**МИНОБРНАУКИ РОССИИ**

**федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ**

**УНИВЕРСИТЕТ»**

**(СПбГЭУ)**

**ОТЧЕТ**

**по учебной практике**

**(научно-исследовательская работа**

**(получение первичных навыков научно-исследовательской работы))**

Наименование организации прохождения практической подготовки:

кафедра ПМ и ЭММ СПбГЭУ

*(наименование организации)*

Направление: 01.03.02 - Прикладная математика и информатика

*(шифр, наименование)*

Направленность: Прикладная математика и информатика в экономике и управлении

*(наименование)*

Обучающийся Яковлев Никита Андреевич

*(Ф.И.О. полностью)*

Группа ПМ-1901 Подпись\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(номер группы)*

Руководитель

по практической подготовке от СПбГЭУ

Лебедева Людмила Николаевна, к.ф.-м.н., доцент

*(Ф.И.О., ученая степень, должность, ученое звание)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись руководителя)*

|  |  |
| --- | --- |
| Оценка по итогам защиты отчета\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ |  |

Санкт-Петербург

2022 г.

**МИНОБРНАУКИ РОССИИ**

**федеральное государственное бюджетное образовательное учреждение**

**высшего образования**

**«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ ЭКОНОМИЧЕСКИЙ**

**УНИВЕРСИТЕТ»**

**(СПбГЭУ)**

|  |  |
| --- | --- |
| СОГЛАСОВАНО:  Руководитель по практической подготовке от профильной организации  *(заполняется в случае прохождения практической подготовки в профильной организации)*  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (Ф.И.О., должность)  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись)  **«\_\_\_\_\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20\_\_\_г.**  **М.П.** | УТВЕРЖДАЮ:  Заведующий кафедрой прикладной математики и экономико-математических методов, доктор технических наук, профессор  Фридман Григорий Морицович  (Ф.И.О.)  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_  (подпись)  **«\_\_\_\_\_\_\_\_\_» \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_20\_\_\_г.** |

**Индивидуальное задание**

**на учебную практику**

(научно-исследовательская работа

получение первичных навыков научно-исследовательской работы)

**Обучающегося: 3 курса обучения Яковлев Никита Андреевич**

(курс обучения) (Ф.И.О. полностью)

**Направление/специальность:** 01.03.02 Прикладная математика и информатика

**Направленность (профиль)/специализация:** Прикладная математика и информатика в экономике и управлении

**Тема ВКР/НКР**: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(заполняется в случае прохождении преддипломной практики)

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Наименование организации прохождения практической подготовки** СПбГЭУ,

кафедра Прикладной математики и экономико-математических методов

**Сроки практической подготовки** 7.06.2022 – 21.06.2022

**Руководитель по практической подготовке от СПбГЭУ**

Лебедева Людмила Николаевна к.ф.-м.н., доцент кафедры Прикладной математики и экономико-математических методов

(Ф.И.О. полностью) (Должность)

Совместный рабочий график

с указанием видов работ, связанных с будущей профессиональной деятельностью

|  |  |  |
| --- | --- | --- |
| **№ п/п** | **Перечень заданий, подлежащих разработке** | **Календарные сроки**  **(даты выполнения)** |
| 1 | Прохождение инструктажа по технике безопасности | 07.06.2022 |
| 2 | Ознакомление с ЛНА | 07.06.2022 |
| 3 | Согласование индивидуального задания | 07.06.2022 |
| 4 | Анализ источников | 09.06.2022 |
| 5 | Сбор данных с интерне ресурса **monitoring.miccedu.ru** | 13.06.2022 |
| 6 | Систематизация и обработка данных | 15.06.2022 |
| 7 | Визуализация данных | 17.06.2022 |
| 8 | Проведение анализа показателей | 19.06.2022 |
| 9 | Подготовка и оформление отчета | 20.06.2022 |

**С заданием ознакомлен(а)** \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись обучающегося)

**Руководитель по практической подготовке от СПбГЭУ**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Подпись) (Расшифровка)

**Руководитель по практической подготовке от профильной организации**

*(заполняется в случае прохождения практической подготовки в профильной организации)*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Подпись) (Расшифровка)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Обучающийся прошел инструктаж по ознакомлению с требованиями охраны труда, техники безопасности, пожарной безопасности, а также с правилами внутреннего распорядка. Вводный инструктаж и инструктаж на рабочем месте пройдены с оформлением установленной документации.

Руководитель по практической подготовке от организации/профильной организации назначен приказом № \_\_\_\_\_ дата \_\_\_\_\_\_\_\_ и соответствует требованиям трудового законодательства Российской Федерации о допуске к педагогической деятельности.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(Ф.И.О. должность)

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

(подпись)

**Содержание**

[Введение 5](#_Toc106730061)

[1. Анализ источников 6](#_Toc106730062)

[2. Сбор и систематизация данных 8](#_Toc106730063)

[2.1. Описание HTML-структуры интернет-ресурса 8](#_Toc106730064)

[2.2. Реализация вспомогательного модуля для сбора информации 12](#_Toc106730065)

[2.3. Реализация парсера 15](#_Toc106730066)

[3. визуализация данных 17](#_Toc106730067)

[3.1. Подготовка данных к визуализации 17](#_Toc106730068)

[3.2. Реализация визуализации 19](#_Toc106730069)

[3.3. Визуализация 20](#_Toc106730070)

[4. Анализ показателей 34](#_Toc106730071)

[4.1. Постановка гипотез о взаимосвязях показателей 34](#_Toc106730072)

[4.2. Построение матриц корреляции и визуализация средствами Python 34](#_Toc106730073)

[4.3. Корреляционный анализ 36](#_Toc106730074)

[Заключение 40](#_Toc106730075)

[Список использованных источников 41](#_Toc106730076)

# Введение

В современном мире множество компаний стремятся собрать различного рода информацию: о ситуации на рынке определённых благ, данные опросов пользователей товаров других компаний, информацию о том, кто совершал целевое действие на сайте и многое другое. Собранная информация в последствии обрабатывается и систематизируется для дальнейшего анализа, прогнозов, кластеризации по признакам или же выявления силы (тесноты) корреляционных связей между различными наборами данных, что позволяет в свою очередь определить цель компании и последовательность действий ведущих к ней.

Целью учебной практики является на примере показателей деятельности ФБОУ ВО «Санкт-Петербургский государственный экономический университет»: закрепить теоретические знания, полученные при изучении учебных дисциплин; улучшить навыки сбора, обработки, систематизации и анализа информации.

# Анализ источников

В качестве первого источника была выбрана книга Райана Митчелла «Скрапинг веб-сайтов с помощью python». В ней автор описывает способы навигации по дереву синтаксического разбора, используя библиотеку  языка программирования Python:

1. Проход сверху вниз по дереву, используя свойства объектов , такие как (хранит генератор всех прямых дочерних элементов), (хранит список всех прямых дочерних элементов), (хранит итератор всех дочерних элементов) или свойства имени тегов;
2. Проход снизу вверх по дереву, используя свойства объектов класса  и ;
3. Проход по элементам одного уровня (проход вбок по дереву), используя свойства (свойство хранит элемент, определённый после элемента, обладающим этим свойством, в HTML-документе на том же уровне), (свойство хранит элемент, определённый до элемента, обладающим этим свойством, в HTML-документе на том же уровне), а также с помощью свойств и.
4. Методы поиска и класса .

Также описываются основные классы библиотеки и способы работы с ними, работу с интерфейсами программирования приложений. Автор показывает работу с библиотекой, которая содержит функции для запроса данных в сети, обработки cookies и изменения метаданных. В книге автор применяет её в основном, для извлечения страниц с помощью функции , и конвертацию в объект для дальнейшего парсинга.

Второй источник – документация библиотеки ( 4.9.0 Documentation). В ней описывается как устроена библиотека, то есть её структура, классы библиотеки, навигация по HTML-документу, функции извлечения содержимого тегов, функции поиска и изменения дерева, объясняется кодировка вывода и смена кодировки содержимого тегов.

Третьим источником является книга американского разработчика программного обеспечения Уэса Маккинни «Python и анализ данных». В данном учебнике автор описал основные библиотеки языка программирования Python используемых для анализа данных:

1. Библиотека , которая является основой огромного числа библиотек;
2. Библиотека , которая предоставляет большое количество инструментов при работе со структурированными данными;
3. Библиотека , необходимая для визуализации различных процессов, благодаря чему можно выдвигать предположения о распределении случайных величин, в какой зависимости находятся различные показатели и многое другое.

Автор даёт базовые знания о группировании данных и временных рядах, упражнениям, которые позволят применить в реальной жизни.

# Сбор и систематизация данных

В ходе работы необходимо было собрать данные с интернет-ресурса «Мониторинг (ВО)» об образовательной деятельности федерального государственного бюджетного образовательного учреждения высшего образования «Санкт-Петербургский государственный экономический университет» с 2015 по 2021 года с последующей конвертацией данных в DataFrame объекты языка программирования Python.

## Описание HTML-структуры интернет-ресурса

Данные представлены в HTML-документе сайта в виде таблиц и определены с помощью тега table. Всего на сайте размещено по десять таблиц в период с 2015 по 2018 год и по девять таблиц в периоде с 2019 по 2021 год, которые находятся в разделах: «Сведения по показателям мониторинга эффективности деятельности», «Результаты мониторинга образовательной организации по направлениям деятельности» и «Роль организации в системе подготовки кадров для региона, дополнительные характеристики образовательной организации».

Таблица, находящаяся в разделе «Сведения по показателям мониторинга эффективности деятельности»:

1. Позиции организации по основным показателям в сравнении с пороговыми значениями.

Таблицы, находящиеся в разделе «Результаты мониторинга образовательной организации по направлениям деятельности»:

1. Образовательная деятельность;
2. Научно-исследовательская деятельность;
3. Международная деятельность;
4. Финансово-экономическая деятельность;
5. Инфраструктура;
6. Трудоустройство (данная таблица представлена лишь на страницах в периоде с 2015 по 2018 год);
7. Кадровый состав.

Таблица, находящаяся в разделе «Роль организации в системе подготовки кадров для региона»:

1. О роли организации в системе подготовки кадров для региона.

Таблица, находящаяся в разделе «Дополнительные характеристики образовательной организации»:

1. О дополнительных характеристиках образовательной организации.

Некоторые вышеуказанные таблицы имеют различные структуры, что требует написать различные методы, которые будут осуществлять сбор информации и её последующее преобразование в объекты языка Python. Все таблицы можно разделить на классы опираясь на их структуру.

Так, например, таблицы из раздела «Результаты мониторинга образовательной организации по направлениям деятельности» имеют одну и ту же структуру: в каждой строке таблицы находится одно и тоже количество элементов. Перейдя в HTML-документ, можно также заметить, что все таблицы из данного раздела имеют один и тот же атрибут со значением (Рисунок №1. Структура таблиц из второго раздела), что позволит осуществлять навигацию по документу проще. Таким образом, таблицы с данного раздела будут образовывать первый класс.

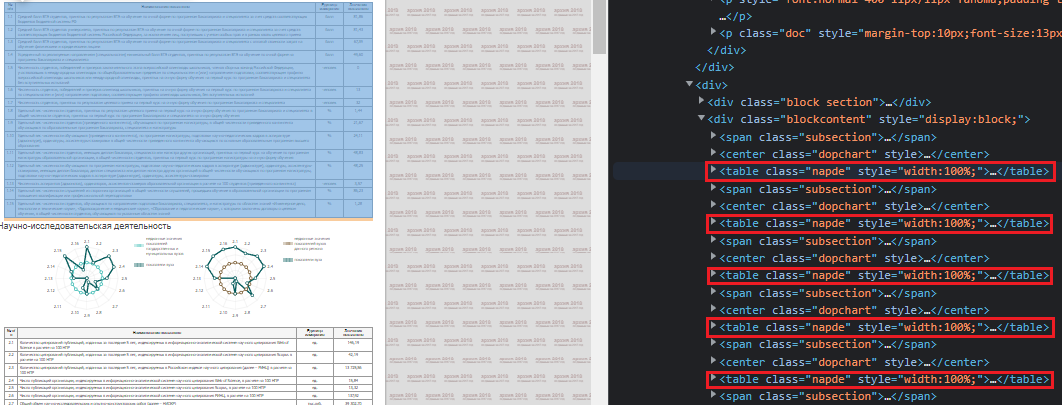


Рисунок №1. Структура таблиц из второго раздела.

Таблицы из третьего и четвёртого разделов имеют одинаковые структуры: в каждой из них существуют строки с одним элементом, который необходим для того, чтобы разделить таблицу на составляющие (Рисунок №2. Структура таблиц из третьего и четвёртого разделов).

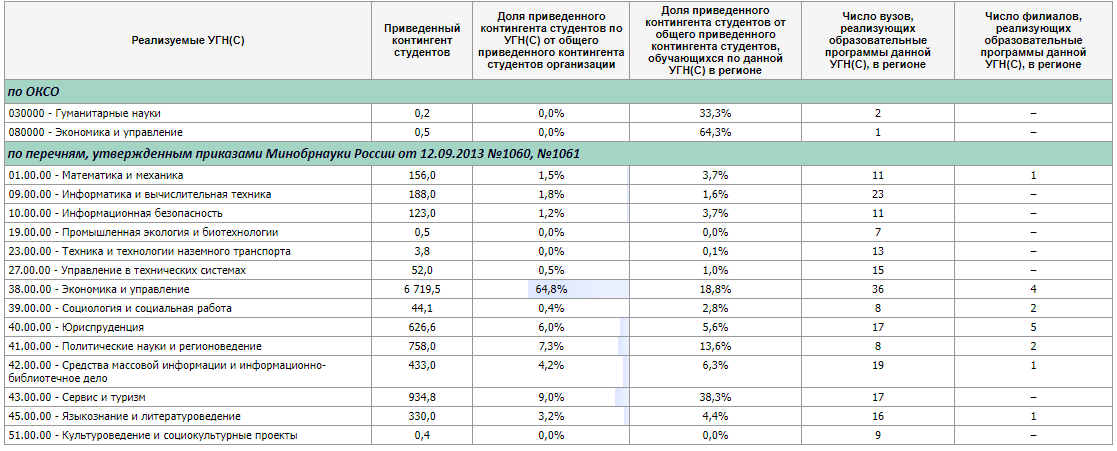


Рисунок №2. Структура таблиц из третьего и четвёртого разделов.

Открыв HTML-документ можно заметить, что данные таблицы имеют уникальные идентификаторы и соответственно (Рисунок №3. Атрибуты таблиц из третьего и четвёртого разделов).

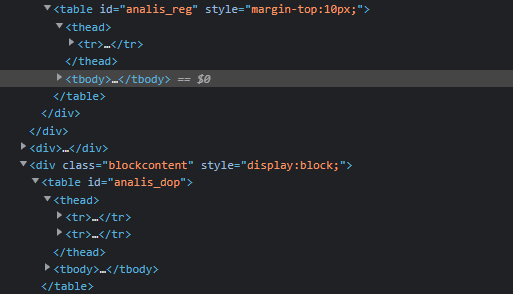


Рисунок №3. Атрибуты таблиц из третьего и четвёртого разделов.

Таблица, находящаяся в первом разделе, имеет разную структуру в зависимости от периода, за который были собраны даны. Например, в данной таблице за 2020 год в теге определены два дочерних тега , то есть имена полей таблицы были введены с помощью двух строк таблицы, сделано это было для того, чтобы разделить один столбец на несколько (Рисунок №5. Структура таблицы из первого раздела за 2020 год). В таблицах за период 2015 по 2019 год, нет разбиения столбца на подстолбцы (Рисунок №6. Структура таблиц из первого раздела с 2015 по 2019 год), а в таблице за 2021 год определены два тега , один из которых пустой. Следовательно, сбор информации с этих таблиц будет осуществляться индивидуально. Важно отметить, что все таблицы имеют атрибут со значением result (Рисунок №4. Атрибуты таблицы из 1 раздела).

Graphical user interface

Description automatically generated with medium confidence

Рисунок №4. Атрибуты таблицы из 1 раздела.

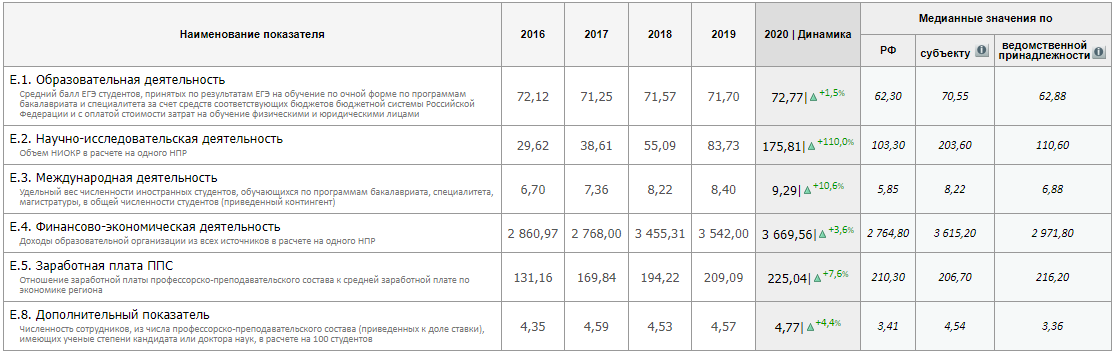


Рисунок №5. Структура таблицы из первого раздела за 2020 год.

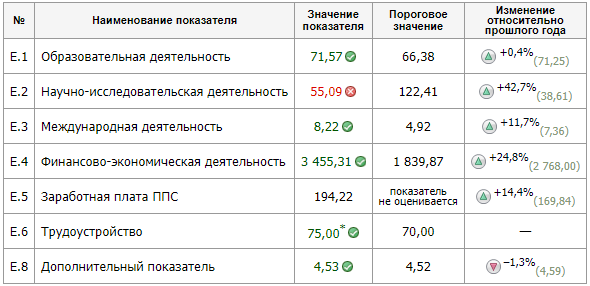


Рисунок №6. Структура таблиц из первого раздела с 2015 по 2019 год.

## Реализация вспомогательного модуля для сбора информации

После того как таблицы были разделены на классы будет удобно описать функции, которые автоматически будут производить сбор информации с таблиц интернет-ресурса и создавать объекты, которые будут хранить полученную информацию в удобной для дальнейшего анализа форме. Также стоит отметить, что данный подход позволяет не нарушать принцип DRY («Don't Repeat Yourself»), так как позволяет не дублировать код и увеличивает его читабельность. Данные функции определяются в отдельном файле .

В реализации парсера используется навигация по дереву HTML документа от предка к дочерним элементам. При таком подходе может возникать эффект, когда знаки табуляции воспринимаются как отдельный HTML элемент, поэтому необходимо написать функцию, которая будет очищать список дочерних элементов от знаков табуляции, полученный с помощью свойства contents. В данной реализации эта функция носит имя (Рисунок №7. Функция clear\_).

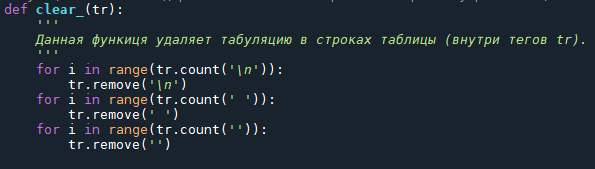


Рисунок №7. Функция .

Также необходимо определить функцию, которая будет возвращать список содержимого тегов: на практике, чтобы получить содержимое тегов (Рисунок №8. Функция ).

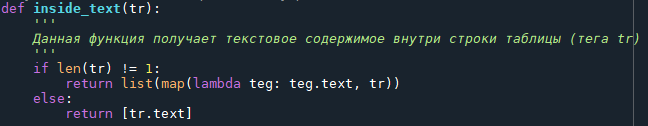


Рисунок №8. Функция

Функция , которая принимает тег (тело таблицы) и создаёт двумерный список с содержимым тегов в разных строках таблицы (Рисунок №9. Функция ).

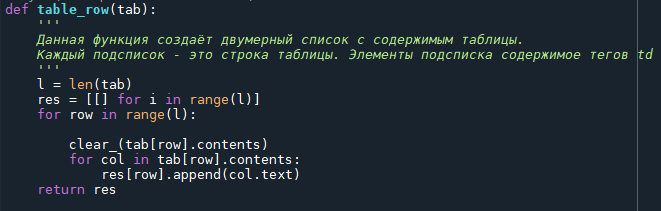


Рисунок №9. Функция .

Функция , которая принимает на вход имена полей таблицы и тело таблицы, преобразует двумерный список с полученными данными из тега в таблицу и возвращает её (Рисунок №10. Функция ).

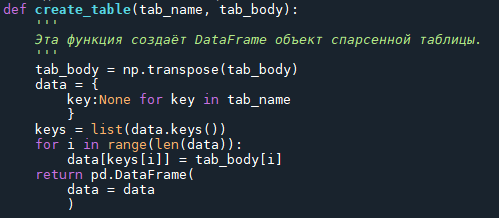


Рисунок №10. Функция .

Функция , которая принимает HTML-документ и количество таблиц из раздела «Результаты мониторинга образовательной организации по направлениям деятельности», преобразует их в объекты и возвращает список получившихся таблиц (Рисунок №11. Функция ).

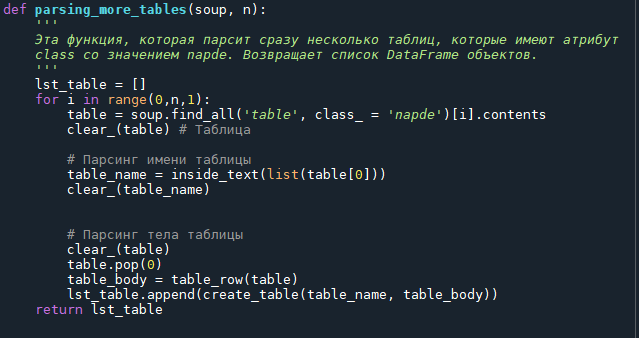


Рисунок №11. Функция .

Функция , которая принимает HTML-документ и тега таблицы, в которой определены строки с одним элементом, необходимым для разбиения таблиц на составляющие (Рисунок №12. Функция ). В ней происходит разбиение таблицы на подтаблицы и преобразование в объекты. Возвращает функция, список объектов, которые являются подтаблицами передаваемой таблицы. Данная функция определена для парсинга таблиц из разделов «Результаты мониторинга образовательной организации по направлениям деятельности» и «Роль организации в системе подготовки кадров для региона, дополнительные характеристики образовательной организации».

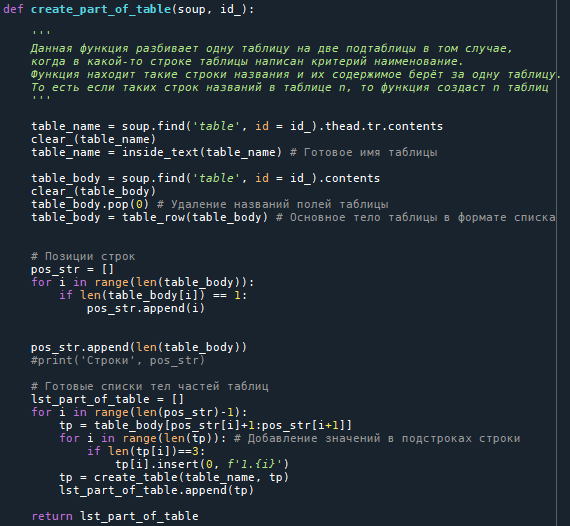


Рисунок №12. Функция .

Также существуют ещё две функции, работающие аналогично функции и их определение связано с тем, что таблицы из вышеуказанных разделов имеют в имени полей таблиц изображение, которое нельзя использовать в качестве наименования полей таблицы. В этих функциях вручную прописывались имена полей для таблиц.

## Реализация парсера

С готовыми инструментами можно переходить к реализации парсера. Импортируются библиотека , конструктор класса и написанный ранее модуль .

Парсер, который будет использоваться для сбора информации, HTML-парсер в . Выбор обусловлен большой скоростью парсера и нестрогостью ( 4.9.0 документация/Установка парсера).

Переменной присваивается строка, содержащая адрес интернет-ресурса, с которого будет происходить сбор информации за определённый год. Осуществляется запрос на необходимый сайт, затем переменной присваивается содержимое HTML-документа, по которому затем ведётся поиск необходимых таблиц, с помощью функций из модуля  и создание объектов, которые хранят информацию из каждой таблицы интернет-ресурса. Как было сказано выше, парсинг таблицы из раздела «Сведения по показателям мониторинга эффективности деятельности» происходит индивидуально. После того, как все таблицы сформированы, их помещают в список. Пример парсера для сбора информации за 2015 год, показано на Рисунке №13. Файл с парсером. Для каждого года создавался отдельный файл, который носит имя

Text

Description automatically generated

Рисунок №13. Файл с парсером.

# визуализация данных

## Подготовка данных к визуализации

После того как все таблицы за каждый год сформированы их необходимо поместить в один массив для дальнейшей работы с ними. В файле  импортируем списки таблиц за все года (Рисунок №14. Файл ). Определяется функция flattenlist для создания одномерного списка и функция для объединения всех списков таблиц по годам в один словарь, ключами которого являются строки с номером года, а значением список таблиц по годам.

Получившийся словарь хранится в переменной (словарь и его элементы представлены на рисунках №15–№17).

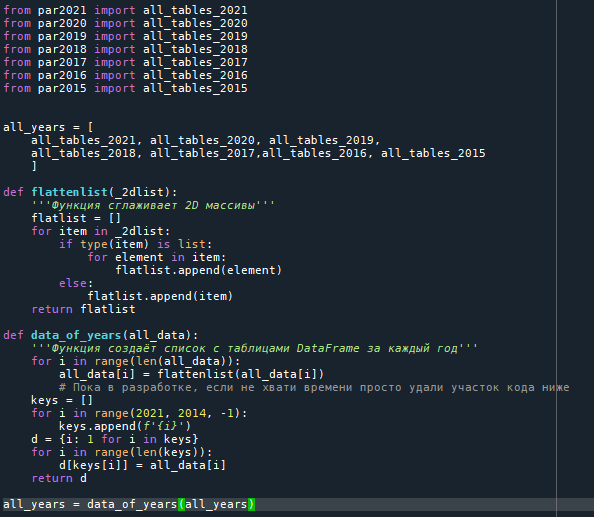


Рисунок №14. Файл .

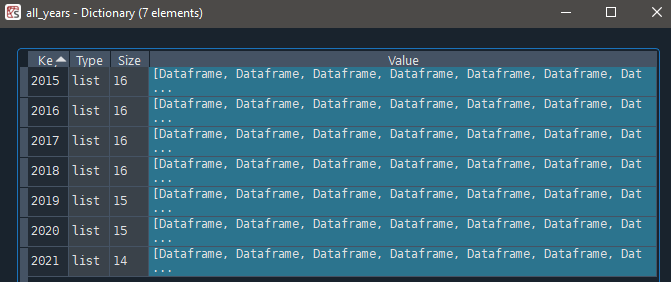


Рисунок №15. Словарь .

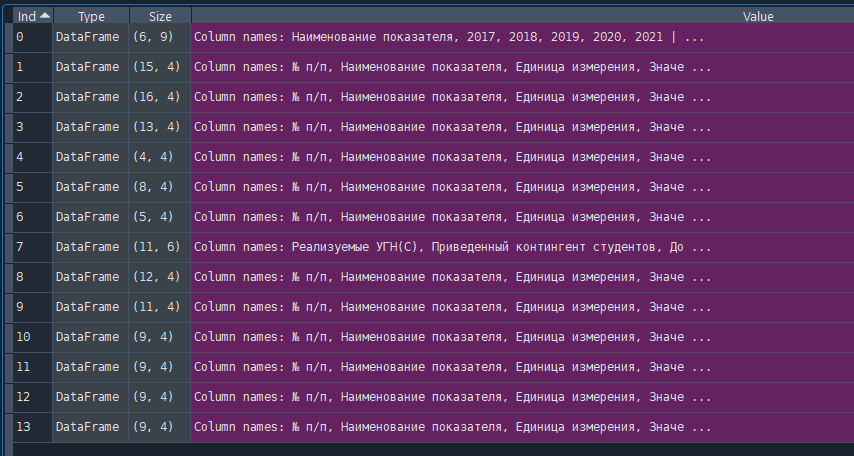


Рисунок №16.Элементы словаря .

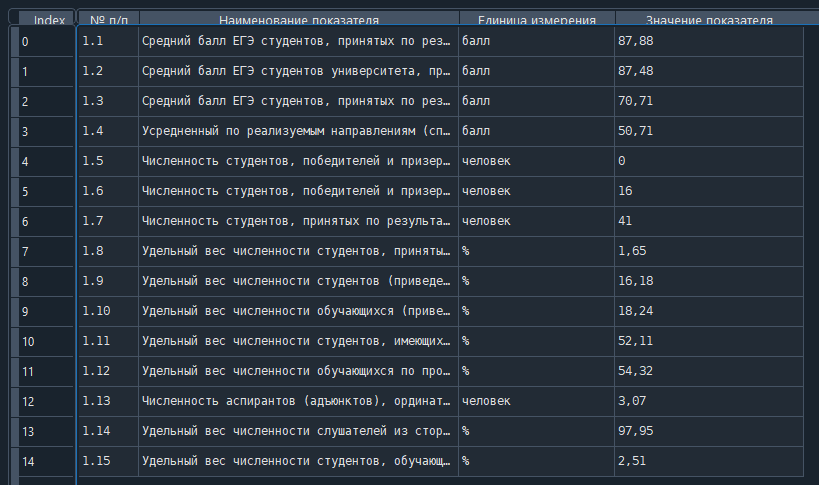


Рисунок №17. Пример таблицы из .

## Реализация визуализации

Все объекты имеют поля «№ п/п» и «Единица измерения», которые необходимо удалить, так как они будут мешать визуализации. В файле создаются словари ключами которых являются строки с номерами года, а значениями являются объекты с индексами «Значения показателя» и «наименования показателя» (Рисунок №19. Определение словаря для визуализации). Далее по словарю изымая элементы под индексом «Значения показателя», строятся гистограммы, с помощью библиотеки (Рисунок №20. Пример кода, отвечающего за визуализацию). Так как элементы под индексами «Значения показателя» типа данных , то их необходимо конвертировать в тип данных . Для этого была написана функция (Рисунок №18. Функция ), которая принимает список со строковыми значениями, осуществляет замену запятой на точку и применяет функцию , для каждого строки из списка, на выходе возвращается список с элементами класса , который в последствии передаётся в функцию для построения диаграммы.

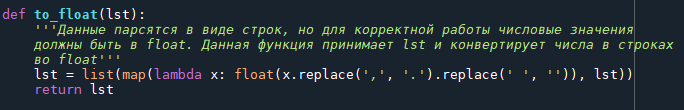


Рисунок №18. Функция .

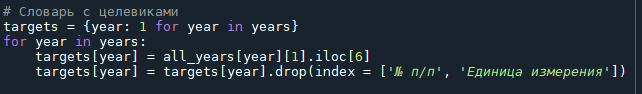


Рисунок №19. Определение словаря для визуализации.

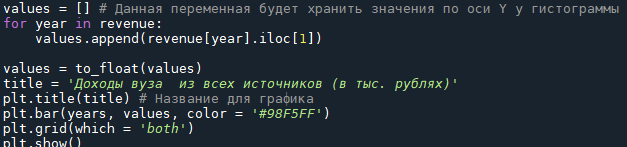


Рисунок №20. Пример кода, отвечающего за визуализацию.

## Визуализация

В результате визуализации получились графики, представленные на рисунках № 21–№47.

Chart, histogram

Description automatically generated

Рисунок №21. Средний балл ЕГЭ студентов, принятых на бюджетное обучение.

Как показывает график средний балл ЕГЭ у студентов, принятых на обучение, с каждым годом увеличивался по сравнению с предыдущим, за исключением 2017 года, когда средний балл ЕГЭ по сравнению с предыдущим упал.

Chart, bar chart

Description automatically generated

Рисунок №22. Средний балл ЕГЭ студентов, принятых на платное обучение.

Среди студентов, которые поступили с оплатой стоимости затрат на обучение физическими и юридическими лицами, можно наблюдать, что с периода 2016 по 2020 год, средний балл ЕГЭ не изменялся, с 2015 по 2016 год и с 2020 по 2021 год произошёл рост.Chart, bar chart

Description automatically generated

Рисунок №23. Численность победителей Всероссийской олимпиады школьников.

Можно видеть, что среди студентов, поступавших на первый курс в период с 2015 по 2021 год, победители Всероссийской олимпиады школьников были только в 2019 году. Число студентов-призёров в 2019 году - 10 человек.

Chart, bar chart

Description automatically generated

Рисунок №24. Численность победителей и призёров олимпиад школьников.

Число студентов победителей и призёров олимпиад школьников, принятых на первый курс, было максимально в 2015 году: около 55 человек. В 2016 году их число сократилось на 10 студентов. В период с 2017 по 2018 год, число победителей и призёров было около 13 человек, в 2019 около 22. В 2020 году число победителей и призёров было минимальным в рассматриваемом промежутке времени: около 4 человек. В 2021 году число победителей и призёров среди первокурсников составило 17 человек.

Chart

Description automatically generated

Рисунок №25. Численность студентов, принятых по результатам целевого приёма.

Chart, bar chart

Description automatically generated

Рисунок №26. Удельный вес численности магистров, имеющих дипломы других образовательных учреждений.

По графику видно, что среди людей, принятых на первый курс по программам магистратуры, удельный вес студентов имеющих дипломы с других университетов постоянно рос, в периоде с 2016 по 2021 год. В 2015 году удельный вес таких студентов составил более 60%.

Chart

Description automatically generated

Рисунок №27. Удельный вес численности магистров и аспирантов, имеющих дипломы других образовательных учреждений.

Почти аналогичная ситуация наблюдается среди студентов, поступивших на первый курс по программам магистратуры и научно-педагогических кадров в аспирантуре.

Chart

Description automatically generated

Рисунок №28. Удельный вес численности магистров от общей численности студентов.

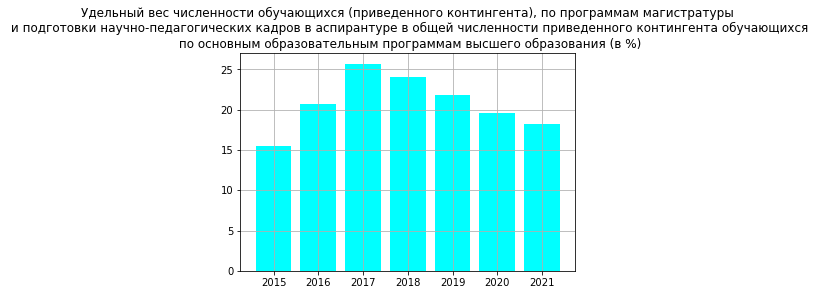


Рисунок №29. Удельный вес численности магистров и аспирантов от общей численности студентов.

Chart

Description automatically generated

Рисунок №30. Общий объём научно-исследовательских и опытно-конструкторских работ.

Chart, histogram

Description automatically generated

Рисунок №31.Удельный вес численности иностранных студентов не из СНГ.

Chart, bar chart, histogram

Description automatically generated

Рисунок №31.Удельный вес численности иностранных студентов из СНГ.

В тот же период времени можно наблюдать, что удельный вес численности иностранных студентов из СНГ с каждым годом понижается.

Chart, bar chart

Description automatically generated

Рисунок №32. Объём средств, полученных на выполнение НИОКР от иностранных граждан и иностранных юридических лиц.

Chart, bar chart

Description automatically generated

Рисунок №33. Объём средств, полученных на от иностранных граждан и иностранных юридических лиц.

Chart

Description automatically generated

Рисунок №34. Доля иностранных студентов в общей численности студентов.

Chart

Description automatically generated

Рисунок №35. Доходы вуза из всех источников.

Chart

Description automatically generated

Рисунок №36. Доходы вуза из внебюджетных источников.

Chart

Description automatically generated

Рисунок №37. Численность студентов на очной форме обучения.

Можно наблюдать, что с 2015 года по 2018 год, число студентов обучающихся на очной форме обучения постоянно снижалась, но после 2018 года наблюдается рост.

Chart, histogram

Description automatically generated

Рисунок №38. Численность студентов на очно-заочной форме обучения.

По данному графику можно видеть, что численность студентов обучающихся по очно-заочной форме, с каждым годом снижалась почти в 2 раза, за исключением 2021 года, когда численность таких студентов осталось такой же как и в 2020 году.

Chart, histogram

Description automatically generated

Рисунок №39. Численность студентов на заочной форме обучения.

Также можно видеть, что с каждым годом происходило снижение численности студентов, обучающихся на заочной форме обучения.

Chart, histogram

Description automatically generated

Рисунок №40. Общее количество публикаций организации в расчёте на 100 НПР.

Chart

Description automatically generated

Рисунок №41. Общая численность аспирантов, интернов, ординаторов, ассистентов-стажёров.

Chart

Description automatically generated

Рисунок №42. Удельный вес НПР, имеющих степень кандидата наук.

Chart, histogram

Description automatically generated

Рисунок №43. Удельный вес НПР, имеющих степень доктора наук.

Chart

Description automatically generated

Рисунок №44. Общая численность профессорско-преподавательского состава.

На данном графике можно наблюдать сокращение числа профессорско-педагогического состава учебного заведения.

Chart

Description automatically generated

Рисунок №45. Общая численность научных работников.

Chart

Description automatically generated

Рисунок №46. Средняя заработная плата ППС.

На графике можно наблюдать ежегодный рост средней зарплаты профессорско-педагогического состава.

Chart, bar chart

Description automatically generated

Рисунок №47. Средняя заработная плата научных работников.

# Анализ показателей

## Постановка гипотез о взаимосвязях показателей

Для проверки предположений о взаимосвязи процессов, происходящих в учебном заведении, будет применяться корреляционный анализ.

По графикам представленными в пункте визуализация было видно, что число профессорско-преподавательского состава сильно сокращалось каждый год. Также было видно сокращение числа студентов обучающихся на очном обучении и особенно сильное сокращение числа студентов, обучающихся на очно-заочном и заочном формах обучения. С помощью матрицы корреляции и тепловой карты можно установить связаны ли были эти события. Также на сокращение числа профессорско-преподавательского состава могли повлиять и другие факторы: изменение зарплат, изменение доходов высшего учебного заведения из всех источников, изменение количества аспирантов каждый год.

В период с 2015 по 2021 год наблюдалось повышение заплат профессорско-преподавательского состава и научных работников. Это могло быть связано с изменениями доходов высшего учебного заведения со всех источников, сокращением числа профессорско-преподавательского состава и научных работников, доходами от Научно-исследовательские и опытно-конструкторские работы (НИОКР).

Также было бы интересно узнать, от чего зависит количество научных публикаций. Связано ли это с числом аспирантов, зарплатами профессорско-преподавательского состава, средствами, выделяемыми на НИОКР, общим объем научно-исследовательских и опытно-конструкторских работ.

## Построение матриц корреляции и визуализация средствами Python

Для проверки предположений о взаимосвязи процессов, происходящих в учебном заведении, будет применяться корреляционный анализ.

Для построения матриц корреляции будут использоваться функции, предоставляемые библиотекой , а для визуализации будут использоваться библиотеки и .

Код написан в отдельном файле: . Происходит импортирование необходимых модулей, также потребуются словари, находящиеся в файле (Рисунок №48. Импорт модулей).

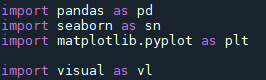


Рисунок №48. Импорт модулей.

Для проверки предположений о взаимосвязях необходимо создать таблицы, которые будут содержать данные об изменениях показателей в периоде с 2015 по 2021 год. Создаётся переменная s, которая хранит список строк, содержащие номера годов (Рисунок №49. Список).



Рисунок №49. Список years.

Далее идёт обработка данных из словаря, а именно с помощью функции идёт преобразование строк, содержащие численные показатели, в объекты класса (Рисунок №50. Обработка словарей с показателями).

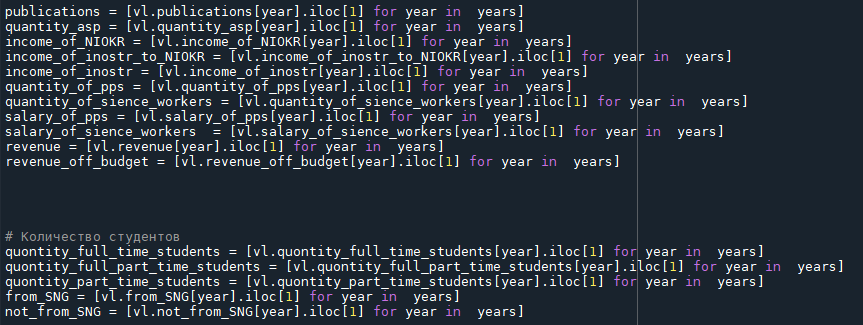


Рисунок №50. Обработка словарей с показателями.

Далее происходит создание таблиц, для построения корреляционных матриц (Рисунок №51. Создание таблиц).

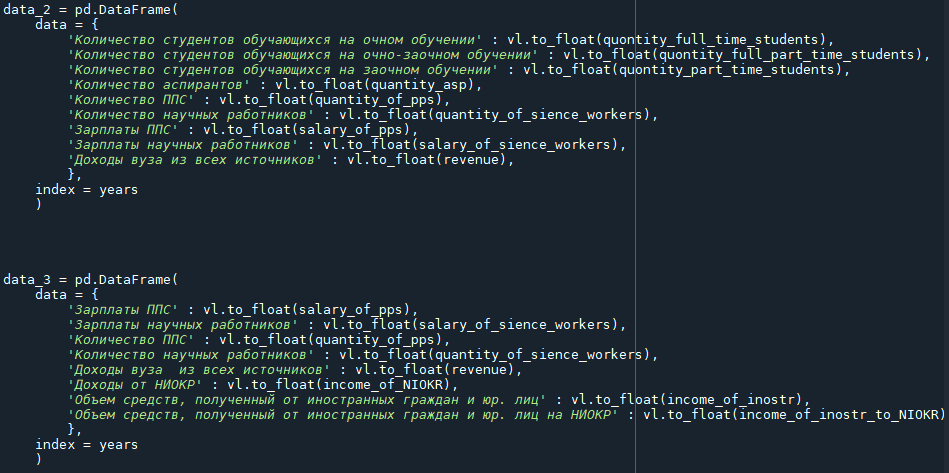


Рисунок №51. Создание таблиц.

С помощью метода класса строится корреляционная матрица, а с помощью функции строится тепловая карта корреляционной матрицы (Рисунок №52. Построение корреляционной матрицы и создание тепловых карт).

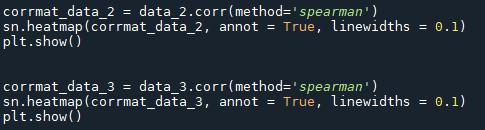


Рисунок №52. Построение корреляционной матрицы и создание тепловых карт.

## Корреляционный анализ

Как показывает тепловая карта (Рисунок №53. Первая тепловая карта), предположение, что сокращение числа преподавательского состава связано с сокращением количества студентов по всем формам обучения оказалось верным: коэффициенты корреляции оказались выше, чем 0.9, а следовательно сила корреляционной связи весьма высокая. Также можно сделать вывод, что сокращение числа профессорско-преподавательского состава не связано с их зарплатами, что тоже логично, так как в периоде с 2015 по 2021 год наблюдался ежегодный рост зарплат, причём достаточно существенный. Также видна заметная корреляция между числом ППС и доходами университета. Корреляция между числом научных сотрудников и доходами университета является высокой, а корреляция между числом научных работников и числом студентов с очно-заочной и заочной форм обучения умеренная.

Chart

Description automatically generated

Рисунок №53. Первая тепловая карта.

Вторая тепловая карта матрицы корреляции показывает, что рост зарплат ППС и научных работников имеют высокую корреляцию с общим объёмом НИОКР и умеренную корреляционную связь с объёмом средств, полученных от иностранных граждан и иностранных юридических лиц. Также видно, что увеличение зарплат никак не связано с количеством профессорско-преподавательского состава и доходами вуза из всех источников (Рисунок №54. Вторая тепловая карта).

Chart, treemap chart

Description automatically generated

Рисунок №54. Вторая тепловая карта.

Как показывает тепловая карта (Рисунок №55. Третья тепловая карта), умеренную корреляционную связь с количеством научных публикаций имеют с зарплатами профессорско-преподавательского состава, средствами, выделяемыми на НИОКР, общим объем научно-исследовательских и опытно-конструкторских работ

Chart

Description automatically generated

Рисунок №55. Третья тепловая карта.

# Заключение

В ходе научно-исследовательской практики были закреплены знания о сборе, обработке, систематизации и визуализации информации, а также библиотек языка программирования Python, также разработаны функции позволяющие осуществлять сбор данных с таблиц, размещённых на интернет-ресурсах и дальнейшем представлении данных в таблицах. Данная практика является хорошим практическим опытом для дальнейшей деятельности.

# Список использованных источников

1. Райан Митчелл. Скрапинг веб-сайтов с помощью Python. – М.: ДМК Пресс, 2016. – 280 с.
2. Уэс Маккинли Python и анализ данных / Пер. с англ. Слинкин А. А. – М.: ДМК Пресс, 2015. – 482 с.: ил.
3. <https://www.crummy.com/software/BeautifulSoup/bs4/doc.ru/bs4ru.html#id53> – BeautifullSoup 4.9.0 документация.