案例名称：网易云音乐歌曲信息采集与分析系统

项目概述

开发一个基于Scrapy框架的网易云音乐歌曲数据采集系统，用于采集网易云音乐热门歌曲的详细信息，并对采集到的数据进行分析和可视化展示。系统将包括数据采集、数据存储、数据分析和数据可视化四个主要模块。

项目目标

1. 实现一个Scrapy爬虫，用于采集网易云音乐热门歌曲的详细信息，包括歌曲名称、歌手、专辑、发行时间、歌词、评论数量和用户评分等。

2. 将采集到的数据保存为JSON格式，并存储在本地文件系统中。

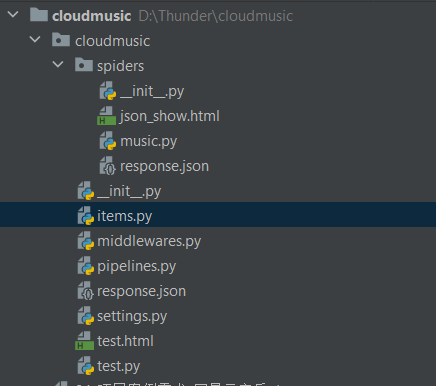
3. 使用ECharts等可视化工具，将json以图表的形式展示出来，包括但不限于歌曲流行度柱状图、用户评分分布图、热门歌手和专辑的饼图等。

项目详细步骤

步骤1：环境搭建和Scrapy项目创建

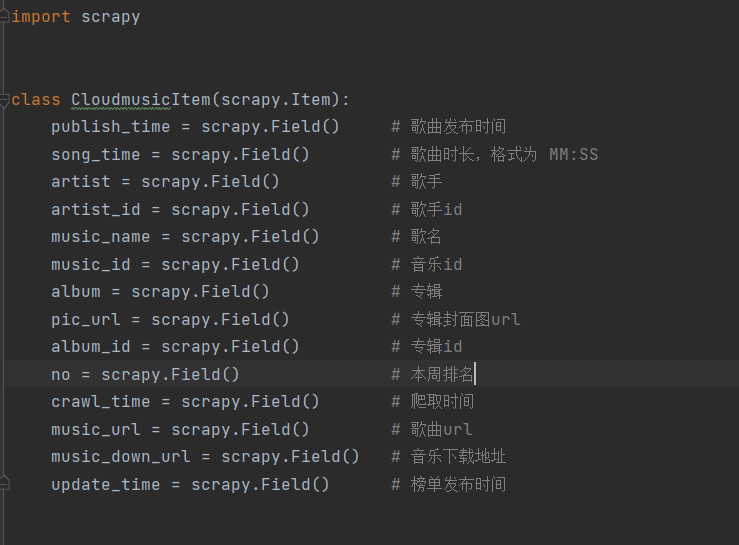
- 安装Python和Scrapy框架。

- 使用`scrapy startproject`命令创建一个新的Scrapy项目，例如命名为`cloudmusic`.



步骤2：定义数据结构

- 在项目的`items.py`文件中定义一个` CloudmusicItem`类，用于存储歌曲信息，如名称、歌手、专辑等。

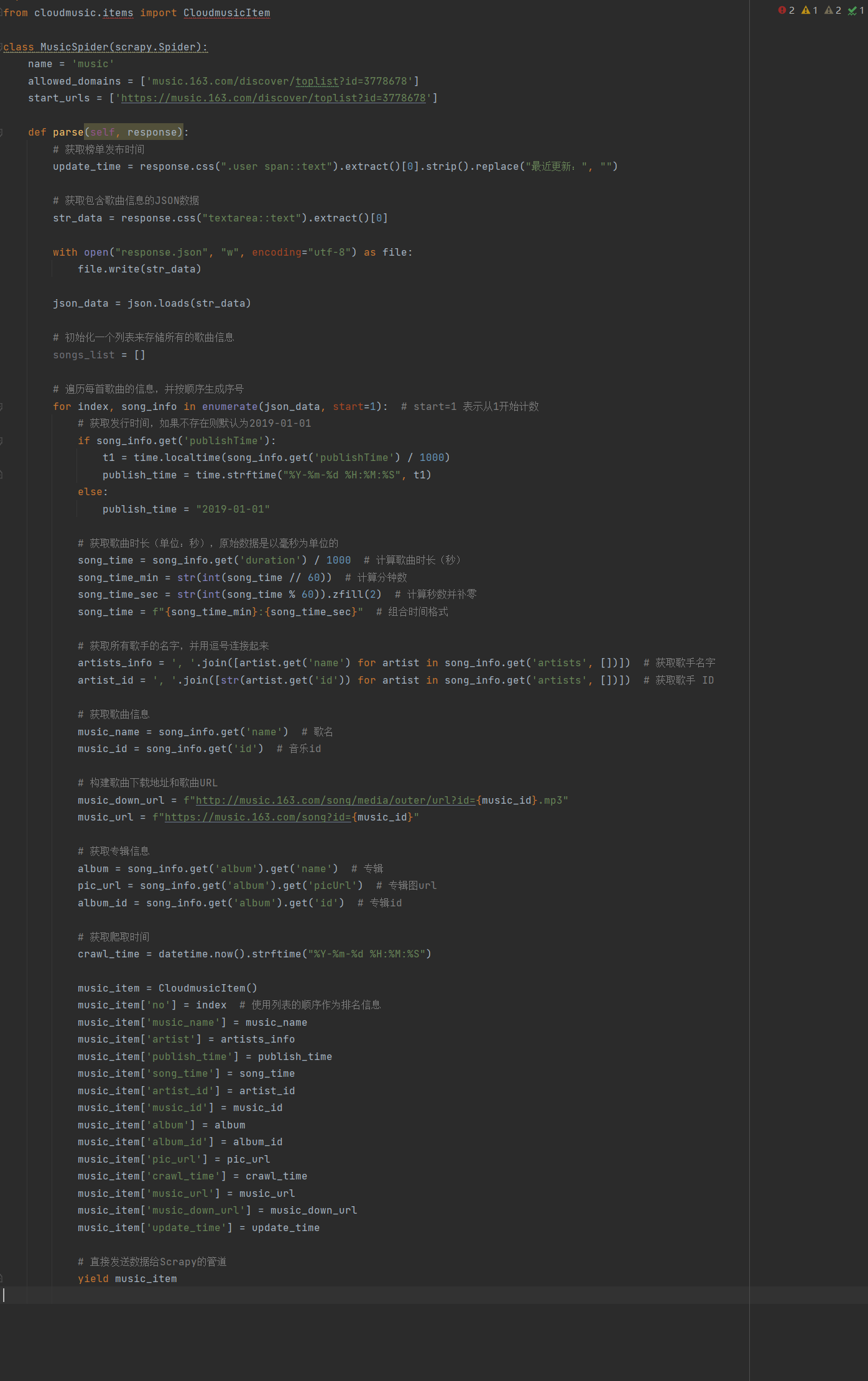


步骤3：创建爬虫

- 在项目的`spiders`目录下创建一个新的爬虫文件，例如`netease\_spider.py`。

- 编写爬虫代码，使用CSS选择器或XPath选择器提取歌曲信息。

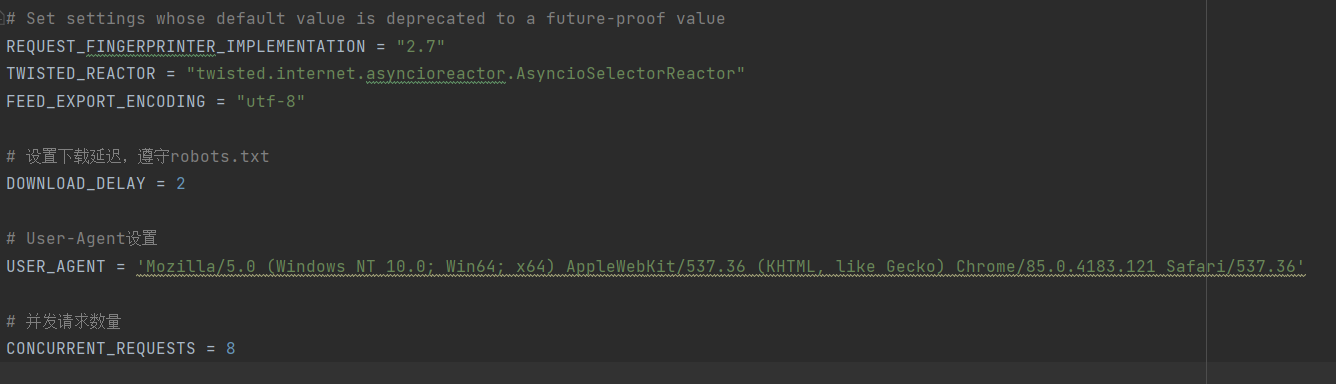
- 实现分页处理逻辑，确保能够遍历所有目标歌曲页面。



步骤4：设置项目配置

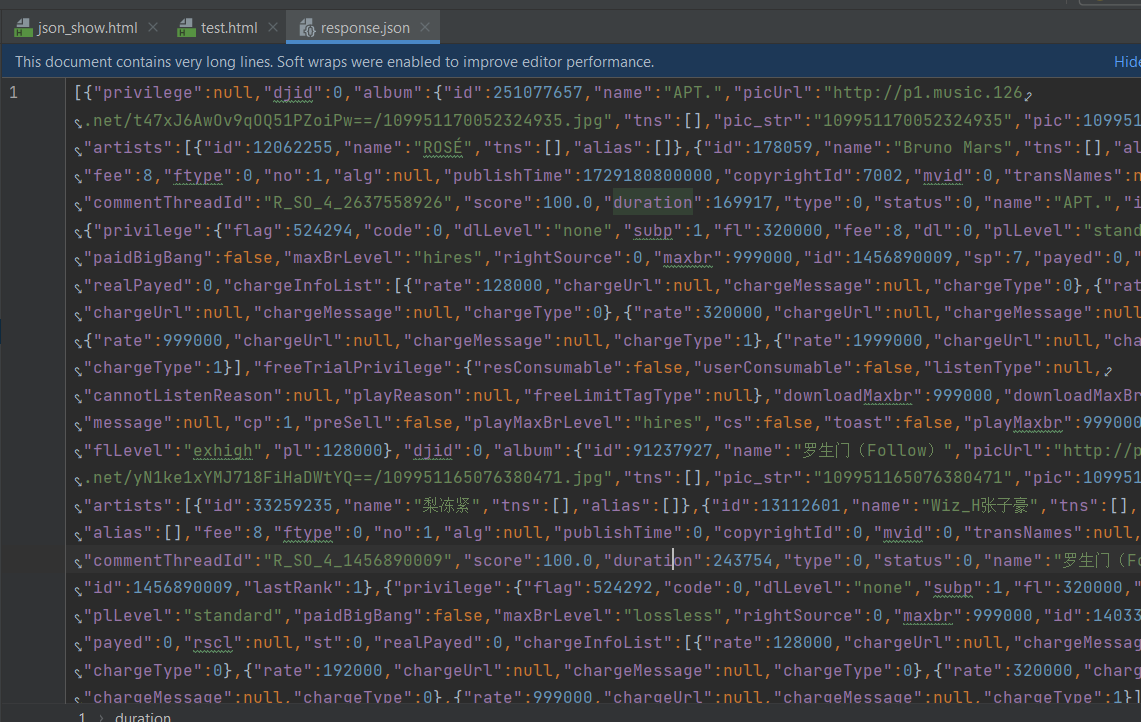
- 在`settings.py`文件中设置并发请求数量、下载延迟等参数，以遵守网易云音乐网站的robots.txt规定并减少对网站的压力。

- 设置User Agent和下载延迟。



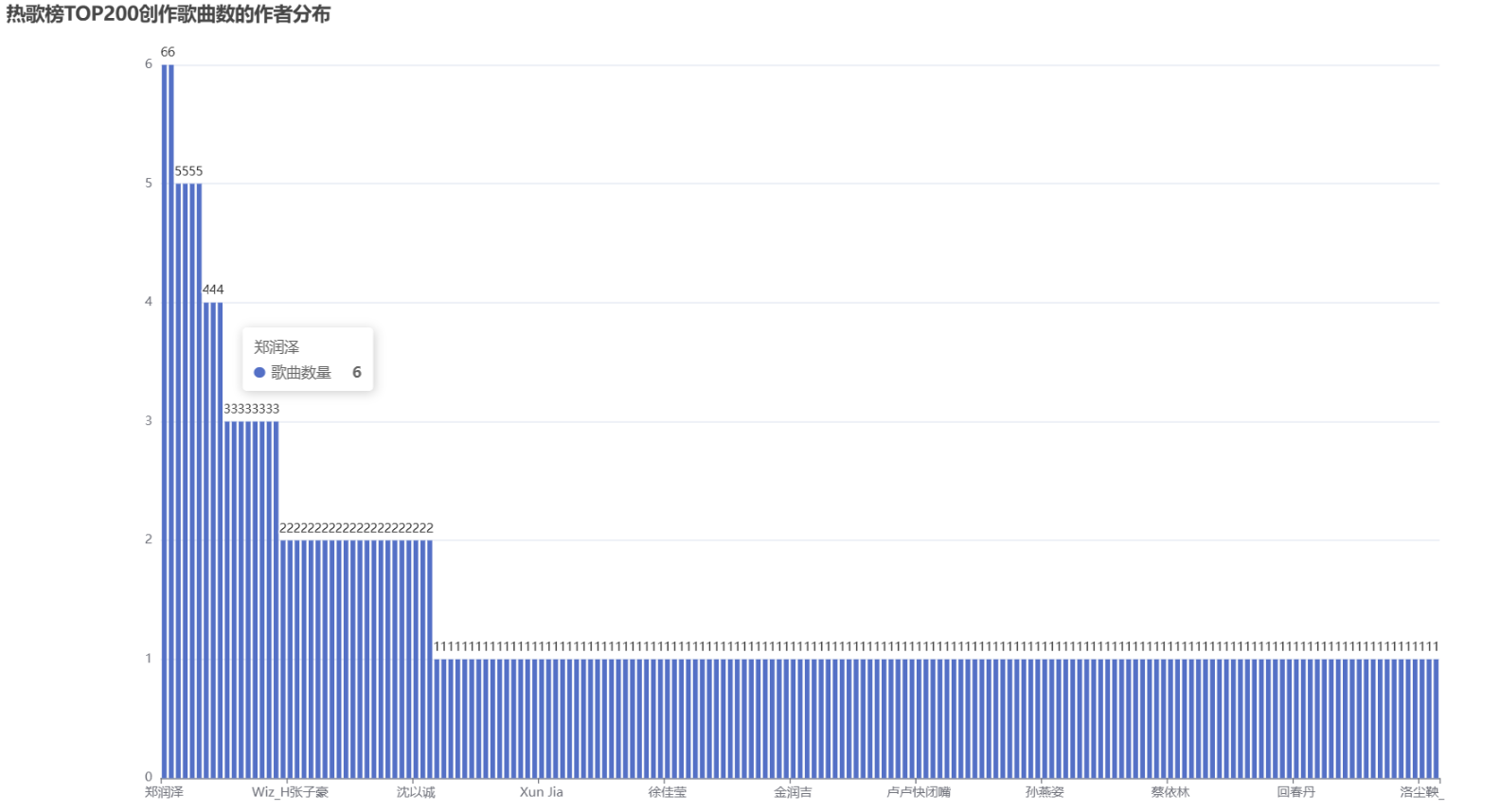
步骤5：运行爬虫并保存数据

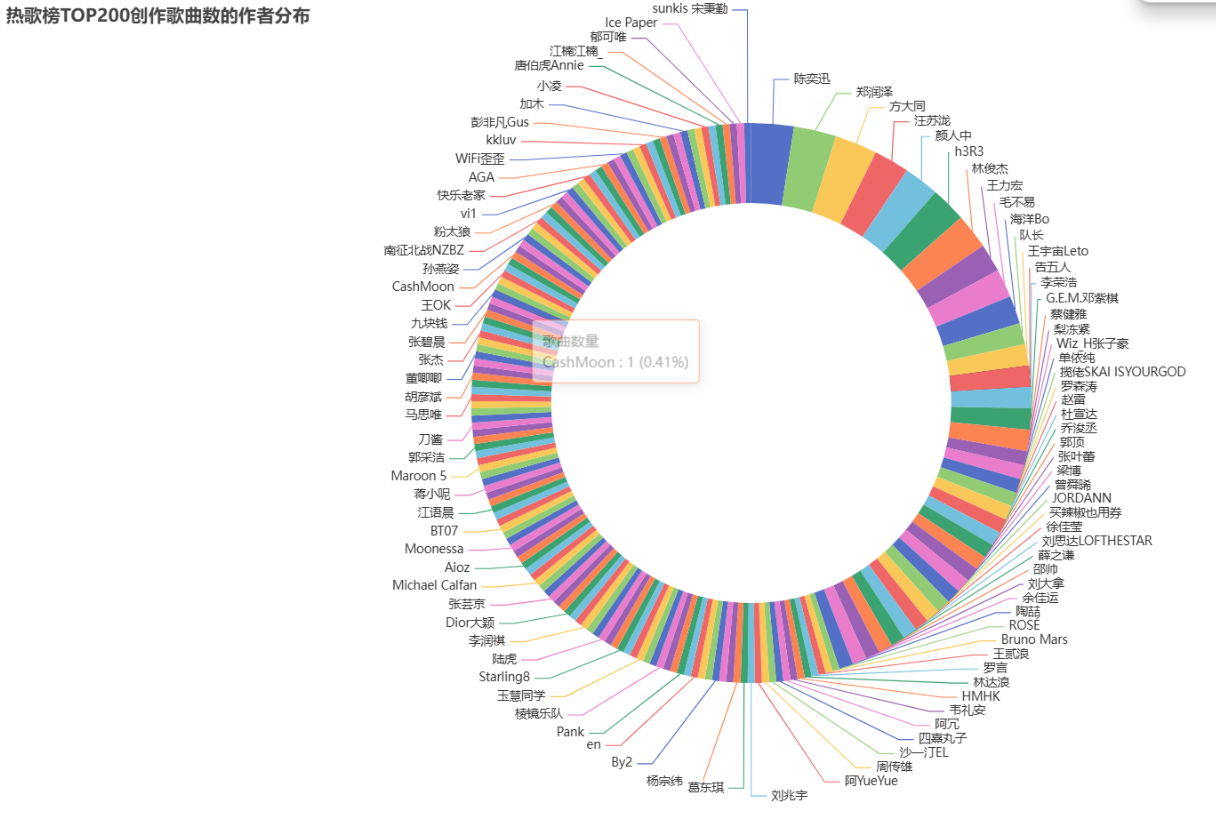
- 在命令行中运行爬虫，使用`-o`选项将结果保存为JSON文件，例如`music.json`。



步骤6：数据可视化

- 使用ECharts创建一个HTML页面，用于展示分析结果。

- 根据分析结果，创建相应的图表，如柱状图、折线图、饼图等。



步骤7：项目部署

- 将项目部署到服务器或本地计算机，确保可以通过Web浏览器访问数据可视化页面。

注意事项

- 确保遵守网易云音乐网站的爬虫政策，避免对网站造成过大压力。

- 考虑到网站的反爬虫机制，可能需要设置请求头、Cookies或使用代理。

- 在进行数据可视化时，确保图表清晰、易于理解。

项目交付物

- 完整的Scrapy爬虫代码。

- 采集到的网易云音乐歌曲的JSON数据文件。

- 使用ECharts创建的数据可视化页面。

通过完成这个项目，你将能够掌握使用Scrapy进行数据采集、使用ECharts进行数据可视化的技能，这对于数据分析和Web开发都是非常有价值的。同时，你也将对国内音乐平台的数据结构和反爬虫策略有更深入的了解。