

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

**ФГБОУ ВО «Кубанский государственный аграрный  
университет имени И.Т. Трубилина»**

**Учетно-финансовый факультет**

**Кафедра экономического анализа**

## **ОСНОВЫ АНАЛИЗА ДАННЫХ НА PYTHON**

### **Методические указания**

**по выполнению контрольной работы для  
обучающихся очно-заочной формы обучения  
по направлению 38.03.01 Экономика,  
направленность «Бизнес-аналитика»**

**Краснодар  
КубГАУ  
2021**

*Составители: Васильева Н.К., Мороз Н.Ю.*

**Основы анализа данных на Python:** метод. указания по выполнению контрольной работы / сост. Васильева Н.К., Мороз Н.Ю.– Краснодар :КубГАУ, 2021. – 17 с.

Методические указания по выполнению контрольной работы по дисциплине «Основы анализа данных на Python» содержат контрольные вопросы и задачи, их распределение по вариантам, а также рекомендуемые источники информации, позволяющие закрепить знания, умения и навыки.

Предназначены для обучающихся по программе бакалавриата 38.03.01 «Экономика», направленность «Бизнес-аналитика».

Рассмотрены и одобрены методической комиссией учетно-финансового факультета Кубанского государственного аграрного университета, протокол № 9 от 16.06.2021.

Председатель методической комиссии

И.Н. Хромова

© Васильева Н.К.,  
Мороз Н.Ю.,  
составление, 2021  
© ФГБОУ ВО «Кубанский  
государственный аграрный  
университет имени  
И. Т. Трубилина», 2021

## ВВЕДЕНИЕ

**Целью** освоения дисциплины «Основы анализа данных на Python» является формирование комплекса знаний о средствах программирования на Python и специализированных библиотек анализа данных, а так же применение полученных знаний для целей бизнес-аналитики.

**Задачи** дисциплины:

- формирование теоретических знаний и практических навыков по основам анализа данных в Python
- формирование навыков использования Python для работы с табличными данными, импортируемыми или экспортируемыми популярными программами для анализа данных
- приобретение практических навыков использования набора библиотек языка Python для решения задач в сфере бизнес-аналитики

В целях обеспечения текущего контроля самостоятельной работы студентов очно-заочной формы обучения по дисциплине ведущим преподавателем выдается задание на контрольную работу.

Задание включает рассмотрение двух теоретических вопросов (раздел 1 методических указаний) и решение одной задачи (из раздела 2).

Вариант обучающегося определяется по номеру зачетной книжки в соответствии с порядком, изложенном в разделе 3.

При выполнении контрольной работы рекомендуется использовать нормативно-правовую информацию и литературные источники, приведенные в разделе 4.

В конце контрольной работы необходимо привести список источников, использованных обучающимся при ее выполнении.

Контрольная работа не ограничивается по объему. Вопросы

должны быть раскрыты полностью. Решение задачи предполагает ответы на все вопросы со ссылками на нормативно-правовые источники.

Оригинальность работ, выполненных на компьютере, проверяется в системе «Антиплагиат».

# 1 Вопросы для контрольной работы

1. Python – стандарт для работы с большими данными
2. Прикладные задачи бизнес-аналитиков, для решения которых применяется язык Python
3. Установка сторонних библиотек в Python
4. Преобразования между типами в Python
5. Работа с последовательностями в Python
6. Работа с текстовыми файлами в Python
7. Работа с модулями в Python
8. Формат иерархических данных JSON
9. Формат иерархических данных XML
10. Модуль для работы с числовыми данными NumPy
11. Особенности типы данных в NumPy
12. Работа с векторами и матрицами
13. Вычисление главных статистических метрик с помощью NumPy
14. Модуль для работы с табличным представлением данных Pandas
15. Преобразование словарей в табличный формат Pandas
16. Инструменты визуализации данных для Python
17. Американский и европейский стандарт записи дат
18. Регулярные выражения regex
19. Сбор данных с произвольных веб-страниц
20. Формирование файла CSV средствами Python
21. Загрузка сформированных в Python данных в программу
22. Прямая интеграция с документами Excel
23. Преимущества использования Python для анализа данных по сравнению с другими инструментами.
24. Прикладные задачи бизнес-аналитики, для решения которых подходит инструментарий Python.
25. Обзор библиотек и инструментов.
26. Программирование на Python: типы данных, основные структуры, методы и функции, условные операторы, циклы, списковые включения, функция map() и анонимные функции, отладка кода (try/except).
27. Алгоритмы: оптимизация и сложность на примере алгоритмов сортировок.
28. Библиотека numpy: векторы и массивы, специальные типы данных.

29. Библиотека pandas.
30. Основы работы с датафреймами: загрузка, очистка, фильтрация, группировка и агрегация.
31. Описательные статистики, распределения, разведывательный анализ, работа с пропущенными значениями, постановка гипотезы.
32. Сбор данных из открытых источников: web-scraping, работа с API, парсинг текста и таблиц, сохранение файлов.
33. Виды графиков, их корректное использование и интерпретация.
34. Принципы хорошей визуализации, основные ошибки при визуализации данных.
35. Основы визуализации в matplotlib.
36. Оформление и кастомизация графиков.
37. Визуализации отфильтрованных и сгруппированных данных.
38. Создание интерактивных визуализаций в Plotly.
39. Нейронные сети. Области применения.
40. Использование существующих решений для собственных задач
41. Американский и европейский стандарт записи дат
42. Формат записи дат ISO
43. Формат хранения дат Unix Time
44. Регулярные выражения regex
45. Сбор данных с произвольных веб-страниц
46. Автоматизация передачи входных данных через аргументы командной строки
47. Функции построения графиков в pandas
48. Недостатки Python
49. Модуль для работы с числовыми данными NumPy
50. Особенности типов данных в NumPy
51. Работа с векторами и матрицами
52. Вычисление главных статистических метрик с помощью NumPy
53. Модуль для работы с табличным представлением данных Pandas
54. Преобразование словарей в табличный формат Pandas
55. Загрузка данных из внешних источников
56. Инструменты визуализации данных для Python
57. Обзор библиотеки Matplotlib
58. Преобразования между типами в Python
59. Работа с последовательностями в Python
60. Работа с текстовыми файлами в Python



## 2 Задачи для контрольной работы

**Задача 1.** Создать следующего вида исходный файл из восьми строк, содержащий в каждой строке значение исходной системы счисления, результирующей системы счисления и пяти различных целых положительных чисел, корректных для исходной системы счисления:

10 2 121 13 10 2018 5

**Задача 2.** Создать следующего вида исходный файл из десяти строк, содержащий в каждой строке ФИО, дату рождения, дату получения паспорта и баллы ЕГЭ по трём предметам:

Апельсинов А.А. 08.02.2000 17.03.2014 90 100 91

Виноградов В.В. 09.03.1999 15.04.2013 67 99 98

Яблоков Я.Я. 10.04.2000 19.05.2014 94 94 94

Морковкин М.М. 11.05.1999 17.06.2013 91 82 73

**Задача 3.** Не используя готовые сторонние подключаемые функции для факториала, `int()`, `bin()` и т.п., написать программу на языке Python 3.x, которая бы вычисляла среднее значение баллов ЕГЭ, сортировала строки по указанной колонке в обратном порядке (от большего к меньшему) и выводила результат следующего вида (для сортировки по дате рождения):

Яблоков Я.Я. | 10.04.2000 | 19.05.2014 | 94 94 94 -> 94

Апельсинов А.А. | 08.02.2000 | 17.03.2014 | 90 100 91 ->  
93,666666

Морковкин М.М. | 11.05.1999 | 17.06.2013 | 91 82 73 -> 82

Виноградов В.В. | 09.03.1999 | 15.04.2013 | 67 99 98 -> 88



**Задача 4.** Импортируйте набор данных о пассажирах «Титаника». Проведите разведывательный анализ данных. Поставьте гипотезу. Проведите необходимые преобразования данных.

**Задача 5.** На основании поставленной гипотезы в задании 1 выберите модели для предсказания. Обоснуйте выбор. Постройте выбранные модели, выберите лучшую. Постройте ансамбль моделей. Улучшились ли предсказания? Оформите результаты в виде отчета.

**Задача 6.** Вариант проекта: соберите данные для анализа с сайта Федеральной службы государственной статистики. Проведите предварительную обработку данных. Поставьте задачу классификации. Проведите необходимые преобразования данных. Постройте модель. Оцените, насколько хорошо модель справляется с поставленной задачей. Опишите результат, назовите возможные причины плохой работы модели (в случае, если результаты неудовлетворительные).

**Задача 7.** Создайте книгу Excel в Python. Подпишите «шапку» таблицы. Заполните все ячейки одним и тем же числом 123. Добавьте строки с суммой и средним значением.

**Задача 8.** Возьмём три строки:

```
s1 = 'daleonov@specialist.ru'
```

```
s2 = 'not an email'
```

```
s3 = "" " sadjflak asdkfjla asklfj daleonov@specialist.ru afasdk  
asdf la asf ;asl dfka
```

```
fas;dlfka user@gmail.com asklfjalsdf asfk info@mail.ru
```

```
sdfkljasl asldkfjal
```

""""

Первая, очевидно, почтой не является. Вторая является. Третья - это текст, содержащий адреса электронной почты, но сам по себе он не почта, а текст. Создайте регулярные выражения.

**Задача 9.** Вы хотите узнать на сайте Центробанка РФ курсы доллара с 1 по 31 января 2021 года, на ваш запрос появилась такая ссылка в браузере:  
[http://www.cbr.ru/scripts/XML\\_dynamic.asp?date\\_req1=01/01/2021&date\\_req2=31/01/2021&VAL\\_NM\\_RQ=R01235](http://www.cbr.ru/scripts/XML_dynamic.asp?date_req1=01/01/2021&date_req2=31/01/2021&VAL_NM_RQ=R01235) Это текст, который содержит интересующие вас данные. Какие? Охарактеризуйте их. Как называется данная выгрузка?

**Задача 10.** Напишите программу Pandas для отображения всех записей файла REGIONS.

**Задача 11.** Напишите программу Pandas для отображения имени, фамилии, оклада и идентификатора менеджера, где идентификаторы менеджера равны нулю.

**Задача 12.** Напишите программу Pandas для отображения всех идентификаторов местоположений из файла местоположений.

**Задача 13.** Напишите программу Pandas для извлечения первых 7 записей из файла сотрудников.

**Задача 14.** Напишите программу Pandas, чтобы выбрать отдельный идентификатор отдела из файла сотрудников.

**Задача 15.** Напишите программу Pandas, в которой будут

отображаться имя и фамилия, а также номер отдела для всех сотрудников, чья фамилия «Иванов».

**Задача 16.** Напишите программу Pandas для отображения имени, фамилии, оклада и идентификатора менеджера, где идентификаторы менеджера не равны нулю.

**Задача 17.** Напишите программу Pandas для создания и отображения логического ряда, где True для not null и False для null-значений или пропущенных значений в столбце state\_province файла местоположений.

**Задача 18.** Напишите программу Pandas для отображения имени, фамилии, оклада и номера отдела для тех сотрудников, менеджеры которых имеют ID 120, 103 или 145.

**Задача 19.** Напишите программу Pandas, чтобы отобразить имя, фамилию, оклад и номер отдела для тех сотрудников, для которых буква n является третьим символом в их имени.

**Задача 20.** Напишите программу Pandas для отображения имени, идентификатора задания, зарплаты и отдела для тех сотрудников, которые не работают в отделах 50, 30 и 80.

### 3 Определение варианта контрольной работы обучающимся

Таблица 1 – Номера контрольных вопросов для выполнения контрольной работы по дисциплине «Основы анализа данных на Python»

Предпоследняя цифра номера	Последняя цифра номера зачетной книжки									
	0	1	2	3	4	5	6	7	8	9
0	1,	2,	3,	4,	5,	6,	7,	8,	9,	10,
	13	14	15	16	17	18	19	20	21	22
1	11,	12,	13,	14,	15,	16,	17,	18,	19,	20,
	23	24	25	26	27	28	29	30	31	32
2	21,	22,	23,	24,	25,	26,	27,	28,	29,	30,
	33	34	35	36	37	38	39	40	41	42
3	10	2,	3,	4,	5,	6,	2,	3,	4,	5,
	43	44	45	46	47	48	49	50	51	52
4	8,	7,	6,	5,	4,	3,	2,	10	61,	60,
	53	54	55	56	57	58	59	60	2	3
5	41,	42,	43,	44,	45,	46,	47,	48,	49,	50,
	4	5	6	7	8	9	10	11	12	13
6	31,	32,	33,	34,	35,	36,	37,	38,	39,	40,
	14	15	16	17	18	19	20	21	22	23
7	21,	22,	23,	24,	25,	26,	27,	28,	29,	30,
	24	25	26	27	28	29	30	31	32	33
8	11,	12,	13,	14,	15,	16,	17,	18,	19,	20,
	34	35	36	37	38	39	40	41	42	43
9	2	3,	4,	5,	6,	7,	8,	9,	10,	11,
	44	45	46	47	48	49	50	51	52	53

Таблица 2 – Номера задач для выполнения контрольной работы по дисциплине «Основы анализа данных на Python»

Предпоследняя цифра номера	Последняя цифра номера зачетной книжки									
	0	1	2	3	4	5	6	7	8	9
0	2	3	4	5	6	7	8	9	10	12
1	13	14	15	16	17	18	19	20	2	1
2	2	3	4	5	6	7	8	9	10	11
3	12	13	14	15	6	17	18	19	20	2
4	1	2	3	4	5	6	7	8	9	10
5	11	12	13	14	15	16	17	18	19	20
6	2	1	2	3	4	5	6	7	8	9
7	10	11	12	13	14	15	16	17	18	19
8	20	2	1	2	3	4	5	6	7	8
9	9	10	11	12	13	14	15	16	17	18

## 4 Рекомендуемые источники информации

1. Блэнд, Д. Тестирование бизнес-идей / Дэвид Блэнд, Алекс Остервальдер ; пер. с англ. - Москва : Альпина Паблишер, 2020. - 354 с. - Режим доступа: <https://znanium.com/catalog/product/1222476>

2. Глебов, В. И. Практикум по математической статистике. Проверка гипотез с использованием Excel, MatCalc, R и Python : учебное пособие / В. И. Глебов, С. Я. Криволапов. — Москва : Прометей, 2019. — 86 с. - Режим доступа: <https://www.iprbookshop.ru/94504.html>

3. Гоулман, Д. Управление бизнесом / Гоулман Д., Друкер П., Дэвенпорт Т. - Москва : Альпина Пабли., 2016. - 290 с. (Harvard Business Review 10 лучших статей) ISBN 978-5-9614-5860-2. - Текст : электронный. - URL: <https://znanium.com/catalog/product/7018001>. Маккинли, Уэс Python и анализ данных / Уэс Маккинли; перевод А. Слинкина. — 2-е изд. — Саратов : Профобразование, 2019. — 482 с. - Режим доступа: <https://www.iprbookshop.ru/88752.html>

4. Гуриков, С. Р. Основы алгоритмизации и программирования на Python: учебное пособие / С.Р. Гуриков. — Москва: ФОРУМ: ИНФРА-М, 2020. — 343 с. — (Высшее образование: Бакалавриат). - Режим доступа: <https://znanium.com/catalog/product/1206074>

5. Кон, М. Agile. Оценка и планирование проектов: Практическое руководство / Кон М. - М.: Альпина Паблишер, 2018. - 418 с Режим доступа: <https://znanium.com/catalog/product/1003486>

6. Коттер, Д. Управление изменениями: Практическое руководство - М.: Альпина Паблишер, 2016. - 226 с.: Режим доступа: <https://znanium.com/catalog/product/1003136>

7. Программные системы статистического анализа. Обнаружение закономерностей в данных с использованием системы R и языка Python : учебное пособие / В. М. Волкова, М. А. Семёнова, Е. С. Четвертакова, С. С. Вожов. — Новосибирск: Новосибирский государственный технический университет, 2017. — 74 с. - Режим доступа: <https://www.iprbookshop.ru/91682.html>

8. Свод знаний по управлению бизнес-процессами. BPM СВОК 3.0: Учебное пособие / Под ред. Белайчук А.А. - Москва :Альпина Пабл., 2016. - 480 с. Режим доступа: <https://znanium.com/catalog/product/558829>

9. Сузи, Р. А. Язык программирования Python : учебное пособие / Р. А. Сузи. — 3-е изд. — Москва : Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 350 с. - Режим доступа: <https://www.iprbookshop.ru/97589.html>

10. Чиркова, Е. В. Как оценить бизнес по аналогии: Пособие по использованию сравнительных рыночных коэффициентов Учебное пособие / Чиркова Е.В., - 4-е изд., исправ. и доп. - М.:Альпина Паблицер, 2017. - 284 с. Режим доступа: <https://znanium.com/catalog/product/1002519>.

11. Шелудько, В. М. Язык программирования высокого уровня Python. Функции, структуры данных, дополнительные модули : учебное пособие / В. М. Шелудько. — Ростов-на-Дону, Таганрог : Издательство Южного федерального университета, 2017. — 107 с. - Режим доступа: <https://www.iprbookshop.ru/87530.html>

## Оглавление

Введение.....	3
1 Вопросы для контрольной работы .....	5
2 Задачи для контрольной работы .....	8
3 Определение варианта контрольной работы обучающимся .....	12
4 Рекомендуемые источники информации.....	14



# **ОСНОВЫ АНАЛИЗА ДАННЫХ НА PYTHON**

*Методические указания*

*Составители:*

**Васильева Надежда Константиновна**  
**Мороз Наталья Юрьевна**

**В авторской редакции**

Подписано в печать \_\_\_\_\_ 2021. Формат бумаги 60 × 84<sup>1/8</sup>  
Усл. печ. л. – 2,0. Уч.-изд. л. – 1,6.

Кубанский государственный аграрный университет.  
350044, г. Краснодар, ул. Калинина, 13