Data Abstraction:

**Dataset**

* Dataset: tĩnh
* Type: Table

**Items**

1. InvoiceNo: Số hoá đơn bao gồm 6 chữ số. Nếu mã này bắt đâu bằng ‘c’, nó chỉ ra một giao dịch bị huỷ bỏ.
2. StockCode: Mã sản phẩm bao gồm 5 chữ số.
3. Description: Tên sản phẩm.
4. Quantity: Số lượng sản phẩm trên mỗi giao dịch.
5. InvoiceDate: Ngày và giờ mỗi khi giao dịch được tạo.
6. UnitPrice: Giá sản phẩm trên một cái.
7. CustomerID: Số khách hàng bao gồm 5 chữ số. Mỗi khách hàng có một ID khách hàng duy nhất.
8. Country: Tên quốc gia nơi mỗi khách hàng cư trú.

**Abstact**

1. InvoiceNo: Ordered – Ordinal, đặc trưng: liên tục, không phân cấp, hướng : tuần tự.
2. StockCode: Categorical, đặc trưng: rời rạc , không phân cấp
3. Description: Categorical, đặc trưng: rời rạc, không phân cấp.
4. Quantity: Ordered - Quantitative – Ratio, đặc trưng: liên tục, không phân cấp, hướng: tuần tự.
5. InvoiceDate: Ordered - Quantitative- Ratio, phân cấp: ngày / tháng / năm, chiều : tuần hoàn (tháng ), đặc trưng: liên tục.
6. UnitPrice: Ordered - Quantitative - Ratio, đặc trưng: liên tục, chiều: tuần tự, không phân cấp.
7. CustomerID: Quantitative – Ordinal, đặc trưng: liên tục, chiều: tuần tự, không phân cấp.
8. Country: Categorical, đặc trưng: tuần tự, không phân cấp.

Profiling:

Task Abstract:

**Câu hỏi a:Cho biết doanh thu của các tháng trong năm 2011**

1. Tính tổng tiền của từng khách hàng.

Action: Analyze - Produce - derive

Target: Attribution – One

1. Tính tổng doanh thu từng tháng trong năm 2011

Action: Analyze – Produce – Derive

Target: Attibution – One

**Câu hỏi b: Cho biết top 5 quốc gia có tổng doanh thu mua hàng cao nhất.**

1. Tính tổng tiền của từng khách hàng.

Action: Analyze – Produce – Derive

Target: Attribution - One

1. Tìm tổng dọanh thu của từng quốc gia.

Target: Attribution – One

Action: Search – Lookup

Action: Analyze - Produce - derive

1. Tìm top 5 quốc gia có doanh thu cao nhất.

Action: Query – Compare

Target: Attribution – One

**Câu hỏi c: Cho biết top 10 sản phẩm bán chạy nhất.**

1. Tính tổng số lượng của từng mặt hàng.

Action: Analyze – Produce – Derive

Target: Attribution - One

1. Tìm top 10 sản phẩm bán chạy nhất.

Action: Query – Compare

Target: Atribution – One

Thiết kế Idiom và Vẽ biểu đồ:

**Câu hỏi a:** Cho biết tỉ lệ giá trị mua hàng của United Kingdom so với giá trị mua hàng của Netherlands.

|  |  |
| --- | --- |
| **Idiom** | **Line Chart** |
| **Data Type** | Table: gồm 2 thuộc tính ordered tự dẫn xuất (Month và Doanh thu) |
| **Encoding** | -Mark: Line và point  -Channel:   * Align: lengths (quantitative) * Separate, order: horizontal region(ordered) * Color |
| **Task** | Thể hiện xu hướng |

Chart, line chart

Description automatically generated

**Nhận xét:**

1. **Nguyên lí biểu đạt:**

* Dựa vào biểu đồ ta có thể thấy được tháng 11/2011 có doanh thu cao nhất.
* Tháng 12 có doanh thu thấp nhất.
* Từ đây có thể giúp ban quản lí đặt ra câu hỏi tạo sao tháng 9,10,11 lại có doanh thu cao nhưng tháng 12 doanh thu lại giảm nhanh chóng.
* Biểu đò sử dụng marks line và point giúp cho biểu đồ có điểm nhấn, giúp người xem có thể dễ dàng phân biệt.

1. **Đánh giá độ hiệu quả:** 
   1. **Chính xác**

Có thể phân biệt giữa encoding items => Độ hiệu quả cao

* 1. **Phân biệt**

Có thể phân biệt giữa các tháng, có sự kết hợp giữa marks line và point làm cho người xem dễ dàng phân biệt, đồng thời kết hợp color làm cho đồ thị rõ ràng hơn.

* 1. **Tách biệt**

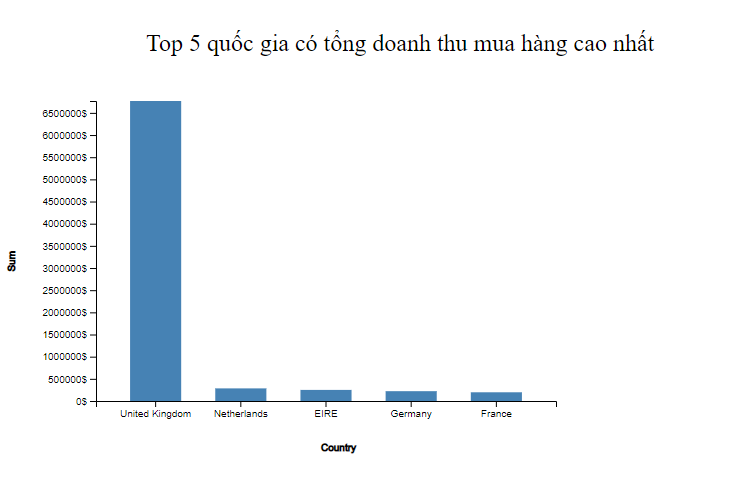
Các encoding items không ảnh hưởng gì đến nhau.

* 1. **Nổi bật**

Mark point và line là sự kết hợp tuyệt vời, giúp người xem dễ dàng nhận biết giữa các tháng.

**Câu hỏi b:** Cho biết top 5 quốc gia có tổng doanh thu mua hàng cao nhất.

|  |  |
| --- | --- |
| **Idiom** | **Bar Chart** |
| **Data Type** | Một thuộc tính categorical có sẵn trong tập dữ liệu, một thuộc tính ordered tự dẫn xuất (Sum) |
| **Encoding** | * Marks: line, area * Categorical (Description) : X position * Ordered(Quantity): Y position * Color * Separate, Align, order : X positon |
| **Task** | Tốt cho việc tìm kiếm và so sánh giữa các giá trị với nhau |



**Nhận xét:**

1. **Nguyên lí biểu đạt**

* Dựa vào biểu đồ ta có thể thấy được doanh thu của United Kingdom là cao nhất.
* Các quốc gia còn lại gần như gần bằng nhau.
* Thuộc tính categorical (Country): có căn chỉnh và sắp xếp giảm dần theo doanh thu
* Thuộc tính ordered(Sum): dùng length và area để thể hiện

1. **Đánh giá độ hiệu quả:** 
   1. **Chính xác**

Có thể phân biệt giữa các encoding itmes => độ chính xác cao.

* 1. **Phân biệt**

Thuộc tính ordered (Sum) dùng length để thể hiện doanh thu, dễ dàng phân biệt doanh thu của các quốc gia khác.

* 1. **Tách biệt**

Sử dụng màu sắc, area, căn chỉnh và sắp xếp giảm dần theo doanh thu nên có sự tách biệt rõ.

* 1. **Nổi bật**

Có sự căn chỉnh ở X position, kết hợp với length làm cho biểu đồ nổi bật, dễ dàng so sánh doanh thu giữa các quốc gia với nhau.

**Câu hỏi c:** Cho biết top 10 sản phẩm bán chạy nhất.

|  |  |
| --- | --- |
| **Idiom** | **Bar Chart** |
| **Data Type** | Một thuộc tính categorical, một thuộc tính ordered (table) |
| **Encoding** | * Marks: line, area * Categorical (Description) : X position * Ordered(Quantity): Y position * Separate, Align, order: X positon |
| **Task** | Tốt cho việc tìm kiếm và so sánh giữa các giá trị với nhau |

**Chart, bar chart, histogram

Description automatically generated**

**Nhận xét:**

1. **Nguyên lí biểu đạt**

* Dựa vào biểu đồ ta có thể thấy được WHITE HANGING HEART T-LIGHT HOLDER được bán nhiêu nhất với số lượng 2070
* Thấp nhất là mặt hàng PACK OF 72 RETROSPOT CAKE CASES, số lượng bán ra là 1080.
* Biểu đồ dùng separate, order, align thể hiện ở trục x
* Dùng length và are thể hiện ở trục y
* Có dùng màu sắc để làm nổi bật cho đồ thị.

1. **Đánh giá độ hiệu quả:** 
   1. **Chính xác**

Có thể phân biệt giữa các encoding itmes => độ chính xác cao.

* 1. **Phân biệt**

Thuộc tính ordered (Quantity) dùng length để thể hiện số lượng, dễ dàng phân biệt số lượng giữa các sản phẩm.

* 1. **Tách biệt**

Sử dụng màu sắc, area, căn chỉnh và sắp xếp giảm dần theo số lượng (Quantity) nên có sự tách biệt rõ.

* 1. **Nổi bật**

Có sự căn chỉnh ở X position, kết hợp với length làm cho biểu đồ nổi bật, dễ dàng so số lượng sản phẩm giữa các mặt hàng với nhau.