

PART 1: Connecting & Shaping the Data

Mở 01 file Power BI Desktop mới, và hoàn thành các bước sau đây:

1) Cập nhật Power BI options và cài đặt như dưới đây:

- Bỏ chọn "Autodetect new relationships after data is loaded" trong tab **Data Load**
- Đảm bảo Locale for import được cài là "English (United States)" trong tab **Regional Settings**

2) Connect tới file **MavenMarket_Customers** (loại file: csv) trong folder **Tài nguyên** (lưu ý: giải nén file zip Tài nguyên)

- Đổi tên bảng là "**Customers**", và đảm bảo có headers cho các trường dữ liệu
- Đảm bảo loại data ở mỗi trường là chuẩn xác (**Note:** "customer_id" là dạng whole numbers, "customer_acct_num" và "customer_postal_code" là dạng text ..., kiểm tra lại các trường khác cho phù hợp)
- Thêm cột mới tên là "full_name" để gộp 2 cột "first_name" và "last_name", ngăn cách bằng dấu space (dấu cách)
- Tạo cột mới tên là "birth_year" để extract dữ liệu là năm (year) từ cột "birthdate", định dạng là text
- Tạo một **conditional column** tên là "has_children", kết quả là "**N**" nếu "total_children" = 0, còn lại là "**Y**"

3) Connect tới file **MavenMarket_Products** (loại file: csv) trong folder **Tài nguyên** (lưu ý: giải nén file zip Tài nguyên)

- Đổi tên bảng là "**Products**" và đảm bảo có headers cho các trường dữ liệu
- Đảm bảo loại data ở mỗi trường là chuẩn xác (**Note:** "product_id" là dạng whole numbers, "product_sku" là dạng text, "product_retail_price" và "product_cost" là dạng decimal numbers)

- Sử dụng Statistics tools để lấy ra số loại sản phẩm của các brands, dựa trên các product names khác nhau (count distinct value)
- **Check:** bạn sẽ thấy kết quả là **111 brands** và **1,560 product names**
- Thêm một cột tên là "discount_price", bằng 90% của original retail price
- Định dạng cột vừa tạo dạng decimal number, và sử dụng rounding tool để làm tròn 2 chữ số ở phần thập phân
- Chọn cột "product_brand" và sử dụng tùy chọn **Group By** option để tính giá trị trung bình của retail price theo brand, và đặt tên cột là "Avg Retail Price"
- **Check:** bạn sẽ thấy kết quả giá retail price trung bình là **\$2.18** cho Washington products, và **\$2.21** cho Green Ribbon
- Xóa step cuối cùng để trả về kết quả là bảng trước khi **Group By**
- Thay thế các giá trị "null" bằng 0 trong cột "recyclable" và "low-fat"

4) Connect tới file **MavenMarket_Stores** (loại file: csv) trong folder **Tài nguyên** (lưu ý: giải nén file zip Tài nguyên)

- Đổi tên bảng là "**Stores**" và đảm bảo có headers cho các trường dữ liệu
- Đảm bảo loại data ở mỗi trường là chuẩn xác (**Note:** "store_id" và "region_id" là dạng whole numbers)
- Thêm cột mới tên là "full_address", được gộp bởi "store_city", "store_state", và "store_country", ngăn cách với nhau bởi dấu phẩy và dấu space (**note:** sử dụng custom separator)
- Thêm cột mới tên là "area_code", bằng cách extract các ký tự trước dấu ("-") trong trường "store_phone"

5) Connect tới file **MavenMarket_Regions** (loại file: csv) trong folder **Tài nguyên** (lưu ý: giải nén file zip Tài nguyên)

- Đổi tên bảng là "**Regions**" và đảm bảo có headers cho các trường dữ liệu
- Đảm bảo loại data ở mỗi trường là chuẩn xác (**Note:** "region_id" là dạng whole numbers)

6) Connect tới file **MavenMarket_Calendar (loại file: csv) trong folder **Tài nguyên** (lưu ý: giải nén file zip Tài nguyên)**

- Đổi tên bảng là "**Calendar**" và đảm bảo có headers cho các trường dữ liệu
- Sử dụng date tools trong query editor để thêm các cột:
- *Start of Week (starting Sunday)*
- *Name of Day*
- *Start of Month*
- *Name of Month*
- *Quarter of Year*
- *Year*

7) Connect tới file **MavenMarket>Returns (loại file: csv) trong folder **Tài nguyên** (lưu ý: giải nén file zip Tài nguyên)**

- Đổi tên bảng là "**Return_Data**" và đảm bảo có headers cho các trường dữ liệu
- Đảm bảo loại data ở mỗi trường là chuẩn xác (các cột ID *quantity* là dạng whole numbers)

8) Thêm một folder mới trên desktop (hoặc trong documents) tên là "MavenMarket Transactions**", bao gồm 2 file **MavenMarket_Transactions_1997** và **MavenMarket_Transactions_1998** (copy từ folder **Tài nguyên**)**

- Kết nối folder với PowerBI, và chọn "Edit" (hoặc Transform Data)
- Click vào header cột "*Content*" (double vào icon) và combine 02 file này, sau đó remove cột "*Source.Name*"
- Đặt tên bảng là "**Transaction_Data**", và đảm bảo có headers cho các trường dữ liệu
- Đảm bảo loại data ở mỗi trường là chuẩn xác (các cột ID *quantity* là dạng whole numbers)
- **Check:** *Bạn sẽ thấy data từ 1/1/1997 tới 12/30/1998 trong cột "transaction_date"*

9) Ngoại trừ 2 bảng loại data tables (tự phân biệt đâu là data tables, đâu là lookup table), disable tính năng "*Include in Report Refresh*", sau đó **Close & Apply**

- Đảm bảo rằng có tất cả 7 bảng dữ liệu được kết nối cả trong **RELATIONSHIPS** view và **DATA** view

10) Lưu file với định dạng .pbix (gợi ý lưu với tên "***MavenMarket_Report***")

Kết quả (tham khảo):

The screenshot displays the Microsoft Power BI Desktop interface. The main view shows a data table with 6 columns and 999+ rows. The columns are: transaction_date, stock_date, product_id, customer_id, store_id, and quantity. The right sidebar shows the 'QUERY SETTINGS' pane with 'Removed Columns' selected under 'APPLIED STEPS'.

| | transaction_date | stock_date | product_id | customer_id | store_id | qu |
|----|------------------|------------|------------|-------------|----------|----|
| 1 | 1/1/1997 | 12/31/1996 | 869 | 3449 | 6 | |
| 2 | 1/1/1997 | 12/31/1996 | 1472 | 3449 | 6 | |
| 3 | 1/1/1997 | 12/28/1996 | 76 | 3449 | 6 | |
| 4 | 1/1/1997 | 12/26/1996 | 320 | 3449 | 6 | |
| 5 | 1/1/1997 | 12/25/1996 | 4 | 3449 | 6 | |
| 6 | 1/1/1997 | 12/30/1996 | 952 | 3449 | 6 | |
| 7 | 1/1/1997 | 12/31/1996 | 1222 | 3449 | 6 | |
| 8 | 1/1/1997 | 12/30/1996 | 517 | 7859 | 6 | |
| 9 | 1/1/1997 | 12/27/1996 | 1359 | 7859 | 6 | |
| 10 | 1/1/1997 | 12/31/1996 | 357 | 106 | 6 | |
| 11 | 1/1/1997 | 12/27/1996 | 1426 | 106 | 6 | |
| 12 | 1/1/1997 | 12/26/1996 | 190 | 106 | 6 | |
| 13 | 1/1/1997 | 12/28/1996 | 367 | 106 | 6 | |
| 14 | 1/1/1997 | 12/29/1996 | 250 | 106 | 6 | |
| 15 | 1/1/1997 | 12/26/1996 | 600 | 106 | 6 | |
| 16 | 1/1/1997 | 12/29/1996 | 702 | 8248 | 6 | |
| 17 | 1/1/1997 | 12/27/1996 | 786 | 8248 | 6 | |
| 18 | 1/1/1997 | 12/30/1996 | 536 | 8248 | 6 | |
| 19 | 1/1/1997 | 12/27/1996 | 1285 | 8248 | 6 | |
| 20 | 1/1/1997 | 12/29/1996 | 596 | 8248 | 6 | |
| 21 | 1/1/1997 | 12/27/1996 | 846 | 1906 | 6 | |
| 22 | 1/1/1997 | 12/31/1996 | 769 | 1906 | 6 | |
| 23 | 1/1/1997 | 12/26/1996 | 586 | 1906 | 6 | |
| 24 | 1/1/1997 | 12/28/1996 | 1135 | 1906 | 6 | |
| 25 | 1/1/1997 | 12/25/1996 | 1046 | 1906 | 6 | |
| 26 | 1/1/1997 | 12/25/1996 | 170 | 1906 | 6 | |