

Analysis Script

This script, `analysis.py`, processes JSON files containing job data to identify and analyze duplicate job postings.

How to Use

1. **Place JSON Files:** Ensure all JSON files containing job data are in the same directory as `analysis.py`.
2. **Run the Script:** Execute the script using Python:

```
python analysis.py
```

What It Does

- **Loads JSON Files:** Reads all JSON files in the directory.
- **Extracts Job Data:** Collects job information from each file.
- **Identifies Duplicates:** Finds duplicate job postings based on `job_id`.
- **Decodes Filenames:** Converts filenames to readable dates and times.
- **Prints Analysis:** Outputs the total number of jobs, duplicates, unique job titles, employers, and locations. Also lists duplicate job titles with their occurrence counts and dates.

Output Example

```
Total jobs: 173
Duplicate jobs: 8
Unique job titles: 159
Unique employers: 130
Unique locations: 55
Duplicate job titles and their occurrences:
Title: Large Commercial Canada Intern, Count: 2, Dates: ['2024-11-27 15:45:20',
'2024-12-03 15:15:55']
Title: Software Engineer Intern (Toronto) - Summer 2025, Count: 2, Dates: ['2024-
11-27 15:45:20', '2024-12-03 15:15:55']
Title: Experiential Education Ambassador/Coordinator (Internship/ Coop), Count: 2,
Dates: ['2024-11-29 06:07:27', '2024-12-02 01:48:05']
Title: Software Engineer Intern (Toronto) - Spring 2025, Count: 2, Dates: ['2024-
11-29 06:07:27', '2024-11-30 15:48:07']
Title: web marketing internship, Count: 3, Dates: ['2024-11-30 15:48:07', '2024-
12-01 06:16:13', '2024-12-02 01:48:05']
Title: Social Media & Marketing Intern, Count: 2, Dates: ['2024-12-01 06:16:13',
'2024-12-02 01:48:05']
Title: Frontend Developer Intern - Winter 2025 Semester (January-April, Remote -
Canada), Count: 2, Dates: ['2024-12-01 06:16:13', '2024-12-02 01:48:05']
Title: Field Compliance Summer Associate - Internship. Job in Saskatoon
LilyLifestyle Jobs, Count: 2, Dates: ['2024-12-02 01:48:05', '2024-12-02
01:48:05']
```

For more details, refer to the script [analysis.py](#).