MLEng Assignment 1: Conda Installation, Pandas Basics, and Data Visualization

Highlight Reel

1 Objectives

By the end of this assignment, you will be able to:

- Install Conda on your operating system

- Set up a Conda environment using a provided YAML file

- Read and manipulate data from a CSV file using pandas

- Create basic visualizations using matplotlib

2 Seeking Help and Collaboration

Throughout this assignment and the course, you may encounter challenges or have questions. Here are some resources for seeking help:

1. **Language Models:** Try asking AI language models like ChatGPT or Claude for help with coding problems, explanations of concepts, or debugging assistance. Remember to:

   - Clearly describe your problem or question
   - Provide relevant code snippets or error messages
   - Be specific about what you've already tried
   - Verify any solutions or explanations provided

2. **Course Discord:** Our course Discord is a great place to:

   - Ask questions and get help from your instructor and peers
   - Share interesting findings or visualizations
   - Collaborate on problem-solving

3. **Official Documentation:** Always refer to the official documentation for Conda, Pandas, and Matplotlib. These resources provide detailed information and examples.

Remember, seeking help and collaborating are important skills in machine learning engineering. Don't hesitate to use these resources, but also challenge yourself to solve problems independently.

3 Part 1: Installing Conda

Follow the instructions for your operating system:

3.1 For Windows

1. Download the Conda installer for Windows from `https://docs.conda.io/en/latest/miniconda.html`

2. Run the installer and follow the prompts

3. Open the Anaconda Prompt from the Start menu to use Conda

3.2 For macOS

1. Open Terminal

2. Run the following commands:

```
curl -O https://repo.anaconda.com/miniconda/Miniconda3-latest-MacOSX-arm64.sh
bash Miniconda3-latest-MacOSX-x86_64.sh
```

3. Follow the prompts to complete the installation

4. Restart your terminal or run `source ~/.bash_profile`

3.3 For Linux

1. Open a terminal window

2. Run the following commands:

```
wget https://repo.anaconda.com/miniconda/Miniconda3-latest-Linux-x86_64.sh
bash Miniconda3-latest-Linux-x86_64.sh
```

3. Follow the prompts to complete the installation

4. Restart your terminal or run `source ~/.bashrc`

4 Part 2: Setting Up the Conda Environment

1. Download the provided `environment.yaml` file

2. Open a terminal (Anaconda Prompt on Windows)

3. Navigate to the directory containing `environment.yaml`

4. Run the following command:

```
conda env create -f environment.yaml
```

5. Activate the new environment:

```
conda activate mleng_env
```

5 Part 3: Reading a CSV with Pandas and Visualizing with Matplotlib

1. Create a new Python file named `data_analysis.py`

2. Add the following code to the file:

```python
import pandas as pd
import matplotlib.pyplot as plt

# Read the CSV file
df = pd.read_csv('provided_data.csv')

# Display the first 5 rows
print(df.head())

# Display basic information about the dataset
print(df.info())

# Calculate and print summary statistics
print(df.describe())

# Plotting
plt.figure(figsize=(10, 6))
plt.plot(df.iloc[:, 0], df.iloc[:, 1])
plt.xlabel('Frame Number')
plt.ylabel('Value')
plt.title('Second Column vs Frame Number')
plt.grid(True)
plt.savefig('plot.png')
plt.show()
```

3. Save the file and run it using:

```
python data_analysis.py
```

6 Part 4: Experimenting with Visualizations

Now that you've created a basic plot, experiment with different types of visualizations using matplotlib. Try the following:

1. Create a scatter plot instead of a line plot

2. Plot multiple columns on the same graph

3. Create a histogram of one of the columns

4. Use subplots to display multiple visualizations in one figure

Be creative and try to find interesting patterns or relationships in the data!

7 Submission

Post to our Discord:

- Screenshots showing successful Conda installation and environment creation

- The output from running `data_analysis.py`

- The plot generated by the script (`plot.png`)

- At least one additional visualization you created while experimenting

- Any challenges you faced during the assignment and how you overcame them

- A brief explanation (2-3 sentences) of what you learned from visualizing the data