Working with API

What is an API?

- API stands for Application Programming Interface, which allows different software systems to communicate and exchange data with each other.
- APIs are widely used for data retrieval, such as booking flights, train tickets, or accessing databases from third-party websites

Why Use APIs in Machine Learning?

- APIs provide real-time, structured data, which is crucial for building machine learning models or datasets.
- When direct data sources (CSV, Excel) are unavailable, APIs are the primary method to obtain required information

Steps to Fetch Data and Create a DataFrame

1. **Import Libraries**: Use `pandas` for DataFrame creation and `requests` for API calls.
2. **Send Request**: Pass the API key and desired endpoint to get data in JSON format.
3. **Parse JSON**: Convert the JSON response into a Python dictionary/list.

4. **Extract Fields**: Select only the fields needed for your DataFrame (e.g., movie ID, title, release date, overview, popularity, vote count).
5. **Loop for Multiple Pages**: If the API returns paginated data, loop through all pages to collect all records.
6. **Create DataFrame**: Use `pandas.DataFrame()` to convert the collected data into a DataFrame.
7. **Save Data**: Export the DataFrame to CSV or upload it to platforms like Kaggle for sharing.

Key Concepts and Tips:

- APIs are the backbone of modern web-based data exchange, enabling real-time access to remote databases.
- Always check API documentation to understand endpoints, required parameters, and rate limits.
- Handling paginated responses is important for large datasets; use loops to aggregate all pages.
- The resulting DataFrame can be used for analysis, modeling, or shared as a dataset for community projects