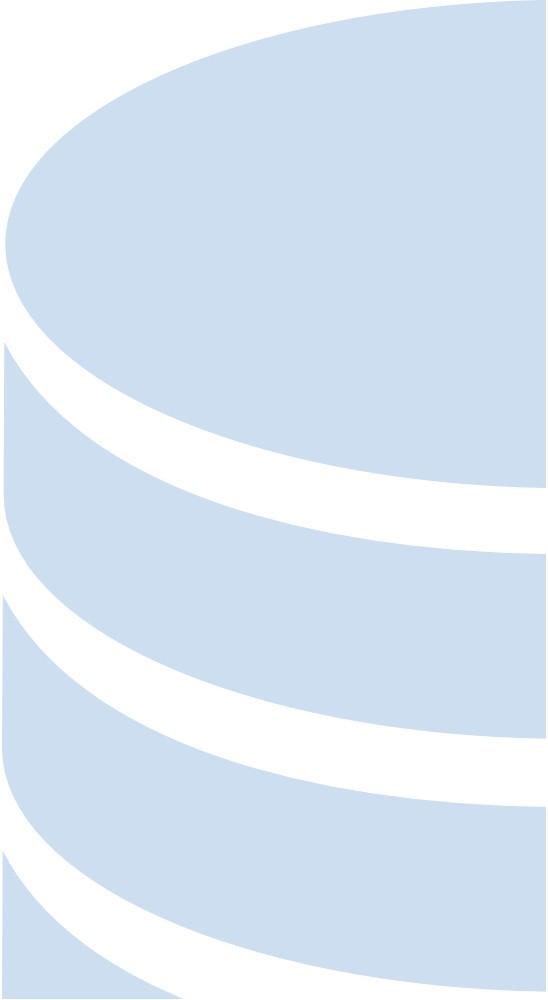
**BÁO CÁO DỰ ÁN CUỐI KHÓA  
PHÂN TÍCH DỮ LIỆU**

**“Phân tích tình hình kinh doanh: So sánh hiệu suất bán hàng trên kênh Reseller và kênh Online của AdventureWorks”**

**Mentor: Công Trần**

**Học Viên: Trần Thanh Tuyền FX18918**

*Hà Nội - 2024*

# LỜI MỞ ĐẦU

Ngày nay có thể nói dữ liệu là một mỏ vàng của thế giới, do đó ngày càng có nhiều doanh nghiệp quan tâm đến việc khai thác, tạo ra giá trị từ dữ liệu. Business Intelligence (BI - Kinh doanh thông minh) là thuật ngữ để thể hiện công việc khai thác dữ liệu mà doanh nghiệp cần để chúng ta có thể phân tích và hình dung dữ liệu, từ đó có được cái nhìn sâu sắc về cách doanh nghiệp hoạt động, giúp doanh nghiệp nắm được những hiểu biết sâu sắc về chiến lược, chiến thuật, vận hành và ra quyết định.

Với định hướng nghề nghiệp trở thành một Data Analyst của các công ty về kinh doanh, sản xuất, bán lẻ vì vậy tôi đã chọn chủ đề phân tích tình hình kinh doanh của công ty AdventureWorks. Bộ dữ liệu AdventureWorks 2019 là một cơ sở dữ liệu mẫu được Microsoft cung cấp để giúp người dùng học cách làm việc với SQL Server và các công nghệ liên quan. Cơ sở dữ liệu này cung cấp các bảng dữ liệu về doanh nghiệp giả định, bao gồm thông tin về khách hàng, sản phẩm, đơn đặt hàng, và nhiều hơn nữa.

**Bản báo cáo dự án cuối khóa được bố cục theo 3 phần chính:**

| **Phần 1 - Tổng quan** | Phần mở đầu của dự án, bao gồm những thông tin khái quát như giới thiệu về công ty, giới thiệu về dự án, một số các yêu cầu của dự án, định hướng các bước để hoàn thành dự án. |
| --- | --- |
| **Phần 2 - Chi Tiết Dự án** | Mô tả chi tiết quá trình đã thực hiện từng bước để hoàn thành dự án. |
| **Phần 3 - Kết Luận** | Tổng kết lại quá trình làm việc và tổng hợp lại những kiến thức, các đặc điểm về sản phẩm, một số khó khăn trong quá trình làm dự án. |

# PHẦN 1 - TỔNG QUAN

## Chương 1 – Giới thiệu

### Giới thiệu công ty:

Công ty AdventureWorks – một công ty chuyên sản xuất xe đạp với nhiều nhãn hiệu xe đạp khác nhau được chia thành 3 nhóm chính là xe đạp đường trường, xe đạp du lịch và xe đạp leo núi. Ngoài ra công ty còn sản xuất và kinh doanh linh kiện, phụ kiện, quần áo khác. Công ty hoạt động tại Canada, Hoa Kỳ, Úc và 1 số nước Châu Âu với hai kênh phân phối là Reseller và Online.

### Giới Thiệu Dự án:

Với mục tiêu cải thiện tình hình kinh doanh trong năm 2014, ban lãnh đạo của công ty AdventureWork đã giao nhiệm vụ xuống phòng Phân tích dữ liệu của công ty và tôi là người được phân công phụ trách xây dựng 1 sản phẩm phân tích tình hình kinh doanh trên hai kênh phân phối là Reseller và Online, dựa trên nguồn dữ liệu của công ty trong 2 năm 2012 & 2013.

Dự án bao gồm những phần chính sau:

* Xây dựng Data WareHouse.
* Xây dựng Dashboard - Báo cáo tự động.

### Mục Tiêu của Dự án:

Mục tiêu của dự án này là chuyển hóa dữ liệu trong quá khứ thành thông tin hữu ích, để công ty có cái nhìn sâu sắc về cách mà họ đã và đang hoạt động. Từ đó có thể đưa ra các quyết định dựa trên dữ liệu nhằm cải thiện doanh thu, lợi nhuận cho công ty.

1. Tạo các báo cáo cho các đối tượng từ cấp quản lý đến nhân viên trong công ty có thể theo dõi tiến độ công việc, sản phẩm, doanh thu, lợi nhuận trên 2 kênh bán hàng trực tuyến và bán lẻ… Hỗ trợ họ đưa ra các quyết định tùy theo tình hình thực tế.
2. Phát hiện ra các vấn đề kinh doanh quá khứ đến hiện tại trong doanh nghiệp thông qua sản phẩm, khách hàng, khu vực, vận chuyển, hệ thống cửa hàng…
3. Đưa ra dự đoán về các chỉ số kinh doanh của công ty.
4. Đề xuất giải pháp cải thiện tình hình kinh doanh của doanh nghiệp.

## Chương 2 - Yêu cầu của dự án

### Yêu cầu đối với DataWareHouse:

* 1. Yêu cầu kinh doanh:
     1. Xác định và hiểu rõ các mục tiêu và yêu cầu kinh doanh.
     2. Đảm bảo rằng dữ liệu trong data warehouse đáp ứng được nhu cầu phân tích và báo cáo của doanh nghiệp.
  2. Dữ liệu đồng nhất và chính xác:
     1. Đảm bảo rằng dữ liệu trong data warehouse được chuẩn hóa và chính xác.
     2. Định nghĩa và duy trì các quy trình để xử lý dữ liệu đảm bảo tính nhất quán và độ tin cậy.
  3. Hiệu suất:
     1. Tối ưu hóa cấu trúc dữ liệu và truy vấn để đảm bảo hiệu suất tốt.
  4. Mở rộng và linh hoạt:
     1. Thiết kế data warehouse để có thể mở rộng dễ dàng khi nhu cầu phát triển.

### Yêu cầu đối với báo cáo

* 1. Bố cục:
     1. Báo cáo bao gồm 3 phần, 1 phần dành cho cấp quản lý, 1 phần cho bộ phận online, 1 phần cho bộ phận reseller.
  2. Chính xác và đáng tin cậy:
     1. Báo cáo cần cung cấp thông tin chính xác và đáng tin cậy để người dùng có thể dựa vào để ra quyết định.
     2. Dữ liệu được sử dụng trong báo cáo cần được kiểm tra và xác thực để đảm bảo tính nhất quán.
  3. Dễ đọc và dễ hiểu:
     1. Giao diện báo cáo cần được thiết kế sao cho dễ đọc và dễ hiểu cho người dùng cuối.
     2. Sử dụng biểu đồ và bảng để minh họa dữ liệu một cách trực quan và rõ ràng.
  4. Tính linh hoạt và tương tác:
     1. Báo cáo nên cho phép người dùng tương tác để phân tích dữ liệu theo các góc độ khác nhau.
     2. Cung cấp các bộ lọc, slicers và điều kiện để người dùng có thể tùy chỉnh báo cáo theo nhu cầu của họ.
  5. Bảo mật và tuân thủ:
     1. Đảm bảo rằng báo cáo tuân thủ các quy định bảo mật và quyền riêng tư.

## Chương 3 - Hướng giải quyết dự án

1. **Dựa vào mục tiêu của dự án chúng ta cần xác định rõ các vấn đề sau:** 
   1. Bài toán chúng ta cần giải quyết là gì?
   2. Các bên liên quan là ai? Những ai sẽ là người dùng cuối sản phẩm khi chúng ta hoàn thành?
2. **Đặt ra các câu hỏi nghiệp vụ, xác định các thước đo để định hướng quá trình thu thập dữ liệu.**
3. **Thu thập dữ liệu**
   1. Chúng ta cần xác định các nguồn có thể thu thập dữ liệu để trả lời các câu hỏi nghiệp vụ.
   2. Chúng ta cần lập 1 kế hoạch thu thập dữ liệu rõ ràng.
   3. Người liên hệ là ai?
   4. Tính bảo mật và quyền riêng tư của dữ liệu như thế nào?
4. **Khám phá dữ liệu**
   1. Để có thể chọn ra các dữ liệu cần thiết phục vụ cho dự án, việc đầu tiên chúng ra cần làm là phải hiểu ý nghĩa của các bảng trong bộ dữ liệu nguồn sau đó chúng ta sẽ chọn ra các bảng chứa nội dung cần thiết cho việc phân tích, việc tìm hiểu dữ liệu từ nhiều nguồn có thể sẽ tốn rất nhiều công sức, tuy nhiên chúng ta phải xác định dữ liệu sẽ là nền tảng quyết định cho sự thành công của 1 dự án. Vì vậy phải khám phá từng bảng, từng cột dữ liệu, từng nội dung, thậm chí là từng ký hiệu.. xem chúng ta có những thông tin gì có thể khai thác phục vụ cho dự án.
5. **Thiết kế ERD**
   1. Khi đã xác định được các bảng và các cột chứa dữ liệu cần thiết cho việc phân tích, chúng ta cần vẽ 1 bản thiết kế ERD để tổng hợp lại trước khi thực hiện quá trình ETL xây dựng Data WareHouse.
6. **Xây dựng Data WareHouse**
   1. Dựa vào các yêu cầu đã nêu ở bên trên để thực hiện xây dựng Data WareHouse.
7. **Kiểm tra lại Data WareHouse**
   1. Sau khi đã xây dựng xong kho dữ liệu, chúng ta kiểm tra lại bằng cách truy vấn vào cơ sở dữ liệu dựa theo 1 số yêu cầu nghiệp vụ trước khi import dữ liệu vào các ứng dụng xây dựng báo cáo.
8. **Xây dựng báo cáo**
   1. Xây dựng báo cáo dựa theo yêu cầu của báo cáo đã nêu ở trên.
9. **Đánh giá kết quả**
   1. Sau khi xây dựng xong báo cáo chúng ta cần đánh giá lại các điểm được và chưa được, sau đó có thể chúng ta sẽ phải thực hiện lặp lại từ bước thu thập dữ liệu cho đến bước này để hoàn thiện sản phẩm.
10. **Trình bày kết quả**
    1. Chuẩn bị bài thuyết trình để trình bày sản phẩm với các bên liên quan.

# PHẦN 2 - CHI TIẾT DỰ ÁN

## Chương 4 – Triển khai dự án

### Xác định vấn đề cần giải quyết:

* 1. Bài toán chúng ta cần giải quyết là gì?
     1. Phân tích tình hình kinh doanh trên 2 kênh bán lẻ Reseller và Online của công ty.
     2. Đưa ra dự đoán các chỉ số kinh doanh trong năm 2014 và đề xuất các biện pháp để cải thiện doanh số và lợi nhuận cho công ty.
  2. Các bên liên quan là ai? Những ai sẽ là người dùng cuối sản phẩm khi chúng ta hoàn thành?
     1. Giám đốc điều hành, Giám đốc kinh doanh.
     2. Nhân viên các bộ phận của kênh phân phối Reseller và Online.

### Đặt ra các câu hỏi nghiệp vụ, xác định các thước đo để định hướng quá trình thu thập dữ liệu.

* Để hiểu cách 1 doanh nghiệp đang hoạt động như thế nào chúng ta cần trả lời 4 câu hỏi sau:
* Bán cái gì?
* Bán cho ai?
* Bán như thế nào?
* Doanh thu, lợi nhuận là bao nhiêu?
* Vì vậy dù là cấp quản lý hay nhân viên thì cách để họ hiểu về công ty, hay bộ phận của mình đang hoạt động thế nào cũng sẽ xoay quanh những câu hỏi trên, chỉ khác nhau về chỉ số kinh doanh.
* Đối với cấp quản lý họ sẽ quan tâm đến các con số ngắn gọn mà thông qua đó họ nhìn thấy bức tranh toàn cảnh của công ty, qua đó họ có thể có những chỉ đạo phù hợp từng thời điểm.
* Đối với nhân viên các bộ phận họ sẽ quan tâm nhiều hơn đến chi tiết, tiến độ để họ có thể điều chỉnh kịp thời nhằm đạt được mục tiêu kinh doanh KPI.
* Bán cái gì? tương đương với sản phẩm.
* Bán cho ai? tương đương với khách hàng.
* Bán như thế nào? tương đương với thị trường, các kênh phân phối, hệ thống cửa hàng…
  1. **Giám đốc điều hành, giám đốc kinh doanh quan tâm tới điều gì?**
     1. Sản phẩm
        1. Công ty đang có những phân loại sản phẩm nào? mặt hàng nào đem lại doanh thu cao nhất?
        2. Sản phẩm nào đang bán chạy nhất?
        3. Sản phẩm nào đang đem lại lợi nhuận nhiều nhất cho công ty?
     2. Khách hàng
        1. Công ty đang có bao nhiêu khách hàng?
        2. Có thêm bao nhiêu khách hàng mới theo tháng và theo năm?
        3. Có bao nhiêu khách hàng cũ từ năm trước vẫn trung thành với công ty?
        4. Doanh thu của các nhóm khách hàng này là bao nhiêu?
     3. Thị trường
        1. Công ty đang hoạt động tại thị trường nào?
        2. Thị trường nào đem lại doanh thu tốt nhất cho công ty?
     4. Kênh phân phối
        1. Có những kênh phân phối nào?doanh thu, lợi nhuận trên từng kênh phân phối là bao nhiêu?
        2. Doanh thu, lợi nhuận tăng trưởng trong các năm của 2 kênh phân phối?
        3. Biên lợi nhuận của từng kênh là bao nhiêu %?
     5. Doanh thu, lợi nhuận
        1. Tổng doanh thu, lợi nhuận của công ty là bao nhiêu?
        2. Có đạt được KPI hàng tháng hay không?
  2. **Cán bộ nhân viên của kênh bán lẻ Reseller quan tâm tới điều gì?**
     1. Sản phẩm
        1. Reseller đang có những loại mặt hàng nào? mặt hàng nào đem lại doanh thu cao nhất?
        2. Các đặc điểm sản phẩm nào đang bán được nhiều nhất?
* Color?
* ProductLine?
* Style?
  + - 1. Đâu là những sản phẩm tiềm năng theo từng khu vực?
    1. Khách hàng
       1. Hàng tháng có bao nhiêu khách hàng mới? có bằng tháng trước hay không?
       2. Phân khúc khách hàng của Reseller? Chất lượng của từng phân khúc?
    2. Thị trường và hệ thống cửa hàng
       1. Reseller đang hoạt động mạnh nhất tại thị trường nào? đem lại doanh thu là bao nhiêu?
       2. Có bao nhiêu cửa hàng tại các thị trường công ty đang hoạt động?
       3. Tình hình kinh doanh tại các cửa hàng?
    3. Doanh thu, lợi nhuận
       1. Doanh thu lợi nhuận tháng này đang là bao nhiêu? Đã sắp đủ KPI hay chưa?
       2. Doanh thu lợi nhuận tăng trưởng theo thời gian?
  1. **Nhân viên các bộ phận của kênh bán lẻ Online quan tâm tới điều gì?**
     1. Sản phẩm
        1. Online đang có những loại mặt hàng nào? mặt hàng nào, sản phẩm nào đem lại doanh thu cao nhất?
        2. Các đặc điểm sản phẩm nào đang bán được nhiều nhất?
* Màu sắc?
* ProductLine?
* Style?
  + - 1. Đâu là những sản phẩm tiềm năng theo từng khu vực?
    1. Khách hàng
       1. Hàng tháng có bao nhiêu khách hàng mới? có đạt được mục tiêu đề ra hay không?
       2. Phân khúc khách hàng của Online? Chất lượng của từng phân khúc?
    2. Thị trường
       1. Online đang hoạt động mạnh nhất tại thị trường nào? đem lại doanh thu là bao nhiêu?
    3. Doanh thu, lợi nhuận
       1. Doanh thu lợi nhuận tháng này đang là bao nhiêu? Đã sắp đủ KPI hay chưa?
       2. Doanh thu lợi nhuận tăng trưởng theo thời gian?

### Xác định các thước đo

* + 1. Số lượng sản phẩm, các loại mặt hàng của công ty.
    2. Số lượng khách hàng của công ty.
    3. Số lượng khách hàng trên từng kênh.
    4. Số lượng khách hàng của từng phân khúc.
    5. Số lượng thị trường công ty đang hoạt động.
    6. Số lượng kênh phân phối.
    7. Số lượng cửa hàng.
    8. Tổng doanh thu của công ty, doanh thu của từng kênh.
    9. Tổng lợi nhuận của công ty, lợi nhuận của từng kênh.
    10. Biên lợi nhuận của công ty, biên lợi nhuận trên từng kênh.

### Thu thập dữ liệu

* 1. Mục tiêu của dự án này là phân tích tình hình kinh doanh của công ty trên 2 kênh phân phối Reseller và Online vì vậy chúng ta xác định nguồn dữ liệu chúng ta cần thu thập đó là dữ liệu chính của AdventureWorks(Dataset AdventureWork 2019 của Microsoft) đã lưu trữ trong 2 năm 2012 và 2013.
  2. Chúng ta không thu thập thêm nguồn dữ liệu thứ cấp hoặc bên thứ 3 lý do liên quan đến yêu cầu kinh doanh của dự án. Trong phạm vi dự án này, chúng ta chỉ xem xét các vấn đề liên quan đến nội bộ công ty, chứ không nghiên cứu thêm về thị trường hoặc đối thủ cạnh tranh, vì vậy việc thu thập thêm dữ liệu là không cần thiết.
  3. Nguồn dữ liệu: [AdventureWorks sample databases - SQL Server | Microsoft Learn](https://learn.microsoft.com/en-us/sql/samples/adventureworks-install-configure?view=sql-server-ver16&tabs=ssms)

### Khám phá dữ liệu

Dataset của công ty AW là cơ sở dữ liệu quan hệ được xây dựng trên nền tảng SSMS thông qua ngôn ngữ SQL.

Vì vậy để khám phá dữ liệu chúng ta sẽ thực hiện các bước như sau:

* 1. Sử dụng SSMS back up lại file AdventureWorks 2019 chứa dữ liệu nguồn. Sau khi back up xong chúng ta có CSDL AdventureWorks 2019.
  2. Truy vấn vào CSDL AdventureWorks 2019 để khám phá cấu trúc, các bảng biểu, các ràng buộc..(VD: select from where, inner join, left join, group by, order by..). (Cụ thể các câu truy vấn lưu tại file Phụ Lục.)
  3. Tìm kiếm các tài liệu tham khảo để hiểu rõ hơn về ý nghĩa các dữ liệu lưu trữ tại các bảng, các cột.

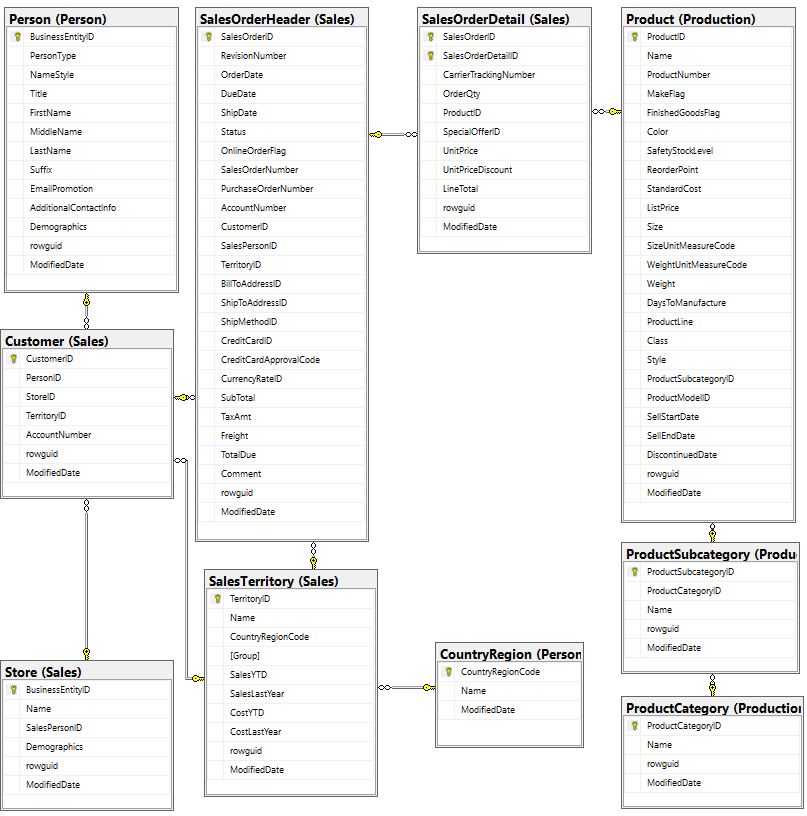
Link tài liệu tham khảo:

1. [AdventureWorks Data Dictionary](https://learn.microsoft.com/en-us/previous-versions/sql/sql-server-2008/ms124438(v=sql.100)?redirectedfrom=MSDN)
2. [AdventureWorks – Data Dictionary](https://dataedo.com/download/AdventureWorks.pdf)
   1. Hiểu về Data source, xác định các bảng chứa thông tin phục vụ cho dự án.

| **STT** | **Schemas** | **Table** | **Nội dung** | **Table Selection** |
| --- | --- | --- | --- | --- |
| 1 | **Human  Resources** | Department | Bảng này chứa thông tin về các phòng ban trong tổ chức như tên phòng ban, mô tả, v.v. | Không |
| 2 | **Human  Resources** | Employee | Bảng này lưu trữ thông tin về các nhân viên trong tổ chức như tên, địa chỉ, ngày sinh, v.v. | Không |
| 3 | **Human  Resources** | EmployeeDepartmentHistory | Bảng này ghi lại lịch sử phòng ban của mỗi nhân viên, bao gồm thông tin như thời gian bắt đầu và kết thúc làm việc ở mỗi phòng ban. | Không |
| 4 | **Human  Resources** | EmployeePayHistory | Bảng này chứa thông tin về lịch sử lương của các nhân viên, bao gồm mức lương, ngày thay đổi, v.v. | Không |
| 5 | **Human  Resources** | JobCandidate | Bảng này lưu trữ thông tin về các ứng viên cho các vị trí công việc trong tổ chức. | Không |
| 6 | **Human  Resources** | Shift | Bảng này chứa thông tin về các ca làm việc, bao gồm thời gian bắt đầu, kết thúc và mô tả. | Không |
| 7 | **Person** | Address | Bảng này lưu trữ thông tin về địa chỉ, bao gồm địa chỉ, thành phố, quận huyện, v.v. | Không |
| 8 | **Person** | AddressType | Bảng này chứa các loại địa chỉ, ví dụ: "Home", "Work", "Shipping", v.v. | Không |
| 9 | **Person** | BusinessEntity | Bảng này chứa thông tin cơ bản về các thực thể kinh doanh, chẳng hạn như ID và ngày tạo. | Không |
| 10 | **Person** | BusinessEntity  Address | Bảng liên kết giữa các thực thể kinh doanh và địa chỉ, để biểu diễn mối quan hệ giữa chúng. | Không |
| 11 | **Person** | BusinessEntity  Contact | Bảng liên kết giữa các thực thể kinh doanh và các liên hệ, để biểu diễn mối quan hệ giữa chúng. | Không |
| 12 | **Person** | ContactType | Bảng này chứa các loại thông tin liên hệ, chẳng hạn như "Email", "Phone", v.v. | Không |
| 13 | **Person** | CountryRegion | Bảng này chứa thông tin về các quốc gia và khu vực. | Chọn |
| 14 | **Person** | EmailAddress | Bảng này lưu trữ địa chỉ email của các thực thể kinh doanh hoặc cá nhân. | Không |
| 15 | **Person** | Password | Bảng này chứa thông tin mật khẩu của người dùng, có thể được sử dụng để xác thực. | Không |
| 16 | **Person** | Person | Bảng này lưu trữ thông tin cá nhân, bao gồm tên, giới tính, ngày sinh, v.v. của khách hàng và các cá thể khác trong công ty. | **Chọn**  Không có thông tin về ngày sinh, thu nhập, học vấn…. |
| 17 | **Person** | PersonPhone | Bảng liên kết giữa các cá nhân và số điện thoại của họ. | Không |
| 18 | **Person** | PhoneNumberType | Bảng này chứa các loại số điện thoại, chẳng hạn như "Home", "Work", "Mobile", v.v. | Không |
| 19 | **Person** | StateProvince | Bảng này chứa thông tin về các bang hoặc tỉnh trong một quốc gia, bao gồm cả mã bang và tên. | Không |
| 20 | **Production** | BillOfMaterials | Chứa thông tin về các hóa đơn vật liệu, đặc biệt là các thành phần và cấu trúc của sản phẩm. | Không |
| 21 | **Production** | Culture | Lưu trữ thông tin về các văn hóa hoặc ngôn ngữ được sử dụng trong hệ thống. | Không |
| 22 | **Production** | Document | Chứa thông tin về các tài liệu hoặc hồ sơ trong hệ thống. | Không |
| 23 | **Production** | Illustration | Lưu trữ hình ảnh minh họa hoặc đồ họa được sử dụng trong các tài liệu hoặc sản phẩm. | Không |
| 24 | **Production** | Location | Chứa thông tin về các địa điểm hoặc vị trí, có thể liên quan đến việc quản lý kho, vận chuyển, v.v. | Không |
| 25 | **Production** | Product | Là bảng cơ bản chứa thông tin về các sản phẩm được bán hoặc sản xuất bởi công ty. | **Chọn** |
| 26 | **Production** | ProductCategory | Chứa các danh mục (categories) cho các sản phẩm. | **Chọn** |
| 27 | **Production** | ProductCostHistory | Lưu trữ lịch sử giá thành sản phẩm. | Không |
| 28 | **Production** | ProductDescription | Chứa mô tả chi tiết về các sản phẩm. | Không |
| 29 | **Production** | ProductInventory | Lưu trữ thông tin về lượng tồn kho của các sản phẩm. | Không |
| 30 | **Production** | ProductListPriceHistory | Chứa thông tin về lịch sử giá bán lẻ niêm yết của nhà sản xuất của các sản phẩm. | Không |
| 31 | **Production** | ProductModel | Định nghĩa các mẫu (models) cho các loại sản phẩm. | Không |
| 32 | **Production** | ProductModelIllustration | Chứa thông tin về mối quan hệ giữa các mô hình sản phẩm và các minh họa được sử dụng để mô tả chúng. | Không |
| 33 | **Production** | ProductModelProductDescriptionCulture | Bảng này lưu trữ thông tin về mối quan hệ giữa các mô hình sản phẩm, mô tả sản phẩm và các văn hóa ngôn ngữ. | Không |
| 34 | **Production** | ProductPhoto | Lưu trữ hình ảnh của sản phẩm. | Không |
| 35 | **Production** | ProductProduct  Photo | Bảng "ProductProductPhoto" là một bảng kết nối (junction table) được sử dụng để ánh xạ mối quan hệ nhiều-nhiều giữa sản phẩm và các hình ảnh của chúng. | Không |
| 36 | **Production** | ProductReview | Chứa thông tin về đánh giá và nhận xét của khách hàng về các sản phẩm. | Không |
| 37 | **Production** | ProductSubcategory | Chứa các phân loại con (subcategories) cho các sản phẩm. | **Chọn** |
| 38 | **Production** | ScrapReason | Lưu trữ các lý do khiến sản phẩm bị hỏng hoặc bị loại bỏ. | Không |
| 39 | **Production** | TransactionHistory | Chứa thông tin về lịch sử giao dịch, bao gồm đơn đặt hàng, xuất nhập kho, v.v. | Không |
| 40 | **Production** | TransactionHistory  Archive | Bảng "TransactionHistoryArchive" t chứa thông tin về lịch sử các giao dịch đã được lưu trữ từ bảng "TransactionHistory" sau khi chúng đã được lưu trữ lâu dài hoặc chuyển sang một bảng lưu trữ dành riêng. | Không |
| 41 | **Production** | UnitMeasure | Định nghĩa các đơn vị đo lường được sử dụng trong hệ thống. | Không |
| 42 | **Production** | WorkOrder | Chứa thông tin về các đơn đặt hàng công việc. | Không |
| 43 | **Production** | WorkOrderRouting | Quản lý thông tin về quy trình làm việc của các đơn đặt hàng công việc. | Không |
| 44 | **Purchasing** | ProductVendor | Bảng này chứa thông tin về mối quan hệ giữa các sản phẩm và các nhà cung cấp. Nó ghi lại các thông tin như giá mua, số lượng tối thiểu, số lượng tối đa, v.v., cho mỗi sản phẩm của mỗi nhà cung cấp. | Không |
| 45 | **Purchasing** | PurchaseOrderDetail | Chứa thông tin chi tiết về các mặt hàng được đặt mua trong các đơn đặt hàng mua hàng, bao gồm thông tin về số lượng, giá, chiết khấu, v.v. | Không |
| 46 | **Purchasing** | PurchaseOrderHeader | Bảng này lưu trữ thông tin tổng quan về các đơn đặt hàng mua hàng, bao gồm thông tin về ngày đặt hàng, ngày dự kiến giao hàng, tổng giá trị, v.v. | Không |
| 47 | **Purchasing** | ShipMethod | Chứa thông tin về các phương thức vận chuyển hoặc giao hàng có sẵn để sử dụng trong quá trình giao hàng của đơn đặt hàng. | Không |
| 48 | **Purchasing** | Vendor | Bảng này chứa thông tin về các nhà cung cấp, bao gồm tên, địa chỉ, thông tin liên lạc, v.v. | Không |
| 49 | **Sales** | CountryRegionCurrency | Chứa thông tin về quan hệ giữa các quốc gia, khu vực và loại tiền tệ. | Không |
| 50 | **Sales** | CreditCard | Lưu trữ thông tin về các thẻ tín dụng được sử dụng trong các giao dịch bán hàng. | Không |
| 51 | **Sales** | Currency | Chứa thông tin về các loại tiền tệ được sử dụng trong các giao dịch. | Không |
| 52 | **Sales** | CurrencyRate | Ghi lại tỷ giá hối đoái giữa các loại tiền tệ. | Không |
| 53 | **Sales** | Customer | Lưu trữ thông tin về các khách hàng, bao gồm tên, địa chỉ, thông tin liên lạc, v.v. | **Chọn** |
| 54 | **Sales** | PersonCreditCard | Chứa thông tin về quan hệ giữa các cá nhân và các thẻ tín dụng. | Không |
| 55 | **Sales** | SalesOrderDetail | Chứa thông tin chi tiết về các mặt hàng được bán trong các đơn đặt hàng. | **Chọn** |
| 56 | **Sales** | SalesOrderHeader | Lưu trữ thông tin tổng quan về các đơn đặt hàng bao gồm thông tin như khách hàng, ngày đặt hàng, ngày dự kiến giao hàng, v.v. | **Chọn** |
| 57 | **Sales** | SalesOrderHeaderSalesReason | Liên kết giữa đơn đặt hàng và các lý do được cung cấp cho việc đặt hàng đó. | Không |
| 58 | **Sales** | SalesPerson | Chứa thông tin về các nhân viên bán hàng. | Không |
| 59 | **Sales** | SalesPersonQuota  History | Ghi lại lịch sử mục tiêu doanh số bán hàng của các nhân viên bán hàng. | Không |
| 60 | **Sales** | SalesReason | Lưu trữ các lý do được cung cấp cho việc thực hiện giao dịch bán hàng. | Không |
| 61 | **Sales** | SalesTaxRate | Chứa thông tin về tỷ lệ thuế được áp dụng cho các giao dịch bán hàng. | Không |
| 62 | **Sales** | SalesTerritory | Lưu trữ thông tin về các khu vực bán hàng và quản lý các quy tắc kinh doanh liên quan đến khu vực đó. | **Chọn** |
| 63 | **Sales** | SalesTerritory  History | Ghi lại lịch sử của các khu vực bán hàng. | Không |
| 64 | **Sales** | ShoppingCartItem | Chứa thông tin về các mặt hàng được thêm vào giỏ hàng của khách hàng. | Không |
| 65 | **Sales** | SpecialOffer | Chứa thông tin về các ưu đãi đặc biệt hoặc chương trình khuyến mãi được cung cấp cho khách hàng. | Không |
| 66 | **Sales** | SpecialOfferProduct | Liên kết giữa các sản phẩm và các ưu đãi đặc biệt. | Không |
| 67 | **Sales** | Store | Lưu trữ thông tin về các cửa hàng hoặc địa điểm bán hàng. | **Chọn** |

### Source Diagram

Tổng hợp lại các bảng đã chọn ở trên chúng ta có source diagram bao gồm 10 bảng chứa thông tin phục vụ cho dự án như sau:

****

### Thiết kế ERD

* 1. Tìm kiếm và chọn lọc trong các bảng đã lựa chọn ở trên các cột cần thiết cho việc phân tích, xác định các bảng DIM, FACT để chuẩn bị xây dựng DWH.

| **STT** | **Source** | **Columns** | **Columns Selection** | **Keys** | **Type** |
| --- | --- | --- | --- | --- | --- |
| 1 | **Production.**  **Product** | ProductID  Name  ProductNumber  MakeFlag  FinishedGoodsFlag  Color  SafetyStockLevel  ReorderPoint  StandardCost  ListPrice  Size  SizeUnitMeasureCode  WeightUnitMeasureCode  Weight  DaysToManufacture  ProductLine  Class  Style  ProductSubcategoryID  ProductModelID  SellStartDate  SellEndDate  DiscontinuedDate  rowguid  ModifiedDate | ProductID  ProductSubcategoryID  ProductName  ProductNumber  MakeFlag  Color  StandardCost  ListPrice  ProductLine  Style | PK: ProductID  FK:ProductSubcategoryID(Link to Production.  Product Subcategory) | DIM |
| 2 | **Production.**  **Product**  **Subcategory** | ProductSubcategoryID  ProductCategoryID  Name  rowguid  ModifiedDate | ProductSubcategoryID  ProductCategoryID  ProductSubcategoryName | PK:ProductSubcategoryID  FK:ProductCategoryID(Link to Production.  Product Category) | DIM |
| 3 | **Production.**  **Product**  **Category** | ProductCategoryID  Name  rowguid  ModifiedDate | Product  CategoryName | PK: ProductCategoryID | DIM |
|  | **Sales.**  **Customer** | CustomerID  PersonID  StoreID  TerritoryID  AccountNumber  rowguid  ModifiedDate | CustomerID  PersonID  StoreID  TerritoryID  Title  FirstName  MiddleName  LastName | PK: CustomerID  FK: PersonID(Person)  FK: StoreID(Store)  FK:TerritoryID(SalesTerritory) | DIM |
| 5 | **Person.**  **Person** | BusinessEntityID  PersonType  NameStyle  Title  FirstName  MiddleName  LastName  Suffix  EmailPromotion  AdditionalContactInfo  Demographics  rowguid  ModifiedDate | Title  FirstName  MiddleName  LastName | PK: BusinessEntityID | Bảng để lấy thông tin về khách hàng |
| 6 | **Sales.Store** | BusinessEntityID  Name  SalesPersonID  Demographics  rowguid  ModifiedDate | StoreID StoreName | PK: StoreID | DIM |
| 7 | **Sales.**  **Sales**  **Territory** | TerritoryID  Name  CountryRegionCode  Group  SalesYTD  SalesLastYear  CostYTD  CostLastYear  rowguid  ModifiedDate | TerritoryID  TerritoryName CountryRegionCode  Group | PK: TerritoryID | DIM |
| 8 | **Person.CountryRegion** |  |  |  | Bảng để lấy thông tin các khu vực |
| 9 | **Sales.Sales**  **Order**  **Header** | SalesOrderID  RevisionNumber  OrderDate  DueDate  ShipDate  Status  OnlineOrderFlag  SalesOrderNumber  PurchaseOrderNumber  AccountNumber  CustomerID  SalesPersonID  TerritoryID  BillToAddressID  ShipToAddressID  ShipMethodID  CreditCardID CreditCardApprovalCode  CurrencyRateID  SubTotal  TaxAmt  Freight  TotalDue  Comment  rowguid  ModifiedDate | SalesOrderID  CustomerID  TerritoryID  OrderDate  DueDate  ShipDate  Status  OnlineOrderFlag | PK: SalesOrderID  FK: CustomerID (Link to SalesCustomer)  FK:TerritoryID(SalesTerritory) | FACT |
| 10 | **Sales.Sales**  **Order**  **Detail** | SalesOrderID SalesOrderDetailID  CarrierTrackingNumber  OrderQty  ProductID  SpecialOfferID  UnitPrice  UnitPriceDiscount  LineTotal  rowguid  ModifiedDate | SalesOrderID  SalesOrderDetailID  ProductID  OrderQty  UnitPrice  LineTotal | PK, FK: SalesOrderID  PK: SalesOrderDetailID  FK: ProductID | FACT |

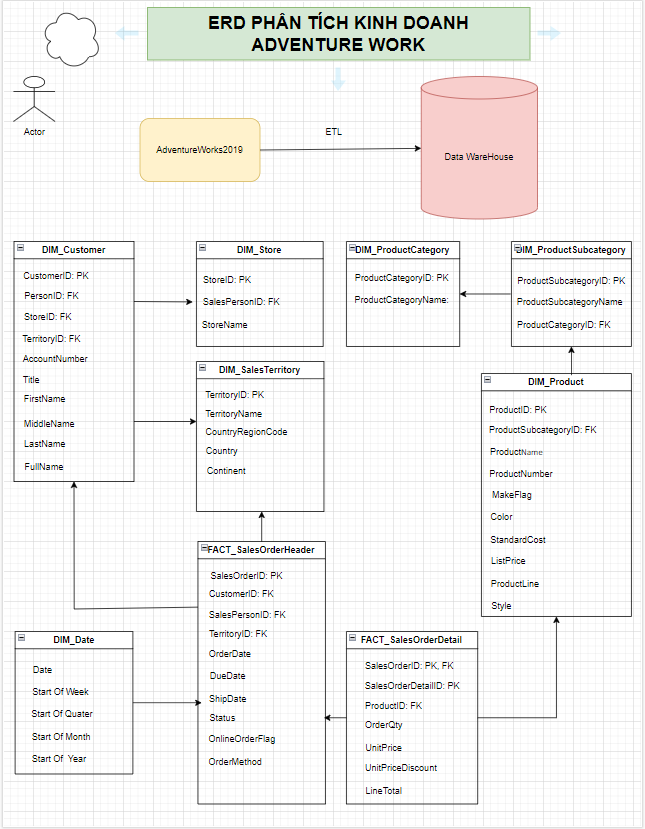
* 1. Chọn kho dữ liệu phù hợp
     1. Dữ liệu nguồn là cơ sở dữ liệu quan hệ có các ràng buộc khóa chính và khóa ngoại
     2. Yêu cầu của dự án là phân tích tình hình kinh doanh chung của công ty trên 2 kênh bán hàng Reseller và Online, có nghĩa là chúng ta đang cần phân tích tất cả các chỉ số kinh doanh nói chung chứ không có yêu cầu chuyên biệt vào 1 chủ đề nào.

⇒ Từ 2 lý do trên thì cấu trúc Centralized Data WareHouse là phù hợp với yêu cầu dự án

* 1. Lựa chọn lược đồ phù hợp
     1. Cơ sở dữ liệu nguồn là cơ sở dữ liệu quan hệ với các ràng buộc khóa chính và khóa ngoại.
     2. Nhìn tổng quan thì CSDL nguồn đang có lược đồ hình bông tuyết, các bảng đã được chuẩn hóa(1 lần tách bảng là 1 lần chuẩn hóa), các bảng lớn đang kết nối với các bảng con qua các khóa ngoại tham chiếu.

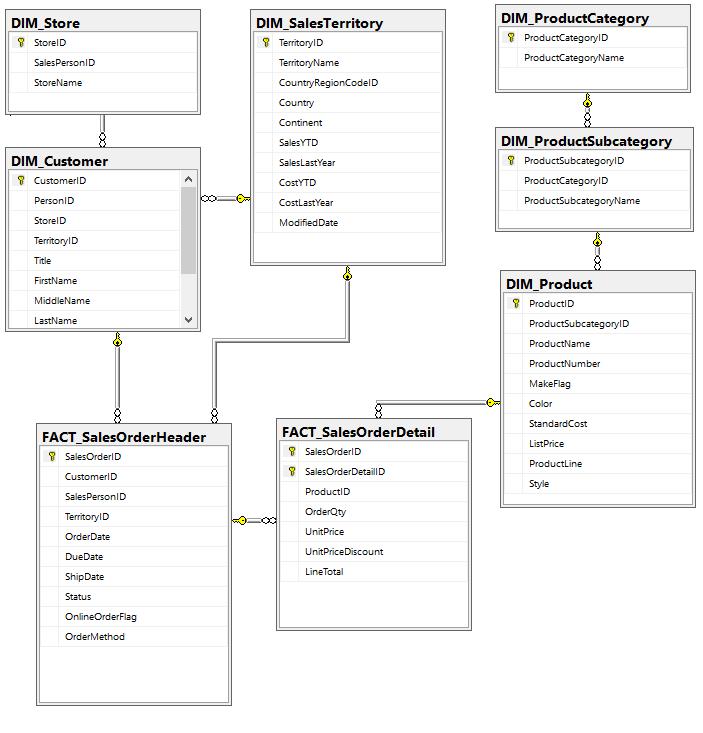
⇒ Từ 2 lý do trên thì lược đồ hình bông tuyết là phù hợp để xây dựng các bảng trong CSDL đích, như vậy sẽ dễ dàng hơn khi có nhu cầu mở rộng DWH phục vụ các nhu cầu kinh doanh bởi vì các bảng trong CSDL nguồn đều có sẵn các ràng buộc khóa chính và khóa ngoại tham chiếu.

* 1. Chúng ta có ERD như sau:  
     ERD bao gồm 6 bảng DIM và 2 bảng FACT được trích xuất từ CSDL nguồn(Source Diagram ở trên gồm 10 bảng). Và 1 bảng DIM\_Date sẽ tạo thêm khi xây dựng báo cáo trong PBI.



### Xây dựng Data WareHouse

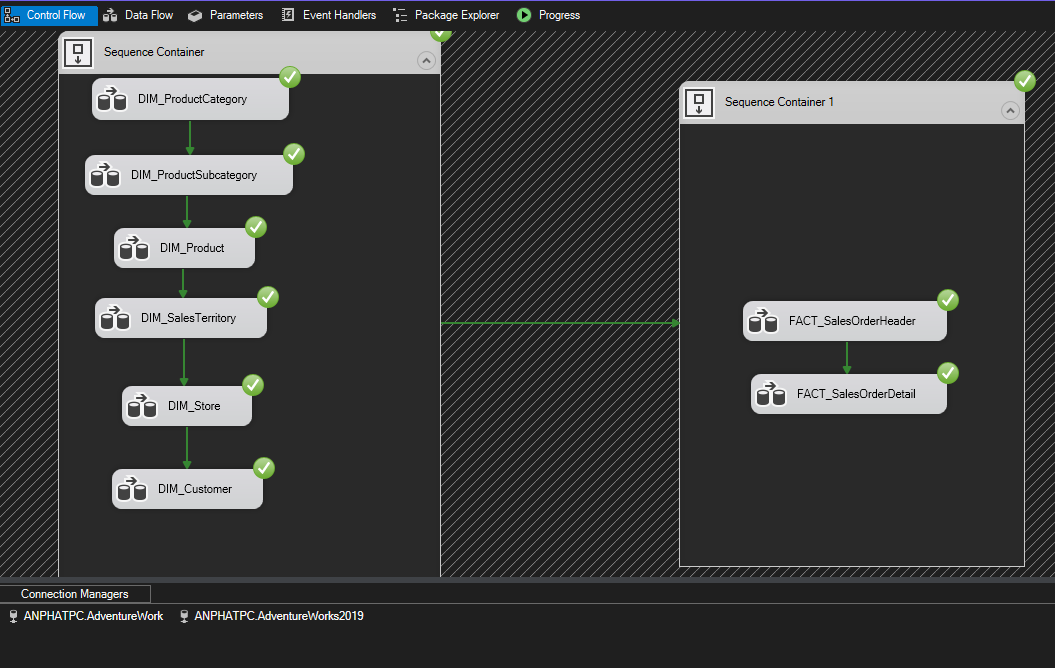
* 1. Xây dựng Destination Database AdventureWorks có các bảng, các cột, ràng buộc khóa chính khóa ngoại theo bản thiết kế ERD đã xây dựng ở trên, để chuẩn bị quá trình ETL vào DWH.



* 1. ETL Data
     1. Đây là CSDL quan hệ đã có đầy đủ các ràng buộc khóa chính vì vậy tại mỗi bảng sẽ không có dữ liệu Duplicates.
     2. Kiểm tra dữ liệu tại các bảng dữ liệu nguồn đã chọn, định hướng quá trình chuyển đổi, chuẩn hóa, đồng nhất, làm sạch dữ liệu.

| **STT** | **Extract From** | **Transform To** | **Tình trạng dữ liệu nguồn** | **Hướng xử lý trong SSIS** |
| --- | --- | --- | --- | --- |
| **1** | **Production.**  **Product** | DIM\_Product | - Có nhiều giá trị NULL tại cột ProductSubcategoryID.  - Các ProductID chứa các giá trị NULL này khi kiểm tra JOIN với bảng SalesOrderDetail cũng không có thông tin về đơn hàng.  - Các cột ProductLine, Style, MakeFlag đang chứa các ký hiệu viết tắt cụ thể:  - ProductLine S: Standard  M: Mountain  R: Road  T: Touring  - Style  U: Universal  M: Man  W: Women  - MakeFlag  0: Purchased  1: Manufactured in-house | Giải pháp giả định thực tế: Xác nhận lại dữ liệu với bên liên quan.  Giải pháp trong phạm vi dự án:  - Chỉ lấy những giá trị NOTNULL. (Sử dụng Conditional Split)  - Chuyển đổi các ký tự viết tắt thành thông số đầy đủ(Sử dụng Derived Column) |
| **2** | **Production.**  **Product**  **Subcategory** | DIM\_ProductSubcategory | OK | ETL đầy đủ dữ liệu từ nguồn vào bảng đích |
| **3** | **Production.**  **Product**  **Category** | DIM\_ProductCategory | OK | ETL đầy đủ dữ liệu từ nguồn vào bảng đích |
| **4** | **Sales.**  **Customer** | DIM\_Customer | Cần lấy thêm thông tin của khách hàng từ bảng Person.Person. | Sử dụng Merge Join để nối 2 bảng Sales.Customer và Person.Person để lấy thông tin về tên và title  Sử dụng Derived Column để chuyển đổi First name, Mid name, Last name thành FullName và chuyển đổi title thành Gender. |
| **5** | **Person.**  **Person** | DIM\_Customer | Dữ liệu khách hàng rất hạn chế khi chỉ có thông tin về tên và title |
| **6** | **Sales.Store** | DIM\_Store | OK | ETL đầy đủ dữ liệu từ nguồn vào bảng đích |
| **7** | **Sales.**  **Sales**  **Territory** | DIM\_Territory | Cần thêm thông tin tên quốc gia từ bảng Person.CountryRegion | Sử dụng Merge Join để nối 2 bảng Sales.SalesTerritory và Person.CountryRegion. |
| **8** | **Person.CountryRegion** | DIM\_Territory | Dữ liệu đầy đủ |
| **9** | **Sales.Sales**  **Order**  **Header** | DIM\_SalesOrderHeader | Không có giá trị NULL tại các cột cần cho dự án.  Cột OnlineOrderFlag đang chứa giá trị viết tắt:  0: Reseller  1: Online  Có thừa dữ liệu ngoài 2 năm 2012 và 2013 | Sử dụng Conditional Split để lọc dữ liệu lấy trong 2 năm 2012&2013.  Sử dụng Derived Column để chuyển đổi cột OnlineOrderFlag thành thông tin đầy đủ. |
| **10** | **Sales.Sales**  **Order**  **Detail** | DIM\_SalesOrderDetail | Không có giá trị NULL  Chúng ta cũng cần lọc ra các SalesOrderDetailID trong 2 năm 2012&2013 tuy nhiên trong bảng không có cột Orderdate | Sử dụng Merge Join nối 2 bảng Sales.SalesOrderHeader và Sales.SalesOrderDetail để lấy cột Orderdate.  Sử dụng Conditional Split để lọc dữ liệu lấy trong 2 năm 2012 & 2013 |

* + 1. Thực hiện ETL trong SSIS



### Kiểm tra lại Data WareHouse

* 1. Sau khi đã xây dựng xong kho dữ liệu, chúng ta kiểm tra lại bằng cách truy vấn vào cơ sở dữ liệu dựa theo 1 số yêu cầu nghiệp vụ trước khi import dữ liệu vào các ứng dụng xây dựng báo cáo.  
     (Chi tiết trong file phụ lục đính kèm)

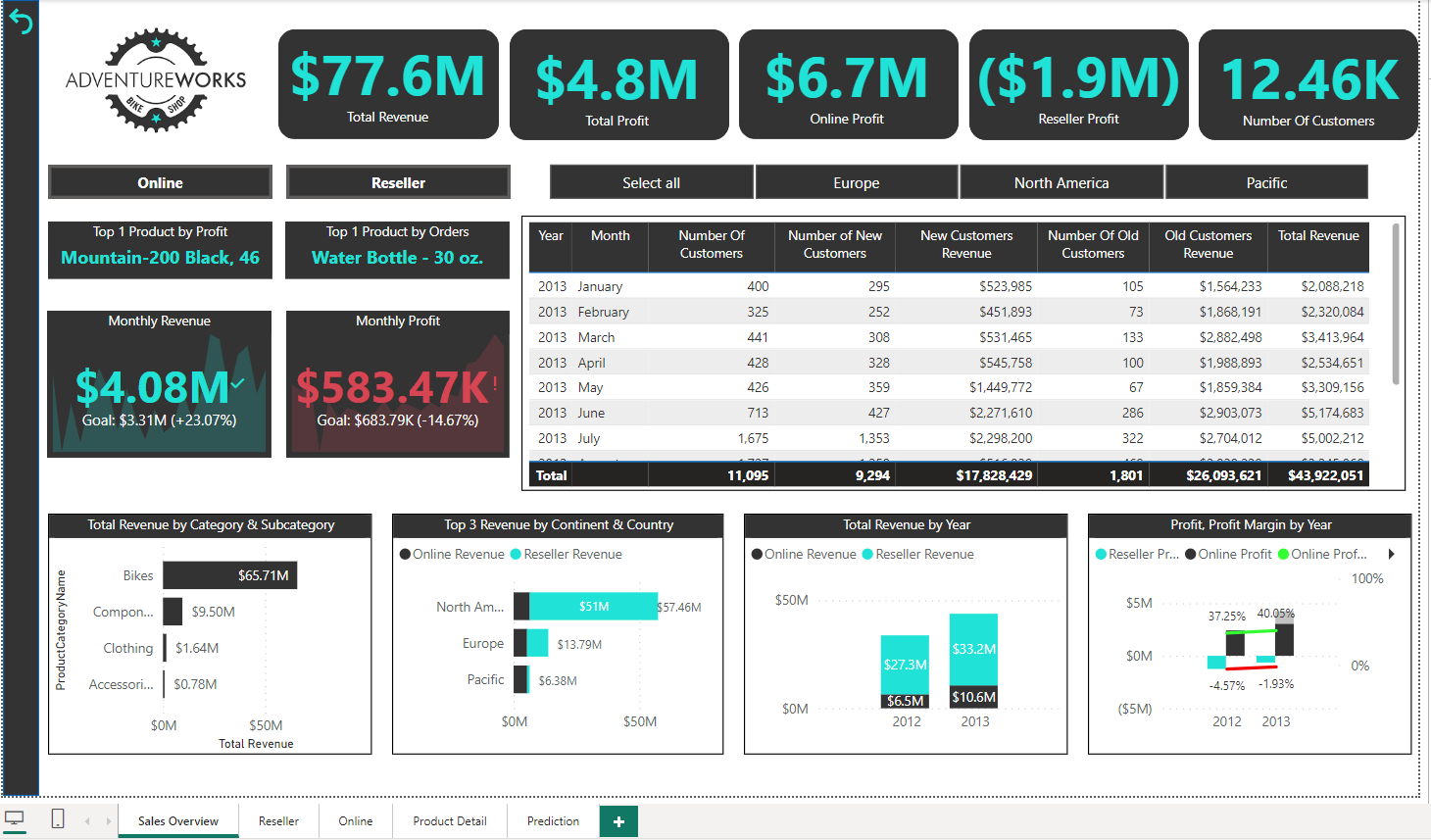
### Xây dựng báo cáo

#### Ý tưởng thiết kế nội dung

* + 1. Sản phẩm sẽ được xây dựng dựa trên 4 loại phân tích dữ liệu chính:
       1. Phân tích mô tả
       2. Phân tích chẩn đoán
       3. Phân tích dự đoán
       4. Phân tích đề xuất

#### Ý tưởng thiết kế giao diện

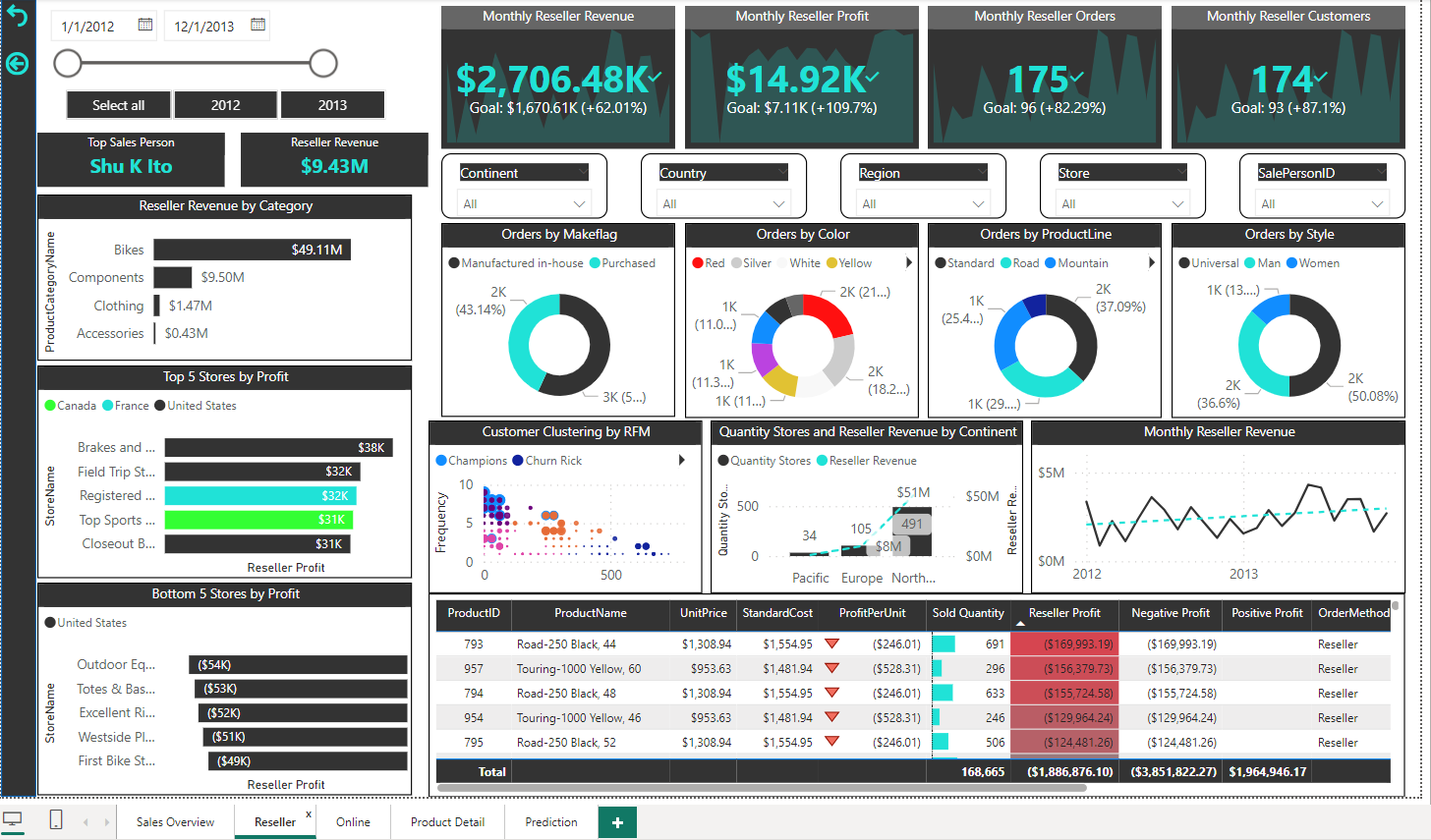
* + 1. Sản phẩm sẽ được xây dựng dựa trên các màu sắc chủ đạo của công ty AdventureWorks là màu đen và màu xanh, giống như Logo của công ty.
    2. Báo sẽ được thiết kế theo view chữ F.
    3. Bố cục báo cáo sẽ gồm 3 phần chính.
       1. Sales OverView(Phân tích mô tả)  
          Đây là trang chứa các thông tin tổng quan phục vụ cho các cấp quản lý như Giám đốc điều hành, giám đốc kinh doanh.



* + - 1. Online Channel(Phân tích chẩn đoán)  
         Đây là trang chứa các thông tin chi tiết phục vụ cho cấp nhân viên hoặc cấp quản lý trực tiếp như trưởng phó phòng, leader của bộ phận Online.

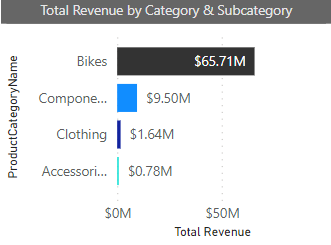


* + - 1. Reseller Channel(Phân tích chẩn đoán)  
         Đây là trang chứa các thông tin chi tiết phục vụ cho cấp nhân viên hoặc cấp quản lý trực tiếp như trưởng phó phòng, leader của bộ phận Reseller.



### Đánh giá kết quả

#### Sử dụng báo cáo thử trả lời các câu hỏi nghiệp vụ đã đặt ở trên.

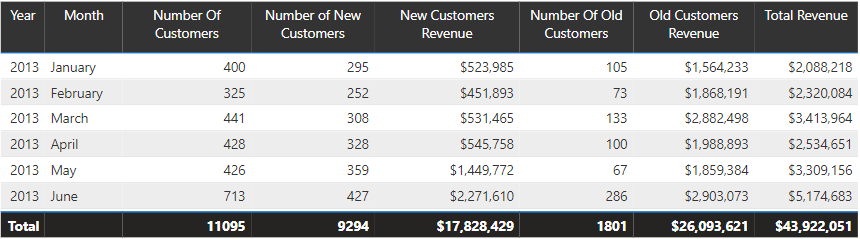
* + 1. Giám đốc điều hành, giám đốc kinh doanh quan tâm tới điều gì?
       1. Sản phẩm
          1. Công ty đang có những phân loại sản phẩm nào? mặt hàng nào đem lại doanh thu cao nhất?  
             
          2. Sản phẩm nào đang bán chạy nhất?



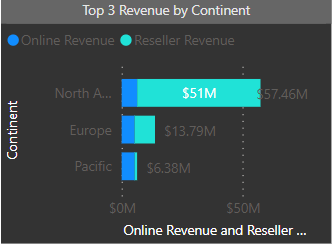
* + - * 1. Sản phẩm nào đang đem lại lợi nhuận nhiều nhất cho công ty?



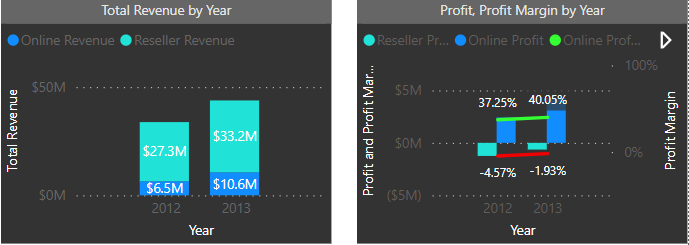
* + - 1. Khách hàng
         1. Công ty đang có bao nhiêu khách hàng?
         2. Có thêm bao nhiêu khách hàng mới theo tháng và theo năm?
         3. Có bao nhiêu khách hàng cũ từ năm trước vẫn trung thành với công ty?
         4. Doanh thu của các nhóm khách hàng này là bao nhiêu?



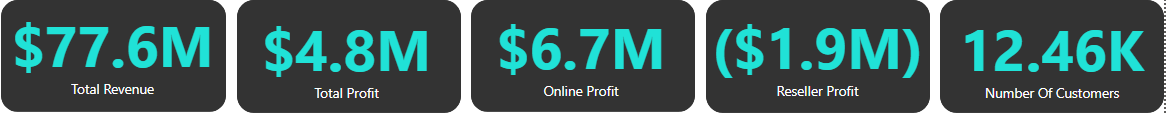
* + - 1. Thị trường
         1. Công ty đang hoạt động tại thị trường nào?
         2. Thị trường nào đem lại doanh thu tốt nhất cho công ty?

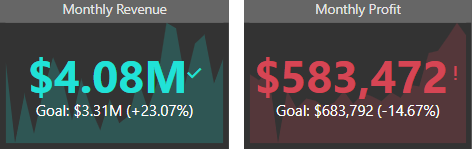


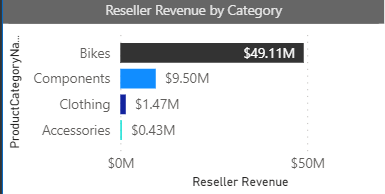
* + - 1. Kênh phân phối
         1. Có những kênh phân phối nào?doanh thu, lợi nhuận trên từng kênh phân phối là bao nhiêu?
         2. Doanh thu, lợi nhuận tăng trưởng trong các năm của 2 kênh phân phối?
         3. Biên lợi nhuận của từng kênh là bao nhiêu %?

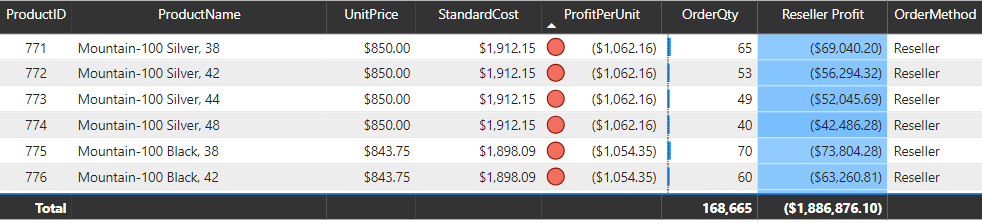


* + - 1. Doanh thu, lợi nhuận
         1. Tổng doanh thu, lợi nhuận của công ty là bao nhiêu?

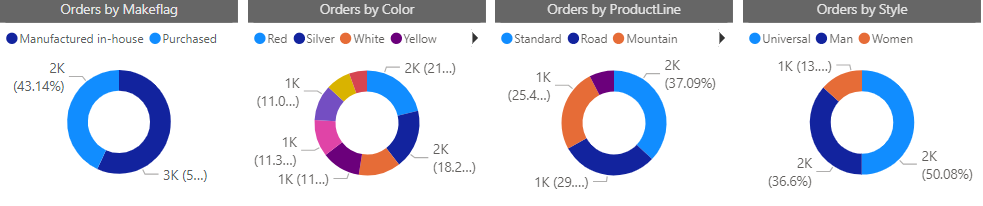


* + - * 1. Có đạt được KPI hàng tháng hay không? 
    1. **Nhân viên các bộ phận của kênh bán lẻ Reseller quan tâm tới điều gì?**
       1. Sản phẩm
          1. Công ty đang có những loại mặt hàng nào? mặt hàng nào đem lại doanh thu cao nhất?

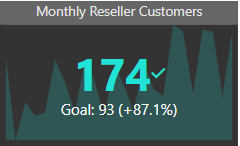




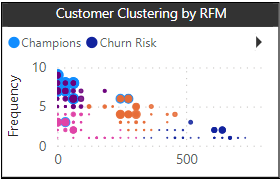
* + - * 1. Các đặc điểm sản phẩm nào đang bán được nhiều nhất?
* Color?
* ProductLine?
* Style?



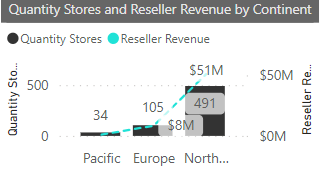
* + - 1. Khách hàng
         1. Hàng tháng có bao nhiêu khách hàng mới? có bằng tháng trước hay không?



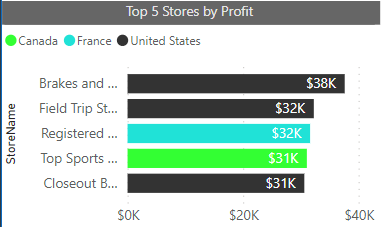
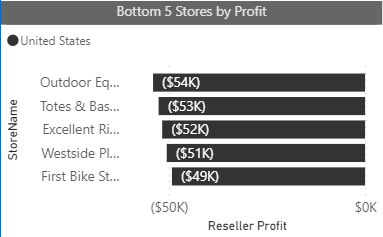
* + - * 1. Phân khúc khách hàng của công ty? Chất lượng của từng phân khúc?

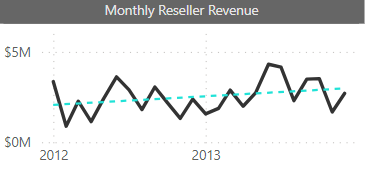


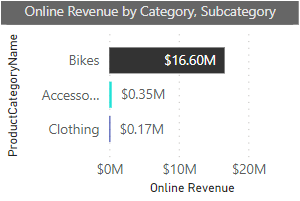
* + - 1. Thị trường và hệ thống cửa hàng
         1. Công ty đang hoạt động mạnh nhất tại thị trường nào? đem lại doanh thu là bao nhiêu?
         2. Có bao nhiêu cửa hàng tại các thị trường công ty đang hoạt động?

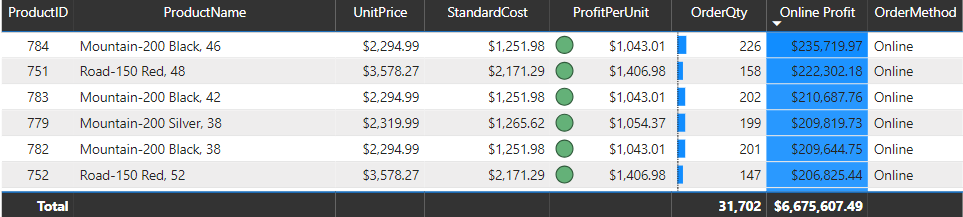


* + - * 1. Tình hình kinh doanh tại các cửa hàng?

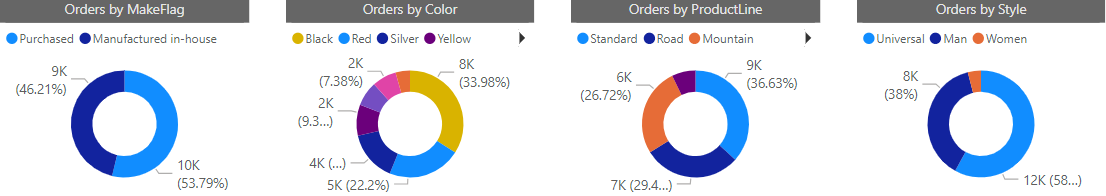


* + - 1. Doanh thu, lợi nhuận
         1. Doanh thu lợi nhuận tháng này đang là bao nhiêu? Đã sắp đủ KPI hay chưa?  
            
         2. Doanh thu lợi nhuận tăng trưởng theo thời gian?  
            
    1. **Nhân viên các bộ phận của kênh bán lẻ Online quan tâm tới điều gì?**
       1. Sản phẩm
          1. Công ty đang có những loại mặt hàng nào? mặt hàng nào, sản phẩm nào đem lại doanh thu cao nhất?

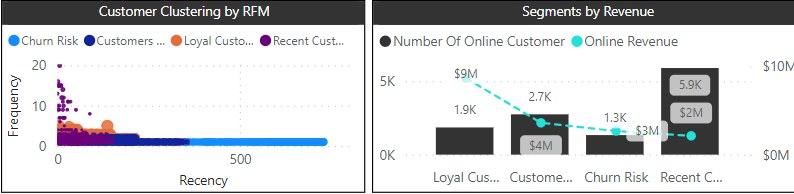


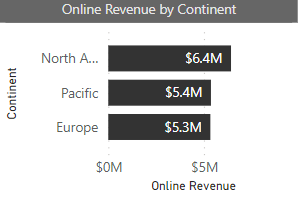


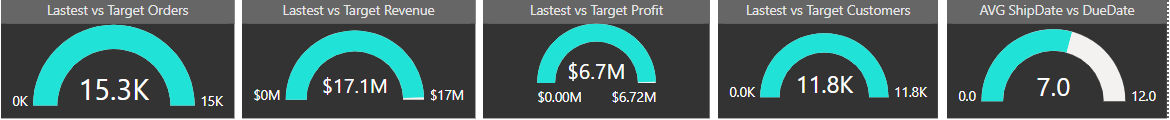
* + - * 1. Các đặc điểm sản phẩm nào đang bán được nhiều nhất?
* Màu sắc?
* ProductLine?
* Style?

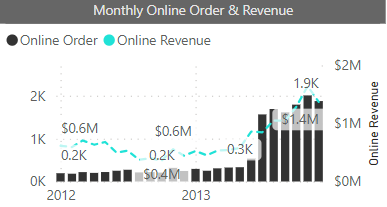


* + - 1. Khách hàng
         1. Hàng tháng có bao nhiêu khách hàng mới? có đạt được mục tiêu đề ra hay không?
         2. Phân khúc khách hàng của công ty? Chất lượng của từng phân khúc?



* + - 1. Thị trường
         1. Công ty đang hoạt động mạnh nhất tại thị trường nào? đem lại doanh thu là bao nhiêu?  
            
      2. Doanh thu, lợi nhuận
         1. Doanh thu lợi nhuận tháng này đang là bao nhiêu? Đã sắp đủ KPI hay chưa?



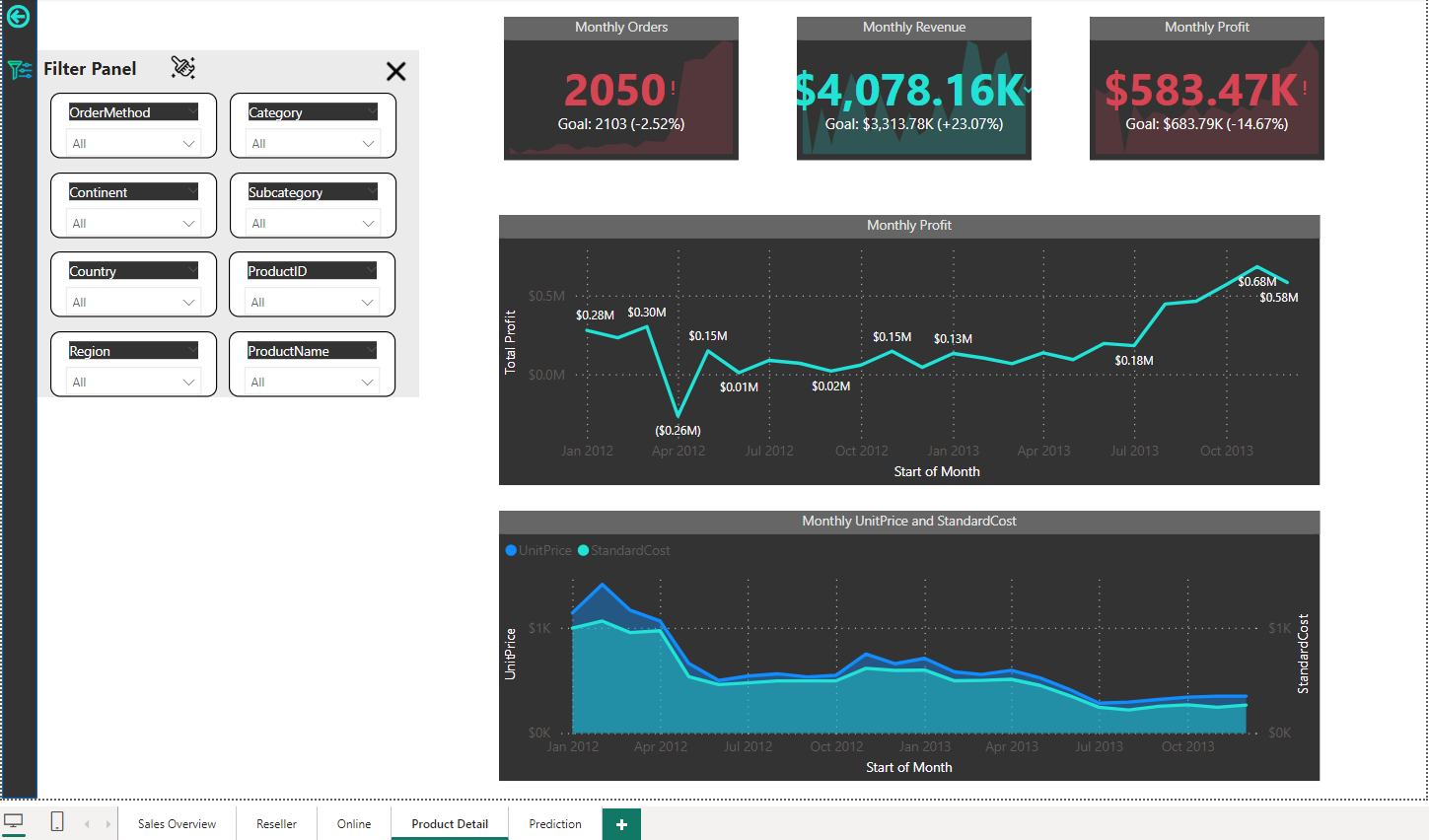
* + - * 1. Doanh thu lợi nhuận theo xu hướng thời gian tăng hay giảm?  
           

#### Dựa vào các yêu cầu của DWH và của báo cáo đã nêu ở trên để đánh giá tổng quan về dự án.

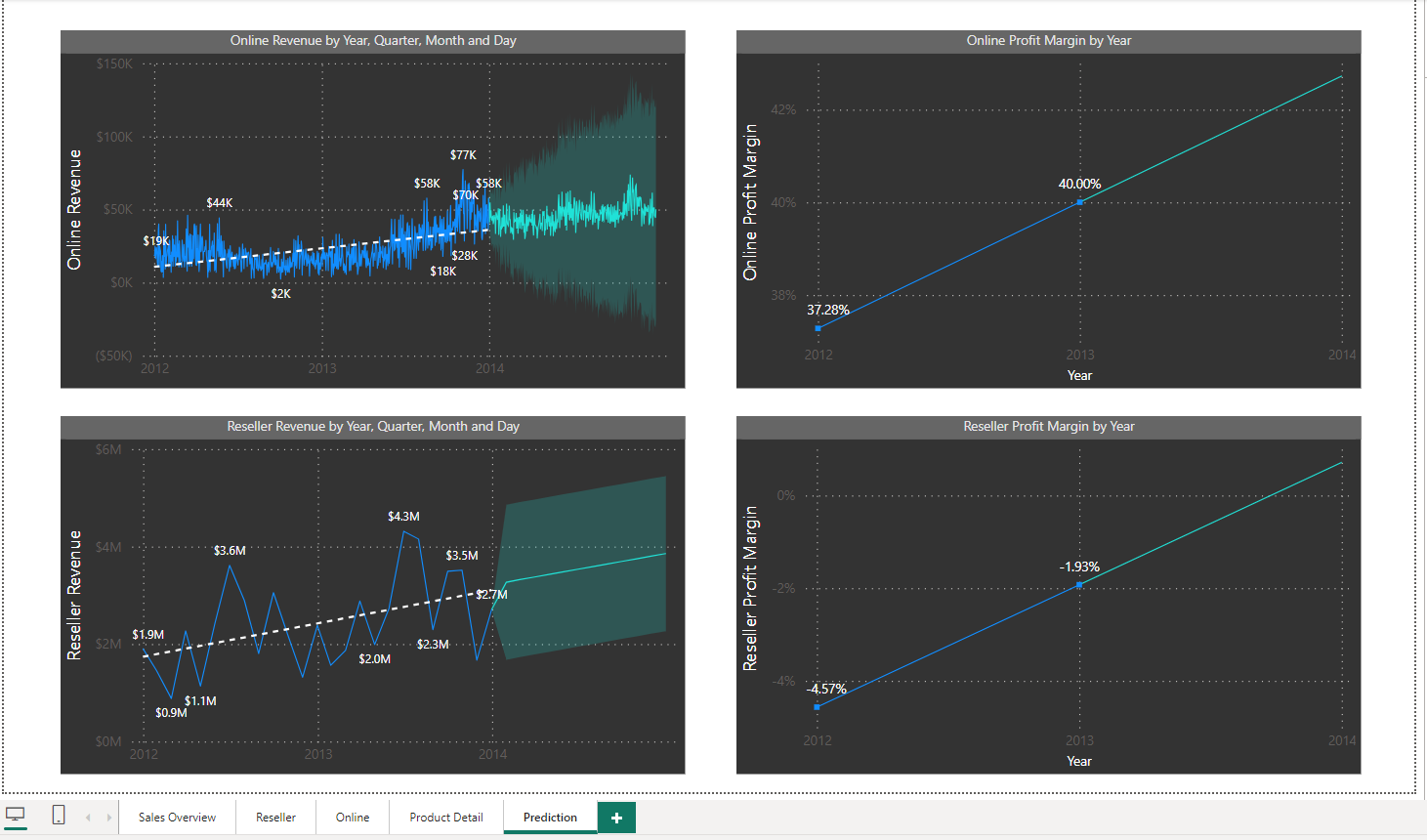
* + 1. Đối với DWH
       1. Đã đáp ứng được yêu cầu kinh doanh, trả lời được tất cả các câu hỏi nghiệp vụ.
       2. Dữ liệu đã được đồng nhất và chuẩn hóa, đã làm sạch tất cả các dữ liệu bị lỗi, giá trị null.
       3. DWH có thể mở rộng rất dễ dàng, tuy nhiên cần tuân theo các ràng buộc của CSDL nguồn.
    2. Đối với Báo cáo
       1. Bố cục: Báo cáo đã đủ 3 phần chính như yêu cầu.
       2. Dữ liệu sử dụng trong báo cáo đảm bảo tính nhất quán và chính xác so với dữ liệu nguồn.
       3. Báo cáo đã được bố cục theo layout hình chữ F, có màu sắc nhất quán tại các trang, nội dung được trực quan hóa bằng biểu đồ đảm bảo dễ đọc và dễ hiểu.
       4. Tại các trang có các slicer đảm bảo tính cơ động, người dùng có thể tương tác bằng các Drill Though hoặc Slicer.

#### Các phần bổ sung thêm để giải quyết 1 số yêu cầu nghiệp vụ.

* + 1. Trong quá trình phân tích đã có những vấn đề phát sinh, có những chỉ số kinh doanh cần báo cáo thêm vì vậy báo cáo đã bổ sung 1 trang Product Detail. Nội dung là phân tích lợi nhuận, giá bán, giá gốc của từng sản phẩm theo thời gian.



* + 1. Ngoài ra sản phẩm cũng được thiết kế thêm 1 trang phụ Prediction để dự báo doanh thu, biên lợi nhuận cho công ty và cho từng kênh phân phối.



* + 1. Bổ sung bảng DIM\_SalesPerSon vào DWH.

Đây là bảng chứa thông tin về nhân viên bán hàng của kênh Reseller được trích xuất từ bảng Sales.SalesPerSon và Person.PerSon.

# PHẦN 3 - KẾT LUẬN

## Chương 5 - Hướng mở rộng dự án

1. Một số vấn đề còn chưa hoàn thành 100% và 1 số hướng mở rộng cho dự án như:
   1. Phân tích khách hàng - Bởi vì dữ liệu bị thiếu rất nhiều thông tin về khách hàng vì vậy báo cáo chưa thể phân tích sâu vào khách hàng. Khi có thêm thông tin khách hàng, có thể phân tích tương quan giữa các đặc điểm khách hàng và hành vi mua hàng..
   2. Phân tích nhân viên bán hàng - Cần thu thập thêm thông tin để xây dựng các mục tiêu theo tháng và theo năm qua đó làm báo cáo chi tiết hơn về nhân viên bán hàng.
   3. Phân tích xu hướng kênh Online: Thu thập thêm dữ liệu để phân tích sâu hơn 6 tháng cuối năm 2013 tại sao doanh thu lại tăng trưởng mạnh so với 6 tháng đầu năm và cùng kỳ năm 2012.
   4. Phân tích hàng tồn kho.
   5. Phân tích hành chính nhân sự.
   6. Phân tích bộ phận Purchase.

## Chương 6 - Tổng hợp những kiến thức sử dụng để hoàn thành dự án.

1. **Kiến thức về nghiệp vụ**
   1. Một chút kiến thức về kinh doanh, ngành bán lẻ, sản phẩm, thị trường để có thể hiểu và phân tích yêu cầu kinh doanh**.**
2. **SQL server.**
   1. Biết cách tạo database, bảng, thiết lập các ràng buộc khóa chính khóa ngoại để có thể tạo cơ sở dữ liệu phục vụ chứa dữ liệu được trích xuất từ nguồn dữ liệu thập được.
      1. Một số câu lệnh: create database, create table, primary key, foreign key.
   2. Biết cách truy vấn để có thể khám phá dữ liệu .
      1. Một số câu lệnh: select from where, inner join, group by, order by.
3. **SSIS**
   1. Hiểu về nguyên lý hoạt động của quy trình ETL từ DIM - >FACT.
   2. Hiểu thế nào là Control Flow, Data flow.
   3. Biết cách sử dụng 1 số công cụ trong SSIS để transform dữ liệu như:
      1. Conditional Split : Dùng để lọc dữ liệu theo điều kiện.
      2. Data Conversion: Dùng để chuyển đổi kiểu dữ liệu.
      3. Derived Column: Dùng để tạo các cột mới bằng cách sử dụng các biểu thức hoặc kết hợp các cột hiện có.
      4. Sort: Dùng để sắp xếp dữ liệu và lọc dữ liệu trùng.
      5. Look up: Dùng để chỉ trích xuất những dữ liệu mới hoặc ánh xạ theo dữ liệu của bảng khác.
      6. Merge join: Dùng để nối dữ liệu từ 2 bảng nguồn thành 1 bảng đích.
      7. OLE DB Source: Chứa nguồn dữ liệu.
      8. OLE DB Destination: Chứa dữ liệu đích sau quá trình ETL.
4. **Power BI**
   1. Biết cách import dữ liệu từ SQL vào PBI để làm báo cáo.
   2. Biết cách vào power query để thực hiện 1 số thao tác với dữ liệu, tạo thêm các cột cần thiết.
   3. Biết cách vào Model View để kiểm tra và tạo các relations cần thiết.
   4. Biết các tạo các measure dựa theo yêu cầu kinh doanh của dự án.
   5. Biết cách sử dụng các loại biểu đồ để trực quan hóa dữ liệu ví dụ như, line chart, column chart, kpi, card…
5. **Machine Learning**
   1. Hiểu về các mô hình học máy để có thể áp dụng vào dự án.
      1. Mô hình hồi quy tuyến tính để dự đoán 1 biến mục tiêu dựa trên các biến độc lập.
      2. Mô hình KMeans để phân khúc khách hàng hoặc sản phẩm.

## Chương 7 – Những đặc điểm nổi trội của dự án

1. Sản phẩm có 1 số yêu cầu nâng cao như tính toán khách hàng mới, khách hàng cũ.
2. Báo cáo có sử dụng học máy Kmeans để phân khúc khách hàng theo RFM theo phương pháp Elbow.
3. Tính tương tác và kết nối.
4. Màu sắc đồng nhất và thẩm mỹ.

## Chương 8 – Những khó khăn khi thực hiện dự án

1. Khó khăn ban đầu khi thực hiện dự án là phải hiểu về bộ dữ liệu nguồn với hàng chục bảng, các bảng thuộc phòng ban nào, chứa dữ liệu gì, sau đó phải hiểu rõ ý nghĩa của các cột trong các bảng, các ký hiệu viết tắt có ý nghĩa thế nào? các mối quan hệ giữa các bảng thế nào? từ đó mới có thể chọn lọc ra các bảng, các cột cần thiết để phục vụ cho dự án.
2. Khi xây dựng Data WareHouse phải tính toán sao cho tối giản nhất về mặt cấu trúc mà vẫn đầy đủ các dữ liệu cần cho bài toán kinh doanh.
3. Khi xây dựng báo cáo cũng luôn phải tư duy biểu đồ nào cần hoặc không cần, có thể tối giản đi hay không? báo cáo có phản ảnh đủ thông tin cho các bên liên quan hay không? cảm nhận của người dùng cuối thế nào khi họ xem báo cáo của mình..