|  |
| --- |
| МИНОБРНАУКИ РОССИИ |
| Федеральное государственное бюджетное образовательное учреждение высшего образования  **«МИРЭА**  **Российский технологический университет»**  **РТУ МИРЭА** |

**Институт информационных технологий (ИИТ) Кафедра прикладной математики (ПМ)**

## ОТЧЕТ ПО ПРАКТИЧЕСКОЙ РАБОТЕ №1

по дисциплине «Языки программирования для статистической обработки данных»

|  |  |
| --- | --- |
| Студент группы | *ИМБО-11-23. Журавлев Ф.А.* |
|  | (подпись) |
| Преподаватель | *Трушин СМ* |
|  | (подпись) |

Москва 2025 г.

# Содержание

1. [Введение 3](#_bookmark0)

[Цель практики: 3](#_bookmark1)

[Задачи практики: 3](#_bookmark2)

[Установка и настройка окружения R, Python, Glarus BI 5](#_bookmark3)

1. [Результаты работы 7](#_bookmark4)

[Работа в Glarus BI: 7](#_bookmark5)

[Посмотрим на реализацию этого алгоритма на языке Python. 9](#_bookmark6)

[Реализация в R: 10](#_bookmark7)

1. [Выводы: 13](#_bookmark8)

# Введение

## Цель практики:

Научиться настраивать рабочее окружение для работы с Python, R и Glarus BI, а также выполнять базовые операции с данными в каждом из инстру- ментов.

## Задачи практики:

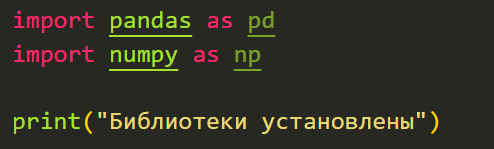
1. Установить и настроить окружение: Python (VSCode), R (VSCode), подключение к Glarus BI.
2. Изучить интерфейс Glarus BI, основные элементы и инструменты для анализа данных.
3. Загрузить предоставленный набор данных в Glarus BI.
4. Выполнить базовые операции: Отфильтровать данные по определённым условиям. Рассчитать среднее значение для одного из столбцов.
5. Провести аналогичные операции в Python и R: загрузка данных, подсчёт среднего значения.

# Установка и настройка окружения R, Python, Glarus BI

## Настройка Python (VSCode):

1. Откроем VScode, запустим программу, убедимся, что библиотеки установлены.

*Листинг 1.1. Команды для проверки установки библиотек*

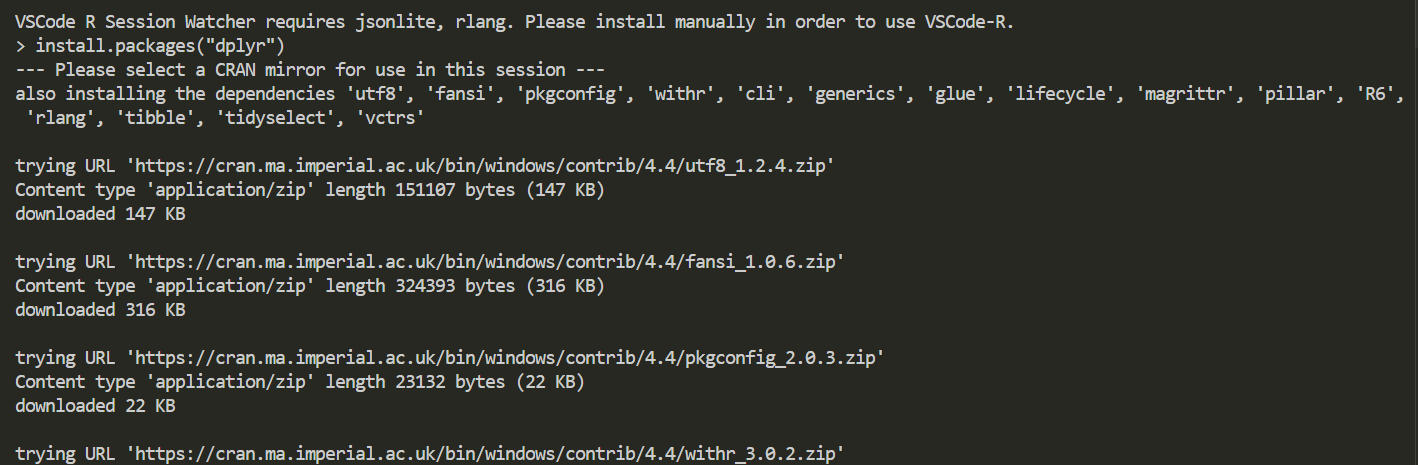
**

## Настройка R (VSCode):

Установим необходимые для анализа данных библиотеки

**dplyr** и **readr**.

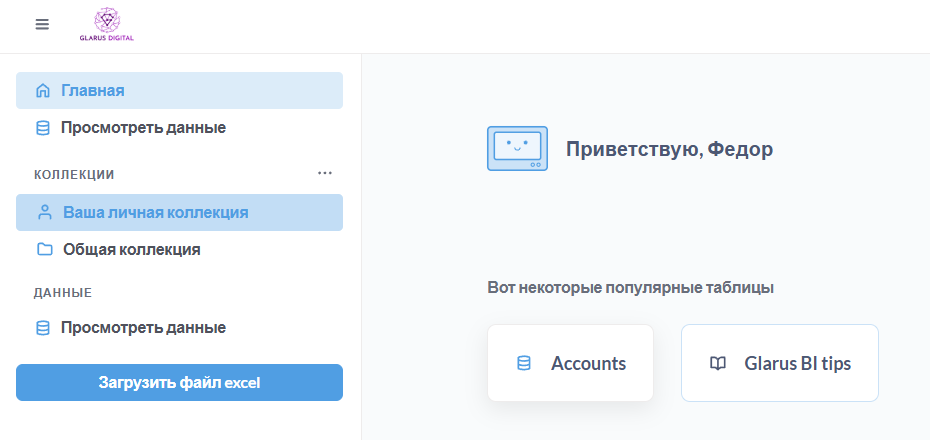
*Листининг 1.2.Установка библиотек dplyr и readr*

**

## Настройка Glarus BI:

1. Откроем веб-интерфейс Glarus BI.
2. Войдем в систему с учётными данными, предоставленными администратором.
3. Проверем доступ к функциям загрузки данных и создания отчётов.

*Листининг 1.3.Авторизация в Glarus BI*

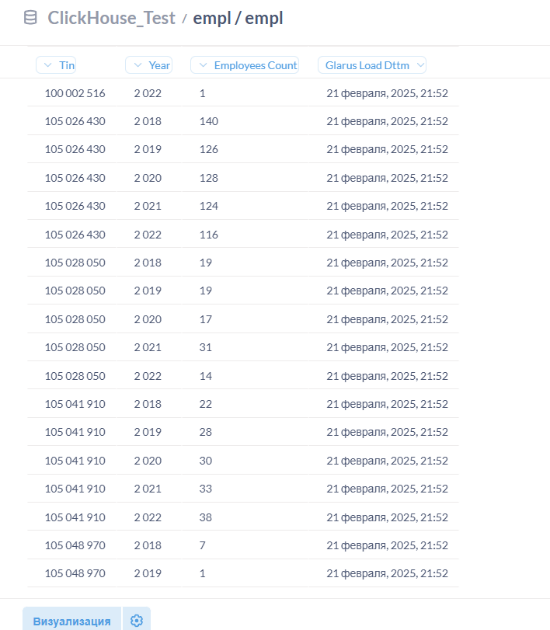
**

# Результаты работы

## Работа в Glarus BI:

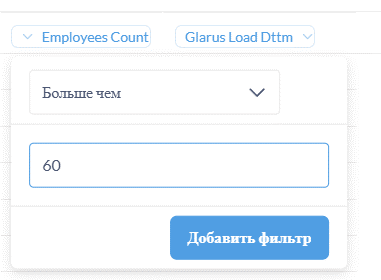
Загрузим нашу таблицу в формале .xls для работы с ней в Glarus BI:

*Листининг 2.1.1.Загрузка набора данных на тему «Количество сотрудников в предприятиях»*



## Фильтрация данных:

Выполним фильтрацию данных на примере количества сотрудников в предприятии. Выделим интересующий нас столбец «Employees Count», затем выберем «больше или равно» , поставим в поле значение «60» и нажмем «добавить фильтр».

*Листининг 2.1.2.Фильтрация данных*

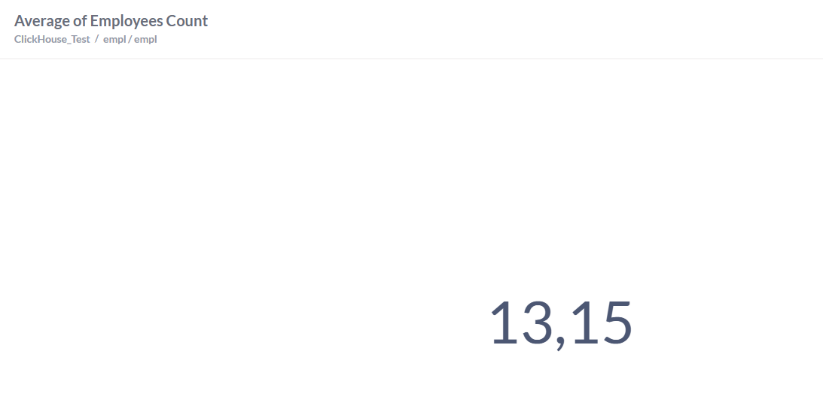
*Листининг 2.1.2.Результат фильтрации.*

**

## Подсчет медианы:

Посчитаем средний балл по математике воспользовавшись специальной функцией Glarus BI:

*Листининг 2.1.3.Среднее количество сотрудников*

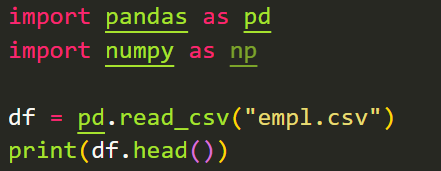
**

# Скрипты Python и R для загрузки данных и подсчёта среднего значения.

## Посмотрим на реализацию этого алгоритма на языке Python.

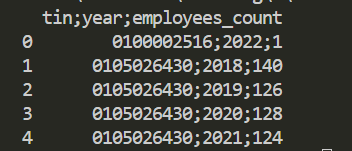
Загрузим в Google Colab наш .csv файл прописав специальный код.

Листининг 2.2.1.Среднее количество сотрудников



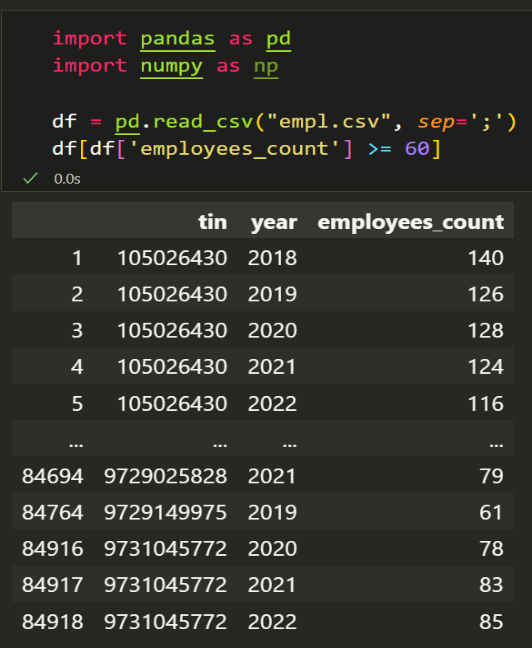
Посмотрим на таблицу, которую нам вывело:

Листининг 2.2.2.Таблица в Python

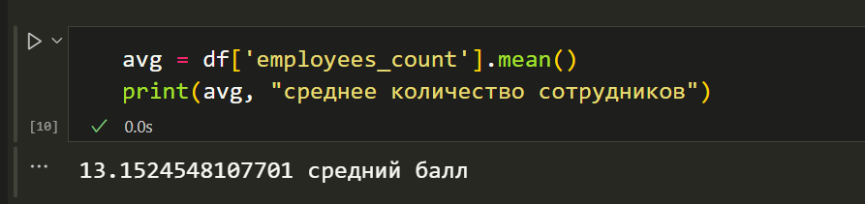


Напишем код, который поможет нам отсортировать количество сотрудников на предприятии, которые больше или равно 60.

*Листининг 2.2.3.Фильтрация в Python*

**

*листининг 2.2.4.Среднее количество сотрудников*

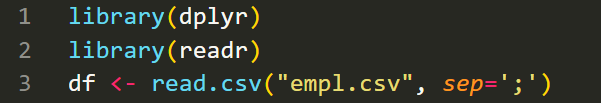
**

Как видно, средний балл такой же, какой и получился в **GLARUS BI**

## Реализация на R:

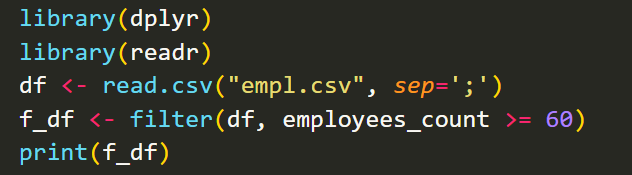
Пропишем код на R, благодаря которому сможем загрузить наш файл .csv.

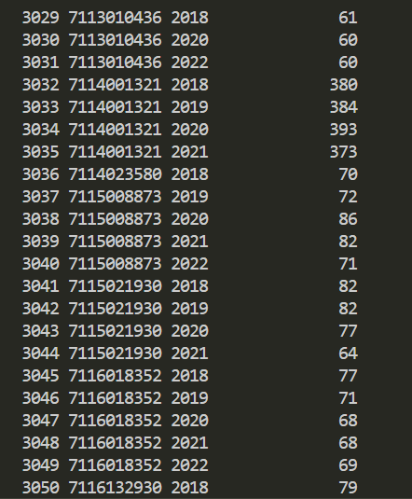
*Листининг 2.2.5.Загрузка в R*

**

После загрузки нашего датасета, напишем код, который отсортирует нашу таблицу по количеству сотрудников, которых больше или равно 60.

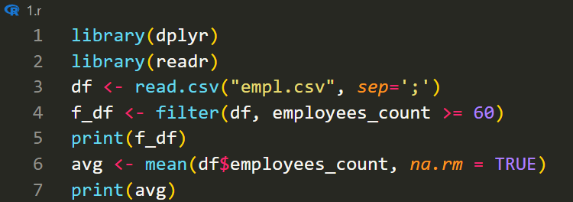
*Листининг 2.2.6.Фильтрация данных.*

**

*Листининг 2.2.7.Отфильтрованная таблица в R*

И напишем код, благодаря которому сможем оценить медиану:

*Листининг 2.2.8.Медиана в R*

**

**

По итогам все значения совпали на всех трех платформах.

# Выводы

В результате первой практической работы был проведен анализ набора данных «Количество сотрудников на предприятиях» на трех разных платформах: R, Python, Glarus BI. Каждый из них имеет свои достоинства. Код на R более понятный и легкий, нежели на питоне, но в Glarus BI ничего не надо писать и для такой простой обработки он самый быстрый и эффективный вариант. Python в свою очередь имеет самый привычный для обывателя код, который будет понятен всем, кто чуть-чуть с ним знаком.