

Astroinformatics I

Project Presentation

José Batista-Mendoza

CITEVA, Universidad de Antofagasta

July 10, 2025

Table of Contents

- 1 **Introduction**
- 2 **Practices 1 & 2: Shell, Linux & Python Fundamentals**
- 3 **Practice 3: Python Programming for Astronomical Data Analysis**
- 4 **Practice 4: Software Project Management & Quality**
- 5 **Conclusions**

Introduction

Motivation:

- In modern astronomy, theoretical knowledge must be coupled with strong computational skills.
- Astroinformatics is precisely this intersection: applying informatics tools to astronomical data.
- Bridge the gap between classroom concepts and their tangible application in real-world data handling and analysis.
- Show how the fundamental computational principles provided in the course become powerful instruments for astronomical discovery.

Introduction (Cont.)

Course Overview:

- Astroinformatics I covered a comprehensive range of essential computational topics.
- From foundational Linux command-line operations and shell for data management,
- Through core Python programming for data manipulation and analysis,
- To specialized astronomical libraries (`astropy`, `pandas`, `numpy`, `matplotlib`) for specific data challenges.
- Finally, integrating crucial software engineering principles for robust and reproducible work.
- The course emphasized hands-on learning through lectures, tutorials, and graded practices.

Introduction (Cont.)

Presentation Goal:

- The primary goal is to illustrate how the theoretical and practical concepts from Lectures and Tutorials were directly applied.
- This presentation will walk through key examples from the **graded practices (Practices 1–4)**.
- This will highlight the practical implementation of tasks like data cleaning, transformation, analysis, visualization, and project organization.
- Ultimately, showcasing a complete workflow for astronomical data processing.

Shell, Linux & Python Fundamentals

Foundation: Based on Lectures and Tutorials 2–5

Key Concepts Applied:

- **File System Navigation & Manipulation**

- Identifying files (`ls`, wildcards like `*.csv`).
- Directory management (`dirname`, variable expansion).
- File creation (`touch`).

- **Text Processing & Data Transformation**

- Changing delimiters (`sed s/,/ /g`).
- Removing columns (`awk '{print $1, $2, $3}'`).
- File renaming (`mv`).

Shell, Linux & Python Fundamentals (Cont.)

Foundation: Based on Lectures and Tutorials 2–5

Key Concepts Applied:

● **Shell Scripting Constructs**

- for loops for batch processing.
- Conditional statements (`[-e "$csv_file"], if`).
- Output Redirection (`>>`).

● **Python Fundamentals**

- Defining and calling functions.
- Performing basic operations, comparisons, data type conversion and string manipulation.
- Applying conditional logic (`if-elif-else`).
- Handling command-line arguments (`sys.argv`).
- Incorporating user interaction (`input()`).
- Implementing error handling (`try-except` blocks).

Practice Showcase

● Practice 1: Listing CSV filenames

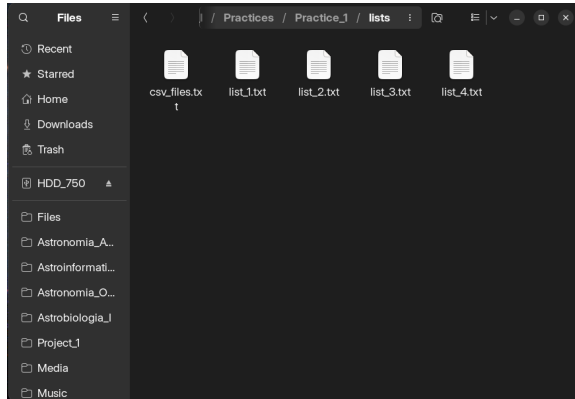
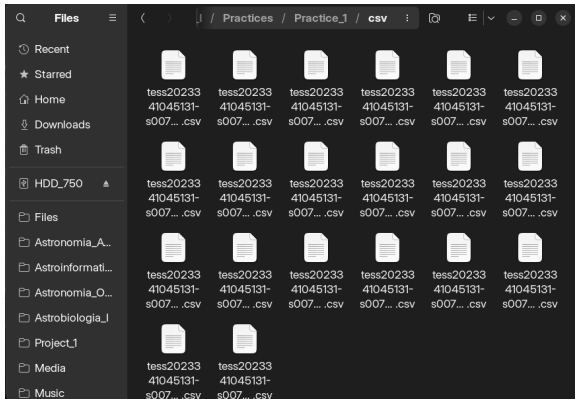
```
joe@joesminipc: ~/Files/Studies/MSc/Semester_1/Astroinformatica_I/P...
joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_1
/codes$ sudo sh list_csv_files.sh
CSV files saved to /csv_files.txt
joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_1
/codes$ sudo sh split_csv_list.sh
csv_files.txt has been split into 4 files.
joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_1
/codes$
```

```
Open ~/Files/Studies/MSc/Semester_1_matica_I/Practices/Practice_1/lists
Ln 1, Col 1

1 tess2023341045131-s0073-0000000001750268-0268-s_lc.csv
2 tess2023341045131-s0073-0000000001755406-0268-s_lc.csv
3 tess2023341045131-s0073-0000000001827744-0268-s_lc.csv
4 tess2023341045131-s0073-0000000001828620-0268-s_lc.csv
5 tess2023341045131-s0073-0000000001840666-0268-s_lc.csv
6 tess2023341045131-s0073-0000000001942416-0268-s_lc.csv
7 tess2023341045131-s0073-0000000001942969-0268-s_lc.csv
8 tess2023341045131-s0073-0000000001947463-0268-s_lc.csv
9 tess2023341045131-s0073-0000000001950736-0268-s_lc.csv
10 tess2023341045131-s0073-0000000001950893-0268-s_lc.csv
11 tess2023341045131-s0073-0000000002006984-0268-s_lc.csv
12 tess2023341045131-s0073-0000000002008765-0268-s_lc.csv
13 tess2023341045131-s0073-0000000002014191-0268-s_lc.csv
14 tess2023341045131-s0073-0000000002102329-0268-s_lc.csv
15 tess2023341045131-s0073-0000000002104696-0268-s_lc.csv
16 tess2023341045131-s0073-0000000002105589-0268-s_lc.csv
17 tess2023341045131-s0073-0000000002143575-0268-s_lc.csv
18 tess2023341045131-s0073-0000000002149979-0268-s_lc.csv
19 tess2023341045131-s0073-0000000002152411-0268-s_lc.csv
20 tess2023341045131-s0073-0000000002234692-0268-s_lc.csv
```


Practice Showcase (Cont.)

● Practice 1: Splitting text files



Practice Showcase (Cont.)

- **Practice 2:** Modifying CSV files (delimiter change, column removal, and extension change)

```
joe@joesminipc: ~/Files/Studies/MSc/Semester_1/Astroinformatica_I/P...
joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/codes$ sudo sh task_1.sh
Starting modification of files in '/home/joe/Nextcloud/joe-new/files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/csv'...
1. Changing delimiter from ',' to '...'
   Delimiter change complete.
2. Changing file extension from '.csv' to '.lc'...
   Extension change complete.
3. Removing extra columns (keeping 'TIME', 'PDCSAP_FLUX', and 'PDCSAP_FLUX_ERR')...
   Column removal complete.
All specified modifications have been applied to the files.
joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/codes$
```

```
Open  tess2023341045131-s0073-0000000001750268-...  Ln 1, Col 1
~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/csv

1 TIME,TIMECORR,CADENCENO,SAP_FLUX,SAP_FLUX_ERR,SAP_BKG,SAP_BKG_ERR,PDCSAP_FLUX,
PDCSAP_FLUX_ERR,QUALITY,PSF_CENTR1,PSF_CENTR1_ERR,PSF_CENTR2,PSF_CENTR2_ERR,M-
OM_CENTR1,MOM_CENTR1_ERR,MOM_CENTR2,MOM_CENTR2_ERR,POS_CORR1,POS_CORR2
2 3285.7989573595028,0.005495878,1482004,,,,,,,,168,,,,,,,,
3 3285.8003462524134,0.0054958826,1482005,,,,,,,,32,,,,,,,,
4 3285.801735145324,0.0054958872,1482006,,,,,,,,32,,,,,,,,
5 3285.8031240382356,0.005495892,1482007,,,,,,,,32,,,,,,,,
6 3285.804512931147,0.0054958966,1482008,2682.9397,7.219023,1142.6876,2.7112477,
2078.9084,11.244134,0,,,,,905.2719761768637,0.0012244291,979.8425420784771,0.0
019117142,0.012102187,0.13282996
7 3285.805901823592,0.0054959008,1482009,2686.4644,7.2265897,1149.2695,2.715973,
2083.7026,11.255919,0,,,,,905.2716320661049,0.0012238288,979.8406725674315,0.0
019090768,0.014106098,0.12953551
8 3285.8072907165038,0.0054959054,1482010,2675.366,7.2166214,1146.6066,2.7137868
,2065.4507,11.240394,0,,,,,905.2724933426068,0.00122752,979.8406864188747,0.00
1916577,0.014835117,0.12700012
9 3285.8086796094153,0.00549591,1482011,2696.2197,7.2292824,1143.683,2.7122424,2
098.1396,11.260113,0,,,,,905.2721938803626,0.0012198925,979.8402739550324,0.00
19031566,0.014181586,0.12540044
10 3285.810068502326,0.0054959147,1482012,2679.299,7.2180204,1141.5579,2.709709,2
070.363,11.242573,0,,,,,905.2704813733113,0.0012250078,979.844814968241,0.0019
112825,0.013755005,0.12885012
```

Practice Showcase (Cont.)

● Practice 2: Modifying CSV files (Cont.) and basic Python scripting

```
Open ▾  tess2023341045131-s0073-0000000001750268-... Ln 1, Col 1
~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/csv

1 TIME PDCSAP_FLUX PDCSAP_FLUX_ERR
2 3285.7989573595028
3 3285.8003462524134
4 3285.801735145324
5 3285.8031240382356
6 3285.804512931147 2078.9084 11.244134
7 3285.805901823592 2083.7026 11.255919
8 3285.8072907165038 2065.4507 11.240394
9 3285.8086796094153 2098.1396 11.260113
10 3285.810068502326 2070.363 11.242573
11 3285.8114573952375 2082.0286 11.252562
12 3285.812846288149 2079.1445 11.240101
13 3285.814235180594 2065.8347 11.235728
14 3285.815624073505 2084.989 11.237153
15 3285.8170129664168 2074.002 11.242372
16 3285.8184018593283 2076.244 11.235358
17 3285.819790752239 2055.999 11.213792
18 3285.8211796451506 2085.9194 11.243473
19 3285.822568538062 2100.8157 11.248517
20 3285.8239574305076 2098.8606 11.244343
21 3285.825346323419 2110.7817 11.246265
22 3285.8267352163302 2102.7793 11.248538
```

```
Joe@joesminipc: ~/Files/Studies/MSc/Semester_1/Astroinformatica_I/P...
joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/codes$ source /home/joe/Files/Studies/MSc/Semester_1/Astroinformatica_I/.venv-linux/bin/activate
(.venv-linux) joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/codes$ python3 task_3.py
-----# Julian Date Calculator #-----

Enter the date in the 'd/m/y' format:
Date: 08/06/2008

Your date is 08/06/2008 (8 June, 2008). Is this correct? (Y/n) Y

The Julian date for 08/06/2008 is 2454625
(.venv-linux) joe@joesminipc:~/Files/Studies/MSc/Semester_1/Astroinformatica_I/Practices/Practice_2/codes$
```

Python for Astronomical Data Analysis

Foundation: Based on Lectures and Tutorials 6–7

Key Concepts Applied:

- **Data Manipulation (pandas, numpy)**

- Loading and structuring data with `pandas.DataFrame`.
- Numerical operations and statistical analysis (numpy for mean, std, median, etc.).

- **Astronomical Data Handling (astropy)**

- `astropy.timeseries.TimeSeries` for light curves.
- Time conversion (`astropy.time.Time`, BTJD to JD TDB, BJD to UTC).
- Lomb-Scargle periodograms (`astropy.timeseries.LombScargle`).

Python for Astronomical Data Analysis (Cont.)

Foundation: Based on Lectures and Tutorials 6–7

Key Concepts Applied:

- **Data Visualization (`matplotlib`)**

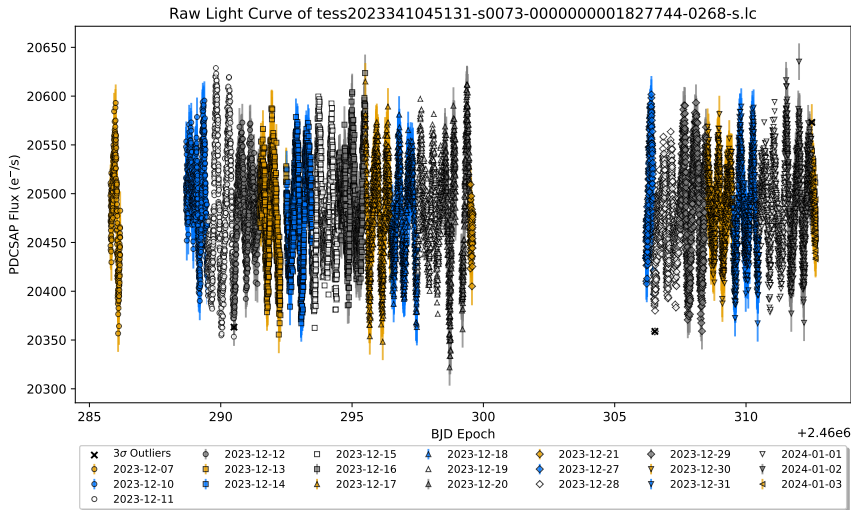
- Generating scatter plots with error bars.
- Plot customization (titles, labels, legends, colorblind-friendly palettes, markers).
- Subplots for combined visualizations (light curve + periodogram).
- Annotations on plots.

- **Astronomical Concepts:**

- Outlier detection, phase-folding and period finding for TESS light curves.

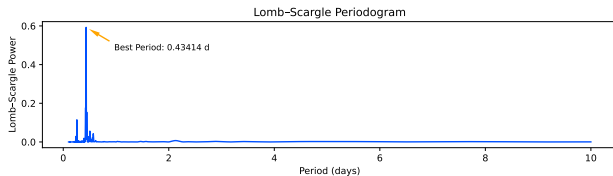
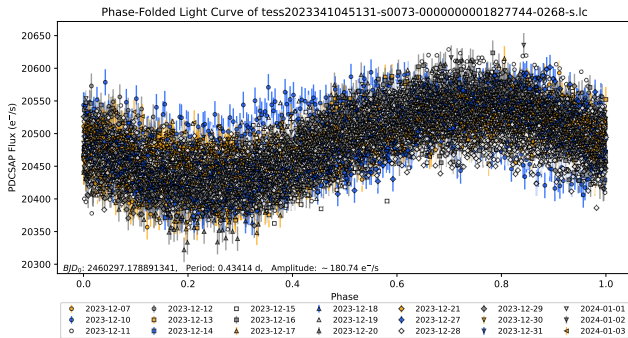
Practice Showcase

● **Practice 3:** Processing TESS light curves (raw plotting and outlier detection)



Practice Showcase (Cont.)

● Practice 3: Processing TESS light curves (phase-folding and period finding)



Software Project Management & Quality

Foundation: Based on Lecture and Tutorial 8

Key Concepts Applied:

● **Version Control (Git & GitHub)**

- Creating and structuring GitHub repositories.
- Organizing project files (/codes, /latex, /csv, etc.).
- Implied Git workflow (`git add`, `git commit`, `git push`).

● **Documentation**

- Importance of basic documentation for code and workflows.
- Role of README.md for project overview.

● **Software Testing**

- Identifying test cases for functions (e.g., `get_lc_data`, `fold_lc`).
- Considering test scenarios (valid data, missing data, known periodicity, no periodicity).

Practice Showcase

● Practice 4: Managing a project (GitHub repository set up)

The screenshot shows a GitHub repository named 'Astroinformatics_I' by user 'josebatistam'. The repository is public and has 0 forks, 0 stars, and 0 watchers. The main branch is 'main'. The repository contains a README.md file and a directory structure with four subdirectories: Practice_1, Practice_2, Practice_3, and Practice_4. The README.md file is open, showing the title 'Astroinformatics I' and a description of the repository's purpose. It also includes a 'Repository Structure' section with a list of files and directories.

Repository Structure

- The **root directory** contains this main `README.md` file and acts as the entry point for the entire collection of practices.
- Individual graded practices** are organized within their own dedicated subdirectories, named `/Practice_1/`, `/Practice_2/`, `/Practice_3/`, and `/Practice_4/`.

Repository Content:

File/Directory	Commit Message	Time Ago
Practice_1	Update split_csv_list.sh	39 minutes ago
Practice_2	Add files via upload	26 minutes ago
Practice_3	Update README.md	8 hours ago
Practice_4	Rename Practice_4/latex/Practice_4.pdf to Practic...	6 hours ago
README.md	Update README.md	6 hours ago

Repository Description:

This repository serves as a centralized collection point for all graded practices and the project presentation submitted as part of the Astroinformatics I course (Semester 1, 2025) at Universidad de Antofagasta. It aims to maintain a well-organized and version-controlled record of all assignments.

Repository Structure:

The repository's layout is designed for clarity and maintainability:

- The **root directory** contains this main `README.md` file and acts as the entry point for the entire collection of practices.
- Individual graded practices** are organized within their own dedicated subdirectories, named `/Practice_1/`, `/Practice_2/`, `/Practice_3/`, and `/Practice_4/`.

Repository Metadata:

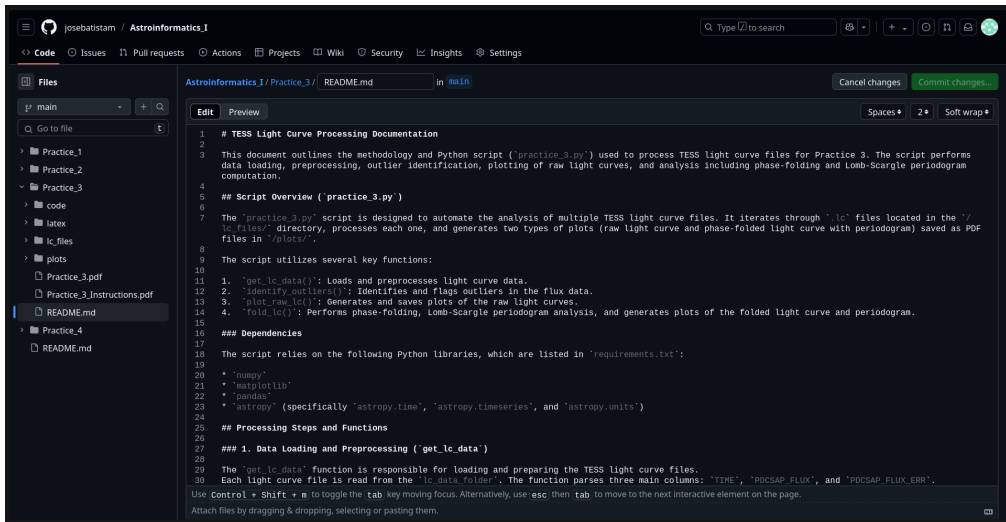
- Repository: Astroinformatics_I (Public)
- Branch: main (1 Branch)
- Tags: 0 Tags
- Commits: 49 Commits
- Language: Shell (96.2%), TeX (3.1%), Python (0.7%)

Repository Actions:

- Pin
- Watch (0)
- Fork (0)
- Star (0)
- Add file
- Code
- About
- Readme
- Activity
- Stars
- Watching
- Forks
- Releases
- Packages
- Languages
- Suggested workflows

Practice Showcase (Cont.)

● Practice 4: Managing a project (GitHub documentation creation)



The screenshot displays the GitHub web interface for a repository named 'Astroinformatics_I' by user 'josebatistam'. The left sidebar shows the file tree with folders 'Practice_1', 'Practice_2', 'Practice_3', and 'Practice_4'. Under 'Practice_3', there are files 'code', 'latex', 'lc_files', 'plots', 'Practice_3.pdf', 'Practice_3_Instructions.pdf', and 'README.md' (which is selected). The main area shows the content of 'README.md' in the 'main' branch. The document is titled '# TESS Light Curve Processing Documentation' and provides an overview of a Python script 'practice_3.py' used for processing TESS light curve data. It lists dependencies (numpy, matplotlib, pandas, astropy) and processing steps, including data loading and preprocessing. The bottom of the page includes instructions on how to toggle the tab key and attach files.

```
1 # TESS Light Curve Processing Documentation
2
3 This document outlines the methodology and Python script ('practice_3.py') used to process TESS light curve files for Practice 3. The script performs
4 data loading, preprocessing, outlier identification, plotting of raw light curves, and analysis including phase-folding and Lomb-Scargle periodogram
5 computation.
6
7 ## Script Overview ('practice_3.py')
8
9 The 'practice_3.py' script is designed to automate the analysis of multiple TESS light curve files. It iterates through '.lc' files located in the '/'
10 lc_files/' directory, processes each one, and generates two types of plots (raw light curve and phase-folded light curve with periodogram) saved as PDF
11 files in '/plots/'.
12
13 The script utilizes several key functions:
14
15 1. 'get_lc_data()': Loads and preprocesses light curve data.
16 2. 'identify_outliers()': Identifies and flags outliers in the flux data.
17 3. 'plot_raw_lc()': Generates and saves plots of the raw light curves.
18 4. 'fold_lc()': Performs phase-folding, Lomb-Scargle periodogram analysis, and generates plots of the folded light curve and periodogram.
19
20 ### Dependencies
21
22 The script relies on the following Python libraries, which are listed in 'requirements.txt':
23
24 * 'numpy'
25 * 'matplotlib'
26 * 'pandas'
27 * 'astropy' (specifically 'astropy.time', 'astropy.timeseries', and 'astropy.units')
28
29 ## Processing Steps and Functions
30
31 ### 1. Data Loading and Preprocessing ('get_lc_data')
32
33 The 'get_lc_data' function is responsible for loading and preparing the TESS light curve files.
34 Each light curve file is read from the 'lc_data_folder'. The function parses three main columns: 'TIME', 'PDCSAP_FLUX', and 'PDCSAP_FLUX_ERR'.
```

Use Control + Shift + m to toggle the tab key moving focus. Alternatively, use esc then tab to move to the next interactive element on the page.

Attach files by dragging & dropping, selecting or pasting them.

Conclusions

- The course succeeds in reinforcing the **interconnectedness** of seemingly disparate tools and principles.
- Linux/Shell provides the essential command-line foundation for data wrangling, file management, and automating initial steps. It's the 'glue' that orchestrates tasks.
- Python serves as the primary computational engine, allowing for sophisticated data analysis, algorithm implementation, and statistical modeling.

Conclusions (Cont.)

- Astronomical Libraries (Astropy, Lightkurve, etc.) bridge general-purpose Python with domain-specific astronomical knowledge, providing robust tools for units, time systems, and specialized analysis (e.g., periodograms, phase-folding).
- Software Engineering Principles (Version Control with Git/GitHub, Documentation, Testing) are paramount for ensuring that the code is **reproducible, maintainable, scalable, and collaborative**. They transform individual scripts into robust scientific tools.
- Together, these elements form a powerful and synergistic toolkit, essential for navigating the complexities of modern astronomical data.