|  |  |
| --- | --- |
| Funix -DAP305  ID: dungpnFX18614  Name: Phạm Ngọc Dũng |  |

**BÁO CÁO**

**ĐỒ ÁN CUỐI KHÓA – PHÂN TÍCH DỮ LIỆU**

# Giới thiệu về dự án

Công ty DA Company là 1 công ty thương mại có trụ sở tại Vương Quốc Anh, kinh doanh về các loại rượu, đồ uống. Công ty có 80 cửa hàng ở nhiều thành phố khác nhau. Nhóm quản lý muốn tiến hành phân tích hàng tồn kho để xác định các lĩnh vực cần cải thiện và tối ưu hóa các hoạt động quản lý hàng tồn kho của họ. Công ty cũng muốn đánh giá lại các cửa hàng hoạt động không hiệu quả, cắt giảm một số cửa hàng.

Dữ liệu dữ án được lấy từ quá trình kinh doanh của công ty DA Company trong năm 2016

* Nguồn dữ liệu: <https://www.kaggle.com/datasets/bhanupratapbiswas/inventory-analysis-case-study>
* Data bao gồm:

1. Bảng mô tả thương hiệu sản phẩm: giá mua, giá bán…

2017PurchasePricesDec.csv

1. Bảng tổng hợp tồn kho đầu năm 2016:

BegInvFINAL12312016.csv

1. Bảng tổng hợp tồn kho cuối năm 2016:

EndInvFINAL12312016.csv

1. Bảng theo dõi quá trình mua hàng (nhập hàng) trong năm 2016:

PurchasesFINAL12312016.csv

1. Bảng hóa đơn mua hàng năm 2016:

InvoicePurchases12312016.csv

1. Bảng theo dõi quá trình bán hàng năm 2016:

SalesFINAL12312016.csv

# Phân tích mục tiêu, yêu cầu, hướng giải quyết dự án

## Mục tiêu

* Báo cáo kết quả kinh doanh, tình hình mua hàng của công ty trong năm 2016
* Phát triển một chiến lược quản lý hàng tồn kho bền vững cho sự tăng trưởng trong tương lai.

## Yêu cầu

* Tổng hợp doanh thu bán hàng, giá trị nhập hàng của công ty
* Đánh giá khen thưởng các cửa hàng có kết quả kinh doanh tốt
* Phân tích ABC: Phân loại các mặt hàng tồn kho dựa trên giá trị và tầm quan trọng của chúng để ưu tiên các nỗ lực quản lý.
* Phân cụm các mặt hàng theo giá bán và doanh thu

## Hướng giải quyết dự án

* Tìm hiểu thông tin cơ bản từ các bảng dữ liệu bằng Python
* Làm sạch, lựa chọn dữ liệu bằng Python, excel
* Thiết lập mô hình cơ sở dữ liệu quan hệ
* Sử dụng SSIS để ETL dữ liệu lên SQL server
* Truy vấn, trích xuất các dữ liệu cần thiết cho quá trình phân tích trên SQL server
* Trình bày báo cáo bằng Power BI

# Quá trình thực hiện

## Tìm hiểu thông tin cơ bản từ các bảng dữ liệu

* **Bảng tổng hợp tồn kho đầu năm:**

A screenshot of a computer

Description automatically generated

A screenshot of a phone

Description automatically generated

+ Bảng có 206529 dòng, 9 cột ;

+ Các giá trị trong cột InventoryId là độc lập và được kết hợp từ 3 giá trị trong 3 cột Store, City và Brand

+ Cột StartDate chỉ ghi nhận ngày đầu tiên trong năm

+ Cột OnHand thể hiện số lượng tồn kho

+ Cột Price thể hiện mức giá bán

+ Các cột không chứa giá trị null

* **Bảng tổng hợp tồn kho cuối năm**

A screenshot of a computer

Description automatically generatedA screenshot of a phone number

Description automatically generated

+ Bảng có 224489 dòng, 9 cột ;

+ Các giá trị trong cột InventoryId là độc lập và được kết hợp từ 3 giá trị trong 3 cột Store, City và Brand

+ Cột endtDate chỉ ghi nhận ngày cuối cùng trong năm

+ Cột OnHand thể hiện số lượng tồn kho

+ Cột Price thể hiện mức giá bán

+ Các cột không chứa giá trị null

* **Bảng mô tả thương hiệu sản phẩm**

**A screenshot of a computer

Description automatically generated** **A screenshot of a computer

Description automatically generated**

+ Bảng có 12.261 dòng, 9 cột ;

+ Các giá trị trong cột Brand là độc lập

+ Cột Price thể hiện giá bán

+ Cột PurchasePrice thể hiện giá nhập (mua hàng)

+ Cột VendorNumber thể hiện mã số của đơn vị cung cấp hàng

+ Cột VendorName thể hiện tên của công ty cung cấp hàng

+ Các cột không chứa giá trị null

+ Số lượng giá trị độc lập của VendorNumber và VendorName không trùng khớp cần kiểm tra lại

* **Bảng hóa đơn mua hàng:**

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

+ Bảng có 5543 dòng, 10 cột;

+ Cột VendorNumber thể hiện mã số của đơn vị cung cấp hàng

+ Cột VendorName thể hiện tên của công ty cung cấp hàng

+ Số lượng giá trị độc lập của VendorNumber và VendorName không trùng khớp cần kiểm tra lại

+ Cột InvoiceDate thể hiện ngày xuất hóa đơn

+ PONumber thể hiện mã số của mỗi lần đặt hàng

+ PoDate thể hiện ngày đặt hàng

+ Cột Quantity thể hiện số lượng đặt hàng

+ Cột Dollars thể hiện giá trị hóa đơn

+ Cột Freight thể hiện chi phí giao hàng

+ Cột Approval thể hiện người kiểm duyệt

* **Bảng theo dõi quá trình mua hàng (nhập hàng):**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**+** Bảng có 2.372.474 dòng, 16 cột;

+ Cột VendorNumber thể hiện mã số của đơn vị cung cấp hàng

+ Cột VendorName thể hiện tên của công ty cung cấp hàng

+ Số lượng giá trị độc lập của VendorNumber và VendorName không trùng khớp cần kiểm tra lại (giống file hóa đơn).

+ Cột ReceivingDate thể hiện ngày nhận hàng

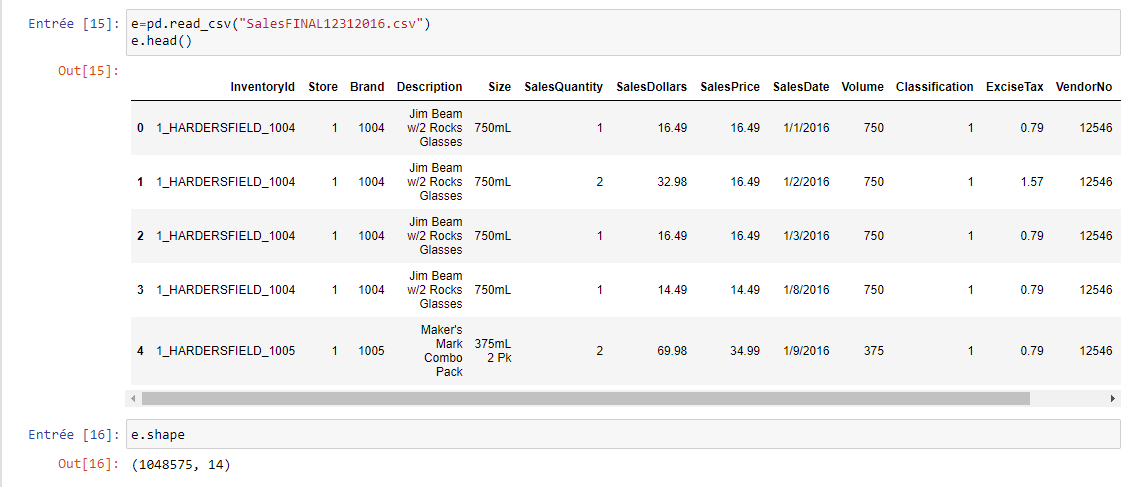
+ PONumber thể hiện mã số của mỗi lần đặt hàng

+ PoDate thể hiện ngày đặt hàng

+ InvoiceDate thể hiện ngày xuất hóa đơn

+ PurchasePrice thể hiện giá nhập

* **Bảng theo dõi quá trình bán hàng:**

**** **A screenshot of a computer

Description automatically generated**

+ Bảng có 1.048.575 dòng, 14 cột;

+ Cột SaleDate thể hiện ngày bán: giá trị lớn nhất là ngày 9-2-2016

* Bảng theo dõi quá trình bán hàng không có ý nghĩa phân tích

## Làm sạch, lựa chọn dữ liệu

1. Bảng mô tả thương hiệu sản phẩm

* Qua quá trình tìm hiểu thông tin ban đầu, chúng ta phát hiện ra rằng số lượng giá trị độc lập của VendorNumber và VendorName không trùng khớp
* Sử dụng công cụ Excel cũng như Python đã phát hiện ra nguyên nhân là do sai sót của các dòng text trong cột VendorName.
* Sử dụng công cụ Excel để thay thế, sửa chữa các lỗi sai về dạng đúng.
* Kết quả có được 1 file Brand.csv đã được làm sạch như sau:

A screenshot of a computer

Description automatically generated

1. Bảng hóa đơn mua hàng

Tương tự với tình trạng và nguyên nhân như ở Bảng mô tả thương hiệu, qua quá trình làm sạch, chúng ta có một bảng hóa đơn mua hàng đẹp đẽ hơn:

A screenshot of a computer

Description automatically generated

1. Bảng theo dõi quá trình mua hàng(nhập hàng)

Bảng theo dõi quá trình mua hàng cũng xảy ra lỗi tương tự với cột VendorName nên khi thiết kế mô hình cơ sở dữ liệu quan hệ, chúng ta sẽ không lựa chọn dữ liệu từ cột này mà lấy từ cột VendorName của bảng hóa đơn mua hàng.

## Thiết lập mô hình cơ sở dữ liệu quan hệ

A computer screen shot of a computer

Description automatically generated

Mô hình được sử dụng: Snowflake

## ETL từ dữ liệu nguồn, tạo lập các bảng trên SQL server

* Tạo lập các bảng để chứa dữ liệu trên SQL server

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

* Sử dụng SSIS để ETL dữ liệu từ các file csv lên SQL server

A screenshot of a computer

Description automatically generated

1. Bảng Vendor\_DIM

A screenshot of a computer

Description automatically generated

* Kết quả:

A white screen with a yellow border

Description automatically generated

1. Bảng Brand\_DIM

A screenshot of a computer

Description automatically generated

* Chỉnh sửa kiểu dữ liệu của các cột trong bảng trên SQL server

A screenshot of a computer code

Description automatically generated

* Kết quả:

A screenshot of a computer

Description automatically generated

1. Bảng InvoicePurchases\_DIM

A screenshot of a computer

Description automatically generated

* Chỉnh sửa kiểu dữ liệu trong các cột của bảng

A screenshot of a computer

Description automatically generated

* Kết quả:

A screenshot of a computer

Description automatically generated

1. Bảng Address\_DIM

A screenshot of a computer

Description automatically generated

* Chỉnh sửa kiểu dữ liệu trong bảng; thêm 3 cột Store, City, Brand

A white background with black text

Description automatically generated

* Kết quả:

A screenshot of a computer

Description automatically generated

1. Bảng BeginInventory\_Fact

A screenshot of a computer

Description automatically generated

* Chỉnh sửa kiểu dữ liệu trong các cột

A screen shot of a computer

Description automatically generated

* Kết quả:

A white screen with a yellow border

Description automatically generated

1. Bảng EndInventory\_Fact

A screenshot of a computer

Description automatically generated

* Chỉnh sửa kiểu dữ liệu trong các cột

A screenshot of a computer code

Description automatically generated

* Kết quả:

A white screen with a yellow border

Description automatically generated

1. Bảng PurchasesFinal\_Fact

A screenshot of a computer

Description automatically generated

* Chỉnh sửa kiểu dữ liệu của các cột

A screenshot of a computer code

Description automatically generated

* Kết quả:

A screen shot of a computer

Description automatically generated

## Truy vấn, tạo thêm một số bảng trên SQL Server

1. Bảng date\_DIM

Tạo một bảng date\_DIM để phân tích quá trình nhập hàng, bán hàng… theo thời gian (giá trị min và max của bảng này được lấy từ giá trị min và max của tất cả các bảng dữ liệu đầu vào).

A screenshot of a computer code

Description automatically generated

A screen shot of a computer

Description automatically generated

1. Tạo bảng Sales

* Do bảng theo dõi quá trình bán hàng SalesFINAL12312016.csv chỉ cung cấp các giao dịch trong tháng 1 và tháng 2 của năm 2016 nên không có ý nghĩa phân tích đối với yêu cầu của dự án này.
* Nhằm khắc phục sự thiếu xót về dữ liệu bán hàng, chúng ta chỉ có thể tạo một bảng tổng hợp kết quả kinh doanh của công ty trong cả năm mà không thể phân tích theo thời gian, dựa vào phương trình kế toán sau:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Trị giá hàng xuất kho trong kỳ | = | Trị giá hàng tồn kho đầu kỳ | + | Trị giá hàng nhập kho trong kỳ | - | Trị giá hàng tồn kho cuối kỳ |

* Chúng ta có 3 bảng hàng tồn kho đầu kỳ, hàng tồn kho cuối kỳ, hàng nhập trong kỳ tương ứng với 3 giá trị trong công thức trên, hoàn toàn có thể tìm ra được lượng bán hàng trong kỳ.

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

* Chúng ta có được một bảng Sales tương ứng với từng loại mặt hàng như sau:

A screenshot of a computer

Description automatically generated

1. Truy vấn, tạo bảng kết quả bán hàng, phân loại hàng tồn kho theo thương hiệu: SaleOnBrand

* Ở dự án này, chúng ta lựa chọn công cụ phân tích là phân tích ABC để phân tích hàng tồn kho theo 3 loại A, B,C để theo dõi và quản lý hàng tồn kho
* Theo Wikipedia, phân tích ABC được định nghĩa như sau:

A screenshot of a computer

Description automatically generated

* Tạo bảng, truy vấn trên SQL server

A screenshot of a computer program

Description automatically generated

A screenshot of a computer code

Description automatically generated

* Kết quả của truy vấn:

Bảng SaleOnBrand sắp xếp các thương hiệu có doanh thu giảm dần, phân loại chúng vào các class tương ứng theo lý thuyết phân tích ABC

A screenshot of a computer

Description automatically generated

1. Truy vấn, tạo bảng kết quả bán hàng, phân loại hàng tồn kho theo thương hiệu: SaleOnStore

* Tạo bảng, truy vấn trên SQL server

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated

* Kết quả của truy vấn:

Bảng SaleOnStore sắp xếp các cửa hàng có doanh thu giảm dần, phân loại chúng vào các class tương ứng theo lý thuyết phân tích ABC

A screen shot of a computer

Description automatically generated

## Trực quan hóa dữ liệu trên Power BI

\*) Import dữ liệu từ SQL server

A screenshot of a computer

Description automatically generated

\*) Thiết lập mô hình cơ sở dữ liệu quan hệ

A screenshot of a computer

Description automatically generated

\*) Tạo các measure cần thiết cho trực quan hóa

A screenshot of a computer

Description automatically generated

\*) Trực quan hóa dữ liệu theo yêu cầu dự án

- Dashboard số 1: kết quả kinh doanh chung của công ty DA trong năm 2016

A screenshot of a computer screen

Description automatically generated

+ Công ty hoạt động tại 69 thành phố, có 80 cửa hàng, 12.261 thương hiệu

+ Tổng doanh thu trong năm 2016 là 455,7 triệu USD

+ Tổng lợi nhuận gộp trong năm 2016 là 142,4 triệu USD

+ Biểu đồ trên cũng thể hiện các thương hiệu được bán tốt nhất, những cửa hàng có doanh thu tốt nhất cũng như bán được nhiều hàng nhất.

+ Tìm ra được 5 cửa hàng có doanh thu tốt nhất để khen thưởng cuối năm: cửa hàng số 34,76,73,50,38.

* Dashboard số 2: Theo dõi quá trình nhập hàng (mua hàng) của công ty DA trong năm 2016

A screenshot of a computer

Description automatically generated

+ Công ty có 131 nhà cung cấp

+ Tổng giá trị nhập trong năm 2016 là 321,9 triệu USD

+ Công ty nhập hàng nhiều nhất vào tháng 7,8 và tháng 11,12

+ Công ty Diago North America Inc là nhà cung cấp lớn nhất của công ty với 51 triệu USD

* Dashboard số 3: Phân loại hàng tồn kho theo thương hiệu

A screenshot of a computer

Description automatically generated

+ Doanh thu của các mặt hàng nhóm A chiếm 70% tổng doanh thu của công ty nhưng số lượng mặt hàng này chỉ chiếm 8,82% tổng số mặt hàng.

+ Doanh thu của các mặt hàng nhóm B chiếm 20% tổng doanh thu của công ty, số lượng mặt hàng này chiếm 14,21% tổng số mặt hàng.

+ Doanh thu của các mặt hàng nhóm C chiếm 10% tổng doanh thu của công ty nhưng số lượng mặt hàng chiếm tới 76,96% tổng số mặt hàng.

* Dashboard số 4: Phân loại hàng tồn kho theo cửa hàng

A screenshot of a computer

Description automatically generated

+ Doanh thu của các cửa hàng nhóm A chiếm 60% tổng doanh thu, số lượng cửa hàng nhóm A là 22 trên tổng số 80 cửa hàng.

+ Doanh thu của các cửa hàng nhóm B chiếm 30% tổng doanh thu, số lượng cửa hàng nhóm B là 32 trên tổng số 80 cửa hàng.

+ Doanh thu của các cửa hàng nhóm C chiếm 10% tổng doanh thu, số lượng cửa hàng nhóm C là 26 trên tổng số 80 cửa hàng.

* Dashboard số 5: Phân cụm thương hiệu theo doanh thu và giá bán

A screen shot of a graph

Description automatically generated

+ Cụm số 1 là các thương hiệu có giá bán thấp và doanh thu thấp.

+ Cụm số 2 là các thương hiệu có doanh thu thấp và trung bình, giá bán trải dài từ trung bình đến cao.

+ Cụm số 3 là các thương hiệu có giá bán trung bình và doanh thu cao

+ Vẫn tồn tại một số thương hiệu có giá bán cao nhưng doanh thu thấp, đây là các mặt hàng xa xỉ nên rất ít khách hàng quan tâm

+ Tệp khách hàng chủ yếu của công ty nằm ở phân khúc giá tầm trung.

# Đặc điểm nổi trội, khó khăn

## Đặc điểm nổi trội

* Dữ liệu của công ty được cập nhật khá lớn tạo nên sự khách quan của dữ liệu: file dữ liệu nhập hàng chứa hơn 2 triệu dòng.

## Khó khăn

* Dữ liệu chưa được làm sạch: các bảng chứa cột VendorName còn sai sót trong các giá trị text
* Dữ liệu thiếu xót, không có ý nghĩa phân tích ở file SalesFinal2016.csv khi chỉ cung cấp dữ liệu bán hàng của công ty trong tháng 1, tháng 2 năm 2016

# Kết luận, giải pháp

## Kết luận

* Tổng doanh thu bán hàng của công ty DA trong năm 2016 là 455,7 triệu USD
* Tổng lợi nhuận gộp của công ty DA trong năm 2016 là 142,4 triệu USD
* Tổng giá trị hàng tồn kho đầu năm là 46,9 triệu USD
* Tổng giá trị hàng tồn kho cuối năm là 55,4 triệu USD
* Tổng giá trị nhập hàng trong năm là 321,9 triệu USD
* 5 cửa hàng có kết quả kinh doanh tốt nhất là cửa hàng số 34,76,73,50,38.
* Qua quá trình phân loại hàng tồn kho, chúng ta có thể thấy doanh thu của công ty trong năm 2016 phụ thuộc nhiều vào một số cửa hàng, thương hiệu chủ yếu.
* Qua quá trình phân cụm, chúng ta nhận ra rằng khách hàng chủ yếu quan tâm đến mặt hàng có giá tầm trung.

## Giải pháp

* Ngừng kinh doanh một số thương hiệu nhóm C không có doanh thu, hoặc doanh thu rất thấp để giảm chi phí nhập hàng, bảo quản.
* Chuyển đổi nhà cung cấp các thương hiệu nhóm C cho các nhà cung cấp thân thiết, có giá trị nhập hàng lớn để giảm chi phí vận chuyển, hưởng chính sách chiết khấu từ nhà cung cấp.
* Tập trung theo dõi chặt chẽ quản lý hàng tồn kho nhóm A vì nhóm này tạo ra tới 70% doanh thu.
* Cân nhắc việc cắt giảm một số cửa hàng hoạt động không hiệu quả.
* Các loại mặt hàng giá cao tạo ra doanh thu rất nhỏ nên cắt giảm kinh doanh các loại mặt hàng này, tập trung vào phân khúc giá tầm trung.