



The research team is aligning its operations with the established process standards of the data team. Consequently, their procedures must adhere to these new standards, emphasizing the necessity for automation and testing. In initiating this migration process, the first step is the development of a dedicated scraper. This tool is intended to systematically extract data from the Internet Crime Complaint Center (IC3) Annual Reports.

A link to 2016 Alabama's State Report:

<https://www.ic3.gov/Media/PDF/AnnualReport/2016State/StateReport.aspx#?s=1>

Requirements:

1. Develop a Python script that scrapes the data from the IC3 Annual Report, specifically targeting state indexes 1 to 57 for the years 2016 to 2022;
2. Extract data into 5 tables (*ic3\_\_crime\_type\_by\_victim\_count*, *ic3\_\_crime\_type\_by\_victim\_loss*, *ic3\_\_crime\_type\_by\_subject\_count*, *ic3\_\_crime\_type\_by\_subject\_loss*, *ic3\_\_victims\_by\_age\_group*);
3. Upload the extracted tables as a parquet file to local partitioned folders;
4. The process should be able to run on demand;
5. Document the solution and put it on GitHub providing the project's URL;

Extra:

1. At least one unit test method that covers part of your script;
2. Monitoring is important, so implement logging to your script;

If you have any questions, you can contact [mantas.satkevicius@surfsharkteam.com](mailto:mantas.satkevicius@surfsharkteam.com)