南豐集團
NAN FUNG GROUP

Data Engineer (International Property)

Technical Assessment

The goal of technical test is to assess your ability to design and implement big data systems. You are free to use any open-source library for the technical question. You are highly encouraged to make use of AWS services for the technical test to build big data systems.

Questions

1. Build a web scraper that scrapes https://www.michaelkors.com/ on a daily cadence. Please focus only on extracting the price/item data for the women handbag category. Handling of page pagination is great to have but not a must. Ideally, you can cache the entire page and have separate programs to extract the item price/product info into a more structural format. The scraper is expected to run on a regular cadence so please incorporate that into your design. Please prepare design diagrams that illustrate key components of your systems, code base, and some collected data samples for discussion. Collected data must include but not limited to the following 3 columns: Product name, Price, timestamp when the data is collected.

Requirements

1. You will be required to use Python / Jupyter Notebook for data manipulation and analysis. Feel free to use libraries e.g. panda, numpy.
2. For your submission, please submit the original Jupyter notebook as well as the PDF form of the notebook. The notebook is expected to run end-to-end successfully without any error. You can also include another file for charts, visualizations, writeups if there is a need.
3. There might be a case debrief after the exercise depending on the deliverables.
4. You are free to use any open-source library. You are highly encouraged to make use of AWS services to complete the task. Please include documentation that will allow us to set up the environment with the required dependencies to run the code successfully on our side. Your code is expected to be fully functional. Python is strongly preferred for implementation.

Thank you for your time and feel free to reach out our HR representative if you do have any question.