

## WEB SCRAPING AND TEXT EXTRACTION ASSIGNMENT- PYTHON

1. Go to URL: <https://www.marketsandmarkets.com/telecom-and-IT-market-research-113.html> (no pagination)
2. Scrape the following details:
  - a) Report Titles (Eg: Artificial Intelligence Market By Type, By Technology, By Region....)
  - b) Date Published
  - c) Link of each report (Eg: <https://www.marketsandmarkets.com/Market-Reports/edge-ai-software-market-70030817.html> )
  - d) Report descriptions (Eg: The Edge AI Software Market is expected to grow from USD 590 million in 2020 to USD 1,835 million by 2026, at a CAGR of 20.8% during the forecast period.)
3. Extract/ match/ find the first occurrence of 'USD 'values and string containing % values in the report description.  
(EG: Ideal Output= USD 590 million, 20.8%)
4. Store all fields i.e report\_title, date\_published, link, report\_description, USD\_value, %\_val in a csv file and share the .csv file as the final output and answer the below questions.
5. What are the different packages or modules used in this assignment to scrape all the results? Which are the different web scraping packages in Python? Among these, which package/ module scrapes websites fastest according to you?