**Федеральное государственное автономное образовательное учреждение  
высшего образования  
"Национальный исследовательский университет  
"Высшая школа экономики"**

Московский институт электроники и математики им. А.Н. Тихонова

**Курс: Python в науке о данных**

Руководство пользователя к приложению по базе данных, предназначенной для оценки вулканической активности в мире

Разработчики:

Студент Подкопаева Полина БИВ191

Студент Баканов Глеб БИВ191

Студент Ковязин Владислав БИВ191

Руководитель: Полякова Марина Васильевна

**Москва 2020**

***Описание решаемой задачи:***

Предлагаемое информационно-аналитическое приложение предназначено для редактирования и анализа базы данных, содержащую информацию о различных параметрах, связанных с вулканической активностью.

Для управления и анализа предусмотрен следующий функционал:

1. Загрузка и сохранение базы данных
2. Редактирование отдельных данных базы
3. Поиск данных в базе по фильтрам и ключевым словам, фильтр по столбцам
4. Статистический анализ отфильтрованных данных, представление результатов в виде графиков, диаграмм и карт с возможностью экспорта (сохранения в пользовательский файл)

***Требования к характеристикам компьютера и ОС:***

1. Оперативная память – не менее 2 Гб
2. Свободная память на диске – не менее 200 Мб
3. Операционная система – Windows 7 и выше
4. Интерпретатор языка Python – Python 3.7 и выше

***Инструкция по установке приложения:***

Пользователю требуется установить на своем компьютере дистрибутив Python 3 (Anaconda или аналогичный), который содержит интерпретатор Python и все необходимые библиотеки, которые используются данным приложением: pandas, matplotlib, numpy, tkinter, pickle, sys, os (библиотеки обычно установлены по умолчанию). Также необходимо установить дополнительную библиотеку для работы с картами folium. Для этого в консоли Spyder (среда, входит в дистрибутив) необходимо набрать:

#Python 3.x  
sudo apt install python3-pip  
pip install folium

***Инструкция по запуску и настройке приложения:***

Для работы программы на компьютере должен быть установлен дистрибутив Python 3. Для её начала необходимо:

* Запустить скрипт main.py, находящийся в каталоге Work/Scripts. Можно использовать командную строку, терминал или любую интегрированную среду разработки.
* В конфигурационном файле constants.ini пользователь может настроить приложение, поправив некоторые параметры либо оставив их по умолчанию.

***Инструкция по использованию функционала приложения:***

1. ***Описание интерфейса программы***

При запуске программы открывается окно с графическим интерфейсом, в окне слева находятся последние открытые файлы с базами данных. Чтобы открыть нужный, необходимо нажать на него двойным кликом правой кнопкой мыши. Если последних открытых файлов нет, то загружается пустая среда (рис. 1).

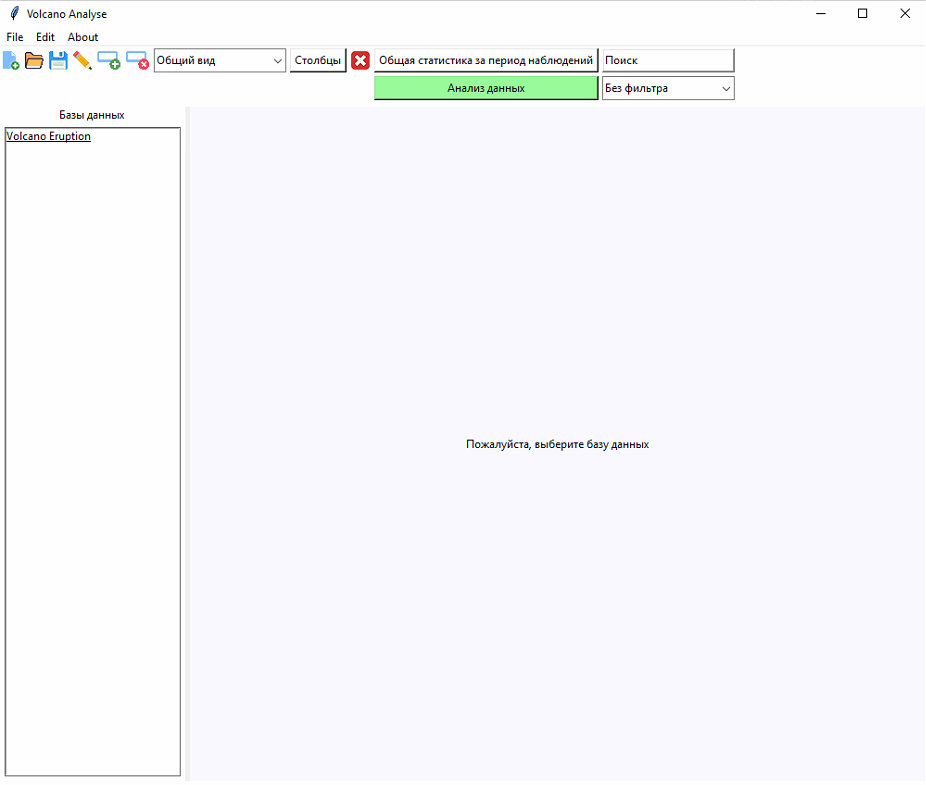


Рис. 1

1. ***Главный интерфейс – Структура***

Главный интерфейс состоит из рабочего поля, меню баз данных и меню функционала (рис.2)



Рис. 2

1. Для того, чтобы «Сохранить как» базу данных, с которой работает сейчас пользователь, необходимо нажать на голубой значок с листом - C:\Users\User\Work\Graphics\add_icon.gif. Программа откроет диалоговое окно, в котором пользователю предоставляется возможность выбрать путь для сохранения таблицы и имя для неё (по умолчанию открывается папка Data). Таблица сохраняется с расширением csv. Аналогично получится, если нажать в левом верхнем углу File -> Load (рис. 3).

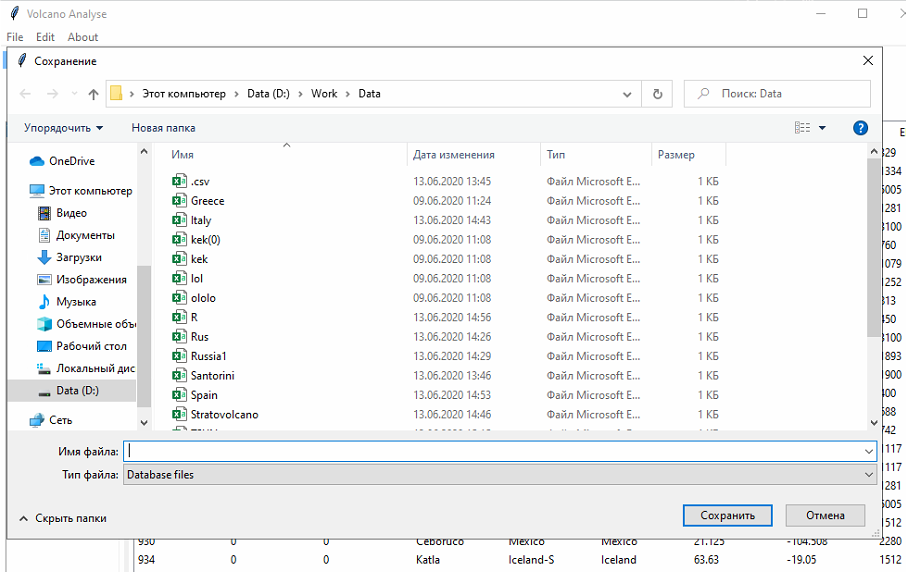


Рис. 3

1. Значок C:\Users\User\Work\Graphics\load_icon.gif предназначен для того, чтобы пользователь мог открыть необходимую ему для работы базу данных из каталога (рис. 4).

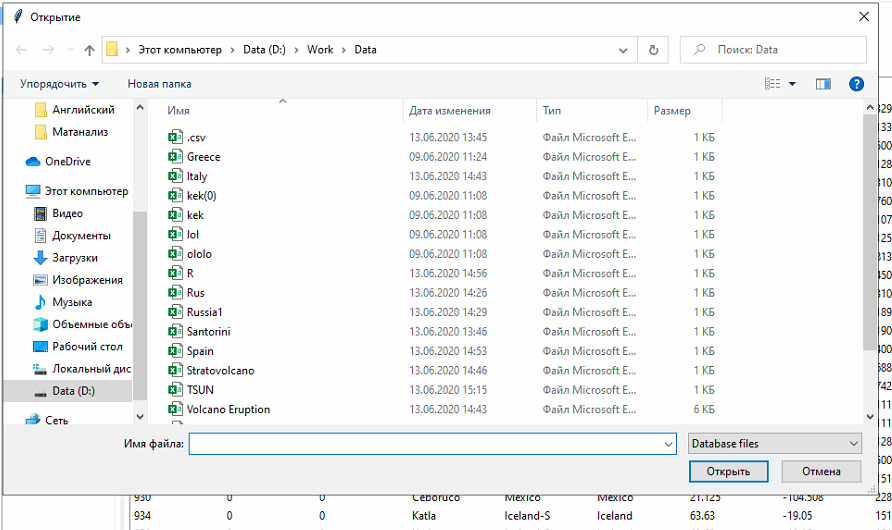


Рис. 4

1. Чтобы сохранить изменения в текущей базе, необходимо нажать кнопку «Сохранить» - C:\Users\User\Work\Graphics\save_icon.gif.
2. Для того, чтобы отредактировать данные в строке необходимо выбрать нужную строку путём нажатия одним кликом мыши (выбранная строка будет выделена синим цветом), затем нажать на кнопку «Редактировать» - C:\Users\User\Work\Graphics\edit_icon.gif. При нажатии откроется окно редактирования, аналогичное окну добавления (рис. 5-8), в зависимости от того, как выбран вид таблицы (см. пункт 5). Здесь пользователь может изменить значения в ячейках.

Редактирование выбранной строки также возможно с помощью двойного клика мыши по шапке любого столбца таблицы.

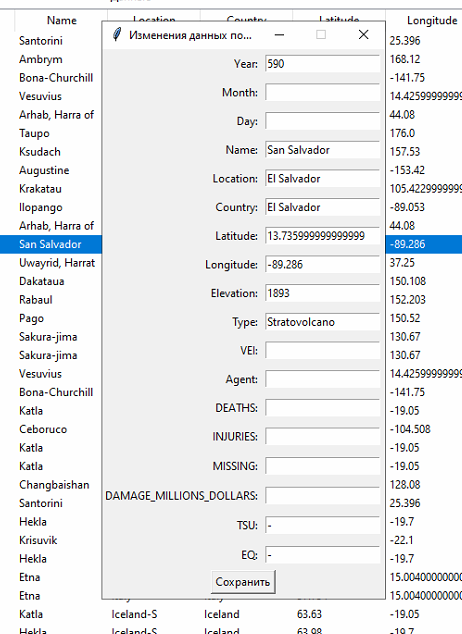


Рис. 5, Редактирование данных, общий вид

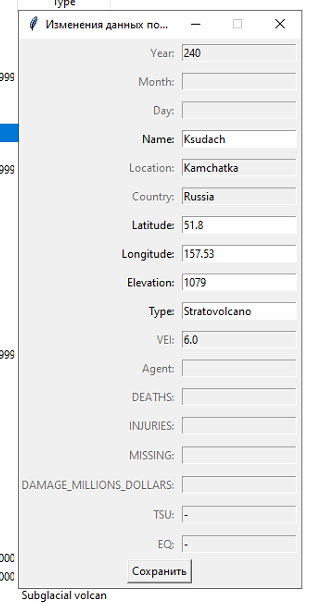


Рис. 6, Редактирование данных, Вид первый

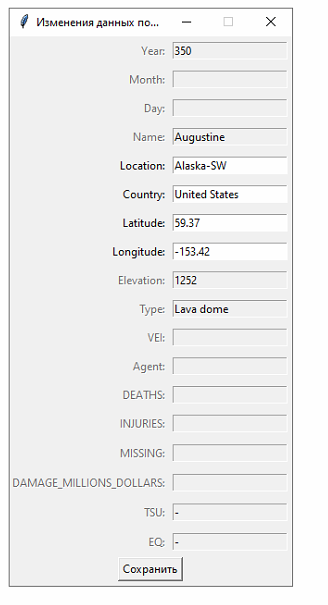


Рис. 7, Редактирование данных, Вид второй

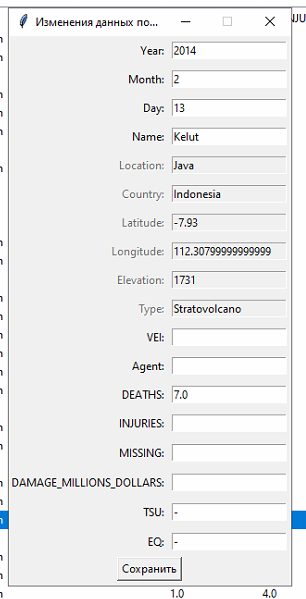


Рис. 8, Редактирование данных, Вид третий

1. В приложении существует 4 режима вида таблиц: Общий вид (формируется соединением трёх видов), Вид первый, Вид второй, Вид третий. Последние три из них представляют собой базу данных в третьей нормальной форме. Центральной является таблица «Вид первый» ключом которой являются «День-Месяц-Год». Две второстепенные таблицы «Вид второй» и «Вид третий» имеют ключи «Название» и «Долгота-Широта» соответственно. Ниже иллюстрирована структура базы данных (Рис. 9)

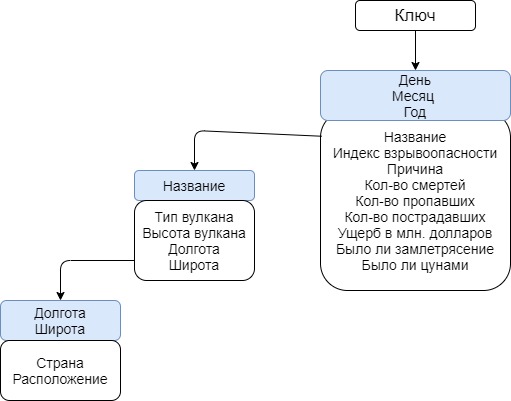


Рис. 9, Структура базы данных

Пользователь имеет возможность выбрать, в каком виде он хочет работать. Для этого в приложении существует OptionMenu, при изменении значения которого автоматически открывается нужный вид (рис. 10).

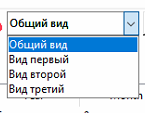


Рис. 10, OptionMenu

1. Для того, чтобы добавить новые данные, необходимо нажать значок «Добавить» - C:\Users\User\Work\Graphics\add_field_icon.gif. Пользователю откроется окно для заполнения, соответствующее выбранному виду таблицы (рис. 11-14).

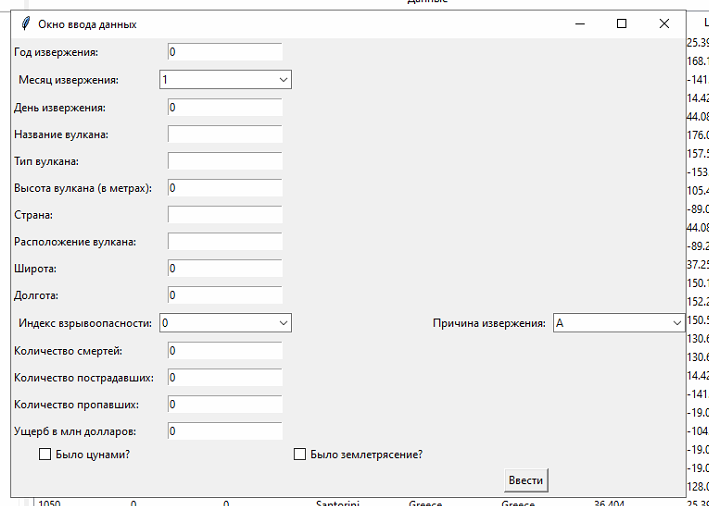


Рис. 11, Окно ввода данных, Общий вид

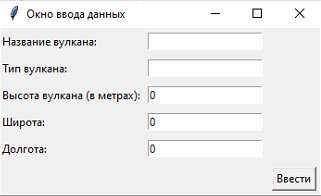


Рис. 12, Окно ввода данных, Вид первый

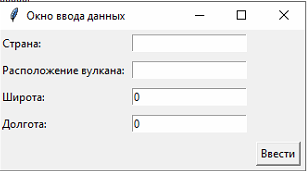


Рис. 13, Окно ввода данных, Вид второй

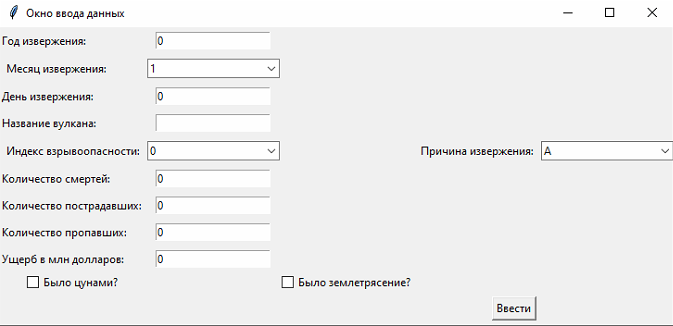


Рис. 14, Окно ввода данных, Вид третий

1. Для того, чтобы удалить данные из таблицы, пользователю достаточно отметить нужное поле и нажать кнопку «Удалить» - C:\Users\User\Work\Graphics\del_field_icon.gif. Замечание: удаление элементов из второстепенных таблиц приведёт к удалению остальных данных, которые их используют. Перед удалением приложение задаст вопрос об окончательном решении с предупреждением (рис.15).

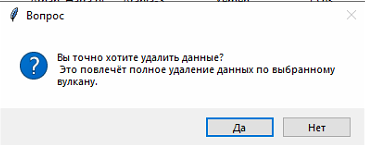


Рис. 15, Окно с предупреждением

1. Для того, чтобы закрыть выбранную базу данных, необходимо нажать значок «Закрыть» - C:\Users\User\Work\Graphics\close_icon.gif.
2. Во всех таблицах предусмотрена удобная сортировка полей по возрастанию. Пользователь может нажать на интересующий его столбец одним кликом правой кнопкой мыши, и программа отсортирует поля по убыванию/возрастанию. В виду большого количества параметров, в таблице присутствует вертикальный и горизонтальный скролл бар, для удобной навигации по базе данных (рис. 16).

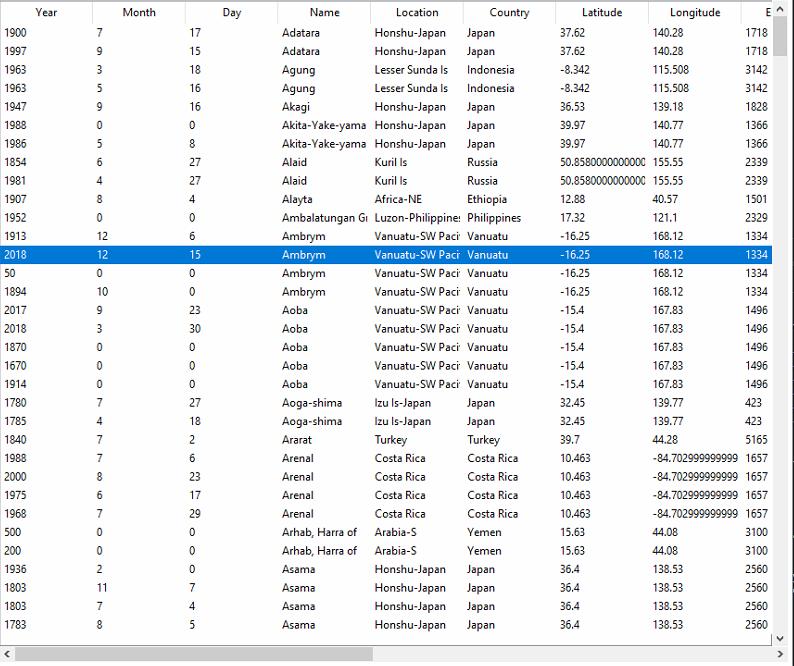
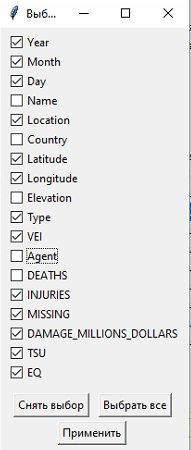


Рис. 16, База данных, отсортированная по названию вулкана

1. Пользователь имеет возможность выбрать определённые столбцы для показа в базе (отфильтровать по столбцам). Для этого необходимо нажать кнопку «Столбцы» в меню. Пользователю откроется окно, где он может очистить выбор, выбрать все. После окончания работы с данной опцией, необходимо нажать «применить» (рис. 17).

 Рис. 17, Окно выбора столбцов

1. ***Инструменты для работы с базой данных***

В программе предусмотрены следующие инструменты для работы с базой данных: анализ данных, общая статистика за период наблюдений, фильтры (рис. 18).



Рис. 18, Инструменты для работы

**Общая статистика за период наблюдений**

В данном разделе пользователь может построить общую статистику за период наблюдений выбранной базы данных.

При выборе данной опции открывается окно с выбором параметра, количественные и качественные разделены горизонтальной линией (рис. 19).

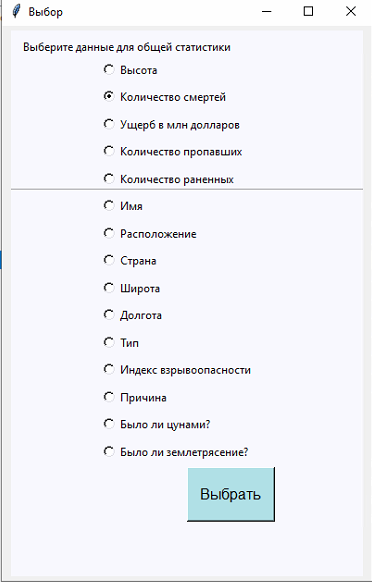


Рис. 19, Окно выбора параметра

После нажатия кнопки «Выбрать» открывается таблицы с соответствующими характеристиками для выбранного параметра.

Характеристики для количественных переменных (рис. 20):

* Среднее арифметическое
* Медиана
* Мода
* Форма плотности распределения
* Выборочная дисперсия
* Нижняя квартиль
* Верхняя квартиль
* Межквартильный размах
* Максимум
* Минимум
* Размах
* Стандартное отклонение (СКО)
* Коэффициент вариации

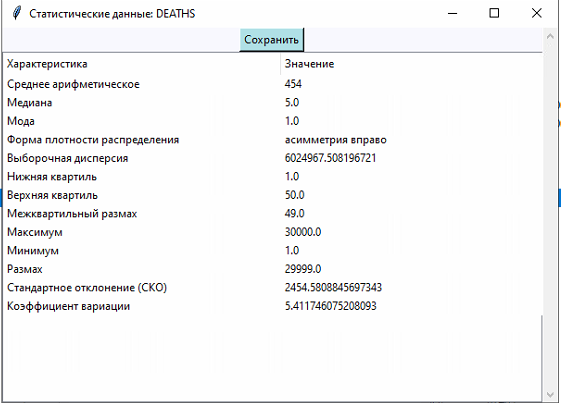


Рис. 20, Пример таблицы для количественного параметра

Характеристики для качественных переменных (рис. 21):

* Частота появления в базе данных
* Процент от общего числа

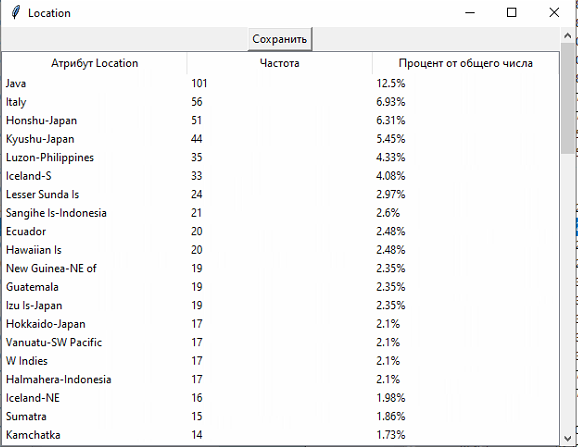


Рис. 21, Пример таблицы для качественного параметра

Построенные таблицы общей статистики пользователь может сохранить в выбранном каталоге в формате csv, нажав соответствующую кнопку вверху таблицы (автоматически изначально открывается «сохранить как» в Output).

**Фильтры**

Приложение обладает функционалом поиска данных по ключевым словам. Данная опция возможна по 5 параметрам: по названию, по типу, по стране, по цунами, по землетрясению. Для того, чтобы воспользоваться функцией, необходимо ввести ключевое слово на англ. языке (или +/- для последних двух параметров), далее выбрать по какому параметру идёт поиск (рис. 22)

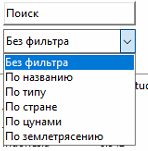


Рис. 22, Окно для поиска

После выбора программа автоматически откроет диалоговое окно, в котором пользователю предоставляется возможность выбрать путь для сохранения таблицы и имя для неё (по умолчанию открывается папка Data). Таблица сохраняется с расширением csv. После сохранения таблица открывается в приложении для работы с базами данных (рис. 23).

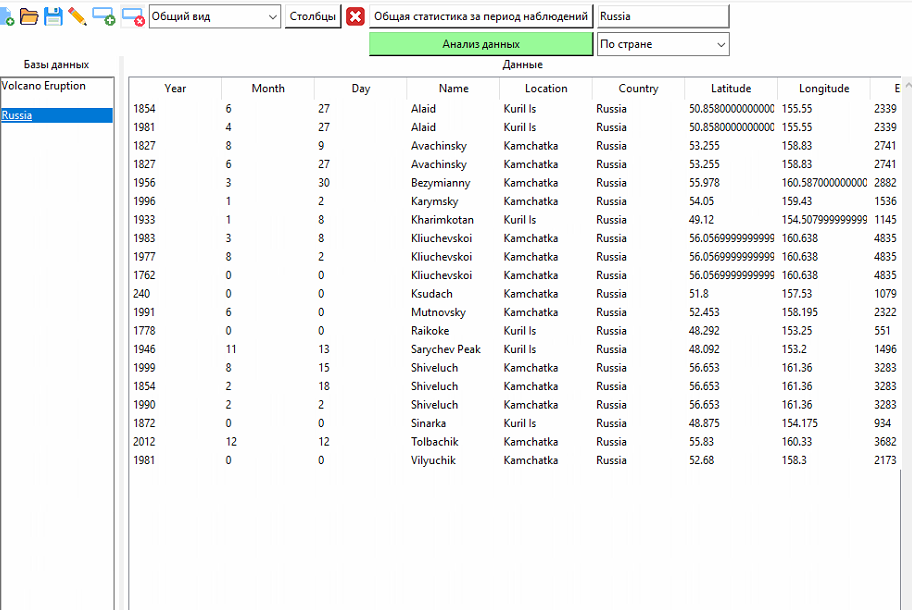


Рис. 23, Пример работы функционала, поиск по стране

**Анализ данных**

Программа позволяет делать анализ по различным параметрам. Могут быть использованы, помимо базовых статистических данных для количественных и качественных переменных, следующие методы анализа данных: «Сводная таблица», «Построение карты», «Построение графиков и диаграмм».

После нажатия опции «Анализ данных» открывается окно для выбора метода (рис. 24).

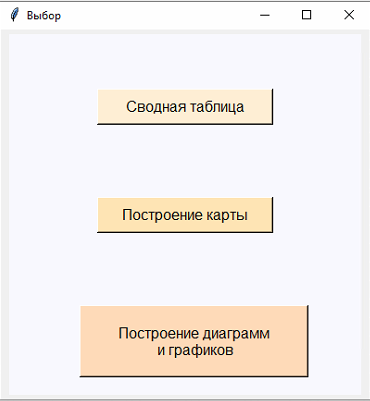


Рис. 24, Окно для выбора метода

1. Сводная таблица

В приложении откроется отдельное окно, в котором находится OptionMenu с возможными вариантами построения сводной таблицы (рис. 25):

* Страна - Средняя смертность
* Страна - Средняя высота вулкана
* Расположение - Средняя смертность
* Тип вулкана - Средняя смертность
* Тип вулкана - Количество пропавших
* Название вулкана - Количество раненных

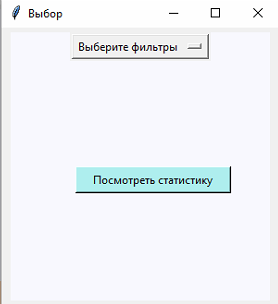


Рис. 25, Окно выбора параметра

После выбора фильтра пользователю необходимо нажать кнопку «Посмотреть статистику». В случае, если ни один параметр не выбран, приложение выдаст сообщение об ошибке. Иначе приложение откроет окно со сводной таблицей (рис. 26).

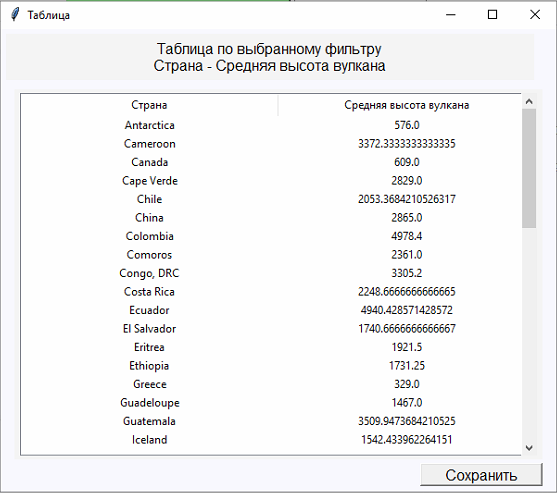


Рис. 26, Пример сводной таблицы

Пользователь может сохранить полученную таблицу, нажав на соответствующую кнопку. Программа автоматически сохранит сводную таблицу с соответствующим фильтру названием в папку Output. Таблица сохраняется в формате csv.

1. Построение карты

В приложении откроется отдельное окно, в котором находится OptionMenu с возможными вариантами построения карты (рис. 27).

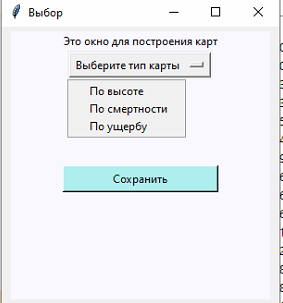


Рис. 27, Окно выбора фильтра для карты

После выбора фильтра пользователю необходимо нажать кнопку «Сохранить». В случае, если ни один параметр не выбран, приложение выдаст сообщение об ошибке. Иначе карта автоматически сохранится в папке Output в формате html (рис. 28).



Рис. 28, Пример построенной карты

1. Построение диаграмм и графиков

В приложении откроется отдельное окно, в котором находится OptionMenu с возможными вариантами для каждого вида графического отчёта. Каждому меню соответствует своя кнопка с соответствующим названием (рис. 29).

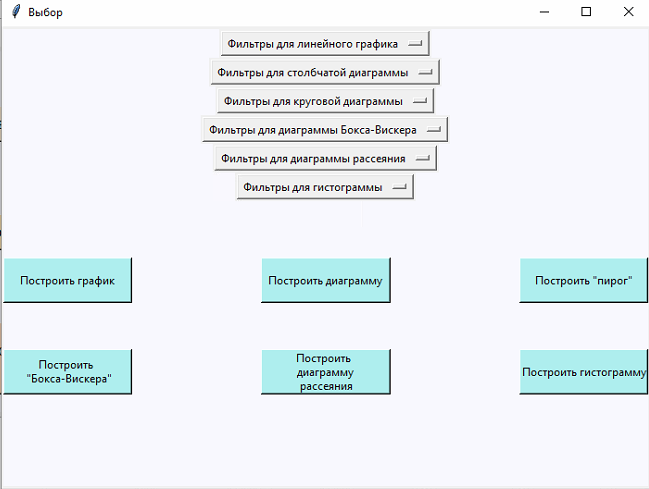


Рис. 29, Окно фильтра для графического отчёта

После выбора фильтра пользователю необходимо нажать соответствующую кнопку. В случае, если ни один параметр не выбран, приложение выдаст сообщение об ошибке. Иначе приложение откроет окно с построенной графикой (рис. 30).

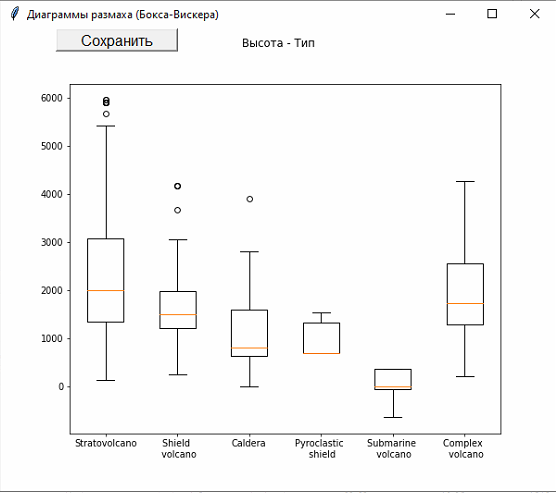


Рис. 30, Пример окна с графическим отчётом

Пользователь может сохранить полученный отчёт, нажав на соответствующую кнопку. Программа автоматически сохранит «картинку» с соответствующим фильтру названием в папку Graphics. Отчёт сохраняется в форматах jpeg и png (зависит от диаграммы/графика).

Ниже приведена таблица фильтров возможных построений для каждого вида графического отчёта, присутствующего в функционале:

|  |  |
| --- | --- |
| Вид графического отчёта | Фильтры |
| Линейный график | * Год – Средняя смертность |
| Столбчатая диаграмма | * Страна – Средняя смертность * Страна – Средняя высота вулкана * Расположение – Средняя смертность * Тип вулкана – Средняя смертность * Тип вулкана – Количество пропавших |
| Круговая диаграмма | * Страна – Средняя смертность * Тип – Средняя смертность * Тип - Ранения |
| Диаграмма размаха Бокса-Вискера | * Высота - Тип |
| Диаграмма рассеяния | * Высота – Смерти – Индекс взрывоопасности * Высота – Смерти - Тип |
| Гистограмма | * Высота – Количество смертей * Высота - Ущерб |