Experiment Tracking with MLflow

Objective:

Set up and use MLflow to track machine learning experiments, compare model performances, and manage experiment metadata.

1. Clone your repository.
2. Initialize the project structure like this:

```
├── data/
├── notebooks/
├── scripts/
├── mlruns/                  # MLflow logs and experiment metadata
├── requirements.txt
├── train.py                 # Main training script
├── README.md
```

3. **Setup MLflow**: Install the necessary packages by creating a `requirements.txt` file.

```
pandas
scikit-learn
mlflow
```

4. Then, install them in your virtual environment:

```
pip install -r requirement.txt
```

5. **Train Two Models (Linear Regression and Random Forest)**: Below is a simplified version of the `train.py` script for training and logging experiments:
6. **Run Experiments and Track with MLflow**: Launch the MLflow UI in your local environment:

```
mlflow ui
```

7. Run the training script:

```
python train.py
```

You should now be able to see the logs of different experiments in the MLflow UI at `http://127.0.0.1:5000`.

8. **Compare Models**: Use the MLflow UI to compare the `mse` (Mean Squared Error) of the `Linear_Regression` and `Random_Forest` models. Screenshots of the comparison and tracked metadata should be included in your report.
9. **Save Model**: The best-performing model is saved and logged in MLflow's Model Registry. Use MLflow's `log_model()` function to version models.

To be submitted:

- **GitHub Repository**: Link to the repository.
- **Report**: Comparisons of `mse` values between models, screenshots of the MLflow UI.