1. **Tiến trình phân tích dữ liệu**

* Khái niệm: ***Tiến trình phân tích dữ liệu*** (*data analysis process*) là quá trình mà mọt 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):** Đặt câu hỏi dể hiểu về bài toán, mục đích của 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,..

+ **Pha chuẩn bị (prepare phase):** Thu thập và chuẩn bị dữ liệu cho quá trình phân tích

+ **Pha xử lý (process phase):** 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

+ **Pha phân tích (analyze phase):** 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

+ **Pha chia sẻ (share phase):** Sử dụng các công cụ trực quan hóa dữ liệu để tạo ra thể hiện đơn giản và dễ hiểu trogn quá trình chỉa sẻ cho các bên liên quan

+ **Pha hành động (act phase):** 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

1. **Understand the data ecosystem**

The process presented as part of the Google Data Analytics Certificate is one that will be valuable to you as you keep moving forward in your career:

1. **Ask**: business challenge, objective, or question
2. **Prepare**: data generation, collection, storage, and data management
3. **Process**: data cleaning and data integrity
4. **Analyze**: data exploration, visualization, and analysis
5. **Share**: communicating and interpreting results
6. **Act**:  putting insights to work to solve the problem

A project-based data analytics process has five simple steps:

1. Identifying the problem
2. Designing data requirements
3. Pre-processing data
4. Performing data analysis
5. Visualizing data

Big data analytics process:

1. Business case evaluation
2. Data identification
3. Data acquisition and filtering
4. Data extraction
5. Data validation and cleaning
6. Data aggregation and representation
7. Data analysis
8. Data visualization
9. Utilization of analysis results

Khái niệm: ***Data ecosystems*** are the various elements that interact with one another in order to produce, manage, store, organize, analyze and share data.

Sự khác nhau giữa Data Science và Data Analysis

* Data Science: creating new ways of modeling and understanding the unknown by using raw data
* Data Analysis: the collection, transformation and organization of data in order to draw conclusions, make predictions and drive infromed decision-making

Data-driven decision-making: using facts to guide business stratehy

Data analysts need to combine with **subject matter experts** to give data-driven decisions.

**Gut instinct *(research more)***

* **Difference between Data analytiicc and Data analysis**

*Data analytics* is a very broad concept that encompasses everything from the job of managing and using data to the tools and methods that data workers use each and every day

*Data analysis* is the collection, transformation, and organization of data in order to draw conclusions, make predictions, and drive informed decision-making

1. **Embrace your data analyst skill**

Analytical skills are qualities and characteristics associated with solving problems using facts

Các kỹ năng phân tích bao gồm

1. ***Tò mò (curiosity***):

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

1. ***Hiểu về ngữ cảnh (understanding context)***

* Hiểu được 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

VD: thông tin mô tả dữ liệu, động cơ của việc thu thập dữ liệu

1. ***Hiểu biết 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

1. ***Thiết kế dữ liệu (data design):*** là cách thức tổ chức thông tin một cách rõ ràng và hợp lý
2. ***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

1. **Analytical thinking for effective outcomes**

Khái niệm: ***Analytical thinking*** is identifying and defining a problem and then solving it by using data in an organized, step-by-step manner.

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): xác định những gì mình muốn đạt và cách để đạt nó
* Hướng bài toán (problem-orientation): định nghĩa, mô tả và tập trung vào bài toán cần phải giải quyết
* Độ tương quan (correlation) : nhìn chi tiết để xác định tính phù hợp, tương thích của mỗi phần
* Tư duy bức tranh lớn (big-picture thinking): nhìn bài toán tổng thể để xác định tính khả thi và cơ hội
* Tư duy hướng chi tiết (detail-oriented thinking): mô tả mối quan hệ của hai hay nhiều mảnh của dữ liệu
* ***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 dề?”, hay “Tại sao có vấn đề?”…
* Để trả lời câu hỏi này, sử dụng phương pháp: “5 lần hỏi tại sao” (5 Whys)
* 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

***Data-driven decision making***

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

Ý nghĩa của 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