# Report 02

Thành viên nhóm:

• SE193829 – Đinh Minh Quân

• SE190380 – Võ Anh Trí

Project Name: Finance Fian

## Phần 1: Báo cáo tiến độ

"Hiện giờ nhóm đã làm được gì rồi, có đủ data và code chưa, gặp khó khăn thuận lợi gì?"

### 1. Tiến độ của nhóm

Hiện tại, nhóm đã lấy được data thành công. Tụi mình có thể truy xuất dữ liệu của bất kỳ công ty nào có trên Yahoo Finance, chỉ cần biết ticker là xong.  
  
Nhóm dùng Gemma 2b – một mô hình NLP nhẹ, đơn giản – để phân tích câu hỏi người dùng và trích xuất ra 3 phần chính:  
- Intent Detection – Hiểu người dùng đang muốn gì. Ví dụ: show\_stock, calculate\_rsi, predict\_gcross,...  
- Ticker Detection – Biến tên công ty thành mã cổ phiếu. Ví dụ: Apple → AAPL, Tesla → TSLA,...  
- Period Detection – Nhận diện khoảng thời gian người dùng muốn xem:  
 • Nếu chỉ xem dữ liệu → giữ nguyên thời gian như người dùng nhập.  
 • Nếu dùng cho tính toán/predict → Fian sẽ tự quyết định khoảng thời gian phù hợp.  
  
Ngoài ra, nhờ thư viện ta của Python, nhóm có thể tự động tính hơn 130 chỉ báo kỹ thuật (Technical Indicators). Tuy nhiên tính năng này vẫn đang thử nghiệm, vì khả năng hiểu nhầm yêu cầu từ người dùng vẫn còn.

Về front-end:  
Hiện nhóm đang tính dùng Streamlit cho phần giao diện. Mặc dù không mạnh lắm về mặt thẩm mỹ, nhưng đủ để có một chatbox cho người dùng tương tác với Fian.  
Luồng xử lý như sau: người dùng nhập yêu cầu → gửi về backend → backend xử lý → Gemma 2b trả lời lại một cách tự nhiên.  
  
Ví dụ khi đưa prompt:  
“Bạn hãy trình bày cho khách hàng phân tích của {RSI} cho ra {72.1}%, tình trạng: {quá mua}”  
Gemma 2b sẽ trả lời:  
“Tôi phân tích được rằng, hiện tại đường RSI đang là 72.1%, đang trong tình trạng quá mua.”

### 2. Đối chiếu với kế hoạch ban đầu & những vấn đề gặp phải

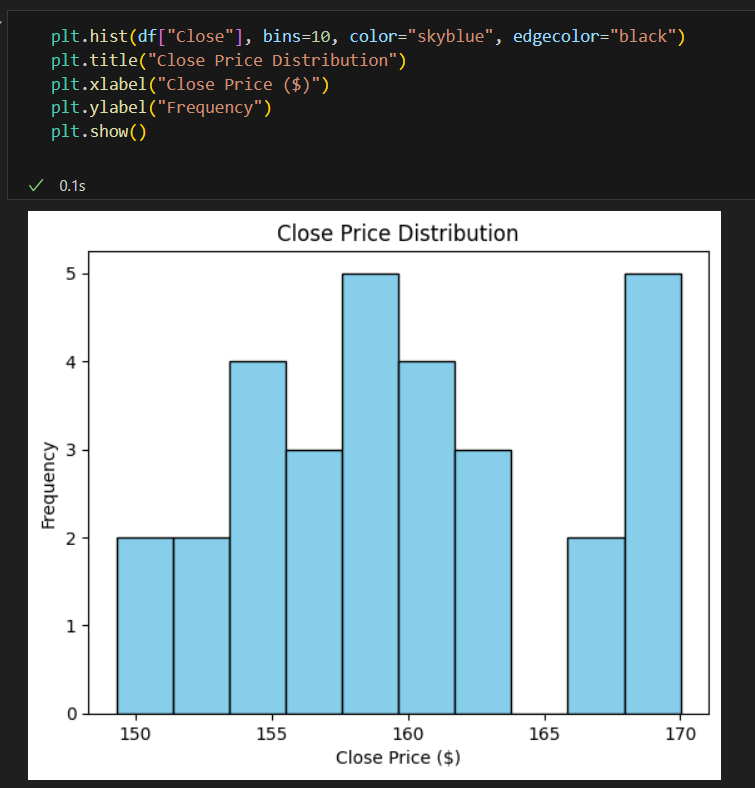
So với plan ban đầu:  
- Tuần 2–3: Data Preparation + bắt đầu chatbot → DONE  
 • Đã trích xuất data bằng yfinance  
 • Đã áp dụng thư viện ta để tính TI  
 • Gemma 2b hoạt động tốt hơn mong đợi, xử lý parsing ổn áp  
  
- Tuần 4: Data Visualization → ĐANG TIẾN HÀNH  
 • Nhóm đang thử nghiệm để xem model nào phù hợp với từng chỉ báo kỹ thuật nhất  
  
→ Nhìn chung, nhóm vẫn đúng tiến độ, thậm chí có một số phần làm trước so với dự kiến.

Thuận lợi:  
- Có thể trích xuất thông tin công ty theo yêu cầu một cách linh hoạt  
- Gemma 2b tuy nhẹ nhưng lại hoạt động hiệu quả bất ngờ  
- Streamlit tỏ ra khá thân thiện, dễ dựng giao diện thử nghiệm

Khó khăn:  
- Có hơn 130 chỉ báo kỹ thuật, mà mỗi cái có cách visualize riêng → Hiện nhóm vẫn chưa rõ nên dùng biểu đồ chung hay phải hard-code từng cái  
- Phần prediction vẫn còn mơ hồ: predict theo chỉ báo nào? Model nào tốt nhất cho từng cái? Không thể predict tất cả 130+ chỉ báo được  
- Gemma 2b dù khá ổn nhưng không phải lúc nào cũng đáng tin. Nhóm vẫn lo lắng nếu nó “đứng hình” lúc demo thật.

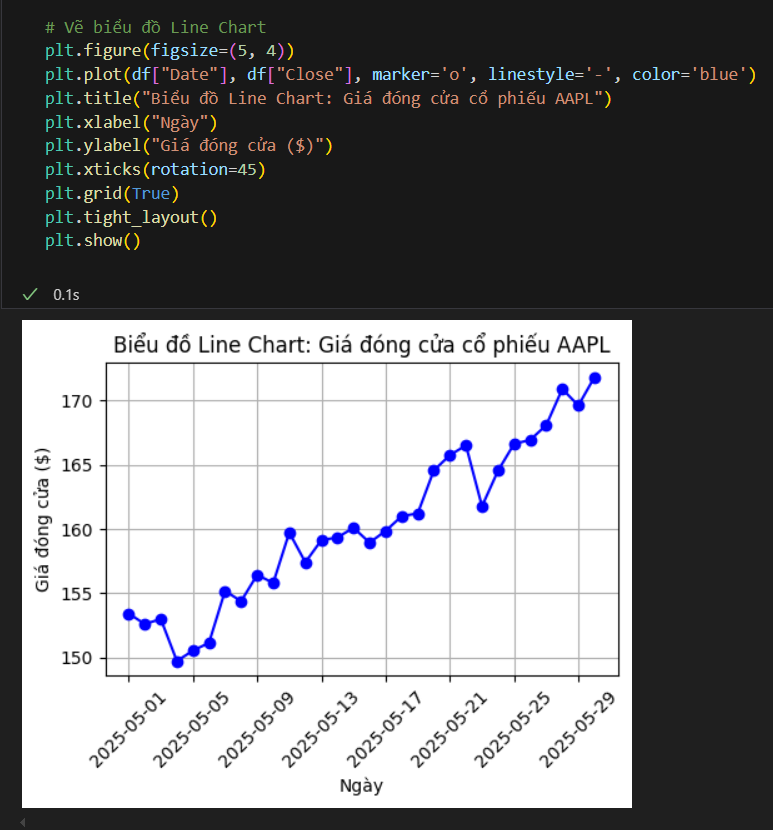
## Phần 2: Bài tập biểu đồ

**1. Bar Chart – Biểu đồ cột**



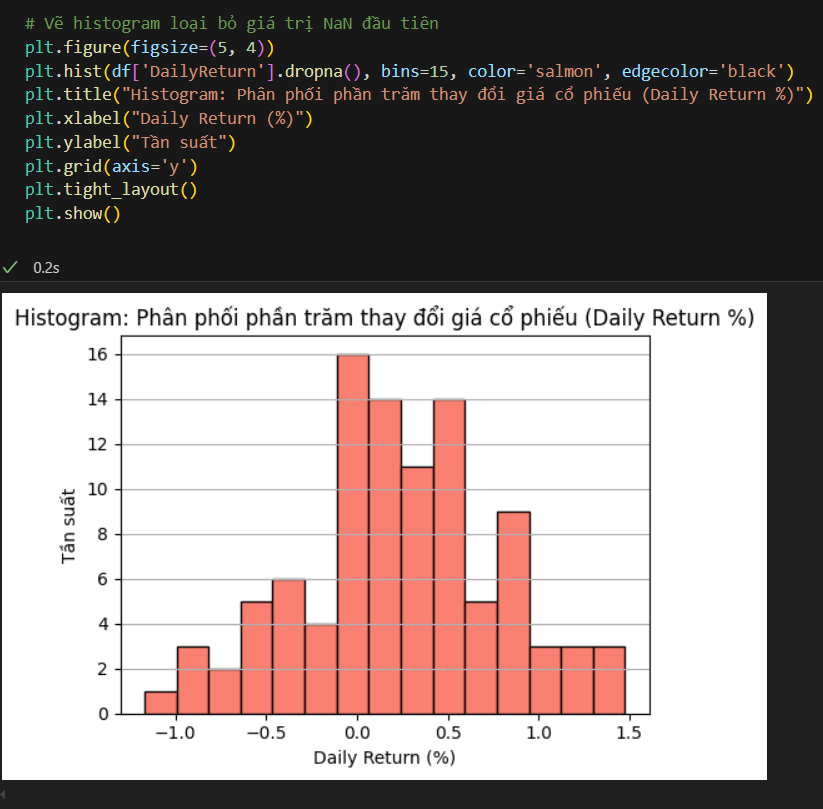
Dùng để làm gì: So sánh định lượng giữa các danh mục (ví dụ: volume giao dịch mỗi ngày).  
Khi nào dùng: Khi bạn muốn nhấn mạnh sự khác biệt giữa các đối tượng rời rạc theo thời gian, theo loại cổ phiếu, hoặc nhóm ngành.

**2. Line Chart – Biểu đồ đường**



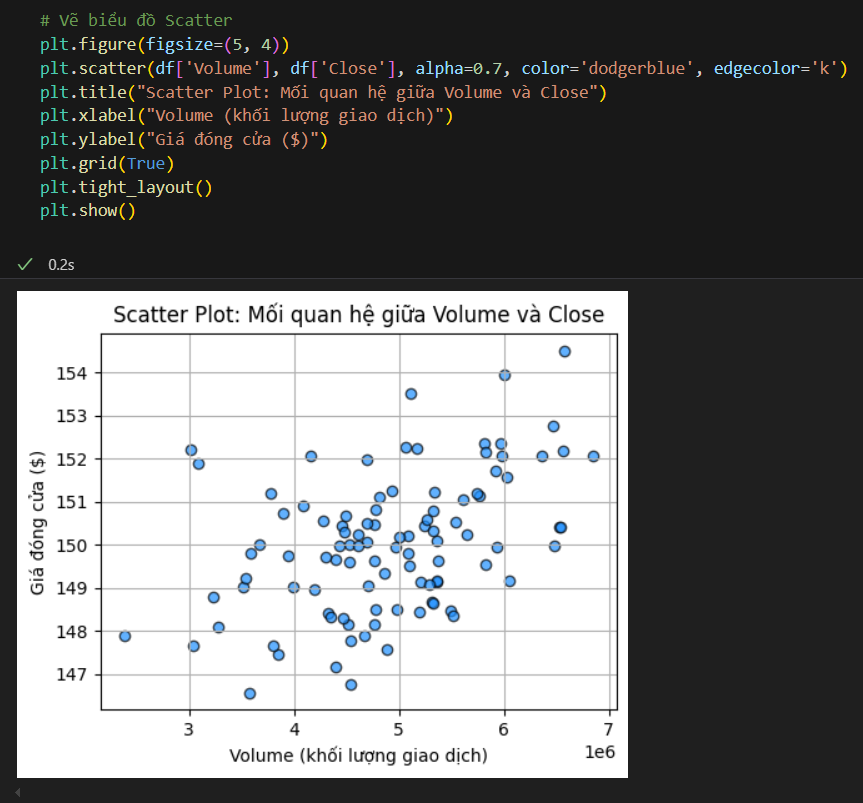
Dùng để làm gì: Theo dõi xu hướng hoặc sự thay đổi theo thời gian.  
Khi nào dùng: Khi bạn có dữ liệu dạng chuỗi thời gian, ví dụ như giá đóng cửa (Close) mỗi ngày.

**3. Histogram – Biểu đồ tần suất**

****

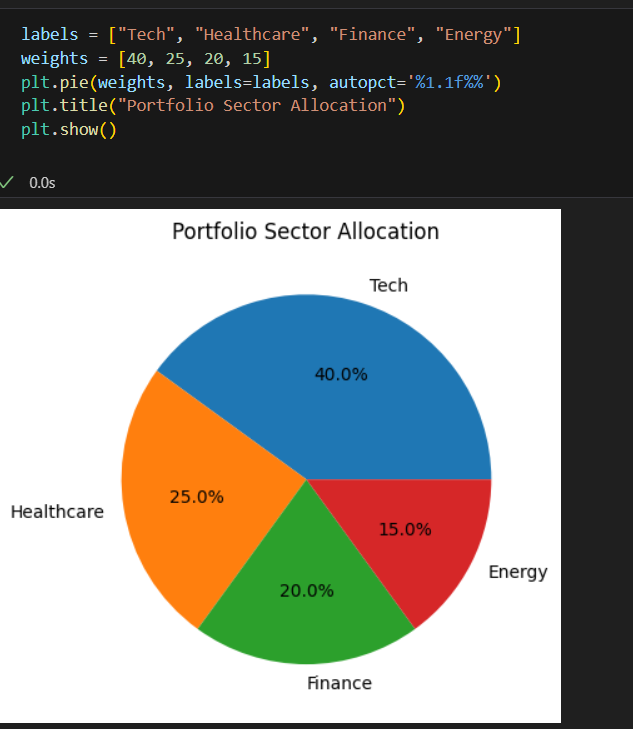
Dùng để làm gì: Phân bố của một biến số (liên tục hoặc rời rạc).  
Khi nào dùng: Khi bạn muốn biết giá trị nào xuất hiện nhiều/lần trong tập dữ liệu (ví dụ như giá Close).

**4. Scatter Plot – Biểu đồ phân tán**

****

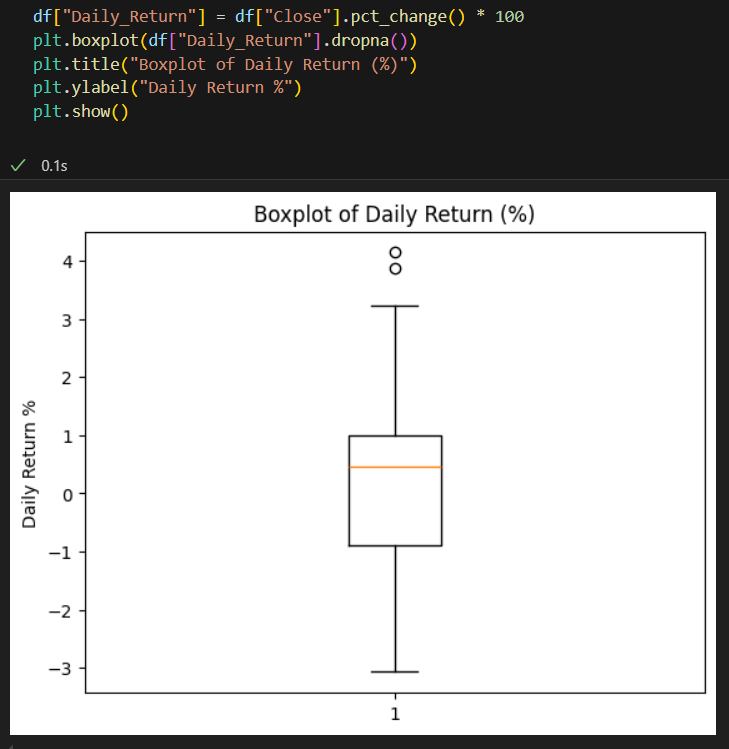
Dùng để làm gì: Xem xét mối tương quan giữa hai biến số.  
Khi nào dùng: Khi bạn muốn xác định xem có quan hệ tuyến tính hay không giữa 2 biến (ví dụ: Volume và Close).

**5. Pie Chart – Biểu đồ tròn**

****

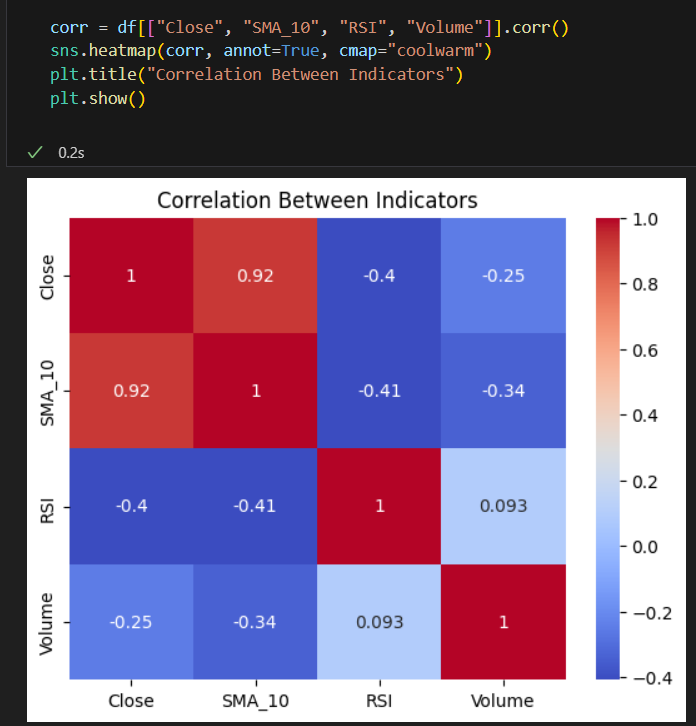
Dùng để làm gì: Thể hiện tỷ lệ phần trăm giữa các phần trong một tổng thể.  
Khi nào dùng: Khi bạn có một biến phân loại thể hiện phân bổ danh mục đầu tư.

**6. Box Plot – Biểu đồ hộp**

****

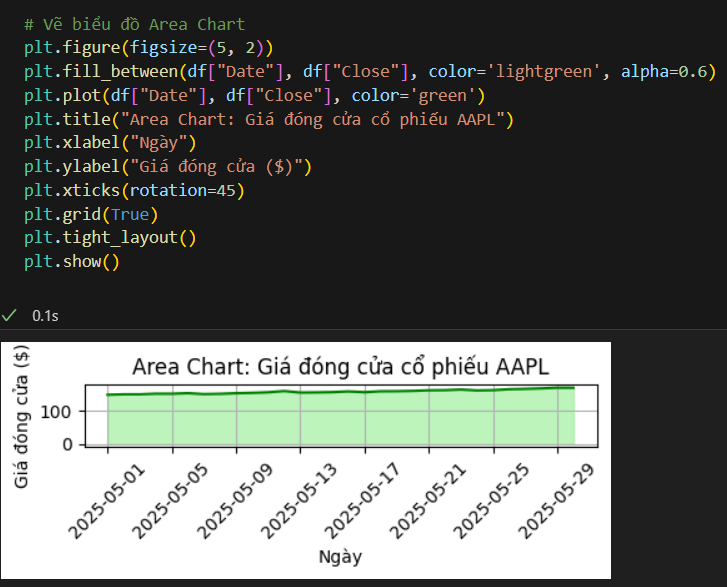
Dùng để làm gì: Thể hiện phân phối, trung vị và phát hiện giá trị ngoại lai.  
Khi nào dùng: Khi cần mô tả đặc trưng phân phối của biến liên tục như phần trăm thay đổi giá mỗi ngày.

**7. Heatmap – Biểu đồ ma trận nhiệt**

****

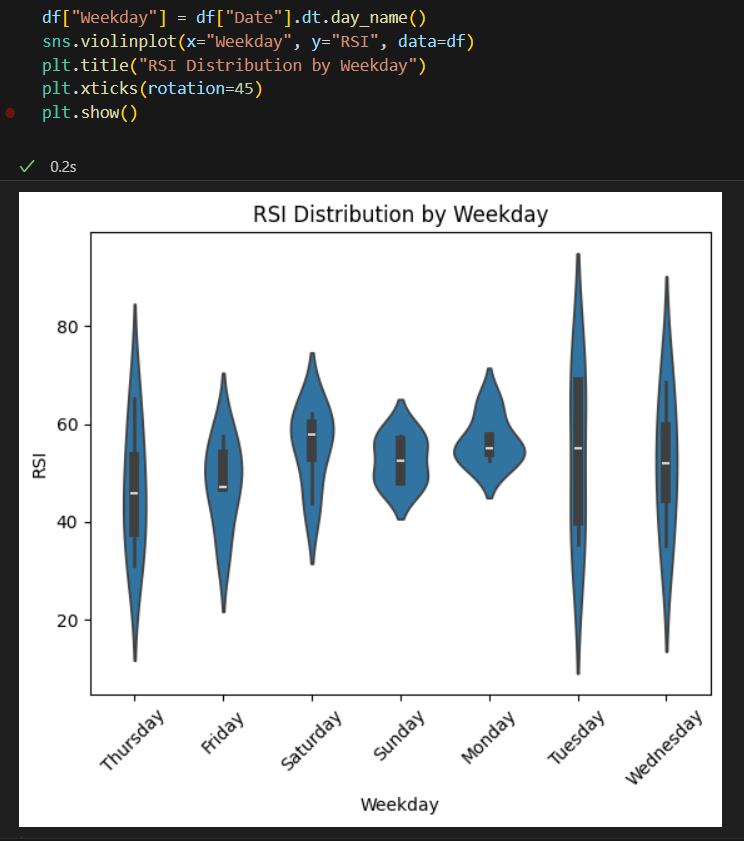
Dùng để làm gì: Thể hiện giá trị của ma trận (như hệ số tương quan) bằng màu sắc.  
Khi nào dùng: Khi bạn có nhiều biến số và muốn đánh giá mối quan hệ giữa chúng.

**8. Area Chart – Biểu đồ vùng**

****

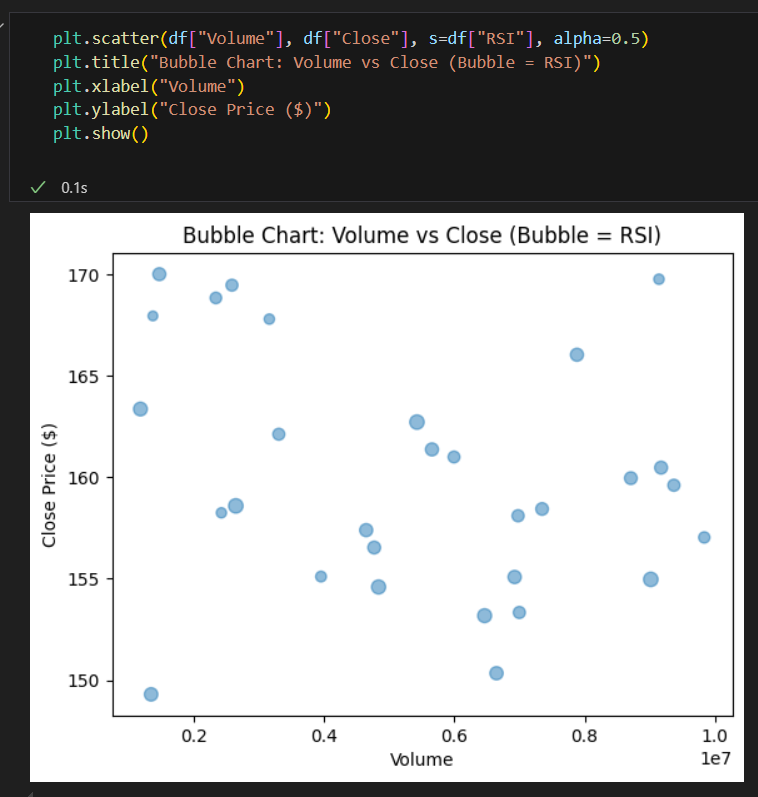
Dùng để làm gì: Giống biểu đồ đường, nhưng nhấn mạnh phần tích lũy hoặc tổng cộng bằng cách tô màu dưới đường.  
Khi nào dùng: Khi cần thể hiện xu hướng tích lũy (ví dụ như lãi gộp).

**9. Violin Plot – Biểu đồ violin**

****

Dùng để làm gì: Kết hợp Boxplot với mật độ phân phối để thể hiện rõ hơn hình dạng phân phối dữ liệu.  
Khi nào dùng: Khi muốn so sánh phân phối của một biến giữa các nhóm.

**10. Bubble Chart – Biểu đồ bong bóng**

****

Dùng để làm gì: Biểu diễn 3 biến cùng lúc (x, y và kích thước bong bóng).  
Khi nào dùng: Khi có thêm một biến muốn biểu diễn như kích thước hoặc độ quan trọng.