Everything You Need To Know About PandasGUI

Kaustubh Gupta

Last Updated : 26 Dec, 2023

This article was published as a part of the Data Science Blogathon

Introduction

How many times you have used Pandas library for your Data Science tasks? Almost every time! Pandas is an essential library for data manipulation and generating insights from the dataset in the form of summary tables, visualizations, and much more.

PandasGUI is a Python-based library that facilitates data manipulation and summary statistics to be applied on the dataset using GUI. That means that all the operations will be performed via the graphical user interface (GUI), but pandas will be used to execute them under the hood.

In this article, I will walk you through all the features of this library, how it generates interactive plots, and how to access the automatic Python code generated for all the operations performed via GUI!

(Image by Author, Made in Canva)

Installation

PandasGUI is a Python package and can be installed via the pip package manager. Note: I would recommend using Python 3.8 and above for installation. Also, make sure to create a separate environment (virtual environment) and install the library in that environment.

Creating virtual environment (conda)

Run these commands in your terminal

```
conda create -n nameofenv python=version
```

Setup

First thing first, load the library. We are interested in the show() function of the pandasGUI library that initializes the application.

```
from pandasgui import show
```

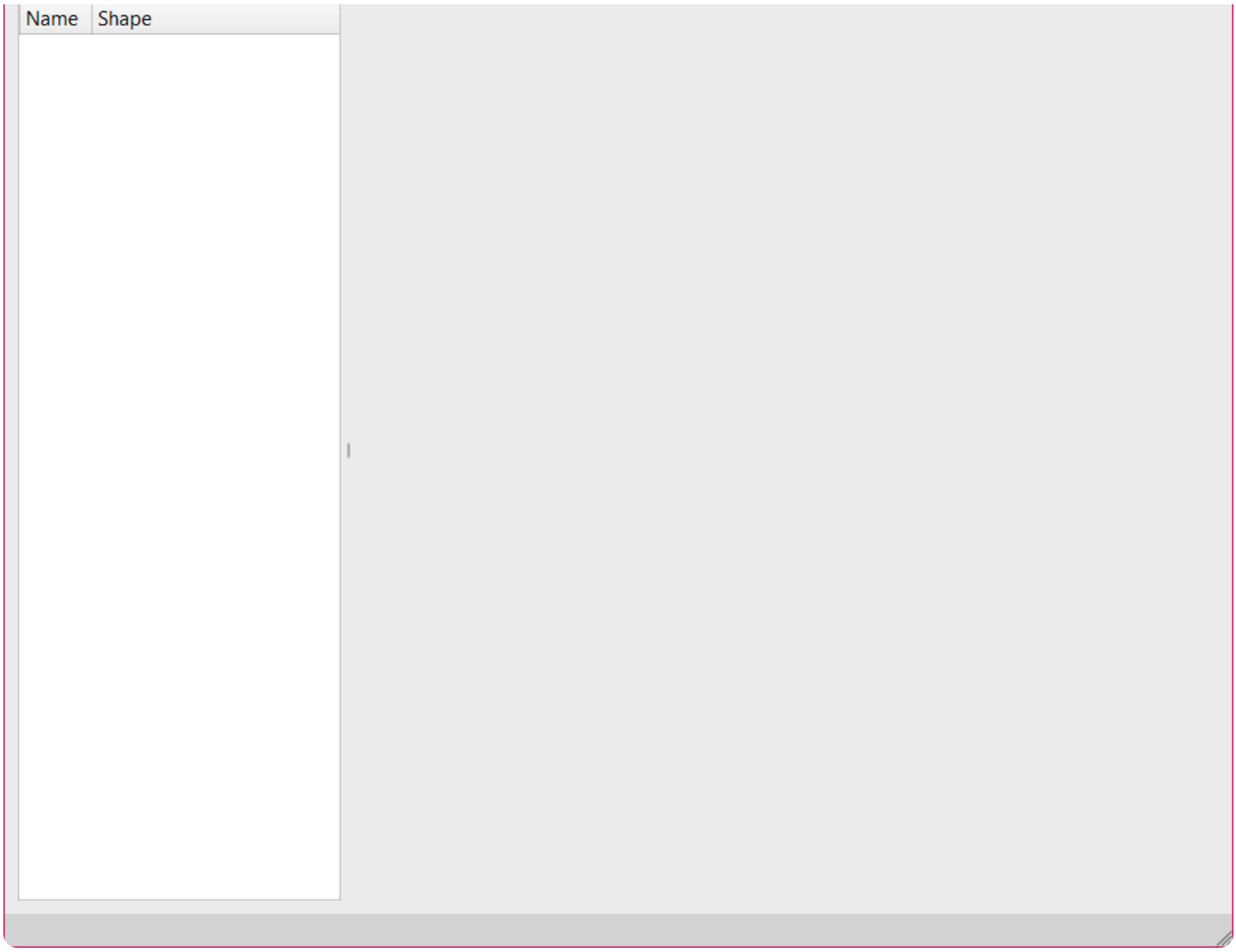
This function can be used in two ways:

1. You can directly instantiate the pandasGUI show function which will open up the application without any dataset.
2. You can pass the dataframe to the function and the application will be populated with the rows and columns of the dataset. This way, you can load multiple datasets at the same time bypassing all the dataframe objects to the show() method.

For now, we will load the application without any dataset,

```
show()
```

And you will get a screen like this:



First-time screen

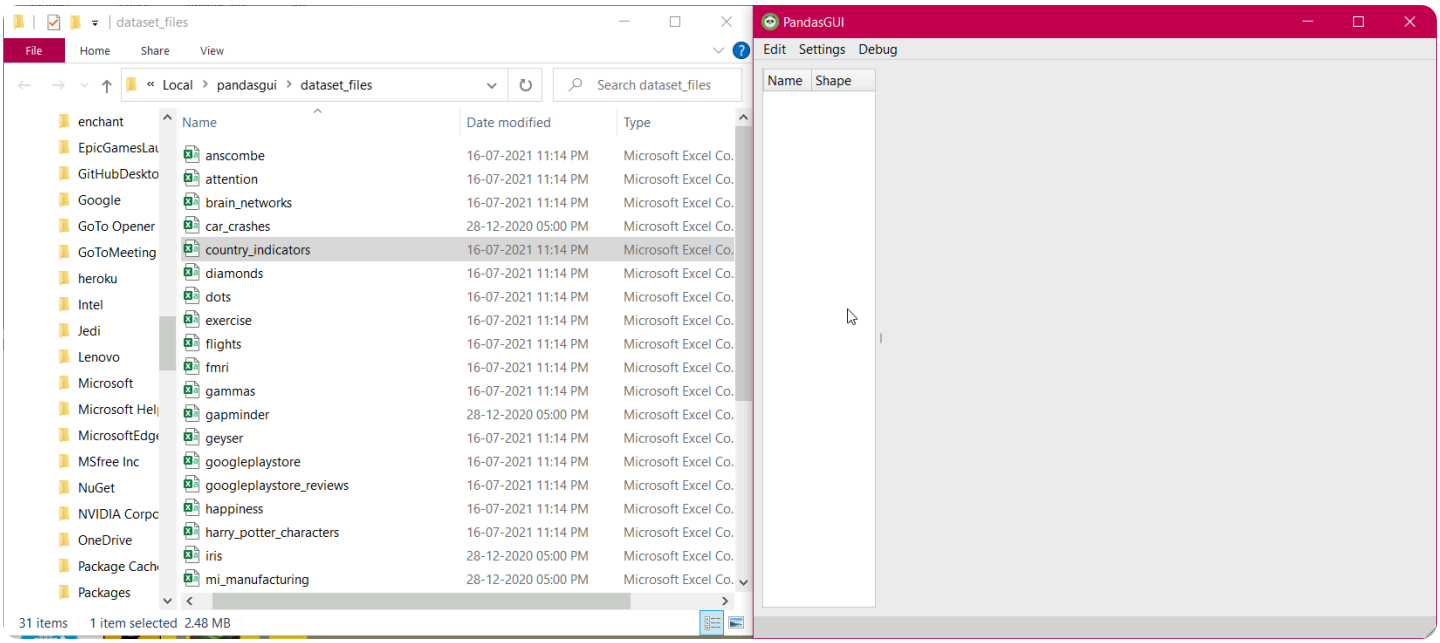
Various features

Now, we will explore all the features of this application step by step.

Loading the Dataset

3. Import dataframe from the clipboard using Edit menu's "Import From Clipboard"
4. Use Debug menu's "Browse Sample Datasets" to load any sample dataset for testing.
5. Drag and drop the CSV files into the application

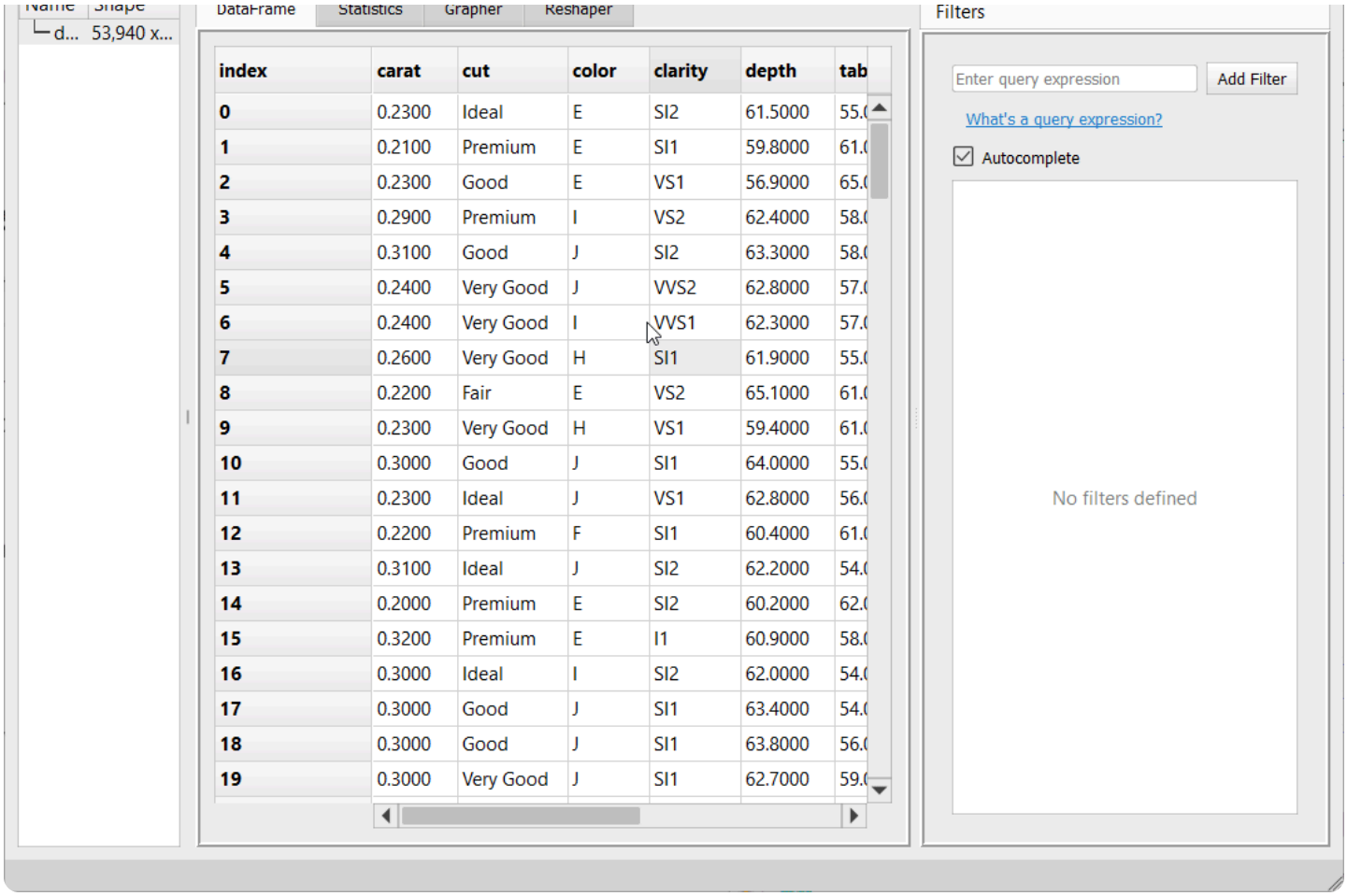
See the GIF below:



On-Screen operations

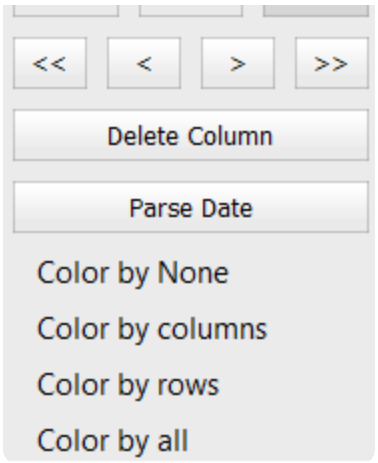
Apart from the various tabs and menus offered by the pandasGUI application, there are some of the operations which you can apply directly by triggering the GUI elements displayed:

1. All the cells of the dataset shown are editable. You can click on any of the cells and



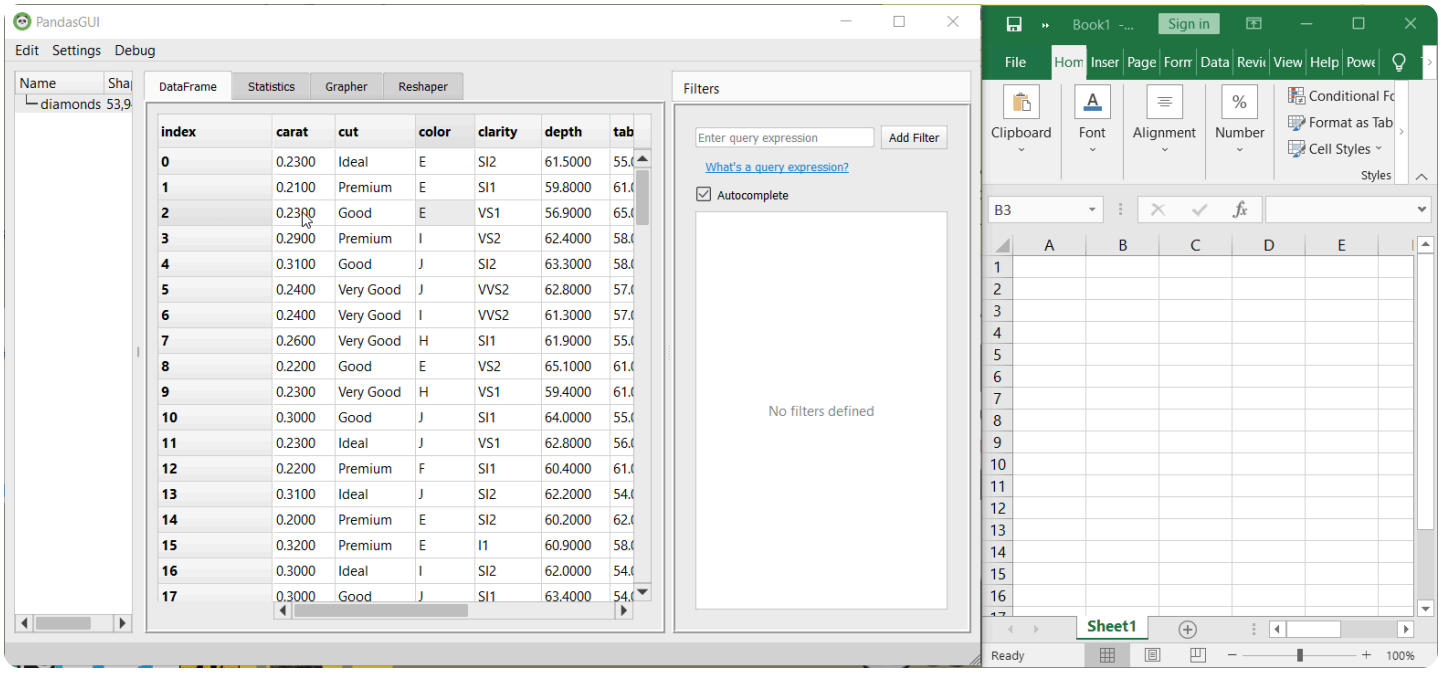
index	carat	cut	color	clarity	depth	tab
0	0.2300	Ideal	E	SI2	61.5000	55.(
1	0.2100	Premium	E	SI1	59.8000	61.(
2	0.2300	Good	E	VS1	56.9000	65.(
3	0.2900	Premium	I	VS2	62.4000	58.(
4	0.3100	Good	J	SI2	63.3000	58.(
5	0.2400	Very Good	J	VVS2	62.8000	57.(
6	0.2400	Very Good	I	VVS1	62.3000	57.(
7	0.2600	Very Good	H	SI1	61.9000	55.(
8	0.2200	Fair	E	VS2	65.1000	61.(
9	0.2300	Very Good	H	VS1	59.4000	61.(
10	0.3000	Good	J	SI1	64.0000	55.(
11	0.2300	Ideal	J	VS1	62.8000	56.(
12	0.2200	Premium	F	SI1	60.4000	61.(
13	0.3100	Ideal	J	SI2	62.2000	54.(
14	0.2000	Premium	E	SI2	60.2000	62.(
15	0.3200	Premium	E	I1	60.9000	58.(
16	0.3000	Ideal	I	SI2	62.0000	54.(
17	0.3000	Good	J	SI1	63.4000	54.(
18	0.3000	Good	J	SI1	63.8000	56.(
19	0.3000	Very Good	J	SI1	62.7000	59.(

2. Right-click on the column headers will give you an interactive column menu containing options for sorting the column by ascending, descending order, moving the column to extreme ends or within the range, parsing the column containing dates as a string to pandas datetime format, and deleting the column option.



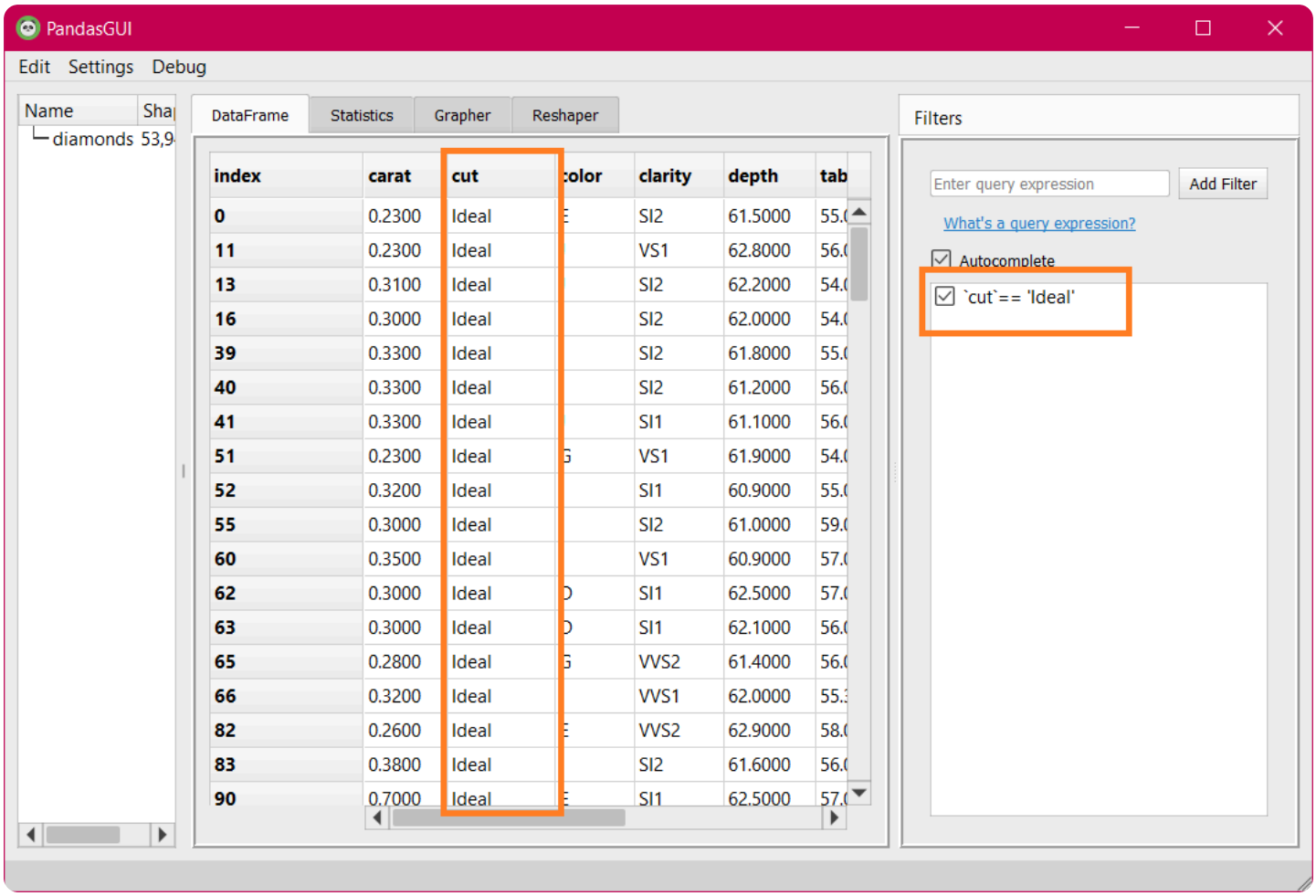
Column Header Options

3. By dragging and selecting any portion of the dataset, you can copy that section and paste it into any cell of the excel sheet. It will be automatically converted to a tabular format as shown in the application itself.

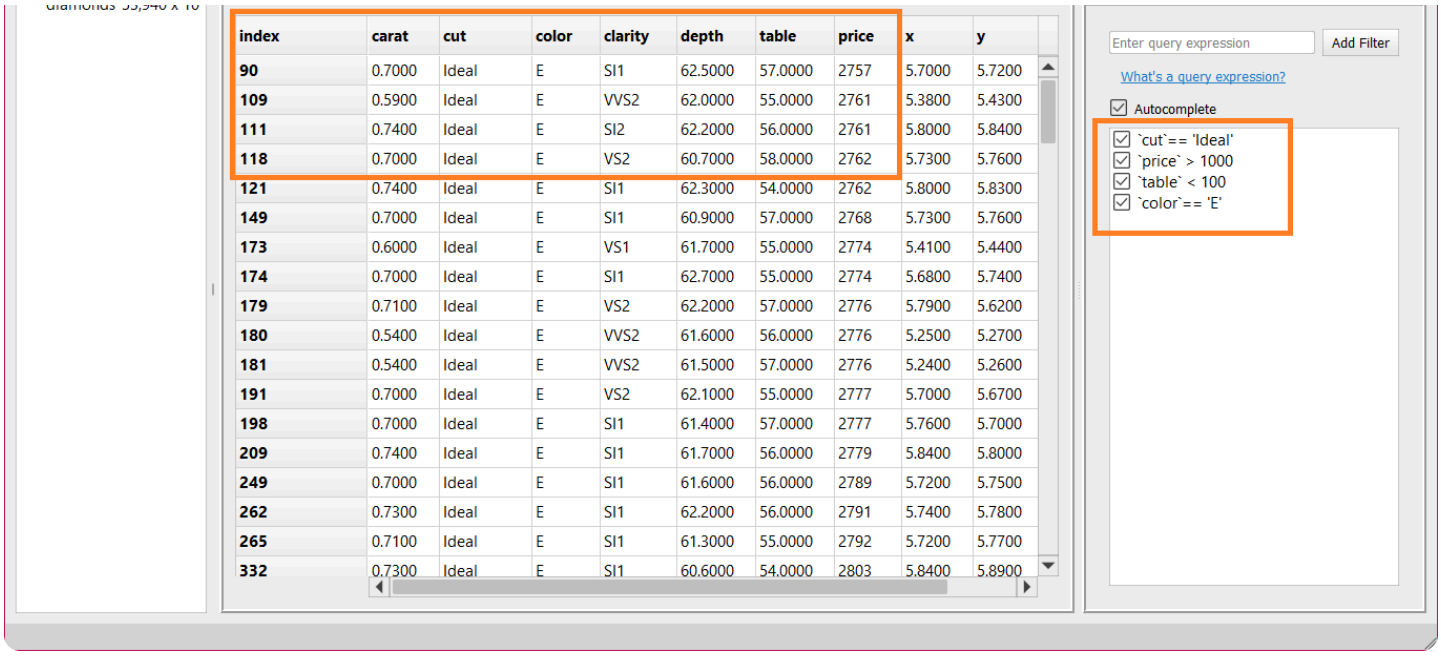


task. It helps in isolating a segment of the dataset and work on that isolated section. In pandas, you usually use comparisons, threshold values with columns, or whole datasets to filter data. In pandasGUI, you can do the same type of filtering using the Filters tab on the right.

Just type out the expression for filtering and only the data that satisfies the condition will be displayed. See an example below:



Filtering data by “cut=ideal”



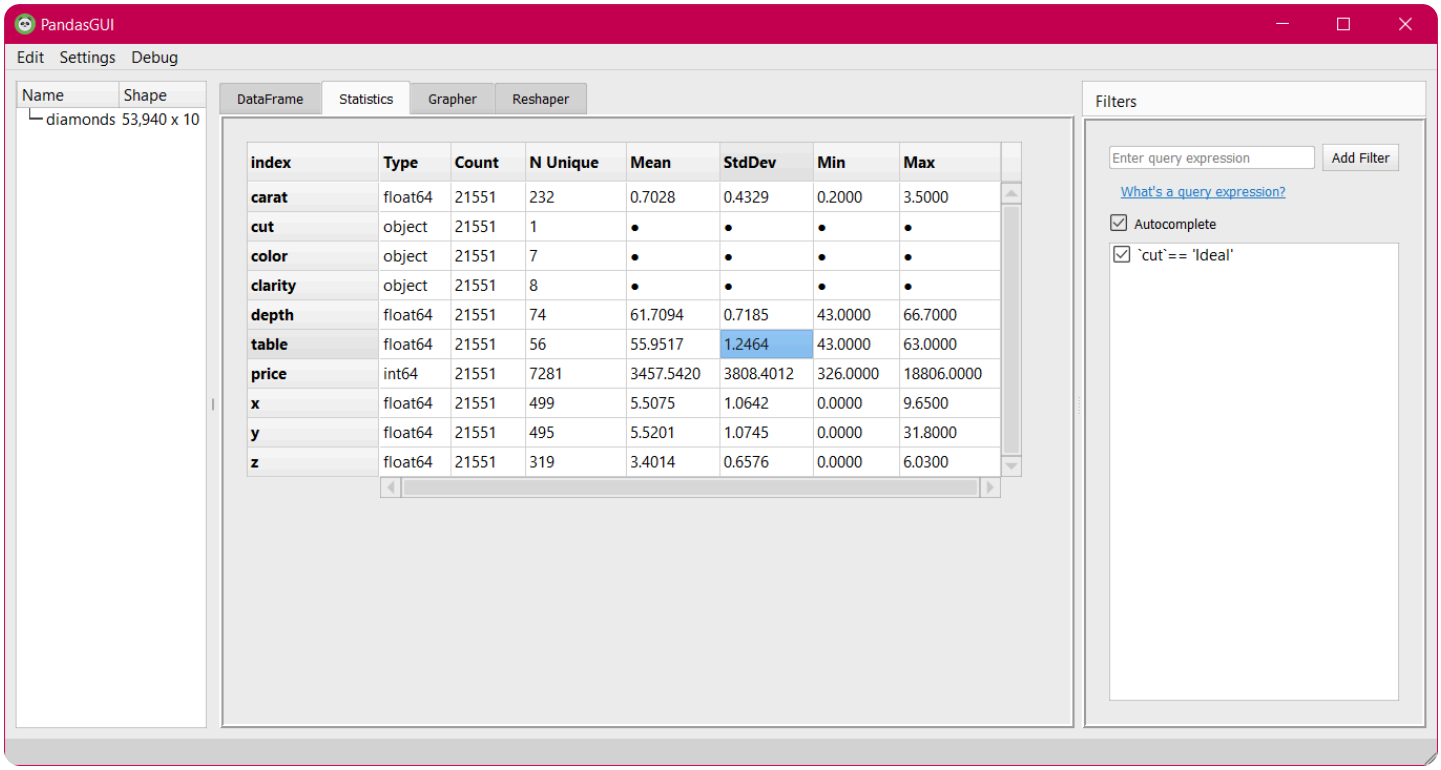
An example where multiple filters are applied to the dataset

> Note: All the filters are applied using “pandas.dataframe.query()” method under the hood.

Statistics Tab

Statistics play an important role in a descriptive view of all the features of the dataset. These contain parameters such as percentiles which helps in getting insight into how the data is spread, mean, which is affected outliers, can still tell us about the center of data and standard deviation tells about how much the column data varies inside. A column with 0 standard deviations will be of no use as this would mean that all the

data type, count, the number of unique values, mean, standard deviation, and min/max. For string type values, numerically calculated parameters such as mean will be null.



index	Type	Count	N Unique	Mean	StdDev	Min	Max
carat	float64	21551	232	0.7028	0.4329	0.2000	3.5000
cut	object	21551	1	•	•	•	•
color	object	21551	7	•	•	•	•
clarity	object	21551	8	•	•	•	•
depth	float64	21551	74	61.7094	0.7185	43.0000	66.7000
table	float64	21551	56	55.9517	1.2464	43.0000	63.0000
price	int64	21551	7281	3457.5420	3808.4012	326.0000	18806.0000
x	float64	21551	499	5.5075	1.0642	0.0000	9.6500
y	float64	21551	495	5.5201	1.0745	0.0000	31.8000
z	float64	21551	319	3.4014	0.6576	0.0000	6.0300

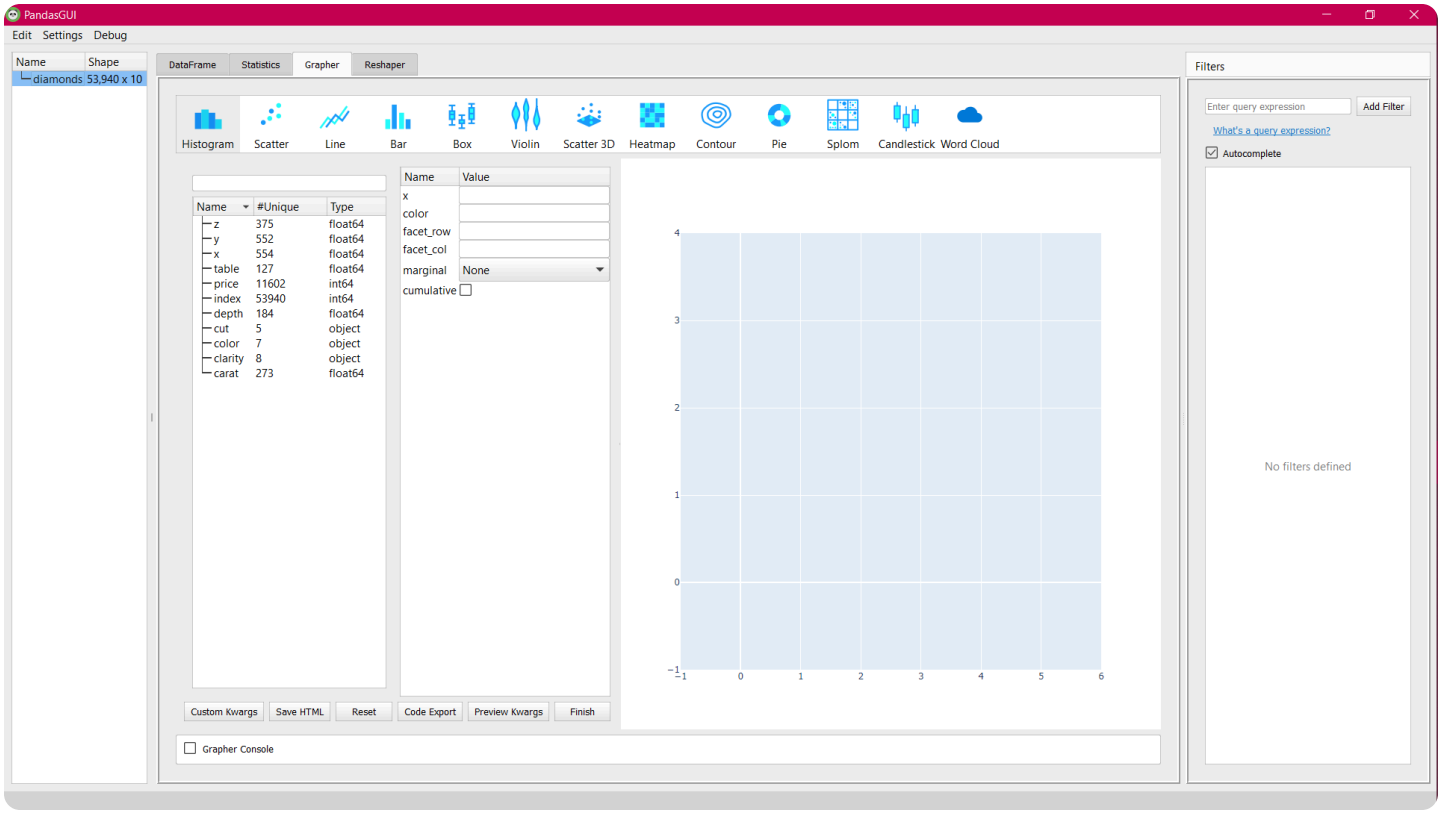
Statistical Summary

Grapher Tab

Graphs and visualizations is a very powerful tool in presenting a summary of the data using visual elements. For instance, Distribution graphs can help in determining whether the column values align with normal distribution/bell-shaped curve which simplifies our task for determining the population parameters and concluding facts

advantages, and therefore, plotting data is important.

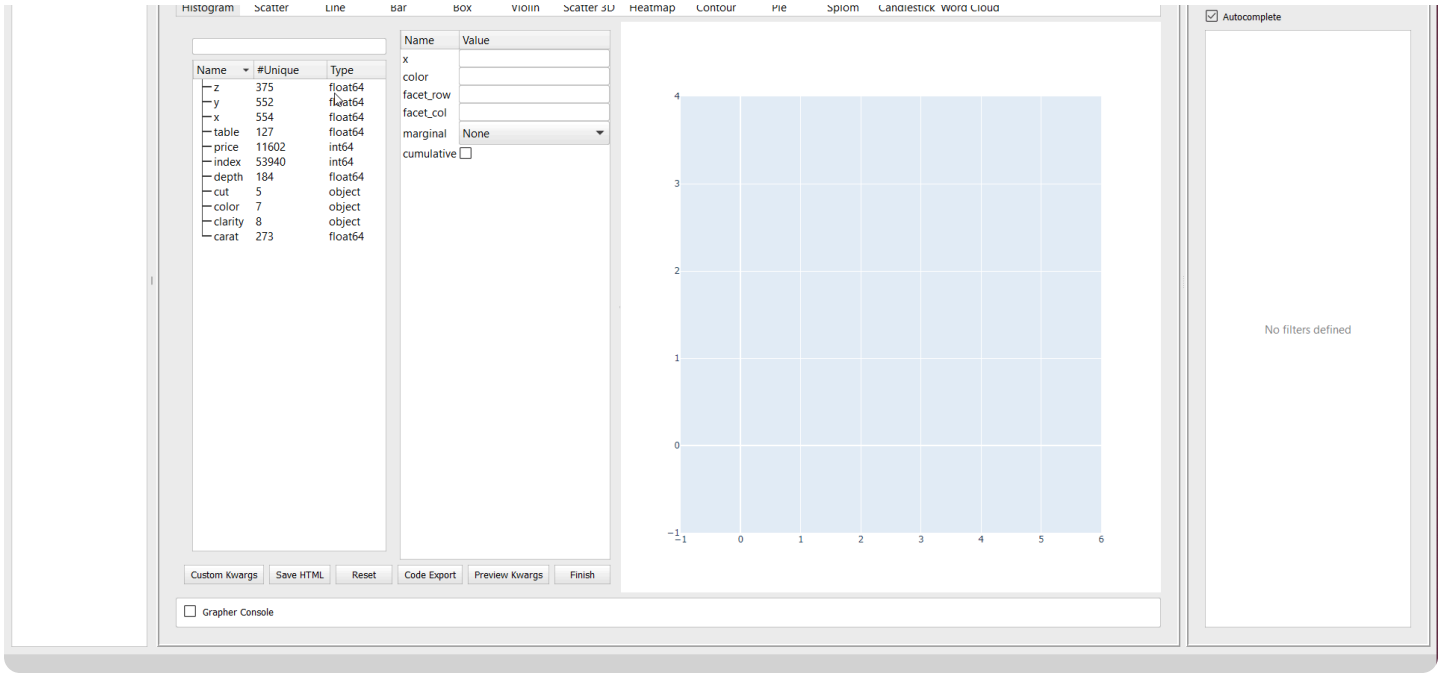
To plot such visualizations using pandasGUI, simply switch to the Grapher tab and you will get an initial screen like this:



Grapher tab initial screen

You can choose from histograms, scatter plots, bar charts, box plots, violin plots, scatter 3D, heatmap, contour plots, pie charts, splom, candlestick, candlestick, and word clouds. All these plots are created using the Plotly library at the backend and therefore, the plots are interactive in nature.

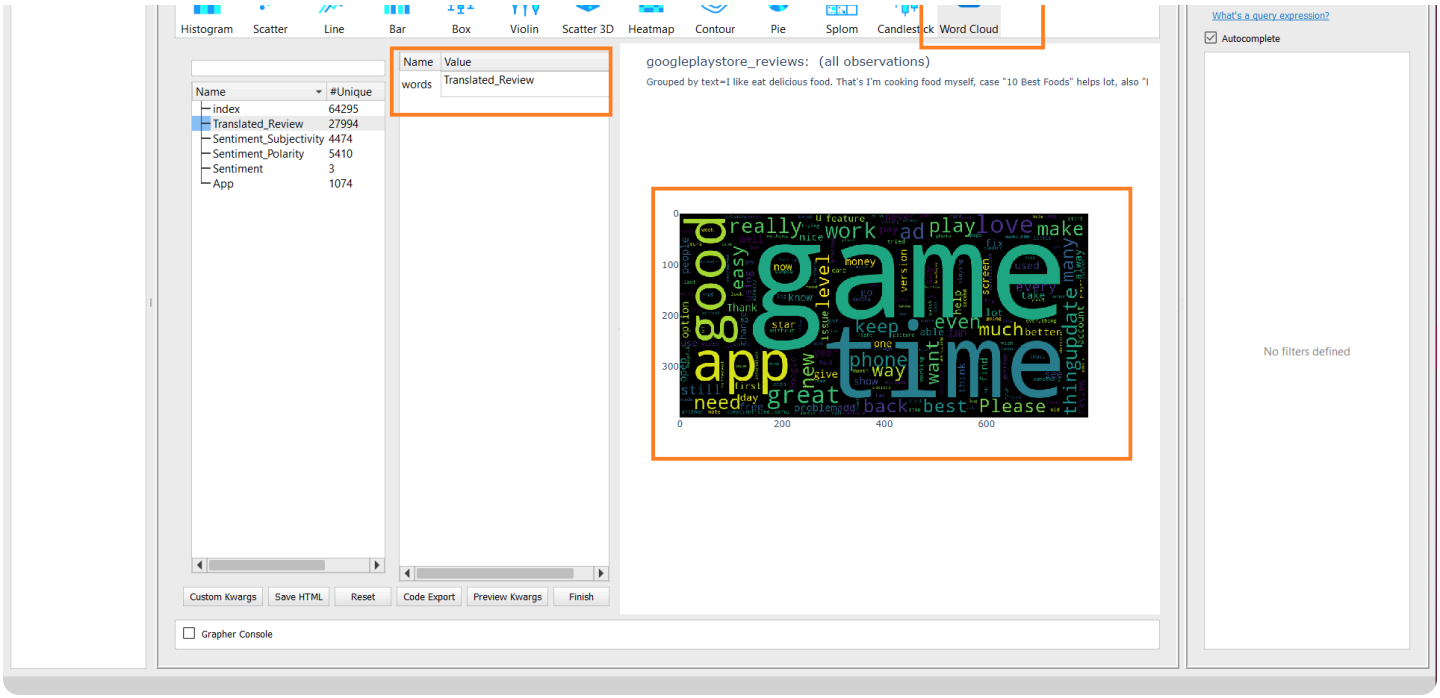
Let's plot some data. To do the plotting, simply select any type of plot, and then drag



Scatter plots

You can see that as “x” increases, the “price” increases exponentially. The plot is interactive in nature due to the plotly backend. The parameters such as “x”, “y”, “color” change as per the plot to be displayed. For instance, you will not get “x”, “y” parameters for the word cloud.

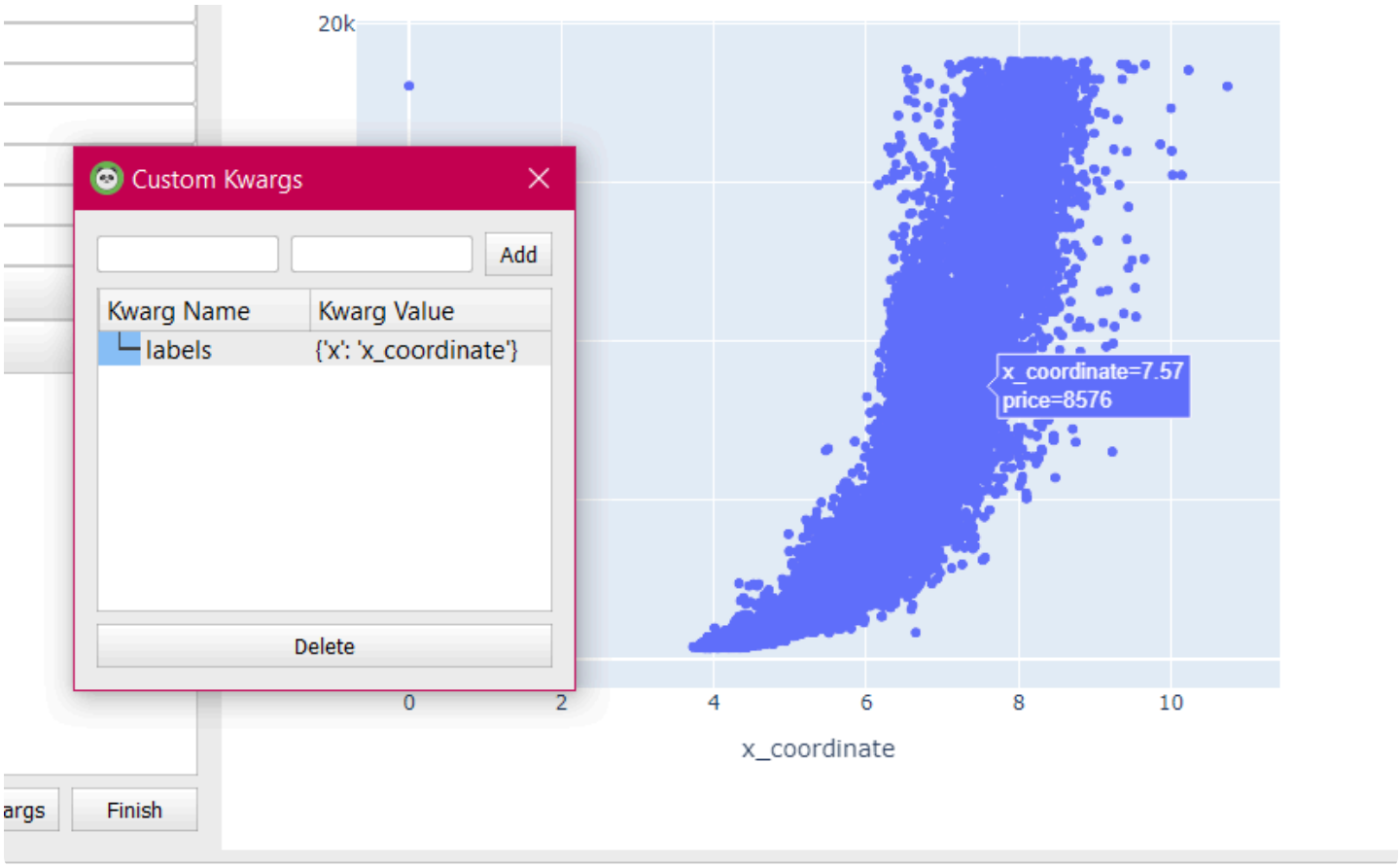
Let’s plot a word cloud for a different dataset, Google Play store app reviews, that contain textual information about the app’s feedback:



Word cloud for Google Play store app reviews

Other common options are available for each plot are available at the bottom of the column and parameters panel. These include:

Custom kwargs: As the plotting is done plotly, you can add all the kwargs supported by plotly express. The current arguments may be limited for best use-case but you can pass your custom arguments using this option. For example, the default name of the column “x” can be changed to something else using the “labels” argument:



Custom kwargs

Save HTML: Using this option, you can save the plot generated into an HTML file. This file can be used independently without the need for any backend. The HTML file displays the graph interactively using JavaScript.

Reset: This option clears all the parameters values.

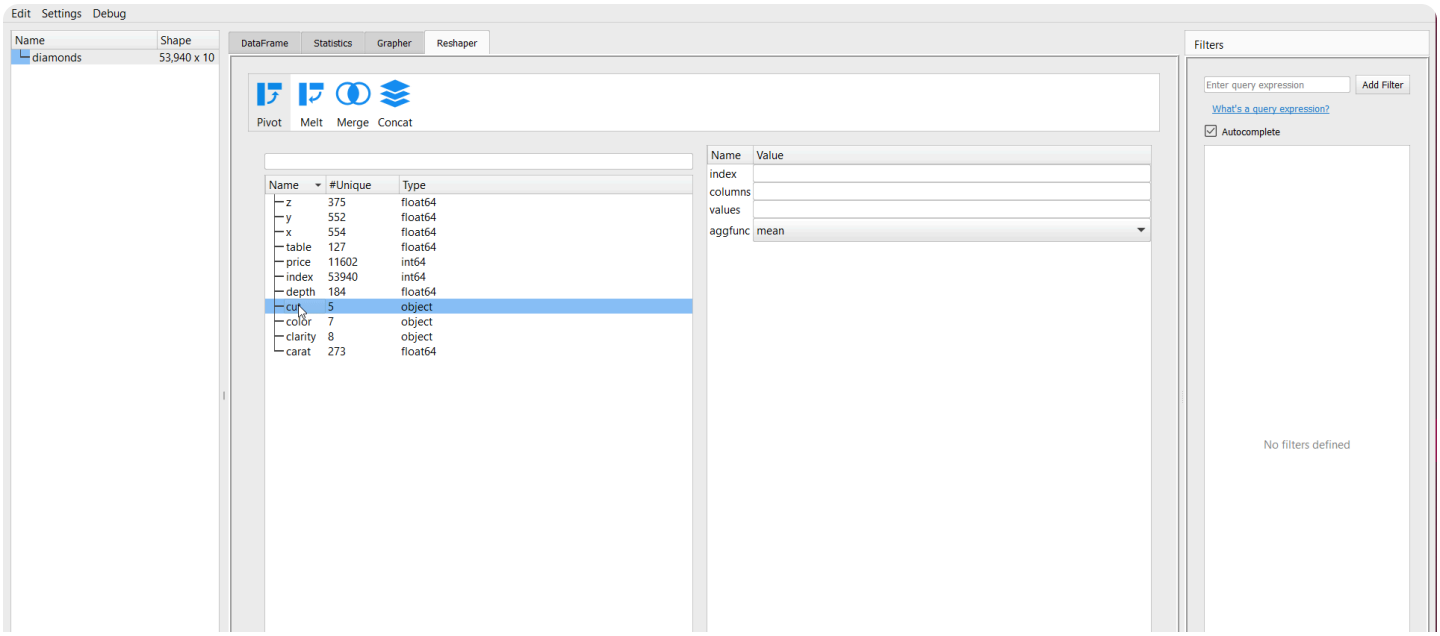
Code Export: You can export the code for plotting the graph using this option. The code will lack the custom keywords provided but the basic code will be helpful to get

This is useful in cases when you have applied the custom arguments and you want to reload the graph with those arguments applied.

Reshaper Tab

Summary tables such as Pivot tables play a crucial role to summarize data based on aggregation functions such as mean, median, sum, etc. You can simply create these tables using the reshaper tab. This tab offers Pivot tables, melt tables, merging and concatenating tables. Simply drag the column names from the left panel to the parameter required and click on finish.

You will get a new dataframe with the summary selected. Let's create a pivot table for summary for "cut" as the index, "color" as the new column, and aggregated mean values "depth" of diamonds dataset:



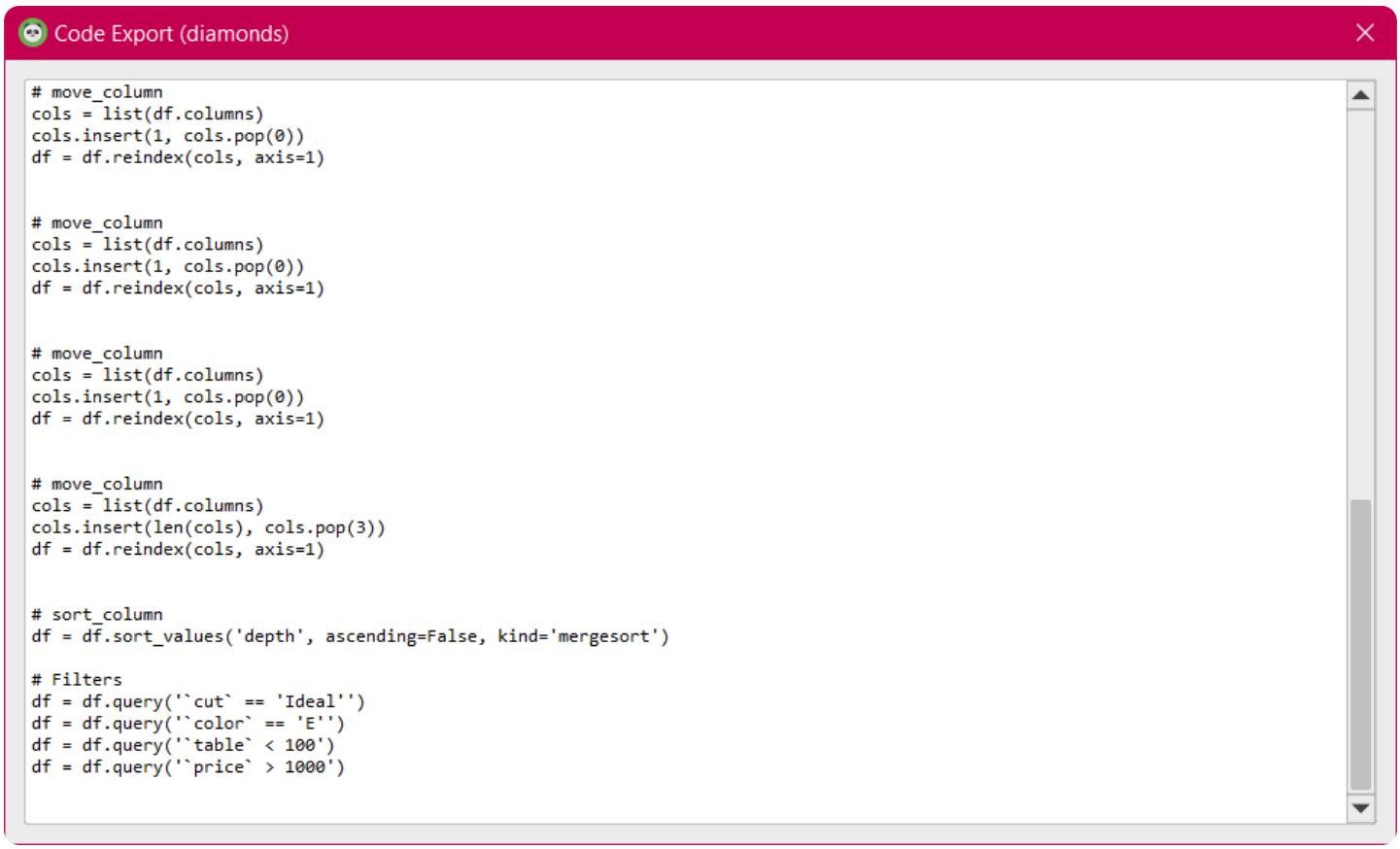
using the setting's menu set preference option. You can change the theme of the application to dark mode, disable the editable on-screen features, set auto_finish to True, change render mode, change default aggregation function from mean to other function, and format the title.

Name	Value
editable	☑ editable
theme	light
refresh_statistics	☑ refresh_statistics
auto_finish	☑ auto_finish
render_mode	auto
aggregation	mean
title_format	('{name}: ' '{title_columns}{title_dimensions}{names}{title_y}{title_z}{over_by}{title_x} ' '{selection} _{{groupings}{filters} {title_trendline}}')

Finish

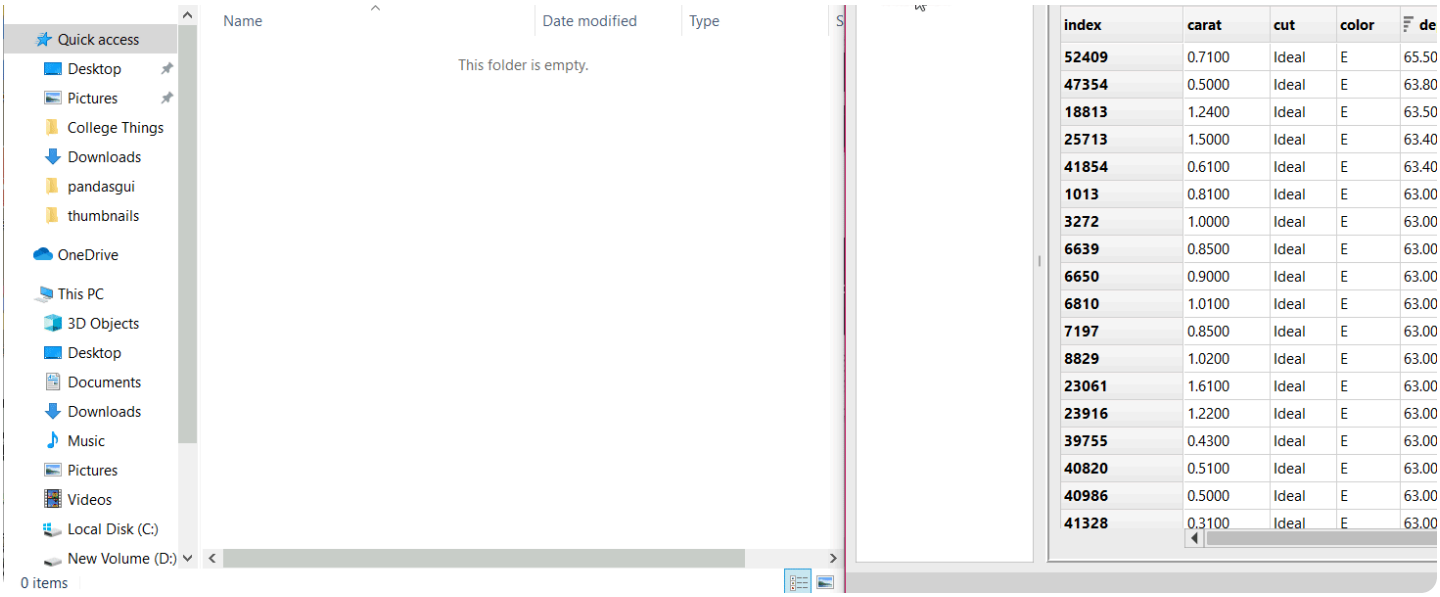
Reset To Defaults

pandas equivalent code for all the operations performed using the edit menu's export code option. This code will be formatted in the order of the operation with appropriate comments for every operation performed.



Code Generation by PandasGUI

In the same edit menu, you have the option of exporting the modified dataframe. You can opt for that option or more easily, you can drag and drop the dataframe name to any folder and a CSV file for that dataframe will be created.



Conclusion

In this detailed article, I introduced you to PandasGUI. It is a very powerful tool to perform data manipulation and exploration via the graphical user interface. We started with the installation of the library, loading the dataset, then explored all the on-screen functions, looked at each tab in greater detail, and then saw some of the miscellaneous functions.

> **If you want to read/explore every article of mine, then head over to my [master article list] which gets updated every time I publish a new article on any platform!**

4. Medium—@kaustubhgupta1828

The media shown in this article are not owned by Analytics Vidhya and are used at the Author's discretion.

Kaustubh Gupta

Kaustubh Gupta is a skilled engineer with a B.Tech in Information Technology from Maharaja Agrasen Institute of Technology. With experience as a CS Analyst and Analyst Intern at Prodigal Technologies, Kaustubh excels in Python, SQL, Libraries, and various engineering tools. He has developed core components of product intent engines, created gold tables in Databricks, and built internal tools and dashboards using Streamlit and Tableau. Recognized as India's Top 5 Community Contributor 2023 by Analytics Vidhya, Kaustubh is also a prolific writer and mentor, contributing significantly to the tech community through speaking sessions and workshops.

Beginner | Libraries | Machine Learning | Pandas | Python | Python

Free Courses

Generative AI - A Way of Life

Explore Generative AI for beginners: create text and images, use top AI tools, learn practical skills, and ethics.

Getting Started with Large Language Models

Master Large Language Models (LLMs) with this course, offering clear guidance in NLP and model training made simple.

Building LLM Applications using Prompt Engineering

This free course guides you on building LLM apps, mastering prompt engineering, and developing chatbots with enterprise data.

Microsoft Excel: Formulas & Functions

Master MS Excel for data analysis with key formulas, functions, and LookUp tools in this comprehensive course.

Personalized GenAI Learning Path 2025✨ Crafted Just for YOU!

What are your thoughts?...

Submit reply

Karen McNicholas

I was diagnosed of Parkinson's Disease a couple of years ago, I had severe fatigue, difficulty with mobility and sleeping. I was given medications which helped but only for a short while. So i decided to try alternative measures and began on Parkinson's HERBAL TREATMENT from Kykuyu Health Clinic, It made a tremendous difference for me (Go to their website www. kykuyuhealthclinic. com). I had improved walking balance, muscle strength and improved vision

sumit kumar

Write for us

Write, captivate, and earn accolades and rewards for your work

- Reach a Global Audience
- Get Expert Feedback
- Build Your Brand & Audience

- Cash In on Your Knowledge
- Join a Thriving Community
- Level Up Your Data Science Game

Generative AI | Large Language Models | Building LLM Applications using Prompt Engineering | Building Your first RAG System using LlamaIndex | Stability.AI | MidJourney | Building Production Ready RAG systems using LlamaIndex | Building LLMs for Code | Deep Learning | Python | Microsoft Excel | Machine Learning | Decision Trees | Pandas for Data Analysis | Ensemble Learning | NLP | NLP using Deep Learning | Neural Networks | Loan Prediction Practice Problem | Time Series Forecasting | Tableau | Business Analytics

Popular Categories

Generative AI | Prompt Engineering | Generative AI Application | News | Technical Guides | AI Tools | Interview Preparation | Research Papers | Success Stories | Quiz | Use Cases | Listicles

Generative AI Tools and Techniques

GANs | VAEs | Transformers | StyleGAN | Pix2Pix | Autoencoders | GPT | BERT | Word2Vec | LSTM | Attention Mechanisms | Diffusion Models | LLMs | SLMs | StyleGAN | Encoder Decoder Models | Prompt Engineering | LangChain | LlamaIndex | RAG | Fine-tuning | LangChain AI Agent | Multimodal Models | RNNs | DCGAN | ProGAN | Text-to-Image Models | DDPM | Document Question Answering | Imagen | T5 (Text-to-Text Transfer Transformer) | Seq2seq Models | WaveNet | Attention Is All You Need (Transformer Architecture)

Popular GenAI Models

Llama 3.1 | Llama 3 | Llama 2 | GPT 4o Mini | GPT 4o | GPT 3 | Claude 3 Haiku | Claude 3.5 Sonnet | Phi 3.5 | Phi 3 | Mistral Large 2 | Mistral NeMo | Mistral-7b | Gemini 1.5 Pro | Gemini Flash 1.5 | Bedrock | Vertex AI | DALL.E | Midjourney | Stable Diffusion

Data Science Tools and Techniques