Reflection Paper

Mashfika Jahan

Madeline Johnson

Yisihaq Yemiru

Working on the group project was a meaningful learning experience that helped us strengthen our technical skills and code effectively as a team. Our group decided to meet in person and do work together, which turned out to be extremely helpful as it helped us communicate better, solve problems more efficiently, and stay organized through each of our phases.

One of our earliest challenges was getting our GitHub set up. When we first created our repository, we forgot to add a ‘ .gitignore’ file. This caused a lot of issues as we had to add it manually later, which was way more complicated. Even after adding, we faced multiple issues as unnecessary files and checkpoints were being pushed into our repo. It created a new branch, which led to us having merge conflicts and confusion about which files were important. Eventually, after phase 2, we added Google Colab instead of Jupyter Notebook as it made it easier to share the notebook without worrying about the push and pull commands, merge conflicts, or version control issues. However, we switched back to Jupyter Notebook after we realized that we needed to show teamwork, version control and proper GitHub usage. This also allowed us to further familiarize ourselves with the traditional use of GitHub. We needed a clear display that would allow us to show collaboration through the notebook.

Throughout the project, we practiced loops, Python fundamentals, version control, file handling, and data cleaning. We used AI to help us understand and work through the phases. Reading, writing CSV files, creating a summary text file, cleaning the dataset, and working with visualization all helped us understand the data analysis workflow. We created a histogram, box plot, scatter plot, and bar chart, which helped strengthen our ability to interpret data visualization and explain our findings.

Something we learned near the end of our group project was how much of an impact color can make on data visualizations. When we first created the histogram for petal length, we had not thought of adding a 'groupby' feature to differentiate between the species, so it only showed the distribution of petal lengths in general. When we got to the last question that asked us to look at our visualizations to determine which flower feature was the strongest indicator of species, we decided to create a new histogram and color code by species. After doing this, our histogram was able to give us a much clearer picture of the data.

Overall, this project helped us sharpen our Python fundamental skills, work with a team, and use GitHub to collaborate on projects, which is essential in real-life settings. This project also taught us the importance of version control, good communication, and shared responsibility. All these skills will be beneficial to our future projects when we work with real-life data in professional environments.