandas, matplotlib, data analysis, data processing, data visualization.

Анализ данных — это процесс обработки структурированной и неструктурированной информации с использованием статистических методов, вычислительных алгоритмов и инструментов визуализации. Его основная цель — выявление значимых закономерностей, извлечение ценной информации и прогнозирование тенденций. Процесс анализа данных

включает несколько ключевых этапов: сбор информации, её предварительная обработка и очистка, исследование и интерпретация результатов, а также визуализацию и представление выводов в удобном формате. В ходе анализа применяются различные методы, такие как описательная статистика, корреляционный и регрессионный анализ, кластеризация, классификация и другие алгоритмы интеллектуального анализа данных [1].

В современном мире анализ данных играет ключевую роль в принятии решений в различных сферах, включая бизнес, медицину, науку и технологии. Python стал одним из наиболее популярных языков программирования для работы с данными благодаря своей простоте, гибкости и мощному набору библиотек, которые позволяют эффективно обрабатывать большие массивы информации. В числе наиболее часто используемых библиотек входит:

Pandas – инструмент для работы с таблицами и структурированными данными, позволяющий загружать, очищать и анализировать данные.

NumPy – библиотека для работы с массивами и матрицами, предоставляющая инструменты для выполнения математических вычислений, таких как сложение, умножение, статистические операции и многое другое.

Scikit-learn – инструмент для машинного обучения, включающий алгоритмы классификации, кластеризации и предсказательного моделирования.

Matplotlib и Seaborn – библиотеки для визуализации данных, позволяющие строить графики, диаграммы и другие виды визуальных представлений информации. С их помощью можно создавать линейные графики, гистограммы, диаграммы рассеяния и другие аналитические визуализации.

Рассмотрим основные инструменты вышеперечисленных библиотек Python, используемые для анализа данных, и покажем их применение на примере анализа демографического состояния населения Кыргызстана. Это позволит продемонстрировать, как методы анализа данных могут быть использованы для выявления ключевых демографических тенденций и прогнозирования будущих изменений. Целесообразно показать возможность языка Python на примере анализа демографической ситуации Кыргызстана. Для этой работы разработан следующий алгоритм:

1. Собрать данные из открытых источников для дальнейшего анализа.
2. Чтение данных из файла, соответствующего определенному формату.
3. Чтобы не перегружать память, извлеките часть данных для обработки, создав так называемый «кадр данных».
4. Выбрать для анализа только необходимые столбцы, включая названия объектов и значения изучаемых факторов.
5. Сгруппируйте объекты с одинаковыми характеристиками и используйте правила обработки значений коэффициентов для каждой группы.
6. Сравните данные по столбцам, чтобы выявить закономерности.
7. Выберите оптимальный тип графического представления данных в зависимости от цели анализа.
8. Создайте выбранный график или диаграмму на основе отсортированных данных.

Мы рассмотрели открытые источники для сбора данных. В основном необходимые данные были получены из открытых источников: сайта Национального статистического комитета (opendata), годовых отчетов [2].

Собранные данные сохранялись в электронной таблице Excel в формате CSV (Рисунок 1).

	A	B	C	D	E
1	oblast,2019 г.,2020 г.,2021 г.,2022 г.,2023 г.				
2	Баткен,13123,11263,10303,11342,12380				
3	Жалал-Абад,26564,23307,21660,22701,25402				
4	Иссык-Куль,7298,6078,5208,6039,7532				
5	Нарын,4109,3579,2977,3197,5597				
6	Ош,29464,25443,21969,24305,29530				
7	Талас,4575,3972,3577,3846,4794				
8	Чуй,18299,14342,12314,14102,15659				
9	город Бишкек,24780,20236,21312,22509,8369				
10	город Ош,11977,9915,11969,10783,5214				

Рисунок 1. Файл данных “grow_of_people.csv”

Python предоставляет обширные возможности для анализа данных, начиная с их загрузки и предварительной обработки. Одним из ключевых этапов аналитического процесса является импорт данных, который в Python можно выполнять с помощью различных библиотек. Они позволяют загружать информацию из множества источников, включая CSV- и Excel-файлы, базы данных SQL, веб-страницы и другие форматы.

Одной из наиболее востребованных библиотек для работы с данными является Pandas. С её помощью были загружены данные из CSV-файла, на основе которых сформированы датафреймы с информацией о естественном приросте населения за 2019–2023 годы. Кроме того, проведён анализ численности населения (включая распределение по полу), данных о внешней миграции (количество прибывших и выбывших), а также показателей рождаемости и смертности [3].

```
import pandas as pd
if __name__ == '__main__':
    data = pd.read_csv('grow_of_people.csv')
    print(data.head(10))
    data_2019 = data[['oblast', '2019 г.']]
    data_2020 = data[['oblast', '2020 г.']]
    data_2021 = data[['oblast', '2021 г.']]
    data_2022 = data[['oblast', '2022 г.']]
    data_2023 = data[['oblast', '2023 г.']]
    plt.plot(oblast,data_19)
    plt.plot(oblast, data_19, label="2019", color="blue", linestyle="-", marker="o")
    plt.plot(oblast, data_20, label="2020", color="red", linestyle="--", marker="s")
    plt.plot(oblast, data_21, label="2021", color="green", linestyle="--", marker="o")
    plt.plot(oblast, data_22, label="2022", color="yellow", linestyle="--", marker="s")
    plt.plot(oblast, data_23, label="2023", color="black", linestyle="-", marker="o")
    plt.legend(loc="upper left", bbox_to_anchor=(1, 1))
    plt.grid(True)
    plt.show()
```

Результаты и визуализация данных

С использованием библиотек Pandas и Matplotlib в языке программирования Python был проведен анализ демографических данных. На Рисунке 2 представлена визуализация естественного прироста населения Кыргызстана за 2019–2023 годы.

Несмотря на стабильный рост численности населения, Кыргызстан остается наименее населенным государством в Центральной Азии, что во многом связано с миграционными процессами (Рисунок 3) [4]. Однако в последние годы наблюдается увеличение числа граждан, возвращающихся в страну, что обусловлено появлением новых рабочих мест (Рисунок 4).

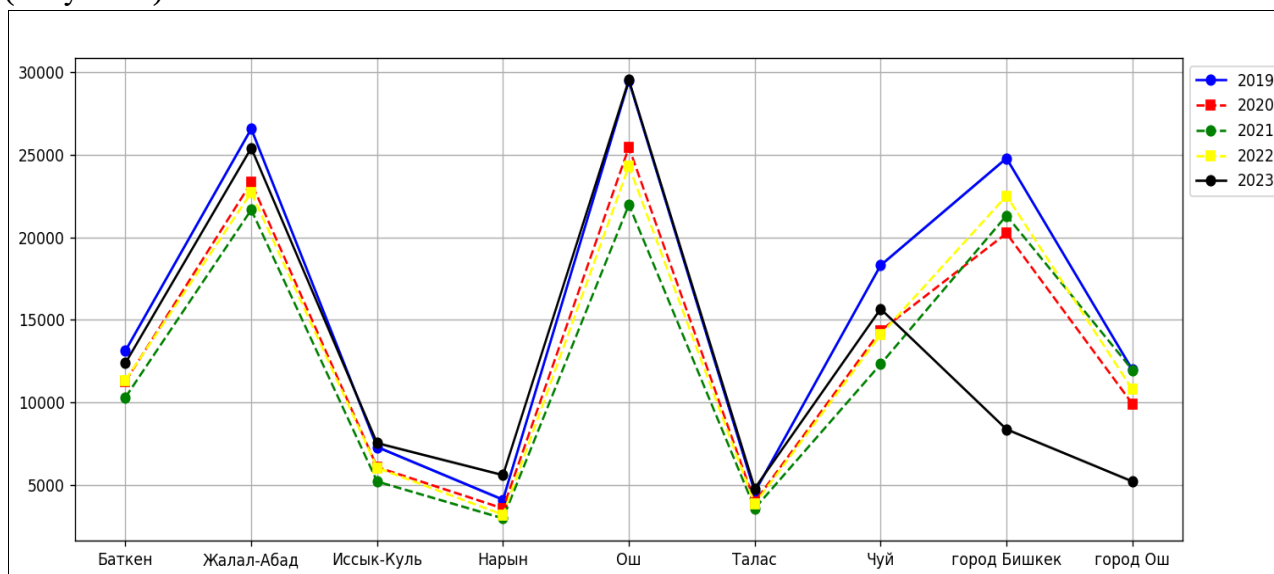


Рисунок 2. Сведения о естественном приросте населения за 2019–2023 годы

```
data = pd.read_csv('migration.csv')
data_obl=data[['country']]
data_2019 = data[['country','2019 г.']]
data_2020 = data[['country', '2020 г.']]
data_2021 = data[['country', '2021 г.']]
data_2022 = data[['country', '2022 г.']]
data_2023 = data[['country', '2023 г.']]
```

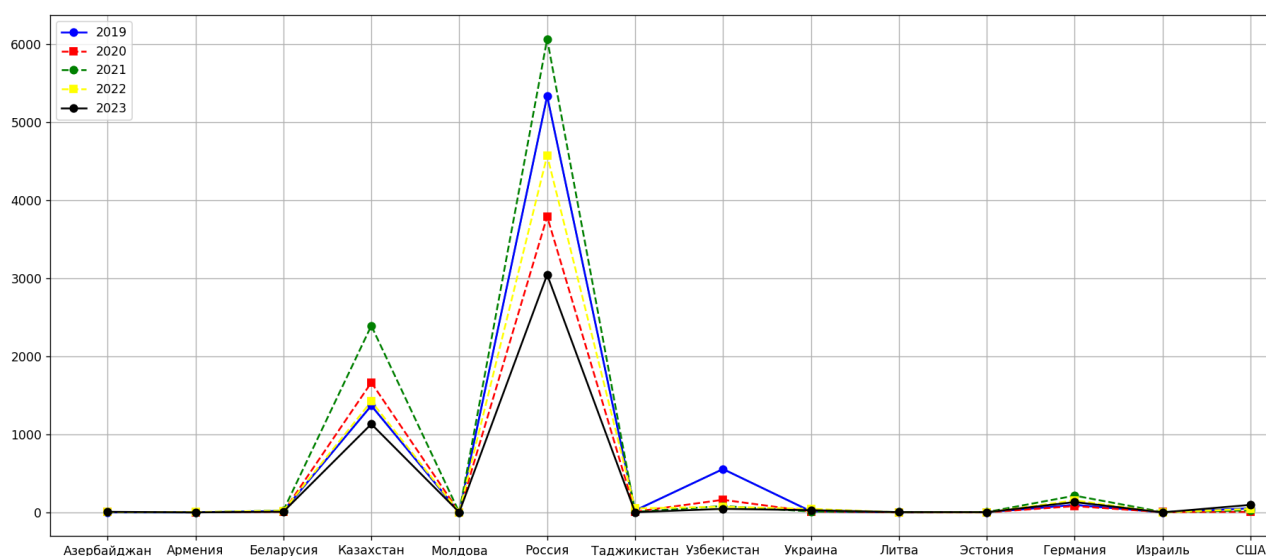


Рисунок 3. Число выбывших из страны за 2019-2023 гг.

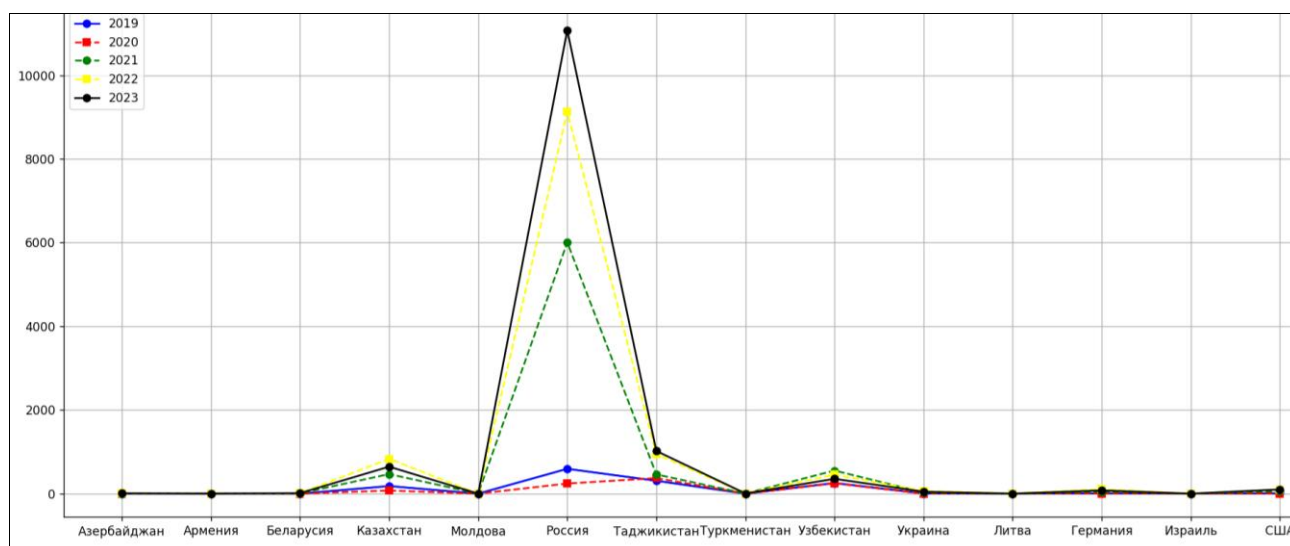


Рисунок 4. Число прибывших в Кыргызстан за 2019-2023 гг.

Полученный график наглядно отражает результаты обработки данных о людях, прибывших в страну за последние годы.

Библиотека Matplotlib предоставляет широкие возможности для настройки внешнего вида графиков, позволяя изменять стиль отображения элементов, цветовые схемы, шрифты и другие параметры. Кроме того, результаты можно представить в табличном формате, что упрощает их дальнейший анализ и объединение с другими данными [5].

Активное и в некоторой степени хаотичное развитие экосистемы Python объясняется его открытым исходным кодом и свободным распространением. Этот фактор способствует его применению в самых разных областях науки и прикладных исследований. Благодаря этому как профессиональные программисты, так и энтузиасты вносят вклад в его развитие, расширяя функциональные возможности языка и его библиотек [6].

Процесс разработки и тестирования прикладного компьютерного приложения подтвердил доступность и простоту реализации алгоритмов обработки данных с использованием Python. Основная цель созданного алгоритма — продемонстрировать возможности библиотеки Pandas, подчеркнуть удобство её использования и наглядность получаемых результатов.

На основе разработанного алгоритма обработки и анализа массивов данных с использованием языка программирования Python и библиотек Pandas и Matplotlib можно выделить следующие преимущества данного подхода:

Весь процесс — от загрузки данных до их графической визуализации — выполняется в рамках одного приложения.

Поддержка работы с массивами данных в различных форматах.

Гибкость в выборе нужных данных из общего массива, что упрощает исследовательский процесс.

Возможность сортировки информации по любым заданным параметрам.

Удобное и наглядное представление результатов в виде графиков, диаграмм или таблиц.

Открытый исходный код, доступность и простота использования.

Учитывая перечисленные преимущества, данный алгоритм можно рекомендовать для применения в различных научных и прикладных исследованиях. В частности, авторы планируют использовать его для обработки баз данных, связанных с параметрами

солнечного излучения и потоками воздуха в атмосфере отдельных регионов. Это позволит анализировать изменения мощности солнечных и ветровых электростанций. Для реализации данной задачи потребуется подключение дополнительных библиотек, доступных на специализированных платформах.

Список литературы:

1. Кочербаева А. А. Особенности демографического развития Кыргызстана: проблемы, их социальные последствия и перспективы решения // Народонаселение. 2016. №1-1. С. 16-24.
2. Авдеев А. А. Анализ демографической ситуации в Кыргызстане: обновление 2020 г. // Демографические вызовы в развитии Кыргызской Республики. Аналитическая записка. ЮНФПА. 2021.
3. Кыргызстан в цифрах. Статистический сборник. Бишкек, 2015. 352 с.
4. Эргешбаев У. Ж., Ешенова С. Проблемы внешней трудовой миграции населения Кыргызстана // Народонаселение. 2016. Т. 1. №71. С. 34-41.
5. Ильичев В. Ю., Юрик Е. А. Анализ массивов данных с использованием библиотеки pandas для Python // Научное обозрение. Технические науки. 2020. №4. С. 41-45.
6. Уэс Маккинли. Python и анализ данных. М.: ДМК Пресс, 2015. 482 с.

References:

1. Kocherbaeva, A. A. (2016). Osobennosti demograficheskogo razvitiya Kyrgyzstana: problemy, ikh sotsial'nye posledstviya i perspektivy resheniya. *Narodonaselenie*, (1-1), 16-24. (in Russian).
2. Avdeev, A. A. (2021). Analiz demograficheskoi situatsii v Kyrgyzstane: obnovlenie 2020 g. Demograficheskie vyzovy v razvitii Kyrgyzskoi Respubliki. *Analiticheskaya zapiska. YuNFPA*. (in Russian).
3. Kyrgyzstan v tsifrah. Statisticheskii sbornik (2015). Bishkek. (in Russian).
4. Ergeshbaev, U. Zh., & Eshenova S. (2016). Problemy vneshnei trudovoi migratsii naseleniya Kyrgyzstana. *Narodonaselenie*, 1(71), 34-41. (in Russian).
5. Il'ichev, V. Yu., & Yurik, E. A. (2020). Analiz massivov dannykh s ispol'zovaniem biblioteki pandas dlya Python. Nauchnoe obozrenie. *Tekhnicheskie nauki*, (4), 41-45. (in Russian).
6. Ues Makkinli. Python i analiz dannykh (2015). Moscow. (in Russian).

*Работа поступила
в редакцию 07.03.2025 г.*

*Принята к публикации
12.03.2025 г.*

Ссылка для цитирования:

Тайырбекова Р., Кочконбаева Б. О., Калбаева Д. И. Анализ данных с использованием языка программирования Python // Бюллетень науки и практики. 2025. Т. 11. №5. С. 139-144. <https://doi.org/10.33619/2414-2948/114/20>

Cite as (APA):

Taiirbekova, R., Kochkonbaeva, B., & Kalbaeva, D. (2025). Data Analysis using Python Programming Language. *Bulletin of Science and Practice*, 11(5), 139-144. (in Russian). <https://doi.org/10.33619/2414-2948/114/20>