

FEDERAL STATE AUTONOMOUS EDUCATIONAL INSTITUTION OF HIGHER EDUCATION

“NATIONAL RESEARCH UNIVERSITY HIGHER

SCHOOL OF ECONOMICS”

MOSCOW INSTITUTE OF ELECTRONICS AND MATHEMATICS

TECHNICAL SPECIFICATION

" Python in Data Science"

Analysis of COVID-19 statistics in different countries

USER MANUAL

Student:

Mohamed Abdelhafiz Salem Abdelhafiz Youssef

Email: abmokhamed@edu.hse.ru

Phone number: +79067678234

Supervisor: Polyakov Konstantin Lvovich

Content table

1.	Install Prerequisites	4
	Python 3.10	4
	Anaconda	4
	MySQL + MySQL Workbench	4
2.	Project	4
3.	Virtual Environment	5
4.	Database.....	7
5.	Python Interpreter	8
6.	Run the Program	8
7.	User App Guide	9
	Loading & Preprocessing.....	9
	Data Filtering	11
	Statistics	11
	Visualization	12
	Configuration.....	12

List of figures

Figure 1 - Project Structure	4
Figure 2 - Anaconda Prompt	5
Figure 3 - Working Folder.....	5
Figure 4 - Virtual environment creation.....	5
Figure 5 - Environment activation	5
Figure 6 - Environment activated	5
Figure 7 - Install dependencies	5
Figure 8 - Environment selection	6
Figure 9 - Spyder installation and launch.....	6
Figure 10 - Spyder IDE	6
Figure 11 - MySQL Workbench.....	7
Figure 12 - Database creation	7
Figure 13 - MySQL configuration.....	7
Figure 14 - Python interpreter	8
Figure 15 - Environment check.....	8
Figure 16 - App	9
Figure 17 - Database Connection	9
Figure 18 - Clear Table.....	10
Figure 19 - Data loading from CSVs	10
Figure 20 - Data cleaning.....	10
Figure 21 - Data insertion.....	11
Figure 22 - Data Filtering Page	11
Figure 23 - Statistics page.....	12
Figure 24 - Visualization page	12
Figure 25 - Configuration page.....	12

1. Install Prerequisites

Python 3.10

Check if Python is already installed in CMD:

- `python --version` on Windows
- `python3 --version` on Linux/MacOs

Something like **Python 3.10.6** should appear

If not installed:

Download the latest stable version (preferably Python 3.10 at least)

Download it from: <https://www.python.org/downloads/>

Run the installer and **check the box that says "Add Python to PATH"** before clicking "Install Now"

Anaconda

Check if Anaconda is already installed in CMD:

- `conda --version`

Something like **conda 24.x.x** should appear

If not installed:

Download it from: <https://www.anaconda.com/download>

Run the installer and accept all default settings.

MySQL + MySQL Workbench

Download it from: <https://dev.mysql.com/downloads/>

Download the one complete version of the installer (300mb+)

Run the installer and select the **FULL** installation and continue with execute and install, make sure to remember the **password**, the **port** and the **user**.

2. Project

After downloading the project ZIP file, Find the downloaded ZIP file (Work.zip) in your **Downloads** folder or wherever you saved it and extract it. The structure of the folder is the following:

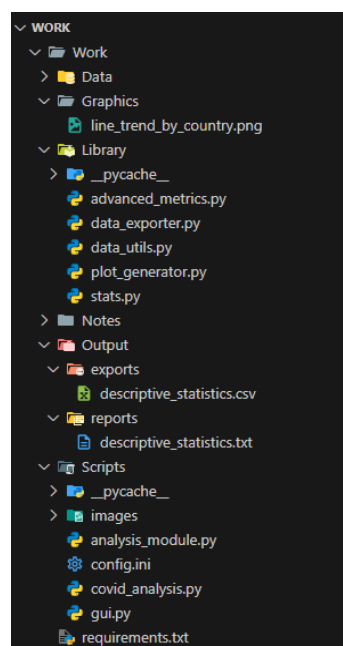


Figure 1 - Project Structure

3. Virtual Environment

After extracting the Zip file, Open **anaconda Prompt**, we will use the **Anaconda Navigator** later.

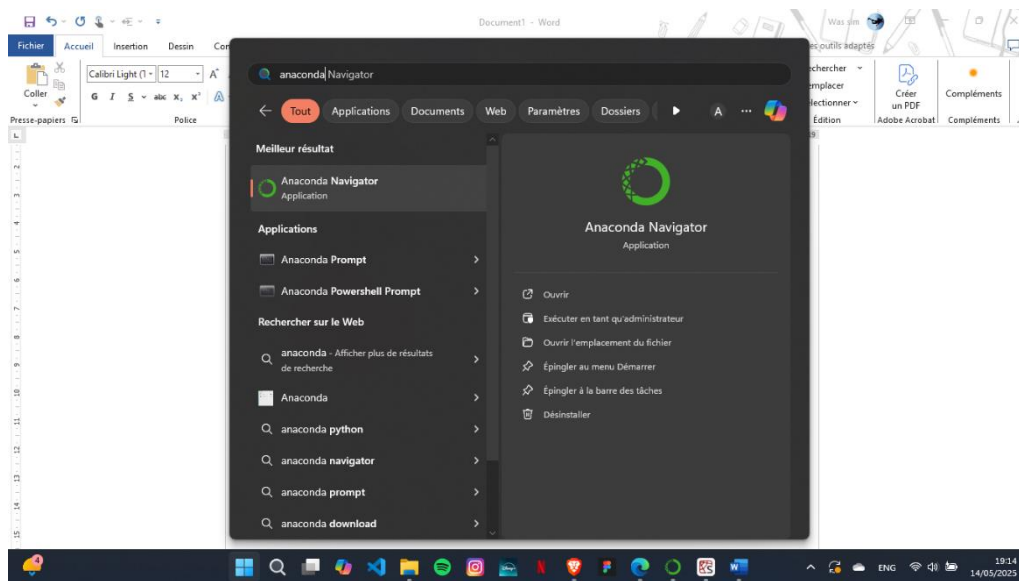


Figure 2 - Anaconda Prompt

Access the path to your extracted Work folder. If the folder was extracted in Downloads, it should be something like: **C:\Users\abahr\Downloads\Work\Work**

- Access Folder: **cd C:\Users\abahr\Downloads\Work\Work**

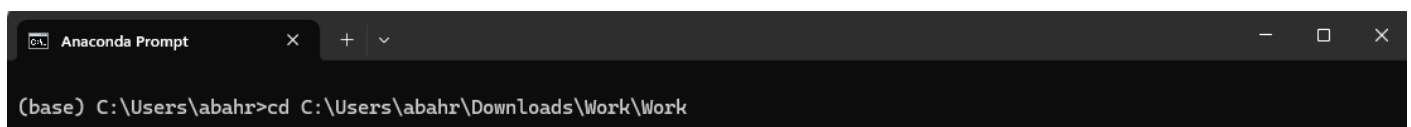


Figure 3 - Working Folder

- Create virtual environment: **conda create --name covid_analysis_env python=3.10**

```
(base) C:\Users\abahr\Downloads\Work\Work>conda create --name covid_analysis_env python=3.10
```

Figure 4 - Virtual environment creation

- Activate it: **conda activate covid_analysis_env**

```
(base) C:\Users\abahr\Downloads\Work\Work>conda activate covid_analysis_env
```

Figure 5 - Environment activation

- **Base** will change into **covid_analysis_env** which means the environment is activated

```
(covid_analysis_env) C:\Users\abahr\Downloads\Work\Work>
```

Figure 6 - Environment activated

- Install all dependencies: **pip install -r requirements.txt**

```
(covid_analysis_env) C:\Users\abahr\Downloads\Work\Work>pip install -r requirements.txt
```

Figure 7 - Install dependencies

Now Open Anaconda Navigator and choose the environment we just created

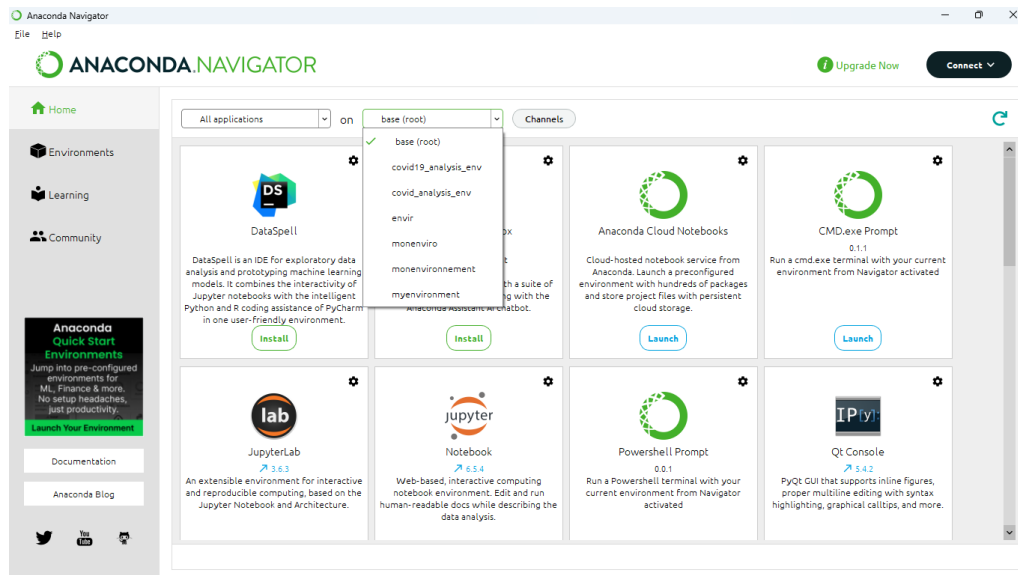


Figure 8 - Environment selection

Search for Spyder, Install It and launch it

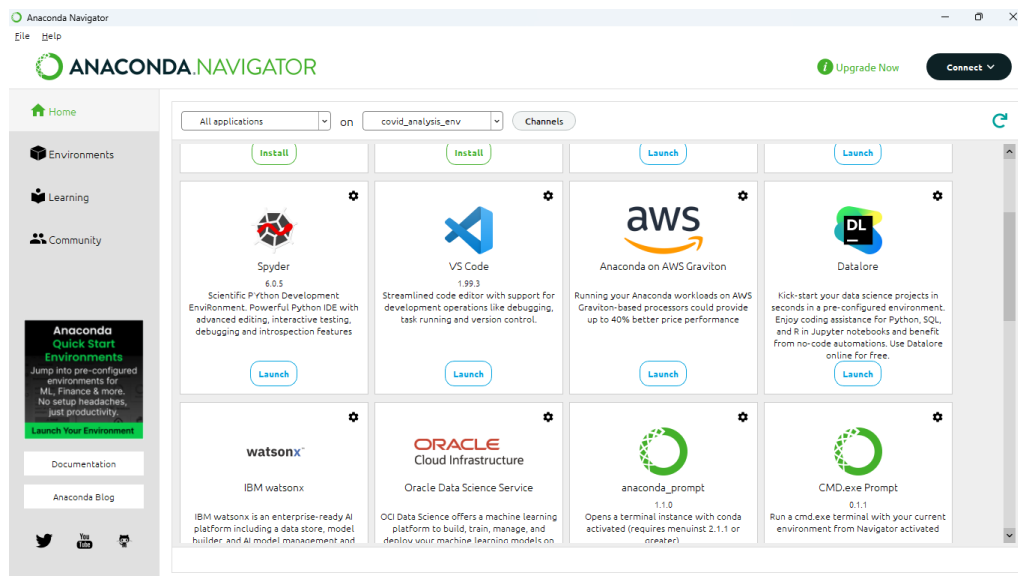


Figure 9 - Spyder installation and launch

Inside Spyder Open the **covid_analysis.py** and the **config.ini** files from **Work/Scripts** folder

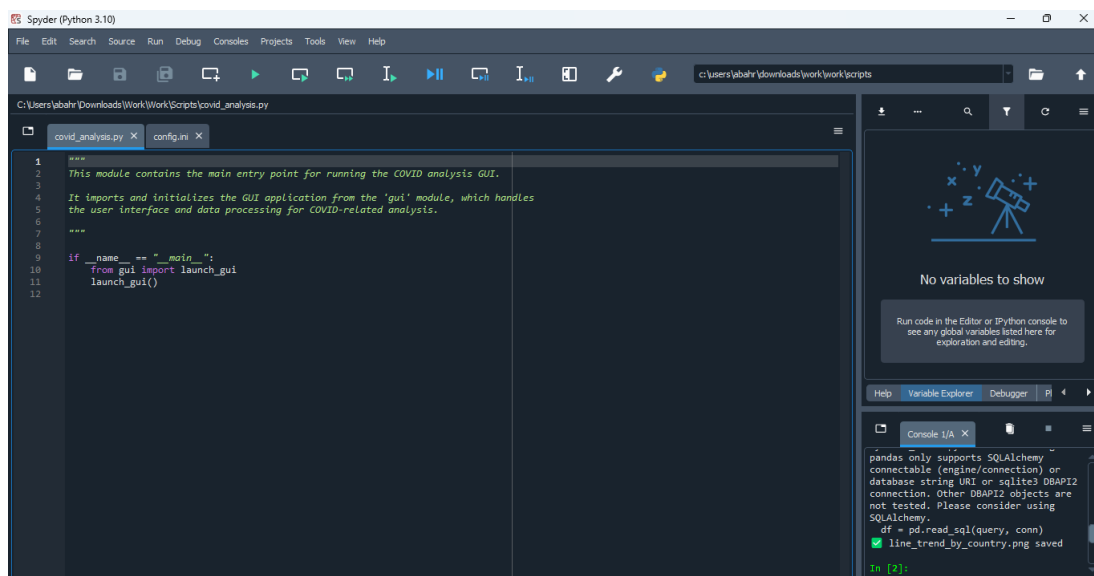


Figure 10 - Spyder IDE

4. Database

Now after Opening `covid_analysis.py` and `config.ini` in spyder we need to configure the **database** information's so:

- Open MySQL Workbench

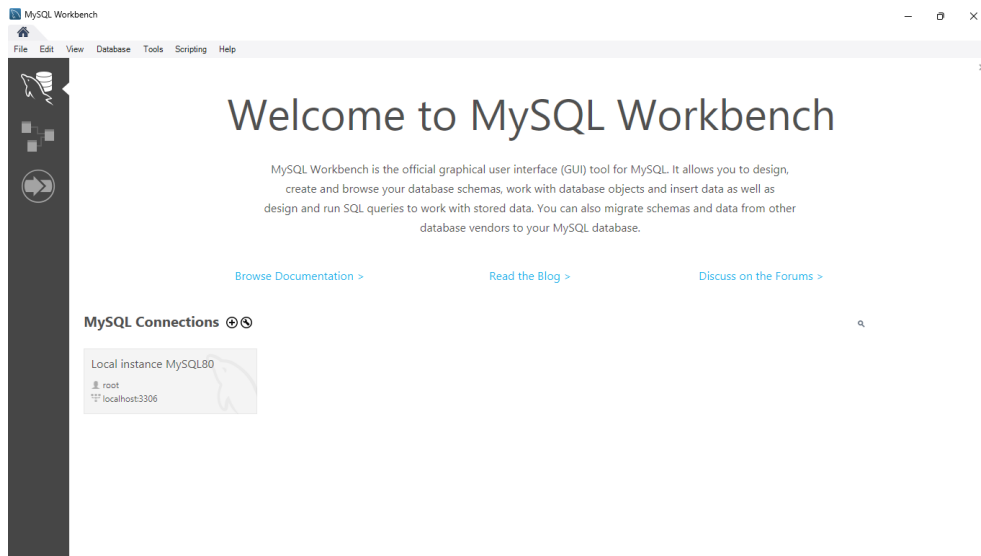


Figure 11 - MySQL Workbench

- At the bottom you can see MySQL localhost connection click on it, it will ask for **user** and **password**, the ones you used during installation, and then it will take you into this interface
- Create a new database schema and name it `covid_analysis_db`, it's empty for now.

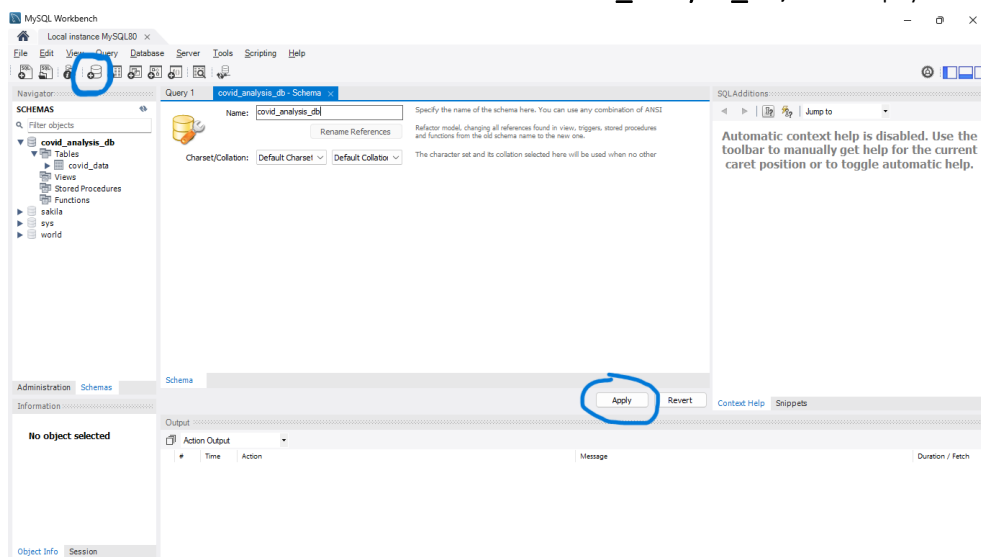


Figure 12 - Database creation

- Now we go back to spyder to update the `config.ini` file for now it's something like this, fill the variables with the right information's: **password** and **user** used during **installation**, and the **name** of the database we just created

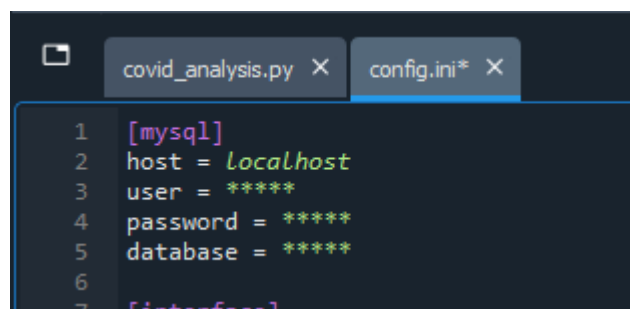


Figure 13 - MySQL configuration

5. Python Interpreter

Inside Spyder in **Tools>Preferences>Python Interpreter** choose the one used by the environment we created previously. (In our example `covid_analysis_env\python.exe`). Click apply, maybe you will need to restart Spyder for the change to take action

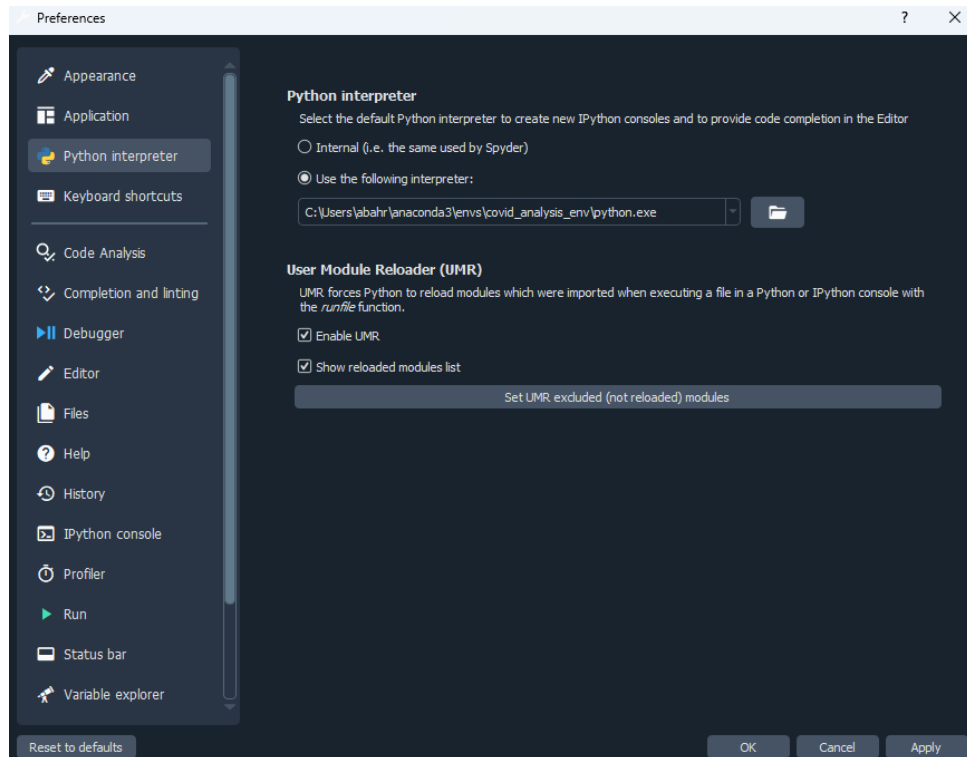


Figure 14 - Python interpreter

You should see the environment we are using at the bottom of Spyder, make sure it's the right one:



Figure 15 - Environment check

6. Run the Program

Now go to the `covid_analysis.py` file opened in Spyder and run it and the GUI must appear

7. User App Guide

Once you run the covid_analysis.py file the tkinter app will appear, a sidebar on the left to make the navigation easy and better. The app has 5 pages one for Data Loading and Preprocessing, a page where you can read data and filter it, a page for statistics, export and report generation, a page for visualization and a page for configuration of the app.

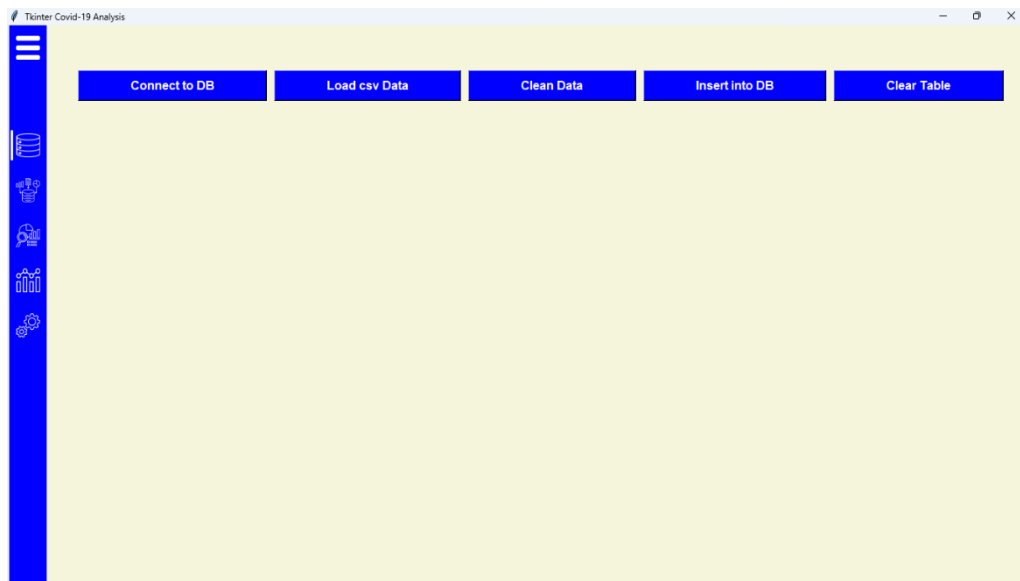


Figure 16 - App

Loading & Preprocessing

The following instructions in **Loading/Preprocessing** page are **mandatory before anything else**, after you have done them, you can skip them on the next launch

- a. At first the user must Connect to the database; this will create the covid data table in database if it doesn't exist.



Figure 17 - Database Connection

- b. The user can use Clear table to delete all rows from database and start fresh.



Figure 18 - Clear Table

- c. Load data will read the data from the csv files inside the DATA folder.

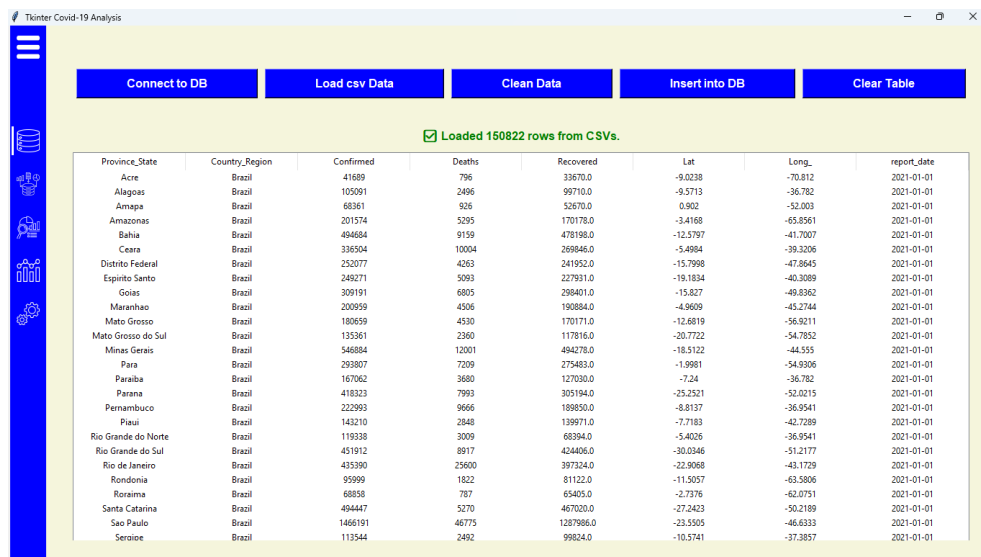


Figure 19 - Data loading from CSVs

- d. Clean data to prevent having null values in database, it helps checking and filling fields with appropriate default data.

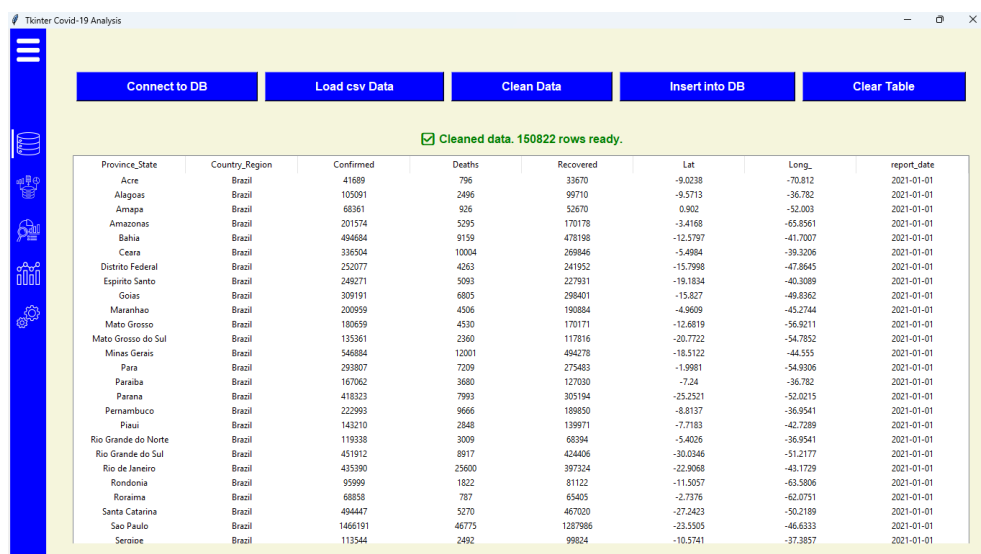
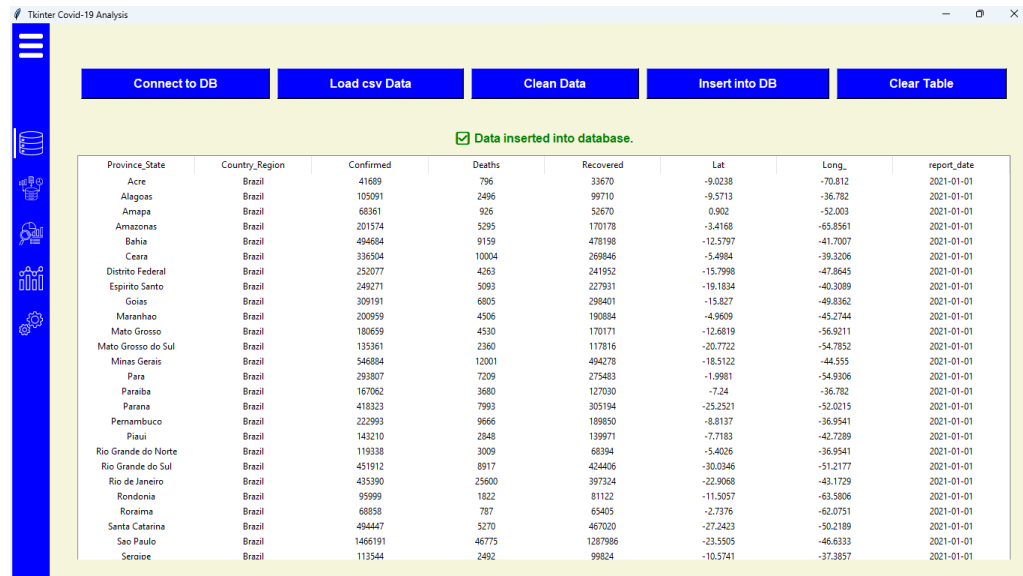


Figure 20 - Data cleaning

e. At the end, the user must insert that data into the database so it can be used later.

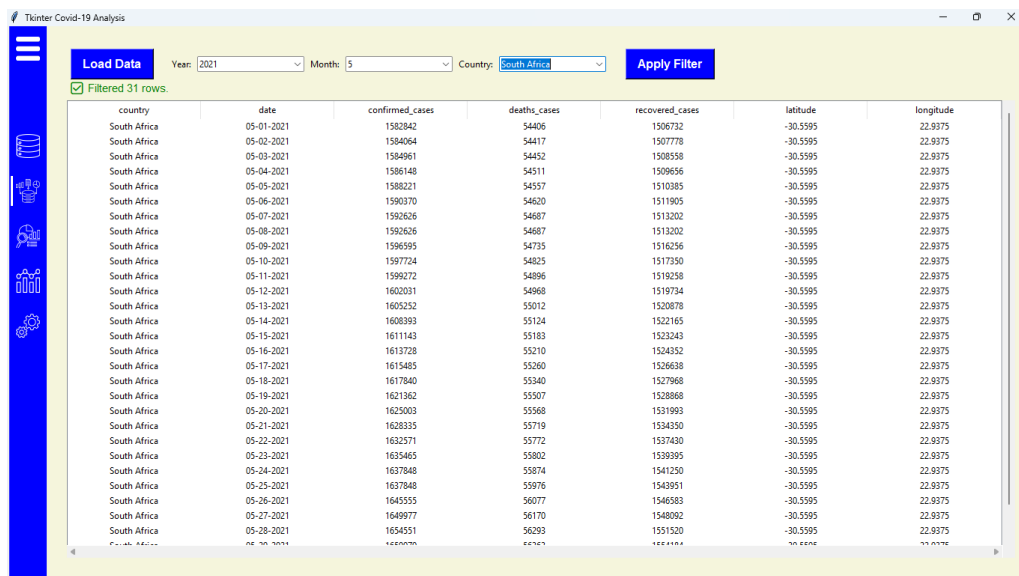


Province_State	Country_Region	Confirmed	Deaths	Recovered	Lat	Long_	report_date
Acre	Brazil	41689	796	33670	-9.0238	-70.812	2021-01-01
Alagoas	Brazil	105091	2496	99710	-9.5713	-36.782	2021-01-01
Amapa	Brazil	68361	926	52670	0.902	-52.003	2021-01-01
Amazonas	Brazil	201574	5295	170178	-3.4168	-65.8561	2021-01-01
Bahia	Brazil	494684	9159	478198	-12.5797	-41.7007	2021-01-01
Ceara	Brazil	336504	10004	269046	-5.4984	-39.3206	2021-01-01
Distrito Federal	Brazil	252077	4263	241952	-15.7998	-47.8645	2021-01-01
Espirito Santo	Brazil	249271	5093	227931	-19.1834	-40.3089	2021-01-01
Goiás	Brazil	309191	6805	298401	-15.827	-49.8362	2021-01-01
Maranhao	Brazil	200959	4506	190884	-4.9609	-45.2744	2021-01-01
Mato Grosso	Brazil	180659	4530	170171	-12.6819	-56.9211	2021-01-01
Mato Grosso do Sul	Brazil	135361	2360	117816	-20.7722	-54.7852	2021-01-01
Minas Gerais	Brazil	546884	12001	494278	-18.5122	-44.555	2021-01-01
Papa	Brazil	293807	7209	275483	-1.9981	-54.9306	2021-01-01
Parabiba	Brazil	167062	3680	127030	-7.24	-36.782	2021-01-01
Parana	Brazil	418323	7993	305194	-25.2521	-52.0215	2021-01-01
Pernambuco	Brazil	222993	9666	189850	-8.8137	-36.9541	2021-01-01
Piaui	Brazil	143210	2848	139971	-7.7183	-42.7289	2021-01-01
Rio Grande do Norte	Brazil	119338	3009	68394	-5.4026	-36.9541	2021-01-01
Rio Grande do Sul	Brazil	451912	8917	424406	-30.0346	-51.2177	2021-01-01
Rio de Janeiro	Brazil	435390	25600	397324	-22.9068	-43.1729	2021-01-01
Rondonia	Brazil	95999	1822	81122	-11.5057	-63.5806	2021-01-01
Roraima	Brazil	68858	787	65405	-2.7376	-62.0751	2021-01-01
Santa Catarina	Brazil	494447	5270	467020	-27.2423	-50.2189	2021-01-01
Sao Paulo	Brazil	1466191	46775	1287986	-23.5505	-46.6333	2021-01-01
Serarioe	Brazil	113544	2492	99824	-10.5741	-37.3857	2021-01-01

Figure 21 - Data insertion

Data Filtering

The next page is **Data Filtering** where the user must first load data and then he can filter by **year**, **country** and **month**, the user can choose either **one**, **two** or **three** at a time to filter.



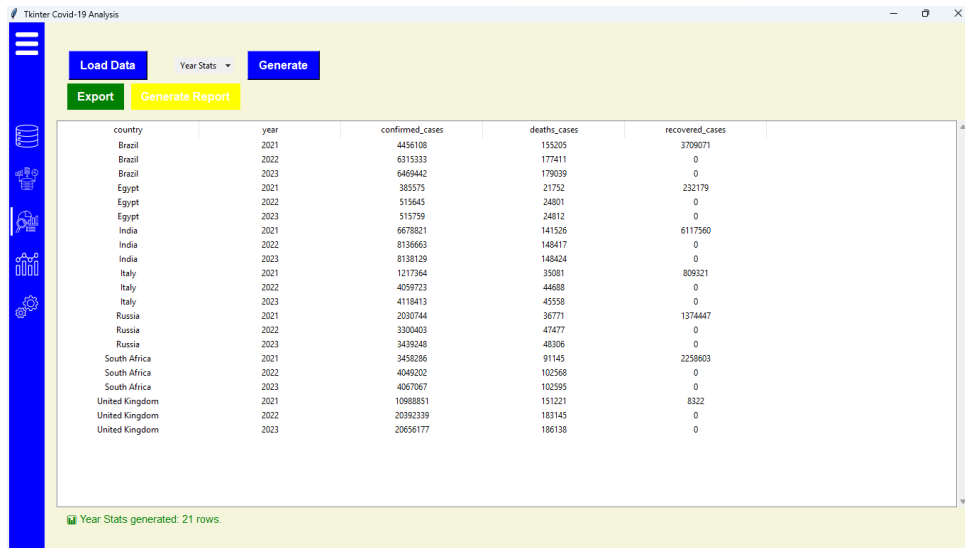
country	date	confirmed_cases	deaths_cases	recovered_cases	latitude	longitude
South Africa	05-01-2021	1582842	54406	1506732	-30.5595	22.9375
South Africa	05-02-2021	1584064	54417	1507778	-30.5595	22.9375
South Africa	05-03-2021	1584961	54452	1508558	-30.5595	22.9375
South Africa	05-04-2021	1586148	54511	1509656	-30.5595	22.9375
South Africa	05-05-2021	1588221	54557	1510385	-30.5595	22.9375
South Africa	05-06-2021	1590370	54620	1511905	-30.5595	22.9375
South Africa	05-07-2021	1592626	54687	1513202	-30.5595	22.9375
South Africa	05-08-2021	1592626	54687	1513202	-30.5595	22.9375
South Africa	05-09-2021	1596595	54735	1516256	-30.5595	22.9375
South Africa	05-10-2021	1597724	54825	1517350	-30.5595	22.9375
South Africa	05-11-2021	1599272	54896	1518258	-30.5595	22.9375
South Africa	05-12-2021	1602031	54968	1519734	-30.5595	22.9375
South Africa	05-13-2021	1605252	55012	1520878	-30.5595	22.9375
South Africa	05-14-2021	1608393	55124	1522165	-30.5595	22.9375
South Africa	05-15-2021	1611143	55183	1523243	-30.5595	22.9375
South Africa	05-16-2021	1613728	55210	1524352	-30.5595	22.9375
South Africa	05-17-2021	1615485	55260	1526638	-30.5595	22.9375
South Africa	05-18-2021	1617840	55340	1527968	-30.5595	22.9375
South Africa	05-19-2021	1621362	55507	1528888	-30.5595	22.9375
South Africa	05-20-2021	1623003	55568	1529199	-30.5595	22.9375
South Africa	05-21-2021	1623835	55719	1534550	-30.5595	22.9375
South Africa	05-22-2021	162571	55772	1537420	-30.5595	22.9375
South Africa	05-23-2021	1635405	55802	1539395	-30.5595	22.9375
South Africa	05-24-2021	1637848	55874	1541250	-30.5595	22.9375
South Africa	05-25-2021	1637848	55976	1543951	-30.5595	22.9375
South Africa	05-26-2021	1645555	56077	1546583	-30.5595	22.9375
South Africa	05-27-2021	1649977	56170	1548092	-30.5595	22.9375
South Africa	05-28-2021	1654551	56293	1551520	-30.5595	22.9375

Figure 22 - Data Filtering Page

Statistics

The next page is **Statistics** where the user must first load data and then can select from a dropdown menu an option for statistics.

- The user can after, generate a report, which will be stored inside **Work/Output/reports** folder.
- The user can after, export the statistic, which will be stored inside **Work/Output/exports** folder.



The screenshot shows the 'Statistics' page of the Tkinter Covid-19 Analysis application. It features a sidebar with icons for menu, database, report, and settings. The main area has buttons for 'Load Data', 'Year Stats', 'Generate', 'Export', and 'Generate Report'. A table displays COVID-19 statistics for various countries from 2021 to 2023. A status message at the bottom indicates 'Year Stats generated: 21 rows'.

country	year	confirmed_cases	deaths_cases	recovered_cases
Brazil	2021	4456108	153205	3709071
Brazil	2022	6315333	177411	0
Brazil	2023	6469442	179039	0
Egypt	2021	385575	21752	232179
Egypt	2022	515645	24801	0
Egypt	2023	515799	24812	0
India	2021	6678821	141526	6117560
India	2022	8136663	148417	0
India	2023	8138129	148424	0
Italy	2021	1217364	35091	806321
Italy	2022	4059723	44688	0
Italy	2023	4118413	45558	0
Russia	2021	2030744	36771	1374447
Russia	2022	3300403	47477	0
Russia	2023	3439248	48306	0
South Africa	2021	3458286	91145	2258603
South Africa	2022	4049202	102568	0
South Africa	2023	4067067	102595	0
United Kingdom	2021	10988851	151221	8322
United Kingdom	2022	20392339	183145	0
United Kingdom	2023	20656177	186138	0

Figure 23 - Statistics page

Visualization

The next page is **Visualization** where the user must first load data and then can select from a dropdown menu an option for visualization.

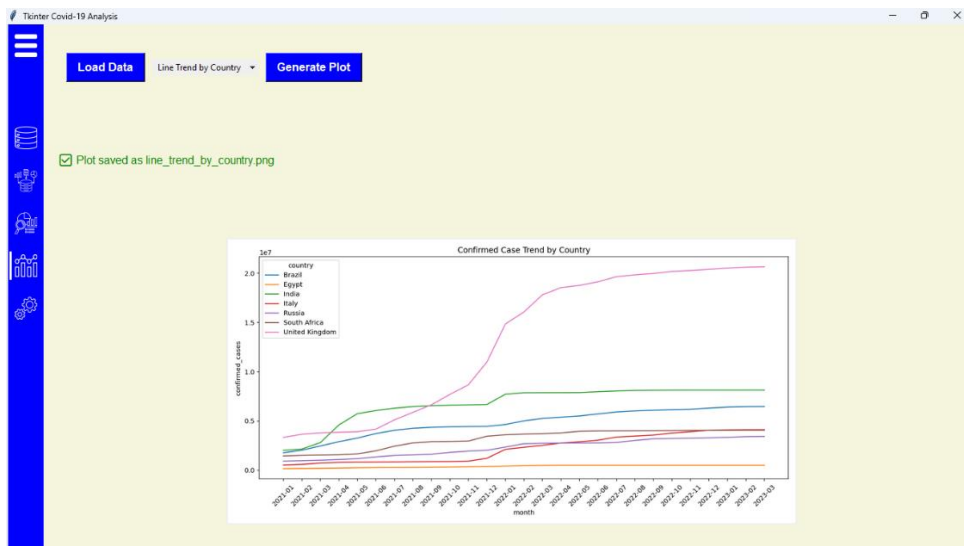
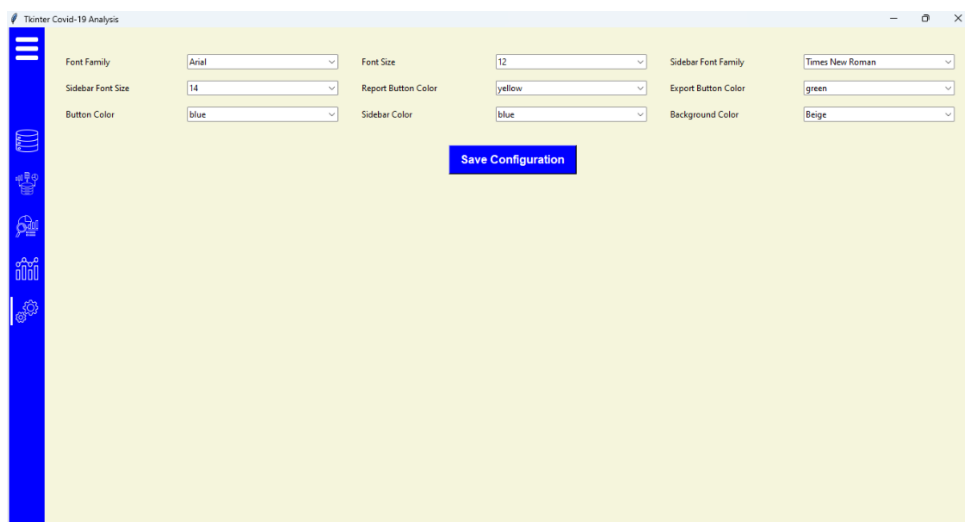


Figure 24 - Visualization page

Configuration

The last page is the **Configuration** page where the user can customize the interface by changing the fonts and colors, when the changes are saved, the app will automatically restart to show updates.



The screenshot shows the 'Configuration' page of the Tkinter Covid-19 Analysis application. It features a sidebar with icons for menu, database, report, and settings. The main area contains configuration options for Font Family, Font Size, Sidebar Font Family, Sidebar Font Size, Report Button Color, Export Button Color, Button Color, Sidebar Color, and Background Color. A 'Save Configuration' button is located at the bottom.

Figure 25 - Configuration page