Twitter Scraper

Technologies used:

- Scrapy has been used in the project which is a popular web-scraping framework.
- Scrapy-selenium
- Beautifulsoup4

Code clarification:

- Created a spider called twitter and initialised the targeted urls.
- Customised my webdriver settings to give a delay of 10 seconds between every link.
- Used start request to initialise my spider request.
- Tried to extract all the tweets by different ways (Beautiful soup - scrapy-selenium)

Final Results:

- The spider was the web page without any errors <200>

```
024-03-16 13:37:47 [scrapy.core.engine] INFO: Spider opened
024-03-16 13:37:47 [scrapy.extensions.logstats] INFO: Crawled 0 pages (at 0 pages/min), scraped 0 items (at 0 items/min)
024-03-16 13:37:47 [scrapy.extensions.telnet] INFO: Telnet console listening on 127.0.0.1:6023
024-03-16 13:37:48 [scrapy.core.engine] DEBUG: Crawled (200) <GET https://twitter.com/Mr_Derivatives> (referer: None)
```

- There is no data returned
- Please check the result in output.csv

Conclusion:

- According to my point of view this is because the Web-driver is not opening in the maximised window that's why there is no data returned probable.
- Even when I tried to disable the headless mode it start giving me exception [Key value: driver]
- Also, I wrote a selenium script to test my web-driver it gave me the same exception although, the version is compatible with my chrome version