mehmetimga / **data_analysis_pipeline_practice**

Code Issues Pull requests Actions Projects Wiki Security Insights Settir

View license

0 stars **0** forks **0** watching Branches Activity
Tags

Public repository · Generated from skysheng7/ia4

1 Branch **0** Tags Go to file t Go to file Add file Code

mehmetimga adding repo url f9c95b2 · 1 minute ago

.github/workflows	Initial commit	20 hours ago
data	Initial commit	20 hours ago
docs	Fix image paths for GitHub Pages	8 minutes ago
report	checkpoint for github pages	12 minutes ago
results	checkpoint for github pages	12 minutes ago
scripts	Initial commit	20 hours ago
.gitignore	Initial commit	20 hours ago
Dockerfile	Initial commit	20 hours ago
LICENSE.md	Initial commit	20 hours ago
Makefile	Fix image paths for GitHub Pages	8 minutes ago
README.md	adding repo url	1 minute ago
conda-linux-64.lock	Initial commit	20 hours ago
conda-lock.yml	Initial commit	20 hours ago
conda-osx-64.lock	Initial commit	20 hours ago
conda-osx-arm64.lock	Initial commit	20 hours ago
conda-win-64.lock	Initial commit	20 hours ago
docker-compose.yml	Initial commit	20 hours ago
environment.yml	Initial commit	20 hours ago
mehmetimga_data_analysis_...	adding pdf	4 minutes ago

runall.sh	Initial commit	20 hours ago

README License

Building a Data Analysis pipeline tutorial

adapted from Software Carpentry

This example data analysis project analyzes the word count for all words in 4 novels. It reports the top 10 most occurring words in each book in a report.

Links

- **GitHub Repository:** https://github.com/mehmetimga/data_analysis_pipeline_practice
- **Report (GitHub Pages):** https://mehmetimga.github.io/data_analysis_pipeline_practice/

Recreate the computational environment

1. Clone repo

Clone this repo, and using the command line, navigate to the root of this project.

```
git clone git@github.com:mehmetimga/data_analysis_pipeline_practice.git
cd data_analysis_pipeline_practice
```

2. Recreate the computational environment

▶ **Option 1: Use `conda-lock.yml`**

▶ **Option 2: Use `environment.yml`**

▶ **Option 3: Use `docker-compose.yml`**

Exercise:

Your task is to add a "smarter" data analysis pipeline using GNU Make! It should accomplish the same task as `bash runall.sh` when you type `make all`.

It should reset the analysis the starting point (the state when you first copied this repo) when you type `make clean`.

Depenedencies

- GNU Make
- Quarto
- Python & Python libraries:
  - `click`
  - `matplotlib`
  - `pandas`

Releases

No releases published
Create a new release

Packages

No packages published
Publish your first package

Deployments 4

github-pages 3 minutes ago

+ 3 deployments

Languages

JavaScript 66.7% **HTML** 19.0% **Python** 10.3% **Makefile** 2.3% **Shell** 1.1% **Dockerfile** 0.6%