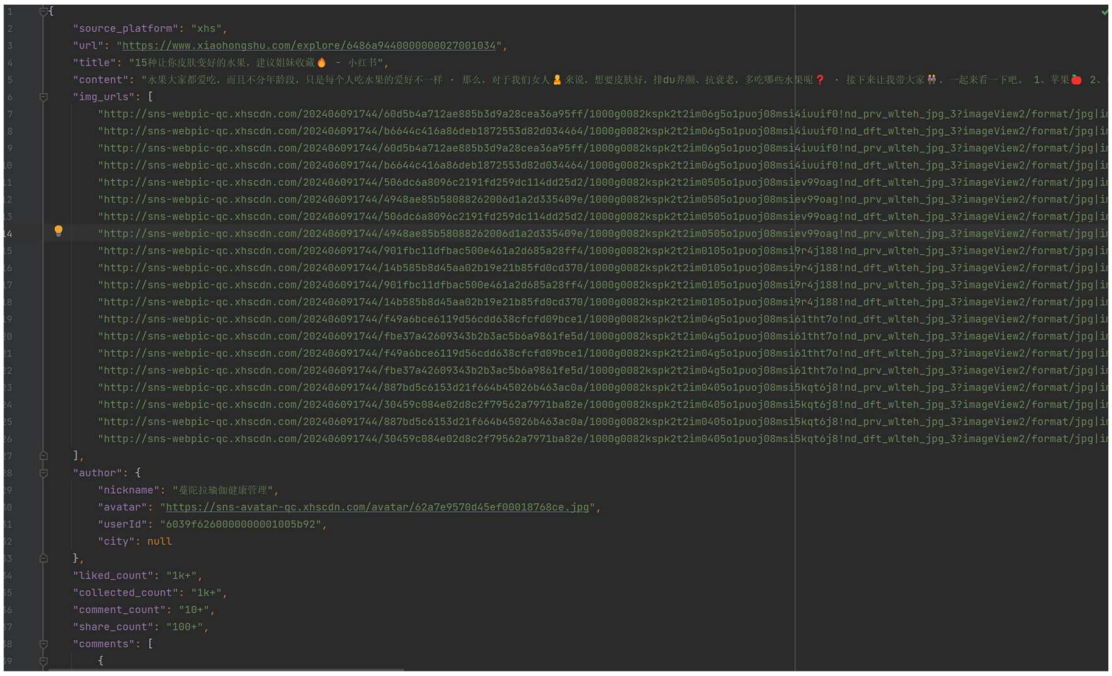
# Data Collection from social media

Xiaohongshu:

1. Do have official API, but it is not open to the public. These APIs are mainly used for internal applications and partners: like E-commerce
2. Some third-party companies provide API
3. Scraper: https://github.com/XiaochenCui/Zhihu-Scraping/blob/master/zhihuScraping.py




X:

1. Have official paid X API:
   1) Free: No posts access with API
   2) Basic: 10,000/month Posts read-limit rate cap; cost: $100 per month
   3) Pro: 1,000,000 posts per month; cost: $5,000 per month

   https://developer.x.com/en/docs/twitter-api

   total posts used in PLIP: 232,067 tweets
2. Does not allow the users to use web scrapers.
   - https://developer.x.com/en/developer-terms/agreement-and-policy
3. Scraper: https://github.com/israel-dryer/Twitter-Scraper/blob/main/scrapper.py

Zhihu:

1. User policy: https://www.zhihu.com/term/zhihu-terms

   您应对您使用本平台的行为负责，除非法律允许或者经知乎事先书面许可，

   您使用本平台不得具有下列行为：

   以任何方式（包括但不限于盗链、冗余盗取、爬取、抓取、模拟下载、深度

链接、假冒、模拟注册等）直接或间接盗取知乎平台的数据和内容；恶意注册知乎账号，包括但不限于频繁、批量注册账号。

2. Official API used for internal applications and partners. There are third-party API but may be not official recognized
3. Scraper: https://github.com/ZWY24/Web-Scrapping/blob/main/scraper.py