## Tool Used: Python, Source: New York **Times API**

request to the

NYT API.

Send a GET

Operation:

Receive JSON

article data. containing response

- Description: Extract York Times Article data from the New
  - Parameters: Search API.
- fetch articles q: Keyword related to
- api-key: Your Paramount.
  - authentication. API key for
    - **Data Retrieved:**
- publication dates, Article headlines, snippets,
  - document types, web URLs, word section name,

## **TRANSFORMATION**

- Jupyter Notebook Name:
- Operations:

'requests' library

**EXTRACTION** 

Parse JSON: Convert JSON

Articles: Stores detailed article data

Destination: AWS RDS MySQL Database.

• Database Schema:

headline, snippet, pub\_date document\_type , news\_desk ,

Fields: article\_id (PK),

section\_name, web\_url,

word\_count.

Sections: Stores unique sections

 Fields: section\_id (PK), section\_name.

- dictionary.
  - o data =
- json.loads(response.text)
- formats in pub\_date .
- Data Filtering:
- irrelevant data based on specific criteria.
- Extract year, month, and day from
  - analysis.

- NYT\_API\_ETL.ipynb
- response into a Python
- Data Cleaning:
- Normalize date
- Clean and standardize text fields like headline and snippet.
- Select articles based on relevant sections if
  - Remove duplicates or necessary.
- · Feature Engineering:
- pub\_date for trend

counts.