

BÁO CÁO NHÓM 6 CASE 11

Đề tài: Phân tích biến động điện năng tiêu thụ hộ gia đình

Lớp/Nhóm: DHKL18A1HN– Nhóm 6 **Thành viên:**

| | |
|-----------------|-------------|
| Nguyễn Vũ An | 24174600025 |
| Hà Giang Nam | 24174600014 |
| Bùi Quang Huy | 24174600018 |
| Vũ Trần Duy Anh | 24174600045 |
| Trần Lương Sỹ | 24174600060 |

Ngày nộp: 16/12/2025

I. Đặt vấn đề

Tại các thành phố và khu dân cư ở Việt Nam, dữ liệu điện năng tiêu thụ được ghi nhận theo từng hộ gia đình dựa trên công tơ điện và các bậc thang giá. Case Study này giúp sinh viên thực hành đầy đủ các kỹ năng Pandas: làm sạch dữ liệu, truy vấn, phân nhóm, merge và xây dựng pivot trong bối cảnh phân tích điện năng tiêu thụ hộ gia đình.

II. Mục tiêu

- Làm sạch và chuẩn hóa dữ liệu điện năng tiêu thụ hộ gia đình. Vận dụng NumPy để xử lý dữ liệu dạng mảng 1 chiều hoặc 2 chiều.
- Thực hành các thao tác xử lý dữ liệu với Pandas: truy vấn, groupby, merge, pivot
- Phân tích mức tiêu thụ điện theo khách hàng, loại nhà và khu vực.
- Ước tính tiền điện theo biểu giá bậc thang.
- Phát hiện các bản ghi tiêu thụ bất thường
- Tổng hợp kết quả phân tích phục vụ báo cáo.

III. Phương pháp

1. Làm sạch dữ liệu (Task 01)

- Chuẩn hóa `custom_id`: loại bỏ khoảng trắng, chuyển về dạng viết hoa thống nhất.
- Chuẩn hóa `house_type`, `region`: loại bỏ khoảng trắng thừa, viết hoa chữ cái đầu.
- Làm sạch `meter_id`: xóa khoảng trắng và ký tự không hợp lệ.
- Chuẩn hóa `reading_date` về một định dạng ngày duy nhất bằng `to_datetime()`.
- Làm sạch `consumption_kwh`: loại bỏ text, chuyển kiểu số, xử lý giá trị âm bất thường.
- Chuẩn hóa các trường `kwh_min`, `kwh_max`, `price_per_kwh` trong bảng giá điện.
- Xuất dữ liệu sạch sang thư mục `data_clean`/ để phục vụ các bước tiếp theo.

2. Truy vấn và thống kê mô tả (Task 02)

- Thống kê số lượng hộ gia đình theo từng `house_type`.
- Tính tổng và mức tiêu thụ điện trung bình theo từng khách hàng.
- Phát hiện các bản ghi tiêu thụ bất thường (âm, bằng 0 hoặc quá lớn).
- Thống kê mức tiêu thụ điện trung bình theo `region`.
- Đếm số lần ghi chi số (số kỳ) của mỗi hộ gia đình.

3. Phân tích GroupBy và tổng hợp (Task 03)

- Phân tích lượng điện tiêu thụ trung bình theo cặp `house_type` × `region`.
- Gán bậc thang giá điện tương ứng để ước tính tiền điện từng kỳ.
- Tính tổng điện năng và tổng tiền điện theo từng khách hàng.
- Xác định nhóm loại nhà có mức tiêu thụ điện cao nhất.
- So sánh chênh lệch tiêu thụ điện giữa các vùng miền.

4. Merge dữ liệu (Task 04)

- Kết hợp bảng `customer_profile` và `consumption_log` theo `customer_id`.
- Xác định bậc thang giá điện phù hợp cho từng bản ghi tiêu thụ.
- Phát hiện các bản ghi không khớp bậc thang hoặc không tồn tại khách hàng.
- Tạo bảng dữ liệu hoàn chỉnh phục vụ phân tích nâng cao.

5. Pivot Table và Stack/Unstack (Task 05)

- Xây dựng pivot tổng điện năng theo `house_type` × `tariff_level`.
- Tạo pivot tiền điện trung bình theo `region` × `house_type`.
- Thực hành `stack()` và `unstack()` để phân tích dữ liệu đa cấp.
- Đánh giá nhóm khách hàng và bậc thang tiêu thụ điện nhiều nhất.

IV. Kết quả

Các kết quả phân tích dữ liệu đã lưu trong price_tariff_clean.csv, customer_profile_clean.csv và consumption_log_lean.csv thể hiện trực quan nhất xu hướng tiêu thụ điện

V. Nhận xét

Các hộ Biệt thự và Nhà phố có xu hướng tiêu thụ điện cao hơn so với Chung cư.Khu vực Nam có mức tiêu thụ điện trung bình cao hơn các khu vực còn lại.Phần lớn điện năng tập trung ở các bậc thang trung – cao, kéo theo chi phí điện tăng mạnh.Một số bản ghi bất thường đã được phát hiện và loại bỏ, giúp kết quả phân tích chính xác hơn.

VI. Nhật ký đóng góp nhóm

| Thành viên | Mã số | Nhiệm vụ cụ thể |
|-----------------|-------------|---|
| Nguyễn Vũ An | 24174600025 | Đọc & làm sạch dữ liệu: Làm sạch và chuẩn hóa dữ liệu thô nhằm đảm bảo tính nhất quán và độ tin cậy trước khi phân tích. |
| Hà Giang Nam | 24174600014 | Merge dữ liệu: Kết hợp các bảng dữ liệu liên quan nhằm tạo bộ dữ liệu hoàn chỉnh phục vụ phân tích chuyên sâu. |
| Bùi Quang Huy | 24174600018 | Truy vấn & thống kê mô tả: Làm sạch và chuẩn hóa dữ liệu thô nhằm đảm bảo tính nhất quán và độ tin cậy trước khi phân tích. |
| Vũ Trần Duy Anh | 24174600045 | GroupBy & Tổng hợp: Phân nhóm và tổng hợp dữ liệu để phân tích hành vi tiêu thụ điện theo loại nhà, khu vực và chi phí. |
| Trần Lương Sỹ | 24174600060 | Pivot Table + Stack/Unstack: Xây dựng pivot table và chuyển đổi cấu trúc dữ liệu để phân tích điện năng tiêu thụ theo nhiều chiều. |

VII. Kết luận

- Dữ liệu điện năng tiêu thụ hộ gia đình đã được làm sạch và chuẩn hóa đầy đủ.
- Phân tích cho thấy mức tiêu thụ điện có sự khác biệt rõ rệt theo loại nhà và khu vực.
- Các bậc thang giá cao tập trung chủ yếu ở một số nhóm khách hàng nhất định.
- Case Study giúp nhóm rèn luyện hiệu quả các kỹ năng xử lý và phân tích dữ liệu với Pandas.