

ПРАКТИЧЕСКАЯ РАБОТА 1: ВВЕДЕНИЕ В PYTHON, R И GLARUS BI ДЛЯ СТАТИСТИКИ

Ожидаемое время выполнения: 2 часа (1 час — настройка и обзор, 1 час — выполнение задач).

Цель практики

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

Задачи практики

1. Установить и настроить окружение: Python (Jupyter Notebook/Google Collab), R (RStudio), 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.

Формат отчета

1. Введение:
 - 1.1. Цель и задачи работы.
 - 1.2. Описание выполненных действий по настройке окружения.
2. Результаты работы:
 - 2.1. Скриншоты из Glarus BI:
 - 2.1.1. Загрузка данных.
 - 2.1.2. Пример выполнения фильтрации данных.
 - 2.1.3. Подсчёт среднего значения.
 - 2.2. Скрипты Python и R для загрузки данных и подсчёта среднего значения.
3. Заключение:
 - 3.1. Краткий анализ: чем удобны Glarus BI, Python и R для базовых операций.

Дополнительные материалы к практической работе 1

Исходные данные: Набор данных в формате CSV (например, данные о продажах или студенческих оценках).

1. ИНИД ИНФРАСТРУКТУРА НАУЧНО-ИССЛЕДОВАТЕЛЬСКИХ ДАННЫХ <https://data.rcsi.science/data-catalog/?page=1>

ПОШАГОВАЯ ИНСТРУКЦИЯ К ПРАКТИЧЕСКОЙ РАБОТЕ 1

Шаг 1. Установка и настройка окружения

1.1. Настройка Python (Google Colab)

1. Откройте браузер и перейдите на [Google Colab](#).
2. Войдите в систему с учётной записью Google.
3. Нажмите **File (Файл) → New Notebook (Новая тетрадь)**.
4. Убедитесь, что библиотеки pandas и numpy установлены (они установлены по умолчанию в Colab). Чтобы проверить это, выполните команду:

Листинг 1.1. Команды для проверки установки библиотек

```
import pandas as pd
import numpy as np
print("Библиотеки загружены успешно.")
```

1.2. Настройка R (RStudio)

1. Установите [RStudio](#) на свой компьютер.
2. Проверьте наличие необходимых библиотек:
 - Откройте RStudio и выполните в консоли команды для установки библиотек:

Листинг 1.2 Установка необходимых библиотек

```
install.packages("dplyr")
install.packages("readr")
```

3. Убедитесь, что библиотеки установлены:

Листинг 1.3 Проверка установки необходимых библиотек

```
library(dplyr)
library(readr)
print("Библиотеки успешно подключены.")
```

1.3. Подключение к Glarus BI

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

Шаг 2. Изучение интерфейса Glarus BI

1. Перейдите в основной раздел Glarus BI после авторизации.
2. Ознакомьтесь с основными элементами:
 - Панель навигации (основные меню и вкладки).
 - Раздел загрузки данных.
 - Возможности фильтрации и агрегации.
3. Откройте раздел справки или встроенное руководство (если доступно) для уточнения функций.

Шаг 3. Загрузка предоставленного набора данных

3.1. Загрузка данных в Glarus BI

1. Перейдите в раздел загрузки данных.
2. Выберите источник данных: xls-файл.¹

3. Загрузите предоставленный xls-файл (например, с данными о продажах).
4. Убедитесь, что данные корректно отобразились в виде таблицы.

3.2. Загрузка данных в Python (Google Colab)

1. Загрузите CSV-файл в Google Colab, используя следующую команду:

Листинг 1.4 Загрузка CSV-файла в Google Colab

```
from google.colab import files
uploaded = files.upload() # Открывает окно загрузки файла

import pandas as pd
df = pd.read_csv(list(uploaded.keys())[0]) # Чтение файла
print(df.head()) # Просмотр первых строк данных
```

3.3. Загрузка данных в R (RStudio)

1. В RStudio загрузите CSV-файл с помощью команды:

Листинг 1.5 Загрузка CSV-файла в RStudio

```
library(readr)
df <- read_csv("путь_к_файлу.csv")
head(df)
```

Шаг 4. Выполнение базовых операций

4.1. Glarus BI

1. Перейдите в раздел анализа данных.
2. Выполните фильтрацию данных:
 - Укажите условие фильтрации (например, выбрать строки, где

продажи превышают 1000).

3. Рассчитайте среднее значение одного из числовых столбцов (например, средней суммы продаж).
4. Сохраните результаты в отчёте или таблице.

4.2. Python (Google Colab)

1. Выполните фильтрацию данных в Python:

Листинг 1.6 Фильтрация данных

```
filtered_df = df[df['Sales'] > 1000] # Фильтрация строк
print(filtered_df.head())
```

2. Рассчитайте среднее значение для одного из столбцов:

Листинг 1.7 Расчет среднего значения для одного столбца

```
mean_value = df['Sales'].mean()
print("Среднее значение продаж:", mean_value)
```

4.3. R (RStudio)

1. Выполните фильтрацию данных в R:

Листинг 1.8 Фильтрация данных

```
library(dplyr)
filtered_df <- filter(df, Sales > 1000) # Фильтрация строк
print(filtered_df)
```

2. Рассчитайте среднее значение для одного из столбцов:

Листинг 1.9 Расчет среднего значения для одного из столбцов

```
mean_value <- mean(df$Sales, na.rm = TRUE)
print(paste("Среднее значение продаж:", mean_value))
```

Шаг 5. Анализ и подготовка отчёта

1. Сделайте скриншоты выполнения основных операций в Glarus BI:
 - Загрузка данных.
 - Фильтрация данных.
 - Подсчёт среднего значения.
2. Сохраните Python- и R-код, использованный для выполнения тех же задач.
3. Подготовьте отчёт в формате, указанном в задании, с разделами:
 - **Введение:** описание цели и задач.
 - **Результаты работы:** скриншоты и код.
 - **Заключение:** краткий анализ удобства работы в каждом инструменте.