Group Project

Plotly

Without the need for JavaScript, Plotly is a powerful data visualization tool that can handle extremely sophisticated analytics and carry out automatic activities. Those that work with large data and researchers who want powerful analysis and AI will find its tools and capabilities suitable. Users may upload, analyze, and visualize an infinite amount of data and outcomes using pre-made templates and accepted formulae. It offers a graphing library that is open-source and receives more than four million monthly downloads. A major factor in the continual advancement of data science is seamless accessibility. Its features and tools are all scaleable. As it is utilized by students and multinational corporations such as Shell, Apple, Cisco, Tesla, and NVIDIA, its usage is virtually limitless. International businesses heavily use Plotly for processing Python models.

Plotly has many data analytical features at its disposal to visualize data sets. One of the main features of Plotly includes the ability to make plots that can visualize a dataset. Plotly works very well with data libraries such as Pandas and NumPy so the dataset can be visualized in a very professional-looking chart. Plotly can be used to create candlestick charts so traders can predict the price of a stock based on previous trends. Plotly also has the ability to create maps. An example of this is to visualize how many citizens of a certain area voted for candidates running in that area. After creating the charts, Plotly has a feature built-in where one can export the graph to an image that can be used in PowerPoint presentations. If images are not the preferred way to save the graphs, one can save the graph as a pdf using Plotly’s write\_image function and make the graph a pdf file. A key attribute of Plotly is that it is open-sourced software, ensuring that the software is free and anyone can view the source code of the software, report issues on it, and even contribute their code to the software so a certain bug or feature set is integrated by a developer.

For this project, we wanted to plot trend lines of data using the Plotly library to see if we could find any correlation between years. The data set we chose to use was the NYPD historic arrest data from the city of New York’s public data website. The years in this data set were 2006 - 2021. The problem we wanted to solve was whether there is a correlation between the number of arrests and a certain month in a given year. The chart type we chose to best display the data was a line chart because it made it the easiest to read and find correlations to the number of arrests in a given month between different years. It also points out years where the number of arrests was abnormally larger than others which could point out significant events in those years. This data set points out how major events whether it was national or internationally affected the data set. For example, at the end of 2019, you can see that there is a steep decline in arrests from December 2019 going into 2020 due to COVID-19. It is very evident in 2020 because the average number of arrests recorded in the whole year was about 14,000 arrests compared to 2019 where the average number of arrests per month hovered right at 20,000 arrests.

While developing the code to convert the data set into readable graphs we ran into multiple errors. The first error we ran into was creating data frames for months and years. When we tried to create the graphs, the library couldn’t plot the data because there were two data frames that had 2 different numbers of indexes. One data frame had an index of 12, while the other had an index of 16. To solve this error we decided to create two new columns in the original data frame one being the months and the other being the years. The second issue we ran into was trying to convert the months from numbers to names which we did successfully but they were sorted in alphabetical order instead of the order of the months. To fix this issue we categorized the month column in the order of the months and then called a sort values function on the Month column. The final two issues we had were after creating a new data frame by grouping by year and month. The issue was the new data frame we wanted to create was actually a series and there was no label for the arrests column. The solution was to call a function that resets the index of the series and converts it into a data frame and gives the index the name of Arrests.

After working on this project, we learned the implications of Plotly and how it can be used effectively in the data analytics field. After implementing the code in our project, we learned how easy it was to manipulate that data and plot the data into a line chart. We also learned that Plotly has two separate imports. One is an express import that makes importing data extremely easy, while the graph\_objects import is a bit more difficult to create plots, but one has more control of the layout of the graph.

Chart, line chart

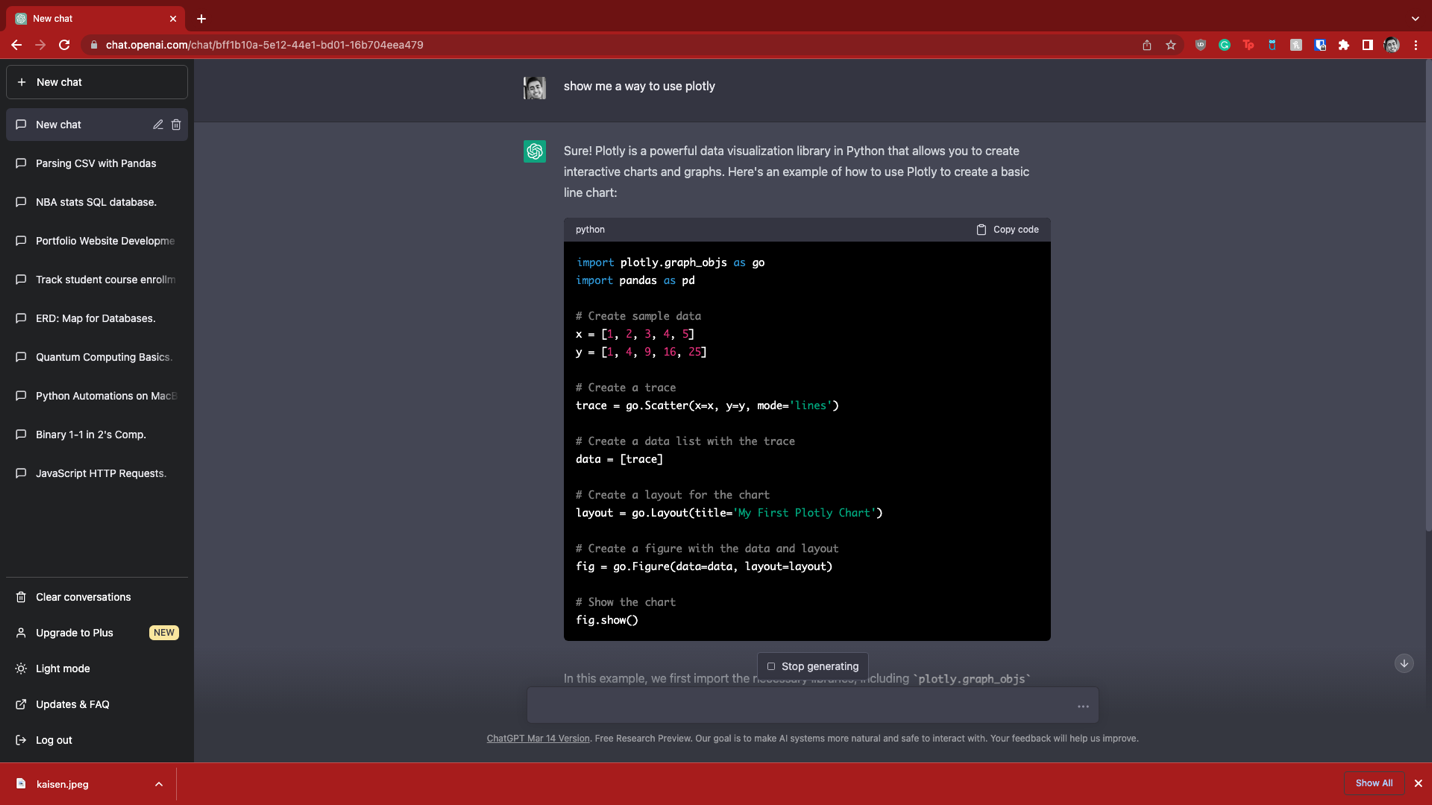
Description automatically generated

This picture is the output of calling Plotly on the Arrests data.

Graphical user interface, text

Description automatically generatedThe following pictures are me asking ChatGPT how to do some things such as parse a CSV file, count number of arrests, change date format, solution to month and year issue and example of Plotly. Graphical user interface, text, application

Description automatically generated

Text

Description automatically generated