**ЛАБОРАТОРНАЯ РАБОТА 6**

**Работа с документами Excel с помощью Python**

Задание 1

- из таблицы лабораторной работы №3 скачать данные с листа 3 в датафрейм с помощью библиотеки Python - pandas

- найти средний балл студентов и вывести на экран

- найти количество студентов, которые не сдали экзамен

- построить гистограмму, отображающую обучение студентов по информатике

Задание 2

- первый лист таблицы из лабораторной работы №3 перевести формулы в значения (копировать – вставить – значения)

- из таблицы лабораторной работы №3 скачать данные с листа 1 в датафрейм с помощью библиотеки Python - pandas

- найти количество студентов, у которых третья степень обучения

- вывести строки, где средний балл студентов >4

Задание 3

- В датафрейм скачать столбцы из итоговой таблицы до черчения

- Добавить в датафрейм столбцы с итоговыми данными и средней оценкой

- построить диаграмму по столбцу со средней оценкой

**Краткая теория**

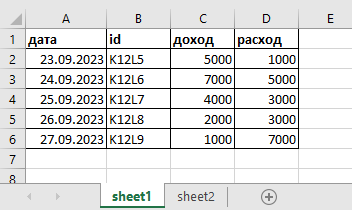
Одним из наиболее удобных инструментов для обработки больших данных является библиотека Pandas. Это мощный инструмент, разработанный для обработки и анализа данных в языке программирования Python. Она предоставляет структуры данных и функции, которые облегчают импорт, манипуляцию и анализ данных, что делает ее популярным выбором среди исследователей, аналитиков данных и научных работников.

Основными структурами данных в Pandas являются DataFrame и Series. DataFrame - это таблица, представляющая собой двумерную структуру данных с метками строк и столбцов, подобную таблице в Excel. Series - это одномерная структура данных, представляющая собой столбец или строку DataFrame. Pandas позволяет импортировать данные из различных источников, таких как файлы Excel, CSV, SQL-базы данных, а также экспортировать данные в различные форматы, что упрощает работу с данными из разных источников. При этом, Pandas интегрируется с библиотекой Matplotlib, что позволяет строить графики и визуализации данных непосредственно в Pandas. Кроме того, данные обработанные в Pandas можно экспортировать в Excel, и таким образом строить графики и диаграммы более привычным способом.

Так как Лист Excel — это двухмерная таблица и объект DataFrame от Pandas также представляет собой двухмерную табличную структуру данных, то их взаимное использование может быть удобным и эффективным. Для этого используется функция read\_excel() модуля Pandas. Рассмотрим примеры загрузки данных из Excel-файла.

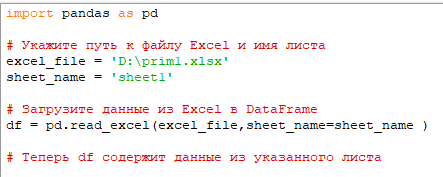
1. Загрузка данных из одно листа excel

На рис.1 приведена таблица доходов и расходов с указанием даты и идентификатора.

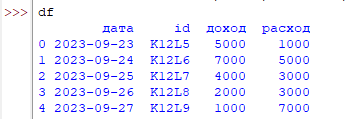


**Рис1. Таблица доходов/расходов**

Для загрузки данных с одного листа Excel используется библиотека Pandas в Python. Процесс начинается с указания пути к файлу Excel и имени листа (рис. 2). После этого, с помощью метода read\_excel, библиотека Pandas читает данные с указанного листа и сохраняет их в объекте DataFrame. Таким образом, данные из листа Excel становятся доступными для дальнейшей обработки и анализа в Python. Результат работы такого кода приведен на рис. 3

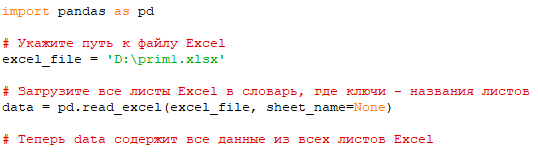


**Рис2. Импорт данных с листа Excel в DataFrame**



**Рис3. Результат работы кода**

1. Загрузка всех листов файла

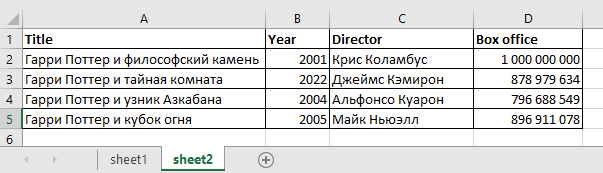


**Рис4. Импорт данных из всего файла Excel**

В данном примере (Рис. 4) мы загружаем данные из нескольких листов Excel. Сначала указываем путь к файлу Excel, а затем, с использованием метода read\_excel и параметра sheet\_name=None, загружаем все листы Excel в словарь, где ключами являются названия листов, а значениями - данные с соответствующих листов.

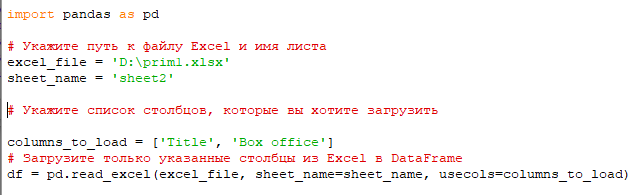
1. Загрузка определенных столбцов файла

На рисунке 5 приведена таблица кассовых сборов фильмов.



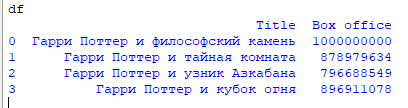
**Рис5. Кассовые сборы фильмов**

Для получения столбцов с наименованием и сборами можно использовать код, приведенный на рис. 6.



**Рис6. Загрузка определенных столбцов**

В примере 3 мы загружаем данные из конкретных столбцов Excel-файла. Сначала мы указываем путь к файлу Excel и имя целевого листа. Затем с помощью параметра usecols, мы указываем список столбцов, которые мы хотим загрузить в DataFrame. Результат представлен на рисунке 7.



**Рис6. Результат работы кода**

**Работа с библиотекой Pandas изучается самостоятельно:**

[**https://blog.skillfactory.ru/kak-nachat-analizirovat-dannye-v-pandas-pervye-shagi/**](https://blog.skillfactory.ru/kak-nachat-analizirovat-dannye-v-pandas-pervye-shagi/)