***I. DataWrangling***

- folder “data” chứa các dữ liệu gốc tải về từ drive đề bài

- folder “cleandata” chứa

+ folder “dataset”: các dataset sau khi clean ( “unilist.csv” là danh sách các trường theo tiêu chí chọn )

+ các file “.ipynb” là các file python để xử lí dữ liệu ( có ghi chú markdown chi tiết trong từng file )

***II. Descriptive statistics & visualization***

- “central\_tendency.xlsx” là file chứa 2 sheet là những bảng thống kê đề bài yêu cầu

- “visualization.ipynb: là file python để xử lí và biểu diễn dữ liệu

- “README.md” hoặc “Statistics.ipynb” là file phân tích dữ liệu

***III. Logical thinking and mini case study***

- “README.md” là file tổng hợp quá trình làm bài và phân tích dữ liệu, trả lời các yêu cầu đề bài dưới dạng Markdown

- “dataEcommerceClean.csv” là dữ liệu sau khi được chuẩn hóa và làm sạch

- “cleanData.ipynb” là file python dùng để xử lí dữ liệu

***IV. Report automation***

- “cleanData.ipynb”: file xử lí dữ liệu

- “cleanData\_out.ipynb”: file sau khi chạy tự động, nếu xuất hiện lỗi sẽ kiểm tra ở file này

- “report\_data.xlsx”: tập dữ liệu gốc (nếu update thêm thông tin sẽ update vào đây)

- “reportDataClean.xlsx”: tập dữ liệu sau khi được xử lí và chuẩn hóa

- “run\_notebooks.py”: file python dùng để chạy notebook

- “runPython.bat”: dùng để chạy tự động hóa, cập nhật dữ liệu cho cho “reportDataClean.xlsx”