Chương III: TRỰC QUAN HÓA DỮ LIỆU

3.1 Giới thiệu về trực quan hóa dữ liệu

Dữ liệu là gì?

* Dữ liệu là thông tin hoặc sự tập hợp của các yếu tố, sự kiện, số liệu, sự mô tả, hay bất kì thứ gì có khả năng được thu nhập, ghi lại hoặc biểu diễn.
* Dữ liệu có thể tồn tại ở nhiều dạng khác nhau, và nó có giá trị khi được xem xét, phân tích, xử lí hoặc sử dụng để hổ trợ ra quyết định hoặc hiểu về hiện tượng nào đó.

Trực quan hóa dữ liệu là gì?

* Trực quan hóa dữ liệu là quá trình biểu diễn dữ liệu bằng các hình ảnh, biểu đồ, đồ thị, hoặc các phương tiện trực quan khác để giúp hiểu rõ dữ liệu một cách dễ dàng và nhanh chóng.
* Trực quan hóa dữ liệu giúp biến dữ liệu trừu tượng và khó hiểu thành hình ảnh hoặc biểu đồ dễ hiểu và hấp dẫn.

Giới thiệu về trực quan hóa dữ liệu

* Trực quan hóa dữ liệu là nghệ thuật và khoa học kể những câu chuyện hấp dẫn bằng dữ liệu.
* Các nhà phát triển và nhà khoa học dữ liệu ngày nay, bất kể lĩnh vực hoạt động, đều đồng ý rằng việc truyền đạt thông tin chi tiết một cách hiệu quả bằng cách sử dụng trực quan hóa dữ liệu là rất quan trọng

3.2 Các loại biểu đồ cơ bản

Biểu đồ cột( Bar chart)

* Biểu đồ cột/thanh là sự thể hiện trực quan các giá trị ở dạng thanh hình chữ nhật ngang hoặc dọc.
* Chiều cao( dài) tỉ lệ thuận với các giá trị được biểu diễn. Được hiển thị theo hai trục, một trục hiển thị phần tử và trục còn lại hiển thị giá trị của phần tử.
* Biểu đồ thanh được hiển thị theo chiều odjc còn được gọi là biểu đồ cột và biểu đồ thanh ngang được gọi là biểu đồ thanh trong một số công cụ như Microsoft Excel.

Biểu đồ đường( line chart)

* Biểu đồ đường là biểu đồ hai chiều của các giá trị được kết nối theo thứ tự. Các giá trị được hiển thị (hoặc phân tán) một cách có trật tự và được kết nối.
* Biểu đồ đường cho thấy xu hướng của một hoặc nhiều phần tử so với thời gian.

Biểu đồ đường có thể là:

* Biểu đồ đường đơn giản( hiển thị giá trị của một phần tử trên một tham chiếu- chẳng hạn như thời gian)
* Nhiều biểu đồ đường – hiển thị nhiều giá trị trên một điểm tham chiếu tương tự - chẳng hạn như giá cổ phiếu của nhiều công ty theo thời gian( hiển thị bằng các màu khác nhau)
* Splines – biểu đồ đường biểu diwwnx đường nối cong của các điểm thay vì đường thẳng.
* Biểu đồ đường bậc thanh- nơi kết nối giauwx các điểm được hiển thị trong một bước.

Biểu đồ tròn (Pie chart)

Biểu đồ tròn là gì?

* Biểu đồ tròn hiển thị tỷ lệ hoặc phần trăm của phần tử dữ liệu ở định dạng hình tròn. Biểu đồ hình tròn được hia thành nhiều phần khác nhau dựa trên giá trị/phần trăm của phần tử dữ liệu cần làm nổi bật.
* Tổng của toàn bộ phần tử tương ứng với 10% giá trị của dữ liệu được hiển thị.
* Biểu đồ tròn có thể biểu diễn ở dạng 3D.

Biểu đồ điểm ( Scatter plot)

* Biểu đồ phân tán là biểu đồ hai chiều hiển thị sự so sánh của hai biến theo tọa độ toán học.
* Trong đó, trục tung Oy là ccs giá trị của biến được dự đoán ( biến phụ thuộc ) và trục hoành Ox là các giá trị của biến dùng để đưa ra dự đoán( biến độc lập).
* Đây là một đồ thị biểu hiện mối tương quan giữa nguyên nhân và kết quả.

Biểu đồ bong bóng( Bubble chart)

* Bubble chart là một biểu đồ phổ biến để biểu thị dữ liệu có mối quan hệ giữa ba chiều: hoành độ, tung độ, và kích cỡ.
* Bubble chart cho phép giải thích mối quan hệ giữa hai biến và thể hiện cả mức dộ sự liên quan giữa hai biến đó. Kích cỡ của các điểm trên biểu đồ có thể được sử dụng để biểu thị một giá trị thứ 3.

Biểu đồ phân phối tần suất ( histogram)

* Biểu đồ histogram là một dạng biểu đồ thể hiện tần suất theo dạng cột. Dữ liệu được hiển thị bằng các cột trên biểu đồ có độ cao khác nhau tùy thuộc vào tần suất ( bao nhiêu lần) phạm vi dữ liệu cụ thể xảy ra.

Biểu đồ Doughnut

* Biểu đồ doughnut có ứng dụng tương tự biểu đồ tròn.
* Tuy nhiên, bởi chứa được nhiều hơn một dải dữ liệu, doughnut chart có thể biểu diễn được sự thay đổi của các phần dữ liệu
* Với mỗi thay đổi, một vòng tròn mới sẽ được thêm vào.

Biểu đồ vùng( area chart)

* Biểu đồ vùng thể hiện mối quan hệ giữa các giá trị và thời gian hoặc các giá khác nhah. Nó biểu thị dữ liệu về tăng giảm và sự thay đổi của các giá trị theo thời gian hoặc them một trục xác định.
* Các giá trị được biểu diễn bằng một chuỗi điểm được kết nối bởi một đường bằng màu sắc cho phép bạn theo dõi dữ liệu dễ dàng.

Biểu đồ mật độ ( density plot)

* Biểu đồ mật độ có liên quan chặt chẽ với biểu đồ histogram và lấy một bộ giá trị số làm đầu vào. Đầu ra của biểu đồ mật độ là sự hiện thị phân bố dữ liệu. Phân phối có thể ở dạng hàm mũ hoặc dạng đường cong hình chuông. Biểu đồ có thể bị lệch sang trái hoặc sang phải dựa trên khối lượng dữ liệu cho một phạm vi dữ liệu cụ thể.

Bản đồ nhiệt( Heatmap)

* Bản đồ nhiệt là một công cụ thể hiện tầm quan trọng của các phần tử dữ liệu bằng cách sử dụng màu sắc.

3.3 Các thành phần cơ bản của biểu đồ