**Nguyễn Ngọc Ánh**

**3120410041**

**BÀI THỰC HÀNH SỐ 3**

1. **File Temperature Readings : IOT Devices - IOT-temp.csv**

**Mô tả tập dữ liệu:**

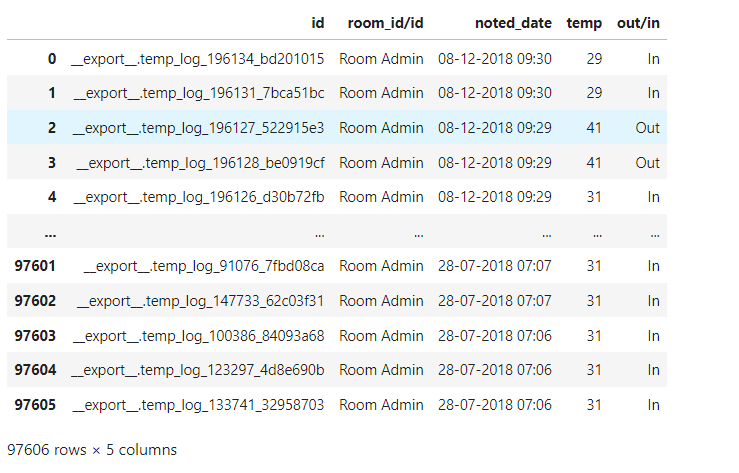
Tập dữ liệu này chứa các chỉ số nhiệt độ từ các thiết bị IOT được cài đặt bên ngoài và bên trong Phòng ẩn danh (giả sử - phòng quản trị). Thiết bị đang trong giai đoạn thử nghiệm alpha. Vì vậy, nó đã được gỡ cài đặt hoặc tắt nhiều lần trong toàn bộ thời gian đọc ( 28-07-2018 đến 08-12-2018).

Mô tả chi tiết các cột dữ liệu với 5 cột và 97605 hàng:

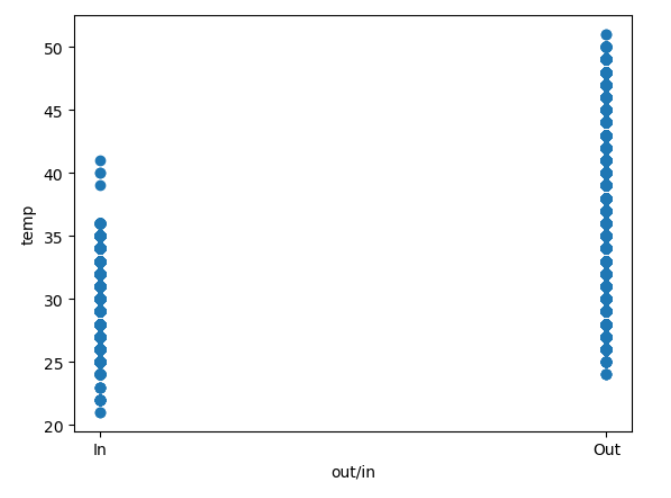
* ID: ID duy nhất cho mỗi lần đọc
* room\_id/id: ID phòng trong đó thiết bị đã được cài đặt (bên trong và / hoặc bên ngoài) -> hiện tại “Admin room” chỉ cho mục đích ví dụ.
* noted\_date: ngày và giờ đọc
* temp: chỉ số nhiệt độ
* out/in: việc đọc được lấy dữ liệu từ thiết bị được cài đặt bên trong hay bên ngoài phòng.

**Mô tả quá trình giả định:**

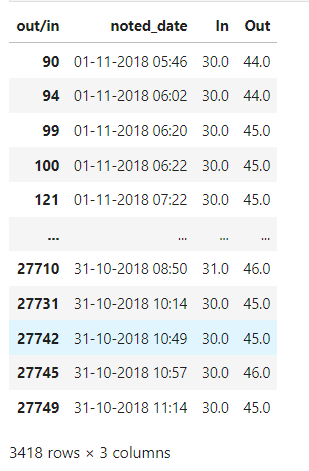
* Xoá bỏ cột dữ liệu không cần thiết



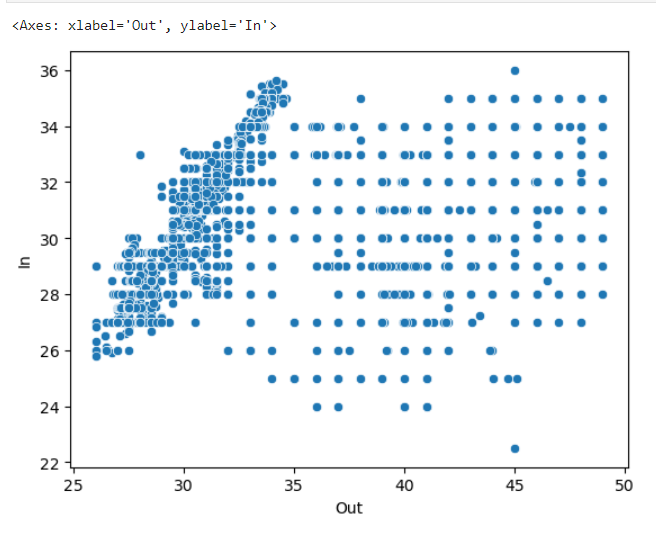
* Vẽ biểu đồ hiện thị nhiệt độ trong phòng và ngoài trời

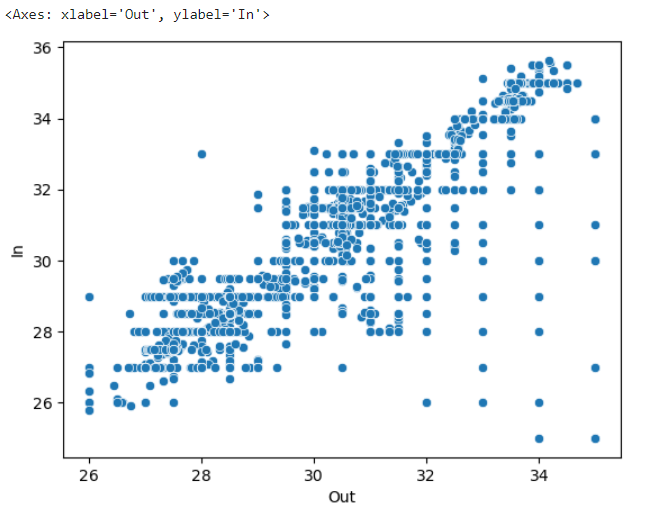


* Tạo data mới dựa trên data cũ

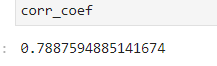


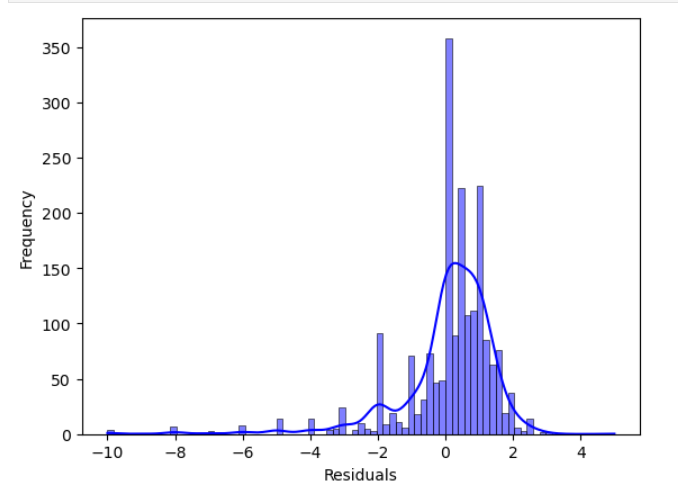
* Vẽ biểu đồ sai số và lọc



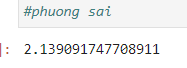


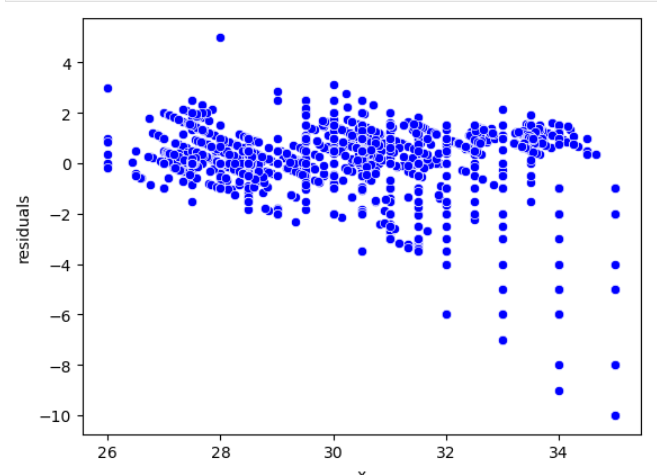
* Tính hệ số tương quan, tính sai số và vẽ biểu đồ phân bố sai số





* Tính phương sai và vẽ biểu đồ phân tán





* Vì không có hai biến độc lập nên không có sự tương quan giữa các biến độc lập

1. **File Laptops and Notebooks from 2020-2023 - mindfactory\_update.csv**

**Mô tả tập dữ liệu:**

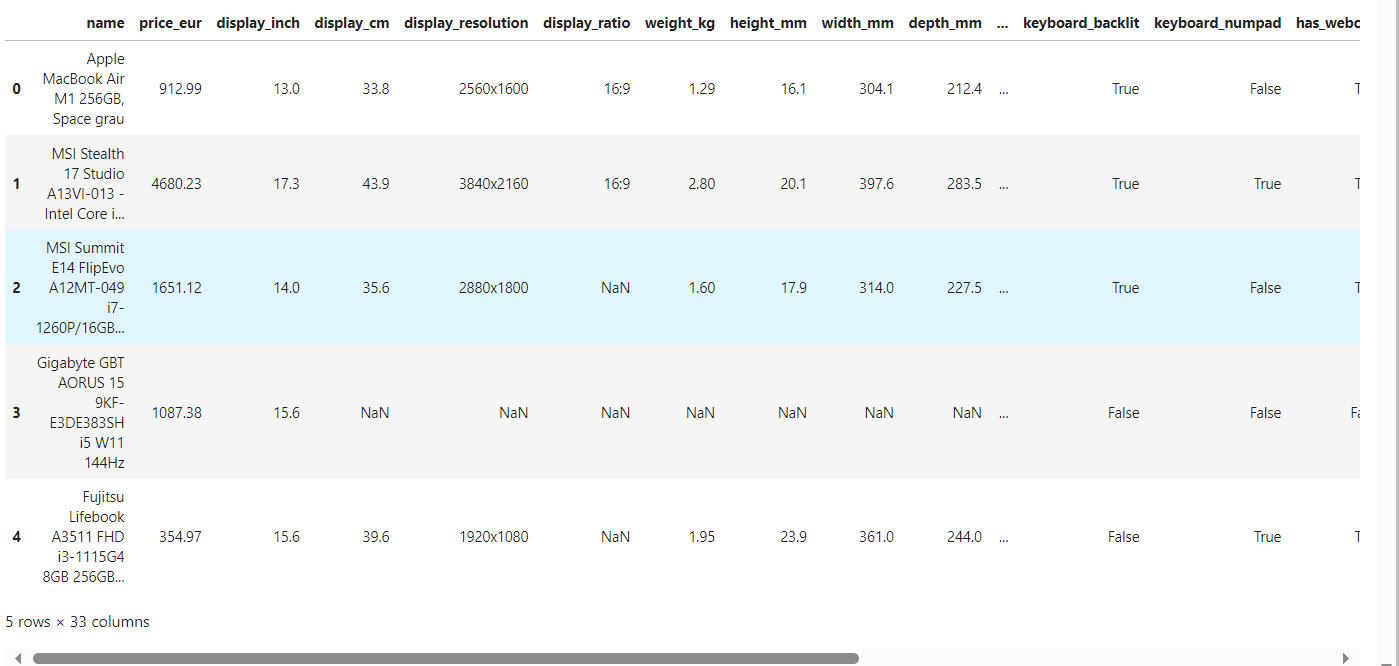
Thông số kỹ thuật chi tiết từ một loạt các máy tính xách tay và máy tính xách tay hiện đại khác nhau, từ Apple MacBook và Lenovo ThinkPad, máy tính xách tay chơi game trên Gigabyte đến máy tính bảng Microsoft Surface.

Mô tả chi tiết dữ liệu với 33 cột và 768 hàng:

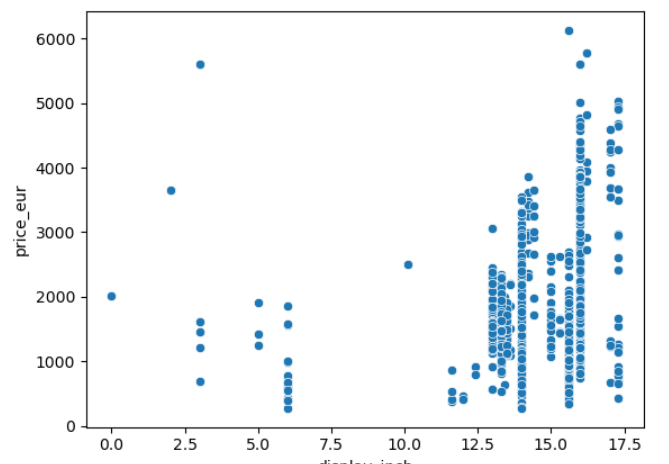
* name: Tên của hãng
* price\_eur:Giá hiện tại bằng Euro
* display\_inch và display\_cm: :Kích thước màn hình tính bằng inch và cm
* display\_resolution và display\_ratio: Độ phân giải màn hình và tỷ lệ khung hình
* weight\_kg, height\_mm, width\_mm, depth\_mm: trọng lượng, chiều cao, chiều rộng, chiều sâu
* operating\_syste, cpu\_processor,gpu\_integrated, 'ram\_memory: Hệ điều hành, CPU, GPU, RAM
* internal\_storage\_gb, ,storage\_type: Loại và kích thước bộ nhớ trong
* battery\_life\_h, battery\_capacity\_wh: tuổi thọ pin và Wh
* psu\_watts: PSU watt
* audio\_system, speakers\_count: Hệ thống âm thanh và số lượng loa
* has\_touchscreen, keyboard\_backlit, keyboard\_numpad, has\_webcam, has\_bluetooth: booleans trên màn hình cảm ứng, đèn nền, numpad, webcam, Bluetooth
* bluetooth\_version:Phiên bản Bluetooth
* wifi\_standard: Tiêu chuẩn WiFi
* product\_ean, product\_sku: EAN và SKU
* release\_year: Năm phát hành (chưa hoàn thành)
* category

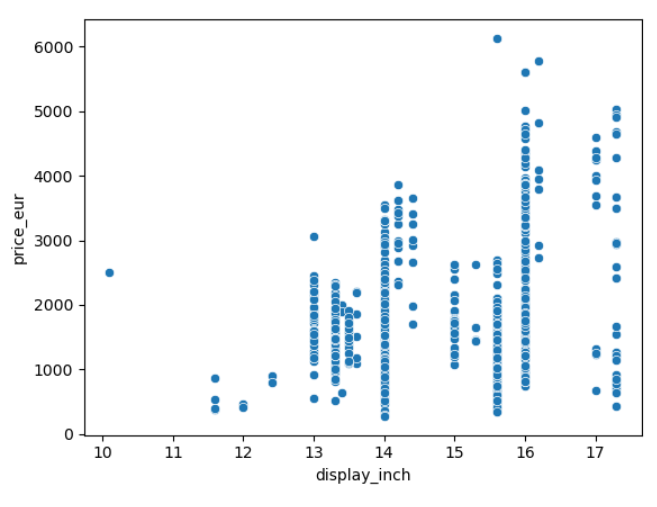
**Mô tả quá trình giả định:**

* Thay thế dữ liệu cột

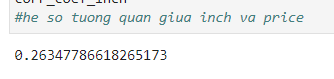


* Biểu đồ phân tán giữa price\_eur và display\_inch

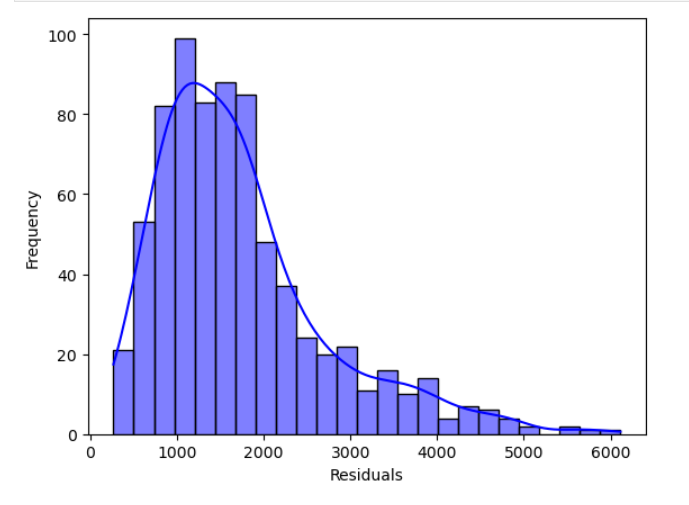




* Tính hệ số tương quan giữa price\_eur và display\_inch

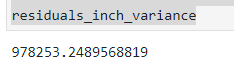


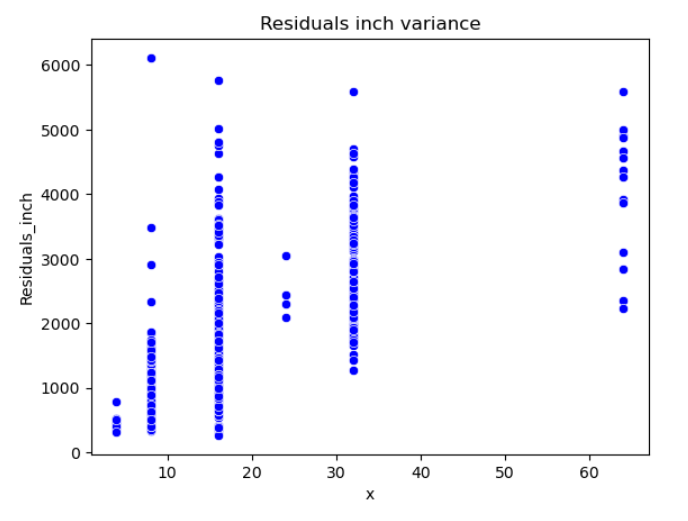
* Vẽ biểu đồ sai số phân phối giữa cột price\_eur và display\_inch



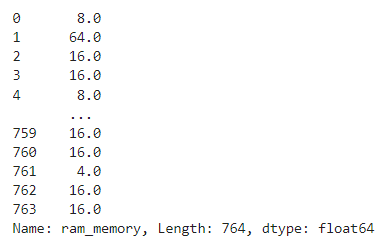
Nhìn biểu đồ ta thấy không có sự phân phối chuẩn giữa price\_eur và display\_inch

* Tính phương sai và vẽ biểu đồ phân tán

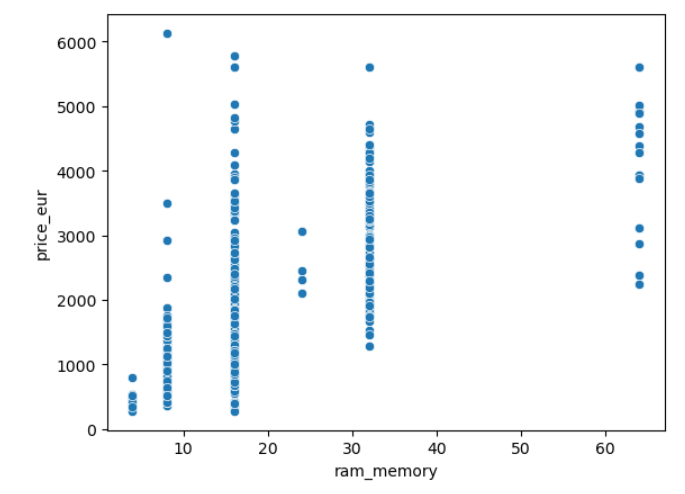




* Lọc và thay thế data ở cột ram\_memory cho đúng dữ liệu

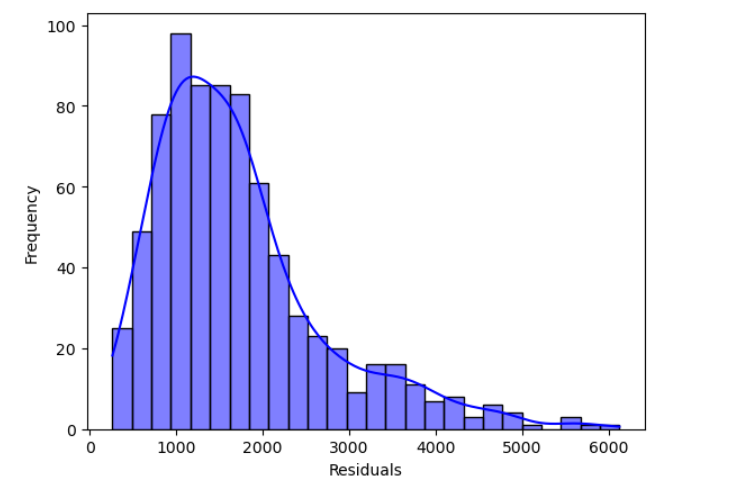


* Biểu đồ phân tán giữa price\_eur và ram\_memory



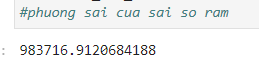
* Tính hệ số tương quan và vẽ biểu đồ sai số phân phối chuẩn

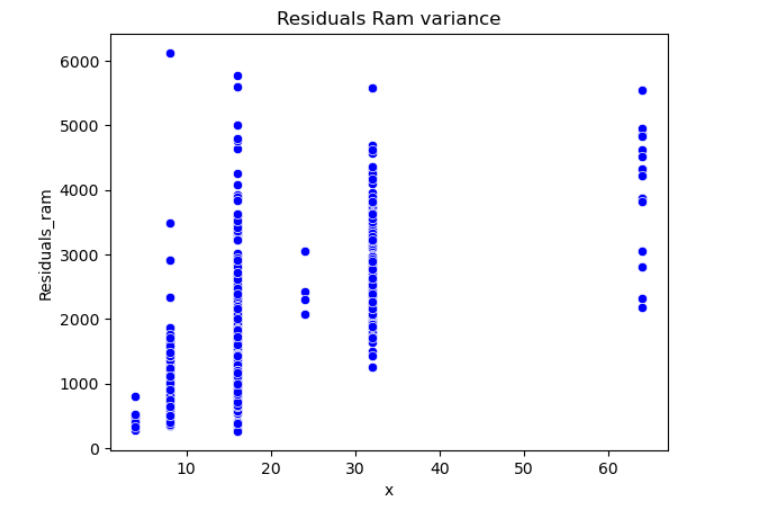




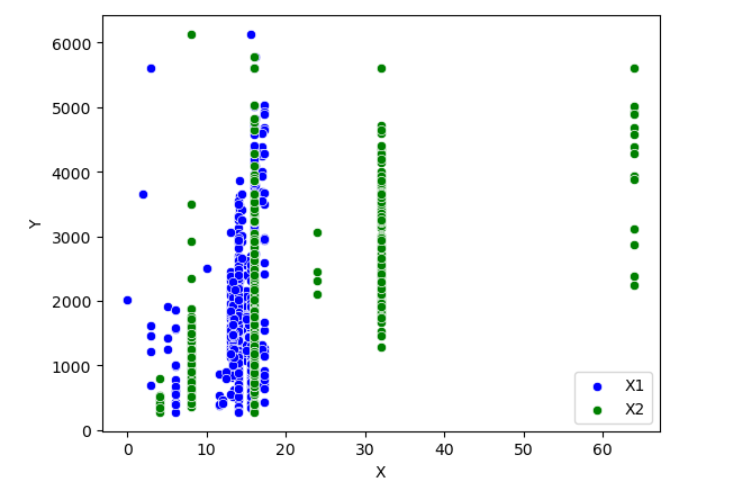
Nhìn biểu đồ ta thấy không có sự phân phối chuẩn giữa price\_eur và ram\_memory

* Tính phương sai và vẽ biểu đồ phân tán





* Vẽ biểu đồ phân tán giữa các biến độc lập



Nhìn biểu đồ, hai biến độc lập có tương quan với nhau.