Faiza Abdullah

L001: Intro to Jupyter and GitHub

ITAI 2373 Natural Language Processing (NLP)

Professor: Patricia Mcmanus

GitHub Repository Link: https://github.com/AbdullahFaiza/jupyter-exploration
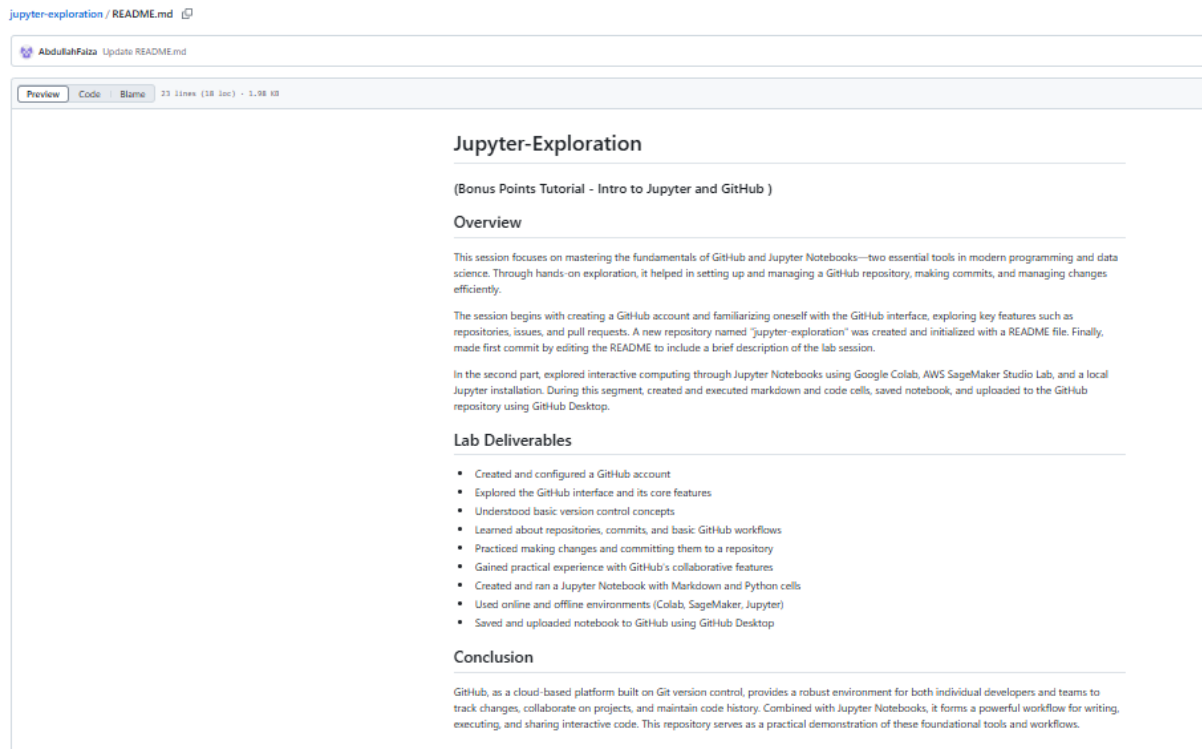
INTRODUCTION:

This lab session provided a structured opportunity to revisit and deepen my understanding of GitHub and Jupyter Notebooks, two indispensable tools in modern programming and data science. Having previously worked with these tools in a Deep Learning course under the same Professor, I approached this exercise with some familiarity. However, this session served as a valuable refresher, allowing me to focus on the foundational aspects of version control and interactive computing that I had not explored in depth previously. Through hands-on activities, I reinforced my skills in repository management, version control workflows, and interactive notebook creation, enhancing my confidence in integrating these tools into my academic and professional workflow.

## WHAT I DID:

GitHub Setup and Repository Management

Given my prior experience with GitHub, I began by accessing my existing account and navigating the platform's interface with ease. I created a new repository named **"jupyter-exploration"** as per the assignment instructions, initializing it with a README file. To align with the lab's objectives, I edited the README to include a concise description of the session's purpose and committed the changes directly to the main branch using GitHub's web interface. This task, while familiar, reminded me of the importance of clear commit messages and maintaining an organized repository structure.

Jupyter-Exploration

(Bonus Points Tutorial - Intro to Jupyter and GitHub)

Overview

This session focuses on mastering the fundamentals of GitHub and Jupyter Notebooks—two essential tools in modern programming and data science. Through hands-on exploration, it helped in setting up and managing a GitHub repository, making commits, and managing changes efficiently.

The session begins with creating a GitHub account and familiarizing oneself with the GitHub interface, exploring key features such as repositories, issues, and pull requests. A new repository named "jupyter-exploration" was created and initialized with a README file. Finally, made first commit by editing the README to include a brief description of the lab session.

In the second part, explored interactive computing through Jupyter Notebooks using Google Colab, AWS SageMaker Studio Lab, and a local Jupyter installation. During this segment, created and executed markdown and code cells, saved notebook, and uploaded to the GitHub repository using GitHub Desktop.

Lab Deliverables

- Created and configured a GitHub account
- Explored the GitHub interface and its core features
- Understood basic version control concepts
- Learned about repositories, commits, and basic GitHub workflows
- Practiced making changes and committing them to a repository
- Gained practical experience with GitHub's collaborative features
- Created and ran a Jupyter Notebook with Markdown and Python cells
- Used online and offline environments (Colab, SageMaker, Jupyter)
- Saved and uploaded notebook to GitHub using GitHub Desktop

Conclusion

GitHub, as a cloud-based platform built on Git version control, provides a robust environment for both individual developers and teams to track changes, collaborate on projects, and maintain code history. Combined with Jupyter Notebooks, it forms a powerful workflow for writing, executing, and sharing interactive code. This repository serves as a practical demonstration of these foundational tools and workflows.

To further integrate version control, I used GitHub Desktop to clone the repository to my local machine. This step allowed me to practice the workflow of cloning, modifying, committing, and pushing changes, reinforcing my understanding of Git's core functionality. I appreciated the streamlined interface of GitHub Desktop, which simplified the process compared to command-line Git operations I had used in my Deep Learning course.
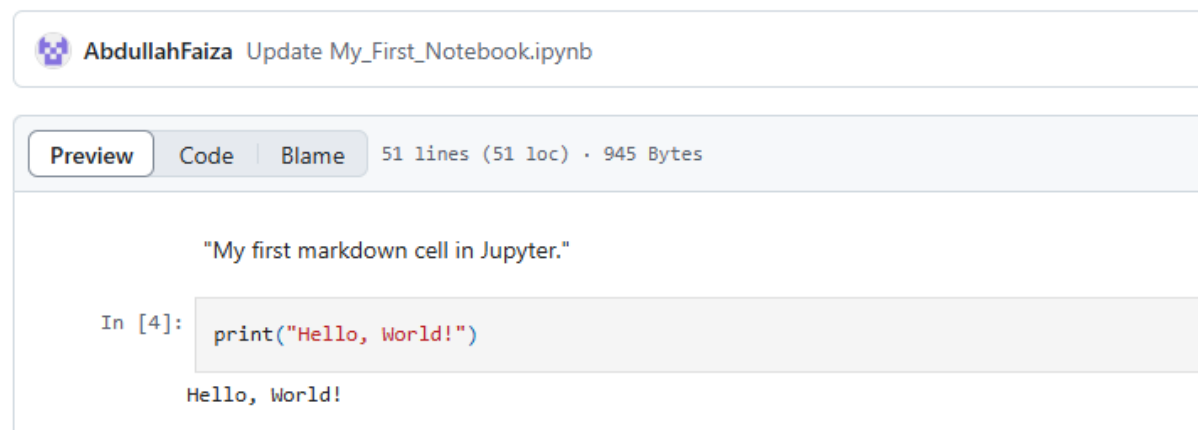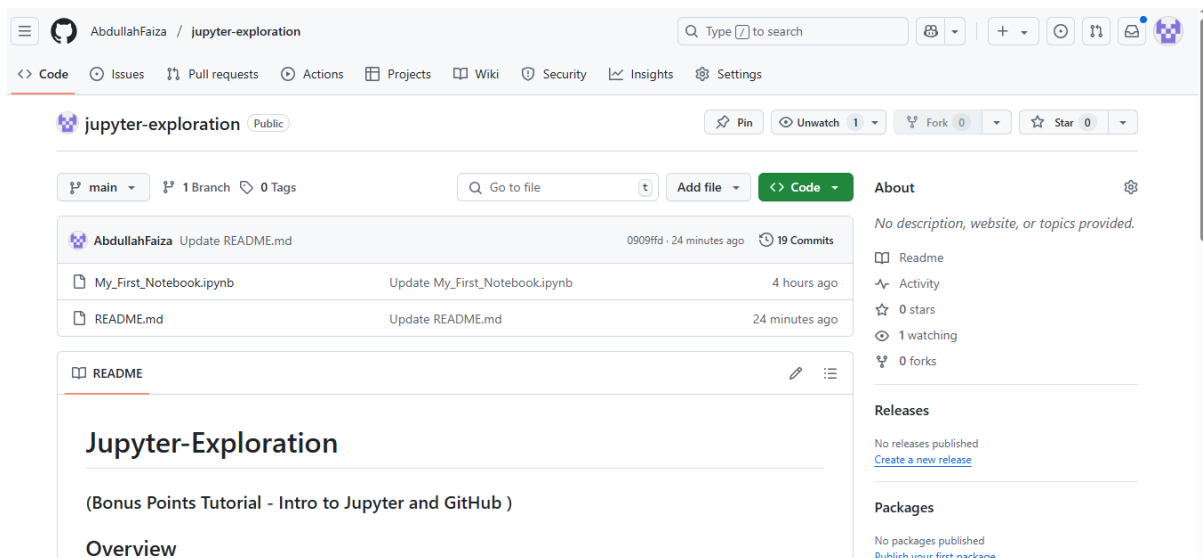
Jupyter Notebook Exploration

For the Jupyter Notebook component, I opted to work in multiple environments to compare their functionalities, leveraging my prior experience. I used Google Colab for its accessibility and AWS SageMaker Studio Lab for its robust cloud-based features, and I also ran a local Jupyter Notebook installation to contrast the offline experience. In each environment, I created a new notebook titled "My_First_Notebook.ipynb." I added a Markdown cell with a brief description of the lab and a code cell containing a simple Python script (print("Hello, World!")). Executing these cells and observing the outputs was straightforward, but it underscored the power of Jupyter's interactive environment for rapid prototyping and documentation.
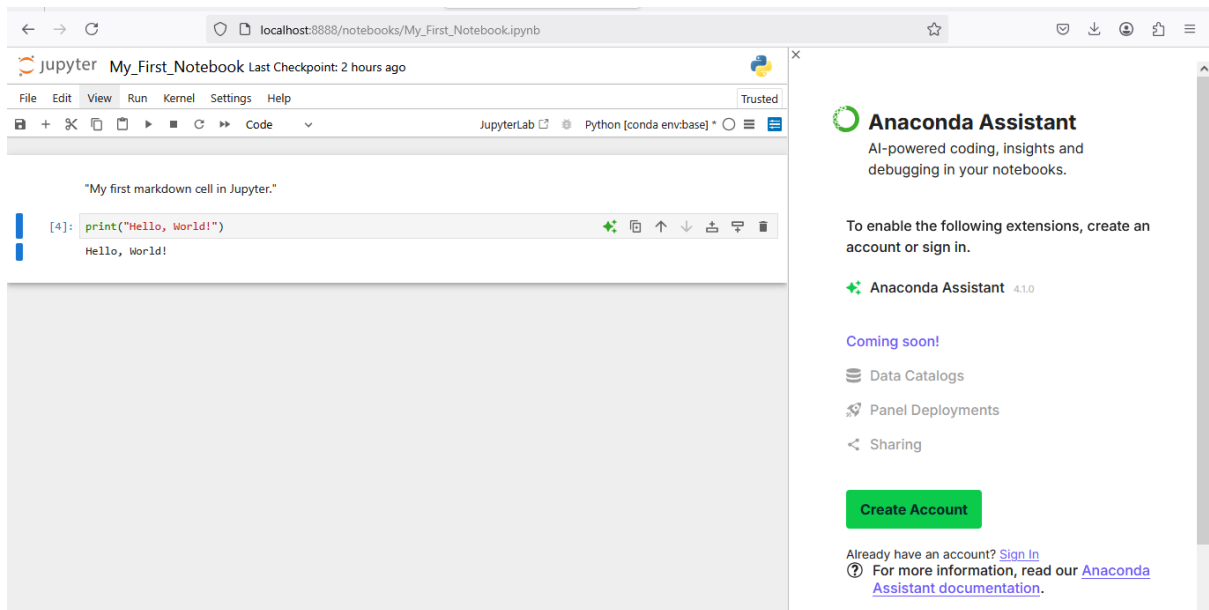
After completing the notebook, I saved it as a .ipynb file and uploaded it to my GitHub repository using GitHub Desktop. This involved cloning the repository locally, adding the notebook file, committing the changes and pushing the updates to the remote repository. This process reinforced the seamless integration of Jupyter Notebooks with GitHub for sharing and collaboration.

WHAT I LEARNED:

This lab session offered a structured refresher on GitHub and Jupyter Notebooks, emphasizing their foundational mechanics, which I had not fully appreciated in my previous Deep Learning course. In that course, my focus was primarily on implementing complex projects, and I used GitHub and Jupyter Notebooks as tools without delving into their workflows. This exercise highlighted the importance of version control for maintaining project organization and collaboration. I gained a deeper appreciation for GitHub's role in tracking changes systematically and enabling collaborative development through features like repositories and commits.

Working with Jupyter Notebooks across multiple environments (Google Colab, AWS SageMaker, and local installation) allowed me to compare their strengths. Google Colab's integration with Google Drive simplified sharing, while AWS SageMaker offered powerful computational resources suitable for larger projects. The local installation, though requiring setup, provided greater control over the environment. This exploration reinforced Jupyter's versatility for combining code, documentation, and visualizations in a single interface, which is particularly valuable for iterative tasks in data science and machine learning.

Challenges and Solutions

While I was comfortable with the tools, I encountered minor challenges in optimizing my workflow. For instance, syncing changes between online and local environments required careful attention to avoid overwriting files. I resolved this by maintaining a clear sequence of cloning, editing, and pushing changes using GitHub Desktop, which minimized confusion. Additionally, I initially overlooked the importance of detailed commit messages, but the lab's emphasis on documentation prompted me to adopt more descriptive summaries, enhancing clarity for future reference.

Questions or Comments

Question: What are the best practices for managing multiple Jupyter Notebooks in a single GitHub repository, especially when dealing with iterative versions or collaborative edits?

Areas for Further Exploration:

- GitHub Advanced Features: I am keen to explore branching strategies and resolving merge conflicts, as these were briefly touched upon in my previous course but seem critical for team-based projects.

- Jupyter Notebook Enhancements: I want to investigate Jupyter extensions, such as those for interactive visualizations or automated testing, to enhance notebook functionality.

- Integration with Other Tools: I am curious about integrating GitHub and Jupyter with CI/CD pipelines for automated testing and deployment of machine learning models.

Feedback:

The lab's structured approach made it an effective refresher, breaking down complex workflows into manageable steps. The flexibility to choose between online and local environments was particularly helpful, as it allowed me to tailor the experience to my

preferences. The hands-on nature of the tasks solidified my understanding, and I appreciated the opportunity to revisit foundational concepts with a fresh perspective.

CONCLUSION:

This lab session was a valuable opportunity to revisit and refine my skills with GitHub and Jupyter Notebooks. While I had prior exposure to these tools, the exercise provided a focused exploration of their fundamental workflows, which I had previously overlooked in favor of application-specific tasks in my Deep Learning course. By reinforcing version control principles and interactive computing techniques, this session equipped me with a stronger foundation for managing code and documentation effectively. Moving forward, I am excited to integrate these tools into my projects, leveraging their capabilities for collaboration, reproducibility, and efficient workflows in my academic and professional endeavors.

CITATIONS:

https://github.com/AbdullahFaiza/jupyter-exploration

https://colab.research.google.com/drive/1rQZ-sMozl782Kj1q4BiVl-xz9BhfLjOp#scrollTo=Fk-NRNi_X8ed

https://studiolab.sagemaker.aws/users/FaizaAbdullah

https://kvjyzhi5whzsjgk.studio.us-east-2.sagemaker.aws/studiolab/default/jupyter/lab/tree/sagemaker-studiolab-notebooks/My_First_Notebook.ipynb

http://localhost:8888/notebooks/My_First_Notebook.ipynb

https://www.youtube.com/watch?v=yjjE-MJD5TI

https://www.youtube.com/watch?v=ktN5tDfQ_g8

https://www.youtube.com/watch?v=Oaj3RBIoGFc

https://eagleonline.hccs.edu/courses/293189/files/75392961?module_item_id=20102496

https://eagleonline.hccs.edu/courses/293189/files/75392965?module_item_id=20102497

https://eagleonline.hccs.edu/courses/293189/files/75392963?module_item_id=20102502