**Sprawozdanie - Hurtownia na podstawie danych sklepu internetowego**

Hurtownie Danych

Prowadzący: dr inż. Dariusz Gąsior

Bączyński Konrad, 211147

Charewicz Jan, 212825

Spis treści

[1. Wybór i opis źródeł 3](#_Toc453023956)

[2. Jakość danych 4](#_Toc453023957)

[3. Projekt bazy analitycznej 5](#_Toc453023958)

[4. ETL 6](#_Toc453023959)

[5. Kostka 8](#_Toc453023960)

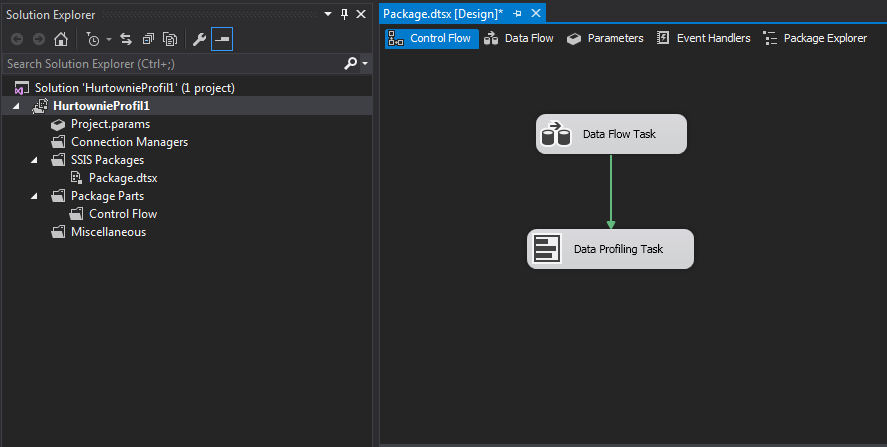
[6. Raporty 9](#_Toc453023961)

# Wybór i opis źródeł

Do naszego projektu, wybraliśmy znalezione w internecie źródło danych, będące sztandarową reprezentacją sklepu internetowego - w tym przypadku - Sklepu Dell’a na terenie Stanów Zjednoczonych. Znaleziony plik .sql - na różnych forach, był po wielokrotnie podawany jako nawet dobry materiał na hurtownie danych. W pliku znaleziono zarówno tworzenie tabel, jak i liczne polecenia INSERT (...). Baza zawierała generowane (nierzadko o rozkładzie zbliżonym do rozkładu równomiernego) informacje na temat klientów, zamówień, pozycji na zamówieniach, produktach.

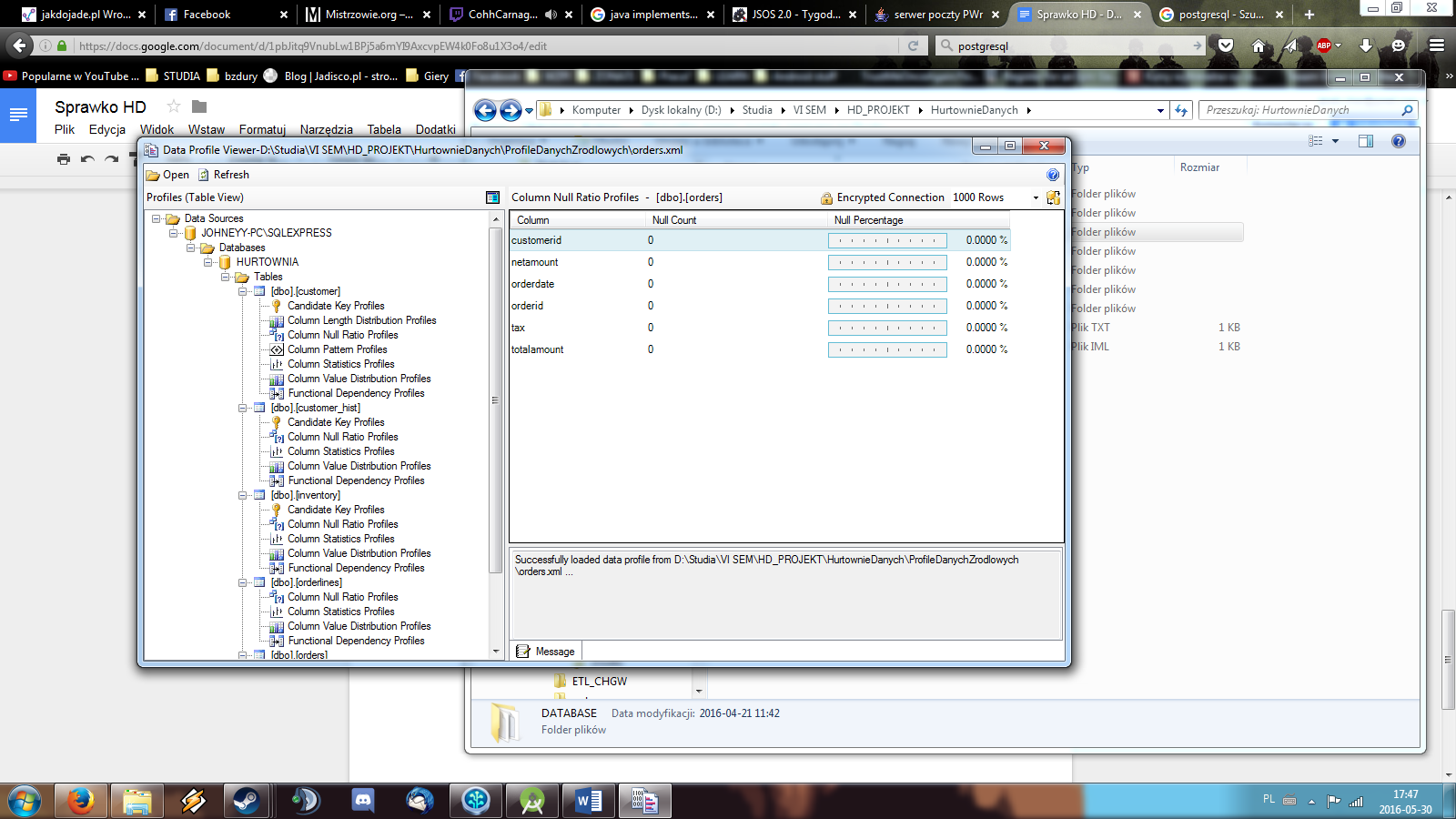
Motywację do wyboru źródła odnaleźliśmy w rekomendacjach tego źródła na forach, a także w bardzo przyjaznym formacie pliku. Niestety, jako, że język użyty w pliku, okazał się PostgreSQL’em, musieliśmy dokonać małych transformacji formatowania źródła, a by lepiej się pracowało w następnych etapach dane umieszczono (przy użyciu tak przeformatowanego i akceptowalnego przez MS pliku .sql) w bazie MS SQL SERVER i to ją wykorzystywaliśmy dalej jako źródło.

# Jakość danych



Rysunek Data Profiling Task w Visual Studio 2015.

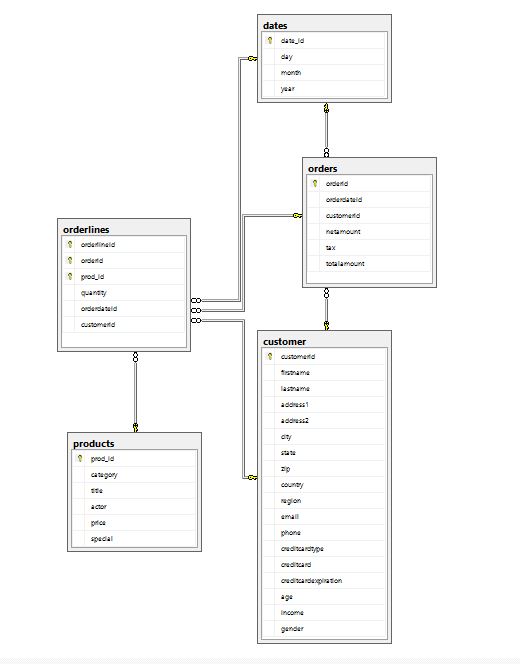
W celu wykonania oceny jakości danych, tworzy się profile na podstawie danych źródłowych, całość eksportowana jest do pliku .xml które to parsuje Data Profile Viewer (rys. poniżej). Całe zadanie opiewało na odpowiednią konfigurację zadań i połączeń w Projekcie Analysis Services.



Rysunek Data Profile Viewer

Jak pokazuje [Rysunek 1] źródło faktycznie okazało się bardzo w porządku, wskaźniki pól o wartości NULL w całej bazie wynosiły okrągłe 0. Z profilowania już można było się całkiem sporo dowiedzieć, np. to, że minimalny wiek klientów wynosił 18, pochodzili oni w około połowie z jednego kraju i częściej robili mniejsze zamówienia.

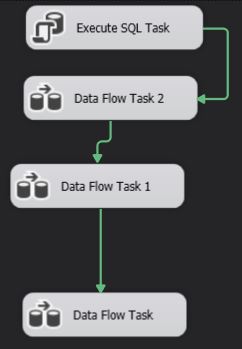
# Projekt bazy analitycznej



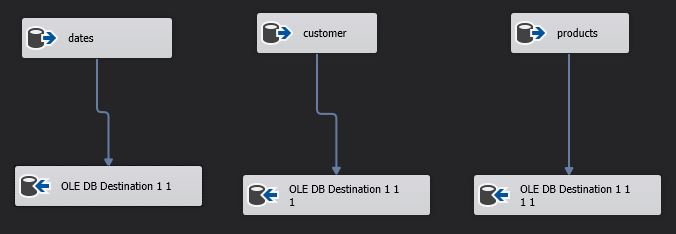
Rysunek Diagram tabel bazy analitycznej.

Powyższy diagram [Rysunek 3] przedstawia zaproponowany po konsultacji z prowadzącym schemat docelowej bazy analitycznej. W niektórych miejscach należało więc zastosować spłaszczenia, by ze schematu z dwoma potencjalnymi tabelami faktów, skupić się na pozycjach na zamówieniu jako głównym wpisie w hurtowni.

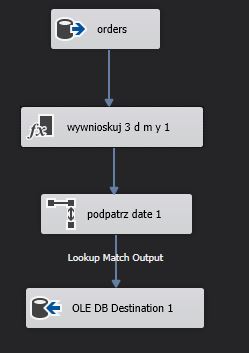
# ETL



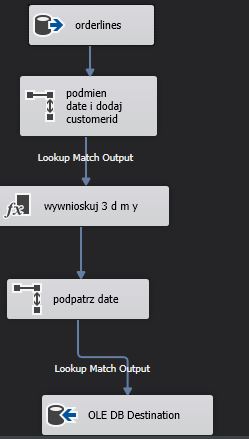
Rysunek Control flow.



Rysunek Data Flow Task 2.



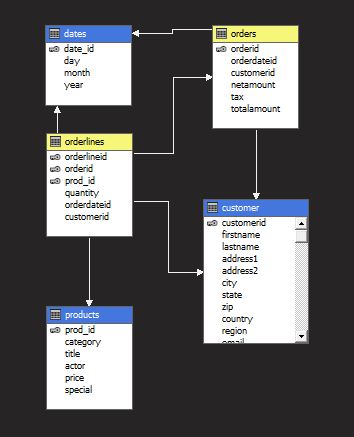
Rysunek Data Flow Task 1



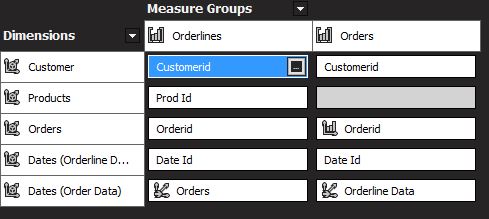
Rysunek Data Flow Task.

Powyższe rysunki przedstawiają realizację procesu ETL. Dzięki niemu uzyskujemy bazę wypełnioną przetworzonymi już częściowo danymi. Na przykład dodajemy dodatkową kolumnę do tabeli i wnioskujemy ją na podstawie innej kolumny w innej tabeli.

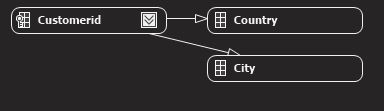
# Kostka



Rysunek Struktura kostki.



Rysunek Wykorzystane wymiary.



Rysunek Relacje między atrybutami Klienta.

Powyższe rysunki przedstawiają utworzoną kostkę. [Rysunek 8] przedstawia strukturę kostki. Kostka pozwala na szybką analizę danych. [Rysunek 9] przedstawia wykorzystane wymiary do budowy kostki oraz dwie tabele faktów. [Rysunek 10] przedstawia wymiar klienta i relacje między jego atrybutami.

# Raporty

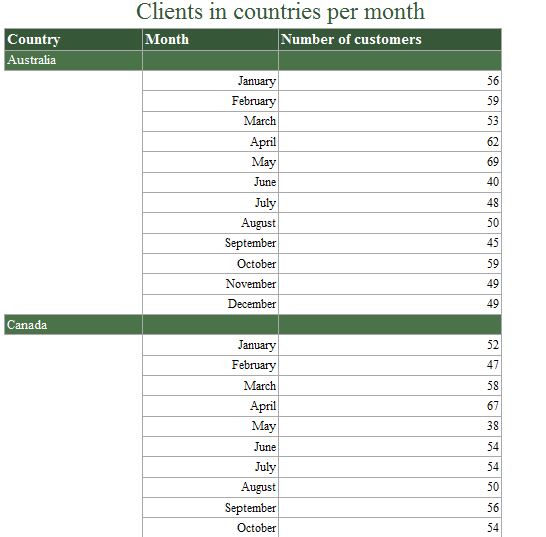
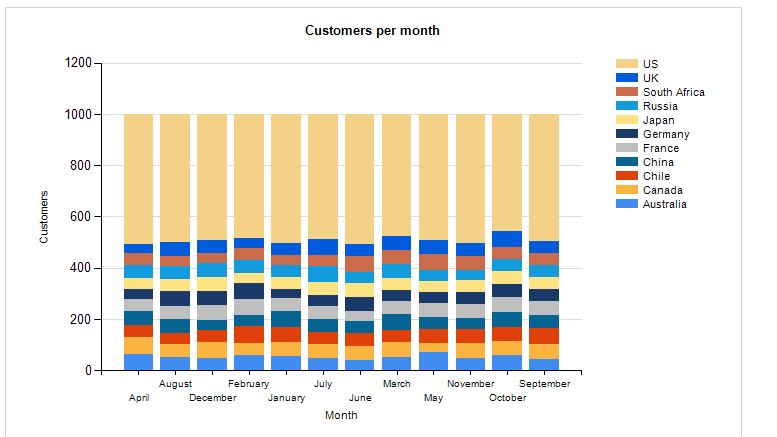
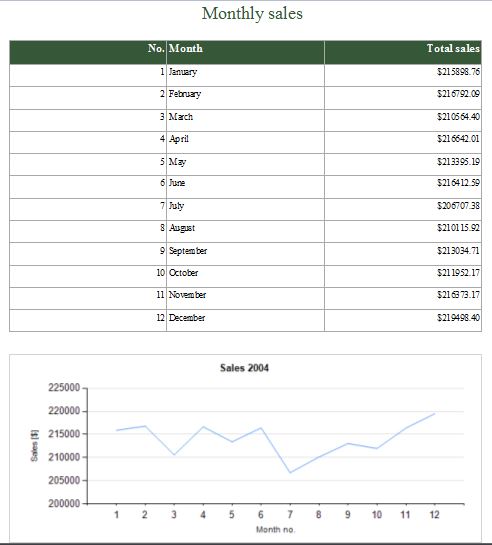


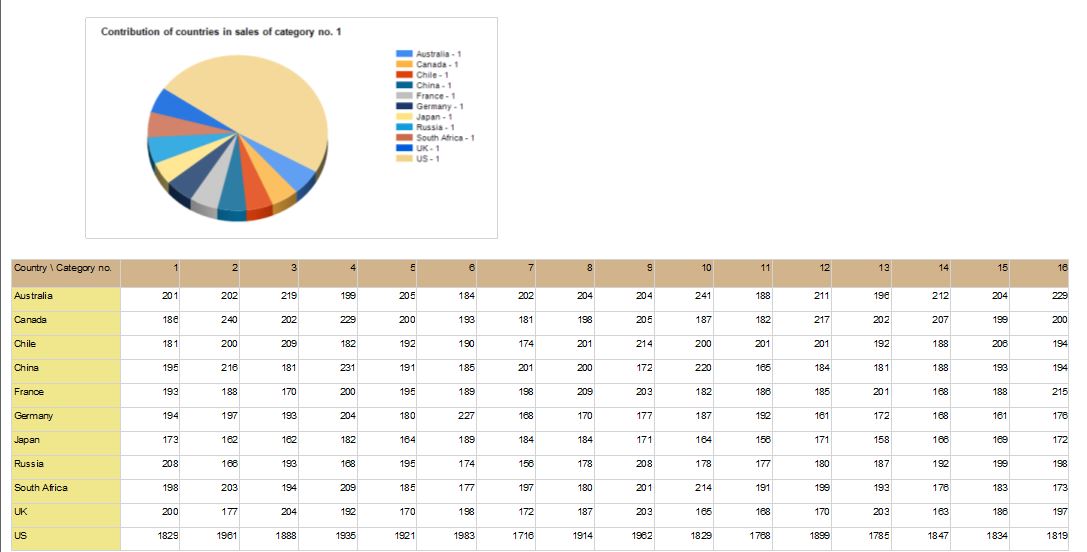
Tabela Liczba klientów w poszczególnych krajach z podziałem na miesiące.



Rysunek Liczba klientów w poszczególnych krajach z podziałem na miesiące.



Rysunek Całkowita miesięczna sprzedaż w poszczególnych miesiącach.



Rysunek Liczba produktów sprzedanych w poszczególnych krajach w danych kategoriach.

Powyższe rysunki przedstawiają raporty wygenerowane dzięki analizie danych. Dzięki nim możemy zorientować się jak dużo klientów miał nasz sklep w każdym z krajów na przestrzeni całego roku. Możemy również dowiedzieć się jaka była całkowita miesięczna sprzedaż. Oraz jak rozkładała się sprzedaż w poszczególnych krajach z podziałem na kategorie.