



# Tổng quan về Phân Tích Dữ Liệu



## Nhóm biên soạn:

1. Lê Ngọc Thành
2. Nguyễn Ngọc Thảo
3. Phạm Trọng Nghĩa
4. Nguyễn Thái Vũ
5. Trương Tấn Khoa

**Năm 2022**



1

# GIỚI THIỆU CHỨNG NHẬN PHÂN TÍCH DỮ LIỆU



# Về chứng nhận phân tích dữ liệu

Chứng nhận được tạo bởi Google để



Hướng dẫn các nền tảng và công cụ liên quan



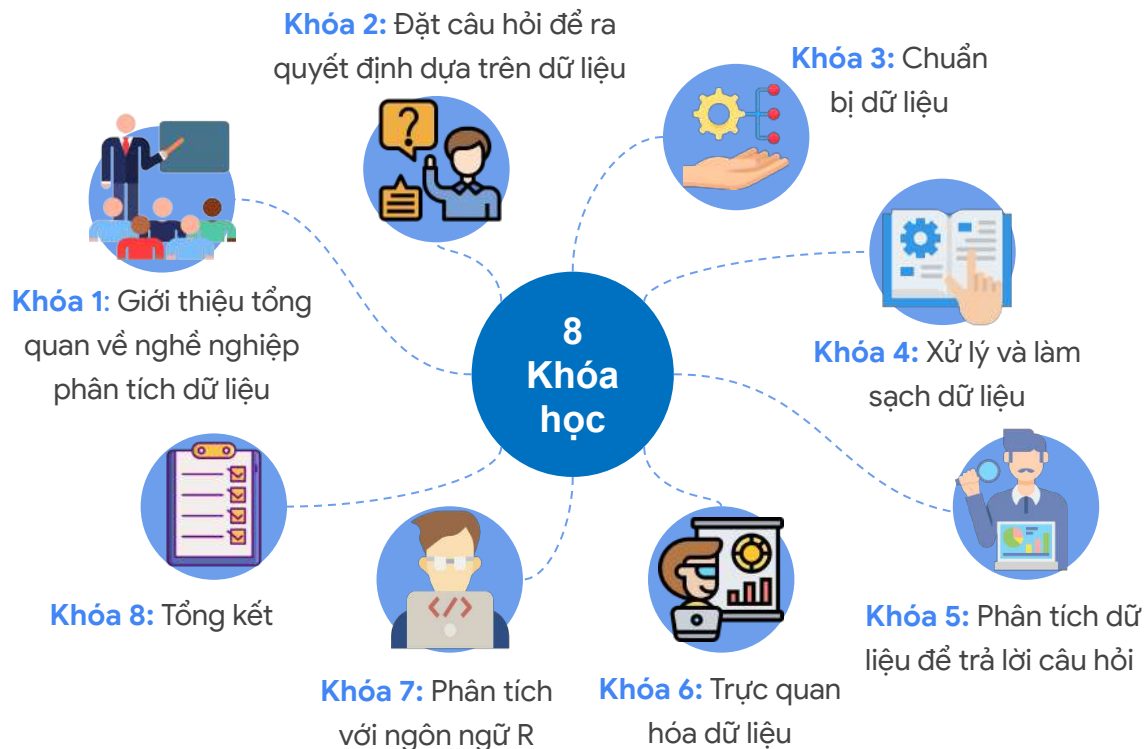
Phát triển các kỹ năng quan trọng trong lĩnh vực phân tích dữ liệu



Giới thiệu nghề nghiệp phân tích dữ liệu và vai trò trong hoạt động doanh nghiệp

# Các khóa học trong chứng chỉ

Chứng chỉ hỗ trợ  
CNTT bao gồm

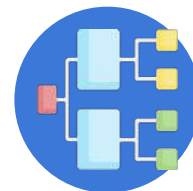


# Tài liệu hỗ trợ

## Mục đích



Tóm tắt những  
điểm quan trọng



Tổ chức các bài đọc để giúp  
người học nắm vững kiến thức



Hướng dẫn thực  
hành các bài tập



Giải đáp những thắc mắc  
mà người học gặp phải



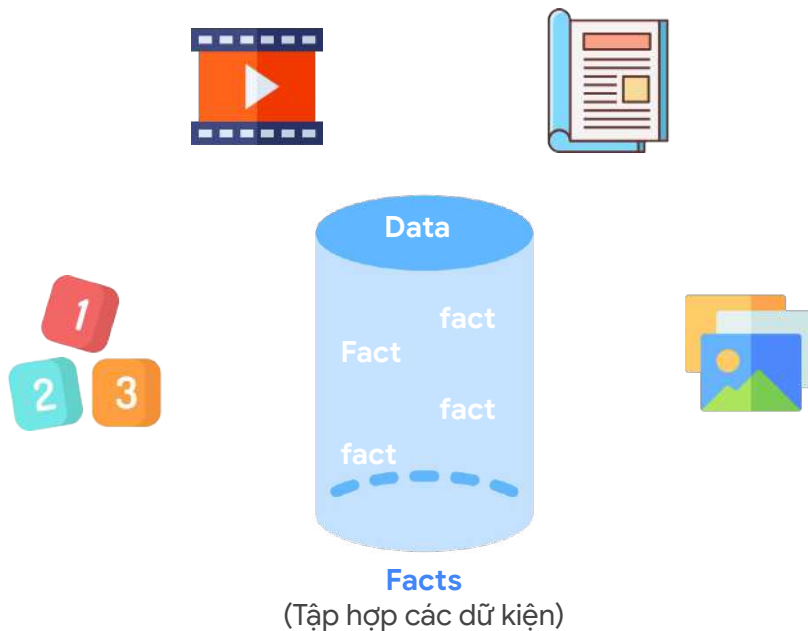
## 2 DỮ LIỆU VÀ PHÂN TÍCH DỮ LIỆU



# Dữ liệu

## Dữ liệu là gì?

- Số
- Hình ảnh
- Video
- Văn bản
- ...



# Dữ liệu đến từ đâu?

Dữ liệu đến từ rất nhiều nguồn

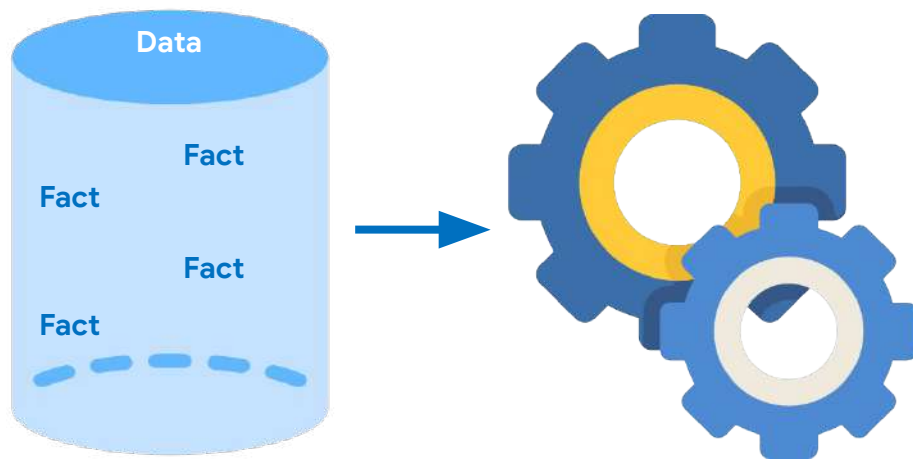




# Hệ sinh thái dữ liệu

Tổng hòa của **nhiều thành phần (phần cứng, phần mềm) tương tác với nhau** để có thể:

- Sản sinh (produce)
- Quản lý (manage)
- Lưu trữ (store)
- Tổ chức (organize)
- **Phân tích (analyze)**
- Chia sẻ (share)



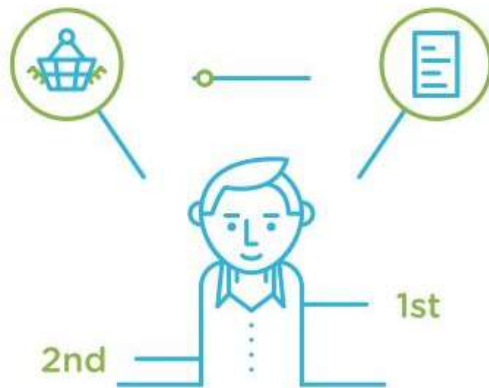
# Phân tích dữ liệu

## Quá trình phân tích dữ liệu



# Lợi ích của phân tích dữ liệu

Giúp thấu hiểu (insight) về hoạt động kinh doanh hiện tại



Nhằm đưa ra quyết định dựa trên dữ liệu (**data-driven decision-making**) và tạo nên chiến lược kinh doanh sử dụng dữ kiện (**business strategy using facts**)

# Trực giác, kiến thức và dữ liệu

## Trực giác/linh tính (**gut instinct**)

- Một cách nhìn chủ quan với ít hoặc không có sự giải thích

## Hiểu biết nghề nghiệp (**business knowledge**)

- Có kiến thức về lĩnh vực và hiểu bài toán kinh doanh
- Chuyên gia về lĩnh vực (**subject matter expert**)

**Quyết định dựa trên trực giác, quyết định dựa trên hiểu biết hay quyết định dựa trên dữ liệu?**

- Kết hợp sẽ mang lại hiệu quả vì hiểu biết về kinh doanh sẽ giúp nắm bắt được nhu cầu phân tích
- Luôn đảm bảo các quyết định cần dựa trên dữ liệu



# Kết hợp như thế nào?

Sau khi phân tích dữ liệu và **trước** khi công bố quyết định đến khách hàng, ta cần:

- Chia sẻ kết quả với các chuyên gia, đồng nghiệp trong lĩnh vực.
- Tham khảo thêm ý kiến của họ để bổ sung vào những thấu hiểu đã được rút ra từ dữ liệu.



# Lợi ích của chuyên gia lĩnh vực

- Chuyên gia lĩnh vực (subject matter expert) sẽ giúp:
  - Xem xét kết quả phân tích dữ liệu và xác định những chỗ không chặt chẽ
  - Phát hiện được các góc khuất
  - Đánh giá các lựa chọn và đưa ra các đề xuất/thấu hiểu cho bài toán

# Nghề nghiệp liên quan đến dữ liệu



## Nhà phân tích dữ liệu (data analyst)

Tìm câu trả lời cho **câu hỏi đang có**  
Tạo ra **sự thấu hiểu** (insight) từ các nguồn dữ liệu



## Nhà khoa học dữ liệu (data scientist)

Tạo ra **câu hỏi mới** để sử dụng dữ liệu  
Tạo ra **những cách mới để mô hình hóa và hiểu biết những gì chưa biết** bằng cách sử dụng dữ liệu thô (raw data)



## Kỹ sư dữ liệu (data engineer)

Xác định các đặc trưng tác động đến kết quả của dữ liệu



## Nhà thiết kế dữ liệu (data designer)

Tổ chức, biến đổi dữ liệu để giúp dễ dàng sử dụng cho các tác vụ liên quan

# Tóm tắt

- Dữ liệu là các dữ kiện được thu thập từ rất nhiều nguồn khác nhau
- Hệ sinh thái dữ liệu bao gồm những thành phần tương tác với nhau để sản sinh, quản lý, lưu trữ, tổ chức, phân tích và chia sẻ.
- Phân tích dữ liệu là quá trình thu thập, biến đổi, tổ chức dữ liệu để rút ra kết luận, dự báo và ra quyết định
- Lợi ích của phân tích dữ liệu mang lại là sự thấu hiểu về hoạt động kinh doanh hiện tại để có thể đưa ra quyết định dựa trên dữ liệu
- Kiến thức, kinh nghiệm của các chuyên gia lĩnh vực cùng với trực giác góp phần làm cho các quyết định dựa trên dữ liệu được tốt hơn





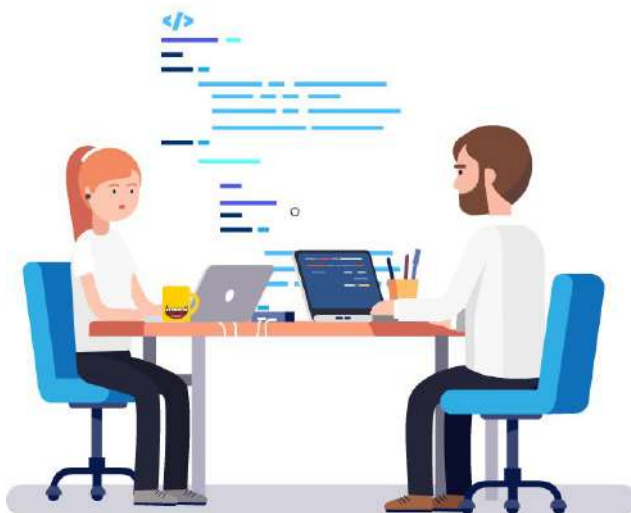


# 3 CÁC KỸ NĂNG PHÂN TÍCH DỮ LIỆU

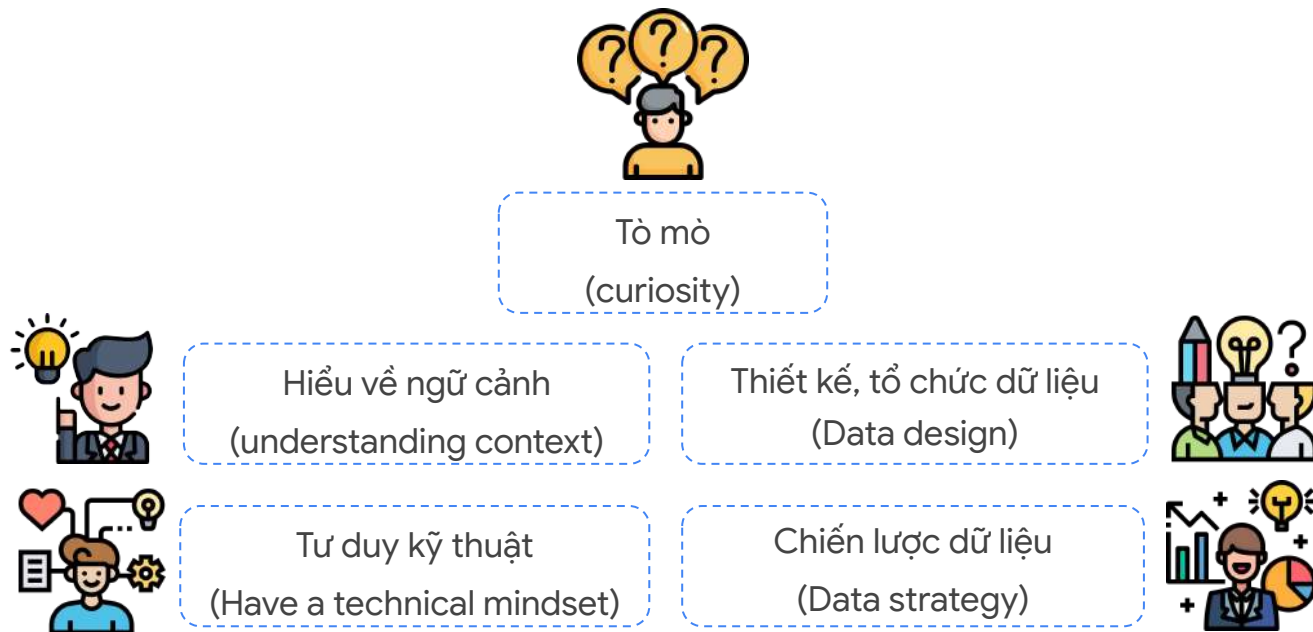


# Kỹ năng phân tích

Kỹ năng phân tích (analytical skill) là những phẩm chất và đặc điểm gắn liền với việc giải quyết vấn đề bằng cách sử dụng dữ kiện



# Các kỹ năng phân tích



# Kỹ năng 1: Sự tò mò

## Tò mò (curiosity)

- Tìm kiếm những **thách thức mới**
- Tìm kiếm những **trải nghiệm mới**
- Mong muốn được **học những điều mới**



# Kỹ năng 2: hiểu về ngữ cảnh

## Hiểu về ngữ cảnh (understanding context)

- Hiểu **bối cảnh**, **đặc điểm** mà một thứ gì đó đang **tồn tại** hay **xảy ra**
- Thấy được **bức tranh tổng thể** mà mỗi sự vật, sự kiện được đặt trong đó

### Ví dụ

- Thông tin mô tả về dữ liệu
- Động cơ của việc thu thập dữ liệu
- ...



# Kỹ năng 3: hiểu biết kỹ thuật

## Tư duy kỹ thuật (technical mindset)

- Biết cách **chia nhỏ bài toán** thành các bài toán nhỏ hơn
- **Xử lý** các bài toán một cách **có thứ tự** và **hợp lý**



# Kỹ năng 4 và 5

- **Thiết kế dữ liệu (data design)** là cách thức **tổ chức thông tin** một cách **rõ ràng** và **hợp lý**.
  - **Chiến lược dữ liệu (data strategy)** liên quan đến cách thức **quản lý con người, tiến trình** và **công cụ** được sử dụng để phân tích dữ liệu.
- Quản lý con người: đảm bảo đội ngũ sử dụng đúng dữ liệu để tìm lời giải cho bài toán
- Quản lý tiến trình: đảm bảo con đường đi đến giải pháp phải rõ ràng và có thể thực hiện được
- Quản lý công cụ: đảm bảo sử dụng đúng công cụ, công nghệ





# 4 TƯ DUY PHÂN TÍCH





# Tư duy phân tích

**Tư duy phân tích** (analytical thinking) là cách chúng ta suy nghĩ để có thể phân tích được bài toán dựa trên dữ kiện.

Bao gồm:

- Xác định và định nghĩa bài toán
- Giải quyết bài toán bằng cách sử dụng dữ liệu một cách có tổ chức (organized) và theo trình tự từng bước (step-by-step)



# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)



Độ tương quan  
(correlation)



Tư duy bức tranh lớn  
(big-picture thinking)



Tư duy hướng chi tiết  
(detail-oriented thinking)

# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)

Thể hiện **thông tin** dưới dạng  
**đồ họa** như **biểu đồ**, **đồ thị**,  
**bản đồ**, ...



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)



Độ tương quan  
(correlation)



Tư duy bức tranh lớn  
(big-picture thinking)



Tư duy hướng chi tiết  
(detail-oriented thinking)

# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)

Xác định những gì mình  
muốn đạt và cách để đạt nó



Độ tương quan  
(correlation)



Tư duy bức tranh lớn  
(big-picture thinking)



Tư duy hướng chi tiết  
(detail-oriented thinking)

# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)

Định nghĩa, mô tả và tập  
trung vào bài toán cần giải  
quyết



Độ tương quan  
(correlation)



Tư duy bức tranh lớn  
(big-picture thinking)



Tư duy hướng chi tiết  
(detail-oriented thinking)

# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)

Nhìn chi tiết để xác định tính  
phù hợp, tương thích của  
mỗi phần



Độ tương quan  
(correlation)



Tư duy bức tranh lớn  
(big-picture thinking)



Tư duy hướng chi tiết  
(detail-oriented thinking)

# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)



Độ tương quan  
(correlation)



Nhìn bài toán **tổng thể** để xác  
**định** tính khả thi và cơ hội

Tư duy bức tranh lớn  
(big-picture thinking)



Tư duy hướng chi tiết  
(detail-oriented thinking)

# Các khía cạnh trong tư duy phân tích



Sự trực quan  
(visualization)



Chiến lược tư duy  
(strategy)



Hướng bài toán  
(problem-orientation)



Độ tương quan  
(correlation)



Tư duy bức tranh lớn  
(big-picture thinking)



Mô tả mối quan hệ của hai  
hay nhiều mảnh của dữ liệu

Tư duy hướng chi tiết  
(detail-oriented thinking)



# Giải pháp cho vấn đề

Để tìm ra giải pháp cho vấn đề, nên bắt đầu bằng cách đặt câu hỏi.

- Một câu hỏi quan trọng: “Nguyên nhân gốc nảy sinh ra vấn đề?” hay “Tại sao lại có vấn đề?”
- Để trả lời câu hỏi này, sử dụng phương pháp: 5 lần hỏi “tại sao” (5 Whys)



# Giải pháp cho vấn đề

**Để tìm ra giải pháp cho vấn đề, nên bắt đầu bằng cách đặt câu hỏi.**

Câu hỏi quan trọng khác: “Liệu có một hố ngăn cách giữa chúng ta với giải pháp?”, “Vậy hố đó là gì?”

**Phân tích khoảng trống** (gap analysis): để **kiểm tra** và **đánh giá trạng thái công việc hiện tại** và **mục tiêu muốn đạt đến** (tốt hơn) ở tương lai





# 5 RA QUYẾT ĐỊNH DỰA TRÊN DỮ LIỆU



Năm 2022



# Ra quyết định dựa trên dữ liệu

**Ra quyết định dựa trên dữ liệu** (data-driven decision making) là việc sử dụng dữ kiện để đưa đến các kết quả đầu ra thành công.

Với dữ liệu, ta có thể:

- Tăng cường những hiểu biết có giá trị về bài toán kinh doanh
- Kiểm chứng được những lý thuyết và giả thuyết
- Hiểu rõ hơn về cơ hội và thách thức
- Giúp đỡ lên kế hoạch hành động
- V.v...



# Ý nghĩa của ra quyết định dựa trên dữ liệu

**Khi ra quyết định dựa trên dữ liệu, chúng ta:**

- Tự tin hơn trên quyết định của mình vì chúng đều có căn cứ từ dữ liệu
- Định hướng được hướng giải một cách tốt hơn và chủ động hơn
- Tiết kiệm thời gian và nguồn lực để đi đến mục tiêu



# Kỹ năng phân tích và quyết định dựa trên dữ liệu

- Các kỹ năng phân tích như sự tò mò, hiểu ngữ cảnh, tư duy kỹ thuật, thiết kế dữ liệu, và chiến lược dữ liệu và ra quyết định có quan hệ tương hỗ lẫn nhau.
- Những kỹ năng phân tích dữ liệu sẽ giúp ra quyết định hướng dữ liệu một cách tốt hơn





# 6 VÒNG ĐỜI DỮ LIỆU

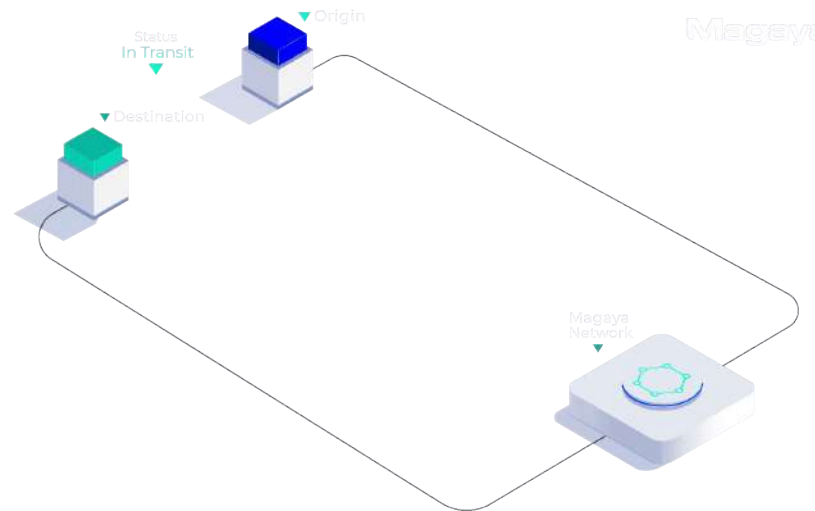


# Vòng đời dữ liệu

**Vòng đời dữ liệu** (data life cycle) là quá trình dữ liệu đi qua từ lúc hình thành, xử lý và phá hủy.

**Mục tiêu xem xét vòng đời dữ liệu để:**

- Biết được những pha mà dữ liệu cần đi qua
- Triển khai những thực thi thích hợp ở mỗi pha





# Các pha trong vòng đời dữ liệu



# Các pha trong vòng đời dữ liệu

## 1. Lên kế hoạch (planning)

Lập kế hoạch cho bài toán, các quyết định cần làm và loại dữ liệu cần có, cách quản lý xuyên suốt vòng đời, người chịu trách nhiệm và kết quả kỳ vọng từ dữ liệu.

# Các pha trong vòng đời dữ liệu

## 2. Thu thập (capture)

Thu thập dữ liệu từ **nơi nào** và **bằng cách nào**, thu thập **từ một nguồn hay nhiều nguồn**, **cách đưa chúng vào** hệ thống quản lý

# Các pha trong vòng đời dữ liệu

## 3. Quản lý (manage)

Cách thức dữ liệu được chăm sóc như lưu trữ dữ liệu như thế nào, cách sử dụng ra sao, bảo mật chúng thế nào và hành động gì để bảo trì, bảo dưỡng chúng.

# Các pha trong vòng đời dữ liệu

Cách thức phân tích dữ liệu, công cụ/công nghệ gì hỗ trợ cho phân tích

4. Phân tích  
(analyze)

# Các pha trong vòng đời dữ liệu

Dữ liệu có dùng tiếp không, xử lý dữ liệu cũ thế nào, thời điểm nào cần quan tâm xử lý chúng

5. Đóng gói  
(archive)

# Các pha trong vòng đời dữ liệu

Dữ liệu có cần phải xóa không? Nếu có thì khi nào và như thế nào?  
Việc xóa và cắt nhỏ dữ liệu còn giúp bảo vệ thông tin riêng tư

6. Phá hủy  
(destroy)



# 7 TIẾN TRÌNH PHÂN TÍCH DỮ LIỆU





# Tiến trình phân tích dữ liệu

Tiến trình phân tích dữ liệu (data analysis process) là quá trình mà một nhà phân tích cần tiến hành để giải quyết bài toán dựa trên dữ liệu.



# Các pha trong tiến trình phân tích dữ liệu



Pha hỏi  
(ask phase)



Pha chuẩn bị  
(prepare phase)



Pha xử lý  
(process phase)



Pha phân tích  
(analyze phase)



Pha chia sẻ  
(share phase)



Pha hành động  
(act phase)

# Các pha trong tiến trình phân tích dữ liệu



Đặt câu hỏi để hiểu về bài toán, mục đích của các bên liên quan, mong đợi của họ, xác định trạng thái hiện tại và khoảng cách đến trạng thái mong đợi, loại dữ liệu cần, ...

# Các pha trong tiến trình phân tích dữ liệu



Thu thập và chuẩn bị dữ liệu cho quá trình phân tích

# Các pha trong tiến trình phân tích dữ liệu



Đánh giá về dữ liệu và các vấn đề như tính đầy đủ, tính cân bằng, nhiễu, lỗi, mâu thuẫn, ...Thực hiện các bước để **xử lý** các vấn đề này

# Các pha trong tiến trình phân tích dữ liệu



Áp dụng các kiến thức, kỹ thuật, công cụ, công nghệ để biến đổi dữ liệu nhằm rút ra các kết luận và các quyết định dựa trên dữ liệu

# Các pha trong tiến trình phân tích dữ liệu



Sử dụng các **công cụ trực quan hóa dữ liệu** để tạo ra **thể hiện đơn giản và dễ hiểu** trong quá trình **chia sẻ** cho các bên liên quan

# Các pha trong tiến trình phân tích dữ liệu



Đem các thấu hiểu, các quyết định vào hành động để giải quyết bài toán đề ra. Các công việc gồm kiểm tra những hiểu biết được rút ra, chọn lựa một chiến lược và triển khai hành động.





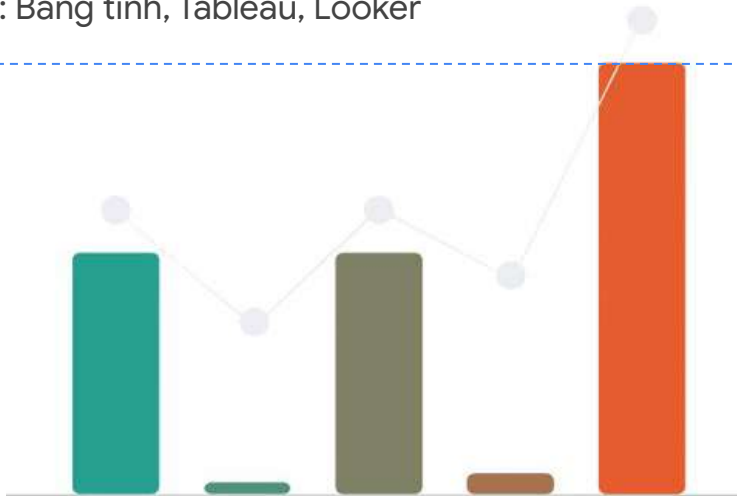
# 8 TRỰC QUAN HOÁ DỮ LIỆU



# Trực quan hóa

**Trực quan hóa** (visualization) là thể hiện dữ liệu dưới dạng đồ thị nhằm tạo sự dễ dàng cho các bên liên quan trong việc hiểu, đưa ra quyết định và chiến lược thực thi.

Một số công cụ phổ biến: Bảng tính, Tableau, Looker



# Những thuận lợi của trực quan hóa

## Thuận lợi chính của trực quan hóa dữ liệu

- Mô tả rõ ràng các quy luật và xu hướng dữ liệu
- Hiểu dữ liệu phức tạp một cách nhanh chóng
- Thể hiện được mối quan hệ giữa các điểm dữ liệu



# Kế hoạch cho trực quan hóa

**Trong quá trình phân tích, chúng ta trải qua:**

- Tìm ra quy luật trong dữ liệu
- Lên kế hoạch cho những gì cần trực quan
- Tạo ra các bản trực quan



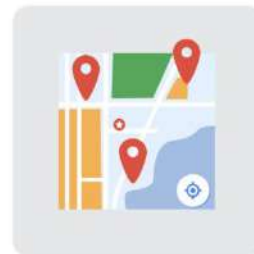
# Các loại biểu đồ

## Một số loại biểu đồ phổ biến:

- Biểu đồ đường (line chart)
- Biểu đồ cột (column chart)
- Biểu đồ điểm (scatter plot)
- Biểu đồ tròn (pie chart)



Line charts can track sales over time



Maps can connect sales to locations



Donut charts can show customer segments



Bar charts can compare total visitors and visitors that make a purchase

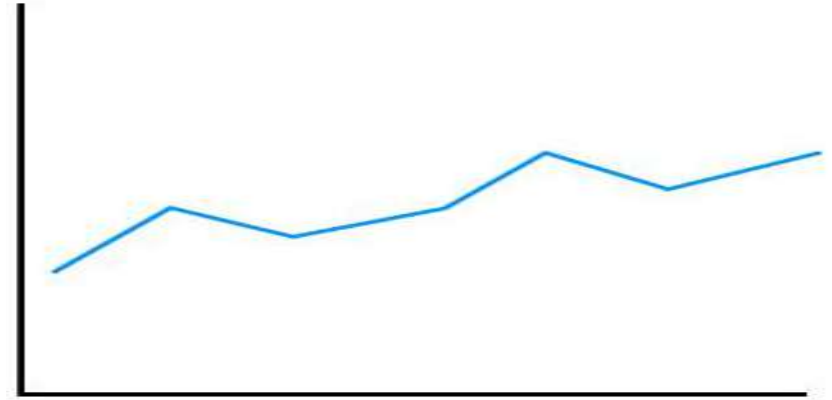
# Các loại biểu đồ

## Một số loại biểu đồ phổ biến:

- Biểu đồ đường (line chart)
- Biểu đồ cột (column chart)
- Biểu đồ điểm (scatter plot)
- Biểu đồ tròn (pie chart)

Thể hiện sự **thay đổi theo thời gian** của đối tượng dữ liệu đang xem xét

- Ví dụ: Tuổi của một người qua thời gian



# Các loại biểu đồ

## Một số loại biểu đồ phổ biến:

- Biểu đồ đường (line chart)
- **Biểu đồ cột (column chart)**
- Biểu đồ điểm (scatter plot)
- Biểu đồ tròn (pie chart)

So sánh giá trị của các loại/lớp đối tượng khác nhau

- Ví dụ: So sánh tuổi giữa nam và nữ



# Các loại biểu đồ

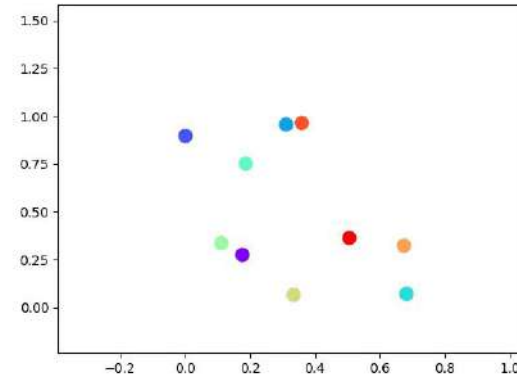
## Một số loại biểu đồ phổ biến:

- Biểu đồ đường (line chart)
- Biểu đồ cột (column chart)
- **Biểu đồ điểm (scatter plot)**
- Biểu đồ tròn (pie chart)



Thể hiện **mối tương quan** giữa 2 hay nhiều thuộc tính của dữ liệu.

- Ví dụ: chúng ta muốn xem tương quan giữa tuổi và thu nhập





# Các loại biểu đồ

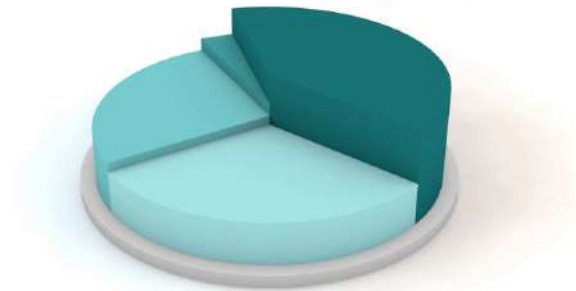
## Một số loại biểu đồ phổ biến:

- Biểu đồ đường (line chart)
- Biểu đồ cột (column chart)
- Biểu đồ điểm (scatter plot)
- Biểu đồ tròn (pie chart)



Thể hiện một **tổng thể** được chia nhỏ thành các **phần như thế nào** hay một thành phần nào đó chiếm tỉ trọng như thế nào so với tổng thể.

- Ví dụ: một tổ chức được thể hiện dưới tỉ lệ phân chia bởi tuổi





# 9 CÔNG CỤ HỖ TRỢ TRONG TIẾN TRÌNH PHÂN TÍCH



# Công cụ hỗ trợ

**Trong quá trình thực hiện phân tích dữ liệu, có nhiều công cụ hỗ trợ như:**

- Bảng tính: Microsoft Excel, Google Sheets
- Truy vấn dữ liệu SQL: MySQL, Microsoft SQL Server
- Trực quan hóa dữ liệu: Tableau, Looker



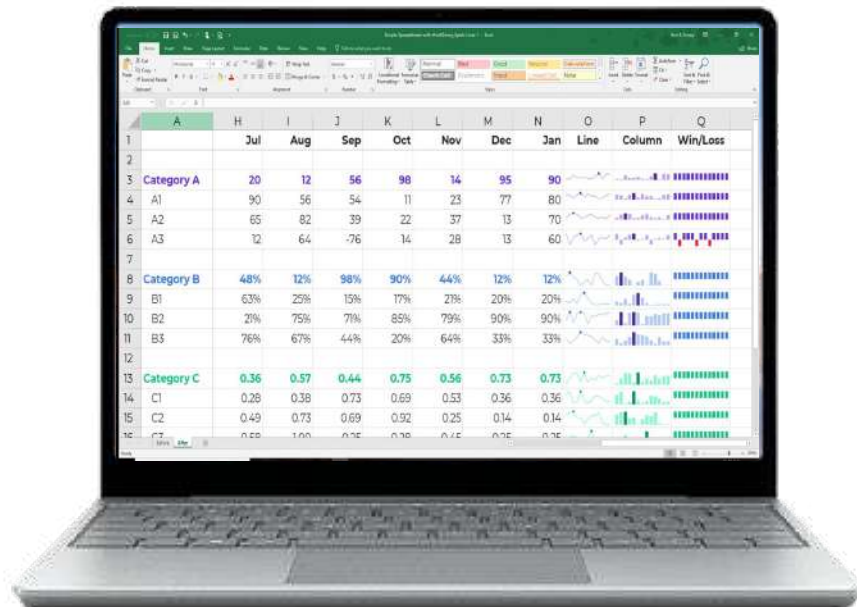
Google docs

# Công cụ bảng tính

**Công cụ bảng tính (spreadsheet)** là công cụ lưu trữ dữ liệu ở mức đơn giản và thông dụng.

**Các đặc điểm chính:**

- Tổ chức dữ liệu dưới dạng cột và dòng
- Sắp xếp và lọc dữ liệu
- Thực hiện tính toán trên dữ liệu thông qua các công thức (formula)
- Thực thi các lệnh, các nhiệm vụ thông qua các hàm (function)



# Truy vấn dữ liệu

**Cơ sở dữ liệu** (database) là cách thức dữ liệu được tập hợp và lưu trữ một cách chặt chẽ và có tổ chức hơn, và được thực hiện bởi một chương trình máy tính.

Dữ liệu lưu trữ trên cơ sở dữ liệu được truy xuất/giao tiếp thông qua **ngôn ngữ truy vấn dữ liệu** như SQL (structured query language – **ngôn ngữ truy vấn có cấu trúc**).

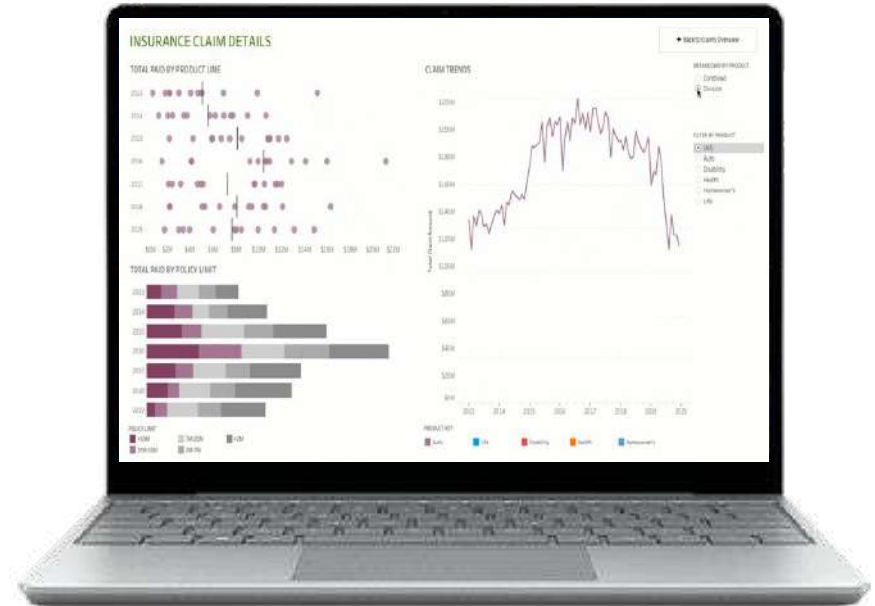
- Để thực hiện yêu cầu (request), cập nhật (update) và truy xuất (retrieve) dữ liệu từ cơ sở dữ liệu, ta sử dụng **câu truy vấn** (query)



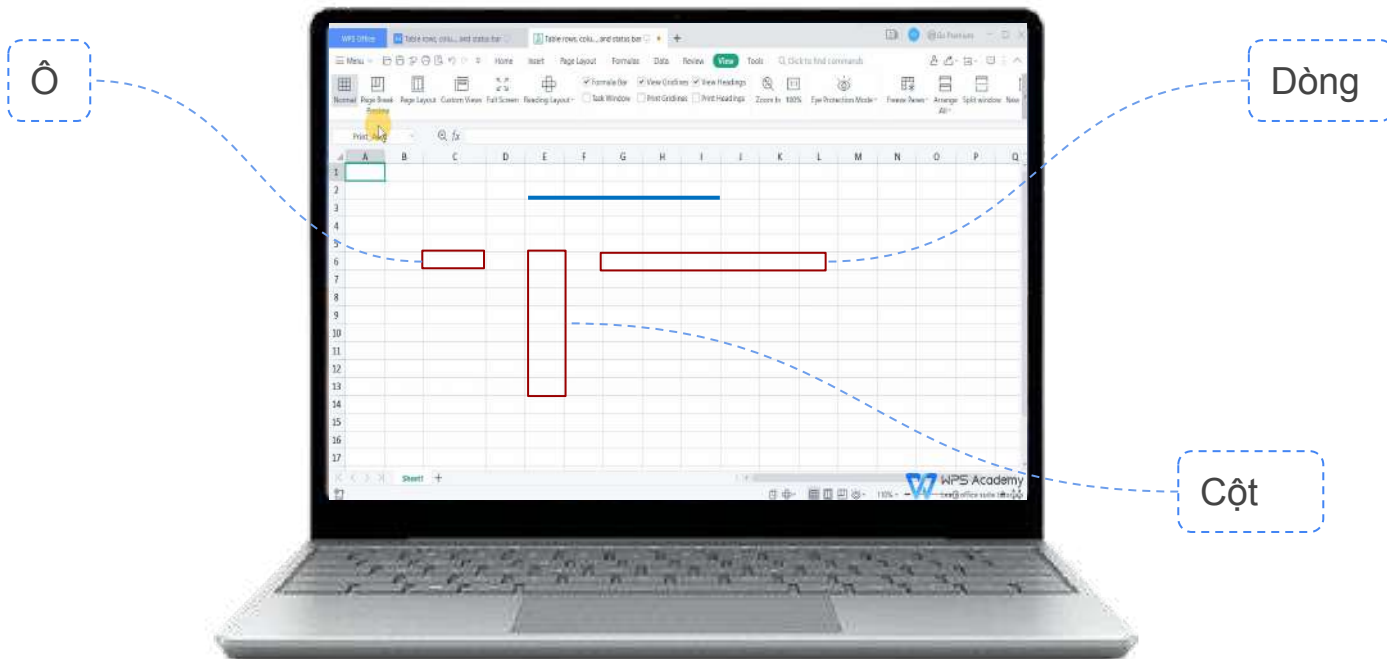
# Công cụ trực quan hóa

**Trực quan hóa** (visualization) là thể hiện dữ liệu dưới dạng đồ thị nhằm tạo sự dễ dàng cho các bên liên quan trong việc hiểu, đưa ra quyết định và chiến lược thực thi.

Một số công cụ phổ biến: Tableau, Looker

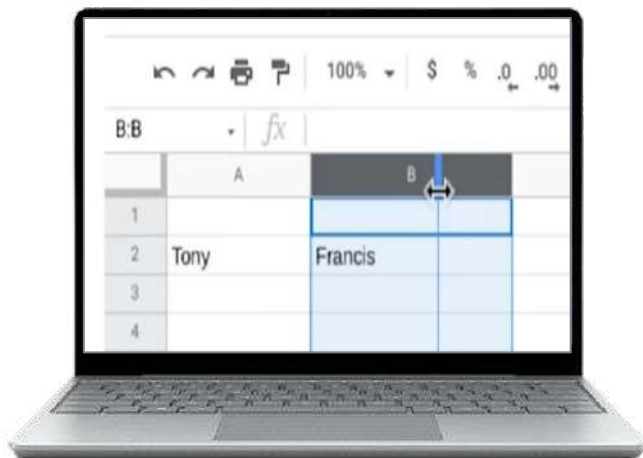


# Bảng tính cơ bản



# Bảng tính cơ bản

Một số thao tác



Thay đổi độ rộng cột



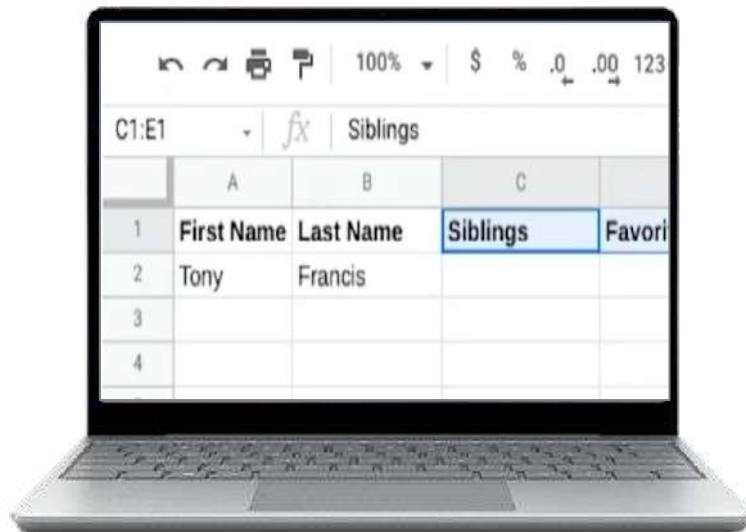
Gói văn bản (text wrapping)



# Bảng tính cơ bản

## Gán nhãn cho từng thuộc tính dữ liệu

- **Thuộc tính** (attribute) là các đặc trưng của từng trường dữ liệu.
- Trong bảng tính, thuộc tính thường được gọi là tên cột (column name), nhãn cột (column label), tiêu đề (header).



# Bảng tính cơ bản

Trong dữ liệu, một dòng được gọi là một **quan sát** (observation).

- Mỗi quan sát thể hiện một mẫu dữ liệu và tập hợp các mẫu tạo thành tập dữ liệu để phân tích.
- Nó chứa thông tin của các thuộc tính đã được định nghĩa.



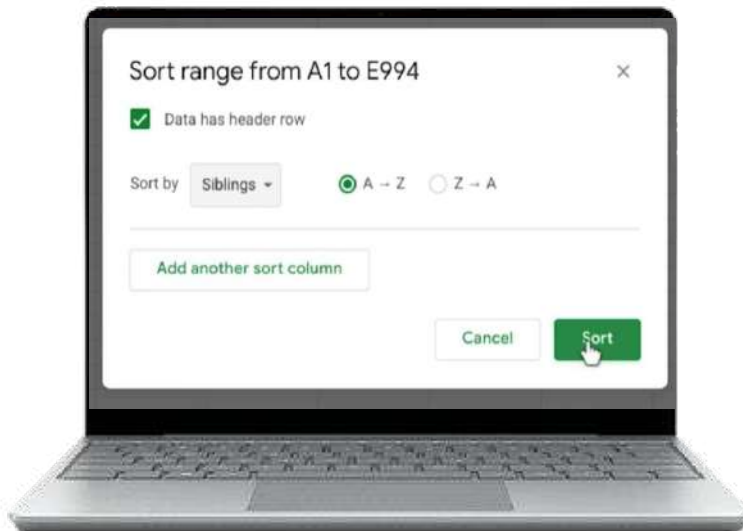
	A	B	C
1	First Name	Last Name	Siblings
2	Tony	Francis	
3	Willa	Stein	
4	Haseem	Galla	
5			

Bảng dữ liệu chứa 3 quan sát

# Bảng tính cơ bản

## Tổ chức, sắp xếp dữ liệu

Sắp xếp giúp nhà phân tích tổ chức và tìm kiếm dữ liệu một cách nhanh hơn

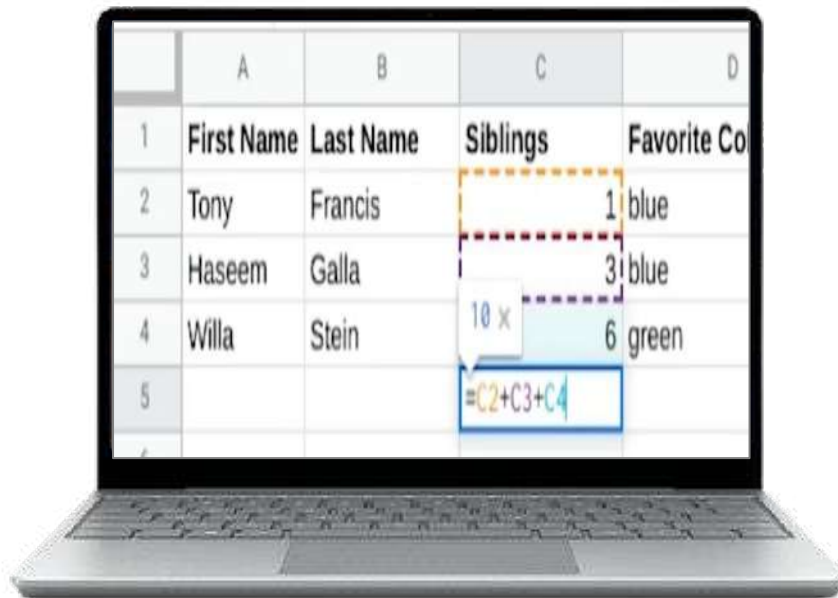


Sắp xếp bảng theo siblings

# Bảng tính cơ bản

**Công thức (formula) thực hiện tính toán, thay đổi trên dữ liệu**

- Nhập công thức bằng cách bắt đầu với dấu bằng (=)
- Chọn công thức muốn thực hiện
- Bên trong là các đối số như các ô, dãy ô, ...



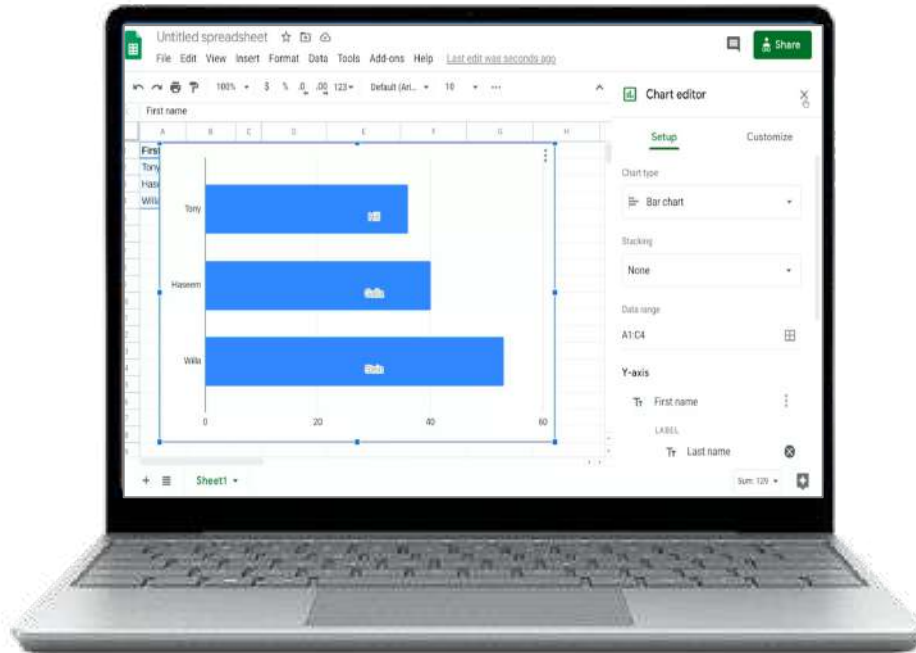
	A	B	C	D
1	First Name	Last Name	Siblings	Favorite Color
2	Tony	Francis	1	blue
3	Haseem	Galla	3	blue
4	Willa	Stein	6	green
5				

Formula bar:  $=C2+C3+C4$

Tooltip: 10 x

Tính tổng của cột siblings

# Trực quan hóa với bảng tính



# Truy vấn dữ liệu

Cú pháp truy vấn trong ngôn ngữ SQL

**SELECT**

Columns you want to look at

**FROM**

Table the data lives in

**WHERE**

Certain condition is met

# Truy vấn dữ liệu

Ví dụ một câu truy vấn

```
SELECT
    first_name
FROM
    customer_data.customer_name
WHERE
    first_name = 'Tony'
```



# 10

## NGHỀ NGHIỆP CHO NHÀ PHÂN TÍCH DỮ LIỆU





# Ngành nghiệp cho nhà phân tích dữ liệu

Nhà phân tích dữ liệu tham gia vào giải quyết bài toán của doanh nghiệp

Mục tiêu chính:

- Sử dụng dữ liệu để cải thiện trải nghiệm người dùng

Các bước trải qua:

- Hiểu về bài toán cần giải quyết
- Xác định dữ liệu cần
- Phân tích dữ liệu để đi đến câu trả lời
- Trong quá trình thực hiện có thể phát sinh các câu hỏi khác cần khám phá thêm



# Vị trí trong nghề phân tích dữ liệu

Một số vai trò, vị trí liên quan đến nghề nghiệp phân tích dữ liệu



Nhà phân tích kinh doanh  
(business analyst)



Nhà hỗ trợ phân tích dữ liệu (data  
analytics consultant)



Nhà phân tích thực thi  
(operation analyst)



Nhà phân tích kinh doanh  
(marketing analyst)



Nhà phân tích tuyển dụng  
(HR/payroll analyst)



Nhà phân tích tài chính  
(Financial analyst)



Nhà phân tích rủi ro  
(Risk analyst)



Vv...

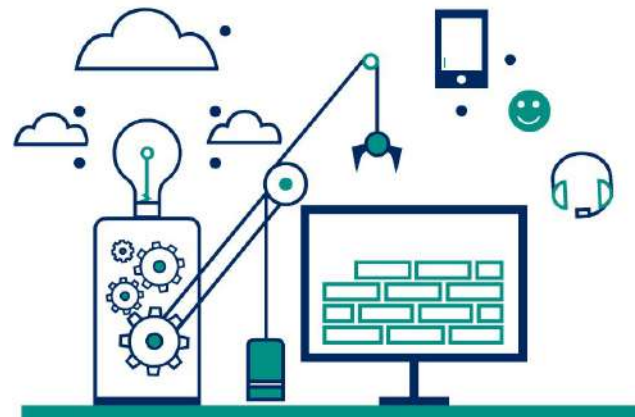
# Đạo đức và công bằng trong phân tích

Quá trình phân tích dữ liệu cần đảm bảo tính hợp lệ và không gây ra tác động tiêu cực đến các bên liên quan.

Luôn đảm bảo công bằng (fairness) trong từng giai đoạn của quá trình phân tích từ lúc thu thập dữ liệu cho đến ra quyết định dựa trên dữ liệu.

Để đảm bảo tính công bằng, cần:

- Xem xét quần thể mẫu dữ liệu (inclusive sample population)
- Hiểu ngữ cảnh xã hội (social context)
- Thu thập dạng dữ liệu tự báo cáo (self-reported data)





# Tổng kết



# Những ý chính cần nắm

- Tầm quan trọng của dữ liệu trong các quyết định hay trong quá trình giải quyết bài toán kinh doanh
- Các kỹ năng cốt lõi được sử dụng bởi các nhà phân tích dữ liệu như làm sạch dữ liệu, phân tích dữ liệu.
- Vai trò của phân tích trong hệ sinh thái dữ liệu
- Tính công bằng và hợp lý trong các thao tác liên quan đến dữ liệu
- Các công cụ hỗ trợ phân tích dữ liệu như bảng tính Google Sheets, Microsoft Excel, ngôn ngữ truy vấn SQL và trực quan hóa dữ liệu Tableau, R.





# THANK YOU

