6 GIAI ĐOẠN PHÂN TÍCH

1. Hỏi

* Hỏi ở đây bao gồm cả việc đặt câu hỏi để có thể thu nhập được dữ liệu một cách trực quan nhất, có thể là các bên liên quan sẽ giao cho bạn chẳng hạn

1. Chuẩn bị

* Sau khi đã có được câu trả lời từ ở bước 1 xong, bạn phải thu nhập dữ liệu xác định dữ liệu và kiểm trả tính trung thực của dữ liệu

1. Xử lý

* Đúng như cái tên nhiệm vụ của bạn phải xử lý những dữ liệu đó thành thông tin để đưa vào bước 4, bao gồm bạn phải điều chỉnh dữ liệu và làm sạch dữ liệu…

1. Phân tích

* Phân tích ở đây dựa vào khối lượng các bên liên quan giao nhiệm vụ cho bạn để tìm ra được giải pháp

1. Chia sẽ

* Đúng như cái tên, Chia sẽ tất là đưa ra những cái mới cho những người báo cáo bằng cách trực quang hóa nó lên dựa vào các công cụ vẽ đồ thị… (vd: Google trang tính, R…)

1. Hành động

* Hành động là bắt đầu đi xâu vào một vấn đề nào đó bằng những phát hiện mới nhằm giải quyết đưa ra những sáng kiến đó giúp giải quyết được những vấn đề đang còn xảy ra.

LỊCH SỬ CỦA PHÂN TÍCH DỮ LIỆU

* Nguồn gốc của phân tích dữ liệu chính là thống kê và thống kê ở đây được những người Ai Cập cổ đại chọn để xác định các kim tự tháp hay nói cách khác đó là người Ai Cập là những người hiểu biết nhất về thống kê. Họ đưa ra các công thức toán học và lý thuyết trên 1 loại giấy papỷi ( vật liệu giống như giấy thời nay ) và hiện tại có thể nói những ứng dụng dễ dàng thấy dựa vào những ghi chép của người Ai Cập đó là Excel,…

HỆ SINH THÁI DỮ LIỆU

* Hệ sinh thái là gì ? Theo wikipidia Hệ sinh thái là một hệ thống hoàng chỉnh, trong đó những thành phần sống được gọi là Quần Xã và những thành phần vô sinh như nước, chất khoáng, …
* Hệ sinh thái dữ liệu là một hệ thống trong đó dữ liệu sẽ được quản lý, lưu trữ, phân tích, chia sẽ…

KỸ NĂNG PHÂN TÍCH DỮ LIỆU QUAN TRỌNG

* Tò mò chính xác là tôi cũng hiểu bạn có khái niệm đó trong đầu rồi, tò mò ở đây chính là tìm kiếm những dữ liệu và có thể đặt câu hỏi để có được đáp án
* Hiểu bối cảnh thường thì mọi người sẽ kh hiểu được khái niệm này, dựa vào ví dụ đếm số từ 1 đến 5, nếu đếm từ 1 2 3 4 5 thì ok các số này thuộc bối cảnh nhưng nếu 1 2 3 5 4 vậy ở đây số 4 sẽ nằm ngoài bối cảnh , đơn giản là hiểu nó không phù hợp với lý thuyết toán đã học
* Tư duy kỹ thuật là việc chia nhỏ ra các chức năng để dễ dàng làm việc..
* Phân tích thiết kế là một khái niệm quá là thân quen rồi nên tôi sẽ kh đề cập nữa
* Chiến lược dữ liệu là một trong các kỹ năng về quản lý con người và dự án việc này giúp đưa ra các chiến lược nhằm thỏa mãn điều kiện hoặc là vấn đề đang giải quyết…

SUY NGHĨ PHÂN TÍCH

* Hình dung là việc bạn giúp cho những người khác có góc nhìn chính xác và đúng đắn hơn hay nói cách khác đó là cho người khác hiểu và giải thích chi tiết hơn.
* Tư duy chiến lược ở bước này có thể hiểu việc tư duy chiến lược giúp chúng ta luôn đi đúng hướng cần giải quyết giúp cải thiện được dữ liệu…

CÁC GIAI ĐOẠN CỦA VÒNG ĐỜI DỮ LIỆU

* Vòng đời dữ liệu là việc y như phân tích dữ liệu cũng là thu nhập phân tích quản lý và lưu trữ…
* **Kế hoạch:** Quyết định loại dữ liệu nào là cần thiết, cách quản lý dữ liệu và ai sẽ chịu trách nhiệm về dữ liệu đó.
* **Thu thập:** Thu thập hoặc mang dữ liệu từ nhiều nguồn khác nhau.
* **Quản lý:** Chăm sóc và duy trì dữ liệu. Điều này bao gồm việc xác định cách thức và nơi nó được lưu trữ cũng như các công cụ được sử dụng để thực hiện việc đó.
* **Phân tích:** Sử dụng dữ liệu để giải quyết vấn đề, đưa ra quyết định và hỗ trợ các mục tiêu kinh doanh.
* **Lưu trữ:** Lưu trữ dữ liệu liên quan để tham khảo lâu dài và trong tương lai.
* **Hủy:** Xóa dữ liệu khỏi bộ lưu trữ và xóa mọi bản sao dữ liệu được chia sẻ.

GIAI ĐOẠN PHÂN TÍCH DỮ LIỆU

**Giai đoạn hỏi**

* Khi bắt đầu bất kỳ phân tích dữ liệu thành công nào, nhà phân tích dữ liệu:
* Cần thời gian để hiểu đầy đủ mong đợi của các bên liên quan
* Xác định vấn đề cần giải quyết
* Quyết định những câu hỏi cần trả lời để giải quyết vấn đề

**Giai đoạn chuẩn bị**

* Trong giai đoạn chuẩn bị, trọng tâm là xác định và định vị dữ liệu bạn có thể sử dụng để trả lời câu hỏi của mình.

**Giai đoạn quá trình**

Trong giai đoạn này, mục đích là để tinh chỉnh dữ liệu. Các nhà phân tích dữ liệu tìm và loại bỏ bất kỳ lỗi và sự không chính xác nào có thể cản trở kết quả. Điều này thường có nghĩa là:

* Làm sạch dữ liệu
* Chuyển đổi dữ liệu sang định dạng hữu ích hơn
* Kết hợp hai hoặc nhiều bộ dữ liệu để làm cho thông tin đầy đủ hơn
* Loại bỏ các ngoại lệ (điểm dữ liệu có thể làm sai lệch thông tin)

**Giai đoạn phân tích**

* Các nhà phân tích dữ liệu sử dụng nhiều công cụ mạnh mẽ trong công việc của họ để biến dữ liệu bạn đã thu thập, chuẩn bị và xử lý thành thông tin hữu ích

**Giai đoạn chia sẻ**

* trong giai đoạn chia sẻ, trực quan hóa là người bạn tốt nhất của nhà phân tích dữ liệu.

MODUN 4

* Trong bối cảnh phân tích dữ liệu, nhiệm vụ kinh doanh là câu hỏi được trả lời hoặc vấn đề được giải quyết bằng phân tích dữ liệu.
* Công bằng có nghĩa là đảm bảo rằng sự thiên vị không được tạo ra hay củng cố bởi quá trình phân tích dữ liệu.
* Để đảm bảo phân tích dữ liệu của họ không thiên vị, các chuyên gia dữ liệu xem xét bối cảnh xã hội phức tạp có thể tạo ra sai lệch, ưu tiên sự công bằng trong quá trình thu thập dữ liệu và đảm bảo mọi điều được chia sẻ với các bên liên quan đều được trình bày trong bối cảnh.
* Kịch bản này mô tả việc lấy mẫu quá mức. Lấy mẫu quá mức là quá trình tăng kích thước mẫu của các nhóm không chiếm ưu thế để thể hiện chúng tốt hơn và giải quyết các tập dữ liệu mất cân bằng.