Hurtownie danych  
Laboratorium Czw 11:15  
  
Lista 2  
  
Kajetan Pynka 254495

**Zad 1.1**

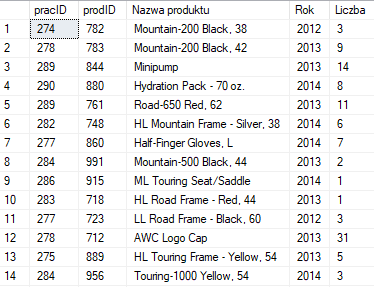
SELECT SOH.SalesPersonID "pracID", SOD.ProductID "prodID", P.Name "Nazwa produktu", Year(SOH.OrderDate) "Rok",

    COUNT(SOD.OrderQty) "Liczba" FROM Sales.SalesOrderHeader SOH

    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID=SOD.SalesOrderID

    JOIN Production.Product P ON P.ProductID=SOD.ProductID

    GROUP BY SOH.SalesPersonID, SOD.ProductID, P.Name, Year(SOH.OrderDate);

****

­**Wniosek:** Ciężko cokolwiek wywnioskować na podstawie powyższych wyników

**Zad 1.1a**

SELECT \* FROM (

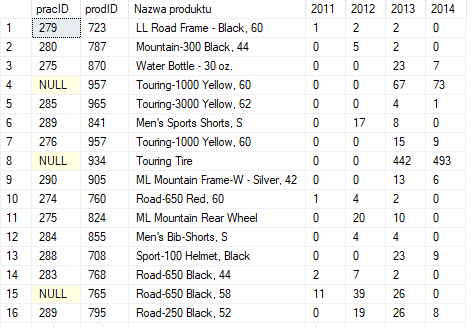
    SELECT SOH.SalesPersonID "pracID", SOD.ProductID "prodID", P.Name "Nazwa produktu", Year(SOH.OrderDate) "Rok",

        SOD.OrderQty "Liczba" FROM Sales.SalesOrderHeader SOH

        JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID=SOD.SalesOrderID

        JOIN Production.Product P ON P.ProductID=SOD.ProductID) S

    PIVOT(COUNT(Liczba) FOR S.Rok IN ([2011], [2012], [2013], [2014])) AS X;

****

**Wniosek:** W 2013 I 2014 bardzo dużo opon zostało zamówionych w ramach transakcji internetowych. Ogólnie ciężko skorelować pracowników z liczbą sprzedawanych produktów w poszczególnych latach na podstawie powyższych wyników.

**Zad 1.1b**

SELECT \* FROM (

SELECT SOH.SalesPersonID "pracID", Year(SOH.OrderDate) "Rok", P.ProductID "NrProd",

    SOD.OrderQty FROM Sales.SalesOrderHeader SOH

    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID=SOD.SalesOrderID

    JOIN Production.Product P ON P.ProductID=SOD.ProductID

    WHERE P.ProductID IN (SELECT ProductID FROM (SELECT DISTINCT TOP 5 SOD.ProductID, MAX(SOD.OrderQty) "Liczba"

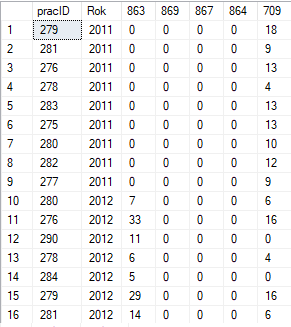
        FROM Sales.SalesOrderDetail SOD

        GROUP BY SOD.ProductID

        ORDER BY 2 DESC) AS Subquery)) S

    PIVOT(COUNT(OrderQty) FOR S.NrProd IN ([863], [869], [867], [864], [709])) X

    ORDER BY 2;

****

**Wniosek:** Produkt o identyfikatorze ‘709’ w zasadzie od początku historii transakcji (Czyli początku 2011 roku) był sprzedawany przez każdego sprzedawcę i to w dużej liczbie sztuk. W roku 2012 produkt o identyfikatorze ‘863’ zaczął się dobrze sprzedawać (możliwie dopiero wtedy został wprowadzony do oferty sklepu).

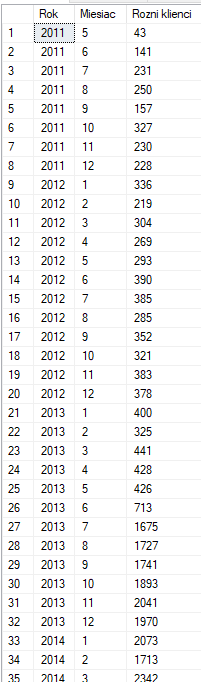
**Zad 1.2**

SELECT YEAR(SOH.OrderDate) "Rok", MONTH(SOH.OrderDate) "Miesiac", COUNT(DISTINCT SOH.CustomerID) "Rozni klienci"

    FROM Sales.SalesOrderHeader SOH

    GROUP BY YEAR(SOH.OrderDate), MONTH(SOH.OrderDate)

    ORDER BY 1, 2;

****

**Wniosek:** Wraz z upływem miesięcy i lat możemy zaobserwować wzorst liczby unikalnych klientów kupujących produkty w sklepie. Na tę liczbę składają się stali klienci oraz nowi klienci.

**Zad 1.2b**

SELECT \* FROM (

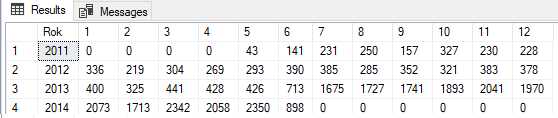
    SELECT DISTINCT YEAR(OrderDate) "Rok", MONTH(OrderDate) "Miesiac", CustomerID "Rozni klienci"

        FROM Sales.SalesOrderHeader

    ) S PIVOT(COUNT([Rozni klienci])

    FOR S.[Miesiac] IN ([1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12])) X

    ORDER BY 1;

****

**Wniosek:** Podobnie do poprzedniego zadania natomiast w bardziej przejrzysty sposób. Aż do czerwca 2014 roku rosła liczba unikalnych klientów kupujących towary (Zdarzały się pewne fluktuacje natomiast ogólnie cały czas następował wzrost).

**Zad 1.3**

SELECT \* FROM (

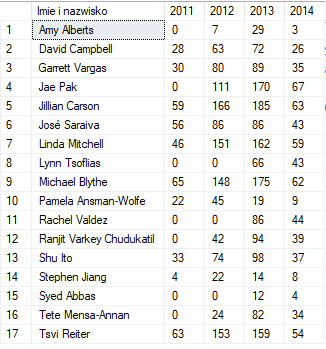
    SELECT Per.FirstName + ' ' + Per.LastName "Imie i nazwisko", YEAR(SOH.OrderDate) "Rok", SOH.SalesOrderID

        FROM Sales.SalesOrderHeader SOH

        JOIN Sales.SalesPerson SP ON SOH.SalesPersonID=SP.BusinessEntityID

        JOIN Person.Person Per ON Per.BusinessEntityID=SP.BusinessEntityID

    ) S PIVOT(COUNT(SalesOrderID) FOR S.Rok IN ([2011], [2012], [2013], [2014])) X;

****

**Wniosek:** Sprzedawcy numer 8,11,15 mogli zacząć pracę w 2013 roku. Ogólnie liczba sprzedawanych produktów może zależeć od regionu, w którym pracują sprzedawcy. Między 2011-2013 prawie każdy sprzedawca zanotował wzrost liczby sprzedawanych produktów.

**Zad 1.4**

SELECT YEAR(SOH.OrderDate) "Rok", MONTH(SOH.OrderDate) "Miesiąc", DAY(SOH.OrderDate) "Dzień",

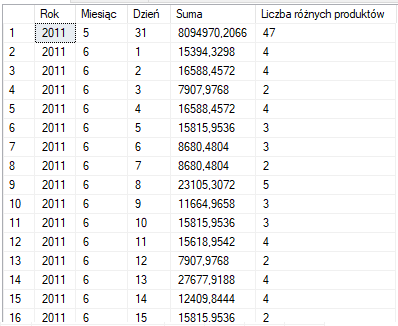
    SUM(SOH.TotalDue) "Suma", COUNT(DISTINCT SOD.ProductID) "Liczba różnych produktów"

    FROM Sales.SalesOrderHeader SOH

    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID=SOD.SalesOrderID

    GROUP BY YEAR(SOH.OrderDate), MONTH(SOH.OrderDate), DAY(SOH.OrderDate)

    ORDER BY 1, 2, 3;

****

**Wniosek:** 31.05 musiał katalogować produkty z wielu dni, ponieważ dość wyraźnie odstaje od reszty następujących po nim dni. Ogólnie jest bardzo dużo rekordów i warto by się posłużyć tabelą przestawną by móc wysnuć lepsze wnioski.

**Zad 1.5**

SELECT CASE

        WHEN MONTH(SOH.OrderDate)=1 THEN 'Styczeń'

        WHEN MONTH(SOH.OrderDate)=2 THEN 'Luty'

        WHEN MONTH(SOH.OrderDate)=3 THEN 'Marzec'

        WHEN MONTH(SOH.OrderDate)=4 THEN 'Kwiecień'

        WHEN MONTH(SOH.OrderDate)=5 THEN 'Maj'

        WHEN MONTH(SOH.OrderDate)=6 THEN 'Czerwiec'

        WHEN MONTH(SOH.OrderDate)=7 THEN 'Lipiec'

        WHEN MONTH(SOH.OrderDate)=8 THEN 'Sierpień'

        WHEN MONTH(SOH.OrderDate)=9 THEN 'Wrzesień'

        WHEN MONTH(SOH.OrderDate)=10 THEN 'Październik'

        WHEN MONTH(SOH.OrderDate)=11 THEN 'Listopad'

        WHEN MONTH(SOH.OrderDate)=12 THEN 'Grudzień'

    END "Miesiąc", SUM(SOH.SubTotal) "Suma",

    COUNT(DISTINCT SOD.ProductID) "Liczba różnych produktów"

    FROM Sales.SalesOrderHeader SOH

    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID=SOD.SalesOrderID

    GROUP BY MONTH(SOH.OrderDate) ORDER BY MONTH(SOH.OrderDate);

****

**Wniosek:** Miesiąc wydaje się nie mieć znaczenia dla liczby sprzedawanych różnych produktów, choć minimalnie większe liczby możemy zaobserwować w okresie letnim.

**Zad 1.5b**

SELECT CASE

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=1 THEN 'Niedziela'

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=2 THEN 'Poniedziałek'

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=3 THEN 'Wtorek'

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=4 THEN 'Środa'

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=5 THEN 'Czwartek'

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=6 THEN 'Piątek'

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=7 THEN 'Sobota'

    END "Dzień tygodnia", SUM(SOH.SubTotal) "Suma",

    COUNT(DISTINCT SOD.ProductID) "Liczba różnych produktów"

    FROM Sales.SalesOrderHeader SOH

    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID=SOD.SalesOrderID

    GROUP BY DATEPART(WEEKDAY, SOH.OrderDate) ORDER BY CASE

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=1 THEN 7

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=2 THEN 1

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=3 THEN 2

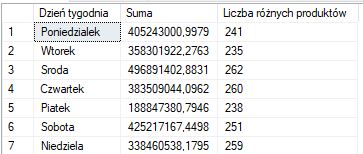
        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=4 THEN 3

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=5 THEN 4

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=6 THEN 5

        WHEN DATEPART(WEEKDAY, SOH.OrderDate)=7 THEN 6

    END;

****

**Wniosek:** Dzień tygodnia wydaje się nie być w żaden sposób skorelowany z z liczbą sprzedawanych różnych produktów, natomiast w piątek możemy zaobserwować znacząco niższą sumę pieniędzy wydanych na zakupy.

**Zad 1.6**

SELECT \* FROM (

SELECT PER.FirstName "Imie", PER.LastName "Nazwisko", COUNT(DISTINCT SOH.SalesOrderID) "Liczba",

    SUM(DISTINCT SOH.TotalDue) "Kwota", CASE

        WHEN (SELECT TOP 1 COUNT(DISTINCT SOH\_2.SalesOrderID)

                FROM Sales.Customer C\_2

                JOIN Sales.SalesOrderHeader SOH\_2 ON SOH\_2.CustomerID=C\_2.CustomerID

                WHERE C\_2.CustomerID=C.CustomerID

                GROUP BY YEAR(SOH\_2.DueDate)

                ORDER BY COUNT(DISTINCT SOH\_2.SalesOrderID)

            ) >= 2 AND (SELECT COUNT(\*) FROM (SELECT COUNT(DISTINCT SOH\_2.SalesOrderID) "test123"

                FROM Sales.Customer C\_2

                JOIN Sales.SalesOrderHeader SOH\_2 ON SOH\_2.CustomerID=C\_2.CustomerID

                WHERE C\_2.CustomerID=C.CustomerID

                GROUP BY YEAR(SOH\_2.DueDate)) something)=4

            AND (SELECT TOP 1 COUNT(DISTINCT SOH\_2.SalesOrderID) FROM Sales.SalesOrderHeader SOH\_2

                    WHERE SOH\_2.CustomerID=C.CustomerID AND

                        SOH\_2.TotalDue > 1.5 \* (SELECT AVG(SOH\_3.TotalDue) FROM Sales.SalesOrderHeader SOH\_3)

                    GROUP BY YEAR(SOH\_2.DueDate) ORDER BY 1) >=  2

            THEN 'Platynowa'

        WHEN (SELECT COUNT(DISTINCT SOH\_2.SalesOrderID) FROM Sales.SalesOrderHeader SOH\_2

                    WHERE SOH\_2.CustomerID=C.CustomerID AND SOH\_2.TotalDue > 1.5 \*

                        (SELECT AVG(SOH\_3.TotalDue) FROM Sales.SalesOrderHeader SOH\_3)) >= 2

            THEN 'Złota'

        WHEN COUNT(DISTINCT SOH.SalesOrderID) >= 5 THEN 'Srebrna'

    END "Karta" FROM Sales.Customer C

    JOIN Sales.SalesOrderHeader SOH ON SOH.CustomerID=C.CustomerID

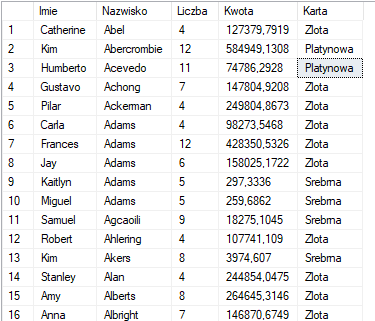
    JOIN Person.Person PER ON PER.BusinessEntityID=C.PersonID

    JOIN Sales.SalesOrderDetail SOD ON SOD.SalesOrderID=SOH.SalesOrderID

    GROUP BY PER.FirstName, PER.LastName, C.CustomerID) MAIN

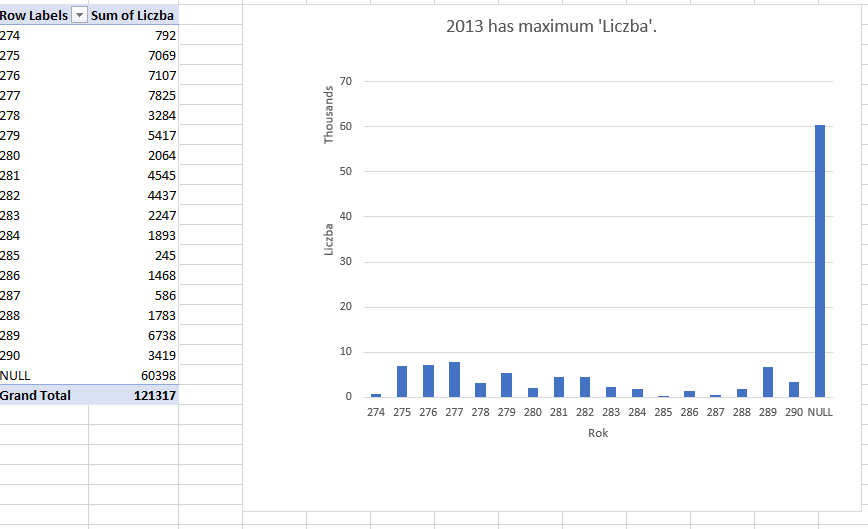
    WHERE Karta IS NOT NULL

    ORDER BY 2, 1;

****

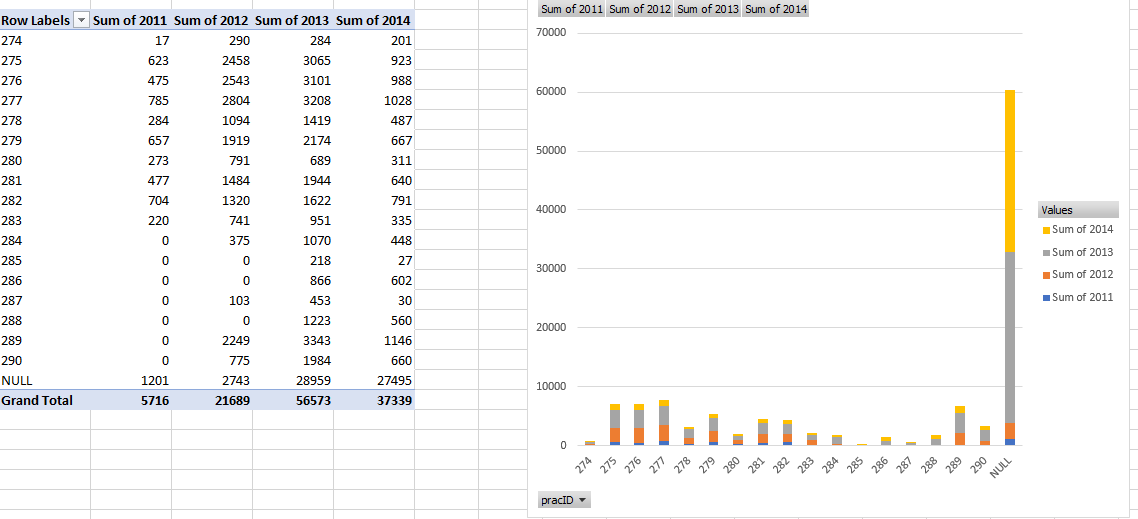
**Wniosek:** Bardzo duża liczba klientów posiada złotą kartę

**Zad 2.1.1**

****

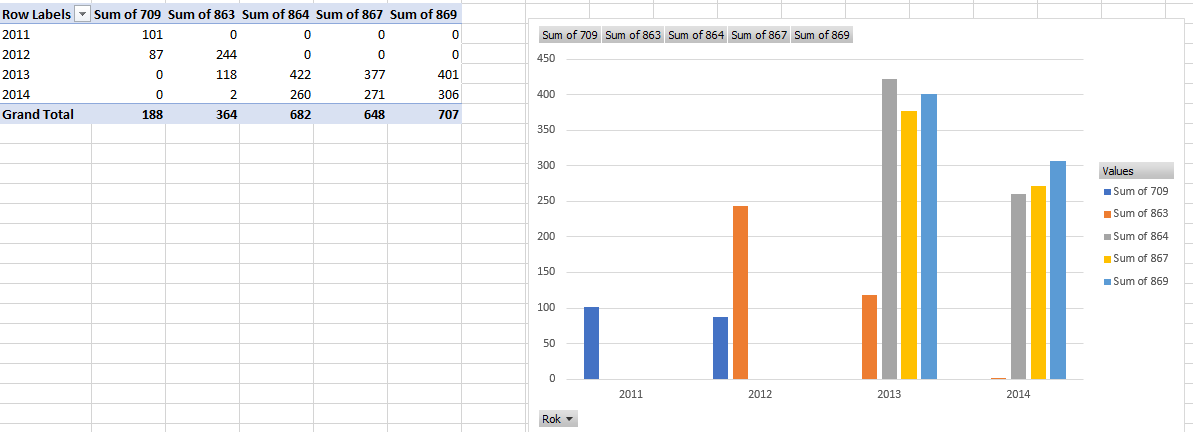
**Wniosek:** W zasadzie połowa produktów została sprzedana przez internet.

**Zad 2.1.1a**

****

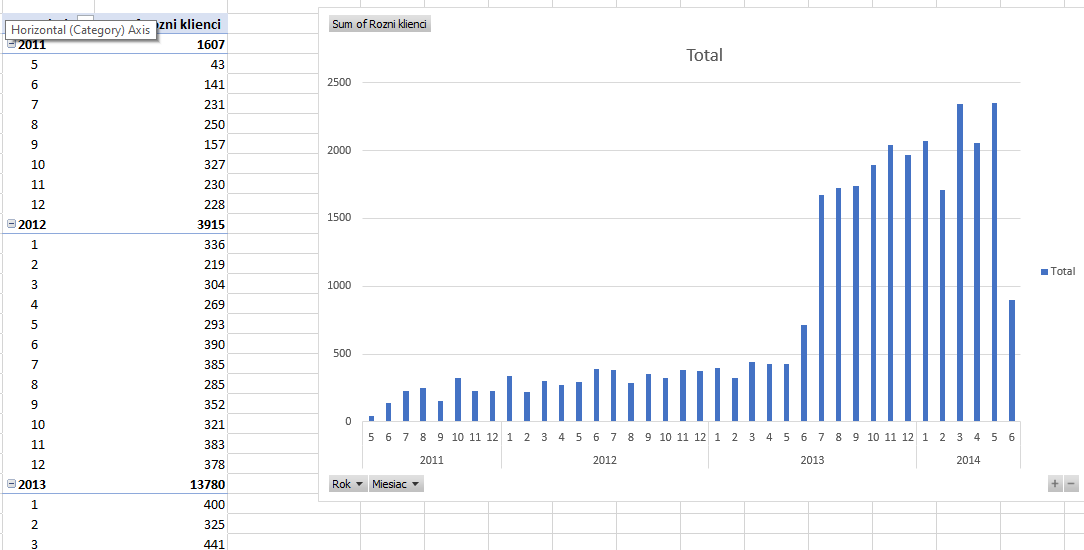
**Wniosek:** Zdecydowanie najmniej produktów zakupionych zostało w 2011 roku. W 2012 i 2013 roku sprzedaż fizyczna była wyższa niż w 2014 roku. Sprzedaż internetowa natomiast znacząco się zwiększyła w 2013 roku i pierwsza połowa 2014 roku osiągnęła taki wynik jak całość 2013r.

**Zad 2.1.1b**

****

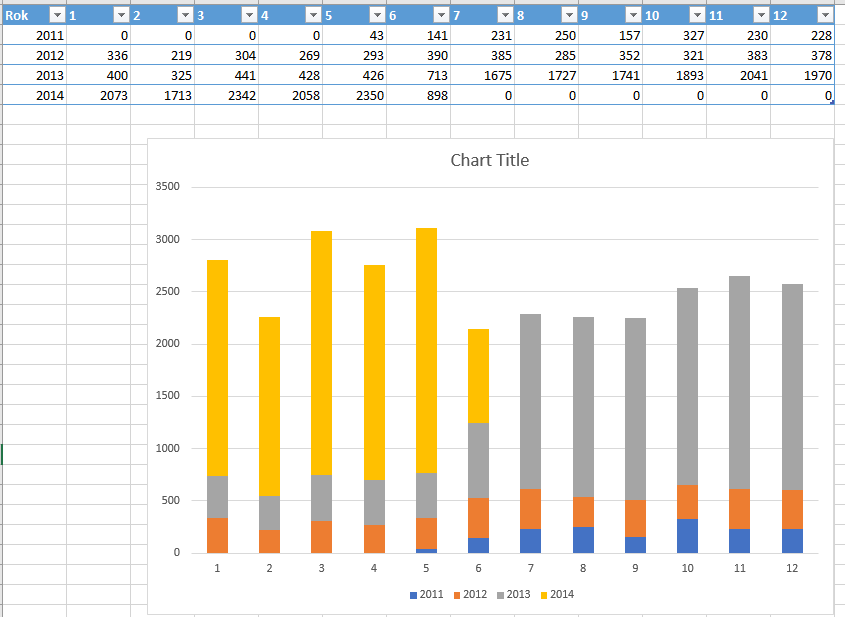
**Wniosek:** Produkt ‘709’ był sprzedawany tylko w 2011 i 2012 roku, natomiast produkt ‘863’ w 2012 i 2013 roku. Pozostałe 3 produkty były sprzedawane w bardzo dużych liczbach w 2013 i 2014 roku.

**Zad 2.1.2**

****

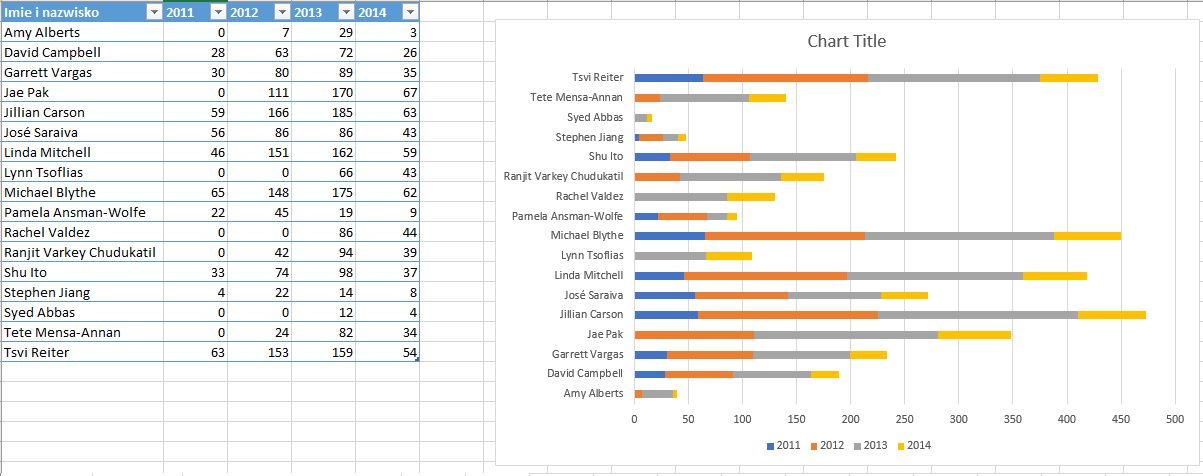
**Wniosek:** Ogólnie panuje tendencja wzrostowa liczby różnych klientów kupujących produkty w sklepie. Definitywny przełom nastąpił w połowie 2013 roku gdy liczba ta zwiększyła się dwukrotnie.

**Zad 2.1.2b**

****

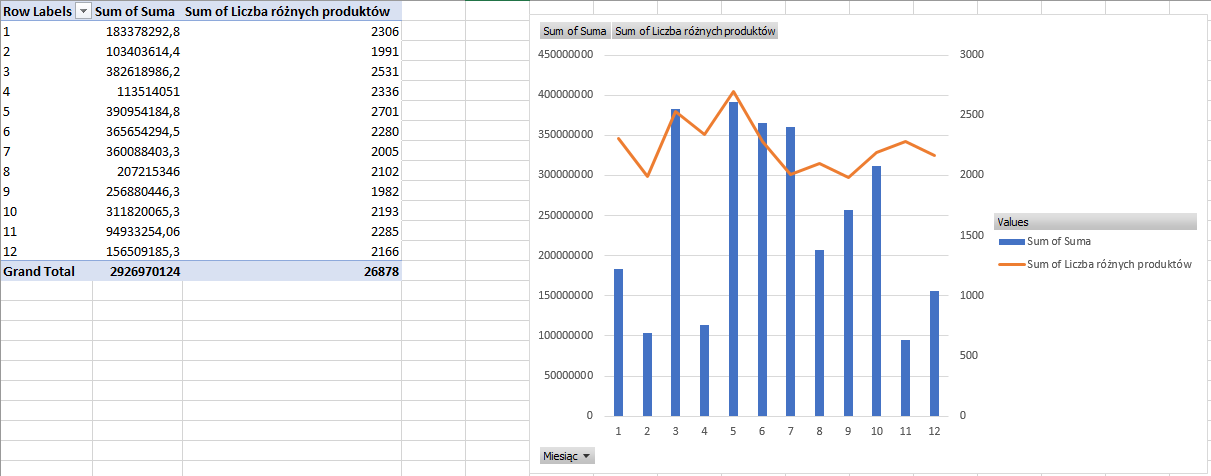
**Wniosek:** W zasadzie jak poprzednio tylko w innej formie wizualnej.

**Zad 2.1.3**

****

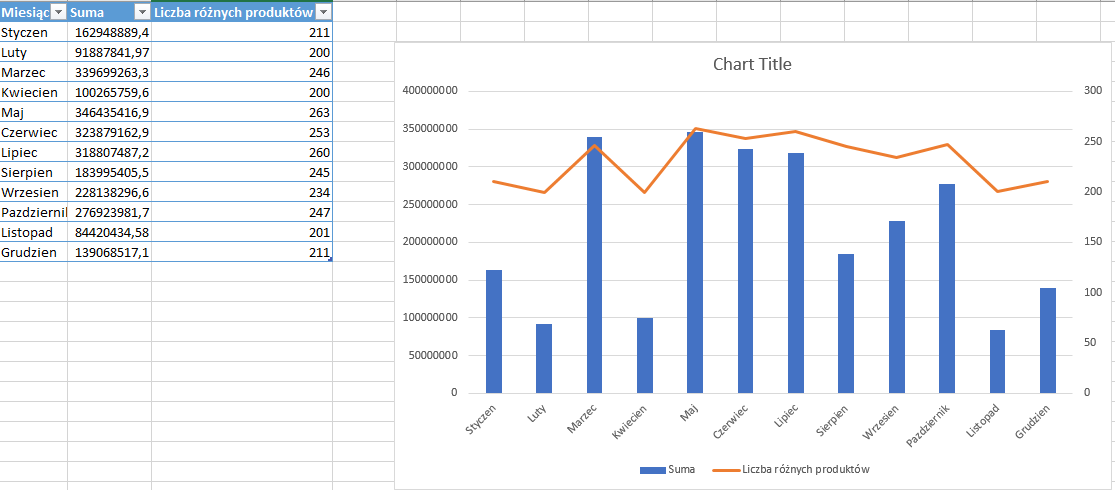
**Wniosek:** Ogólnie sprzedawcy z upływem lat zaczęli sprzedawać coraz więcej produktów. Różnice pomiędzy nimi wynikają prawdopodobnie z regionu, w którym pracują.

**Zad 2.1.4**

****

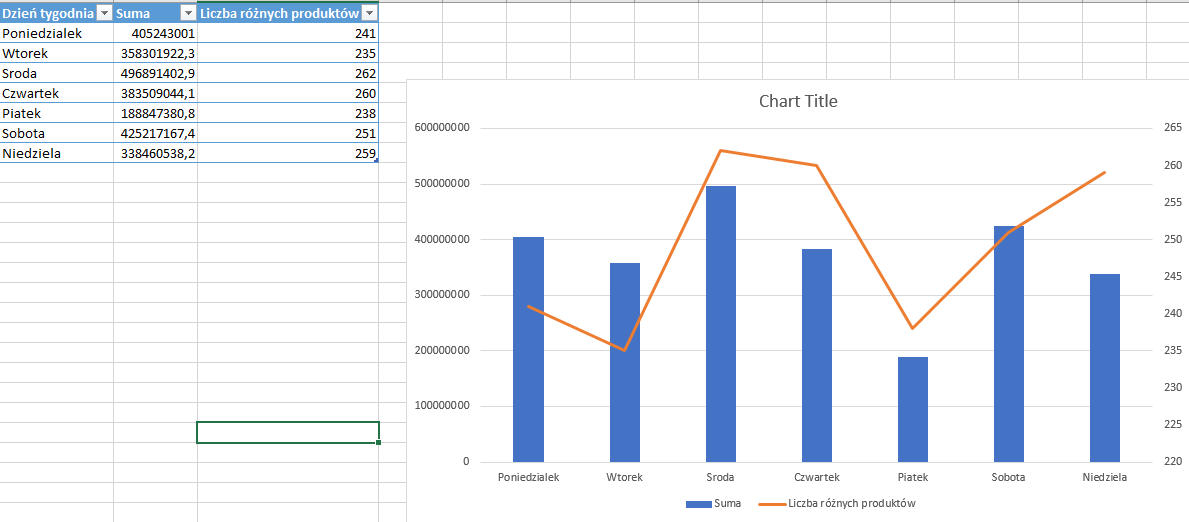
**Wniosek:** Najwięcej pieniędzy na różne produkty klienci przeznaczają w okresie wakacyjnym. Liczba sztuk sprzedawanych różnych produktów wydaje się być natomiast najwyższa na wiosnę.

**Zad 2.1.5**

****

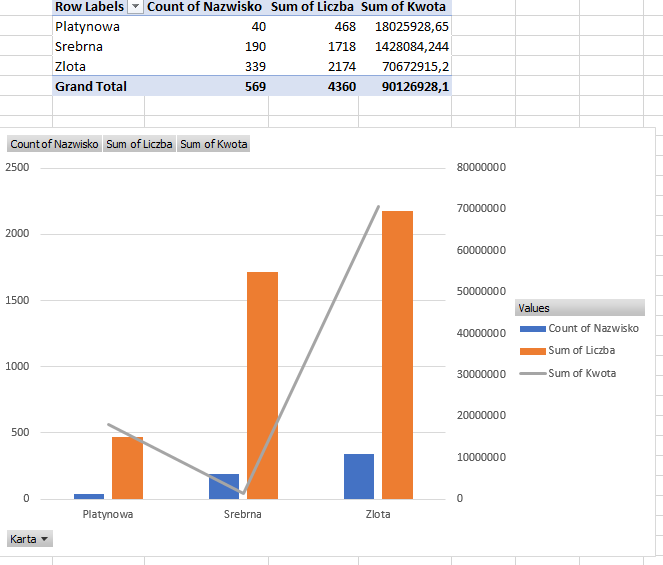
**Wniosek:** Biorąc pod uwagę wszystkie lata to suma wydawanych przez klientów pieniędzy nadal wydaje się być najwyższa w lecie. Liczba rożnych produktów wydaje się nie mieć uzależnienia od miesiąca (ewentualnie jest najniższa zimą).

**Zad 2.1.5b**

****

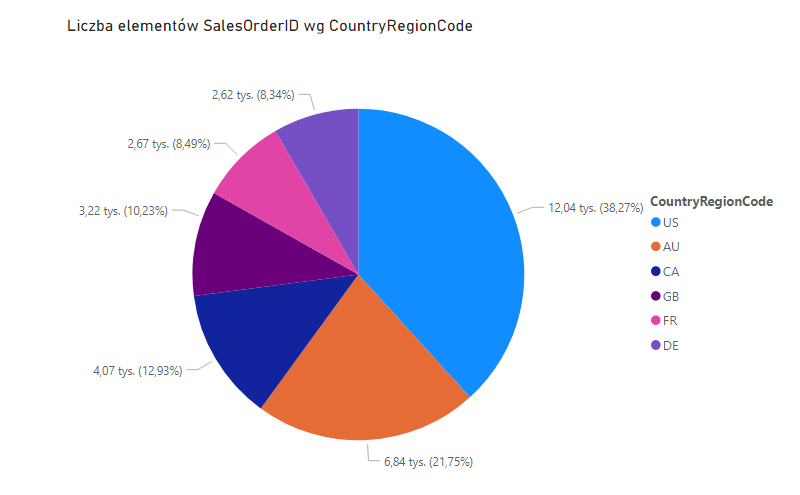
**Wniosek:** W piątek klienci wydają najmniej na różne produkty. Najwięcej różnych produktów klienci wydają się kupować w środę, czwartek i niedzielę.

**Zad 2.1.6**

****

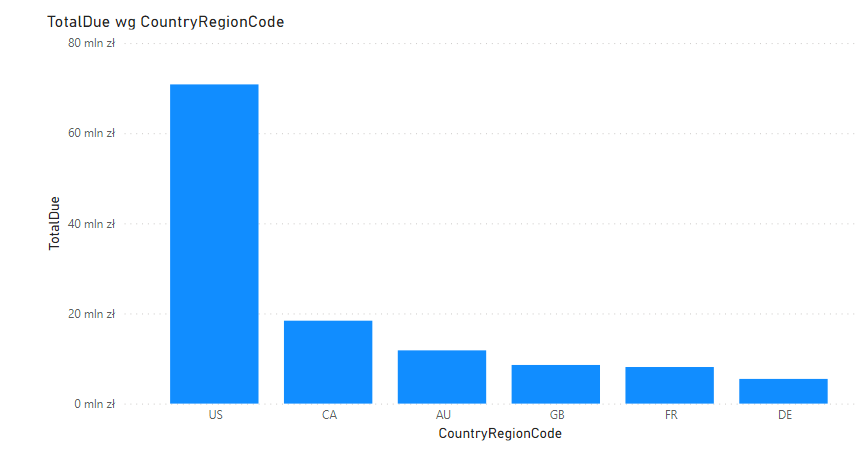
**Wniosek:** Ze wszystkich kart najczęstsza jest złota a najrzadsza platynowa. Najwięcej transakcji zostało zanotowanych dla klientów ze srebrną i złotą kartą. Zdecydowanie najwięcej na transakcjach wydali klienci ze złotą kartą, a najmniej klienci ze srebrną.

**Zad 2.2.1**

****

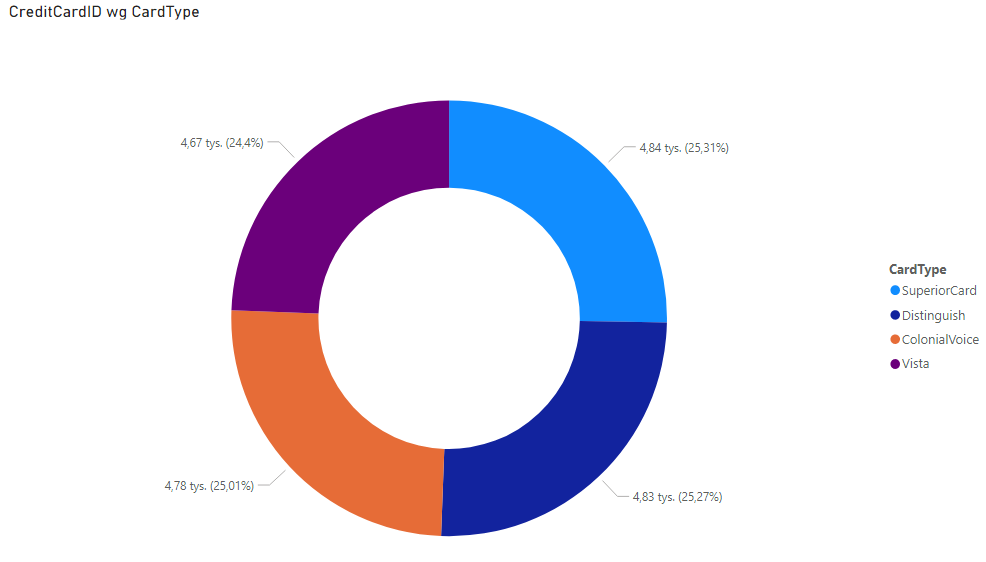
**Wniosek:** Najwięcej transakcji wykonanych zostało w Stanach Zjednoczonych i Australii.

**Zad 2.2.2**

****

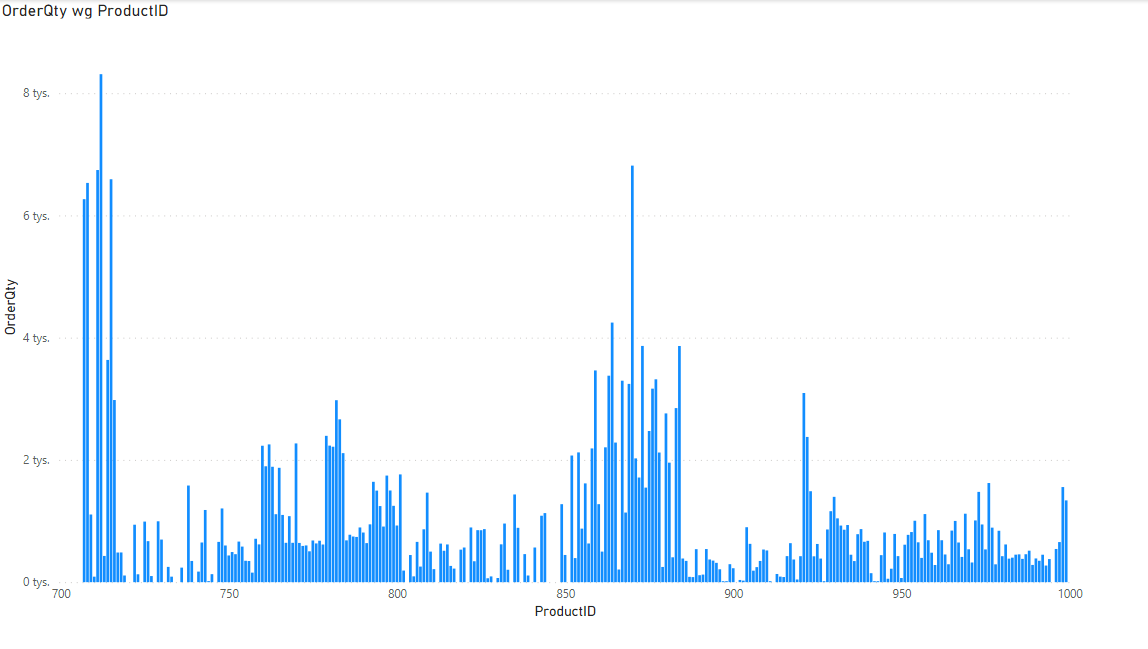
**Wniosek:** Największa suma z transakcji pochodzi ze Stanów Zjednoczonych natomiast na drugim miejscu jest Kanada (wydają więcej niż Australijczycy).

**Zad 2.2.3**

****

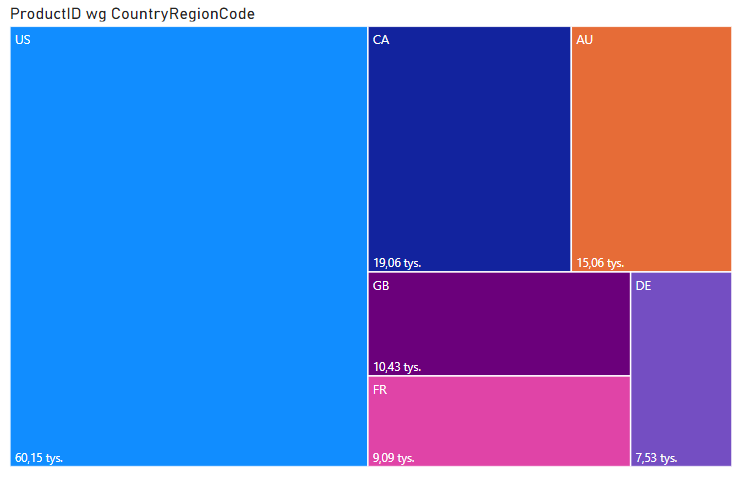
**Wniosek:** Wszystkie rodzaje kart kredytowych są równie popularne.

**Zad 2.2.4**

****

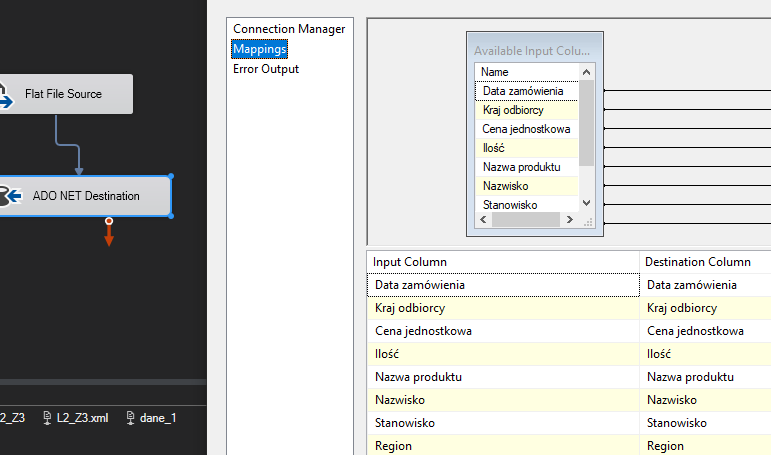
**Wniosek:** Produkty o identyfikatorach między 700-710 oraz 850-900 są sprzedawane w największych liczbach sztuk.

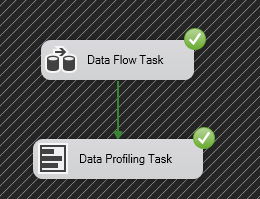
**Zad 2.2.5**

****

**Wniosek:** Najwięcej produktów sprzedanych zostało w Stanach Zjednoczonych, natomiast najmniej w Niemczech.

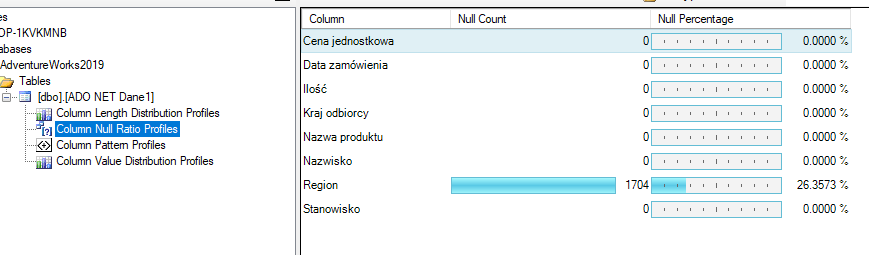
**Zad 3.1**

****

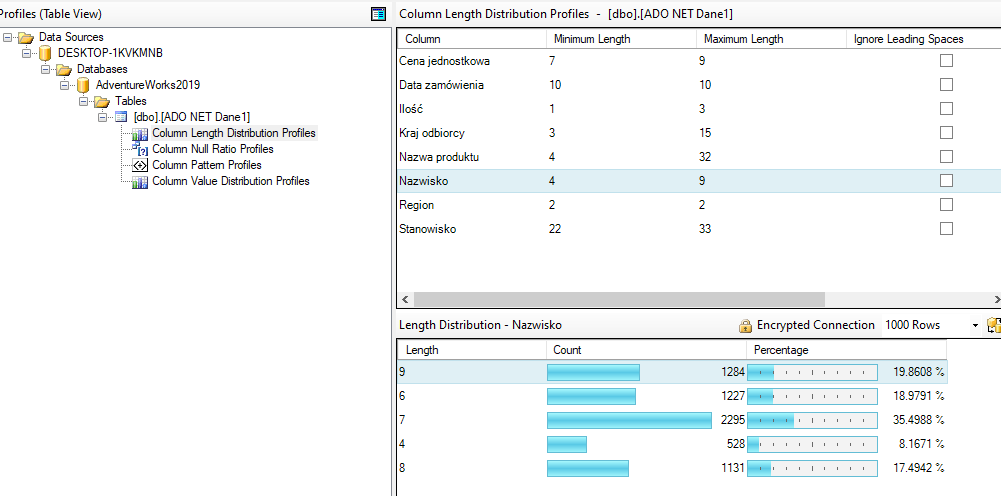
****

**Wniosek:** Profilowanie przebiegło pomyślnie dla pliku .txt oraz .csv.

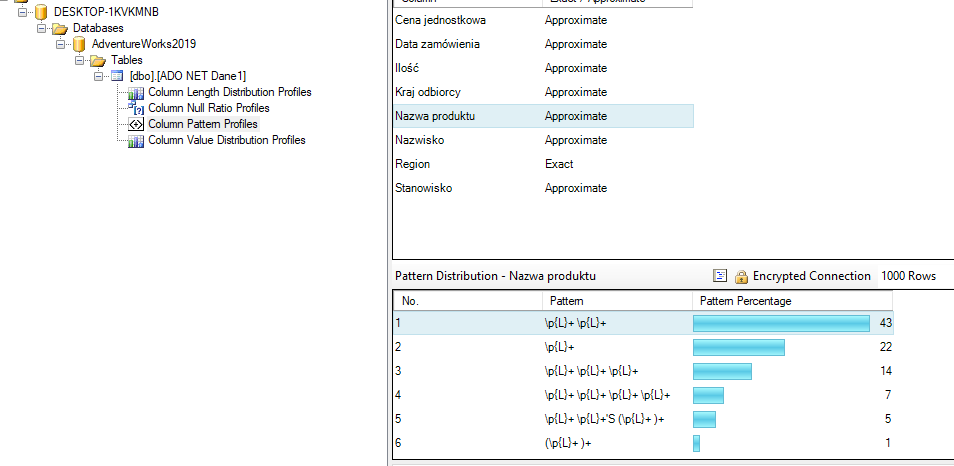
**Zad 3.2**



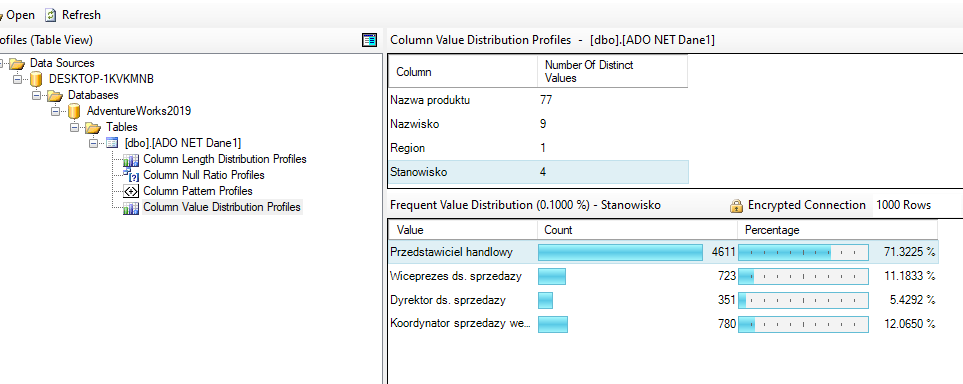
**Wniosek:** Region jest opcjonalny natomiast nie ma aż tak dużo wartości NULL.

****

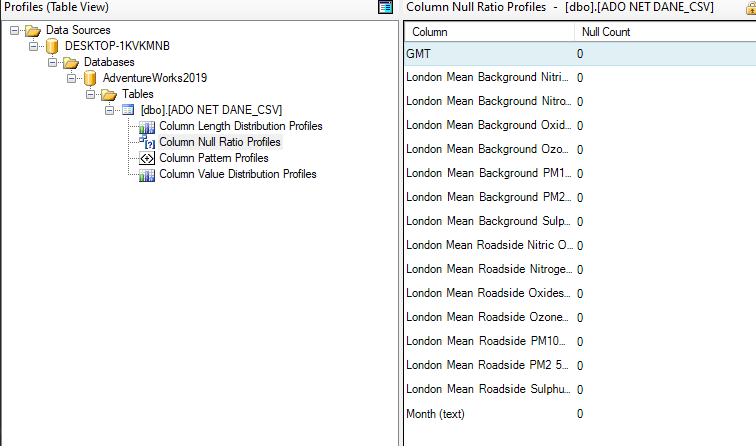
**Wniosek:** Najmniej krótkich nazwisk ale ograniczenie długości ma sens bo widać, że trochę jest osób z maksymalną dopuszczalną długością.



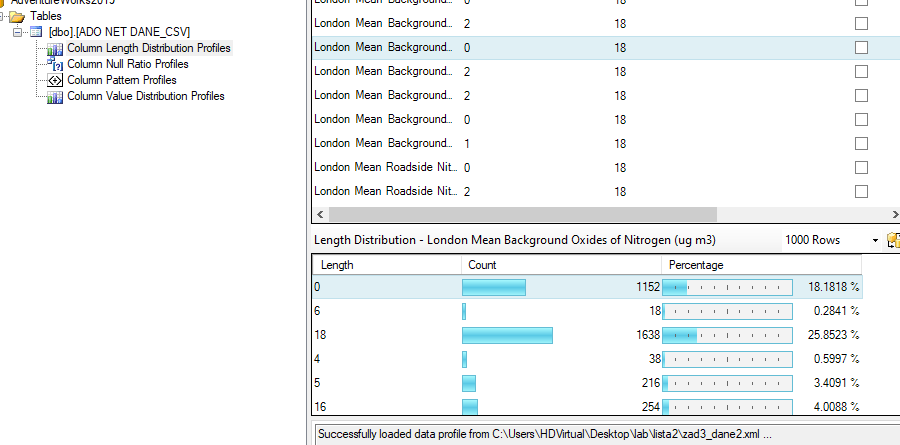
**Wniosek:** Większość produktów ma dwuwyrazowe nazwy (z użyciem liter polskiego alfabetu).



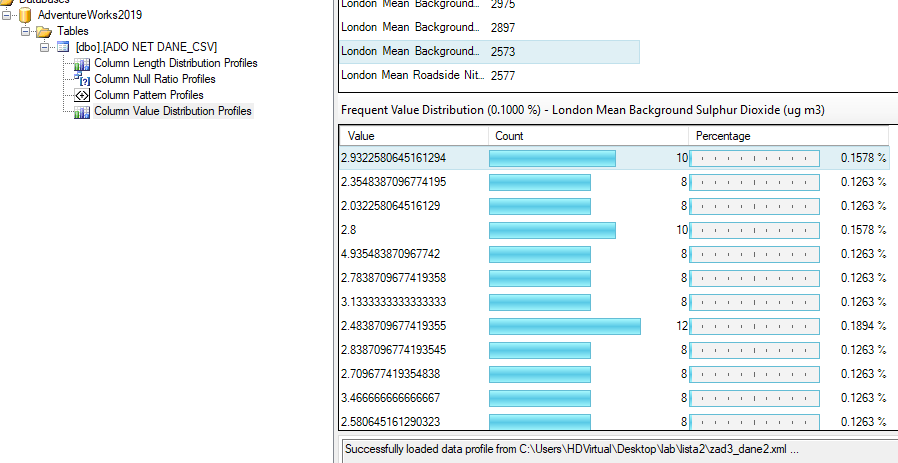
**Wniosek:** Większość pracowników jest na stanowisku przedstawiciela handlowego.



**Wniosek:** Wszystkie rekordy są w pełni zapełnione (tzn. nie występuje wartość NULL – brak opcjonalnych pól).



**Wniosek:** Długość zapisu tlenków wodoru to najczęściej albo 0 albo 18 znaków.



**Wniosek:** Brak wartości dominującej

**Wnioski:**

* Polecenia ‘CASE’ oraz ‘PIVOT’ w ramach SQL’a pozwalają na uzyskanie tabel przestawnych, które w bardzo czytelny sposób przedstawiają zależności między danymi w naszej hurtowni
* Narzędzia takie jak PowerBI, Tableu czy też Excel zapewniają możliwość wygenerowania wielu rodzajów wykresów. Pozwala to przedstawić pewne trendy zachodzące wśród danych w bardzo obrazowy i przejrzysty sposób
* SQL Server Integration Services (SSIS) pozwala wyprofilować nasze źródło danych pod wieloma względami. Zapewnia to unikalne spojrzenie na strukturę naszych danych i pozwala przemyśleć czy aktualna architektura na pewno jest sensowna (np. ze względu na procentowo duży udział wartości NULL w ramach jakiejś kolumny LUB ze względu na procentowy udział długości danych)