

Excel for Data Analysis – Part I

Basic Statistics & Visualization

Lecturer: Nguyễn Thái Hà (Ph.D)

Supporter: Nguyễn Thọ Anh Khoa (Ph.D Candidate)

Date: 02/08/2025 (Sat)



Xây dựng kiến thức phân tích dữ liệu nền tảng qua việc sử dụng Excel

Nắm vững các tính năng và công cụ của Excel từ căn bản đến chuyên sâu cho phân tích dữ liệu



Sử Dụng Excel để Tìm Hiểu Xu Hướng Dữ Liệu

Vận dụng các phương pháp thống kê để nhận diện, phân tích và diễn giải chính xác các xu hướng ẩn trong dữ liệu doanh nghiệp



Sử Dụng Excel Trực Quan Hóa Dữ Liệu

Thành thạo sử dụng công cụ trực quan hóa dữ liệu, tạo ra các biểu đồ và dashboard thuyết phục cho việc ra quyết định

Phần 1: Hiểu Tổng Quan Về Phân Tích Dữ Liệu

Khái niệm cơ bản và ứng dụng Excel trong phân tích dữ liệu.

Phần 2: Hiểu Xu Hướng Dữ Liệu Qua Thống Kê Cơ Bản

Phương pháp thống kê phát hiện xu hướng trong dữ liệu.

Phần 3: Trực Quan Hóa Dữ Liệu

Tạo biểu đồ và dashboard hiệu quả từ dữ liệu.

Phần 1: Hiểu Tổng Quan Về Phân Tích Dữ Liệu

Khái niệm cơ bản và ứng dụng Excel trong phân tích dữ liệu.

Phần 2: Hiểu Xu Hướng Dữ Liệu Qua Thống Kê Cơ Bản

Phương pháp thống kê phát hiện xu hướng trong dữ liệu.

Phần 3: Trực Quan Hóa Dữ Liệu

Tạo biểu đồ và dashboard hiệu quả từ dữ liệu.

Dữ Liệu là gì?

Dữ liệu (data) là các sự kiện thô, các con số, hoặc văn bản chưa được xử lý và tổ chức. Dữ liệu tự thân không mang nhiều ý nghĩa cho đến khi được phân tích và diễn giải.

Dữ liệu (Data)

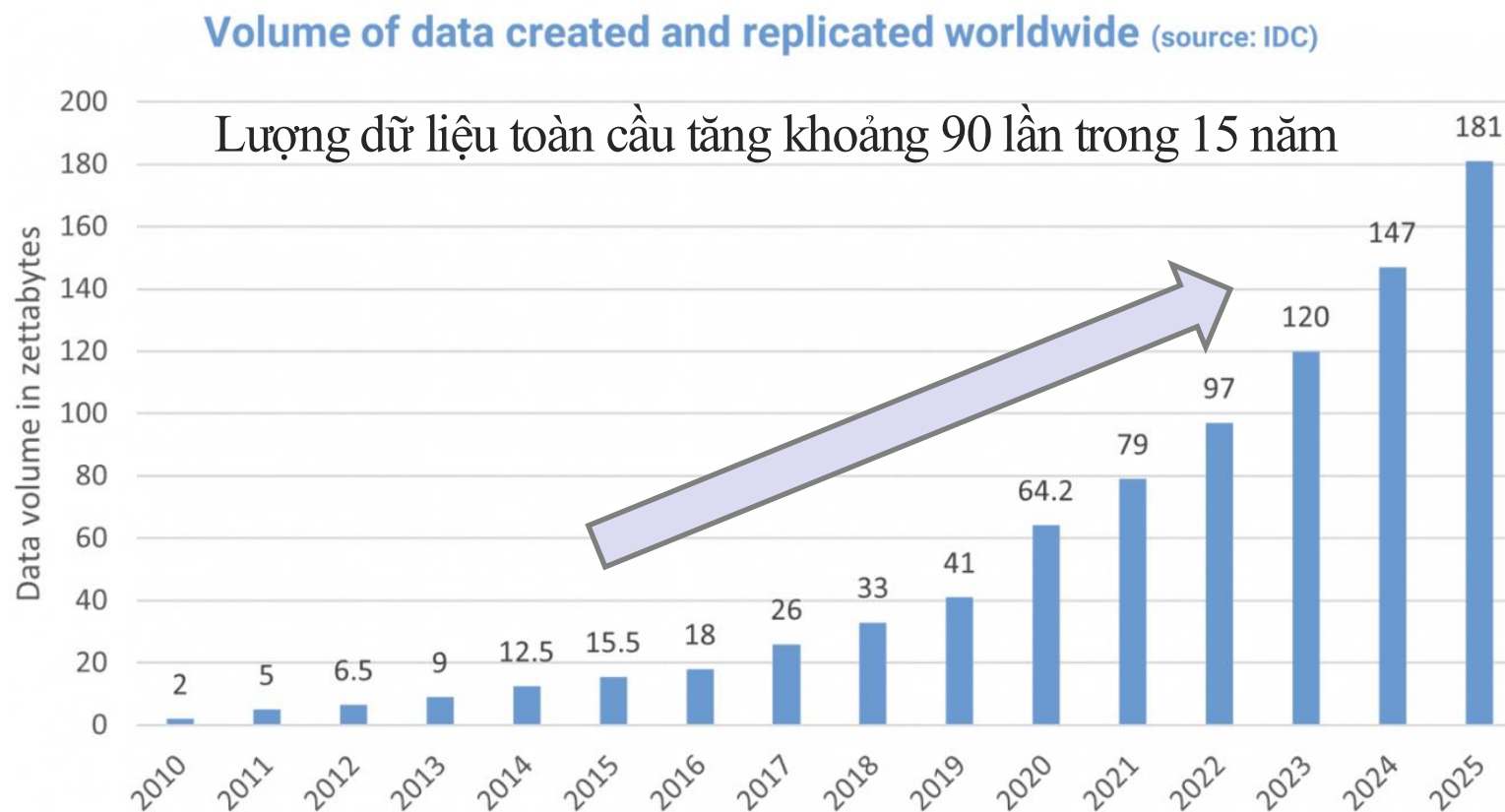
- Dạng thô, chưa qua xử lý
- Thường ở dạng số, ký tự, hình ảnh
- Không có ngữ cảnh hoặc ý nghĩa rõ ràng
- Ví dụ: 15.000.000, 42%, "TPHCM"
- Thường được lưu trữ trong cơ sở dữ liệu

Thông tin (Information)

- Dữ liệu đã được xử lý và tổ chức
- Có ngữ cảnh và ý nghĩa rõ ràng
- Hỗ trợ việc đưa ra quyết định
- Ví dụ: "Doanh thu tháng 6 là 15 triệu đồng, tăng 42% so với TPHCM"
- Được trình bày dưới dạng báo cáo, dashboard

Sự bùng nổ của dữ liệu toàn cầu

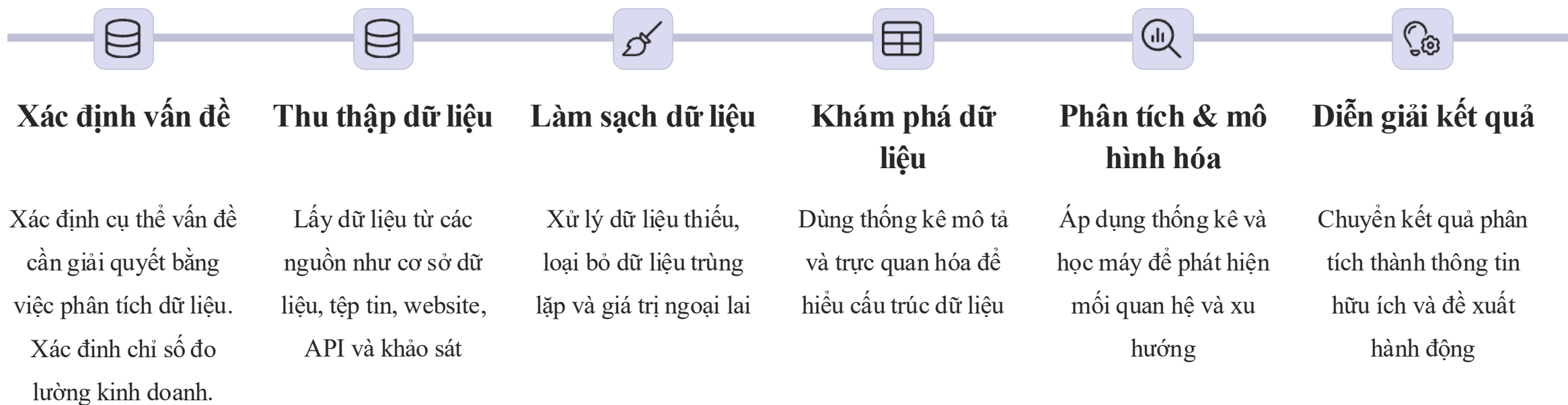
Trong hơn một thập kỷ qua, khối lượng dữ liệu toàn cầu đã tăng trưởng theo cấp số nhân, từ chỉ 2 zettabytes năm 2010 lên đến 181 zettabytes dự kiến vào năm 2025. Phân tích dữ liệu là chìa khoá để khai thác dữ liệu một cách hiệu quả.



Phân Tích Dữ Liệu là gì?

Phân tích dữ liệu là quá trình kiểm tra, biến đổi và phân tích dữ liệu nhằm khám phá thông tin giá trị, rút ra kết luận và hỗ trợ đưa ra quyết định chính xác dựa trên dữ liệu, thay vì cảm tính.

Quy Trình Phân Tích Dữ Liệu



Lợi ích của Phân Tích Dữ Liệu

Phân tích dữ liệu chuyển biến dữ liệu thành thông tin có ích giúp ra quyết định chính xác dựa trên bằng chứng xác thực, nhận diện xu hướng và tối ưu hiệu quả kinh doanh



Ra quyết định dựa trên bằng chứng

Thay vì dựa vào trực giác, các tổ chức sử dụng phân tích dữ liệu để đưa ra quyết định đầu tư và chiến lược dựa trên bằng chứng cụ thể.



Phát hiện xu hướng mới

Phân tích mạng xã hội giúp nhận biết sớm sự thay đổi trong sở thích người tiêu dùng.



Dự đoán nhu cầu thị trường

Sử dụng dữ liệu quá khứ để dự báo nhu cầu và tối ưu hàng tồn kho theo khu vực.



Tối ưu quy trình sản xuất

Áp dụng phân tích dữ liệu để giảm thiểu lãng phí trong sản xuất.



Cá nhân hóa trải nghiệm khách hàng

Phân tích hành vi người dùng giúp đề xuất nội dung phù hợp, nâng tỷ lệ giữ chân khách hàng.

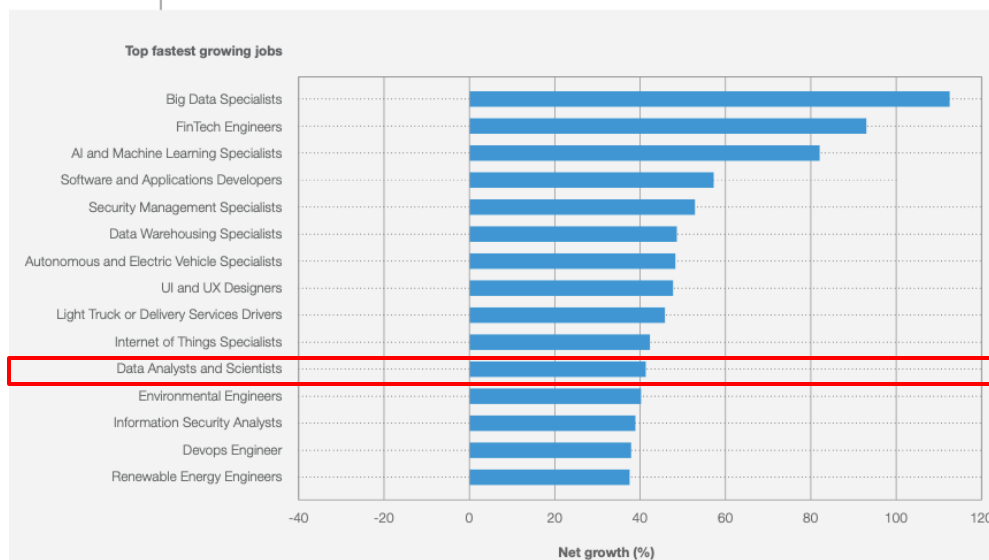
Lý do nên học Phân Tích Dữ Liệu

Phân tích dữ liệu mở ra cơ hội nghề nghiệp và phát triển tư duy phân tích - kỹ năng cốt lõi được nhiều nhà tuyển dụng đánh giá cao.

Cơ hội nghề nghiệp rộng mở

FIGURE 2.2 Fastest-growing and fastest-declining jobs, 2025-2030

Top jobs by fastest net growth and net decline, projected by surveyed employers



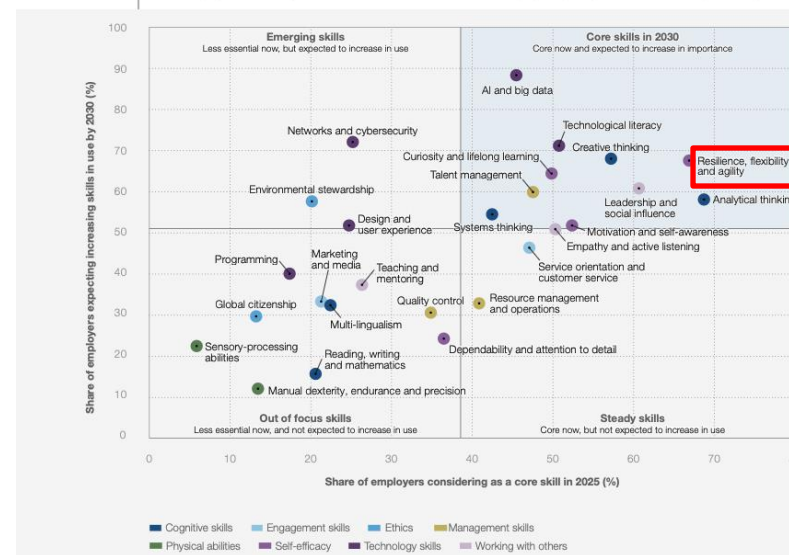
"Các nhà phân tích và khoa học dữ liệu thuộc nhóm nghề phát triển nhanh nhất" - *Future of Jobs Report*, tr.19

"Tại Nhật Bản, Chuyên gia An ninh thông tin và Phân tích dữ liệu được dự đoán là nghề tăng trưởng hàng đầu" - *Future of Jobs Report*, tr.65

Tư duy phân tích là kỹ năng được săn đón nhất

FIGURE 3.6 Core skills in 2030




Share of employers considering skills to be a core skill in 2025 and share of employers expecting skills to increase in importance by 2030.



"Tư duy phân tích vẫn là kỹ năng cốt lõi được các nhà tuyển dụng săn đón nhất" - *Future of Jobs Report*, tr.6

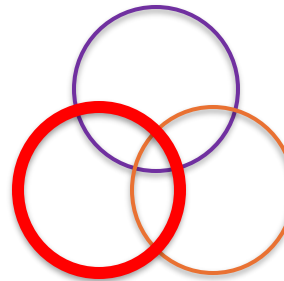
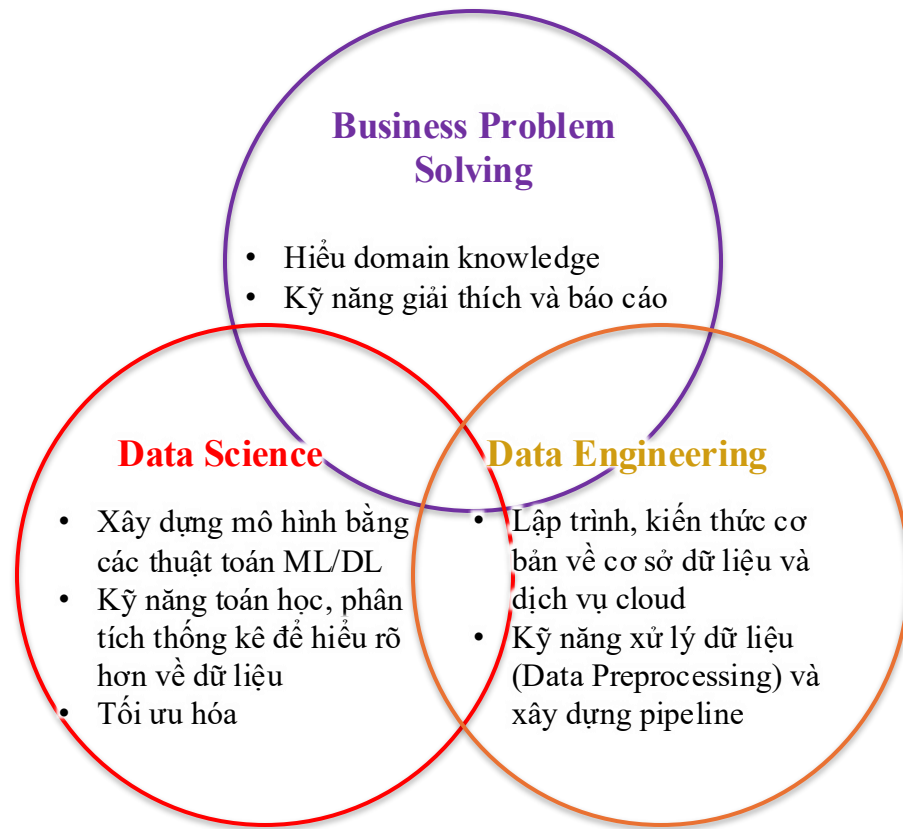
Công việc trong ngành phân tích dữ liệu

Ngành phân tích dữ liệu thường chia thành ba vai trò chính, mỗi vai trò có thế mạnh riêng và cùng nhau tạo nên một hệ sinh thái dữ liệu toàn diện.

Vai trò	Vai trò	Điểm mạnh	Công cụ phổ biến
Data scientist 	<ul style="list-style-type: none">Phân tích dữ liệu nâng cao để xây dựng mô hình và hỗ trợ đưa ra các quyết định kinh doanh	<ul style="list-style-type: none">Tư duy về dữ liệu tốt, khả năng làm việc với nhiều loại dữ liệuCó kỹ năng xây dựng mô hình và tối ưu mô hình	<ul style="list-style-type: none">Excel, Python, SQL, R, SparkCloud (AWS, GCP, Azure, Databricks)
Data Analyst BI Analyst 	<ul style="list-style-type: none">Phân tích dữ liệu để tạo ra các báo cáoCung cấp các insight về dữ liệu để hỗ trợ quyết định kinh doanh	<ul style="list-style-type: none">Khả năng sử dụng đa dạng các công cụ phân tích như Excel, SQL, Power BI và TableauKỹ năng mềm tốtAm hiểu lĩnh vực kinh doanh	<ul style="list-style-type: none">Excel, SQL, Tableau, Power BI, PythonPower Point, Presentation Tools
Data Engineer 	<ul style="list-style-type: none">Xây dựng pipeline dữ liệuXử lý, chuyển đổi dữ liệu (ETL)Quản lý hệ thống dữ liệu	<ul style="list-style-type: none">Kỹ năng lập trình tốt, am hiểu hệ thống cơ sở dữ liệu, có khả năng xử lý dữ liệu lớnKhả năng xây dựng kiến trúc dữ liệu và đảm bảo dữ liệu được ổn định	<ul style="list-style-type: none">Python, SQL, Spark, AirflowCloud (AWS, GCP, Azure, Databricks)

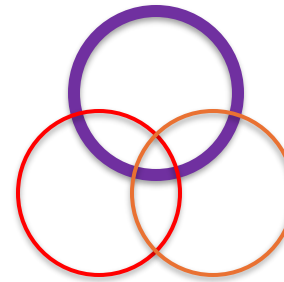
Bộ kỹ năng cần thiết tương ứng với từng vị trí AI VIETNAM @aivietnam.edu.vn

Công việc trong ngành phân tích dữ liệu thường yêu cầu 3 kỹ năng chính: Am hiểu về lĩnh vực chuyên môn, kỹ thuật xử lý dữ liệu, và kỹ năng về khoa học dữ liệu. Yêu cầu và phân bổ sẽ thay đổi dựa vào vị trí.



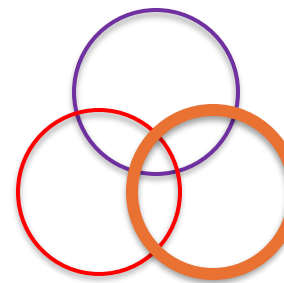
Data Scientist

Trọng tâm: Thiên về kỹ thuật – Xây dựng mô hình hoá – Phân tích thống kê sâu để hiểu dữ liệu - Tối ưu hoá



Data Analyst/Business Intelligence Analyst

Trọng tâm: Hiểu và phân tích bài toán kinh doanh – Truy vấn dữ liệu (SQL) - Trực quan hóa - Báo cáo – Giao tiếp và truyền đạt kết quả với bộ phận kinh doanh



Data Engineer

Trọng tâm: Xây dựng pipeline xử lý dữ liệu – Làm việc với cloud (AWS, GCP, Azure, Databricks) – Tối ưu hoá cơ sở dữ liệu và truy xuất dữ liệu

Đối tượng nên học phân tích dữ liệu

Phân tích dữ liệu là kỹ năng thiết yếu trong thời đại số, không chỉ giới hạn cho chuyên gia hay những người có chuyên môn về dữ liệu.



Nhân viên văn phòng

Tự động hóa quy trình, nâng cao giá trị chuyên môn và tạo lợi thế cạnh tranh trong công việc



Chủ doanh nghiệp nhỏ

Tối ưu chi phí, hiểu hành vi khách hàng và xây dựng chiến lược kinh doanh dựa trên dữ liệu thực tế



Sinh viên và học sinh

Phát triển kỹ năng được nhà tuyển dụng sẵn đón, nâng cao khả năng nghiên cứu và chuẩn bị cho kỷ nguyên số hoá



Tất cả mọi người

Đưa ra quyết định dựa trên dữ liệu thay vì cảm tính, phát triển tư duy phản biện và giải quyết vấn đề hiệu quả

Các trường hợp ứng dụng phân tích dữ liệu

Phân tích dữ liệu được ứng dụng trong nhiều lĩnh vực với mục tiêu cụ thể:

Ngành giải trí

- Netflix phân tích thói quen xem để đề xuất nội dung
- Spotify tạo danh sách phát cá nhân hóa
- Disney tối ưu trải nghiệm công viên từ phản hồi

Thể thao

- NBA sử dụng dữ liệu cho chiến thuật và tuyển dụng
- CLB bóng đá theo dõi hiệu suất qua dữ liệu GPS
- Phân tích thống kê tối ưu chiến thuật thi đấu

Thương mại

- Amazon đề xuất sản phẩm từ lịch sử mua hàng
- Walmart tối ưu hàng tồn kho từ dữ liệu bán hàng
- Starbucks cá nhân hóa chương trình khuyến mãi

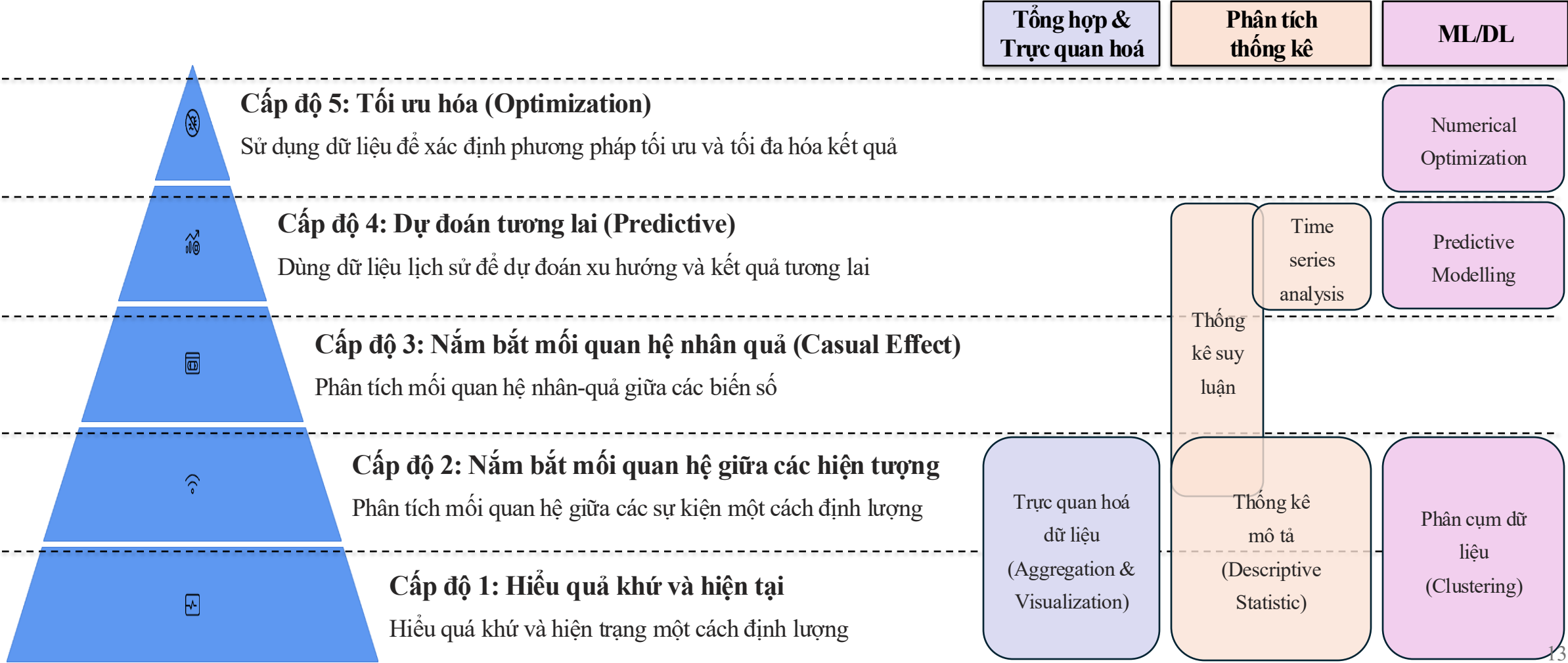
Y tế và giáo dục

- Bệnh viện dự đoán tỷ lệ tái nhập viện
- Trường học theo dõi tiến bộ để can thiệp sớm
- Cơ quan chính phủ phân bổ nguồn lực hiệu quả

Tham Khảo: Coursera, Data Analytics Foundations – Module 1

5 Cấp Độ Của Việc Phân Tích Dữ Liệu

Phân tích dữ liệu phát triển qua 5 cấp độ tăng dần về độ phức tạp: từ theo dõi cơ bản, nhận diện mối quan hệ, phân tích nhân quả, đến dự báo tương lai và tối ưu hóa.



Lý do sử dụng Excel trong Phân Tích Dữ Liệu

Excel là công cụ cơ bản để phân tích dữ liệu, tạo hình ảnh trực quan và là kỹ năng quan trọng cho nhiều công việc



Công Cụ Tính Toán Phân Tích Thông Dụng

Excel được sử dụng rộng rãi, giúp bạn phân tích dữ liệu nhanh chóng.



Tạo Biểu Đồ Từ Dữ Liệu

Dễ dàng tạo biểu đồ giúp hiểu và trình bày thông tin rõ ràng hơn.



Mô Phỏng Đơn Giản

Giúp bạn hiểu các khái niệm xác suất và thống kê qua thực hành.



Tự Động Hóa Công Việc

Dùng macro để làm nhanh các công việc lặp đi lặp lại, tiết kiệm thời gian.

Excel vs Ngôn Ngữ Lập Trình Cho Phân Tích Dữ Liệu

Trong thực tế, nhiều người phân tích dữ liệu dùng Excel cho phân tích nhanh, đơn giản và ngôn ngữ như Python, R hoặc SQL cho việc phức tạp hơn.

Nên Dùng Excel Khi

- Dữ liệu nhỏ hoặc vừa (dưới 1 triệu dòng)
- Cần phân tích nhanh và tạo biểu đồ đơn giản
- Làm việc với người không chuyên môn về kỹ thuật
- Làm các phép tính thống kê cơ bản
- Cần tạo báo cáo nhanh và dễ chia sẻ
- Không cần tự động hóa phức tạp

Nên Dùng Ngôn Ngữ Lập Trình Khi

- Dữ liệu lớn (hàng triệu dòng trở lên)
- Cần tạo mô hình phức tạp (machine learning, AI)
- Cần tự động hóa công việc thường xuyên
- Cần kết nối với hệ thống và API khác
- Phân tích phức tạp với nhiều nguồn dữ liệu
- Muốn tạo ứng dụng hoặc bảng thông tin tương tác



Cài đặt công cụ phân tích

Cài đặt và kích hoạt "Data Analysis Tool" và "Solver" để mở rộng khả năng phân tích dữ liệu của Excel.



Thiết lập môi trường làm việc

Tùy chỉnh thanh công cụ và trang tính với các công thức thường xuyên sử dụng để tăng hiệu quả làm việc.



Tổ chức dữ liệu

Cấu trúc dữ liệu thành bảng có định dạng rõ ràng với cột tiêu đề và loại dữ liệu phù hợp.



Add-ins bổ sung

Cài đặt thêm các tiện ích mở rộng như Power Query và Power Pivot để xử lý dữ liệu phức tạp.

Thất Bại Khi Dùng Công Cụ Không Phù Hợp

Chọn đúng công cụ phù hợp với quy mô dữ liệu và thời gian. Với dữ liệu nhỏ và thời gian hạn chế, Excel thường là lựa chọn hiệu quả hơn so với viết code Python.



Yêu Cầu Ban Đầu

Sếp yêu cầu phân tích bộ dữ liệu nhỏ (50.000 dòng) về doanh số trong vòng 2 giờ để chuẩn bị cho cuộc họp



Lựa Chọn Sai Công Cụ

Quyết định sử dụng Python thay vì Excel dù dữ liệu đơn giản và thời gian hạn chế



Gặp Lỗi Trong Code

Code xử lý dữ liệu bị lỗi cú pháp và import thư viện, mất thời gian để debug



Kết Quả

Trễ deadline, không có kết quả phân tích để trình bày, trong khi Excel có thể hoàn thành trong 30 phút

Phân tích dữ liệu cần dựa trên sự thật, phân biệt rõ giữa dữ liệu và diễn giải, bắt đầu bằng câu hỏi đúng, và luôn xác định trước kết quả mong muốn.



Phân biệt rõ ràng giữa Fact và Inference

Báo cáo kết quả phân tích dữ liệu phải dựa trên dữ liệu thực tế, được kiểm chứng và có thể xác minh.



Phân biệt rõ ràng giữa sự thật và suy luận

Luôn chỉ rõ đâu là dữ liệu thực tế và đâu là kết luận dựa trên sự diễn giải của bạn để người đọc không bị nhầm lẫn.



Bắt đầu bằng câu hỏi "Tại sao"

Trước khi phân tích, hãy đặt câu hỏi "Tại sao" chúng ta cần phân tích dữ liệu này và chúng ta muốn tìm hiểu điều gì.



Phác thảo kết quả mong muốn

Xác định trước các kết quả dự kiến để định hướng quá trình phân tích và đảm bảo đạt được mục tiêu đề ra.

Phân biệt rõ ràng giữa Fact và Inference

Phân tích dữ liệu đáng tin cậy phải dựa trên sự thật rõ ràng, chính xác từ nguồn đáng tin cậy. Cần phân biệt rõ giữa dữ liệu thực tế và diễn giải trong báo cáo.

- **Sự thật (Fact):** Thông tin khách quan, có thể kiểm chứng.
- **Suy luận (Inference):** Kết luận dựa trên diễn giải dữ liệu, thường mang tính chủ quan.

Ví dụ trong ngành đầu tư tài chính

- **Sự thật (Fact):** “Giá cổ phiếu Công ty B đã tăng từ 40.000 VNĐ lên 52.000 VNĐ trong vòng 1 tháng, tương đương mức tăng 30%.”
- **Suy luận (Interpretation):** “Giá cổ phiếu tăng mạnh chứng tỏ Công ty B sắp được một quỹ đầu tư lớn rót vốn.”

Hậu quả có thể xảy ra

- **Gây hiểu lầm cho nhà đầu tư:** Nhà đầu tư không phân biệt đâu là dữ liệu, đâu là nhận định, có thể ra quyết định mua cổ phiếu chỉ vì một giả định chưa được xác thực.
- **Dẫn đến thua lỗ tài chính:** Nếu suy luận sai (ví dụ: không có quỹ nào rót vốn cả), cổ phiếu giảm mạnh sau đó → nhà đầu tư lỗ nặng vì quyết định dựa trên "sự thật tưởng tượng".

Bắt Đầu Bằng Việc Đặt Câu Hỏi

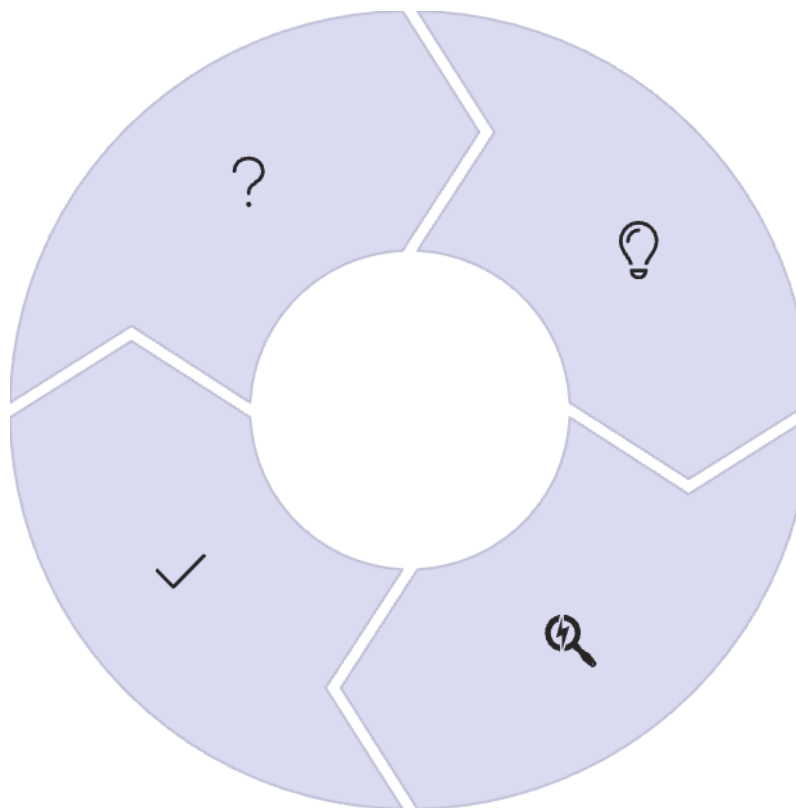
Phân tích dữ liệu hiệu quả bắt đầu từ câu hỏi đúng, giả thuyết rõ ràng và kiểm chứng có phương pháp để ra quyết định chính xác.

Đặt Câu Hỏi

Xác định câu hỏi đúng để định hướng phân tích và xác định thông tin cần thiết.

Kiểm Chứng

Phân tích dữ liệu để kiểm chứng giả thuyết và ra quyết định dựa trên bằng chứng.



Đặt Giả Thuyết

Xây dựng giả thuyết từ quan sát ban đầu để tạo hướng đi cụ thể.

Xác Định Cụ Thể

Thu hẹp phạm vi phân tích để tìm kiếm câu trả lời chính xác.

Bắt Đầu Bằng Việc Đặt Câu Hỏi

Nếu không đặt câu hỏi đầu tiên, cửa hàng có thể mất thời gian xem xét quá nhiều dữ liệu không cần thiết. Với câu hỏi rõ ràng, họ tập trung đúng vấn đề và nhanh chóng điều chỉnh giá hoặc chiến lược bán hàng.



1. Đặt Câu Hỏi

Xác định rõ vấn đề cần tìm hiểu: "Tại sao doanh số cửa hàng A giảm 15% trong quý vừa qua?"



2. Đặt Giả Thuyết

Các lý do có thể giải thích vấn đề: "Có thể do đối thủ mới, giá thay đổi, hoặc khách hàng không hài lòng."



3. Xác Định Cụ Thể

Phân tích dữ liệu: "Cần xem dữ liệu bán hàng theo sản phẩm, thời gian, đánh giá khách hàng, và giá đối thủ."



4. Kiểm Chứng

"Phân tích cho thấy doanh số giảm ở một loại sản phẩm, đúng lúc đối thủ giảm giá 20%."

Phác thảo kết quả mong muốn trước khi phân tích

Xác định rõ mục tiêu trước khi phân tích giúp tiết kiệm thời gian, nâng cao hiệu quả và định hướng đúng cho quá trình làm việc. Điều này đảm bảo thu thập đúng dữ liệu và áp dụng phương pháp phù hợp.



Phác hoạ kết quả

Xác định đầu ra cần thiết: biểu đồ, bảng tổng hợp hoặc báo cáo định hướng.



Thu thập dữ liệu

Tập hợp, làm sạch và chuẩn hóa dữ liệu với công cụ phù hợp.



Trực quan và insight

Tạo biểu đồ, bảng tổng hợp để hiển thị kết quả và rút ra thông tin cho quyết định.

Phân tích doanh số bán hàng

Case study: Tổng hợp và phân tích doanh số theo sản phẩm, khu vực, và thời gian để đưa ra chiến lược kinh doanh hiệu quả.

Công cụ sử dụng:

- Sử dụng **SUMIFS** để tính tổng doanh số theo nhiều điều kiện
- Áp dụng **Pivot Table** để phân tích xu hướng theo quý/tháng
- Tạo **Dashboard** với biểu đồ so sánh hiệu suất bán hàng

Quản lý ngân sách và dự báo tài chính

Case study: Theo dõi chi phí, dự báo ngân sách, và tính toán chỉ số tài chính để kiểm soát tình hình tài chính doanh nghiệp.

Công cụ sử dụng:

- Dùng công thức **IF** và **VLOOKUP** để tự động phân loại chi phí
- Áp dụng **FORECAST** để dự báo doanh thu trong tương lai
- Sử dụng **Conditional Formatting** để đánh dấu các khoản chi vượt ngân sách

Thiết Kế Phân Tích Dữ Liệu Framework 5W2H



Framework 5W2H Trong Excel: Framework 5W2H giúp cấu trúc quá trình phân tích dữ liệu một cách logic và toàn diện.

Framework 5W2H

- **What (Cái gì):** Xác định vấn đề và mục tiêu cần phân tích
- **Why (Tại sao):** Lý do thực hiện phân tích, giá trị mang lại
- **Who (Ai):** Đối tượng liên quan và người dùng kết quả
- **When (Khi nào):** Khung thời gian phân tích và deadline
- **Where (Ở đâu):** Nguồn dữ liệu và phạm vi phân tích
- **How (Làm thế nào):** Phương pháp và công cụ Excel sử dụng
- **How much (Bao nhiêu):** Chi phí và nguồn lực cần thiết

Ví dụ: Phân tích doanh số cửa hàng bán lẻ

- **What:** Phân tích nguyên nhân doanh số giảm 15% trong quý gần nhất
- **Why:** Để xác định chiến lược cải thiện doanh số và điều chỉnh danh mục sản phẩm
- **Who:** Phòng kinh doanh và quản lý cấp cao sẽ sử dụng kết quả
- **When:** Dữ liệu 3 quý gần nhất, cần kết quả trong 1 tuần
- **Where:** Dữ liệu từ hệ thống POS và báo cáo bán hàng (Excel files)
- **How:** Sử dụng Pivot Tables, VLOOKUP, và biểu đồ so sánh trong Excel
- **How much:** Cần 1 chuyên viên phân tích làm việc toàn thời gian trong 1 tuần

Phần 1: Hiểu Tổng Quan Về Phân Tích Dữ Liệu

Khái niệm cơ bản và ứng dụng Excel trong phân tích dữ liệu.

Phần 2: Hiểu Xu Hướng Dữ Liệu Qua Thống Kê Cơ Bản

Phương pháp thống kê phát hiện xu hướng trong dữ liệu.

Phần 3: Trực Quan Hóa Dữ Liệu

Tạo biểu đồ và dashboard hiệu quả từ dữ liệu.

Tại sao phải hiểu xu hướng dữ liệu?

Hiểu xu hướng dữ liệu không chỉ là nền tảng cho quyết định kinh doanh đúng đắn mà còn giúp doanh nghiệp dự đoán thay đổi, phát hiện bất thường và tối ưu hóa hiệu suất hoạt động.

Case study: Hai cửa hàng điện tử A và B cùng có doanh số trung bình 50 triệu/tháng trong Q1/2025

Cửa hàng A

- Doanh số ổn định: 48-52 triệu đồng/tháng (biến thiên $\pm 4\%$)
- Biến động nhỏ theo tuần (độ lệch chuẩn: 1.2 triệu)
- Khách hàng thân thiết: 70% doanh số (trung bình 35 triệu/tháng)
- Lợi nhuận biên: 22% (cao hơn mức trung bình ngành 3%)

⇒ Cần chiến lược chăm sóc khách hàng hiện tại và chương trình loyalty với mục tiêu tăng giá trị đơn hàng trung bình thêm 15%

Cửa hàng B

- Doanh số dao động mạnh: 30-70 triệu đồng/tháng (biến thiên $\pm 40\%$)
- Tăng 60% vào cuối tuần, giảm 35% đầu tuần (mẫu hình tuần rõ rệt)
- Khách hàng mới: 60% doanh số (trung bình 30 triệu/tháng)
- Lợi nhuận biên: 18% (thấp hơn do chi phí marketing cao)

⇒ Cần điều chỉnh nhân sự theo giờ cao điểm, tối ưu chiến dịch marketing theo ngày, và phát triển chiến lược chuyển đổi khách hàng mới thành khách hàng thường xuyên

Không Chỉ Là Giá Trị Trung Bình

Các chỉ số thống kê cơ bản cho ta nhiều góc nhìn khác nhau về dữ liệu. Chỉ dùng mỗi giá trị trung bình khi phân tích có thể gây hiểu nhầm.



Trung bình (Mean)

Là nền tảng quan trọng của phân tích dữ liệu nhưng không nên chỉ dựa vào chỉ số này.

Trong Excel: Sử dụng hàm **AVERAGE(range)** hoặc **=SUM(range)/COUNT(range)** để tính giá trị trung bình.



Trung vị (Median)

Giúp bạn thấy rõ hơn về phân bố dữ liệu, đặc biệt khi có giá trị ngoại lệ.

Trong Excel: Áp dụng hàm **MEDIAN(range)** để xác định giá trị nằm ở vị trí trung tâm của tập dữ liệu.



Phương sai và độ lệch chuẩn

Cho biết mức độ phân tán của dữ liệu, giúp tránh những kết luận sai lầm.

Trong Excel: Dùng **VAR.P(range)** cho phương sai và **STDEV.P(range)** cho độ lệch chuẩn của toàn bộ dữ liệu (hoặc **VAR.S/STDEV.S** cho mẫu).

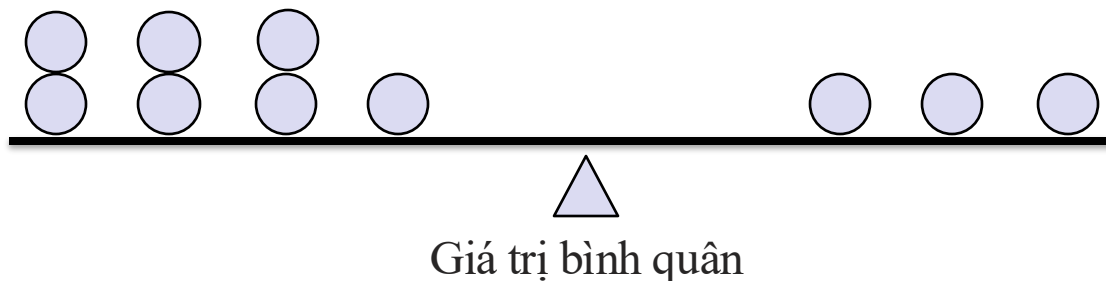
Hình ảnh giá trị trung bình

Giá trị trung bình (Mean) là tổng các số chia cho số lượng phần tử. Khi giá trị bình quân được phác hoạ bằng hình ảnh ta có thể hình dung giá trị trung bình chính là điểm cân bằng.

⌚ Công thức toán học

$$\text{Mean} = \frac{\text{Tổng dữ liệu}}{\text{Số lượng dữ liệu (n)}}$$

⌚ Hình ảnh tưởng tượng



Hạn chế khi chỉ nhìn vào giá trị trung bình

- **Dễ bị mất thông tin quan trọng:** Nếu chỉ nhìn vào giá trị trung bình (ví dụ 50 triệu), ta sẽ không biết được giá trị thấp nhất hay cao nhất là bao nhiêu, hay doanh số có ổn định hay dao động mạnh

Dễ bị hiểu nhầm

- Nó không phải lúc nào cũng là điểm cân bằng (số lượng bên trái và bên phải không nhất thiết phải bằng nhau)
- Dữ liệu không nhất thiết tập trung nhiều quanh giá trị trung bình
- Và giá trị trung bình không luôn đại diện chính xác cho toàn bộ dữ liệu

Độ Phân Tán (Phương Sai, Độ Lệch Chuẩn)

Các chỉ số phân tán giúp hiểu rõ mức độ biến động của dữ liệu - yếu tố quan trọng khi đưa ra quyết định dựa trên dữ liệu

Phương sai (Variance)	Độ lệch chuẩn (Standard Deviation)	Case study: So sánh 2 nhóm sản phẩm
<ul style="list-style-type: none">Đo lường mức độ phân tán của dữ liệu xung quanh giá trị trung bình.Trong Excel: Sử dụng hàm VAR.S(range) cho mẫu hoặc VAR.P(range) cho toàn bộ dữ liệu.Ví dụ: Doanh thu hàng ngày (triệu VND): 50, 52, 49, 51, 48<ul style="list-style-type: none">Trung bình: 50 triệuPhương sai: 2.5 → Biến động thấp	<ul style="list-style-type: none">Là căn bậc hai của phương sai, có cùng đơn vị với dữ liệu gốc nên dễ hiểu hơn.Trong Excel: Sử dụng hàm STDEV.S(range) hoặc STDEV.P(range)Ví dụ: Với phương sai 2.5, độ lệch chuẩn $= \sqrt{2.5} \approx 1.58$<ul style="list-style-type: none">Khoảng 68% dữ liệu nằm trong khoảng 50 ± 1.58 triệu (48.42 - 51.58 triệu)	<ul style="list-style-type: none">Nhóm A: 100, 105, 95, 102, 98 triệu → Độ lệch chuẩn: 3.8Nhóm B: 80, 120, 60, 140, 100 triệu → Độ lệch chuẩn: 31.6Dù cùng trung bình 100 triệu, Nhóm B có độ rủi ro cao hơn vì biến động lớn hơn nhiều.



Hàm MIN và MAX

Dùng MIN, MAX để xác định giá trị nhỏ nhất và lớn nhất trong tập dữ liệu, giúp bạn nhanh chóng nắm bắt phạm vi của dữ liệu.

Ví dụ: Với doanh số 5 cửa hàng:
120, 85, 160, 95, 110 (triệu VND)

=MIN(A1:A5) → 85 triệu |

=MAX(A1:A5) → 160 triệu

→ Nhanh chóng thấy cửa hàng hiệu quả nhất và kém nhất



Hàm LARGE và SMALL

Sử dụng LARGE, SMALL để tìm giá trị lớn thứ k hoặc nhỏ thứ k, hữu ích khi bạn muốn xác định top 5, top 10 giá trị cao nhất hoặc thấp nhất.

Ví dụ: 20 sản phẩm có doanh thu trong tháng:

=LARGE(B1:B20,3) → Top 3 sản phẩm bán chạy nhất

=SMALL(B1:B20,5) → 5 sản phẩm kém nhất cần cải thiện



Thiết lập cảnh báo

Thiết lập cảnh báo tự động khi giá trị vượt ngưỡng, giúp phát hiện sớm các bất thường trong dữ liệu kinh doanh.

Ví dụ: Định dạng có điều kiện → Quy tắc mới → Sử dụng công thức:

=OR(D2<\$G\$1, D2>\$G\$2) → Tự động tô đỏ các ô có giá trị nằm ngoài ngưỡng an toàn

→ Nhanh chóng phát hiện chỉ số bất thường để xử lý kịp thời

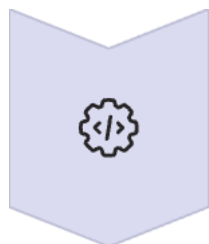
Các Hàm Thống Kê Cơ Bản Trong Excel

Chỉ Số Thống Kê	Hàm Excel	Cách sử dụng	Ví dụ cụ thể
Mean (Giá trị trung bình)	AVERAGE()	=AVERAGE(dãy_số)	=AVERAGE(B1:B10) → Tính giá trị trung bình điểm số của 10 học sinh
Median (Trung vị)	MEDIAN()	=MEDIAN(dãy_số)	=MEDIAN(D1:D15) → Xác định mức lương trung vị của 15 nhân viên
Mode	MODE()	=MODE(dãy_số)	=MODE(E1:E50) → Tìm kích cỡ giày phổ biến nhất trong 50 khách hàng
Giá trị lớn nhất - Giá trị nhỏ nhất	MAX()/MIN()	=MAX(dãy_số) hoặc =MIN(dãy_số)	=MAX(J1:J40) → Tìm doanh thu cao nhất trong 40 cửa hàng
Tổng	SUM()	=SUM(dãy_số)	=SUM(K1:K12) → Tính tổng chi phí hoạt động trong 12 tháng
Số lượng dữ liệu	COUNT()/ COUNTA()	=COUNT(dãy_số) hoặc =COUNTA(dãy_ô)	=COUNT(L1:L200) → Đếm số lượng giao dịch đã ghi nhận trong bảng dữ liệu
Variance (Phương sai)	VAR()	=VAR(dãy_số)	=VAR(G1:G25) → Tính độ phân tán của lợi nhuận 25 sản phẩm
Standard Deviation (Độ lệch chuẩn)	STDEV()	=STDEV(dãy_số)	=STDEV(F1:F30) → Đo mức độ biến động của doanh số trong 30 ngày
Standard Error (Sai số chuẩn)	STDEV()/SQRT(COUNT())	=STDEV(dãy_số)/SQRT(COUNT(dãy_số))	=STDEV(C1:C20)/SQRT(COUNT(C1:C20)) → Tính sai số chuẩn cho 20 mẫu đo lường
Kurtosis (Độ nhọn)	KURT()	=KURT(dãy_số)	=KURT(H1:H100) → Phân tích mức độ tập trung của giá trị trong 100 mẫu
Skewness (Độ lệch)	SKEW()	=SKEW(dãy_số)	=SKEW(I1:I80) → Kiểm tra tính đối xứng của phân phối thu nhập của 80 hộ gia đình

Các hàm Excel quan trọng

Phương thức	Hàm Excel	Cách sử dụng	Ví dụ
Kiểm tra điều kiện	IF()	=IF(điều_kiện, giá_trị_nếu_đúng, giá_trị_nếu_sai)	=IF(A1>100,"Cao","Thấp") → Nếu A1 lớn hơn 100 trả về "Cao", ngược lại trả về "Thấp"
Lồng nhiều điều kiện	IFS()	=IFS(điều_kiện1, giá_trị1, điều_kiện2, giá_trị2...)	=IFS(A1<50,"Thấp",A1<100,"Trung bình",TRUE,"Cao") → Phân loại giá trị dựa trên các mức
Tổng có điều kiện	SUMIF()	=SUMIF(phạm_vi, tiêu_chí, phạm_vi_tổng)	=SUMIF(B1:B10,"Hà Nội",C1:C10) → Tổng doanh số của các cửa hàng ở Hà Nội
Tổng nhiều điều kiện	SUMIFS()	=SUMIFS(phạm_vi_tổng, phạm_vi1, tiêu_chí1, phạm_vi2, tiêu_chí2...)	=SUMIFS(D1:D20,B1:B20,"Hà Nội",C1:C20,"Q1") → Tổng doanh số ở Hà Nội trong quý 1
Đếm có điều kiện	COUNTIF()	=COUNTIF(phạm_vi, tiêu_chí)	=COUNTIF(B1:B50,">100") → Đếm số sản phẩm có doanh số lớn hơn 100
Đếm nhiều điều kiện	COUNTIFS()	=COUNTIFS(phạm_vi1, tiêu_chí1, phạm_vi2, tiêu_chí2...)	=COUNTIFS(B1:B20,"Nam",C1:C20,">30") → Đếm số khách hàng nam trên 30 tuổi
Trung bình có điều kiện	AVERAGEIF()	=AVERAGEIF(phạm_vi, tiêu_chí, phạm_vi_trung_bình)	=AVERAGEIF(B1:B10,"Laptop",C1:C10) → Tính giá trung bình các mặt hàng laptop
Trung bình nhiều điều kiện	AVERAGEIFS()	=AVERAGEIFS(phạm_vi_trung_bình, phạm_vi1, tiêu_chí1...)	=AVERAGEIFS(D1:D10,B1:B10,"Laptop",C1:C10,">5000000") → Giá trung bình laptop trên 5 triệu
Tìm kiếm theo hàng	HLOOKUP()	=HLOOKUP(giá_trị_tìm, bảng_tìm, chỉ_số_hàng, [chính_xác])	=HLOOKUP("Q1",A1:E5,3,FALSE) → Tìm giá trị ở hàng 3 dưới cột "Q1"
Tìm kiếm theo cột	VLOOKUP()	=VLOOKUP(giá_trị_tìm, bảng_tìm, chỉ_số_cột, [chính_xác])	=VLOOKUP("SP001",A1:F20,3,FALSE) → Tìm giá trị ở cột 3 của sản phẩm "SP001"
Truy xuất dữ liệu theo vị trí	INDEX()	=INDEX(mảng, số_hàng, [số_cột])	=INDEX(A1:D10,3,2) → Trả về giá trị ở hàng 3, cột 2 trong phạm vi A1:D10
Tìm vị trí của dữ liệu	MATCH()	=MATCH(giá_trị_tìm, phạm_vi_tìm, [kiểu_đối_chiếu])	=MATCH("SP005",A1:A20,0) → Trả về vị trí hàng của "SP005" trong phạm vi A1:A20

Với Data Analysis Tool > Descriptive Statistics, bạn nhận được đầy đủ các chỉ số như trung bình, mode, độ lệch chuẩn một cách nhanh chóng mà không cần nhập nhiều công thức.



Kích hoạt Data Analysis Tool

Bật công cụ này trong Excel



Chọn Descriptive Statistics

Dùng tính năng này để có kết quả nhanh



Diễn giải kết quả

Hiểu ý nghĩa các chỉ số thu được

Pivot Table là công cụ mạnh mẽ giúp tổng hợp và phân tích lượng lớn dữ liệu một cách trực quan thông qua việc tái cấu trúc dữ liệu từ dạng bảng thành báo cáo có ý nghĩa.

1 Tạo Pivot Table cơ bản

Chọn dữ liệu → Insert → PivotTable → Kéo thả các trường vào 4 vùng: Filters, Columns, Rows và Values

2 Lọc và nhóm dữ liệu

Sử dụng Slicers để lọc → Nhóm theo thời gian → Tạo Calculated Fields → Hiển thị dữ liệu dưới dạng % với Show Values As

3 Tạo báo cáo trực quan

Chuyển sang PivotChart → Kết hợp nhiều Pivot Table trong Dashboard → Tự động cập nhật khi nguồn thay đổi → Tạo từ nhiều nguồn với Data Model

Nguyên Lý Hoạt Động Của Pivot Table

Pivot Table sắp xếp lại dữ liệu bằng cách gom nhóm và tính toán dựa trên các thành phần được đặt vào 4 khu vực chính:



Filters (Bộ lọc)

Chọn xem dữ liệu nào được hiển thị. Ví dụ:
Chỉ xem số liệu của "Quý 1" hoặc "Khu vực miền Nam".



Columns (Cột)

Tạo các cột trong báo cáo. Ví dụ: Đặt "Tháng" vào đây sẽ tạo một cột cho mỗi tháng, giúp xem số liệu theo thời gian.



Rows (Hàng)

Tạo các hàng trong báo cáo. Ví dụ: Đặt "Sản phẩm" vào đây sẽ hiển thị mỗi sản phẩm trên một hàng, giúp so sánh giữa các sản phẩm.



Values (Giá trị)

Tính toán kết quả (tổng, trung bình, đếm...). Ví dụ: Kéo "Doanh thu" vào đây và chọn SUM sẽ hiển thị tổng doanh thu.

i Ưu điểm của Pivot Table là khả năng tự động tính toán lại khi bạn kéo thả các mục khác nhau vào các khu vực này, giúp xem dữ liệu từ nhiều góc độ khác nhau.

Phím Tắt Windows	Phím Tắt Mac	Chức Năng
Ctrl + C	Command (⌘) + C	Sao chép dữ liệu
Ctrl + V	Command (⌘) + V	Dán dữ liệu
Ctrl + Z	Command (⌘) + Z	Hoàn tác thao tác
Ctrl + E	Command (⌘) + E	Flash Fill (Tự động điền dữ liệu theo mẫu)
Ctrl + Y	Command (⌘) + Y	Làm lại thao tác
Ctrl + S	Command (⌘) + S	Lưu tập tin
F4	Command (⌘) + T	Lặp lại thao tác cuối cùng / Cố định tham chiếu trong công thức
F2	Control + U	Chỉnh sửa ô
Ctrl + Home	Command (⌘) + Home	Di chuyển đến ô A1
Ctrl + End	Command (⌘) + End	Di chuyển đến ô cuối có dữ liệu
Ctrl + →	Command (⌘) + →	Di chuyển đến cột cuối có dữ liệu
Ctrl + ↓	Command (⌘) + ↓	Di chuyển đến hàng cuối có dữ liệu
Alt +=	Command (⌘) + Shift + T	Tự động tính tổng
Ctrl + Shift + L	Command (⌘) + Shift + F	Bật/tắt bộ lọc
Ctrl + Space	Control + Space	Chọn toàn bộ cột
Shift + Space	Shift + Space	Chọn toàn bộ hàng
Ctrl + F / Ctrl + H	Command (⌘) + F / Command (⌘) + Shift + H	Tìm kiếm / Tìm và thay thế
Alt → H → O → I/or A	Home → Format → AutoFit Column Width/Row Height	Tự động điều chỉnh chiều rộng cột, hàng
Alt → I → R or C	Command (⌘) + Shift ++	Chèn hàng (R) hoặc cột (C)
Alt → H → H	unknown	Định dạng ô
Alt → H → F → C	unknown	Định dạng có điều kiện

Phần 1: Hiểu Tổng Quan Về Phân Tích Dữ Liệu

Khái niệm cơ bản và ứng dụng Excel trong phân tích dữ liệu.

Phần 2: Hiểu Xu Hướng Dữ Liệu Qua Thống Kê Cơ Bản

Phương pháp thống kê phát hiện xu hướng trong dữ liệu.

Phần 3: Trực Quan Hóa Dữ Liệu

Tạo biểu đồ và dashboard hiệu quả từ dữ liệu.

Vì Sao Cần Trực Quan Dữ Liệu?

Trực quan hóa giúp nhận diện xu hướng nhanh, truyền đạt thông tin hiệu quả và hỗ trợ ra quyết định dựa trên bằng chứng trực quan.

1 Nhận biết xu hướng nhanh chóng

Biểu đồ Excel chuyển số liệu thành hình ảnh, giúp phát hiện ngay xu hướng doanh số và mối tương quan mà bảng số không thể hiện rõ.

2 Truyền đạt thông điệp hiệu quả

Dashboard với biểu đồ trực quan giúp trình bày kết quả phân tích trong 2-3 phút thay vì 15 phút giải thích bảng số.

3 Hỗ trợ ra quyết định

Biểu đồ phân tán và nhiệt thể hiện tương quan dữ liệu, giúp đưa ra quyết định chính xác về phân bổ ngân sách và nguồn lực.

Doanh thu mặt hàng quần áo

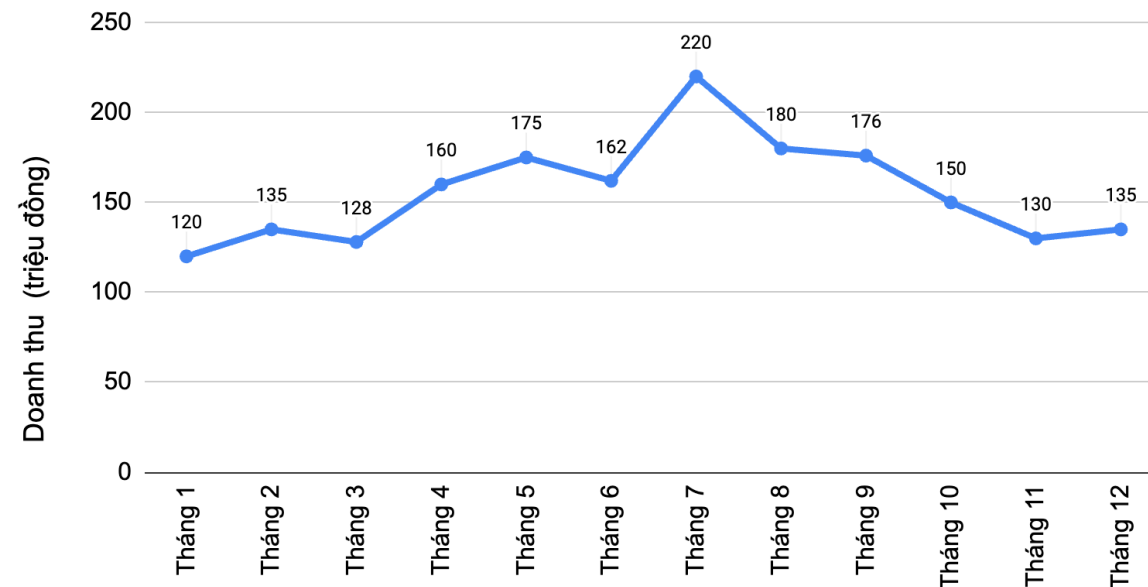
Trực quan hóa dữ liệu giúp chuyển đổi bảng số phức tạp thành biểu đồ dễ hiểu, cho phép nhận biết nhanh xu hướng và hỗ trợ ra quyết định hiệu quả.

✖ Khó nắm bắt xu hướng

Tháng	Doanh thu (triệu đồng)
Tháng 1	120
Tháng 2	135
Tháng 3	128
Tháng 4	160
Tháng 5	175
Tháng 6	162
Tháng 7	220
Tháng 8	180
Tháng 9	176
Tháng 10	150
Tháng 11	130
Tháng 12	135

○ Dễ dàng nắm bắt xu hướng thay đổi doanh số theo mùa

Doanh thu có xu hướng tăng vào mùa hè và bắt đầu giảm khi chuyển sang thu và đông



Các Loại Biểu Đồ Thường Được Sử Dụng



Biểu đồ cột (Bar Chart)

So sánh giá trị giữa các nhóm



Biểu đồ đường (Line Chart)

Hiển thị xu hướng theo thời gian



Biểu đồ phân tán (Scatter Plot)

Thể hiện mối liên hệ giữa hai biến số



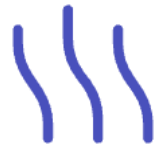
Biểu đồ tròn (Pie Chart)

Thể hiện tỷ lệ các phần trong tổng thể



Biểu đồ hộp (Box Plot)

Cho thấy phân phối dữ liệu và phát hiện giá trị bất thường



Heatmap

Biểu diễn dữ liệu bằng màu sắc, thể hiện mức độ quan hệ



Biểu đồ Phân Phối (Histogram)

Thể hiện tần suất xuất hiện của dữ liệu

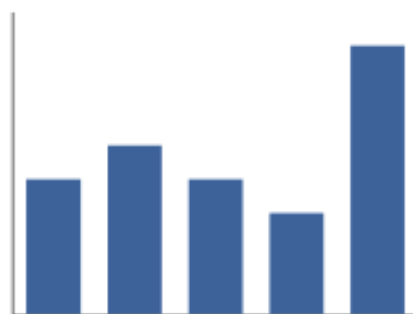


Biểu đồ Water Flow

Thể hiện sự thay đổi giá trị qua các giai đoạn

Biểu Đồ Cột (Bar Plot)

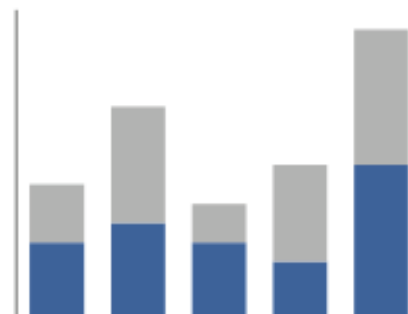
Có 4 loại biểu đồ cột được sử dụng thông dụng: vertical bar, horizontal bar, stacked vertical bar, stacked horizontal bar. Biểu đồ cột được sử dụng để so sánh giá trị giữa các danh mục, với chiều cao cột tương ứng với giá trị.



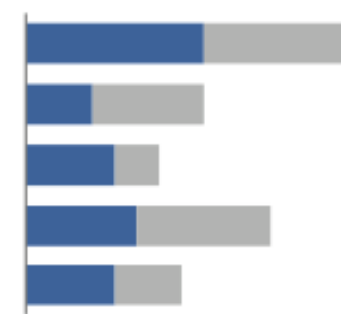
Vertical bar



Horizontal bar



Stacked vertical bar



Stacked horizontal bar

Image source: Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic



Dễ đọc

Trực quan cho mọi đối tượng



So sánh hiệu quả

Phân biệt rõ giữa các nhóm



Linh hoạt

Có thể so sánh nhiều biến cùng lúc

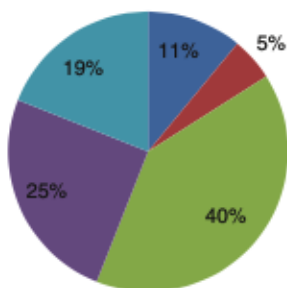
Tips Để Tạo Biểu Đồ Cột Hiệu Quả

Biểu đồ thiết kế không tốt

Survey results: summer learning program on science

PRE: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited



POST: How do you feel about doing science?

■ Bored ■ Not great ■ OK ■ Kind of interested ■ Excited

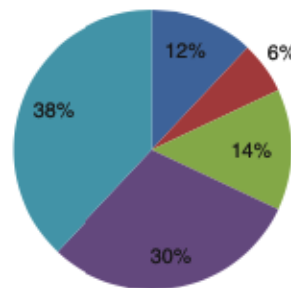


FIGURE 9.28 Original visual

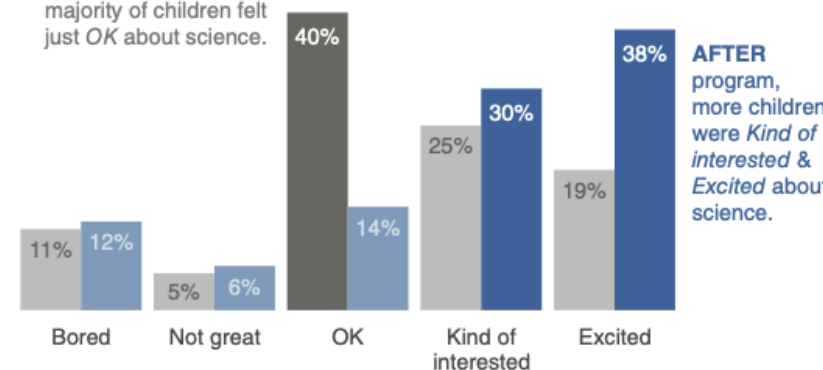
Image source: Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic

Biểu đồ được thiết kế hiệu quả hơn

Pilot program was a success

How do you feel about science?

BEFORE program, the majority of children felt just OK about science.



Based on survey of 100 students conducted before and after pilot program (100% response rate on both surveys).

FIGURE 9.30 Simple bar graph

1. Chọn đúng loại biểu đồ

- Chọn đúng biểu đồ phù hợp

2. Đơn Giản Hóa Nội Dung

- Loại bỏ thông tin dư thừa
- Sắp xếp dữ liệu theo thứ tự logic

3. Gắn Nhãn Rõ Ràng

- Đặt tiêu đề ngắn gọn truyền đạt thông điệp chính
- Thêm chú thích khi cần

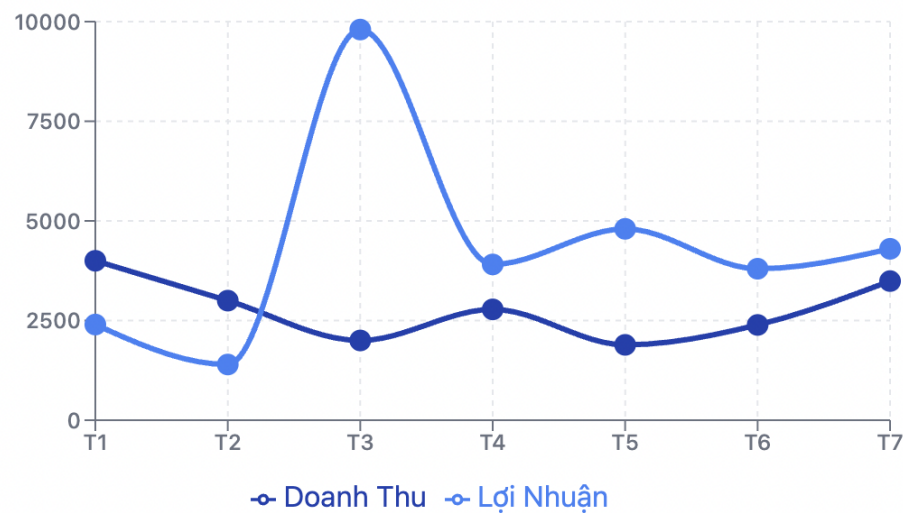
4. Sử Dụng Màu Sắc Phù Hợp

- Chọn màu phù hợp
- Dùng màu tương phản
- Giữ nhất quán màu sắc

Biểu Đồ Đường (Line Chart)

Biểu đồ đường giúp xem sự thay đổi của dữ liệu theo thời gian một cách rõ ràng.

Biểu đồ đường: Thay đổi doanh thu theo tháng



Tips sử dụng biểu đồ đường



Xem xu hướng qua thời gian

Thích hợp để hiển thị dữ liệu theo thời gian như doanh số theo tháng



So sánh nhiều loại dữ liệu

Dễ xem xu hướng của nhiều đối tượng trên cùng một biểu đồ



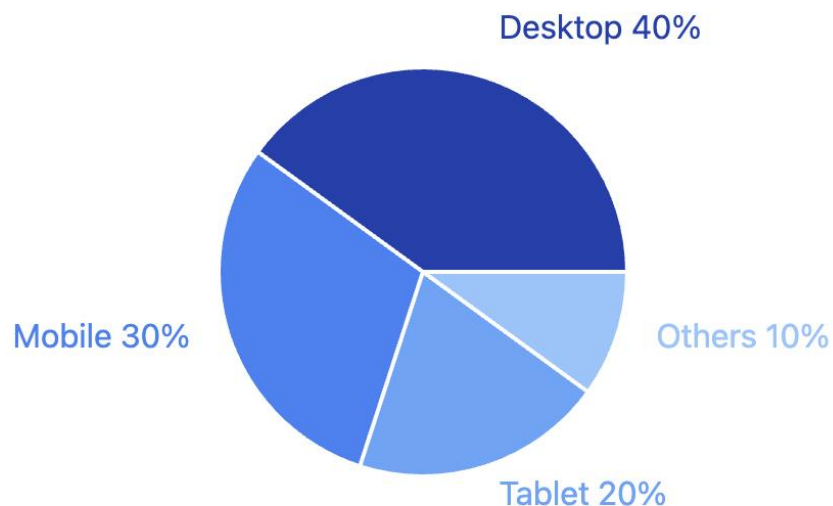
Các kiểu biểu đồ đường trong Excel

Có nhiều loại: biểu đồ đường đơn giản, biểu đồ có điểm đánh dấu, biểu đồ 2D/3D




Biểu Đồ Tròn (Pie Chart)

Biểu đồ tròn thể hiện tỷ lệ các phần trong một tổng thể. Hiệu quả khi so sánh doanh thu, thị phần hoặc phân bổ ngân sách.

Biểu đồ tròn: Phân bố thiết bị truy cập



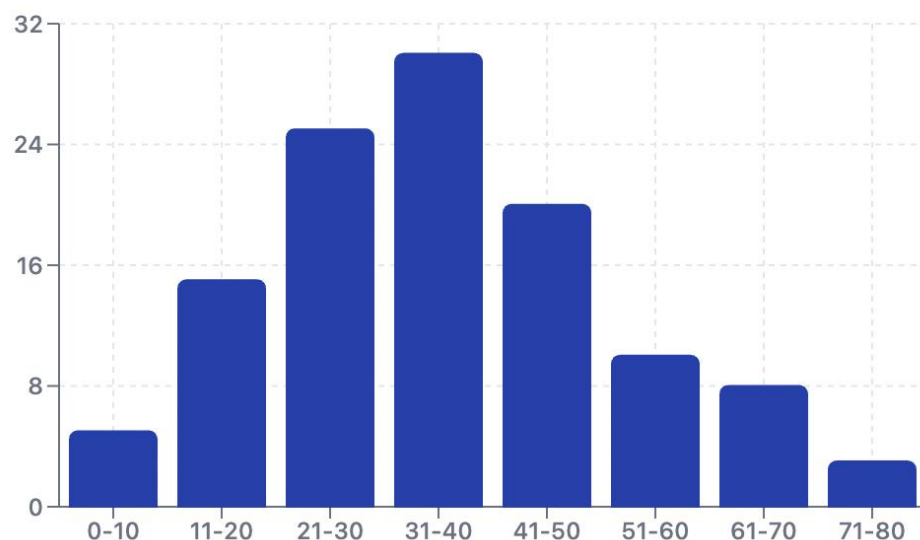
Tips sử dụng biểu đồ tròn

-  **Hiển thị tỷ lệ phần trăm**
Giúp người xem nhanh chóng nắm bắt tỷ lệ của từng phần
-  **Tối ưu với dưới 7 phần**
Hiệu quả nhất khi hiển thị không quá 7 phân khúc
-  **Các biến thể**
Gồm biểu đồ tròn tiêu chuẩn, biểu đồ phân tách và biểu đồ hình khuyên

Biểu đồ Phân Phối (Histogram)

Biểu đồ histogram hiển thị phân phối tần suất của dữ liệu số bằng cách chia thành các khoảng và đếm số lượng điểm dữ liệu trong mỗi khoảng.

Biểu Đồ Histogram: Phân Bố Tần Suất



Tips sử dụng biểu đồ Histogram



Đánh giá hình dạng phân phối

Nhận biết dữ liệu đối xứng, lệch phải hoặc lệch trái



Phát hiện giá trị ngoại lai

Xác định outliers không phù hợp với xu hướng chung



Kiểm tra tính chuẩn

Đánh giá sự phù hợp với phân phối chuẩn



Xác định mô hình thống kê

Chọn mô hình phù hợp dựa trên dạng phân phối

Heatmap là công cụ trực quan hóa dữ liệu bằng màu sắc, giúp nhận biết nhanh các mẫu và xu hướng thông qua thang màu đậm nhạt.

Heat map đơn giản

Tháng	Doanh thu	Lợi nhuận
Tháng 1	120	-12
Tháng 2	135	-7
Tháng 3	128	13
Tháng 4	160	16
Tháng 5	175	18
Tháng 6	162	16
Tháng 7	220	33
Tháng 8	180	18
Tháng 9	176	18
Tháng 10	150	15
Tháng 11	130	13
Tháng 12	135	14

※ Đơn vị: Triệu đồng

Tips tạo và sử dụng heatmap

Tạo heatmap bằng Conditional Formatting

- Sử dụng Color Scales trong Conditional Formatting để tạo heatmap, giúp nhanh chóng nhận biết xu hướng dữ liệu gradient màu sắc.

Heatmap giúp phân tích hiệu suất nhanh chóng

- So sánh hiệu suất giữa khu vực, sản phẩm hoặc thời gian thông qua màu sắc đậm nhạt, nhanh chóng xác định điểm mạnh và yếu.

Biểu đồ Water Flow (hay còn gọi là biểu đồ thác nước) là dạng biểu đồ trực quan hóa thể hiện sự thay đổi giá trị lũy kế theo từng giai đoạn, giúp người xem hiểu được các yếu tố đóng góp tích cực hoặc tiêu cực vào kết quả cuối cùng.

Biểu đồ Water Flow

2014 Headcount math

Though more employees transferred out of the team than transferred in, aggressive hiring means overall headcount (HC) increased 16% over the course of the year.

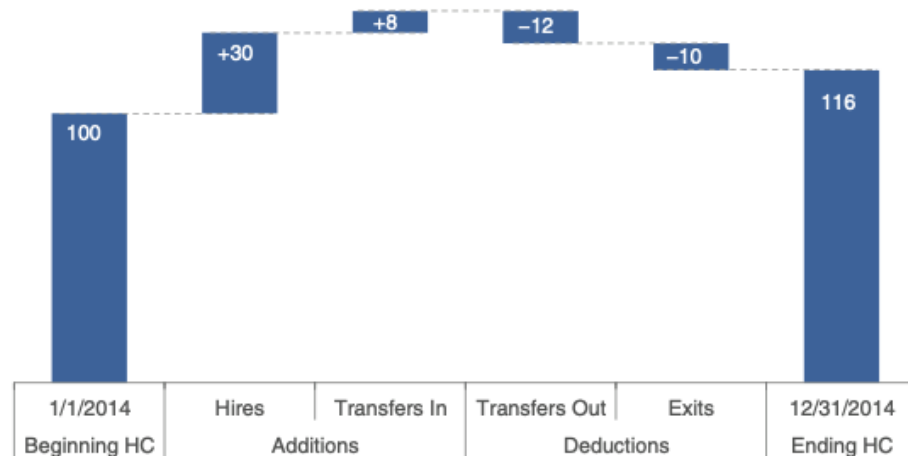


FIGURE 2.17 Waterfall chart

Image source: Storytelling with Data: A Data Visualization Guide for Business Professionals, Cole Nussbaumer Knaflic

Tips tạo và sử dụng Water Flow

Đặc điểm chính

- Hiện thị sự biến động giữa hai điểm dữ liệu - điểm bắt đầu và kết thúc
- Các thanh tăng thường hiển thị màu xanh lá, các thanh giảm thường màu đỏ
- Thanh cuối cùng hiển thị giá trị tổng sau tất cả biến động

Khi nào nên sử dụng

- Phân tích tài chính: theo dõi sự thay đổi doanh thu, chi phí, lợi nhuận
- Phân tích KPI: so sánh chỉ số hiệu suất giữa các giai đoạn
- Phân tích nguyên nhân: xác định các yếu tố tác động tích cực/tiêu cực
- Nghiên cứu thị trường: theo dõi biến động thị phần, khách hàng

Tạo Biểu Đồ Water Flow Trong Excel

Biểu đồ Water Flow (hay biểu đồ thác nước) giúp trực quan hóa sự thay đổi giá trị lũy kế qua các giai đoạn, thể hiện các yếu tố tăng/giảm ảnh hưởng đến kết quả cuối cùng.

1

Chuẩn bị dữ liệu

Tổ chức dữ liệu thành các cột: giá trị ban đầu, các thay đổi (tăng/giảm), và giá trị cuối. Mỗi thay đổi được đặt trong một cột riêng biệt.

2

Tạo biểu đồ cột chồng

Chọn dữ liệu → Insert → Column Chart → Stacked Column. Đây là nền tảng cho biểu đồ Water Flow.

3

Chỉnh sửa hiển thị

Chuyển các giá trị âm thành màu đỏ, giá trị dương thành màu xanh lá. Thêm đường kết nối giữa các cột để tạo hiệu ứng "dòng chảy".

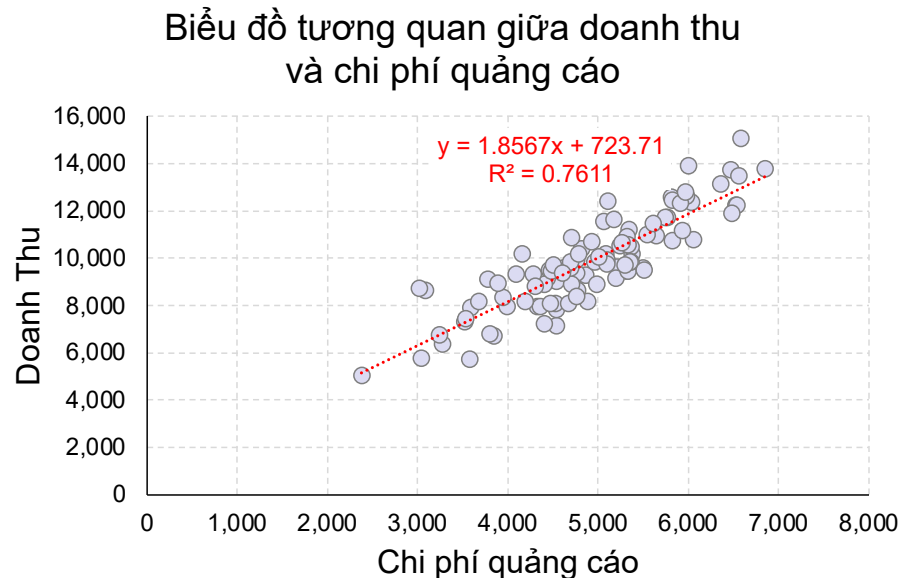
4

Tinh chỉnh định dạng

Thêm nhãn dữ liệu, điều chỉnh màu sắc và căn chỉnh để biểu đồ dễ đọc. Giá trị cuối cùng thường được đánh dấu bằng màu khác biệt.

Biểu đồ phân tán hiển thị mối quan hệ giữa hai biến số, giúp phát hiện mẫu và xu hướng trong dữ liệu.

Biểu Đồ Scatter Plot - Mối Quan Hệ Giữa Các Biến



Tips tạo và sử dụng scatter plot

Tạo scatter plot

- Mỗi điểm trên biểu đồ đại diện cho một cặp giá trị (x,y), giúp nhận diện mẫu và xu hướng.
- Trong Excel: chọn dữ liệu, sử dụng tùy chọn Scatter, thêm nhãn cho các điểm quan trọng.

Thêm trendline và phân tích

- Đường xu hướng làm rõ mối quan hệ giữa các biến. Excel hỗ trợ nhiều dạng: tuyến tính, đa thức, logarithm.
- Hiển thị phương trình và R^2 để đo lường độ mạnh của mối quan hệ và dự đoán giá trị.

3 Tips Để Tạo Scatter Plot Đẹp

1 Cài đặt độ trong suốt khi quá nhiều điểm dữ liệu

- Giảm độ đậm của điểm dữ liệu (opacity) khi visualize dataset lớn giúp nhìn rõ các khu vực tập trung cao và tránh hiện tượng chồng chéo.
- Trong Excel, điều chỉnh này được thực hiện qua Format Data Series.

2 Đặt giới hạn trên và dưới cho dữ liệu

- Thiết lập giới hạn trục x và y phù hợp giúp loại bỏ các giá trị ngoại lệ và tập trung vào phạm vi dữ liệu quan trọng, làm nổi bật xu hướng chính và tăng độ chính xác trong phân tích.

3 Hiển thị đường fitting line

- Thêm đường xu hướng (trendline) để làm rõ mối quan hệ giữa các biến.
- Hiển thị thêm phương trình và giá trị R^2 để đánh giá độ mạnh của mối quan hệ và khả năng dự đoán của mô hình.

Hệ số tương quan đo lường mối quan hệ tuyến tính giữa hai biến, dao động từ -1 đến +1. Giá trị dương thể hiện quan hệ thuận, âm thể hiện quan hệ nghịch, và gần 0 cho thấy không có tương quan.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Công thức tính hệ số Pearson, với x_i , y_i là giá trị của hai biến, \bar{x} , \bar{y} là giá trị trung bình

+1

Tương quan dương hoàn hảo

Hai biến cùng tăng hoặc cùng giảm

0

Không có tương quan

Không có liên hệ tuyến tính

-1

Tương quan âm hoàn hảo

Một biến tăng, biến kia giảm

Dùng hàm CORREL hoặc PEARSON trong Excel để đo lường mối liên hệ tuyến tính giữa hai biến

Ma trận tương quan phân tích mối quan hệ giữa nhiều biến cùng lúc, phát hiện đa cộng tuyến, và xác định các nhóm biến liên quan. Đây là công cụ thiết yếu trong phân tích dữ liệu đa biến.

Tạo ma trận tương quan

- Ma trận tương quan phân tích mối quan hệ giữa nhiều biến đồng thời. Trong Excel, bạn có thể tạo ma trận này bằng công thức CORREL hoặc VBA để tự động hóa quá trình.
- Ma trận tương quan hoàn chỉnh là ma trận đối xứng, với giá trị 1 trên đường chéo chính.

Phân tích ma trận tương quan

- Ma trận tương quan giúp phát hiện đa cộng tuyến và xác định nhóm biến có liên quan chặt chẽ.
- Việc phát hiện đa cộng tuyến rất quan trọng trong phân tích hồi quy, ảnh hưởng đến độ chính xác của mô hình. Nếu hai biến độc lập có tương quan cao (>0.7), nên cân nhắc loại bỏ một trong hai.

Tạo Excel Template giúp tiết kiệm thời gian và đảm bảo nhất quán trong báo cáo doanh nghiệp



Bắt đầu từ template phù hợp

Sử dụng mẫu Dashboard, KPI Tracker hoặc Sales Report. Tổ chức dữ liệu thành Tables và dùng Named Ranges để kết nối dữ liệu với biểu đồ



Giữ dữ liệu – biểu đồ tách biệt

Tách dữ liệu thô, phân tích và biểu đồ vào các sheet riêng. Sử dụng INDIRECT hoặc OFFSET tạo liên kết động để tự động cập nhật biểu đồ



Tùy chỉnh màu sắc và phong cách

Tạo bộ màu nhất quán 4-5 màu. Dùng định dạng có điều kiện để tự thay đổi màu theo giá trị. Loại bỏ đường lưới, sử dụng font Sans-serif để tăng khả năng đọc



Lưu template biểu đồ tùy chỉnh

Chuột phải vào biểu đồ → "Lưu làm template" → lưu trong thư mục Chart Templates.

Phần 1: Hiểu Tổng Quan Về Phân Tích Dữ Liệu

- ✓ Tìm hiểu về dữ liệu, phân tích dữ liệu, lợi ích của nó trong kinh doanh.
- ✓ Nắm rõ tầm quan trọng của Excel và các chuẩn bị các công cụ cho việc phân tích dữ liệu

Phần 2: Hiểu xu hướng dữ liệu qua thống kê cơ bản

- ✓ Ôn lại các khái niệm thống kê cơ bản
- ✓ Áp dụng các hàm thống kê, Pivot Table trong Excel để hiểu rõ xu hướng dữ liệu

Phần 3: Trực quan hoá dữ liệu

- ✓ Học cách trực quan hoá và sử dụng các loại biểu đồ phổ biến