



## Bài-tập-1of4- -Kết-nối-dữ-liệu

Xác xuất thống kê (Trường Đại học Văn Lang)

# Bài tập 1: Kết nối và Xử lý Data

*Sử dụng bộ file thực hành để hoàn thành các nhiệm vụ sau:*

## 1) Kết nối dữ liệu từ file **Product\_Categories** và **Product\_Subcategories** từ bộ file

**thực hành:**

- Đặt tên queries là **Product\_Category\_Lookup** và **Product\_Subcategory\_Lookup**  
(Data -> chọn tên cần đổi -> dấu “...” -> rename ; or nhấn chuột trái 2 lần)
- Đảm bảo rằng tất cả các trường đều có header và loại dữ liệu trong các trường dữ liệu là chuẩn xác (Xem loại dữ liệu, ví dụ cột đó có phải ngày hay chưa)  
(Transform -> Transform (trang mới) -> data type)
- Disable chức năng refresh cho cả 2 bảng dữ liệu (cập nhật dữ liệu trên Excel nhưng không cho chạy lên BI)  
(Transform -> cột trái -> chọn bảng -> chuột phải -> bỏ tích ở include in report refresh)

## 2) Tùy biến dữ liệu trong Query Editor cho **Product\_Lookup**:

- Thêm một calculated column để chiết xuất ra toàn bộ ký tự trước dấu gạch ngang (“-”) trong trường (cột) **ProductSKU**, đặt tên là “**SKUType**” (Tạo cột cạnh bên)  
(Transform -> bôi đen cột cần -> add column -> extract -> Text before delimiter -> đánh dấu “-” -> Ok)
- Cập nhật tính toán cho trường **SKUType** để chiết xuất các ký tự trước dấu gạch ngang thứ 2, thay vì dấu gạch ngang đầu tiên  
(Transform -> bôi đen cột cần -> add column -> extract -> Text before delimiter -> đánh dấu “-” và kéo xuống đánh số 1 -> OK)

- Thay thế các giá trị 0 trong trường **ProductStyle** bằng “NA”  
Downloaded by Mai Nguyen?n th? (maidg1302@gmail.com)

(Chọn ProductStyle -> chuột phải -> chọn replace value -> VALUE TO FIND: 0 và

REPLACE WITH : NA -> OK)

- Bổ sung trường dữ liệu **DiscountPrice** để tính giá khuyến mại 12%, bằng cách nhân **ProductPrice** với 0.88

### 3) Sử dụng Statistics tools trong Query Editor, thực hiện các nhiệm vụ sau:

- Tính chi phí sản phẩm trung bình product cost (Kết quả: *tự so sánh lẫn nhau*)
- Đếm số màu sản phẩm (distinct product colors) (Kết quả: *tự so sánh lẫn nhau*)
- Đếm số tên khách hàng (distinct customer names) (*tự tải file cần thiết*) (Kết quả: *tự so sánh lẫn nhau*)

Tải file Sale\_2017

- Thu nhập lớn nhất của một khách hàng (Kết quả: *tự so sánh lẫn nhau*)
- Đếm số đặt hàng (Kết quả: *tự so sánh lẫn nhau*)
- Đếm số đặt hàng khác nhau (Kết quả: *tự so sánh lẫn nhau*)

### 4) Tùy biến dữ liệu trong Query Editor cho Customer\_Lookup

- Thêm một calculated column để tính năm sinh (đặt tên là "**BirthYear**"), dựa trên dữ liệu từ trường **BirthDate**
- Thêm một conditional column để phân loại khách hàng theo thu nhập (đặt tên là "**IncomeLevel**"), dựa trên các điều kiện sau:
  - Nếu **AnnualIncome** >= \$150,000, thì **IncomeLevel** = "*Very High*"
  - Nếu **AnnualIncome** >= \$100,000, thì **IncomeLevel** = "*High*"
  - Nếu **AnnualIncome** >= \$50,000, thì **IncomeLevel** = "*Average*"
  - Các trường hợp còn lại **IncomeLevel** = "*Low*"

**5) Apply các thay đổi, và đảm bảo các trường và các bảng đều có trong cả Data view và Relationships view. Lưu lại file PBI./.**