

Exercises on CSV/JSON/XML

Tiziano Fagni
BDSCC course - 2024/2025

Ex1

Given the input file 'lego_sets.jsonl' in JSONL format, extract all records that match these conditions:

- Lego sets included in date [2000, 2010]
- Number of pieces in the set ≥ 70

Write the output to another JSONL file.

Use the library **json** to process data.

Ex2

Given the input file 'lego_sets.xml' in XML format, extract all records that match these conditions:

- Lego set including 'station' in its name
- Theme is "Duplo"

Write the output to another XML file.

Use the library **BeautifulSoup4** to process data.

Ex3

Given the input CSV file 'lego_sets.csv', extract all records that match these conditions:

- Lego sets included in date [2000, 2010]
- Number of pieces in the set ≥ 70

Write the output to three different files: a JSON file, a JSONL file, and an XML file.

Use the libraries **csv**, **json** and **BeautifulSoup4** to process data.

Ex4

Given the input CSV file 'lego_sets.csv', filter and export sets that match these conditions:

1. Sets from the 'Modern day' themeGroup
2. With more than 500 pieces
3. Sort results by number of pieces (ascending)
4. Export results to a CSV file containing all original columns

Ex5

Given the input CSV file 'lego_sets.csv', perform the following analysis:

1. Calculate the following statistics for sets between 1990 and 2000:
 - Number of sets per year
 - Average number of pieces per year
 - Average price per year (if available)
2. Export results (one row for each year) to a CSV file with columns: year, total_sets, avg_pieces, avg_price
3. Sort results by year ascending