Project Documentation: newsdata.io API Integration

Project Overview

This project focuses on analyzing news data using the newsdata.io API. The goal is to fetch real-time and historical news articles across various categories like business, technology, health, etc., and use them for generating insights or reports. Instead of manually scraping news websiteswhich can be complex and riskythis project leverages APIs to ensure legal, clean, and scalable access to structured news data.

What is an API?

API stands for Application Programming Interface. It acts as a bridge between two applications to communicate with each other. Think of it like a waiter at a restaurant: you place an order (request), the waiter takes it to the kitchen (server), and brings the food (response) back to your table. You dont need to know how the kitchen worksjust how to place your order.

Explanation of newsdata.io

newsdata.io is a news aggregation service that offers structured access to global news data through a simple API. It collects and delivers real-time articles from various publishers, allowing developers to use this data in their applications without having to manually gather or format the news content.

How to Use newsdata.io

1. Register on the newsdata.io platform and get your API key.

2. Use tools like Python, Postman, or even a browser to make requests to newsdata.ios endpoints.

3. Each 'request' means asking the API to give you news based on your query (e.g., keyword, date, language).

4. The API responds with data in JSON format containing headlines, links, content snippets, and more.

Free vs Paid Plans

newsdata.io offers both free and paid tiers depending on your needs:

 Free Plan: Good for testing with limited requests and basic metadata.

 Paid Plans: Allow more requests and unlock premium features like full article content and historical data.

Pricing

 Free Plan: $0/month (Approx 0) 100-250 requests/day.

 Pro Plan: $29/month (Approx 2,400) 10,000 requests/month.

 Ultra Plan: $99/month (Approx 8,200) 50,000 requests/month, full article access.

 Business Plan: $299+/month (Approx 24,800+) Custom access, support, and archives.

Note: A 'request' is each time your app asks the API for data. For example, requesting news articles for 'India' once counts as one request.

Why APIs are Useful in Your Project

 Get structured news data easily without writing complex code.

 Real-time updates and automatic access to the latest headlines.

 Avoid legal issues of scraping.

Scalable for high-volume data use.

Easy to integrate with Python, MongoDB, or Power BI.

What is Scraping and Why It's Not Recommended

Web scraping means extracting data directly from websites using code. It involves reading HTML pages and collecting specific information.

Problems with scraping:

May violate websites' terms of service (illegal in some cases).

Websites can block scrapers or change their layout.

Very hard to scale and maintain.

May miss or corrupt important data.

Takes time to develop and test.

Final Conclusion

Scraping large datasets is complex, risky, and often not allowed. It also requires a lot of maintenance and server handling.

Instead, using APIs like newsdata.io is a legal, scalable, and reliable way to fetch real-time or historical news data. APIs ensure your project gets clean and well-structured content without the risks or technical difficulties of scraping.