Data Mining Task

Task 1

- Create crawler/Scraper using Python [Preferable using Scrapy] to scrape data from https://www.forbes.com/lists/global2000/?sh=45f017755ac0 and save data in JSON format, attribute-wise.
- Get the list of all 2000 companies.
- Final Json document should contain
    - Rank
    - Name
    - Country
    - Sales
    - Profit
    - Assets
    - Market Value
    - Link to individual company profile

E,g :

{
rank : 1,
name : Berkshire Hathaway,
country : United States,
sales : $276.09B,
profit : $89.8B,
assets : $959.4B,
market_value : $700B,
link : https://www.forbes.com/companies/berkshire-hathaway/?list=global2000&sh=4616376fbef8
}

Task 2

- Crawl Top 20 company profiles using the above extracted links based on ranking.
- Store in a json file with proper formatting.

Note : Use your own understanding on what kind of data would need to be scraped based on proper reasoning and relevance.

(PS : It's always good to have more data, so don’t restrict yourselves on a small number of parameters, however relevance and use case for whatever data we scrape is important too)

Bonus:

- Use parallel processing in the script.
- Any out-of-the-box ideas would also be well appreciated.

Submission guideline: Upload the code to Github and email the link.

Timeline: 2-4 days