

Tài liệu đọc

Phân Tích Dữ Liệu

Khóa 1: Tổng quan về phân tích dữ liệu

Phần 1: Tài liệu đọc bổ trợ

Bài đọc 1	Cách thức sử dụng tài liệu *
	1.1 Khóa học chứng chỉ phân tích dữ liệu Mô tả về các khóa học trong chứng chỉ phân tích dữ liệu của Google. 1.2 Tài liệu bổ trợ Mô tả về tài liệu bổ trợ 1.3 Hướng dẫn học Cách thức học và chiến lược làm các bài đánh giá Phương pháp học với tài liệu bổ trợ (đọc tài liệu + xem video)
Bài đọc 2	Dữ liệu và hệ sinh thái dữ liệu
	2.1 Dữ liệu Mô tả về dữ liệu, nguồn dữ liệu, đặc điểm và phân loại dữ liệu. Cách thức thu thập, lưu trữ Các vấn đề có thể xảy ra đối với dữ liệu Tầm quan trọng của dữ liệu đối với doanh nghiệp 2.2 Hệ sinh thái dữ liệu Định nghĩa về hệ sinh thái dữ liệu, vai trò của chúng Cung cấp các ví dụ liên quan đến thực tế 2.3 Vòng đời của dữ liệu Mô tả về vòng đời của dữ liệu và ý nghĩa của từng pha trong vòng đời

Bài đọc 3	Phân tích dữ liệu
	3.1 Giới thiệu về phân tích dữ liệu Mô tả về phân tích dữ liệu, tầm quan trọng của phân tích dữ liệu. Mô tả về phân tích khoảng trống Phân biệt sự khác nhau của phân tích dựa trên dữ liệu, dựa trên kinh nghiệm, trực giác Vai trò của chuyên gia lĩnh vực 3.2 Quy trình phân tích dữ liệu Mô tả các bước trong quy trình, đặc điểm của chúng 3.3 Ngành nghề liên quan đến phân tích dữ liệu Dựa trên nhu cầu thực tế để liên hệ ngành nghề có liên quan đến phân tích dữ liệu
Bài đọc 4	Tư duy phân tích
	4.1 Giới thiệu tư duy phân tích Khái niệm, các khía cạnh của tư duy phân tích. Vai trò của tư duy phân tích đối với nhà phân tích dữ liệu 4.2 Kỹ năng của nhà phân tích dữ liệu Mô tả các kỹ năng, liên hệ các kỹ năng đối với dữ liệu thực tế 4.3 Ra quyết định dựa trên dữ liệu Vai trò và ý nghĩa của ra quyết định dựa trên dữ liệu. Mối tương quan giữa phân tích dữ liệu và ra quyết định dựa trên dữ liệu
Bài đọc 5	Trực quan hóa dữ liệu
	5.1 Giới thiệu về trực quan hóa dữ liệu Mô tả khái niệm về trực quan hóa dữ liệu, vai trò và sức mạnh của trực quan hóa dữ liệu. Xác định các mẫu xuất hiện trong dữ liệu và mối quan hệ cần trực quan. Tóm tắt một số dạng trực quan hóa dữ liệu. 5.2 Công cụ hỗ trợ trực quan hóa dữ liệu Giới thiệu công cụ trực quan từ đơn giản cho đến các phần mềm chuyên nghiệp, thậm chí là thông qua lập trình với các thư viện hỗ trợ.
Bài đọc 6	Các công cụ hỗ trợ cho nhà phân tích dữ liệu
	6.1 Tổng quan về các công cụ

	<p>Giới thiệu các công cụ, phân loại và đặc điểm của mỗi loại</p> <p><u>6.2 Giới thiệu bảng tính</u></p> <p>Mô tả về đặc điểm và các chức năng, hàm cơ bản</p> <p><u>6.3 Truy vấn dữ liệu</u></p> <p>Giới thiệu về ngôn ngữ SQL, cách thức truy vấn dữ liệu với SQL</p>
--	---

Phần 2: Hướng dẫn trả lời câu hỏi - Quiz

Phần 1

TÀI LIỆU ĐỌC BỔ TRỢ

Bài đọc 1: Cách thức sử dụng tài liệu

1. Khóa học chứng chỉ phân tích dữ liệu

Khóa học phân tích dữ liệu nhằm trang bị cho người học các kiến thức và kỹ năng quan trọng trong lĩnh vực phân tích dữ liệu. Đây là một lĩnh vực mà hầu như tất cả các tổ chức, công ty đều cần đến để giúp họ cải tiến quy trình, xác định cơ hội và xu hướng, tạo ra sản phẩm mới và đưa ra các quyết định hợp lý. Đặc biệt, người học sẽ được giới thiệu về thế giới của phân tích dữ liệu thông qua các bài thực hành được thiết kế bởi Google.

Chứng chỉ phân tích dữ liệu được Google chia thành 8 khóa học để giúp người học có được trải nghiệm từ cơ bản đến nâng cao, từ tổng quan đến chi tiết thực hiện. Bắt đầu với khóa 1, Google giới thiệu những kiến thức tổng thể về dữ liệu và phân tích dữ liệu, vai trò của chúng trong các hoạt động doanh nghiệp. Song song với đó, các kỹ năng và công cụ cần thiết được tóm tắt để người học hình dung những gì sẽ được trải nghiệm xuyên suốt chương trình. Các khóa học sau sẽ đi vào cụ thể từng bước trong quy trình phân tích dữ liệu chẳng hạn như cách đặt câu hỏi, cách chuẩn bị dữ liệu, hay cách xử lý và làm sạch dữ liệu. Tiếp theo là các kỹ năng phân tích dữ liệu và trực quan hóa dữ liệu. Khóa học cuối cùng sẽ tổng kết và đưa ra các bài tập để người học có thể vận dụng các kiến thức và kỹ năng đã được trình bày trong các phần trước.

2. Tài liệu bổ trợ

Nhằm giúp người học tại Việt Nam có được sự hỗ trợ tốt hơn, Google và các đối tác đã nỗ lực tạo thêm các tài liệu bổ sung nhằm tóm tắt những điểm quan trọng, hệ thống lại những kiến thức mà người học cần nắm vững. Bên cạnh đó, tài liệu bổ trợ diễn giải sâu thêm các kiến thức đã được giới thiệu trong khóa học.

Tài liệu đọc là một phần trong những tài liệu được bổ sung. Nội dung các tài liệu đọc sẽ là bao gồm cả các phần trong tài liệu thuyết trình nhưng được diễn giải một cách cặn kẽ hơn, cùng với các ví dụ để giúp người học dễ dàng nắm bắt được kiến thức. Bên cạnh đó, một số tài liệu có nội dung mở rộng để cung cấp thêm kiến thức.

3. Hướng dẫn học

Sau khi đăng ký khóa học chính thức trên Coursera, người học nếu chưa có kiến thức trước đây về lĩnh vực Khoa học Dữ liệu cần trải qua các bài học theo trình tự được thiết kế của Google. Mỗi khóa học được chia theo từng tuần để người học có thể sắp xếp thời gian học phù hợp với công việc của mình. Trong mỗi tuần, Google thiết kế các video trình bày nội dung và xen kẽ đó là các tài liệu đọc thêm. Thông thường, mỗi tuần sẽ kết thúc với các bài kiểm tra đánh giá để củng cố kiến thức. Bên cạnh đó, Google cũng duy trì các diễn đàn để người học có thể thảo luận thêm các vấn đề liên quan, các vấn đề còn thắc mắc.

Trong quá trình xem các video được trình bày bởi các chuyên gia đến từ Google, người học nên xem xét thêm tài liệu bổ trợ liên quan ứng với tuần đó trước khi thực hiện các bài đánh giá. Việc này nhằm giúp củng cố và hệ thống lại kiến thức giúp người học tránh bị bỏ lỡ kiến thức nào đã được nhắc tới trong bài. Thậm chí, trong quá trình xem, nếu có chỗ nào trong khóa học gốc người học cảm thấy chưa hiểu, chúng ta có thể tìm kiếm sự giải thích trong các tài liệu bổ trợ.

Hi vọng với tài liệu bổ trợ này, chúng tôi mang lại những trải nghiệm đầy đủ hơn, gần gũi hơn đối với cộng đồng người học tại Việt Nam

Bài đọc 2: Dữ Liệu và Hệ Sinh Thái Dữ Liệu

1. Dữ liệu

Trong kỷ nguyên của máy tính, chúng ta chứng kiến sự bùng nổ của dữ liệu. Trước thời điểm đó, dữ liệu chủ yếu được lưu trữ dưới dạng giấy nên việc thu thập và lưu trữ chúng là một khó khăn. Với sự phát triển của công nghệ, chúng ta không chỉ có nhiều dữ liệu hơn mà còn đa dạng hơn, tức thời hơn. Sự tăng trưởng dữ liệu này tác động rất nhiều đến các công ty và mô hình kinh doanh của họ. Nếu công ty chậm chạp hay sai lầm khi đưa một quyết định có thể dẫn đến bị đối thủ chiếm mất thị phần. Việc đổi mới không chỉ nằm ở việc cho ra đời những sản phẩm đáp ứng thị yếu của người dùng mà còn ở dịch vụ chăm sóc khách hàng. Việc phân tích các ý kiến phản ánh để đưa ra các dịch vụ tốt hơn, làm hài lòng khách hàng hơn đôi khi lại chính là mấu chốt cho sự thành công và đứng vững trên thị trường.

Để có được những thay đổi bắt kịp thị trường, chắc chắn chúng ta cần phải hiểu về nó, biết được những gì nó đang diễn ra và sắp diễn ra. Từ đó chúng ta mới có những chiến lược phù hợp. Nhiều suy nghĩ cho rằng, chỉ cần dựa trên kinh nghiệm để nắm bắt xu thế của thị trường, của sự phát triển. Tuy nhiên, thế giới thay đổi theo nhiều quy luật khác nhau, và những gì trong quá khứ có thể không còn đúng hoàn toàn trong bối cảnh hiện tại. Hay nói cách khác, kinh nghiệm là điều kiện cần nhưng không thể đủ để chúng ta đưa ra các quyết định tốt.

Như vậy, việc phân tích dữ liệu trở thành yêu cầu cấp thiết của tất cả các doanh nghiệp ở mọi cấp độ. Và các kiến thức cũng như kỹ năng về dữ liệu là một trong những hành trang hữu ích cho những đội ngũ ứng viên cho các công ty này trong tương lai.

Mô tả về Dữ liệu

Trước khi phân tích dữ liệu, chúng ta cần tìm hiểu dữ liệu là gì? Nói một cách đơn giản, dữ liệu là tất cả các sự vật, sự kiện xung quanh chúng ta, được hệ thống máy tính lưu trữ lại và chờ xử lý. Sự vật, sự kiện nói một cách ngắn gọn là dữ kiện (fact).

Bạn có thể hỏi dữ liệu trên giấy thì thế nào. Câu trả lời là những dữ liệu trên giấy cần phải số hóa để tạo thuận lợi cho quá trình phân tích. Việc tồn tại dưới dạng giấy thường không mang lại nhiều ý nghĩa vì những quyết định tốt cần phải nhiều dữ liệu. Ngoài ra, chúng ta cũng cần sự nhanh lẹ và làm việc mọi lúc mọi nơi của hệ thống máy tính để thay thế sức con người trong một số công đoạn của quá trình phân tích.

Chính vì vậy, nhiều tổ chức trên thế giới đều đặt ra yêu cầu số hóa mọi công đoạn của họ. Không phải chỉ dừng ở sự thuận tiện cho con người trong quá trình nhập thông tin mà còn nằm ở quá trình phân tích chúng.

Nguồn dữ liệu

Mỗi hình thức kinh doanh sẽ có nguồn dữ liệu khác nhau. Ví dụ, nếu bạn mở một cửa hàng bán đồ gia dụng, bạn có thể có thông tin của khách hàng, những món hàng họ đã mua, thời điểm họ mua, v.v... và thậm chí bạn có thể biết được tâm trạng của họ. Đó là một nguồn dữ liệu. Nếu bạn mở rộng mô hình kinh doanh chẳng hạn như kinh doanh trực tuyến, bạn lại có thêm thông tin về sở thích mua hàng của họ, những người bạn họ giới thiệu, v.v... Trong mô hình kinh doanh khác như cung cấp điện đến nhà người dân, bạn có thể có được lượng tiêu thụ từng giờ của họ, sự cố vừa mới xảy ra do thiên tai, vị trí hao hụt điện năng.

Để đơn giản, chúng ta có thể tóm tắt một số nguồn có thể thu thập dữ liệu. Thứ nhất, dữ liệu đến từ mạng Internet như trang tin tức, trang bán hàng. Nguồn thứ hai đến từ quá trình giao tiếp trực tiếp với khách hàng. Nguồn thứ ba đến từ hệ thống các thiết bị cảm biến như cảm biến ánh sáng, cảm biến âm thanh, cảm biến nhiệt độ, độ ẩm.

Các dữ liệu này có thể phân thành các loại để thiết kế công cụ phù hợp xử lý nó. Các loại chính bao gồm văn bản, hình ảnh, video, và âm thanh. Trong mỗi loại, có thể tiếp tục phân thành những loại con như chữ và số trong văn bản. Mỗi loại dữ liệu đều mang lại ý nghĩa quan trọng cho quá trình phân tích. Tuy nhiên, trong giới hạn của khóa học này, chúng ta chỉ tập trung trên dữ liệu dạng văn bản.

Cách thức thu thập và lưu trữ

Sau khi xác định được nguồn thu thập, chúng ta cần thiết kế công cụ phù hợp để lấy được chúng. Một dạng tự nhiên là dữ liệu đến từ ngay các hoạt động kinh doanh hàng ngày của tổ chức, doanh nghiệp. Điều này là trường hợp bình thường. Chúng ta nói đến các dạng khác.

Làm bảng khảo sát và gửi đến những đối tượng liên quan là một dạng truyền thống để chúng ta thu thập dữ liệu. Tuy nhiên, cách này khá nhiều hạn chế như tập trung vào một thời điểm nhất định, việc điền thủ công có thể tốn thời gian cho người được thu thập và đối tượng chính là con người. Một cách thức khác là tạo ra những nơi để người dùng bình luận khi họ mong muốn. Hay tạo ra các nền tảng công nghệ để trích xuất dữ liệu từ các cảm biến và gửi về máy chủ. Tuy nhiên, dù cách thu

thập dữ liệu nào, người dùng cần được biết họ bị thu thập những gì và mục đích của việc thu thập đó cũng như họ được quyền từ chối để cung cấp.

Dữ liệu khi thu thập về cần xem xét tiếp cách thức tổ chức lưu trữ. Các tiêu chí lưu trữ cần quan tâm như khả năng lưu trữ được kích thước dữ liệu lớn, đảm bảo dữ liệu không bị mất mát bởi các sự cố nhất định, dữ liệu luôn sẵn sàng cho quá trình phân tích và các vấn đề liên quan đến an ninh, an toàn cho dữ liệu.

Các vấn đề có thể xảy ra đối với dữ liệu

Do dữ liệu đến từ nhiều nguồn nên chúng thường có nhiều vấn đề. Một số vấn đề liên quan đến quá trình phân tích như dữ liệu bị thiếu, nhiễu hay mâu thuẫn. Ví dụ, khi khảo sát sở thích của người dùng đối với sản phẩm nào đó, chúng ta quên hỏi về độ tuổi của họ hoặc có thể họ không muốn cung cấp. Điều này gây ra vấn đề thiếu dữ liệu. Ở mức độ sâu xa hơn, chúng ta thậm chí thiếu cả mẫu dữ liệu. Ví dụ, đúng ra chúng ta phải khảo sát thêm người có thu nhập thấp nhưng lại bỏ sót những đối tượng này.

Nhiều có thể đến từ lỗi khách quan và chủ quan. Ví dụ, người dùng không muốn điền thu nhập của mình nhưng chúng ta lại bắt buộc họ cung cấp, họ có thể cho một con số nào đó. Hay họ không muốn làm khảo sát mà chúng ta lại bắt họ làm, thì dữ liệu chúng ta mang về chỉ là vô ích và thậm chí còn gây cho toàn bộ mô hình phân tích của chúng ta bị lệch hướng. Các thiết bị cảm biến cũng có thể tạo ra nhiễu. Ví dụ, thiết bị cảm biến nhịp tim được gắn trên các đồng hồ thông minh có thể ghi nhận sai chỉ số mạch đập của con người. Nguyên nhân có thể do chất lượng, vật cản hay khoảng cách xa so với cổ tay.

Mâu thuẫn nghĩa là sự khác nhau của các yếu tố dữ liệu. Ví dụ, sự khác nhau của ngày sinh và độ tuổi, sự khác nhau của hai lần lấy dữ liệu của đối tượng. Bản thân mâu thuẫn có thể đến từ con người nhưng cũng có thể đến từ hệ thống thu thập hoạt động không đúng hoặc mọi thứ đều thay đổi theo thời gian.

Tầm quan trọng của dữ liệu đối với doanh nghiệp

Hoạt động sản xuất, kinh doanh hằng ngày đều góp phần tạo ra dữ liệu. Ngoài vấn đề thống kê để biết tình hình hoạt động kinh doanh ra sao, các doanh nghiệp còn đối diện với bài toán về sự đổi mới. Nếu không đổi mới, doanh nghiệp đối diện với việc bị tụt hậu vì cạnh tranh thị trường ngày càng gay gắt. Đổi mới cần diễn ra ở mọi mặt từ sản phẩm, dịch vụ cho đến quy trình sản xuất. Để có được các đổi mới, những người đứng đầu doanh nghiệp vừa cần phải ra các quyết định gần như tức

thời vừa phải đảm bảo các quyết định đúng đắn. Dữ liệu và phân tích dữ liệu là một yếu tố quan trọng để đạt được cả hai mục tiêu đó.

Ví dụ, công ty Microsoft đã phân tích dữ liệu và nhận thấy mọi người có nhiều khả năng cộng tác hơn khi họ ở gần nhau hơn. Sau đó, họ tiến hành sắp xếp lại tổ chức để giảm khoảng cách mà nhân viên phải di chuyển đến cuộc họp. Kết quả thu được là mỗi tuần họ tiết kiệm được 100 giờ, tương đương 520.000 đô la, số cuộc họp cũng tăng lên từ 14 lên tới 18.

2. Hệ sinh thái dữ liệu

Hệ sinh thái dữ liệu (data ecosystem) là tổng hòa của nhiều thành phần từ cơ sở hạ tầng cho đến các phần mềm hỗ trợ và cả người sử dụng chúng. Chúng phối hợp và tương tác với nhau để có thể tạo ra dữ liệu, quản lý dữ liệu, lưu trữ và tổ chức chúng một cách khoa học. Quá trình sau đó là phân tích dữ liệu này để rút ra được các kết luận quan trọng và chia sẻ các kết luận với các bên liên quan. Đám mây đóng một vai trò quan trọng trong hệ sinh thái dữ liệu và với vai trò là một nhà phân tích dữ liệu, nhiệm vụ của chúng ta là khai thác sức mạnh của hệ sinh thái dữ liệu đó, tìm thông tin phù hợp và cung cấp cho nhóm phân tích giúp họ đưa ra quyết định thông minh.

Ví dụ, cơ sở dữ liệu của cửa hàng bán lẻ là một hệ sinh thái chứa tên khách hàng, địa chỉ, các giao dịch mua trước đó và đánh giá của khách hàng. Là một nhà phân tích dữ liệu, bạn có thể sử dụng thông tin này để dự đoán những khách hàng này sẽ mua gì trong tương lai và đảm bảo cửa hàng có sản phẩm và còn hàng khi họ cần.

Một ví dụ nữa, hệ sinh thái dữ liệu được sử dụng bởi một bộ phận nhân sự. Hệ sinh thái này sẽ bao gồm thông tin như các bài đăng từ các trang web việc làm, số liệu thống kê về thị trường lao động hiện tại, tỷ lệ việc làm và dữ liệu truyền thông xã hội về nhân lực tiềm năng. Một nhà phân tích dữ liệu có thể sử dụng thông tin này để giúp nhóm của họ tuyển dụng nhân viên mới và cải thiện tỷ lệ tương tác và giữ chân của nhân viên.

Tuy nhiên, hệ sinh thái dữ liệu không chỉ dành cho các cửa hàng và văn phòng. Chúng còn có mặt trong lĩnh vực nông nghiệp. Các công ty nông nghiệp thường xuyên sử dụng hệ sinh thái dữ liệu bao gồm thông tin địa chất trong các chuyển động của thời tiết. Các nhà phân tích dữ liệu có thể sử dụng dữ liệu này để giúp nông dân dự đoán năng suất cây trồng. Một số nhà phân tích dữ liệu thậm chí đang

sử dụng hệ sinh thái dữ liệu để cứu các hệ sinh thái môi trường thực. Tại viện hải dương học Scripps, các rạn san hô trên khắp thế giới được theo dõi. Qua đó họ có thể xem các sinh vật thay đổi như thế nào theo thời gian, theo dõi sự phát triển của chúng và đo lường bất kỳ sự gia tăng hoặc suy giảm nào trong các thuộc địa riêng lẻ.

3. Vòng đời của dữ liệu

Dữ liệu không phải tồn tại mãi ở một trạng thái ban đầu lúc chúng ta tạo ra hay thu thập, mà chúng đi qua những chặng đường biến đổi, được xử lý bởi những bước trong quá trình phân tích, và thậm chí là bị phá hủy. Chúng ta gọi nó là vòng đời dữ liệu (data life cycle). Việc mô tả rõ ràng về vòng đời dữ liệu giúp chúng ta có những xử lý thích hợp để có thể tận dụng tất cả khía cạnh của chúng cũng như giúp cho quá trình phân tích trở nên thuận lợi hơn.

Vòng đời dữ liệu được thể hiện qua 6 pha chính gồm lên kế hoạch, thu thập, quản lý, phân tích, đóng gói, và phá hủy.

- Lên kế hoạch dữ liệu là việc lập ra kế hoạch cụ thể cho bài toán, các quyết định mà chúng ta sẽ làm trên dữ liệu, loại dữ liệu cần có, cách chúng ta quản lý chúng xuyên suốt vòng đời. Ngoài ra, lên kế hoạch còn phải phân công người chịu trách nhiệm về dữ liệu và kết quả mà chúng ta kỳ vọng từ dữ liệu. Bước này có thể xem như khá quan trọng vì nó đảm bảo chúng ta có đúng và đủ dữ liệu cho quá trình phân tích về sau.
- Trong pha thu thập, chúng ta cần trả lời một số câu hỏi trước khi thực hiện thu thập dữ liệu như dữ liệu được lấy từ nơi nào, bằng cách nào, thu thập từ một nguồn hay nhiều nguồn khác nhau, cách chúng được đưa vào hệ thống quản lý của chúng ta như thế nào. Sau khi đã định hình những điều này, chúng ta sẽ tiến hành thu thập dữ liệu.
- Khi đã có dữ liệu, chúng ta cần tổ chức quản lý chúng. Các vấn đề chính cần quan tâm trong pha quản lý dữ liệu gồm cách thức lưu trữ, cách thức sử dụng, công cụ và phương pháp để giữ chúng an toàn, bảo mật và những hành động để bảo trì dữ liệu thích hợp.

- Pha phân tích là pha chính sử dụng dữ liệu để phân tích nhằm mục tiêu ra các quyết định phục vụ mục tiêu kinh doanh. Trong pha này, chúng ta cũng cần chuẩn bị các công cụ và công nghệ để hỗ trợ cho quá trình phân tích.
- Sau khi phân tích, liệu chúng ta có cần dùng dữ liệu này cho các phân tích khác không, có dùng chúng trong tương lai không? Đóng gói sẽ giúp dữ liệu gọn gàng hơn, tập trung vào dữ liệu quan trọng hơn, v.v... Chúng ta cũng cần xác định thời điểm nào để thực hiện việc đóng gói này.
- Pha cuối cùng trong vòng đời dữ liệu là phá hủy. Trong pha này, chúng ta trả lời câu hỏi như dữ liệu có cần xóa không, nếu có thì khi nào và xóa ra sao. Đôi khi chúng ta cần phải xóa để bảo vệ tính riêng tư của người dùng.

Bài đọc 3: Phân Tích Dữ Liệu

1. Giới thiệu về phân tích dữ liệu

Phân tích dữ liệu và tầm quan trọng

Phân tích dữ liệu (data analysis) là quá trình chúng ta đi thu thập dữ liệu, sau đó biến đổi chúng về dạng phù hợp, chứa chúng một cách có tổ chức nhằm mục đích có thể rút ra được kết luận có ý nghĩa, xa hơn là có thể dự đoán những gì sẽ xảy ra cũng như dựa trên đó chúng ta đưa ra quyết định nhằm đạt được những lợi ích trong kinh doanh.

Rõ ràng chúng ta có thể làm được nhiều thứ sau khi phân tích nhưng điều trước tiên là sự thấu hiểu về dữ liệu chúng ta đang có. Đó có thể là sự thấu hiểu về tình trạng của hoạt động kinh doanh hiện tại, về bối cảnh thị trường, về người sử dụng, v.v... Tuy nhiên, thứ quan trọng là dẫn ta đến một kết luận hay dự đoán. Đó là những quyết định cuối cùng dành cho những sản phẩm mới, dịch vụ mới của công ty, doanh nghiệp. Những quyết định này hoàn toàn dựa trên dữ liệu mà chúng ta phân tích. Qua đó, chúng ta có thuật ngữ data-driven decision-making, nghĩa là ra quyết định dựa trên dữ liệu. Một thuật ngữ khác cũng có ý tương đương đó là chiến lược kinh doanh sử dụng dữ kiện (business strategy using facts).

Phân tích khoảng trống

Các khía cạnh trong tư duy phân tích giúp chúng ta có nhiều cách tiếp cận khác nhau và cùng phối hợp để có được chiến lược tư duy tốt. Tuy nhiên, các khía cạnh này lại không cho ta biết nên bắt đầu từ đâu trên con đường tìm ra giải pháp cho một bài toán hay vấn đề cho trước.

Điểm bắt đầu thông thường mà các nhà phân tích sử dụng là đặt câu hỏi. Và đương nhiên, chúng ta sẽ hỏi là câu hỏi gì? Câu trả lời cho câu hỏi này là “Nguyên nhân nảy sinh ra vấn đề?” hay cách nói khác “Tại sao lại xảy ra vấn đề?” Để trả lời cho câu hỏi này chúng ta cần trải qua phương pháp: 5 lần hỏi “tại sao”. Ví dụ, giả sử chúng ta gặp phải vấn đề là “hôm nay, tôi không thể làm bánh mứt dâu”. Vậy một loạt câu hỏi tại sao: “Tại sao tôi không làm được bánh mứt dâu”. Câu trả lời, do không có dâu, vậy tại sao tôi không có dâu? Vì các cửa hàng hết hàng, v.v...

Ngoài cách đặt câu hỏi để tìm ra nguyên nhân gốc của vấn đề và giải quyết nó thì một cách tiếp cận khác là trả lời câu hỏi “Liệu có một hố hay một khoảng trống ngăn cách giữa chúng ta với mục tiêu muốn đạt đến không?” và “Khoảng cách hay hố đó là gì?”. Từ đó ra đời thuật ngữ phân tích khoảng trống (gap analysis). Phân tích

khoảng trống đó là quá trình kiểm tra và đánh giá trạng thái hiện tại của hệ thống và xác định mục tiêu tốt hơn muốn đạt đến. Sau đó, tìm cách để lấp đầy khoảng trống này. Ví dụ, chúng ta phân tích khoảng trống để cải thiện khả năng truy cập một ứng dụng giáo dục dựa trên chức năng hiện tại của nó, hay tăng hiệu quả của nhà sản xuất ô tô dựa trên quy trình lắp ráp hiện tại của họ. Một ví dụ khác liên quan đến môi trường, phân tích lượng khí thải carbon của một công ty đang ở mức bao nhiêu và mục tiêu để giảm nó trong tương lai.

Phân tích dựa trên dữ liệu và dựa trên kinh nghiệm, trực giác

Bên cạnh các quyết định dựa trên dữ liệu, chúng ta còn thấy nhiều quyết định trong cuộc sống dựa trên trực giác mách bảo hay linh tính mách bảo. Đó được xem như một cách nhìn chủ quan với ít hoặc không có sự giải thích một cách rõ ràng hợp lý. Một dạng quyết định khác nữa là quyết định dựa trên kinh nghiệm hay dựa trên kiến thức về một lĩnh vực đã đúc kết qua nhiều năm tháng. Những người có kiến thức hay kinh nghiệm được xem là các chuyên gia về lĩnh vực (subject matter expert). Các chuyên gia lĩnh vực sẽ giúp chúng ta rất nhiều thứ như họ đọc kết quả và chỉ ra cho chúng ta những chỗ chưa chặt chẽ của kết quả phân tích, hay phát hiện ra các góc khuất mà chúng ta chưa chạm tới mặc dù nó có thể vẫn ở đó trong dữ liệu. Và kể cả khi mọi thứ đều ổn, việc góp ý cho các lựa chọn và đề xuất hướng đi cho bài toán trở thành những điểm mà chúng ta không thể thiếu ở các quyết định cuối cùng.

Vậy câu hỏi đặt ra là khi ra quyết định, chúng ta sẽ dựa trên cái nào? Câu trả lời ngắn gọn đó là chúng ta nên kết hợp chúng lại để có được hiệu quả cao nhất. Vì việc hiểu biết về mô hình kinh doanh sẽ giúp nắm bắt được nhu cầu của phân tích, định hướng cho phân tích luôn đi đúng hướng. Và đương nhiên trong mọi tình huống thì việc quyết định đều cần phải có sự tham gia của dữ liệu, đều cần phải dựa trên dữ liệu.

Cách thức phối hợp có thể tiến hành theo cách sau. Khi quá trình phân tích dựa trên dữ liệu hoàn thành và trước khi công bố quyết định đến khách hàng, chúng ta cần có các hoạt động như chia sẻ kết quả với các chuyên gia, đồng nghiệp trong lĩnh vực, tham khảo thêm ý kiến của họ để đảm bảo những gì được rút trích từ dữ liệu được hiểu một cách đúng đắn và phù hợp với mô hình kinh doanh.

2. Quy 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. Tiến trình phân tích được chia thành 6 pha chính gồm pha hỏi, pha chuẩn bị, pha xử lý, pha phân tích, pha chia sẻ, pha hành động. Và sau đây chúng ta sẽ cùng tìm hiểu chính xác mỗi pha là gì và cần làm gì trong mỗi pha đó.

Pha hỏi

Pha hỏi chính là điểm đầu tiên mà các nhà phân tích cần đi qua, mục tiêu của pha này là đặt câu hỏi để hiểu về bài toán mình đang giải quyết là gì, mục đích hay mong đợi của các bên liên quan, loại dữ liệu cần thu thập, ... Đặc biệt, trong pha này chúng ta cần biết mình đang ở đâu hay trạng thái hiện tại. Nơi nào mình muốn đạt đến hay trạng thái lý tưởng của bài toán và khoảng cách giữa hiện tại và mong đợi nó như thế nào. Tất cả điều này sẽ giúp bài toán đi đúng hướng và có thể lên kế hoạch để giải quyết chúng.

Để minh họa, chúng ta cùng xem xét trên một ngữ cảnh sau. Một tổ chức đang có tỷ lệ doanh thu cao từ các nhân viên mới. Tuy nhiên, nhiều nhân viên đã rời công ty trước khi kết thúc năm đầu tiên làm việc. Các nhà phân tích đã sử dụng quá trình phân tích dữ liệu để trả lời câu hỏi sau: Làm thế nào tổ chức có thể cải thiện tỷ lệ giữ chân nhân viên mới?

Đầu tiên, các nhà phân tích cần xác định dự án trông như thế nào và điều gì sẽ đủ điều kiện để đạt được kết quả thành công. Vì vậy, để xác định những điều này, họ đã đặt những câu hỏi hiệu quả và hợp tác với các nhà lãnh đạo và quản lý, những người quan tâm đến kết quả phân tích nhân sự của họ. Đây là những loại câu hỏi họ có thể đặt ra:

- Bạn nghĩ nhân viên mới cần học gì để thành công trong năm đầu tiên làm việc?
- Bạn đã thu thập dữ liệu từ nhân viên mới trước đây chưa? Nếu vậy, chúng tôi có thể có quyền truy cập vào dữ liệu lịch sử không?
- Bạn có tin rằng những người quản lý có tỷ lệ giữ chân cao hơn sẽ cung cấp cho nhân viên mới điều gì đó bổ sung hoặc độc đáo không?
- Bạn nghi ngờ điều gì là nguyên nhân hàng đầu khiến nhân viên mới không hài lòng?
- Bạn muốn tỷ lệ giữ chân nhân viên tăng lên bao nhiêu phần trăm trong năm tài chính tiếp theo?

Pha chuẩn bị

Pha chuẩn bị là pha tiến hành thu thập và lưu trữ dữ liệu để chuẩn bị cho quá trình phân tích. Trong pha này, chúng ta cần thiết kế các công cụ để thu thập dữ liệu, sau đó tiến hành phân loại và lưu trữ chúng ở dạng thích hợp. Chúng ta cũng cần đánh giá sơ bộ liệu dữ liệu có giúp giải quyết vấn đề không, dữ liệu có khách quan và không bị lệch không, v.v...

Chúng ta tiếp tục trên ngữ cảnh liên quan đến phân tích dữ liệu để tìm cách giữ chân nhân viên mới. Đội ngũ phân tích xác định dữ liệu họ cần để đạt được kết quả thành công mà họ đã xác định ở bước trước. Trong trường hợp này, các nhà phân tích chọn thu thập dữ liệu từ một cuộc khảo sát trực tuyến về nhân viên mới. Đây là những thứ họ làm để chuẩn bị:

- Họ phát triển các câu hỏi cụ thể để hỏi về mức độ hài lòng của nhân viên với các quy trình kinh doanh khác nhau, chẳng hạn như tuyển dụng và giới thiệu, cũng như mức lương thưởng.
- Họ thiết lập các quy tắc về những người sẽ có quyền truy cập vào dữ liệu được thu thập - trong trường hợp này, bất kỳ ai bên ngoài nhóm sẽ không có quyền truy cập vào dữ liệu thô, nhưng có thể xem dữ liệu tóm tắt hoặc tổng hợp. Ví dụ: sẽ không biết thù lao cho mỗi cá nhân, nhưng có thể xem được mức lương cho các nhóm cá nhân.
- Các nhà phân tích đã suy nghĩ về các vấn đề có thể có liên quan đến dữ liệu và dự án cũng như cách tránh chúng.

Pha xử lý

Pha xử lý tập trung vào đánh giá dữ liệu thu thập được, kiểm tra các vấn đề có xảy ra đối với dữ liệu không. Một số vấn đề có thể như dữ liệu bị thiếu, nhiễu, sai, mâu thuẫn. Ngoài ra tính cân bằng hay công bằng giữa các nhóm dữ liệu cũng cần được xem xét. Nếu vấn đề xảy ra, chúng ta cần thực hiện các bước để khắc phục nó để chúng ta có được dữ liệu tốt, dữ liệu sạch trước khi phân tích chúng.

Đối với ngữ cảnh ví dụ, nhóm phân tích gửi bản khảo sát đến các nhân viên. Vì nhân viên cần cung cấp dữ liệu nên điều quan trọng là phải đảm bảo tất cả nhân viên đồng ý tham gia. Các nhà phân tích dữ liệu cũng đảm bảo rằng nhân viên hiểu cách dữ liệu của họ sẽ được thu thập, lưu trữ, quản lý và bảo vệ. Để duy trì tính bảo mật cũng như bảo vệ và lưu trữ dữ liệu một cách hiệu quả, họ đã thực hiện các bước sau:

- Hạn chế quyền truy cập vào dữ liệu đối với một số nhà phân tích.

- Làm sạch dữ liệu để đảm bảo dữ liệu đầy đủ, chính xác và phù hợp. Một số dữ liệu nhất định đã được tổng hợp và tóm tắt mà không tiết lộ thông tin riêng lẻ.
- Tải dữ liệu thô lên kho dữ liệu nội bộ để có thêm một lớp bảo mật.

Pha phân tích

Sau khi đã có dữ liệu tốt, chúng ta sẽ áp dụng các kỹ thuật để phân tích chúng. Các nhà phân tích cũng cần vận dụng các công cụ hỗ trợ như bảng tính, ngôn ngữ truy vấn để biến đổi dữ liệu về dạng mà có thể đưa ra các kết luận và các quyết định cho bài toán đặt ra.

Đối với tình huống trên, sau khi các cuộc khảo sát đã hoàn thành, các nhà phân tích dữ liệu đã phát hiện ra rằng trải nghiệm của một nhân viên với các quy trình nhất định là một chỉ số quan trọng để đánh giá sự hài lòng tổng thể trong công việc. Đây là những phát hiện của họ:

- Những nhân viên đã trải qua một quá trình tuyển dụng phức tạp và lâu dài có nhiều khả năng rời công ty.
- Những nhân viên đã trải qua quá trình đánh giá và phản hồi hiệu quả và minh bạch có nhiều khả năng ở lại với công ty.
- Nhóm biết rằng điều quan trọng là phải ghi lại chính xác những gì họ tìm thấy trong phân tích, bất kể kết quả như thế nào. Làm theo cách khác sẽ làm giảm sự tin tưởng vào quá trình khảo sát và giảm khả năng thu thập dữ liệu trung thực từ nhân viên của họ trong tương lai.

Pha chia sẻ

Kết quả sau khi phân tích có thể phức tạp để hiểu cho những người không phải chuyên gia, thậm chí kể cả những người là chuyên gia nhưng họ không theo đuổi bài toán từ đầu. Các công cụ trực quan hóa như biểu đồ, đồ thị sẽ tạo ra các bản thể hiện đơn giản và dễ hiểu để có thể chia sẻ kết quả cho các bên liên quan.

Đây là cách các nhà phân tích dữ liệu trong tình huống ví dụ chia sẻ những phát hiện của mình:

- Trình bày kết quả với các nhà quản lý để đảm bảo rằng họ có bức tranh đầy đủ.
- Yêu cầu những người quản lý phải đích thân giao kết quả cho đội của họ.

Quá trình này đã tạo cơ hội cho các nhà quản lý truyền đạt kết quả với bối cảnh phù hợp. Do đó, họ có thể có các cuộc trò chuyện nhóm hiệu quả về các bước tiếp theo để cải thiện sự ở lại của nhân viên mới.

Pha hành động

Cuối cùng là pha hành động. Sau khi đã đạt được sự đồng thuận của các bên liên quan, chúng ta sẽ tiến hành đem những hiểu biết hay sự thấu hiểu về dữ liệu, các quyết định dựa trên dữ liệu vào triển khai thực tế. Mục tiêu kỳ vọng là chúng có thể giải quyết bài toán đã đề ra. Một số công việc thường thực hiện trong pha hành động như kiểm tra, đánh giá lại lần nữa những hiểu biết được rút ra, từ đó chọn lựa ra một chiến lược hành động và triển khai hành động đó.

Một số đề xuất của các nhà phân tích dữ liệu cho tình huống ví dụ như sau:

- Chuẩn hóa quy trình tuyển dụng và đánh giá nhân viên dựa trên các quy định minh bạch và hiệu quả nhất.
- Tiến hành cùng một cuộc khảo sát hàng năm và so sánh kết quả với những kết quả từ năm trước.

Một năm sau, bản khảo sát tương tự được gửi cho các nhân viên. Các nhà phân tích dự đoán rằng sự so sánh giữa hai bộ kết quả sẽ chỉ ra rằng kế hoạch hành động có hiệu quả ra sao.

3. Ngành nghề liên quan đến phân tích dữ liệu

Có rất nhiều vị trí, vai trò liên quan đến nghề phân tích dữ liệu. Họ có thể thực hiện chuyên một phần hay toàn bộ giai đoạn trong quá trình phân tích dữ liệu. Ngoài ra, mỗi lĩnh vực sẽ có đặc trưng riêng, đòi hỏi nhà phân tích dữ liệu biến đổi để phù hợp với lĩnh vực đó. Chúng ta liệt kê ở đây một số chức danh nghề nghiệp có liên quan đến phân tích dữ liệu. Cụ thể:

- Nhà phân tích kinh doanh: xem xét các báo cáo doanh thu, đánh giá và phân tích để cải thiện mô hình kinh doanh
- Nhà hỗ trợ phân tích dữ liệu: chuẩn bị dữ liệu như thu thập, tiền xử lý, làm sạch dữ liệu, hỗ trợ chạy mô hình và phân tích các số liệu
- Nhà phân tích thực thi: thu thập dữ liệu liên quan đến quá trình triển khai sản phẩm, các ý kiến từ khách hàng, các độ đo hiệu năng sản phẩm để đưa ra các kế hoạch hành động.
- Nhà phân tích tuyển dụng: thu thập các thông tin tuyển dụng từ thị trường, xem xét các phản hồi từ nhân viên, đánh giá chất lượng đội ngũ, phân tích sự rời đi của nhân viên. Từ đó, cải thiện môi trường làm việc và chất lượng đội ngũ tham gia.
- Nhà phân tích tài chính: xem xét và đánh giá tất cả các hoạt động liên quan đến tình hình tài chính của công ty, các khoản chi trả xuyên suốt hoạt động, các

khoản kinh doanh để từ đó tìm các vấn đề có thể gặp phải, đưa ra các dự báo tài chính trong thời gian sắp tới.

- Nhà phân tích rủi ro: thu thập mọi quy trình, các biến động trên thị trường để đánh giá rủi ro mà công ty có thể gặp phải trong ngắn hạn và dài hạn. Qua đó, đề xuất và đánh giá các giải pháp phòng tránh các rủi ro có thể xảy ra.

Bài đọc 4: Tư Duy Phân Tích

1. Giới thiệu 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. Cụ thể, tư duy phân tích bao gồm các bước xác định và định nghĩa bài toán. Sau đó, giải quyết bài toán bằng cách sử dụng dữ liệu một cách có tổ chức và theo trình tự từng bước.

Khi nói về tư duy phân tích, chúng ta thường đề cập 5 khía cạnh chính của nó gồm sự trực quan, chiến lược tư duy, tính hướng bài toán, độ tương quan, tư duy bức tranh lớn và tư duy hướng chi tiết. Trực quan đó là cách chúng ta thể hiện thông tin dưới dạng đồ họa như biểu đồ (chart), đồ thị (graph), bản đồ (map), v.v... Trong khi đó, chiến lược tư duy tập trung vào xác định những gì mình muốn đạt và cách để đạt nó. Hướng bài toán là xác định, mô tả và tập trung vào bài toán cần giải quyết. Khía cạnh độ tương quan lại đi sâu vào mô tả mối quan hệ của hai hay nhiều mảnh của dữ liệu. Tư duy bức tranh lớn là cách nhìn bài toán một cách tổng thể để xác định tính khả thi và cơ hội cho mô hình hiện tại. Ngược lại, tư duy hướng chi tiết lại tập trung vào chi tiết để xác định tính phù hợp, tương thích của mỗi phần, cũng như chuẩn bị cho những phần kế tiếp. Mỗi khía cạnh này đều góp phần làm cho tư duy được rõ ràng và liền mạch.

2. Kỹ năng của nhà phân tích dữ liệu

Kỹ năng phân tích có thể nói một cách tóm tắt đó là những phẩm chất và đặc điểm cần có để chúng ta giải quyết vấn đề với những dữ liệu, dữ kiện mà chúng ta thu thập được. Để hoàn thành việc gì tốt, chúng ta đều cần phải có những kiến thức và kỹ năng nhất định. Phân tích dữ liệu cũng vậy, một bộ kỹ năng cần được định nghĩa để giúp cho các nhà phân tích tương lai đạt được mục tiêu của mình.

Có 5 kỹ năng liên quan trực tiếp đến phân tích dữ liệu, đó là sự tò mò, hiểu biết về ngữ cảnh, nắm bắt được kỹ thuật, có khả năng thiết kế, tổ chức dữ liệu, cũng như có các chiến lược dữ liệu hợp lý. Chúng ta cũng đi sâu vào mô tả từng kỹ năng:

- Kỹ năng đầu tiên để thành công đó là sự tò mò. Có thể nói chính nhờ sự tò mò cộng với nỗ lực hết mình, chúng ta có được những bước tiến quan trọng trong lịch sử. Tò mò đó là sự khao khát được học những điều mới mẻ trong cuộc sống, tìm kiếm cho mình những thách thức mới, thậm chí đó những trải nghiệm mới.
- Kỹ năng thứ hai là hiểu về ngữ cảnh. Kỹ năng này đòi hỏi chúng ta nắm bắt được bối cảnh, đặc điểm của những thứ mà chúng ta đang làm việc với. Qua đó, chúng ta có thể xâu chuỗi chúng lại để thấy được toàn bộ bức tranh tổng thể. Điều này giúp chúng ta tránh bỏ sót hay hiểu sai về dữ liệu hay bài toán mà chúng ta đang theo đuổi. Một số ví dụ về hiểu ngữ cảnh chẳng hạn chúng ta mô tả thông tin về dữ liệu như loại, tính chất, ngữ nghĩa. Một ví dụ khác là nắm bắt được động cơ của việc thu thập dữ liệu trước khi tiến hành.
- Tư duy kỹ thuật đó là khả năng chia nhỏ bài toán thành các bài toán nhỏ hơn, sau đó xử lý chúng một cách có thứ tự và hợp lý.
- Thiết kế dữ liệu đòi hỏi chúng ta biết cách tổ chức thông tin một cách rõ ràng và hợp lý. Để dễ nhớ, chúng ta có thể liên tưởng thiết kế dữ liệu đến việc thiết kế một cơ sở dữ liệu thỏa mãn các chuẩn.
- Chiến lược dữ liệu đó là chiến lược hay cách thức mà chúng ta quản lý con người để đảm bảo họ dùng đúng dữ liệu, đúng tiến trình và đúng công nghệ để giải quyết bài toán cho trước.

3. Ra quyết định dựa trên dữ liệu

Ra quyết định dựa trên dữ liệu là cách thức mà chúng ta luôn sử dụng dữ kiện để đưa đến các quyết định, các kết quả đầu ra. Khi xem xét bài toán trên dữ liệu, chúng ta có thể hiểu biết cận kề hơn về bài toán mà chúng ta đang giải quyết, chúng ta có thể kiểm chứng những giả thuyết được đề ra. Qua dữ liệu, chúng ta sẽ thấy rõ hơn về cơ hội lẫn thách thức của bài toán, từ đó giúp đỡ lên kế hoạch hành động để đạt được mục tiêu mong đợi cuối cùng. Khi ra quyết định dựa trên dữ liệu, chúng ta có căn cứ cho mọi quyết định của mình, dễ dàng lý giải tại sao đó là quyết định phù hợp và kết quả là chúng ta sẽ tự tin hơn khi thực hiện các hành động. Các quyết định này cũng làm cho việc định hướng đến lời giải một cách tốt hơn, chúng ta cũng chủ động khi biết những gì cần thực hiện kế tiếp. Bên cạnh đó, quyết định dựa trên dữ liệu giúp chúng ta tiết kiệm thời gian và nguồn lực để đi đến mục tiêu mong muốn. Nếu không có dữ liệu, chúng ta không biết bắt đầu từ đâu, cái gì sẽ định hướng, nếu dựa trên kinh nghiệm thì có thể kinh nghiệm đó không phù hợp với ngữ cảnh hiện tại và dẫn đến việc thử và sai gây tốn kém. Người ta nhận thấy giữa các kỹ năng phân

tích dữ liệu giúp cho việc ra quyết định hướng dữ liệu một cách tốt hơn và ngược lại quyết định hướng dữ liệu giúp nhà phân tích có thông tin để cải thiện các kỹ năng phân tích của mình.

Bài đọc 5: Trực Quan Hóa Dữ Liệu

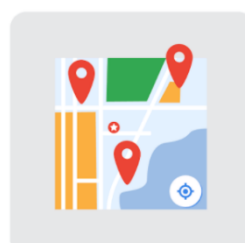
1. Giới thiệu về trực quan hóa dữ liệu, đặc điểm và ý nghĩa

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. Bằng cách trực quan, chúng ta có thể mô tả rõ ràng các quy luật và xu hướng của dữ liệu như cách chúng thay đổi, những gì nổi trội, v.v... Trực quan cũng giúp chúng ta có thể hiểu những dữ liệu phức tạp một cách nhanh chóng cũng như nắm bắt được mối quan hệ giữa các đối tượng có trong dữ liệu.

Trong quá trình phân tích dữ liệu, chúng ta thường trải qua các bước như tìm ra các quy luật trong dữ liệu qua quá trình phân tích. Sau đó, chúng ta mong muốn thể hiện quy luật này, thay vì trình chiếu một loạt các bảng dữ liệu và trình bày trên đó, chúng ta cần trực quan để tạo sự dễ hiểu cho các bên liên quan. Do đó, chúng ta lên kế hoạch những gì chúng ta muốn thể hiện và sau đó tiến hành tạo ra các bản trực



Line charts can track sales over time



Maps can connect sales to locations



Donut charts can show customer segments



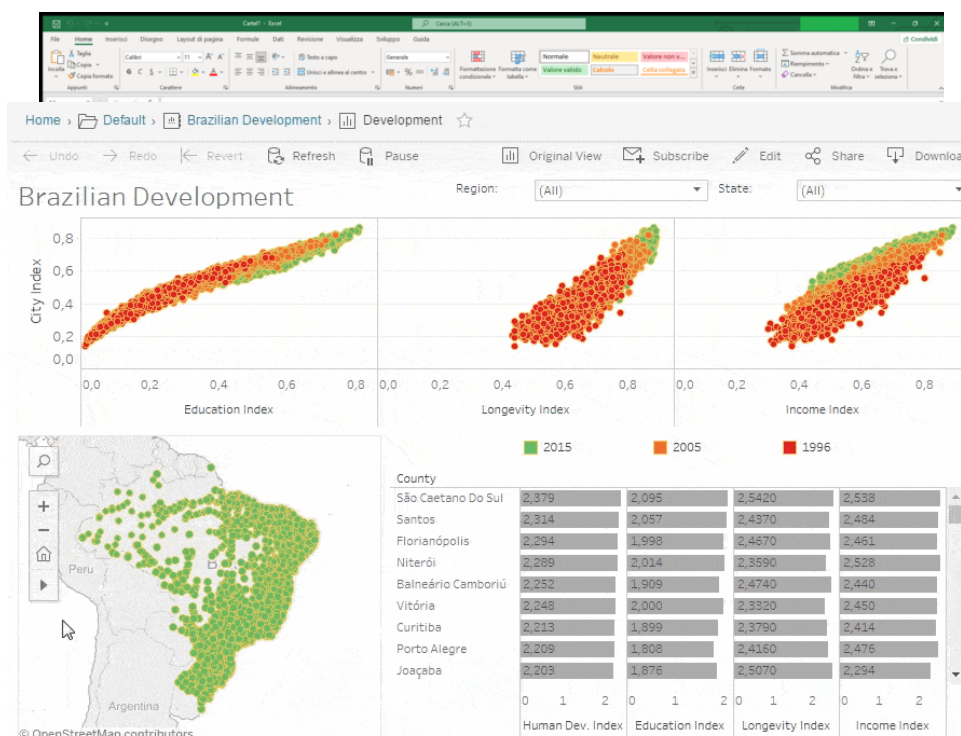
Bar charts can compare total visitors and visitors that make a purchase

quan này.

Có rất nhiều loại biểu đồ để chúng ta có thể trực quan. Trong khuôn khổ khóa học này, chúng ta tập trung trên 4 loại đó là biểu đồ đường, biểu đồ cột, biểu đồ điểm và biểu đồ tròn. Biểu đồ đường tập trung thể hiện sự thay đổi theo thời gian của đối tượng đang xem xét ví dụ tuổi của một người. Trong khi đó, biểu đồ cột lại hướng đến so sánh giá trị thống kê của các loại hay các lớp đối tượng khác nhau. Ví dụ chúng ta muốn so sánh sự khác nhau của tuổi trung bình của nam và nữ. Biểu đồ điểm thể hiện mối tương quan giữa 2 hay nhiều thuộc tính của dữ liệu, ví dụ như chúng ta muốn xem tương quan giữa tuổi và thu nhập, liệu tuổi càng lớn thì thu nhập càng cao hay không? Số lượng điểm dữ liệu nằm ở vùng nào sẽ kết luận cho nhận định đó. Khi chúng ta muốn xác định các thành phần trong một tổng thể, tỷ trọng mỗi thành phần thể nào thì chúng ta sử dụng biểu đồ tròn. Ví dụ, chúng ta muốn biết tỉ lệ các lứa tuổi trong một tổ chức như thế nào, ta dùng biểu đồ tròn mà ở mỗi phần trên hình tròn đó thể hiện tỉ trọng theo từng lứa tuổi.

2. Công cụ hỗ trợ trực quan hóa dữ liệu

Một số công cụ trực quan hóa phổ biến như bảng tính, Tableau, Looker.

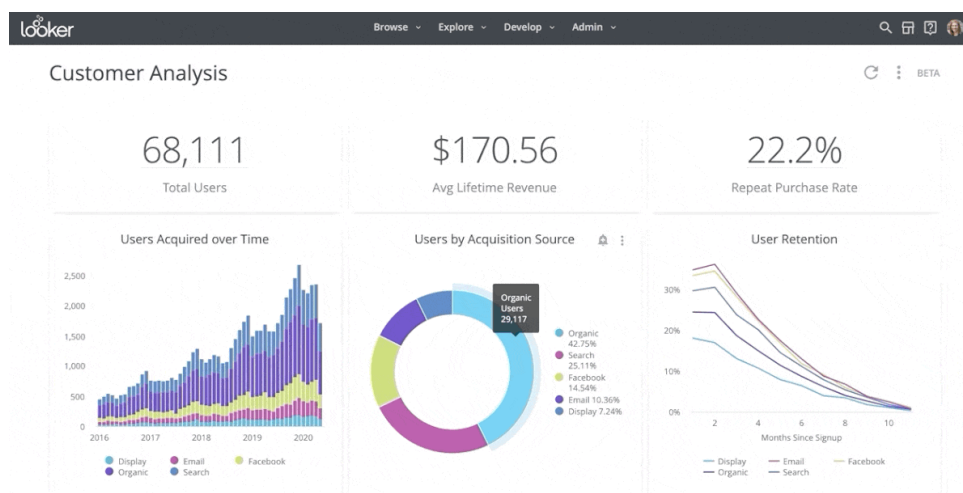


Về bảng tính, có rất nhiều giải pháp bảng tính khác nhau, nhưng hai lựa chọn phổ biến là và Google Trang tính. Bảng tính là một bảng số liệu kỹ thuật số. Nó lưu trữ, tổ chức và sắp xếp dữ liệu. Khi đưa dữ liệu vào bảng tính, chúng ta có thể thấy các

mẫu, thông tin nhóm và dễ dàng tìm thấy thông tin cần. Bảng tính cũng có một số tính năng thực sự hữu ích được gọi là công thức và hàm. Công thức là một tập hợp các hướng dẫn thực hiện một phép tính cụ thể bằng cách sử dụng dữ liệu trong bảng tính. Công thức có thể làm những việc cơ bản như cộng, trừ, nhân và chia, nhưng chúng không chỉ dừng ở đó. Chúng ta cũng có thể sử dụng công thức để tìm giá trị trung bình của một tập hợp số. Tra cứu một giá trị cụ thể, trả về tổng của một tập hợp các giá trị đáp ứng một quy tắc cụ thể, v.v. Hàm là một lệnh đặt trước tự động thực hiện một quy trình hoặc tác vụ cụ thể bằng cách sử dụng dữ liệu trong bảng tính. Bên cạnh đó, bảng tính cũng trang bị các chức năng vẽ biểu đồ dựa trên dữ liệu có trong từng thẻ. Hầu hết các biểu đồ được đề cập trong phần trước đều được tích hợp vào bảng tính.

Công cụ phổ biến thứ hai là Tableau. Tableau được thành lập vào năm 2003 là kết quả của một dự án khoa học máy tính tại Stanford nhằm mục đích cải thiện luồng phân tích và làm cho dữ liệu dễ tiếp cận hơn với mọi người thông qua trực quan. Các nhà phân tích dữ liệu thích sử dụng Tableau vì nó giúp họ tạo ra những hình ảnh trực quan rất dễ hiểu. Điều này có nghĩa là ngay cả những người dùng không chuyên về kỹ thuật cũng có thể nhận được thông tin họ cần.

Looker cũng phổ biến với các nhà phân tích dữ liệu vì nó cung cấp cho họ một cách dễ dàng để tạo hình ảnh trực quan dựa trên kết quả của một truy vấn. Với Looker, bạn có thể cung cấp cho các bên liên quan bức tranh toàn cảnh về công việc của mình bằng cách hiển thị cho họ dữ liệu trực quan và dữ liệu thực tế liên quan đến công việc đó.



Bài đọc 6: Công Cụ Hỗ Trợ Cho Nhà Phân Tích Dữ Liệu

1. Tổng quan về các công cụ

Trong quá trình thực hiện phân tích dữ liệu, có nhiều công cụ hỗ trợ như công cụ lưu trữ và thống kê dữ liệu theo dạng bảng như Microsoft Excel, Google Sheets. Khi dữ liệu lưu trữ trên cơ sở dữ liệu, chúng ta có các ngôn ngữ truy vấn dữ liệu và các phần mềm hỗ trợ như MySQL, Microsoft SQL Server. Các công cụ trực quan hóa dữ liệu như Tableau, Looker.

Công cụ bảng tính được xem là một công cụ lưu trữ dữ liệu đơn giản và dễ sử dụng nhất. Điều đó được minh chứng về tính phổ biến của chúng trong mọi hoạt động của các công ty, cơ sở sản xuất. Chúng có các đặc điểm như tổ chức dữ liệu dưới dạng cột và dòng, sắp xếp và lọc dữ liệu theo tiêu chí xác định, cũng như thực hiện các tính toán và thao tác trên dữ liệu thông qua các công thức và hàm đã được định nghĩa trước.

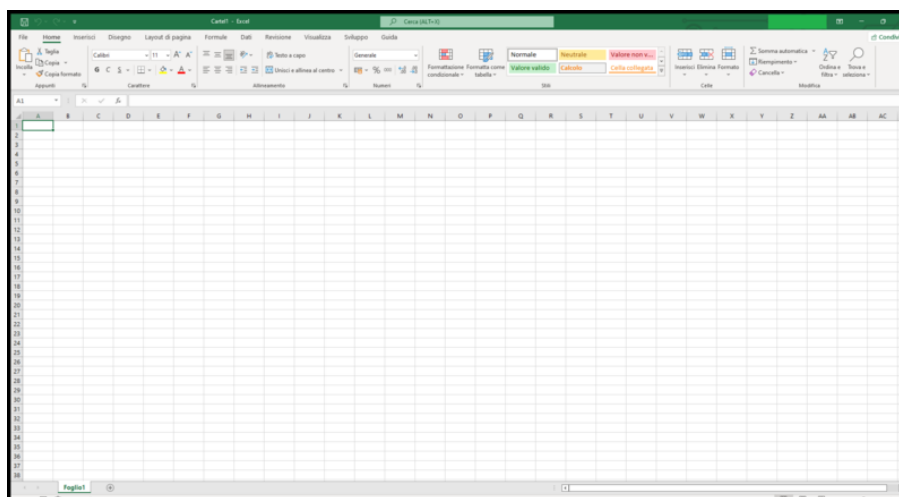
Khi dữ liệu nhiều lên và phục vụ nhiều bài toán khác nhau, chúng ta không thể quản lý chúng bằng công cụ bảng tính được. Lúc này chúng ta cần một cơ sở dữ liệu. Cơ sở dữ liệu là cách thức dữ liệu được tập hợp, lưu trữ một cách chặt chẽ, có tổ chức. Một chương trình máy tính được phát triển để hình thành hệ thống cơ sở dữ liệu và thực hiện lưu trữ lên đó. Để thao tác với dữ liệu như yêu cầu, cập nhật, truy xuất, chúng ta cần một ngôn ngữ truy vấn chẳng hạn như SQL. Mỗi lần truy vấn chúng ta viết ra một câu truy vấn.

2. Giới thiệu bảng tính

Đối với bảng tính, có rất nhiều giải pháp bảng tính khác nhau, nhưng hai lựa chọn phổ biến là Excel và Google Trang tính. Bảng tính là một bảng số liệu kỹ thuật số. Nó lưu trữ, tổ chức và sắp xếp dữ liệu. Khi đưa dữ liệu vào bảng tính, chúng ta có thể thấy các mẫu, thông tin nhóm và dễ dàng tìm thấy thông tin cần.

Cấu trúc cơ bản của một bảng tính bao gồm 3 thành phần chính là cột, dòng và ô. Có nhiều thao tác liên quan đến bảng tính như thay đổi độ rộng của cột, độ cao của dòng, gói văn bản, v.v. Khi thiết lập chức năng gói văn bản (text wrapping) thì bảng tính sẽ tự động thay đổi chiều cao của ô để mà cho phép tất cả nội dung nằm vừa bên trong ô đó, không bị tràn sang ô khác. Các dữ liệu được tổ chức dưới dạng bảng thường có các thuộc tính, đó là đặc trưng của từng trường dữ liệu. Khi thể hiện

trong bảng tính, chúng thường được gọi là tên cột, nhãn cột, hay dòng tiêu đề. Chúng ta thường định dạng cho dòng tiêu đề khác biệt với dữ liệu thực sự bằng cách in đậm, tô màu nền, v.v... 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.

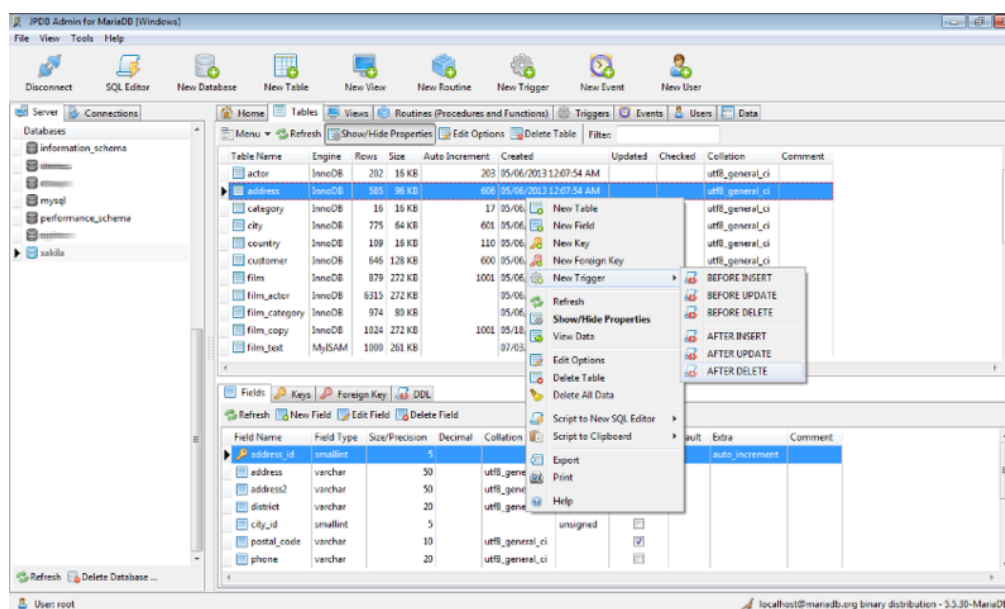


Chức năng sắp xếp dữ liệu theo một trật tự nào đó, ví dụ như theo ký tự chữ cái (A-Z) sẽ giúp nhà phân tích tổ chức và tìm dữ liệu một cách nhanh hơn. Trong bảng tính chúng ta chọn bảng dữ liệu cần sắp xếp, sau đó chọn cột và các tiêu chí sắp xếp khác.

Bảng tính cũng có một số tính năng thực sự hữu ích được gọi là công thức và hàm. Công thức là một tập hợp các hướng dẫn thực hiện một phép tính cụ thể bằng cách sử dụng dữ liệu trong bảng tính. Công thức có thể làm những việc cơ bản như cộng, trừ, nhân và chia, nhưng chúng không chỉ dừng ở đó. Chúng ta cũng có thể sử dụng công thức để tìm giá trị trung bình của một tập hợp số. Tra cứu một giá trị cụ thể, trả về tổng của một tập hợp các giá trị đáp ứng một quy tắc cụ thể, v.v. Hàm là một lệnh đặt trước tự động thực hiện một quy trình hoặc tác vụ cụ thể bằng cách sử dụng dữ liệu trong bảng tính.

3. Truy vấn dữ liệu

Khi dữ liệu lưu trữ trong một cơ sở dữ liệu, chúng ta thường dùng ngôn ngữ truy vấn để thao tác với dữ liệu. Ngôn ngữ truy vấn thông dụng là SQL (structured query language).



Chúng ta cùng điểm qua một số cú pháp cơ bản để truy vấn dữ liệu bằng ngôn ngữ SQL. Cú pháp đầu tiên là cú pháp SELECT-FROM-WHERE.

SELECT

Columns you want to look at

FROM

Table the data lives in

WHERE

Certain condition is met

Cú pháp này dùng để lấy dữ liệu từ một bảng trong cơ sở dữ liệu theo một điều kiện xác định. Phần SELECT là nơi mô tả các cột chứa dữ liệu chúng ta muốn lấy, trong trường hợp chúng ta muốn chọn tất cả các cột, chúng ta ghi dấu hoa thị (*). Phần thứ 2, FROM, để chỉ định bảng dữ liệu, và phần WHERE chứa các điều kiện mà dữ liệu cần thỏa.

Ví dụ một câu truy vấn:

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

Trong ví dụ, chúng ta muốn lấy dữ liệu chỉ ở cột `first_name` từ bảng `customer_name` của cơ sở dữ liệu `customer_data`. Chúng ta mong muốn chỉ lấy những dòng nào có `first_name` là 'Tony' nên tại `WHERE` chúng ta đặt ra điều kiện đó.

Tương tự như bảng tính, truy vấn bằng ngôn ngữ SQL còn nhiều tính năng nữa và đòi hỏi chúng ta tìm hiểu thêm để có thể sử dụng tốt cho quá trình phân tích. Các bài đọc khác sẽ đi sâu hơn vào cách thức sử dụng truy vấn dữ liệu.

Phần 2

HƯỚNG DẪN TRẢ LỜI CÂU HỎI

Phân tích dữ liệu

1. Mô tả nào sau đây đúng nhất về **dữ liệu**?

- A. Dữ liệu là một loạt các câu hỏi
- B. Dữ liệu là một tiến trình kinh doanh
- C. Dữ liệu là kết quả của tính toán và thống kê
- D. Dữ liệu là tập hợp của các dữ kiện

Đáp án: D

2. Một nhà bán lẻ quần áo thu thập và lưu trữ dữ liệu về doanh thu bán hàng. Điều nào sau đây sẽ là một phần của **hệ sinh thái dữ liệu của họ**? Chọn tất cả những gì mà bạn thấy phù hợp.

- A. Cơ sở dữ liệu về doanh thu bán hàng
- B. Đám mây lưu trữ cơ sở dữ liệu
- C. Bản ghi về hàng tồn kho
- D. Cơ sở dữ liệu của các đối thủ của họ

Đáp án: A, B, C

Một hệ sinh thái như định nghĩa đó là tổng hòa của nhiều thành phần từ bản thân dữ liệu liên quan trực tiếp đến hoạt động kinh doanh cho đến phần mềm, phần cứng phục vụ cho các thao tác trên dữ liệu này. Như vậy, chúng ta có thể thấy A,B,C là những thành phần có liên quan đến hệ sinh thái của doanh nghiệp bán lẻ này. Đối với cơ sở dữ liệu của đối thủ thường thực tế sẽ không thể có và cũng không phục vụ quá nhiều vào việc phân tích hiện trạng bài toán mà doanh nghiệp thường đối diện.

3. Điền vào chỗ trống: Trong phân tích dữ liệu, **hệ sinh thái dữ liệu** đề cập đến các thành phần tương tác với nhau để sản sinh, quản lý, lưu trữ, _____, phân tích và chia sẻ dữ liệu.

- A. Xác định (locate)
- B. Trộn (merge)
- C. Tiêu hóa (ingest)

D. Tổ chức (organize)

Đáp án: D

4. Tiến trình hướng **chiến lược kinh doanh** vào sử dụng dữ kiện là gì?

- A. Tiến trình ra quyết định dựa trên dữ liệu
- B. Tiến trình cải thiện chiến lược
- C. Tiến trình xác định dữ liệu và quyết định
- D. Tiến trình lên kế hoạch phân tích

Đáp án: A

5. Thuật ngữ nào sau đây đề cập đến việc **thu thập, biến đổi và tổ chức** dữ liệu để **đưa ra kết luận, dự đoán** và thúc đẩy việc **ra quyết định** sáng suốt?

- A. Vòng đời dữ liệu
- B. Phân tích dữ liệu
- C. Thành phần dữ liệu
- D. Thấu hiểu dữ liệu

Đáp án: B

6. Một hãng hàng không thu thập, quan sát và phân tích các hành vi trực tuyến của khách hàng. Sau đó, họ sử dụng những hiểu biết sâu sắc có được để chọn ra những sản phẩm và dịch vụ mới để đề xuất. Tình huống này mô tả loại tiến trình kinh doanh nào?

- A. Hợp tác với các bên liên quan
- B. Đo lường năng lực
- C. Làm quyết định dựa trên dữ liệu
- D. Tư duy phân tích

Đáp án: C

7. Mô tả nào sau đây đề cập đến **phân tích dữ liệu**?

- A. Tạo ra những mô hình và cách hiểu mới về những gì chưa biết bằng cách sử dụng dữ liệu thô
- B. Sử dụng các dữ kiện để hướng dẫn chiến lược kinh doanh
- C. Các yếu tố tương tác với nhau để cung cấp, quản lý, lưu trữ, tổ chức, phân tích và chia sẻ dữ liệu
- D. Việc thu thập, chuyển đổi và tổ chức dữ liệu để đưa ra kết luận, dự đoán và thúc đẩy việc ra quyết định sáng suốt

Đáp án: D

8. Việc thu thập, biến đổi và tổ chức dữ liệu để đưa ra kết luận, dự đoán và thúc đẩy việc ra quyết định sáng suốt mô tả về điều gì?

- A. Khoa học dữ liệu (data science)
- B. Vòng đời dữ liệu (data life cycle)
- C. Phân tích dữ liệu (data analysis)
- D. Lĩnh vực phân tích dữ liệu (data analytics)

Đáp án: C

- Data Analytics và Data Analysis dịch ra đều có nghĩa là phân tích dữ liệu. Vì thế nhiều người lầm tưởng rằng nghĩa của 2 từ này hoàn toàn giống nhau.
- Về cơ bản, sự khác biệt giữa 2 từ này chính là quy mô, phạm vi. Data Analytics có nghĩa rộng hơn data analysis và data analysis là tập hợp con trong Data Analytics.
- Data analysis **mô tả quá trình** kiểm tra, chuyển đổi và sắp xếp dữ liệu theo trật tự nhất định để tìm hiểu và rút ra các thông tin hữu ích.
- Còn Data Analytics **mô tả một lĩnh vực** bao gồm quá trình quản lý toàn bộ dữ liệu. Quá trình này không chỉ bao gồm việc phân tích (data analysis) mà còn các công cụ và kỹ thuật liên quan.

9. Trong phân tích dữ liệu, **một mô hình** là một nhóm các thành phần tương tác với nhau.

- A. Đúng
- B. Sai

Đáp án: B

(Một hệ sinh thái dữ liệu (Data ecosystem) mới là một nhóm các thành phần tương tác

với nhau.)

10. Điền vào chỗ trống: Mục tiêu chính của _____ là tìm câu trả lời cho các câu hỏi hiện có bằng cách tạo ra sự thấu hiểu các nguồn dữ liệu.

- A. Nhà phân tích dữ liệu (analyst)
- B. Nhà khoa học dữ liệu (scientist)
- C. Kỹ sư dữ liệu (engineer)
- D. Nhà thiết kế dữ liệu (designer)

Đáp án: A

11. Điền vào chỗ trống: Mục tiêu chính của _____ là tạo ra các câu hỏi mới bằng cách sử dụng dữ liệu.

- A. Nhà khoa học dữ liệu (data scientist)
- B. Nhà phân tích dữ liệu (data analyst)
- C. Kỹ sư dữ liệu (data engineer)
- D. Nhà thiết kế dữ liệu (data designer)

Đáp án: A

12. Điền vào chỗ trống: _____ liên quan đến việc tạo ra những cách mới để mô hình hóa và hiểu biết những điều chưa biết bằng cách sử dụng dữ liệu thô.

- A. Khoa học dữ liệu (data science)
- B. Phân tích dữ liệu (data analysis)
- C. Kỹ nghệ dữ liệu (data engineering)
- D. Thiết kế dữ liệu (data design)

Đáp án: A

13. Phân tích dữ liệu là các thành phần khác nhau tương tác với nhau để sản sinh, quản lý, lưu trữ, tổ chức, phân tích và chia sẻ dữ liệu.

- A. Đúng

B. Sai

Đáp án: B

14. Điền vào chỗ trống: Thuật ngữ _____ được định nghĩa là sự hiểu biết trực quan về điều gì đó mà ít hoặc không có lời giải thích.

- A. Quan điểm cá nhân (personal opinion)
- B. Tư duy hợp lý (rational thought)
- C. Nhận thức (awareness)
- D. Trực giác, linh tính (gut instinct)

Đáp án: D

15. Một công ty phát biểu một bài toán mà họ muốn giải quyết. Sau đó, một nhà phân tích dữ liệu thu thập dữ liệu có liên quan, phân tích và sử dụng nó để đưa ra kết luận. Nhà phân tích chia sẻ kết quả của họ với các chuyên gia lĩnh vực, người mà sẽ kiểm tra các kết quả này. Cuối cùng, một kế hoạch được triển khai. Kịch bản này mô tả điều gì?

- A. Ra quyết định dựa trên dữ liệu
- B. Khoa học dữ liệu
- C. Dịch vụ khách hàng
- D. Xác định xu thế

Đáp án: A

16. Để ra quyết định dựa trên dữ liệu một cách trọn vẹn, điều quan trọng là phải xem xét thêm những hiểu biết từ những người rất quen thuộc với vấn đề kinh doanh. Hãy cho biết những người này được gọi là gì?

- A. Các khách hàng (customers)
- B. Các bên liên quan (stakeholders)
- C. Chuyên gia lĩnh vực (subject-matter experts)
- D. Các đối thủ cạnh tranh (competitors)

Đáp án: C

17. Bạn vừa hoàn thành việc phân tích dữ liệu cho một dự án tiếp thị sản phẩm. Trước khi tiếp tục, bạn chia sẻ kết quả của mình với các thành viên của nhóm tiếp thị để xem liệu họ có thể bổ sung thêm hiểu biết về vấn đề kinh doanh hay không. Điều này đề cập đến một hoạt động của?

- A. Khoa học dữ liệu
- B. Ra quyết định dựa trên dữ liệu
- C. Quản lý dữ liệu
- D. Phân tích dữ liệu

Đáp án: B

18. Điền vào chỗ trống: Hệ sinh thái dữ liệu được tạo thành từ các thành phần _____. Điều này giúp chúng có thể sản sinh, quản lý, lưu trữ, tổ chức, phân tích và chia sẻ dữ liệu.

- A. Tương tác với nhau (interact with each other)
- B. Giải quyết vấn đề cùng nhau (problem-solve)
- C. Độc lập (unite)

Tăng cường sự thấu hiểu (gain insights).

Đáp án: A

19. Chọn mô tả tốt nhất về trực giác (gut instinct).

- A. Một cách hiểu trực quan về một thứ gì đó với ít hoặc không có sự giải thích
- B. Chọn lựa các dữ kiện mà bổ sung vào kinh nghiệm cá nhân
- C. Sử dụng khả năng bẩm sinh để phân tích dữ liệu
- D. Thay đổi dữ liệu để chúng khớp với trực giác

Đáp án: A

20. Một nhà sản xuất đồ nội thất muốn tìm một giải pháp thân thiện hơn với môi trường để sản xuất các sản phẩm của mình. Nhà phân tích dữ liệu được thuê để giải quyết vấn đề này đã thu thập dữ liệu có liên quan, phân tích và sử dụng nó để đưa ra kết luận. Sau đó, nhà phân tích chia sẻ phân tích của họ với các chuyên gia từ nhóm sản xuất, người sẽ kiểm tra đánh giá những kết quả của họ. Cuối cùng, một kế hoạch được triển khai.

Tình huống này mô tả về khoa học dữ liệu?

- A. Đúng
- B. Sai

Đáp án: B

21. Điền vào chỗ trống: Những người rất quen thuộc với lĩnh vực kinh doanh được gọi là _____. Đây là một phần quan trọng của quá trình ra quyết định dựa trên dữ liệu.

- A. Các bên liên quan (stakeholders)
- B. Các đối thủ cạnh tranh (competitors)
- C. Các khách hàng (customers)
- D. Chuyên gia lĩnh vực (subject-matter experts)

Đáp án: D

22. Chia sẻ kết quả phân tích của bạn với đồng nghiệp, những người quen thuộc với bài toán kinh doanh. Hoạt động này là một phần của cái gì?

- A. Ra quyết định dựa trên dữ liệu
- B. Khoa học dữ liệu
- C. Phân tích dữ liệu
- D. Quản lý dữ liệu

Đáp án: A

23. Trong phân tích dữ liệu, thuật ngữ nào mô tả một tập hợp các thành phần tương tác với nhau?

- A. Một hệ thống mô hình hóa (a modelling system)
- B. Một cơ sở dữ liệu (a database)
- C. Một môi trường đám mây (the cloud environment)
- D. Một hệ sinh thái dữ liệu (a data ecosystem)

Đáp án: D.

24. Billings Upholstery phát biểu một vấn đề cần giải quyết: Tìm cách sản xuất đồ nội thất thân thiện với môi trường hơn. Một nhà phân tích dữ liệu được giao nhiệm vụ thu thập dữ liệu liên quan, phân tích nó và sử dụng nó để đưa ra kết luận. Sau đó, nhà phân tích chia sẻ phân tích của họ với các chuyên gia từ nhóm sản xuất, người sẽ đánh giá các kết quả. Cuối cùng, một kế hoạch được triển khai. Kịch bản này mô tả quá trình nào?

- A. Ra quyết định dựa trên dữ liệu
- B. Dịch vụ khách hàng
- C. Khoa học dữ liệu
- D. Xác định xu thế

Đáp án: A

25. Nhà phân tích dữ liệu có thể sử dụng chiến thuật nào để kết hợp hiệu quả giữa trực giác với dữ kiện? Chọn tất cả những đáp án phù hợp.

- A. Áp dụng những kinh nghiệm quý giá trong quá khứ của họ vào công việc hiện tại, đồng thời luôn quan tâm đến câu chuyện mà dữ liệu đang muốn kể.
- B. Tham khảo cách xác định thành công cho một dự án, nhưng vẫn chủ yếu dựa vào quan điểm cá nhân.
- C. Sử dụng kiến thức của họ về cách công ty hoạt động để hiểu hơn về nhu cầu kinh doanh
- D. Hoàn toàn dựa vào trực giác để chọn dữ liệu cần thu thập và cách phân tích dữ liệu đó.

Đáp án: A,C

26. Một nhà phân tích dữ liệu kết thúc việc phân tích cho một dự án tiếp thị. Kết quả khá rõ ràng, vì vậy họ trình bày những phát hiện và khuyến nghị cho khách hàng. Họ nên làm gì trước buổi thuyết trình đó?

- A. Đóng gói dữ liệu để giữ chúng an toàn
- B. Khảo sát khách hàng về kết quả, kết luận và khuyến nghị
- C. Tạo một mô hình dựa trên kết quả phân tích
- D. Chia sẻ kết quả với các chuyên gia trong lĩnh vực từ nhóm tiếp thị để họ cung

cấp thêm thông tin

Đáp án: D

27. Các chuyên gia lĩnh vực làm gì để hỗ trợ việc ra quyết định dựa trên dữ liệu? Chọn tất cả các câu phù hợp.

- A. Xem xét kết quả phân tích dữ liệu và xác định bất kỳ điểm mâu thuẫn nào
- B. Đánh giá các lựa chọn được đưa ra từ quá trình thấu hiểu dữ liệu
- C. Thu thập, biến đổi và sắp xếp dữ liệu
- D. Đề xuất những sự thấu hiểu cho bài toán kinh doanh

Đáp án: A,B,D

Diễn đàn thảo luận

1. Ví dụ nào sau đây là cách sử dụng thích hợp của diễn đàn thảo luận?

- A. Đăng câu trả lời của các bài kiểm tra trong khóa học
- B. Đặt câu hỏi về nội dung của một bài học trong chứng chỉ này
- C. Chia sẻ ảnh thú cưng
- D. Nhờ người học khác gửi lý lịch của mình đến nhà tuyển dụng

Đáp án: B

2. Để tạo ra các cuộc thảo luận rõ ràng và hấp dẫn trong diễn đàn, bạn nên sử dụng kiểu viết nào? Chọn tất cả những câu phù hợp.

- A. Viết bằng ngôn ngữ tin nhắn trao đổi, chẳng hạn như BTW cho “By the way”
- B. Sử dụng dấu câu thích hợp, chẳng hạn như dấu phẩy và dấu chấm ở vị trí thích hợp
- C. Gõ tắt cả ở dạng chữ cái viết thường
- D. Viết câu hoàn chỉnh

Đáp án: B,D

3. Khi đăng bài trong diễn đàn thảo luận, cư xử nào được chấp nhận?

- A. Tể nhị, tốt bụng và cởi mở
- B. Thể hiện ý kiến bằng cách tranh luận chống lại quan điểm của người khác
- C. Sử dụng ngôn ngữ thô tục để nhấn mạnh quan điểm của bạn
- D. Chia sẻ quảng cáo và khuyến mãi sản phẩm

Đáp án: A

4. Gần đây bạn đã đăng ký một tạp chí điện tử thuộc lĩnh vực phân tích dữ liệu. Bạn thực sự thích một bài báo và muốn chia sẻ nó trong diễn đàn thảo luận. Điều nào sau đây sẽ phù hợp cho bài đăng của bạn? Chọn tất cả những câu phù hợp.

- A. Trình bày những suy nghĩ của bạn về bài báo
- B. Ghi công cho tác giả gốc
- C. Kiểm tra lỗi chính tả hoặc ngữ pháp trong bài đăng
- D. Đi kèm một quảng cáo về cách đăng ký tạp chí phân tích dữ liệu

Đáp án: A,B,C

5. Bạn đọc một bài báo thú vị về phân tích dữ liệu trên một tạp chí và muốn chia sẻ một số ý tưởng từ bài báo trong diễn đàn thảo luận. Trong bài đăng của bạn, bạn đã trích dẫn tác giả và liên kết đến bài viết gốc. Đây là cách sử dụng diễn đàn không phù hợp?

- A. Đúng
- B. Sai

Đáp án: B

6. Bạn đọc một bài báo thú vị trên tạp chí và muốn chia sẻ nó trong diễn đàn thảo luận. Bạn nên làm gì khi đăng bài? Chọn tất cả những câu phù hợp.

- A. Đảm bảo rằng bài báo có liên quan đến phân tích dữ liệu
- B. Đính kèm địa chỉ email của bạn để mọi người gửi câu hỏi hoặc nhận xét
- C. Kiểm tra lỗi chính tả hoặc ngữ pháp trong bài đăng của bạn
- D. Ghi công cho việc tạo ra bài báo

Đáp án: A,C

Kỹ năng phân tích

1. Điền vào chỗ trống: Sự tò mò, hiểu ngữ cảnh và tư duy kỹ thuật là tất cả các ví dụ về _____ được sử dụng trong quá trình ra quyết định dựa trên dữ liệu.

- A. Các kỹ năng phân tích
- B. Chiến lược kinh doanh
- C. Tiến trình suy nghĩ
- D. Mô hình hóa dữ liệu

Đáp án: A

2. Trong phân tích dữ liệu, kỹ năng phân tích nào liên quan đến việc quản lý con người, quy trình và công cụ?

- A. Điều khiển dữ liệu
- B. Chiến lược dữ liệu
- C. Thiết kế dữ liệu
- D. Phân tích dữ liệu

Đáp án: B

3. Tò mò là một kỹ năng phân tích liên quan đến điều nào sau đây?

- A. Làm việc với các dữ kiện một cách có trật tự
- B. Sử dụng trực giác của mình
- C. Tìm kiếm thách thức và trải nghiệm mới
- D. Hợp tác để giải quyết bài toán

Đáp án: C

4. Thêm tiêu đề mô tả vào các cột dữ liệu trong bảng tính là một ví dụ về kỹ năng phân tích nào?

- A. Hiểu ngữ cảnh
- B. Chiến lược dữ liệu

- C. Tò mò
- D. Tư duy bức tranh lớn

Đáp án: A

5. Tư duy kỹ thuật là một kỹ năng phân tích liên quan đến điều gì?

- A. Quản lý con người, tiến trình và công cụ
- B. Hiểu điều kiện mà một số thứ tồn tại hay xảy ra
- C. Chia các thứ thành những bước hay những mảnh nhỏ hơn
- D. Cân bằng giữa vai trò và trách nhiệm

Đáp án: C

6. Điền vào chỗ trống: Chiến lược dữ liệu liên quan đến ____ con người, quy trình và công cụ được sử dụng trong phân tích dữ liệu.

- A. Chọn lựa (choosing)
- B. Quản lý (managing)
- C. Giám sát (supervising)
- D. Hình dung (visualizing)

Đáp án: B

7. Thiết kế dữ liệu là cách bạn tổ chức thông tin; chiến lược dữ liệu là việc quản lý con người, quy trình và công cụ được sử dụng trong phân tích dữ liệu. Phát biểu trên đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

8. Tìm kiếm những thách thức và kinh nghiệm mới để học hỏi là một ví dụ của kỹ năng phân tích nào?

- A. Có một tư duy kỹ thuật

- B. Hiểu ngữ cảnh
- C. Tò mò
- D. Chiến lược dữ liệu

Đáp án: C

9. Hiểu ngữ cảnh là kỹ năng phân tích được mô tả tốt nhất bằng câu nào sau đây? Chọn tất cả những gì phù hợp.

- A. Thu thập thông tin bổ sung về dữ liệu để hiểu được bức tranh toàn cảnh
- B. Thêm tiêu đề mô tả vào các cột dữ liệu trong bảng tính
- C. Xác định động cơ đằng sau việc thu thập tập dữ liệu
- D. Làm việc với các dữ kiện một cách có trật tự

Đáp án: A,B,C

10. Điền vào chỗ trống: _____ liên quan đến khả năng chia nhỏ mọi thứ thành các bước hoặc phần nhỏ hơn và làm việc với chúng một cách có trật tự và hợp lý.

- A. Tò mò
- B. Tư duy kỹ thuật
- C. Chiến lược dữ liệu
- D. Ngữ cảnh

Đáp án: B

11. Kỹ năng phân tích nào liên quan đến việc quản lý con người, quy trình và công cụ được sử dụng trong phân tích dữ liệu?

- A. Chiến lược dữ liệu
- B. Thiết kế dữ liệu
- C. Hiểu ngữ cảnh
- D. Tò mò

Đáp án: A

12. Một nhà phân tích dữ liệu làm việc cho một nhà sản xuất thiết bị. Năm ngoái, lợi nhuận của công ty giảm. Lợi nhuận thấp hơn có thể là kết quả của việc ít người mua thiết bị hơn, chi phí sản xuất thiết bị cao hơn hoặc cả hai. Nhà phân tích nhận ra rằng đó là những vấn đề lớn cần giải quyết, vì vậy họ chia nhỏ vấn đề thành nhiều phần nhỏ hơn để phân tích một cách có trật tự. Kỹ năng phân tích nào họ đang sử dụng?

- A. Tư duy kỹ thuật
- B. Tò mò
- C. Hiểu ngữ cảnh
- D. Chiến lược dữ liệu

Đáp án: A

13. Một nhà phân tích dữ liệu đang tìm kiếm những trải nghiệm mới để có được kiến thức. Họ xem video và đọc các bài báo về phân tích dữ liệu. Họ đặt câu hỏi cho các chuyên gia. Kỹ năng phân tích nào họ đang sử dụng?

- A. Tư duy kỹ thuật
- B. Chiến lược dữ liệu
- C. Hiểu ngữ cảnh
- D. Tò mò

Đáp án: D

14. Xác định động cơ đằng sau việc thu thập dữ liệu và thu thập thông tin bổ sung là ví dụ về kỹ năng phân tích nào?

- A. Thiết kế dữ liệu
- B. Chiến lược dữ liệu
- C. Tư duy kỹ thuật
- D. Hiểu ngữ cảnh

Đáp án: D

15. Các nhà phân tích dữ liệu sử dụng 5 kỹ năng phân tích là tò mò, hiểu ngữ cảnh, có tư

duy kỹ thuật, thiết kế dữ liệu và chiến lược dữ liệu để đưa ra quyết định _____

- A. Trực giác (intuitive)
- B. Hiệu quả hơn (more efficient)
- C. Tầm nhìn xa (forward-looking)
- D. Hướng dữ liệu (data-driven)

Đáp án: D

16. Một nhà phân tích dữ liệu thêm các tiêu đề mô tả vào các cột dữ liệu trong bảng tính. Điều này cải thiện bảng tính như thế nào?

- A. Nó thêm ngữ cảnh
- B. Nó làm rõ chiến lược kinh doanh
- C. Nó cải thiện tính thẩm mỹ
- D. Nó bỏ đi những chi tiết không cần thiết

Đáp án: A

Tư duy phân tích

1. Xác định mô hình từ trái sang phải trong tập hợp các khối bên dưới và cố gắng dự đoán khối nào sẽ thay thế khối bằng dấu chấm hỏi.

Đáp án: D

(Số chấm đen tăng dần trong mỗi ô)

2. Xác định mô hình từ trái sang phải trong các hình ảnh dưới đây và cố gắng dự đoán hình ảnh nào sẽ xuất hiện tiếp theo.

Đáp án: A

(Hai mẫu có trong chuỗi: số cạnh và số chấm. Di chuyển từ trái sang phải, cả hai đều giảm một. Với những mẫu này, nếu khối trước đó chứa một hình có bốn cạnh và ba chấm, thì hình tiếp theo phải có ba cạnh và hai chấm.)

3. Chọn số tiếp theo trong chuỗi:

9, 13, 17, 21, 25, 29, _____

- A. 33
- B. 10
- C. 55
- D. 25

Đáp án: A

(Các số đều tăng và hiệu của mỗi số là 4 nên số tiếp theo trong dãy là 33)

4. Các số sau đây theo thứ tự từ trái sang phải. Xác định quy luật và quyết định số nào sẽ xuất hiện tiếp theo.



4, 9, 16, 25, 36, 49, _____

- A. 64
- B. 30
- C. 81
- D. 62

Đáp án: A

(Có hai quy luật trong dãy số. Một là mỗi số được bình phương và sau đó số bình phương được tăng lên một (ví dụ: 2^2 , 3^2 , 4^2 , 5^2 , 6^2 , 7^2). Quy luật thứ hai là sự khác biệt giữa các số trong dãy: $9 - 4 = 5$, $16 - 9 = 7$, $25 - 16 = 9$, v.v.)

5. Câu hỏi sau đây là về việc nhận biết và kết hợp các quy luật của các hình giống nhau, nhưng được nhìn từ các góc độ khác nhau.

Hai hình tương tự nhau khi một hình có thể trở thành hình kia sau khi quay theo chiều kim đồng hồ  hoặc ngược chiều kim đồng hồ , hoặc hình phản chiếu theo chiều ngang \leftrightarrow và/hoặc chiều dọc \updownarrow .

Nhiệm vụ của bạn là chọn hình hoàn thành chỗ trống. Hãy chú ý đến quy luật mà hình thứ nhất và thứ hai có liên quan với nhau, sau đó tìm ra lựa chọn nào phù hợp với hình 3.

Đáp án: 5tt

5tt.

Đáp án: B

(Hình đầu tiên trong mẫu là phản chiếu của hình thứ hai. Câu trả lời sẽ là một hình ảnh phản chiếu của hình ảnh thứ ba.)

6. Câu hỏi sau đây là về việc nhận biết và kết hợp các quy luật của những hình giống nhau, nhưng được nhìn từ các góc độ khác nhau. Hai hình tương tự nhau khi một hình có thể trở thành hình kia sau khi xoay theo chiều kim đồng hồ \odot hoặc ngược chiều kim đồng hồ \ominus , hoặc hình phản chiếu theo chiều ngang \leftrightarrow và/hoặc chiều dọc \updownarrow .

Nhiệm vụ của bạn là chọn hình hoàn thành chuỗi sau:

Đáp án: 6tt

6tt.

Đáp án: C

(Vì mô hình trong hình đầu tiên được xoay 90 độ ngược chiều kim đồng hồ, hình C hoàn thành chuỗi trên)

7. Dãy mã sau đây theo thứ tự từ trái sang phải. Có một quy luật lặp lại mà bạn sẽ cần tìm ra. Xác định quy luật và quyết định mã nào sẽ xuất hiện tiếp theo.

A1, B3, C5, D7, E9, F11, G13, _____

- A. H15
- B. H16
- C. D17
- D. J15

Đáp án: A

(Quy luật là các chữ cái được liệt kê theo thứ tự bảng chữ cái và các số tăng lên hai với mỗi bộ mới. Do đó, mã tiếp theo sẽ là H15.)

8. Dãy mã sau đây theo thứ tự từ trái sang phải. Có một quy luật lặp lại mà bạn sẽ cần tìm ra. Xác định quy luật và quyết định chuỗi nào sẽ xuất hiện tiếp theo.

A, AA, AAA, B, BA, BAA, BAAA, BB, BBA, BBAA, BBAAA, BBB, _____

- A. BBBB

- B. BBAA
- C. BBBA
- D. BBAAA

Đáp án: C

(Các quy luật trong trình tự này theo sau chữ A. A được thêm vào cho đến khi có ba A, đó là khi chữ B thay thế cho A trước đó và quy luật tiếp tục. Do đó, BBBA là phần tiếp theo trong chuỗi.)

9. Bây giờ, hãy xác định các quy luật trong một bài toán đố bằng cách sử dụng trực quan hóa dữ liệu. Có 12 viên socola trong một hộp: tám viên có nhân caramel, sáu viên có nhân dừa và hai viên có cả caramel và nhân dừa. Chọn hình ảnh phù hợp nhất mô tả hộp sôcôla này.

Đáp án: B

(Sơ đồ này mô tả sáu sôcôla chỉ có nhân caramel, bốn sôcôla chỉ có nhân dừa, hai sôcôla có cả caramen và dừa, và tổng số sôcôla là 12.)

10. Có 10 trẻ trong một lớp và tất cả đều mang theo bánh mì để ăn trưa: 5 trẻ ăn bánh mì với bơ đậu phộng, 6 trẻ ăn bánh mì với thạch, và 3 trẻ có bánh mì với cả bơ đậu phộng và thạch. Tìm xem có bao nhiêu trẻ em có bánh mì kẹp không có bơ đậu phộng và thạch và chọn hình mô tả tình huống tốt nhất.

Đáp án: A

(Trong sơ đồ này, có sáu bánh mì kẹp với thạch, năm bánh mì kẹp bơ đậu phộng và ba bánh mì kẹp cả hai. Điều này có nghĩa là có $(5 + 6 - 3 = 8)$ tám chiếc bánh mì sandwich với bơ đậu phộng hoặc thạch. Có tổng cộng 10 trẻ em. Xét: $10 - 8 = 2$. Điều này có nghĩa là hai đứa trẻ không có bơ đậu phộng và thạch trong bánh mì của chúng.)

11. Quá trình nào liên quan đến việc xác định, định nghĩa và giải quyết một vấn đề bằng cách sử dụng dữ liệu một cách có tổ chức và từng bước?

- A. Tư duy phân tích
- B. Ngữ cảnh
- C. Thiết kế dữ liệu

D. Trực quan hóa

Đáp án: A

12. Cái nào sau đây là ví dụ về trực quan hóa dữ liệu? Chọn tất cả các câu phù hợp.

A. Đồ thị

B. Bản đồ

C. Biểu đồ

D. Báo cáo

Đáp án: A,B,C

13. Phân tích khoảng trống được sử dụng để kiểm tra và đánh giá cách một quy trình hiện đang hoạt động với mục tiêu đạt được vị trí mong muốn trong tương lai. Phát biểu trên có đúng không?

A. Đúng

B. Sai

Đáp án: A

14. Khía cạnh nào của tư duy phân tích liên quan đến việc xác định mối quan hệ giữa hai hoặc nhiều dữ liệu?

A. Độ tương quan (correlation)

B. Thiết kế dữ liệu (data design)

C. Ngữ cảnh (context)

D. Trực quan hóa (visualization)

Đáp án: A

15. Phương pháp nào liên quan đến việc đặt ra nhiều câu hỏi để tìm ra nguyên nhân gốc rễ của vấn đề?

A. Điều tra (inquiry)

B. Lập chiến lược (strategizing)

C. Tò mò (curiosity)

D. 5 “tại sao” (5 whys)

Đáp án: D

16. Một sân bay muốn làm cho quá trình xử lý hành lý của mình nhanh hơn và đơn giản hơn cho du khách. Một nhà phân tích dữ liệu kiểm tra và đánh giá cách thức hoạt động của quy trình hiện tại để đạt được mục tiêu là quy trình hiệu quả hơn. Họ sử dụng phương pháp nào?

- A. Phân tích khoảng trống
- B. Chiến lược
- C. Trực quan hóa dữ liệu
- D. 5 “tại sao”

Đáp án: A

17. Độ tương quan là khía cạnh của tư duy phân tích liên quan đến việc tìm ra các chi tiết cụ thể giúp bạn thực hiện một kế hoạch. Phát biểu trên có đúng không?

- A. Đúng
- B. Sai

Đáp án: B

18. Điền vào chỗ trống: Xác định mối quan hệ giữa hai hoặc nhiều phần dữ liệu mô tả ____.

- A. Độ tương quan (correlation)
- B. Tư duy hướng chi tiết (detail-oriented thinking)
- C. Trực quan hóa (visualization)
- D. Hướng bài toán (problem-orientation)

Đáp án: A

19. Phương pháp nào liên quan đến việc kiểm tra và đánh giá cách một quy trình hiện tại hoạt động để đưa nó đến nơi bạn muốn trong tương lai?

- A. Trực quan hóa dữ liệu (data visualization)
- B. Chiến lược (strategy)

C. Phân tích khoảng trống (gap analysis)

5 “tại sao” (5 whys)

Đáp án: C

20. Người quản lý của một cửa hàng âm nhạc nhận thấy rằng có nhiều kèn được sửa chữa hơn vào những ngày mà Alex và Jasmine làm việc cùng một ca. Sau khi điều tra, người quản lý phát hiện ra rằng Alex rất giỏi trong việc sửa thanh trượt, còn Jasmine thì rất giỏi trong việc tạo hình miệng kèn. Làm việc cùng nhau, Alex và Jasmine sửa chữa các vết lõm nhanh hơn. Người quản lý rất vui khi phát hiện ra mối quan hệ này và quyết định luôn xếp lịch cho Alex và Jasmine vào những ca giống nhau. Trong tình huống này, nhà quản lý đã sử dụng đặc điểm nào của tư duy phân tích?

- A. Độ tương quan (correlation)
- B. Hướng bài toán (problem-orientation)
- C. Trực quan hóa (visualization)
- D. Tư duy bức tranh lớn (big-picture thinking)

Đáp án: A

21. Các nhà phân tích dữ liệu hỏi, “Tại sao?” năm lần để tìm ra nguyên nhân gốc rễ của vấn đề. Nhận định này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A.

22. Phân tích khoảng trống là một phương pháp để kiểm tra và đánh giá cách một quy trình hoạt động hiện tại để đạt được vị trí bạn muốn trong tương lai. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

23. 5 “tại sao” là một kỹ thuật liên quan đến việc hỏi, “Tại sao?” năm lần để đạt được mục tiêu gì?

- A. Xác định nguyên nhân gốc của vấn đề
- B. Trực quan hóa một tiến trình sẽ như thế nào trong tương lai
- C. Biến kế hoạch thành hành động
- D. Sử dụng dữ kiện để hướng dẫn chiến lược kinh doanh

Đáp án: A

24. Phân tích khoảng trống là một quá trình có thể giúp hoàn thành nhiệm vụ nào sau đây? Chọn tất cả các câu phù hợp.

- A. Cải thiện khả năng truy cập cho một ứng dụng giáo dục dựa trên chức năng hiện tại của nó
- B. Tăng hiệu quả của nhà sản xuất ô tô dựa trên quy trình lắp ráp hiện tại của họ.
- C. Giảm lượng khí thải carbon của một công ty dựa trên lượng khí thải hiện tại của công ty
- D. Nâng cao nhận thức về bất bình đẳng thu nhập dựa trên mức lương địa phương

Đáp án: A,B,C

25. Chủ một cửa hàng giày trượt băng để ý rằng mỗi khi một nhân viên nào đó có ca làm việc thì cuối ngày sẽ có doanh số bán hàng cao hơn. Sau một số điều tra, người chủ nhận ra rằng kể từ khi nhân viên được thuê, cửa hàng kiếm được thêm 15% mỗi tháng. Trong tình huống này, nhà quản lý đã sử dụng tính chất nào của tư duy phân tích?

- A. Tính tương quan (correlation)
- B. Tính trực quan (visualization)
- C. Hướng bài toán (problem-orientation)
- D. Tư duy bức tranh lớn (big-picture thinking)

Đáp án: A

Ra quyết định dựa trên dữ liệu

1. Điền vào chỗ trống: Sự tò mò, hiểu ngữ cảnh, có tư duy kỹ thuật, thiết kế dữ liệu và chiến lược dữ liệu là _____. Chúng cho phép các nhà phân tích dữ liệu giải quyết các vấn đề bằng cách sử dụng các dữ kiện.

- A. Kỹ năng phân tích
- B. Hiểu biết cá nhân
- C. Kỹ năng kinh doanh
- D. Tiến trình suy nghĩ

Đáp án: A

2. Một công ty nhận được những bình luận tiêu cực trên mạng xã hội về sản phẩm của họ. Để giải quyết vấn đề này, một nhà phân tích dữ liệu sử dụng mỗi kỹ năng trong số 5 kỹ năng phân tích: tò mò, hiểu ngữ cảnh, có tư duy kỹ thuật, thiết kế dữ liệu và chiến lược dữ liệu. Điều này giúp nhà phân tích có thể sử dụng các dữ kiện để định hướng chiến lược kinh doanh và tìm ra cách cải thiện sự hài lòng của khách hàng. Đây là một ví dụ về cái gì?

- A. Trực quan hóa dữ liệu
- B. Ra quyết định dựa trên dữ liệu
- C. Phân tích khoảng trống
- D. Khoa học dữ liệu

Đáp án: B

3. Ra quyết định dựa trên dữ liệu bao gồm 5 kỹ năng phân tích: tò mò, hiểu bối cảnh, có tư duy kỹ thuật, thiết kế dữ liệu và chiến lược dữ liệu. Mỗi cái đóng một vai trò trong việc ra quyết định theo hướng dữ liệu. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

4. Khảo sát khách hàng về sở thích của họ và sử dụng thông tin đó để lên chiến lược kinh doanh là một ví dụ về ra quyết định dựa trên dữ liệu. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

Vòng đời dữ liệu

1. Điền vào chỗ trống: Trong pha ____ của vòng đời dữ liệu, một doanh nghiệp quyết định loại dữ liệu mình cần, cách thức quản lý, ai sẽ chịu trách nhiệm về nó và các kết quả mong đợi.

- A. Lập kế hoạch (planning)
- B. Đóng gói (archive)
- C. Quản lý (manage)
- D. Thu thập (capture)

Đáp án: A

2. Trong vòng đời dữ liệu, giai đoạn nào liên quan đến việc thu thập dữ liệu từ nhiều nguồn khác nhau và đưa chúng vào tổ chức?

- A. Quản lý (manage)
- B. Đóng gói (archive)
- C. Phân tích (analyze)
- D. Thu thập (capture)

Đáp án: D

3. Một nhà phân tích dữ liệu kết thúc việc sử dụng tập dữ liệu, vì vậy họ xóa hoặc cắt nhỏ các tệp để bảo vệ thông tin cá nhân. Đây được gọi là pha đóng gói?

- A. Đúng

B. Sai

Đáp án: B

(Đây là pha phá hủy)

4. Một nông dân chăn nuôi bò sữa quyết định mở một cửa hàng kem trong trang trại của mình. Sau khi khảo sát cộng đồng địa phương về hương vị yêu thích của mọi người, cô ta lấy dữ liệu họ cung cấp và lưu trữ trong ổ cứng bảo mật để có thể bảo quản an toàn trên máy tính của mình. Đây là một phần của pha nào của vòng đời dữ liệu?

- A. Đóng gói (archive)
- B. Phân tích (analyze)
- C. Lập kế hoạch (plan)
- D. Quản lý (manage)

Đáp án: D

5. Sau khi mở cửa hàng kem trong trang trại của mình, một nông dân sau đó khảo sát cộng đồng địa phương về hương vị yêu thích của họ. Cô ta sử dụng dữ liệu thu thập được để xác định rằng 5 hương vị hàng đầu là dâu tây, vani, sô cô la, bạc hà và bơ đậu phộng. Cô ta cảm thấy tự tin vào quyết định bán những hương liệu này của mình. Đây là một phần của pha nào trong vòng đời dữ liệu?

- A. Đóng gói (archive)
- B. Phân tích (analyze)
- C. Thu thập (capture)
- D. Lập kế hoạch (plan)

Đáp án: B

6. Pha quản lý của vòng đời dữ liệu mô tả về việc doanh nghiệp quyết định loại dữ liệu mình cần, cách dữ liệu sẽ được xử lý và ai sẽ chịu trách nhiệm về nó. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: B

(Đây là pha lập kế hoạch)

7. Pha phá hủy của vòng đời dữ liệu liên quan đến các hành động nào sau đây? Chọn tất cả những câu phù hợp.

- A. Tải dữ liệu lên đám mây
- B. Cắt nhỏ giấy tờ
- C. Sử dụng phần mềm xóa dữ liệu
- D. Lưu trữ dữ liệu cho sử dụng sau này

Đáp án: B,C

8. Trong pha phân tích của vòng đời dữ liệu, nhà phân tích dữ liệu làm những gì? Chọn tất cả những câu phù hợp.

- A. Sử dụng một công thức để tính toán
- B. Tạo một bản báo cáo từ dữ liệu
- C. Sử dụng phần mềm bảng tính để tổng hợp dữ liệu
- D. Chọn định dạng của tiêu đề trong bảng tính

Đáp án: A,B,C

9. Vòng đời dữ liệu đề cập đến các pha mà dữ liệu trải qua; phân tích dữ liệu bao gồm việc tuân theo một quy trình để phân tích dữ liệu. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

10. Điền vào chỗ trống: Một doanh nghiệp quyết định loại dữ liệu mà họ cần, cách dữ liệu sẽ được quản lý và ai sẽ chịu trách nhiệm về chúng trong pha ____ của vòng đời dữ liệu.

- A. Quản lý (manage)
- B. Lập kế hoạch (plan)

C. Thu thập (capture)

D. Phân tích (analyze)

Đáp án: B

11. Một nhà phân tích dữ liệu đang làm việc tại một công ty khởi nghiệp công nghệ nhỏ. Họ vừa hoàn thành một dự án phân tích, liên quan đến thông tin của công ty tư nhân về việc ra mắt sản phẩm mới. Để giữ thông tin an toàn, nhà phân tích sử dụng phần mềm xóa dữ liệu an toàn cho các tệp kỹ thuật số và một máy hủy tài liệu cho các giấy tờ. Điều này mô tả pha nào của vòng đời dữ liệu?

A. Phá hủy (destroy)

B. Lập kế hoạch (plan)

C. Đóng gói (archive)

D. Quản lý (manage)

Đáp án: A

12. Trong pha thu thập của vòng đời dữ liệu, nhà phân tích dữ liệu có thể sử dụng bảng tính để tổng hợp dữ liệu. Phát biểu này đúng hay sai?

A. Đúng

B. Sai

Đáp án: B

(Đây là pha phân tích)

13. Trong pha nào của vòng đời dữ liệu, một doanh nghiệp quyết định loại dữ liệu nào mà họ cần, dữ liệu sẽ được quản lý như thế nào và ai sẽ chịu trách nhiệm về nó?

A. Lập kế hoạch (plan)

B. Quản lý (manage)

C. Thu thập (capture)

D. Phân tích (analyze)

Đáp án: A

14. Một nhà phân tích dữ liệu đã hoàn thành một dự án phân tích liên quan đến dữ liệu của công ty tư nhân. Họ xóa các tệp kỹ thuật số để giữ thông tin an toàn. Điều này mô tả pha nào của vòng đời dữ liệu?

- A. Đóng gói (archive)
- B. Phá hủy (destroy)
- C. Quản lý (manage)
- D. Lập kế hoạch (plan)

Đáp án: B

15. Một nhà phân tích dữ liệu sử dụng một hàm trong bảng tính để tổng hợp dữ liệu. Sau đó, họ thêm một bảng tổng hợp để trình bày các tổng từ nhỏ nhất đến lớn nhất. Điều này sẽ xảy ra trong pha nào của vòng đời dữ liệu?

- A. Quản lý (manage)
- B. Thu thập (capture)
- C. Lập kế hoạch (plan)
- D. Phân tích (analyze)

Đáp án: D

16. Bạn đóng vai trò là nhà phân tích dữ liệu, hãy tư vấn cho một công ty kem cách thu thập dữ liệu về sở thích hương vị của khách hàng. Viết khoảng 2-3 câu cùng với giải thích của bạn.

Các ý cần xem xét:

- Dữ liệu gì họ cần, ở đâu để lấy
- Thu thập dữ liệu như thế nào, cái gì sẽ dùng nó, cái gì không
- Bảo mật dữ liệu như thế nào, xử lý dữ liệu cũ ra sao

Đáp án: Viết đoạn 2-3 câu

Tiến trình phân tích dữ liệu

1. Các pha của quá trình phân tích dữ liệu gồm hỏi, chuẩn bị, xử lý, phân tích, chia sẻ và hành động. Các nhà phân tích dữ liệu làm gì trong pha hỏi?

- A. Định nghĩa bài toán cần giải quyết
- B. Thu thập và lưu trữ dữ liệu
- C. Làm sạch dữ liệu
- D. Tạo ra bản trực quan hóa dữ liệu

Đáp án: A

2. Trong pha xử lý, nhà phân tích dữ liệu sẽ làm sạch dữ liệu để đảm bảo dữ liệu được đầy đủ và chính xác. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

3. Trong pha phân tích dữ liệu nào, một nhà phân tích sẽ sử dụng bảng tính hoặc ngôn ngữ truy vấn để biến đổi dữ liệu nhằm đưa ra các kết luận?

- A. Xử lý
- B. Hành động
- C. Phân tích
- D. Chuẩn bị

Đáp án: C

4. Trong pha phân tích dữ liệu nào, một nhà phân tích dữ liệu sẽ sử dụng hình ảnh như biểu đồ hoặc đồ thị để đơn giản hóa dữ liệu phức tạp nhằm hiểu rõ hơn?

- A. Xử lý
- B. Chia sẻ
- C. Hành động

D. Chuẩn bị

Đáp án: B

5. Một nhà phân tích dữ liệu chia sẻ những hiểu biết sâu sắc từ phân tích của họ trong một buổi thuyết trình chính thức cho các bên liên quan. Trong bản trình chiếu, họ đưa ra đề xuất dựa trên dữ liệu về cách giải quyết vấn đề kinh doanh. Pha nào của quá trình phân tích dữ liệu sẽ đến tiếp theo?

A. Chuẩn bị

B. Xử lý

C. Hỏi

D. Hành động

Đáp án: D

6. Nhóm phân tích dữ liệu có thể thực hiện những hành động nào trong giai đoạn hành động của quá trình phân tích dữ liệu? Chọn tất cả những câu phù hợp.

A. Kiểm tra, đánh giá những kết luận được cung cấp bởi nhà phân tích dữ liệu

B. Chia sẻ kết quả sử dụng công cụ trực quan hóa

C. Chốt một chiến lược dựa trên quá trình phân tích

D. Triển khai hành động để giải quyết bài toán

Đáp án: A,C,D

7. Mối quan hệ giữa vòng đời dữ liệu và quá trình phân tích dữ liệu là gì? Làm thế nào là hai quá trình giống nhau? Chúng khác nhau như thế nào?

Viết 2-3 câu trình bày ý kiến của bạn về các câu hỏi trên.

Đáp án:

Gợi ý nội dung: Tiến trình phân tích dữ liệu hướng dẫn cách thức để đạt được mục tiêu kinh doanh nhưng chúng ta cần hiểu vòng đời dữ liệu để mà sử dụng cho tiến trình này. Vì để phân tích dữ liệu tốt, chúng ta cần hiểu về nó, thu thập tất cả dữ liệu cần thiết. Ở chiều ngược lại, dữ liệu chỉ có ích khi chúng ta có kế hoạch để phân tích nó

8. Mối liên hệ giữa pha hỏi của tiến trình phân tích dữ liệu và pha lên kế hoạch của vòng đời dữ liệu là gì? Chúng giống và khác nhau những gì?

Viết 2-3 câu trình bày ý kiến của bạn về các câu hỏi trên.

Đáp án:

Gợi ý: Cả hai đều bao gồm lên kế hoạch và hỏi những câu hỏi, nhưng chúng giải quyết các đối tượng khác nhau. Pha Ask trong tiến trình phân tích dữ liệu tập trung vào tư duy chiến thuật trên bức tranh tổng thể của các mục tiêu kinh doanh. Trong khi đó, pha Plan tập trung các nền tảng của dự án như dữ liệu gì cần, ở đâu để lấy nó.

9. Một công ty nhận kết quả do nhóm phân tích dữ liệu của mình cung cấp, kiểm tra và đánh giá lại chúng và hoàn thiện một chiến lược. Sau đó, họ thực hiện một kế hoạch để giải quyết vấn đề kinh doanh ban đầu. Phần này mô tả bước chia sẻ của quá trình phân tích dữ liệu, đúng hay sai?

A. Đúng

B. Sai

Đáp án: B

10. Điền vào chỗ trống: Vòng đời dữ liệu có 6 giai đoạn của dữ liệu, trong khi phân tích dữ liệu có 6 ____.

A. Loại dữ liệu (data types)

B. Câu hỏi khóa (key questions)

C. Công cụ phân tích dữ liệu (data analytics tools)

D. Bước xử lý (process steps)

Đáp án: D

11. Vòng đời dữ liệu khác với phân tích dữ liệu như thế nào?

A. Vòng đời dữ liệu liên quan đến việc đưa ra các quyết định dựa trên dữ liệu; phân tích dữ liệu là sử dụng các công cụ để biến đổi dữ liệu.

B. Vòng đời dữ liệu đề cập đến các giai đoạn mà dữ liệu trải qua trong suốt chu kỳ tồn tại của nó; phân tích dữ liệu là quá trình phân tích dữ liệu.

C. Vòng đời dữ liệu liên quan đến việc xác định dữ liệu tốt nhất để giải quyết một

vấn đề; phân tích dữ liệu là đặt câu hỏi hiệu quả.

- D. Vòng đời dữ liệu liên quan đến việc chuyển đổi và xác minh dữ liệu; phân tích dữ liệu bằng cách sử dụng thông tin chi tiết thu được từ dữ liệu.

Đáp án: B

12. Một công ty nhận kết quả do nhóm phân tích dữ liệu của mình cung cấp, kiểm tra và đánh giá chúng và hoàn thiện một chiến lược. Sau đó, họ thực hiện một kế hoạch để giải quyết vấn đề kinh doanh ban đầu. Điều này mô tả pha nào của quá trình phân tích dữ liệu?

- A. Hành động
- B. Xử lý
- C. Chia sẻ
- D. Phân tích

Đáp án: A

13. Một công ty quảng cáo đã sử dụng thông tin chi tiết từ nhóm phân tích của mình để tạo chiến lược cải thiện doanh số bán hàng. Giờ đây, họ thực hiện kế hoạch tăng doanh thu hàng năm. Công ty đang ở bước nào của quá trình phân tích dữ liệu?

- A. Tiến trình
- B. Hành động
- C. Chia sẻ
- D. Phân tích

Đáp án: B

Trực quan hóa dữ liệu

1. Điền vào chỗ trống: Trực quan hóa dữ liệu là sự thể hiện _____ của thông tin.

- A. Thuộc tính (attributed)
- B. Đồ họa (graphical)

- C. Ngữ cảnh (contextual)
- D. Dạng bảng (tabulated)

Đáp án: B

2. Trường hợp nào biểu đồ tròn được sử dụng hiệu quả cho trực quan?

- A. Khi thể hiện tuổi của nam so với nữ
- B. Khi thể hiện sự thay đổi về tuổi của một người nào đó theo thời gian
- C. Khi thể hiện mối quan hệ giữa tuổi và thu nhập
- D. Khi thể hiện một lớp học được chia nhỏ theo độ tuổi

Đáp án: D

3. Những lợi ích chính của trực quan hóa dữ liệu là gì? Chọn tất cả các câu phù hợp.

- A. Chúng có thể thể hiện rõ ràng các quy luật và xu hướng
- B. Chúng có thể giúp các bên liên quan hiểu dữ liệu phức tạp nhanh hơn
- C. Chúng có thể minh họa mối quan hệ giữa các điểm dữ liệu
- D. Chúng có thể đảm bảo rằng bạn nhận được ít câu hỏi hơn về phân tích của mình

Đáp án: A,B,C

4. Một nhà phân tích dữ liệu tạo ra hình ảnh trực quan sau đây để thể hiện Charlotte đông dân hơn bao nhiêu so với thành phố lớn thứ hai ở phía Bắc Carolina, Raleigh. Đó là loại biểu đồ gì?

- A. Biểu đồ tròn
- B. Biểu đồ điểm
- C. Biểu đồ đường
- D. Biểu đồ cột

Đáp án: D

5. Một nhà phân tích dữ liệu tạo ra hình ảnh trực quan sau đây để thể hiện Charlotte đông dân hơn bao nhiêu so với thành phố lớn thứ hai ở phía Bắc Carolina, Raleigh. Đó là

biểu đồ đường, đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: B

6. Một nhà phân tích dữ liệu muốn thể hiện xu hướng về một thứ đã thay đổi như thế nào theo thời gian. Loại biểu đồ nào là tốt nhất cho nhiệm vụ này?

- A. Đường (line)
- B. Cột (column)
- C. Thanh (bar)
- D. Miền (area)

Đáp án: A

7. Một nhà phân tích dữ liệu muốn thể hiện dân số ở Charlotte tăng theo thời gian thế nào. Đó là ví dụ của lược đồ miền?

- A. Đúng
- B. Sai

Đáp án: B

8. Một nhà phân tích dữ liệu muốn thể hiện dân số ở Charlotte tăng theo thời gian thế nào. Họ tạo ra lược đồ như bên dưới. Đó là loại lược đồ gì?

- A. Lược đồ thanh (bar chart)
- B. Lược đồ cột (column chart)
- C. Lược đồ miền (area chart)

Lược đồ đường (line chart).

Đáp án: D

9. Hãy tưởng tượng bạn đang chia sẻ dữ liệu của mình với một bên liên quan của công ty. Tại sao bạn có thể thể hiện dữ liệu bằng trực quan hóa thay vì bảng? Chọn tất cả những câu phù hợp.

- A. Nó đẹp về mặt thẩm mỹ
- B. Nó dễ hiểu
- C. Nó giúp họ xác định xu hướng một cách nhanh chóng
- D. Nó mô tả kỹ lưỡng từng điểm dữ liệu

Đáp án: A,B,C

Công cụ hỗ trợ trong tiến trình phân tích

1. Dựa trên những gì bạn đã học trong khóa học này, bảng tính là dạng kỹ thuật số cho phép các nhà phân tích dữ liệu thực hiện tác vụ nào sau đây? Chọn tất cả những câu phù hợp.

- A. Chứa dữ liệu
- B. Tổ chức dữ liệu dưới dạng cột và dòng
- C. Chọn chủ đề để phân tích
- D. Sắp xếp và lọc dữ liệu

Đáp án: A,B,D

2. Điền vào chỗ trống: Tập hợp các lệnh thực hiện một phép tính cụ thể trên dữ liệu bảng tính được gọi là ____.

- A. Công thức (formula)
- B. Chương trình (program)
- C. Báo cáo (report)
- D. Thực thi (operation)

Đáp án: A.

3. Cơ sở dữ liệu là một tập hợp dữ liệu được lưu trữ trong hệ thống máy tính. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

4. Trong phân tích dữ liệu, SQL là một từ viết tắt có nghĩa là ngôn ngữ truy vấn _____.

- A. Có cấu trúc (structured)
- B. Phần mềm (software)
- C. Thống kê (statistical)
- D. Cú pháp (syntax)

Đáp án: A

5. Thuật ngữ thể hiện dữ liệu dưới dạng đồ họa là gì?

- A. Thu thập dữ liệu (data collection)
- B. Trực quan hóa dữ liệu (data visualization)
- C. Tóm tắt dữ liệu (data summary)
- D. Ngôn ngữ dữ liệu (data language)

Đáp án: B

6. Điền vào chỗ trống: Công thức (formula) là một tập hợp các lệnh được sử dụng để thực hiện một phép tính xác định; trong khi một hàm (function) là _____.

- A. Một câu hỏi được viết bởi người dùng
- B. Một thực thi đã được định nghĩa trước
- C. Một ngôn ngữ lập trình máy tính
- D. Một giá trị cụ thể

Đáp án: B

7. Điền vào chỗ trống: Một hàm (function) là một thực thi được định nghĩa trước, trong khi một công thức (formula) là _____.

- A. Một phát biểu đã được định nghĩa trước
- B. Một giá trị cụ thể

- C. Một ngôn ngữ lập trình máy tính
- D. Một tập các hướng dẫn được sử dụng để thực thi một tính toán cụ thể.

Đáp án: D

8. Sự khác biệt chính giữa công thức (formula) và hàm (function) là gì?

- A. Một công thức được sử dụng để cộng hoặc trừ; một hàm được sử dụng để nhân và chia
- B. Một công thức là tập các lệnh để thực hiện một tính toán cụ thể; một hàm là một lệnh được thiết lập trước đó để tự động thực hiện một tiến trình cụ thể.
- C. Một công thức bắt đầu với ký hiệu bằng (=); một hàm bắt đầu với ký hiệu hoa thị (*)
- D. Một công thức có thể được sử dụng nhiều lần trong bảng tính; một hàm chỉ có thể dùng một lần duy nhất.

Đáp án: B

9. Điền vào chỗ trống: Để yêu cầu, truy xuất và cập nhật thông tin trong cơ sở dữ liệu, các nhà phân tích dữ liệu sử dụng ____.

- A. Câu truy vấn (query)
- B. Công thức (formula)
- C. Bảng điều khiển (dashboard)
- D. Phép tính (calculation)

Đáp án: A

10. Điền vào chỗ trống: Một truy vấn (query) được sử dụng để ____ thông tin từ cơ sở dữ liệu. Chọn tất cả những câu phù hợp.

- A. Trực quan (visualize)
- B. Yêu cầu (request)
- C. Cập nhật (update)
- D. Truy xuất (retrieve)

Đáp án: B,C,D

11. Các nhà phân tích dữ liệu sử dụng các truy vấn để yêu cầu, thu nhận và cập nhật thông tin trong cơ sở dữ liệu. Phát biểu này đúng hay sai?

A. Đúng

B. Sai

Đáp án: A

12. Một nhà phân tích dữ liệu muốn lấy thông tin từ cơ sở dữ liệu. Chọn một công cụ thích hợp từ bộ đồ nghề của nhà phân tích dữ liệu.

A. Bảng điều khiển (dashboard)

B. Truy vấn (query)

C. Bảng tính (spreadsheet)

D. Trực quan (visualization)

Đáp án: B

13. Ngôn ngữ truy vấn có cấu trúc (SQL) cho phép các nhà phân tích dữ liệu giao tiếp với cơ sở dữ liệu. Phát biểu này đúng hay sai?

A. Đúng

B. Sai

Đáp án: A

14. Điền vào chỗ trống: Ngôn ngữ truy vấn có cấu trúc (SQL) cho phép các nhà phân tích dữ liệu _____ thông tin trong cơ sở dữ liệu. Chọn tất cả những câu phù hợp.

A. Trực quan (visualize)

B. Truy xuất (retrieve)

C. Cập nhật (update)

D. Yêu cầu (request)

Đáp án: B,C,D

15. Sử dụng công cụ bảng tính để nhập và tạo biểu đồ giống như sau:

Đáp án: A....

16. Sau khi bạn thêm dữ liệu vào bảng tính, bạn tạo biểu đồ. Bạn sẽ thực hiện bước nào để làm cho biểu đồ có tính mô tả hơn?

- A. Thêm một tiêu đề cho biểu đồ
- B. Thay đổi màu sắc cho biểu đồ
- C. Xóa cột ngày khỏi biểu đồ
- D. Chuyển đổi biểu đồ dạng cột sang biểu đồ tròn

Đáp án: A

17. Điền vào chỗ trống: Một nhà phân tích dữ liệu tạo một bảng, nhưng họ nhận ra rằng đây không phải là thể hiện trực quan tốt nhất cho dữ liệu của họ. Để khắc phục, họ quyết định sử dụng tính năng _____ để thay đổi nó thành biểu đồ cột.

- A. rename
- B. image
- C. chart editor
- D. filter view

Đáp án: C

18.

- Bảng tính đã giúp gì cho quá trình chuẩn bị dữ liệu của bạn?
- Định dạng biểu đồ như thế nào để giúp bạn phân tích dữ liệu?

Viết 2-3 câu để trả lời cho từng câu hỏi trên..

Đáp án:

Lợi ích của spreadsheet trong việc chuẩn bị (prepare) dữ liệu?

- Giúp cấu trúc dữ liệu dưới dạng dòng và cột
- Tạo ra trực quan hóa dữ liệu

Cần định dạng biểu đồ như thế nào để hỗ trợ cho quá trình phân tích?

- Các biểu đồ (chart) cần được clean up để nó dễ hiểu và thu hút
- Thêm các mô tả về biểu đồ và các trục trong biểu đồ

19. Trong bảng tính, gói văn bản (text wrapping) được sử dụng để làm gì?

- A. Để tự động thay đổi chiều cao ô để cho phép nội dung văn bản vừa bên trong

- B. Để cắt bớt nội dung văn bản trong một ô để nó không tràn sang ô liền kề
- C. Để cho phép văn bản tràn vào một ô liền kề
- D. Để loại bỏ văn bản quá dài để có thể vừa trong một ô

Đáp án: A

20. Các cột trong bảng tính được sắp xếp theo chữ cái và các hàng được sắp xếp theo số. Phát biểu này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

21. Điền vào chỗ trống: Trong bảng dữ liệu, một hàng được gọi là quan sát. Một quan sát bao gồm tất cả _____ được chứa trong hàng.

- A. Thuộc tính (attributes)
- B. Chẩn đoán (diagnostics)
- C. Tên (names)
- D. Điểm chung (commonalities)

Đáp án: A

22. Dấu hoa thị (*) sau khi SELECT yêu cầu cơ sở dữ liệu làm gì trong truy vấn sau?

- A. Chọn tất cả dữ liệu đáp ứng các tiêu chí như đã nêu trong truy vấn
- B. Chọn tất cả các cột từ bảng employee
- C. Chọn tất cả dữ liệu đáp ứng các tiêu chí như đã nêu trong truy vấn, sau đó nhân nó
- D. Chọn cột LastName từ bảng nhân viên

Đáp án: B

23. Trong câu truy vấn sau, nhà phân tích dữ liệu muốn nhận dữ liệu từ bảng nào?

- A. employee
- B. LastName

- C. James
- D. jobCode

Đáp án: A.

24. Trong câu truy vấn sau, cái gì được nhận từ cơ sở dữ liệu?

- A. Tất cả dữ liệu từ bảng employee, trong đó jobCode là FTE và tên là James
- B. Tất cả dữ liệu từ bảng FTE, trong đó Last Name là James
- C. Tất cả dữ liệu từ bảng jobCode, trong đó jobCod là FTE và nhân viên có tên khác James
- D. Tất cả dữ liệu từ bảng employee, trong đó jobCode là FTE và nhân viên có tên khác James

Đáp án: A

25. Trong bảng tính bên, các từ như rank, name, population, county được gọi là gì?

- A. Tiêu chí (criteria)
- B. Thuộc tính (attributes)
- C. Mô tả (descriptors)
- D. Đặc tính (characteristics)

Đáp án: B

26. Trong dòng 8 của bảng tính bên, chúng ta có thể thấy ____ của Cary?

- A. Thuộc tính (attribute)
- B. Định dạng (format)
- C. Quan sát (observation)
- D. Tiêu chí (criteria)

Đáp án: C

27. Trong bảng tính bên, đặc trưng _____ được sử dụng để sắp xếp tên thành phố trong cột B theo thứ tự chữ cái?

- A. Miền tổ chức (organize range)
- B. Miền sắp xếp (sort range)
- C. Miền ngẫu nhiên (randomize range)
- D. Miền tên (name range)

Đáp án: B

28. Nhà phân tích dữ liệu ghi
= POPULATION (C2: C11) để tìm dân số

trung bình của các thành phố trong bảng tính bên. Tuy nhiên, họ nhận ra rằng đã sử dụng sai công thức. Cú pháp nào sẽ sửa công thức này?

- A. =AVERAGE(C2:C11)
- B. AVERAGE(C2:C11)
- C. =AVERAGE(C2-C11)
- D. AVERAGE(C2-C11)

Đáp án: A

Nghề nghiệp cho nhà phân tích dữ liệu

1. Nhà phân tích dữ liệu thực hiện những bước nào để đảm bảo tính công bằng khi thu thập dữ liệu? Chọn tất cả những câu phù hợp.

- A. Sử dụng quần thể mẫu liên quan
- B. Hiểu ngữ cảnh xã hội
- C. Hướng đến dữ liệu tự báo cáo

D. Làm sạch dữ liệu

Đáp án: A,B,C

2. Avens Engineering cần nhiều kỹ sư hơn, vì vậy họ mua quảng cáo trên trang web tìm kiếm việc làm. Dữ liệu của trang web cho thấy 86% kỹ sư là nam giới. Dựa trên con số đó, một nhà phân tích quyết định rằng nam giới có nhiều khả năng trở thành ứng viên thành công hơn, vì vậy họ nhắm mục tiêu quảng cáo đến những người tìm việc là nam giới. Nhà phân tích nên làm gì sẽ tốt hơn?

- A. Để Avens Engineering quyết định loại ứng viên nào sẽ nhắm mục tiêu quảng cáo.
- B. Chỉ hiển thị quảng cáo về công việc kỹ thuật cho phụ nữ
- C. Đảm bảo rằng đề xuất của họ không tạo ra hoặc gia tăng sự thiên lệch
- D. Từ chối chấp nhận quảng cáo từ Avens Engineering vì lo ngại về tính công bằng

Đáp án: C.

3. Trên một tuyến đường sắt, lượng hành khách cao điểm xảy ra từ 7:00 sáng đến 5:00 chiều. Tính công bằng của cuộc khảo sát hành khách có thể được cải thiện bằng cách lấy mẫu thêm từ nhóm nào?

- A. Các hành khách nữ
- B. Các hành khách nam
- C. Người đi ban ngày
- D. Người đi ban đêm

Đáp án: D

4. Công ty bất động sản cần tuyển trợ lý nhân sự. Chủ sở hữu yêu cầu một nhà phân tích dữ liệu để giúp họ quyết định nơi quảng cáo việc làm. Nhà phân tích biết rằng phần lớn các chuyên gia nhân sự là phụ nữ, xác nhận kết quả này bằng nghiên cứu và nhắm mục tiêu quảng cáo đến một trường cao đẳng cộng đồng dành cho phụ nữ. Điều này là công bằng vì nhà phân tích đã tiến hành nghiên cứu để đảm bảo thông tin về phân tích giới của các chuyên gia nhân sự là chính xác. Cách làm này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: B

(Điều này không công bằng. Công bằng có nghĩa là đảm bảo rằng phân tích không tạo ra hoặc gia tăng sự thiên lệch. Là một nhà phân tích dữ liệu, điều quan trọng là phải giúp tạo ra các hệ thống công bằng và hòa nhập cho tất cả mọi người.)

5. Một tạp chí trực tuyến về chủ đề vườn muốn hiểu tại sao số lượng người đăng ký ngày càng tăng. Nhà phân tích dữ liệu có thể cung cấp loại báo cáo nào để giúp trả lời câu hỏi này? Chọn tất cả các câu phù hợp.

- A. Các báo cáo so sánh các quy luật thời tiết trong quá khứ với số lượng người đặt câu hỏi về việc làm vườn trên mạng xã hội
- B. Các báo cáo dự đoán thành công của việc bán hàng dẫn đến việc đảm bảo người đăng ký trong tương lai
- C. Các báo cáo mô tả số lượng khách hàng đã chia sẻ nhận xét tích cực về tạp chí làm vườn trên mạng xã hội trong năm qua
- D. Các báo cáo kiểm tra mức độ giảm giá 50% gần đây ảnh hưởng như thế nào đến số lượng đăng ký mua

Đáp án: A,C,D

6. Một văn phòng bác sĩ đã phát hiện ra rằng bệnh nhân đang đợi cuộc hẹn của họ lâu hơn 20 phút so với những năm trước. Một nhà phân tích dữ liệu có thể giúp giải quyết vấn đề này bằng cách phân tích số lượng bác sĩ và y tá có trong nhân viên tại một thời điểm nhất định so với số lượng bệnh nhân đặt lịch hẹn. Cách làm này có đúng không?

- A. Đúng
- B. Sai

Đáp án: A

7. Một bài toán là một trở ngại cần được giải quyết, một vấn đề là một chủ đề để điều tra và một câu hỏi được thiết kế để khám phá thông tin.

- A. Đúng
- B. Sai

Đáp án: A

8. Một câu hỏi hay bài toán mà nhà phân tích dữ liệu trả lời cho một doanh nghiệp là gì?

- A. Sự phàn nàn (complaint)
- B. Nhiệm vụ kinh doanh (business task)
- C. Giả thuyết (hypothesis)
- D. Phát biểu nhiệm vụ (mission statement)

Đáp án: B

9. Ra quyết định dựa trên dữ liệu là sử dụng các dữ kiện để định hướng chiến lược kinh doanh. Những lợi ích mang lại bao gồm những điều nào sau đây? Chọn tất cả những câu phù hợp.

- A. Sử dụng phân tích dữ liệu để tìm ra giải pháp tốt nhất có thể cho một vấn đề
- B. Kết hợp quan sát với dữ liệu khách quan
- C. Tận dụng tối đa trực giác và bản năng
- D. Thấy được một bức tranh toàn cảnh về vấn đề và nguyên nhân của nó

Đáp án: A,B,D

10. Quá trình sử dụng các dữ kiện để định hướng chiến lược kinh doanh là gì?

- A. Trực quan hóa dữ liệu
- B. Ra quyết định hướng dữ liệu
- C. Lập trình dữ liệu
- D. Đạo đức dữ liệu

Đáp án: B

11. Các nhà phân tích dữ liệu đảm bảo phân tích của họ là công bằng vì lý do gì?

- A. Công bằng giúp họ lấy và chọn dữ liệu nào sẽ đưa vào từ tập dữ liệu
- B. Công bằng giúp họ giao tiếp với các bên liên quan
- C. Công bằng giúp họ tránh được những kết luận thiên vị
- D. Công bằng giúp họ luôn ở trạng thái có tổ chức

Đáp án: C

12. Ví dụ nào sau đây mô tả sự công bằng trong phân tích dữ liệu? Chọn tất cả những câu phù hợp.

- A. Xem xét các yếu tố hệ thống có thể ảnh hưởng đến dữ liệu
- B. Lấy và chọn dữ liệu để đưa vào từ một tập dữ liệu
- C. Các yếu tố trong ngữ cảnh xã hội có thể tạo ra sự sai lệch trong kết luận
- D. Đảm bảo quần thể mẫu đại diện cho tất cả các nhóm

Đáp án: A,C,D

13. Điền vào chỗ trống: Công bằng đạt được khi phân tích dữ liệu không tạo ra hoặc _____ thiên vị.

- A. Tô điểm (highlight)
- B. Gia tăng (reinforce)
- C. Ràng buộc (constrain)
- D. Giải quyết (resolve)

Đáp án: B

14. Một nhà phân tích dữ liệu đang phân tích việc bán trái cây và rau quả tại một cửa hàng tạp hóa. Họ có thể tìm thấy dữ liệu về mọi thứ ngoại trừ hành tím. Cách tốt nhất của hành động là gì?

- A. Loại trừ hành tím khỏi phân tích
- B. Loại trừ tất cả các giống hành tây khỏi phân tích
- C. Nhờ đội ngũ giúp tìm dữ liệu về hành tím
- D. Thay vào đó, sử dụng dữ liệu về hành tây trắng, vì chúng đều là giống hành tây

Đáp án: C

15. Một phòng tập thể dục muốn bắt đầu cung cấp các lớp tập thể dục. Một nhà phân tích dữ liệu có kế hoạch khảo sát 10 người để xác định xem lớp nào sẽ được yêu thích nhất. Để đảm bảo dữ liệu được thu thập là công bằng, họ nên thực hiện những bước

nào? Chọn tất cả những câu phù hợp.

- A. Đảm bảo những người tham gia đại diện cho nhiều loại hồ sơ và lý lịch
- B. Thu thập dữ liệu ẩn danh
- C. Tăng số lượng người tham gia
- D. Chỉ khảo sát những người hiện không đến phòng tập thể dục

Đáp án: A,B,C.

16. Cộng tác với một nhà khoa học xã hội để cung cấp thông tin chi tiết về thành kiến của con người và bối cảnh xã hội là một cách hiệu quả để tránh sai lệch trong dữ liệu của bạn. Nhận định này đúng hay sai?

- A. Đúng
- B. Sai

Đáp án: A

17. Một chuỗi khách sạn lớn đón khoảng 500 khách hàng mỗi tuần. Một nhà phân tích dữ liệu làm việc ở đó đang thu thập dữ liệu thông qua các cuộc khảo sát về mức độ hài lòng của khách hàng. Họ nóng lòng muốn bắt đầu phân tích, vì vậy họ bắt đầu phân tích dữ liệu ngay khi nhận được 50 câu trả lời khảo sát. Đây là một ví dụ về cái gì? Chọn tất cả những câu phù hợp.

- A. Thất bại trong đa dạng của dữ liệu thu thập
- B. Thất bại trong thu thập dữ liệu ẩn danh
- C. Thất bại trong trao thưởng cho khách hàng vì đã tham gia khảo sát
- D. Thất bại trong việc có kích thước dữ liệu đủ lớn

Đáp án: A,D

18. Điền vào chỗ trống: Một nhiệm vụ kinh doanh được mô tả là bài toán hoặc _____ mà một nhà phân tích dữ liệu trả lời cho một doanh nghiệp.

- A. Phàn nàn (complaint)
- B. Giải pháp (solution)
- C. Câu hỏi (question)

D. Góp ý (comment)

Đáp án: C

19. Mô tả sự khác biệt giữa một câu hỏi và một bài toán trong phân tích dữ liệu.

- A. Một câu hỏi không chắc chắn, trong khi một bài toán được xác định rõ ràng
- B. Một câu hỏi có thể có nhiều câu trả lời, trong khi một bài toán chỉ có một giải pháp
- C. Một câu hỏi được thiết kế để khám phá thông tin, trong khi một bài toán là một trở ngại hoặc phức tạp cần được giải quyết
- D. Một câu hỏi là một chủ đề để điều tra, trong khi một bài toán là một đối tượng để điều tra

Đáp án: C

20. Văn phòng bác sĩ phát hiện ra rằng bệnh nhân đang đợi cuộc hẹn của họ lâu hơn 20 phút so với những năm trước. Nhà phân tích dữ liệu có thể giúp giải quyết vấn đề này bằng những cách nào? Chọn tất cả những câu phù hợp.

- A. Phân tích độ dài trung bình của một cuộc hẹn trong năm nay so với những năm trước
- B. Phân tích xem có bao nhiêu bác sĩ và y tá có nhân viên tại một thời điểm nhất định so với số lượng bệnh nhân có cuộc hẹn
- C. Phân tích sự thay đổi gần đây về xếp hạng trung bình cho phòng khám bác sĩ trên mạng xã hội
- D. Phân tích số lượng bệnh nhân khám mỗi ngày so với những năm trước

Đáp án: A,B,D

21. Một tạp chí làm vườn trực tuyến muốn hiểu tại sao số lượng người đăng ký của họ ngày càng tăng. Một nhà phân tích dữ liệu phát hiện ra rằng có nhiều người đăng ký hơn đáng kể khi tạp chí có chương trình giảm giá 50% hàng năm. Đây là một ví dụ về cái gì?

- A. Phân tích mức độ tương tác trên mạng xã hội
- B. Phân tích sở thích của người tiêu dùng bằng trí tuệ nhân tạo
- C. Phân tích hành vi mua của khách hàng

D. Phân tích số lượng khách hàng bằng cách tính lưu lượng truy cập hàng ngày

Đáp án: C

22. Một tạp chí muốn hiểu lý do tại sao số người đăng ký của họ ngày càng tăng. Một nhà phân tích dữ liệu có thể giúp trả lời câu hỏi đó bằng một báo cáo dự đoán kết quả của việc bán nửa giá đối với tỷ lệ đăng ký trong tương lai. Cách làm này đúng hay không?

A. Đúng

B. Sai

Đáp án: B

23. Một nhà phân tích dữ liệu đang phân tích việc bán trái cây và rau quả tại một cửa hàng tạp hóa. Họ có thể tìm thấy dữ liệu về mọi thứ ngoại trừ hành tím. Nếu họ loại trừ hành tím khỏi phân tích, đây sẽ là một ví dụ về việc tạo ra hoặc gia tăng sự thiên vị. Nhận định này đúng hay sai?

A. Đúng

B. Sai

Đáp án: A.

24. Các nhà phân tích dữ liệu trả lời câu hỏi và giải quyết vấn đề. Đây được gọi là nhiệm vụ kinh doanh (business task). Nhận định này đúng hay sai?

A. Đúng

B. Sai

Đáp án: A

25. Điền vào chỗ trống: Văn phòng bác sĩ đã phát hiện ra rằng bệnh nhân đang đợi cuộc hẹn của họ lâu hơn 20 phút so với những năm trước. Để giúp giải quyết vấn đề này, một nhà phân tích dữ liệu có thể điều tra xem có bao nhiêu y tá trong đội ngũ tại một thời điểm nhất định so với số ____.

A. Bác sĩ cùng thời điểm

B. Bệnh nhân đặt lịch hẹn

C. Bác sĩ trông nom bệnh nhân mới

D. Những góp ý tiêu cực về thời gian đợi trên mạng xã hội

Đáp án: B

26. Điền vào chỗ trống: Một bài toán là một trở ngại cần được giải quyết, một vấn đề là một chủ đề để điều tra và _____ được thiết kế để khám phá thông tin.

A. Câu hỏi (question)

B. Chủ đề (theme)

C. Nhiệm vụ kinh doanh (business task)

D. Đột phá (breakthrough)

Đáp án: A
