ĐẠI HỌC QUỐC GIA TPHCM – TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN

KHOA CÔNG NGHỆ THÔNG TIN – BỘ MÔN HỆ THỐNG THÔNG TIN

A picture containing logo

Description automatically generated

**BÀI TẬP NHÓM**

**(HW#4)**

**Môn học: Hệ thống thông tin phục vụ trí tuệ kinh doanh**

*GVHD: Hồ Thị Hoàng Vy*

|  |  |  |
| --- | --- | --- |
| **Mã nhóm** | **MSSV** | **Họ và tên** |
| **TTKD-24** | 18120227 | Phạm Văn Minh Phương |
| 18120299 | Trương Công Quốc Cường |
| 18120035 | Đoàn Nguyễn Tấn Hưng |
| 18120534 | Hoàng Công Sơn |

**BẢNG PHÂN CÔNG**

|  |  |  |  |
| --- | --- | --- | --- |
| MSSV | Họ và tên | Phân công | Đánh giá % |
| 18120227 | Phạm Văn Minh Phương | Phân loại thuộc tính cần thiết, kiểm và sửa lỗi, làm báo cáo | 100% |
| 18120299 | Trương Công Quốc Cường | Giải thích thuộc tính từ bảng Accidents | 100% |
| 18120035 | Đoàn Nguyễn Tấn Hưng | Giải thích thuộc tính từ bảng Casualties và bảng LSOA-Postcode Mapping | 100% |
| 18120534 | Hoàng Công Sơn | Giải thích thuộc tính từ bảng Vehicles và bảng Postcodes | 100% |

Mục lục

[I. Giải thích ý nghĩa thuộc tính nguồn dữ liệu 4](#_Toc85060512)

# **Case study**

1. Hiện trạng

Công ty Amadeus Entertainment có 3 nguồn dữ liệu:

* Jade (DB2) – Product Sale Offline
* Jupiter (SQL server) – Inventory & products
* Webtower9 (Oracle) – Product sales (online)

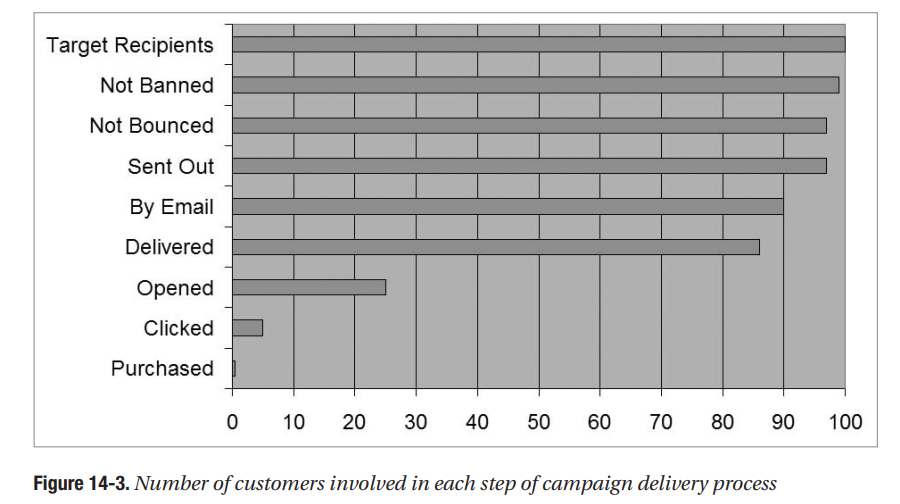
Công ty là một nhà bán lẻ chủ yếu: films, audio books, music.

* Có tám cửa hàng online vận hành ở Mỹ, Đức, pháp, Anh, Tây Ban Nha, Úc, Nhật và Ấn Độ.
* Có 96 cửa hàng offline cũng trên những quốc gia này.
* Khách hàng có thể mua hàng cá nhân với các sản phẩm: bài hát, audio book, phim, hoặc đăng ký gói cho phép download một số sản phẩm trong 1 khoảng thời gian nhất định
* Công ty có nhiều kênh phân phối: TV, internet, mobile phone…
* Có nhiều phương thức chi trả: trả hàng năm, trả trước hàng tháng…

1. Nhu cầu
2. Người dùng nghiệp vụ cần phân tích bán hàng (product sales). Cần biết được doanh số chi phí, lợi nhuận theo một loại tiền tệ ($)
3. Người dùng nghiệp vụ cần phân tích “bán hàng qua hình thức đăng ký” (subscription sales). Cần biết được doanh số, chi phí, lợi nhuận được đánh giá hằng ngày trong khoảng 1 tháng
4. Cấp cửa hàng, có thể xem dữ liệu hằng ngày trong một vài tuần trước đó ở mức tổng quát hoặc chi tiết về doanh số, chi phí, lợi nhuận để hiểu nguyên nhân của việc doanh số thấp, lợi nhuận thấp, sản phẩm nào, khách hàng nào… là nguyên nhân của vấn đề
5. Cấp quản lý toàn cục có thể hiểu được xu hướng chung tổng thể hoặc từng quốc gia. Liệu một cửa hàng hay một quốc gia đang có vấn đề với một sản phẩm nào để có thểtương tác sớm với các cửa hàng nhất có thể
6. Hệ thống có thể hiển thị các hình, biểu đồ cho phép in được, có thể xuất ra excel hoặc csv..
7. Phục vụ phân tích các chiến dịch chăm sóc khách hàng qua việc xem các đo lường của mỗi chiến dịch sau khi gởi cho khách hàng: số lượng thông điệp được gởi của mỗi kênh (mobile message, email, post…), số thông điệp phân phối thành công, thất bại, tỷ lệ mở, tỷ lệ bỏ qua, than phiền, đánh spam, …
8. Phục vụ phân đoạn khách hàng cho các chiến dịch chăm sóc khách hàng dựa vào quyền thông tin, đặc trưng địa lý, thông tin khách hàng, sở thích, lịch sử mua hàng…
9. Phân tích hiệu suất nhà cung ứng (supplier performance) – trung bình chi phí, giá trị trả về, từ chối, thời gian ngừng hoạt động..

# **Phân tích yêu cầu**

Giả sử khi chiến dịch được thực hiện, đã có 100.000 người nhận mục tiêu. Giả sử 1.000 trong số những người nhận mục tiêu này đã bị lọc ra do bị ban; 2.000 bị lọc ra vì địa chỉ e-mail của người nhận đã bị trả lại bốn lần. Vì vậy, 97.000 quảng cáo đã thực sự được gửi đi: 90.000 qua email; 4.000 qua đường bưu điện; 2.000 bởi RSS; và 1.000 bằng tin nhắn văn bản. Trong số 90.000 email, có 86.000 thư đến được người nhận dự định và 4.000 thư chưa được gửi (vì miền email không tồn tại, vì email không hợp lệ, vì hộp thư đã đầy, v.v.). Trong số 86.000 e-mail đã được gửi đi, 25.000 đã được mở / đọc và 5.000 được nhấp; nghĩa là, khách hàng đã truy cập vào các trang web được chỉ định. Cụ thể xem hình biểu diễn tỉ lệ dưới đây:



Số lượng tin nhắn e-mail đã được gửi đi, số lượng tin nhắn đã gửi, số lượng tin nhắn không được gửi (bị trả lại) và số lượng e-mail không được gửi đi (trong danh sách bị cấm hoặc danh sách bị trả lại) được gọi là dữ liệu phân phối chiến dịch và số lượng tin nhắn e-mail đã mở / đọc, số lượng tin nhắn được khách hàng nhấp vào và số lượng khách hàng đã mua một số sản phẩm được gọi dữ liệu phản hồi của chiến dịch. Cả dữ liệu phân phối chiến dịch và dữ liệu phản hồi của chiến dịch đều quan trọng vì chúng cho biết sự thành công hay thất bại của chiến dịch; ví dụ: nếu tỷ lệ mở và tỷ lệ nhấp (CTR) cao nghĩa là chiến dịch thành công.

* Cần thiết phải lưu thông tin về trạng thái của các email (đã gửi, đã nhận, bị trả lại, đã xem, được click, bị phàn nàn, bị đánh dấu spam…) và các thông tin liên quan tới email đó, đặc biệt là các thông tin phục vụ nghiên cứu thói quen mua sắm của khách hàng (ai mua cái gì ở đâu, khi nào, qua đâu, số tiền bỏ ra, hình thức…) và phân chia các thông tin đó theo từng danh mục khác nhau tương ứng để phục vụ việc thống kê, báo cáo, vẽ biểu đồ, hỗ trợ ra quyết định và đáp ứng các nhu cầu kinh doanh khác.

Chart, bar chart

Description automatically generated

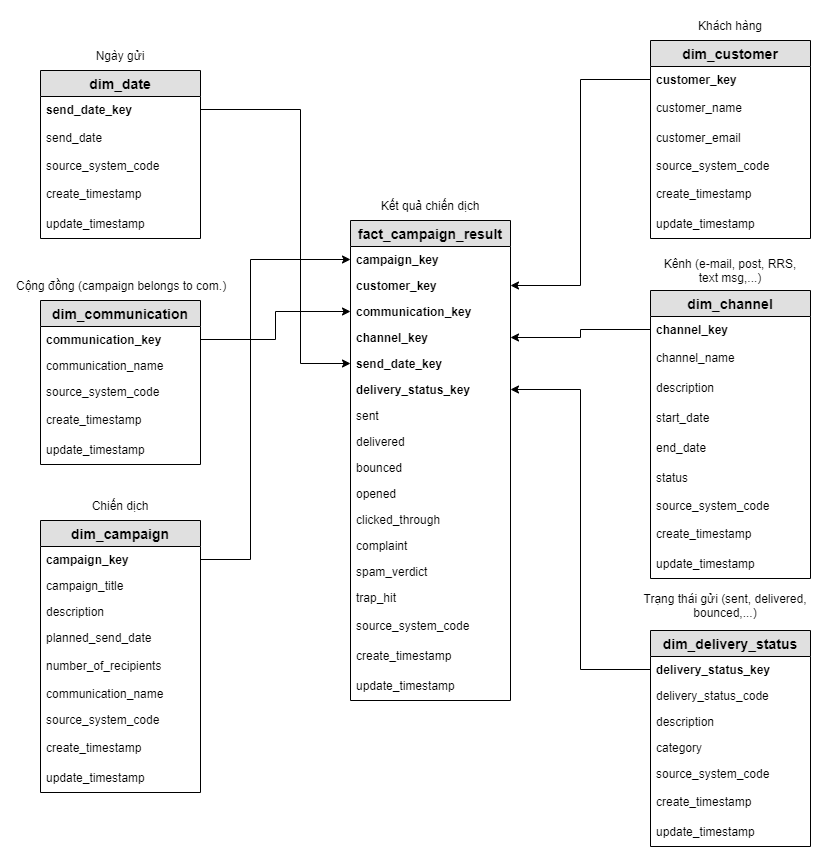
Chart, bubble chart

Description automatically generated

# **Thiết kế**

1. Normalized Data Schema (NDS)

Lược đồ dữ liệu được chuẩn hóa tới chuẩn 3NF để phục vụ việc truy xuất, đọc, ghi nhanh, tránh dư thừa dữ liệu cho các ứng dụng hoạt động, hệ thống OLTP.



1. Dimension Data Schema (DDS)