|  |
| --- |
|  |
| 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 №2**

по дисциплине «Языки программирования для статистической обработки данных»

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

Москва 2025 г.

# ЦЕЛЬ И ЗАДАЧИ

## Цель практической работы:

Изучить методы загрузки и очистки данных в Python, R и Glarus BI, а также освоить базовые инструменты для подготовки данных к анализу.

## Задачи практической работы:

1. Загрузить данные из различных источников (CSV, Excel, базы данных) в Python, R
2. Выполнить очистку данных:

* Обнаружить и обработать пропуски (удаление, замена на среднее/медиану).
* Обнаружить и удалить дубликаты.
* Преобразовать типы данных (например, текст в даты).

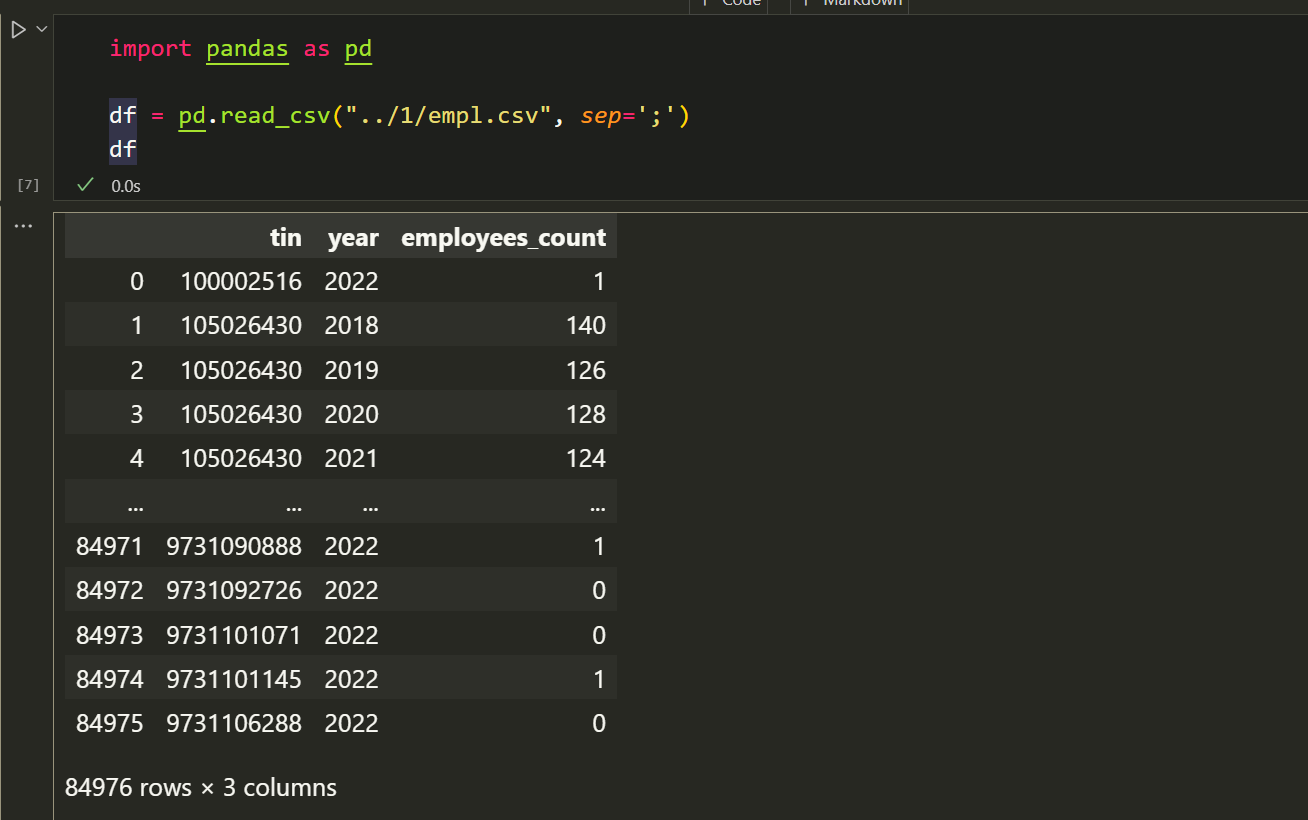
1. Сравнить подходы к очистке данных в Python, R.

# РЕЗУЛЬТАТЫ ПРАКТИКИ

## Шаг 1) Загрузка данных из различных источников

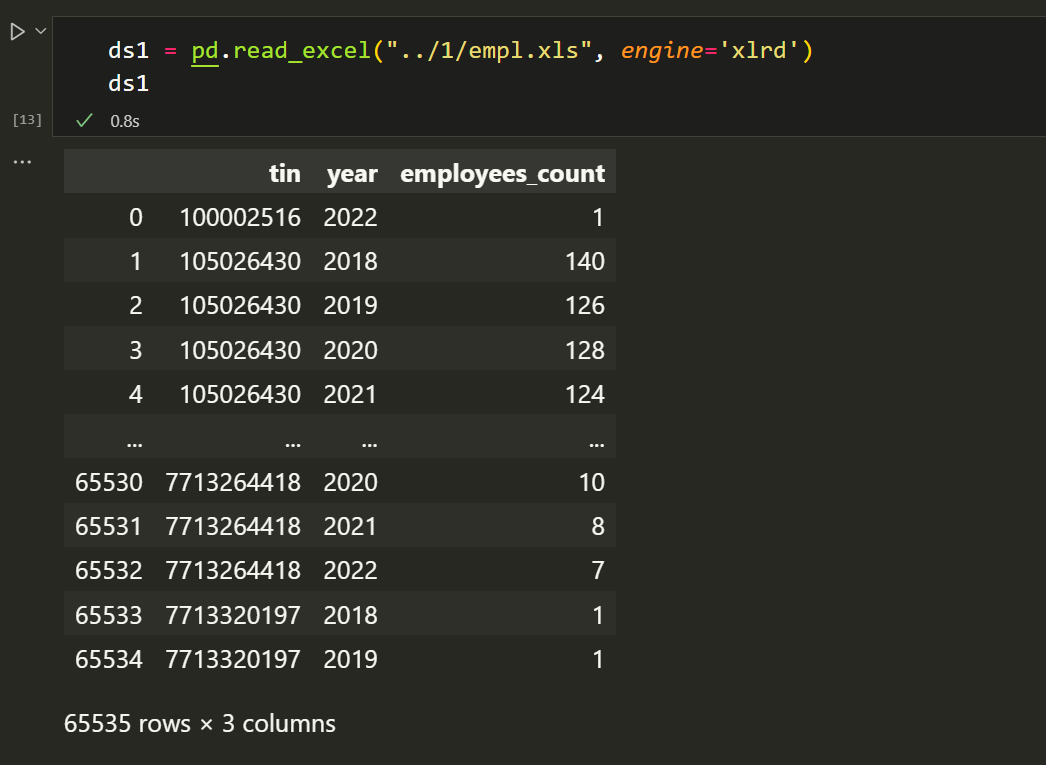
### Загрузка данных в Python (VS Code):

*Рисунок 1.1 — загрузка данных в Python*

**

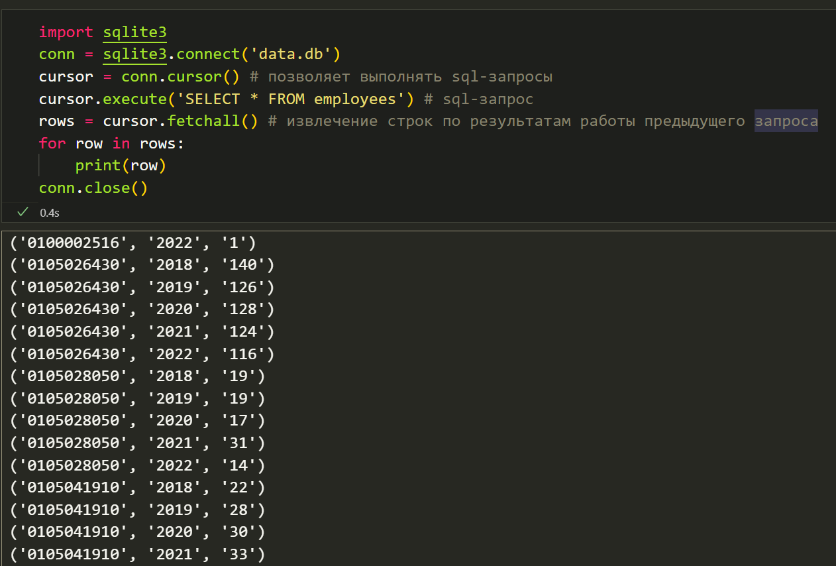
Загрузили исходную таблицу данных в формате .csv.

*Рисунок 1.2 — загрузка данных в Python*

**

Загрузка исходного набора данных в формате .xls

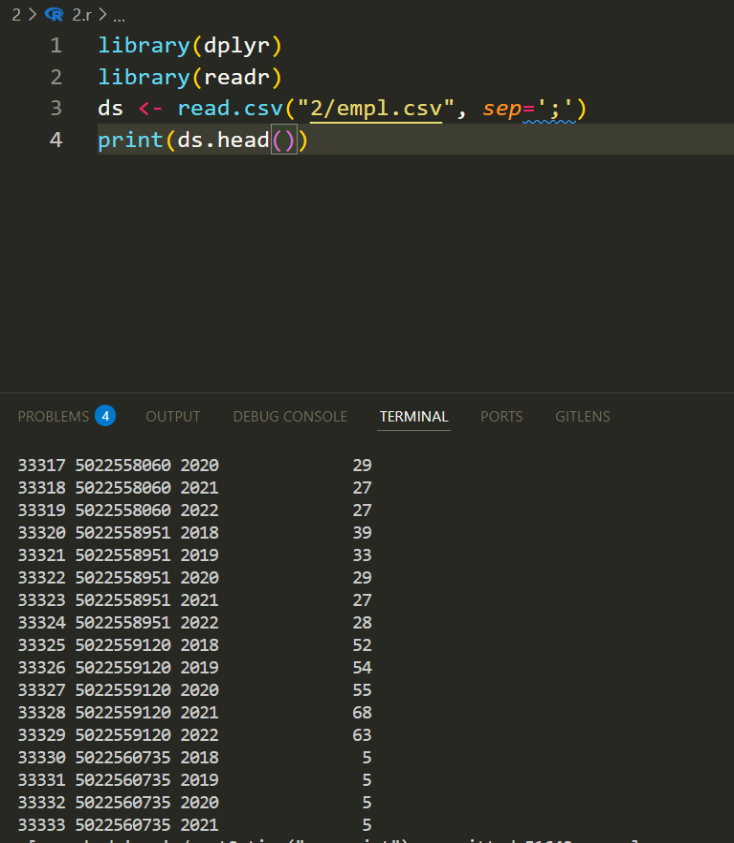
*Рисунок 1.3 — загрузка данных в Python*

**

Загрузка исходного набора данных в формате SQLite

## Загрузка данных в R (VS Code)

*рисунок 1.4 — код для загрузки .csv файла в RStuido*

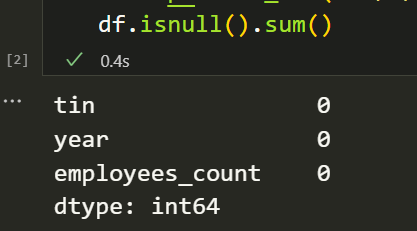
**

## Шаг 2) Очистка данных

## Очистка данных в Python:

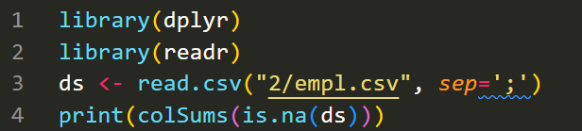
Найдем количество пустых строк в нашем наборе данных:

*Рисунок 2.1.1 – количество пустых строк*

**

## Отчистка данных в R:

*Рисунок 2.2.1 – находимым пустые строчки*

**

*Рисунок 2.2.2 — Набор данных с пустыми строками.*

**

# ВЫВОДЫ

Результате практической работы была произведена очистка данных, заменены пустые строчки, удалены дубликаты и т. д. В Python чуть-чуть удобнее работать с этим, так как код более тривиальный, нежели в R, но у R есть огромный плюс — очень удобная работа с таблицами и установкой нужных пактов прямиком из директории программы. Glarus BI неудобен из-за отсутствия нормального функционала, лоу-код щит.