

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙ-  
СКОЙ ФЕДЕРАЦИИ**

**Федеральное государственное автономное  
образовательное учреждение высшего образования  
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»**

**Кафедра инфокоммуникаций**

**Анализ данных**

**Отчет по лабораторной работе №3.2**

**Тема: «Основы работы с библиотекой NumPy»**

Выполнил студент группы

ИВТ-б-о-21-1

Горшков В.И. «    » \_\_\_\_\_ 20\_\_ г.

Подпись студента \_\_\_\_\_

Работа защищена «    » \_\_\_\_\_ 20\_\_ г.

Проверил доцент

Кафедры инфокоммуникаций, старший  
преподаватель

Воронкин Р.А.

\_\_\_\_\_  
(подпись)

Ставрополь 2023

**Цель работы:** исследовать базовые возможности библиотеки NumPy языка программирования Python.

### Ход работы:

**1. Создал репозиторий в GitHub,** дополнил правила в .gitignore для работы с IDE PyCharm с ЯП Python, выбрал лицензию MIT, клонировал его на компьютер и организовал в соответствии с моделью ветвления git-flow.

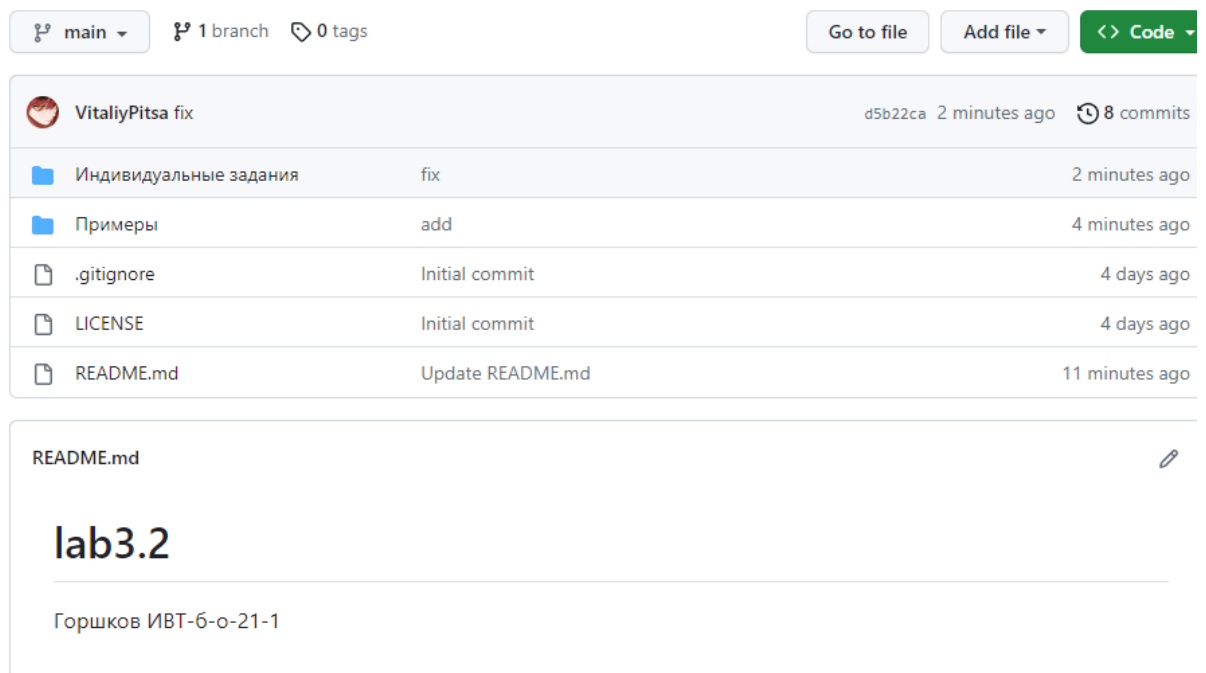
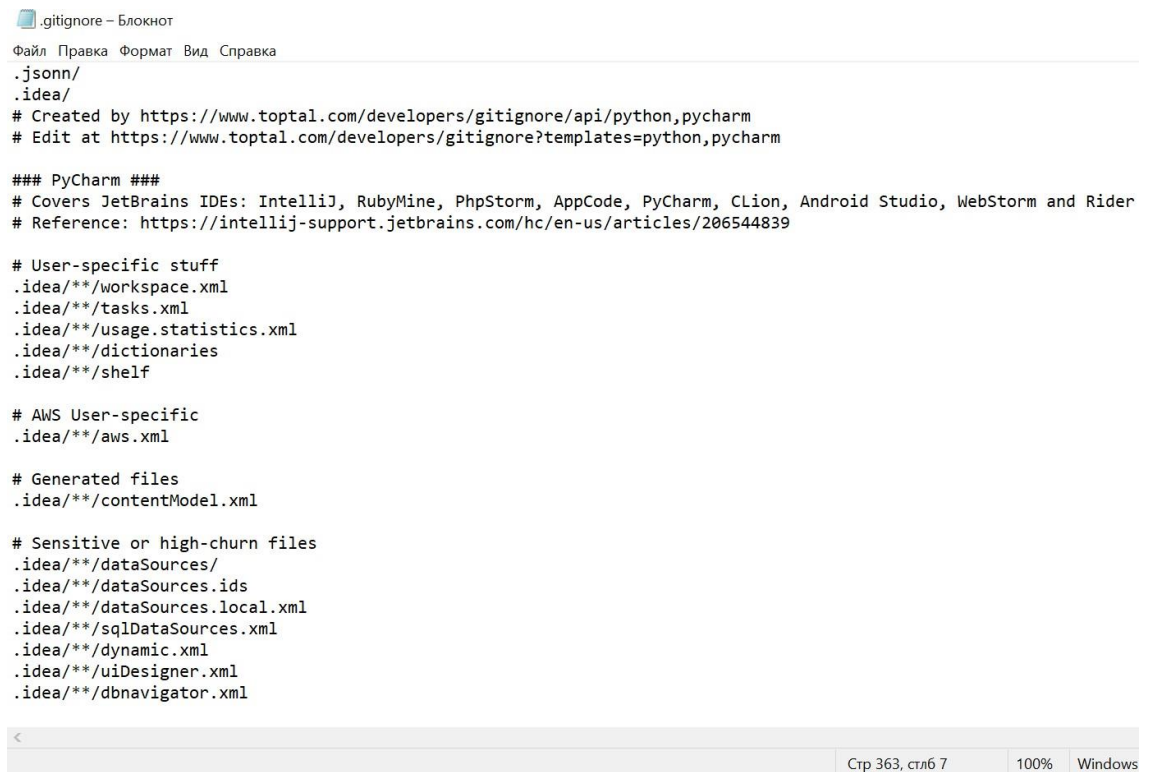


Рисунок 1.1 – Созданный репозиторий

A screenshot of a text editor window titled ".gitignore - Блокнот". The window contains a .gitignore file with various entries. The entries include: ".jsonn/", ".idea/", a comment about creation and editing links, a section for PyCharm IDEs, user-specific files like workspace.xml, tasks.xml, usage.statistics.xml, dictionaries, and shelf; AWS user-specific files; generated files like contentModel.xml; and sensitive or high-churn files like dataSources/, dataSources.ids, dataSources.local.xml, sqlDataSources.xml, dynamic.xml, uiDesigner.xml, and dbnavigator.xml. The editor has a menu bar with "Файл", "Правка", "Формат", "Вид", and "Справка". At the bottom, there is a status bar showing "Стр 363, столб 7", "100%", and "Windows".

```
.gitignore - Блокнот
Файл  Правка  Формат  Вид  Справка
.jsonn/
.idea/
# Created by https://www.toptal.com/developers/gitignore/api/python,pycharm
# Edit at https://www.toptal.com/developers/gitignore?templates=python,pycharm

### PyCharm ###
# Covers JetBrains IDEs: IntelliJ, RubyMine, PhpStorm, AppCode, PyCharm, CLion, Android Studio, WebStorm and Rider
# Reference: https://intellij-support.jetbrains.com/hc/en-us/articles/206544839

# User-specific stuff
.idea/**/workspace.xml
.idea/**/tasks.xml
.idea/**/usage.statistics.xml
.idea/**/dictionaries
.idea/**/shelf

# AWS User-specific
.idea/**/aws.xml

# Generated files
.idea/**/contentModel.xml

# Sensitive or high-churn files
.idea/**/dataSources/
.idea/**/dataSources.ids
.idea/**/dataSources.local.xml
.idea/**/sqlDataSources.xml
.idea/**/dynamic.xml
.idea/**/uiDesigner.xml
.idea/**/dbnavigator.xml
```

Рисунок 1.2 – Изменения в .gitignore

```

c:\Users\Admin\Desktop\git>git clone https://github.com/ItsMyLife1337/Python3.2.git
Cloning into 'Python3.2'...
remote: Enumerating objects: 4, done.
remote: Counting objects: 100% (4/4), done.
remote: Compressing objects: 100% (4/4), done.
remote: Total 4 (delta 0), reused 0 (delta 0), pack-reused 0
Receiving objects: 100% (4/4), done.

c:\Users\Admin\Desktop\git>cd /d c:\users\admin\desktop\git\Python3.2

c:\Users\Admin\Desktop\git\Python3.2>git flow init

Which branch should be used for bringing forth production releases?
- main
Branch name for production releases: [main]
Branch name for "next release" development: [develop]

How to name your supporting branch prefixes?
Feature branches? [feature/]
Bugfix branches? [bugfix/]
Release branches? [release/]
Hotfix branches? [hotfix/]
Support branches? [support/]
Version tag prefix? []
Hooks and filters directory? [C:/Users/Admin/Desktop/git/Python3.2/.git/hooks]

c:\Users\Admin\Desktop\git\Python3.2>

```

Рисунок 1.3 – Организация репозитория в соответствии с моделью ветвления git-flow

2. Создал 4 ноутбука и проработал примеры лабораторной работы.

<input type="checkbox"/>	 <a href="#">Пример №1 (Lab-3.2).ipynb</a>
<input type="checkbox"/>	 <a href="#">Пример №2 Lab-3.2.ipynb</a>
<input type="checkbox"/>	 <a href="#">Пример №3 Lab-3.2.ipynb</a>
<input type="checkbox"/>	 <a href="#">Пример №4 Lab-3.2.ipynb</a>

Рисунок 2 – Примеры

3. Решить задания в ноутбуках выданным преподавателем.

## Лабораторная работа 3.2. Знакомство с NumPy

Библиотека NumPy -- быстрая библиотека для математики в Python, основная структура данных -- массив `numpy.array`:

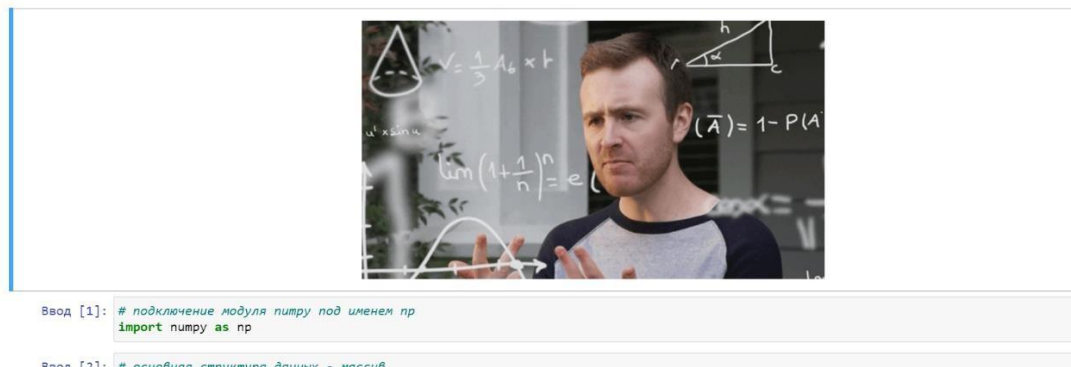


Рисунок 3 – Выполненные задания

### Задание №2

- Найдите интересный для вас датасет. Например, можно выбрать датасет тут: <http://data.un.org/Explorer.aspx> (выбираете датасет, жмете на view data, потом download, выбирайте csv формат)
- Рассчитайте подходящие описательные статистики для признаков объектов в выбранном датасете
- Проанализируйте и прокомментируйте содержательно получившиеся результаты
- Все комментарии оформляйте строго в ячейках формата markdown

```
Ввод [15]: import csv
import numpy as np

Ввод [16]: with open('data.csv', 'r', newline='') as csvf:
    data = csv.reader(csvf, delimiter=',')
    values = []
    years = []
    for row in data:
        if row[0] == "Country or Area":
            continue
        values.append(float(row[5]))
        years.append(int(row[2]))

values_arr = np.array(values)
years_arr = np.array(years)
```

Рисунок 4 – Выполненные задания во втором ноутбуке

**4. Создать ноутбук, в котором выполнить решение индивидуального задания.** Ноутбук должен содержать условие индивидуального задания. При решении индивидуального задания не должны быть использованы условный оператор `if`, а также операторы циклов `while` и `for`, а только средства библиотеки NumPy. Привести в ноутбуке обоснование принятых решений.

### В – 1. Дана целочисленная прямоугольная матрица. Определить:

- количество строк, не содержащих ни одного нулевого элемента;
- максимальное из чисел, встречающихся в заданной матрице более одного раза.

```

Ввод [9]: # 1, что необходимо сделать, это объявить матрицу и заполнить её случайными значениями. Для себя выбрал матрицу 4x5.

import numpy as np

a = np.random.randint(0, 50, (4, 5))
a

Out[9]: array([[47, 29, 43, 28, 19],
               [16, 19, 46, 31, 19],
               [24, 23, 44, 34, 17],
               [ 4, 40,  5, 13, 46]])

Ввод [10]: # Ищу количество строк не содержащих ни одного 0-вого элемента

rows, columns = a[np.where(a.min(axis=1) != 0)].shape
print('Количество строк количество строк, не содержащих ни одного нулевого элемента:', rows)

Количество строк количество строк, не содержащих ни одного нулевого элемента: 4

Ввод [11]: # Ищу максимальное число, которое встретиться в матрице больше 1 раза

vals, counts = np.unique(a, return_counts=True)
print(max(vals[counts > 1]))

46

```

Рисунок 5 – Индивидуальное задание №1

**5. Создать ноутбук, в котором выполнить решение вычислительной задачи** (например, задачи из области физики, экономики, математики, статистики и т. д.), условие которой предварительно необходимо согласовать с преподавателем.

Индивидуальное задание №2: найти площадь круга с радиусом  $r=3$ .

Для нахождения площади круга необходимо вычислить интеграл от функции  $f(x)=\sqrt{r^2 - x^2}$  на отрезке  $[-r, r]$ :  $S = 2 * \int(f(x), -r, r)$ .

```

jupyter Individual2 (автосохранение) Logout

File Edit View Insert Cell Kernel Widgets Help Не доверять Python 3 (ipykernel)

Ввод [1]: # Импортируем необходимые библиотеки
import numpy as np
from scipy import integrate
import matplotlib.pyplot as plt

Ввод [2]: # Радиус круга
r = 3

Ввод [3]: # Определяем функцию для подынтегрального выражения
def f(x):
    return np.sqrt(r**2 - x**2)

Ввод [4]: # Вычисляем интеграл
S, err = integrate.quad(f, -r, r)
print(f"Площадь круга с радиусом {r} равна {S:.2f} ед.")

Площадь круга с радиусом 3 равна 14.14 ед.

Ввод [6]: # Строим график подынтегральной функции
x = np.linspace(-r, r, 100)
y = f(x)
plt.plot(x, y, 'r-', label='f(x)')

# Подписываем оси и легенду
plt.xlabel('x')
plt.ylabel('y')
plt.legend(loc='best')

```

Рисунок 6 – Индивидуальное задание №2

**Вывод:** в результате выполнения лабораторной работы были получены базовые навыки работы с библиотекой NumPy языка программирования Python.

**Ответы на контрольные вопросы:**

## **1. Каково назначение библиотеки NumPy?**

Numpy – это библиотека для языка программирования Python, которая предоставляет в распоряжение разработчика инструменты для эффективно работы с многомерными массивами и высокопроизводительные вычислительные алгоритмы.

## **2. Что такое массивы ndarray?**

Ndarray — это (обычно фиксированный размер) многомерный контейнер элементов одного типа и размера. Количество измерений и элементов в массиве определяется его формой, которая является кортежем из N натуральных чисел, которые определяют размеры каждого измерения.

## **3. Как осуществляется доступ к частям многомерного массива?**

Через срезы:

- Произвольный элемент ( $m[i,j]$ )
- Строка ( $m[i, :]$ )
- Столбец матрицы ( $m[:, j]$ )
- Часть строки/столбца матрицы ( $m[i, j:], m[0:i, j]$ )
- Непрерывная часть матрицы ( $m[i1:i2, j1:j2]$ )
- Произвольные столбцы/строки матрицы ( $col = [0, 1, 2]; m[:, col]$ )

## **4. Как осуществляется расчет статистик по данным?**

shape – Размерность массива

argmax – Индексы элементов с максимальным значением (по осям)

argmin – Индексы элементов с минимальным значением (по осям)

max – Максимальные значения элементов (по осям)

min – Минимальные значения элементов (по осям)

mean – Средние значения элементов (по осям)

prod – Произведение всех элементов (по осям)

std – Стандартное отклонение (по осям)

sum – Сумма всех элементов (по осям)

var – Дисперсия (по осям)

## **5. Как выполняется выборка данных из массивов ndarray?**

Если мы переменную, содержащую boolean-значение передадим в качестве списка индексов для массива (nums), то получим массив, в котором будут содержаться элементы из nums с индексами равными индексам True позиций boolean-массива, графически это будет выглядеть так.

