

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);
```

| | pracID | prodID | Nazwa produktu | Rok | Liczba |
|----|--------|--------|--------------------------------|------|--------|
| 1 | 274 | 782 | Mountain-200 Black, 38 | 2012 | 3 |
| 2 | 278 | 783 | Mountain-200 Black, 42 | 2013 | 9 |
| 3 | 289 | 844 | Minipump | 2013 | 14 |
| 4 | 290 | 880 | Hydration Pack - 70 oz. | 2014 | 8 |
| 5 | 289 | 761 | Road-650 Red, 62 | 2013 | 11 |
| 6 | 282 | 748 | HL Mountain Frame - Silver, 38 | 2014 | 6 |
| 7 | 277 | 860 | Half-Finger Gloves, L | 2014 | 7 |
| 8 | 284 | 991 | Mountain-500 Black, 44 | 2013 | 2 |
| 9 | 286 | 915 | ML Touring Seat/Saddle | 2014 | 1 |
| 10 | 283 | 718 | HL Road Frame - Red, 44 | 2013 | 1 |
| 11 | 277 | 723 | LL Road Frame - Black, 60 | 2012 | 3 |
| 12 | 278 | 712 | AWC Logo Cap | 2013 | 31 |
| 13 | 275 | 889 | HL Touring Frame - Yellow, 54 | 2013 | 5 |
| 14 | 284 | 956 | Touring-1000 Yellow, 54 | 2014 | 3 |

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;
```

| | pracID | prodID | Nazwa produktu | 2011 | 2012 | 2013 | 2014 |
|----|--------|--------|----------------------------------|------|------|------|------|
| 1 | 279 | 723 | LL Road Frame - Black, 60 | 1 | 2 | 2 | 0 |
| 2 | 280 | 787 | Mountain-300 Black, 44 | 0 | 5 | 2 | 0 |
| 3 | 275 | 870 | Water Bottle - 30 oz. | 0 | 0 | 23 | 7 |
| 4 | NULL | 957 | Touring-1000 Yellow, 60 | 0 | 0 | 67 | 73 |
| 5 | 285 | 965 | Touring-3000 Yellow, 62 | 0 | 0 | 4 | 1 |
| 6 | 289 | 841 | Men's Sports Shorts, S | 0 | 17 | 8 | 0 |
| 7 | 276 | 957 | Touring-1000 Yellow, 60 | 0 | 0 | 15 | 9 |
| 8 | NULL | 934 | Touring Tire | 0 | 0 | 442 | 493 |
| 9 | 290 | 905 | ML Mountain Frame-W - Silver, 42 | 0 | 0 | 13 | 6 |
| 10 | 274 | 760 | Road-650 Red, 60 | 1 | 4 | 2 | 0 |
| 11 | 275 | 824 | ML Mountain Rear Wheel | 0 | 20 | 10 | 0 |
| 12 | 284 | 855 | Men's Bib-Shorts, S | 0 | 4 | 4 | 0 |
| 13 | 288 | 708 | Sport-100 Helmet, Black | 0 | 0 | 23 | 9 |
| 14 | 283 | 768 | Road-650 Black, 44 | 2 | 7 | 2 | 0 |
| 15 | NULL | 765 | Road-650 Black, 58 | 11 | 39 | 26 | 0 |
| 16 | 289 | 795 | Road-250 Black, 52 | 0 | 19 | 26 | 8 |

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;
```

| | pracID | Rok | 863 | 869 | 867 | 864 | 709 |
|----|--------|------|-----|-----|-----|-----|-----|
| 1 | 279 | 2011 | 0 | 0 | 0 | 0 | 18 |
| 2 | 281 | 2011 | 0 | 0 | 0 | 0 | 9 |
| 3 | 276 | 2011 | 0 | 0 | 0 | 0 | 13 |
| 4 | 278 | 2011 | 0 | 0 | 0 | 0 | 4 |
| 5 | 283 | 2011 | 0 | 0 | 0 | 0 | 13 |
| 6 | 275 | 2011 | 0 | 0 | 0 | 0 | 13 |
| 7 | 280 | 2011 | 0 | 0 | 0 | 0 | 10 |
| 8 | 282 | 2011 | 0 | 0 | 0 | 0 | 12 |
| 9 | 277 | 2011 | 0 | 0 | 0 | 0 | 9 |
| 10 | 280 | 2012 | 7 | 0 | 0 | 0 | 6 |
| 11 | 276 | 2012 | 33 | 0 | 0 | 0 | 16 |
| 12 | 290 | 2012 | 11 | 0 | 0 | 0 | 0 |
| 13 | 278 | 2012 | 6 | 0 | 0 | 0 | 4 |
| 14 | 284 | 2012 | 5 | 0 | 0 | 0 | 0 |
| 15 | 279 | 2012 | 29 | 0 | 0 | 0 | 16 |
| 16 | 281 | 2012 | 14 | 0 | 0 | 0 | 6 |

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;
```

| | Rok | Miesiac | Rozni klienci |
|----|------|---------|---------------|
| 1 | 2011 | 5 | 43 |
| 2 | 2011 | 6 | 141 |
| 3 | 2011 | 7 | 231 |
| 4 | 2011 | 8 | 250 |
| 5 | 2011 | 9 | 157 |
| 6 | 2011 | 10 | 327 |
| 7 | 2011 | 11 | 230 |
| 8 | 2011 | 12 | 228 |
| 9 | 2012 | 1 | 336 |
| 10 | 2012 | 2 | 219 |
| 11 | 2012 | 3 | 304 |
| 12 | 2012 | 4 | 269 |
| 13 | 2012 | 5 | 293 |
| 14 | 2012 | 6 | 390 |
| 15 | 2012 | 7 | 385 |
| 16 | 2012 | 8 | 285 |
| 17 | 2012 | 9 | 352 |
| 18 | 2012 | 10 | 321 |
| 19 | 2012 | 11 | 383 |
| 20 | 2012 | 12 | 378 |
| 21 | 2013 | 1 | 400 |
| 22 | 2013 | 2 | 325 |
| 23 | 2013 | 3 | 441 |
| 24 | 2013 | 4 | 428 |
| 25 | 2013 | 5 | 426 |
| 26 | 2013 | 6 | 713 |
| 27 | 2013 | 7 | 1675 |
| 28 | 2013 | 8 | 1727 |
| 29 | 2013 | 9 | 1741 |
| 30 | 2013 | 10 | 1893 |
| 31 | 2013 | 11 | 2041 |
| 32 | 2013 | 12 | 1970 |
| 33 | 2014 | 1 | 2073 |
| 34 | 2014 | 2 | 1713 |
| 35 | 2014 | 3 | 2342 |

Wniosek: Wraz z upływem miesięcy i lat możemy zaobserwować wzrost 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;
```

| | Rok | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|------|------|------|------|------|------|-----|------|------|------|------|------|------|
| 1 | 2011 | 0 | 0 | 0 | 0 | 43 | 141 | 231 | 250 | 157 | 327 | 230 | 228 |
| 2 | 2012 | 336 | 219 | 304 | 269 | 293 | 390 | 385 | 285 | 352 | 321 | 383 | 378 |
| 3 | 2013 | 400 | 325 | 441 | 428 | 426 | 713 | 1675 | 1727 | 1741 | 1893 | 2041 | 1970 |
| 4 | 2014 | 2073 | 1713 | 2342 | 2058 | 2350 | 898 | 0 | 0 | 0 | 0 | 0 | 0 |

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;
```

| | Imie i nazwisko | 2011 | 2012 | 2013 | 2014 |
|----|--------------------------|------|------|------|------|
| 1 | Amy Alberts | 0 | 7 | 29 | 3 |
| 2 | David Campbell | 28 | 63 | 72 | 26 |
| 3 | Garrett Vargas | 30 | 80 | 89 | 35 |
| 4 | Jae Pak | 0 | 111 | 170 | 67 |
| 5 | Jillian Carson | 59 | 166 | 185 | 63 |
| 6 | José Saraiva | 56 | 86 | 86 | 43 |
| 7 | Linda Mitchell | 46 | 151 | 162 | 59 |
| 8 | Lynn Tsoulias | 0 | 0 | 66 | 43 |
| 9 | Michael Blythe | 65 | 148 | 175 | 62 |
| 10 | Pamela Ansman-Wolfe | 22 | 45 | 19 | 9 |
| 11 | Rachel Valdez | 0 | 0 | 86 | 44 |
| 12 | Ranjit Varkey Chudukatil | 0 | 42 | 94 | 39 |
| 13 | Shu Ito | 33 | 74 | 98 | 37 |
| 14 | Stephen Jiang | 4 | 22 | 14 | 8 |
| 15 | Syed Abbas | 0 | 0 | 12 | 4 |
| 16 | Tete Mensa-Annan | 0 | 24 | 82 | 34 |
| 17 | Tsvi Reiter | 63 | 153 | 159 | 54 |

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;
```

| | Rok | Miesiąc | Dzień | Suma | Liczba różnych produktów |
|----|------|---------|-------|--------------|--------------------------|
| 1 | 2011 | 5 | 31 | 8094970,2066 | 47 |
| 2 | 2011 | 6 | 1 | 15394,3298 | 4 |
| 3 | 2011 | 6 | 2 | 16588,4572 | 4 |
| 4 | 2011 | 6 | 3 | 7907,9768 | 2 |
| 5 | 2011 | 6 | 4 | 16588,4572 | 4 |
| 6 | 2011 | 6 | 5 | 15815,9536 | 3 |
| 7 | 2011 | 6 | 6 | 8680,4804 | 3 |
| 8 | 2011 | 6 | 7 | 8680,4804 | 2 |
| 9 | 2011 | 6 | 8 | 23105,3072 | 5 |
| 10 | 2011 | 6 | 9 | 11664,9658 | 3 |
| 11 | 2011 | 6 | 10 | 15815,9536 | 3 |
| 12 | 2011 | 6 | 11 | 15618,9542 | 4 |
| 13 | 2011 | 6 | 12 | 7907,9768 | 2 |
| 14 | 2011 | 6 | 13 | 27677,9188 | 4 |
| 15 | 2011 | 6 | 14 | 12409,8444 | 4 |
| 16 | 2011 | 6 | 15 | 15815,9536 | 2 |

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);
```

| | Miesiąc | Suma | Liczba różnych produktów |
|----|-------------|----------------|--------------------------|
| 1 | Styczen | 162948889,3944 | 211 |
| 2 | Luty | 91887841,9651 | 200 |
| 3 | Marzec | 339699263,2856 | 246 |
| 4 | Kwiecien | 100265759,5521 | 200 |
| 5 | Maj | 346435416,8907 | 263 |
| 6 | Czerwiec | 323879162,85 | 253 |
| 7 | Lipiec | 318807487,1861 | 260 |
| 8 | Sierpien | 183995405,5416 | 245 |
| 9 | Wrzesien | 228138296,6322 | 234 |
| 10 | Pazdziernik | 276923981,722 | 247 |
| 11 | Listopad | 84420434,5777 | 201 |
| 12 | Grudzien | 139068517,0799 | 211 |

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;
```

| | Dzień tygodnia | Suma | Liczba różnych produktów |
|---|----------------|----------------|--------------------------|
| 1 | Poniedziałek | 405243000,9979 | 241 |
| 2 | Wtorek | 358301922,2763 | 235 |
| 3 | Sroda | 496891402,8831 | 262 |
| 4 | Czwartek | 383509044,0962 | 260 |
| 5 | Piatek | 188847380,7946 | 238 |
| 6 | Sobota | 425217167,4498 | 251 |
| 7 | Niedziela | 338460538,1795 | 259 |

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.

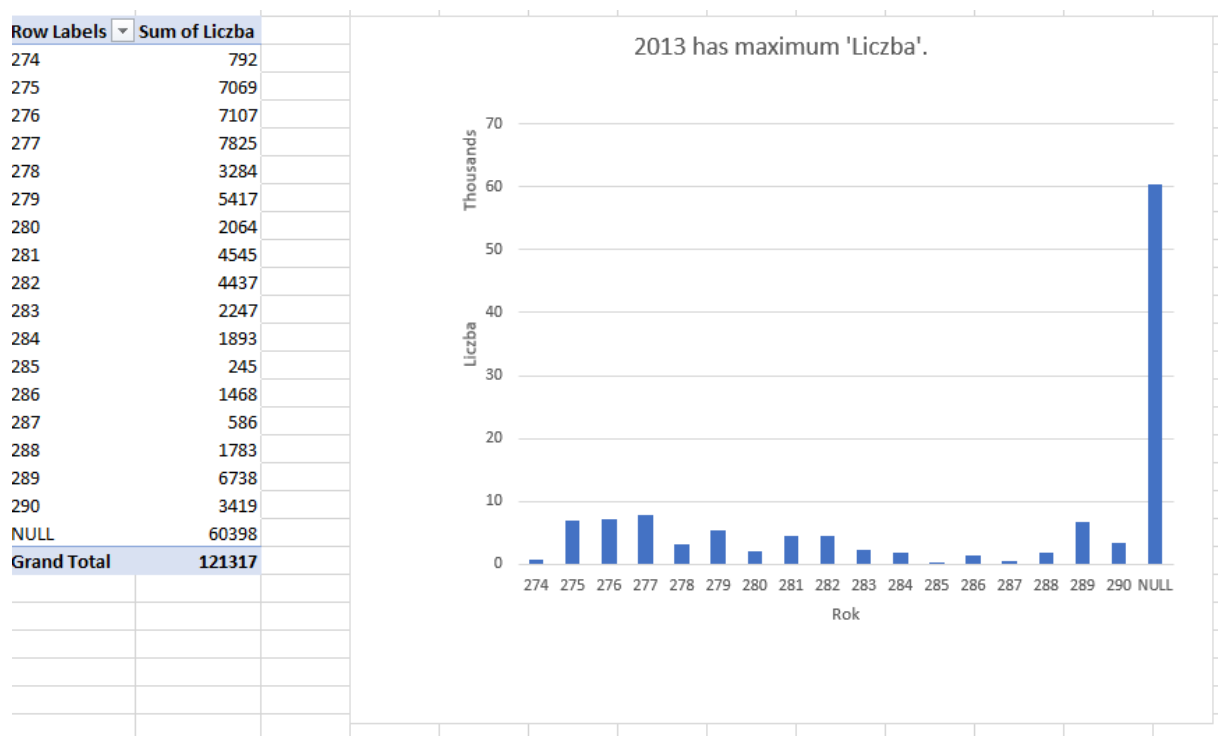
Zad 1.6

```
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;
```

| | Imie | Nazwisko | Liczba | Kwota | Karta |
|----|-----------|-------------|--------|-------------|-----------|
| 1 | Catherine | Abel | 4 | 127379,7919 | Zlota |
| 2 | Kim | Abercrombie | 12 | 584949,1308 | Platynowa |
| 3 | Humberto | Acevedo | 11 | 74786,2928 | Platynowa |
| 4 | Gustavo | Achong | 7 | 147804,9208 | Zlota |
| 5 | Pilar | Ackeman | 4 | 249804,8673 | Zlota |
| 6 | Carla | Adams | 4 | 98273,5468 | Zlota |
| 7 | Frances | Adams | 12 | 428350,5326 | Zlota |
| 8 | Jay | Adams | 6 | 158025,1722 | Zlota |
| 9 | Kaitlyn | Adams | 5 | 297,3336 | Srebna |
| 10 | Miguel | Adams | 5 | 259,6862 | Srebna |
| 11 | Samuel | Agcaoli | 9 | 18275,1045 | Srebna |
| 12 | Robert | Ahlering | 4 | 107741,109 | Zlota |
| 13 | Kim | Akers | 8 | 3974,607 | Srebna |
| 14 | Stanley | Alan | 4 | 244854,0475 | Zlota |
| 15 | Amy | Alberts | 8 | 264645,3146 | Zlota |
| 16 | Anna | Albright | 7 | 146870,6749 | Zlota |

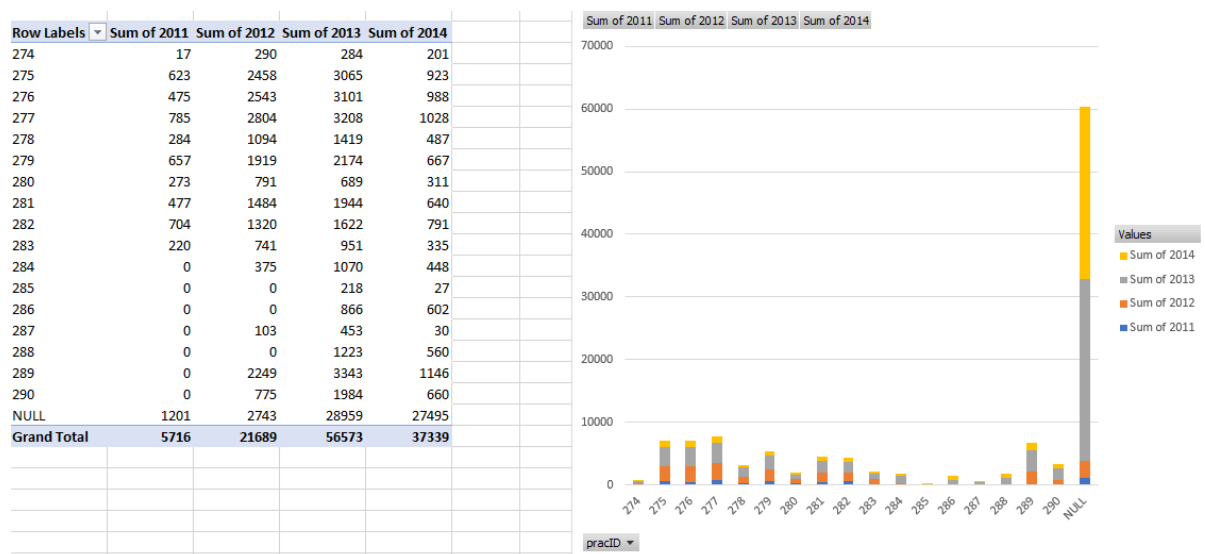
Wniosek: Bardzo duża liczba klientów posiada złotą kartę

Zad 2.1.1



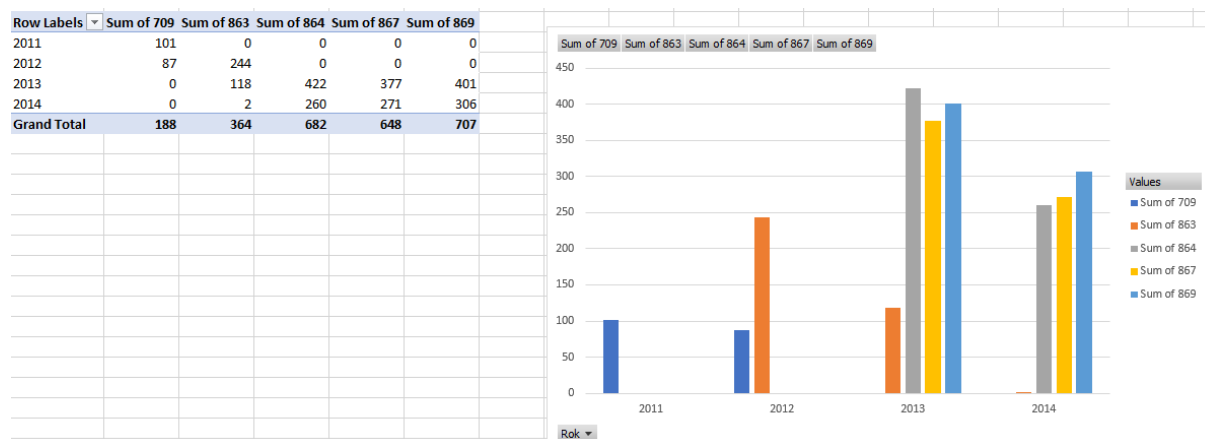
Wniosek: W zasadzie połowa produktów została sprzedana przez internet.

Zad 2.1.1a



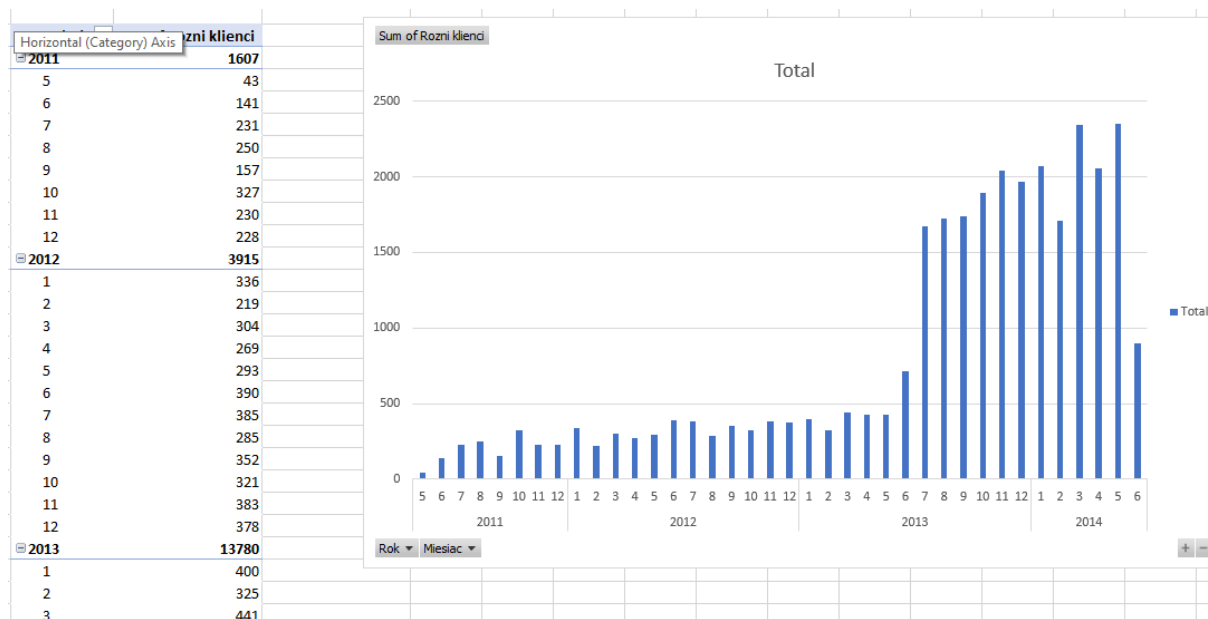
Wniosek: Zdecydowanie najmniej produktów zakupionych zostało w 2011 roku. W 2012 i 2013 roku sprzedaż fizyczna była wyższa niż w 2014 roku. Sprzedaż internetowa natomiast znacząco się zwiększyła w 2013 roku i pierwsza połowa 2014 roku osiągnęła taki wynik jak całość 2013r.

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



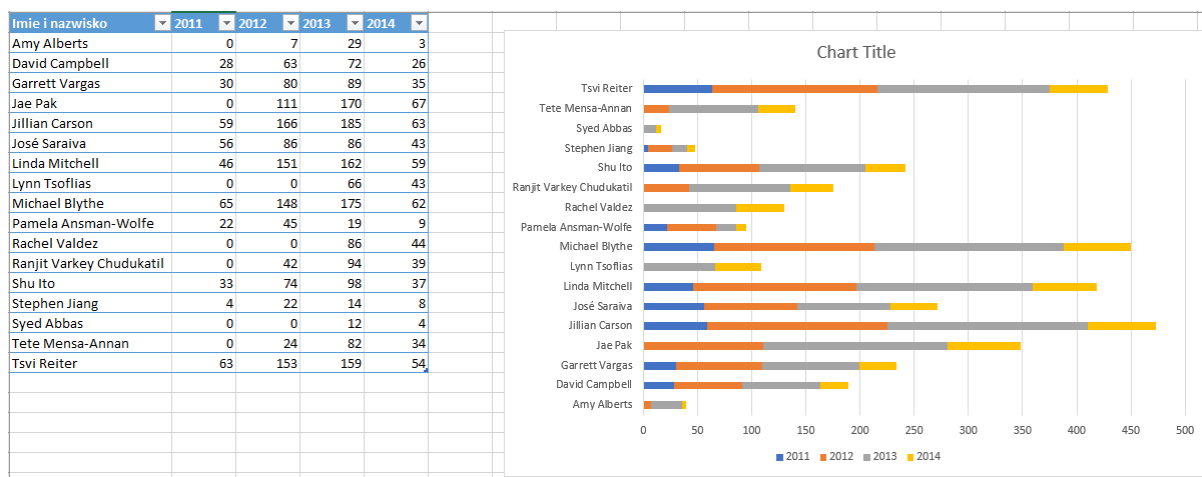
Wniosek: Ogólnie panuje tendencja wzrostowa liczby różnych klientów kupujących produkty w sklepie. Definitywny przełom nastąpił w połowie 2013 roku gdy liczba ta zwiększyła się dwukrotnie.

Zad 2.1.2b



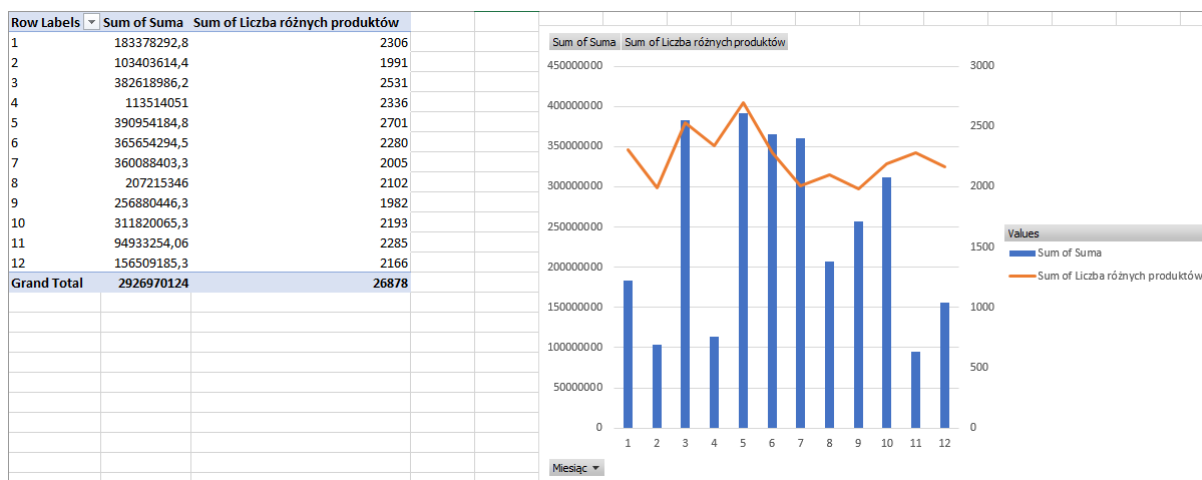
Wniosek: W zasadzie jak poprzednio tylko w innej formie wizualnej.

Zad 2.1.3



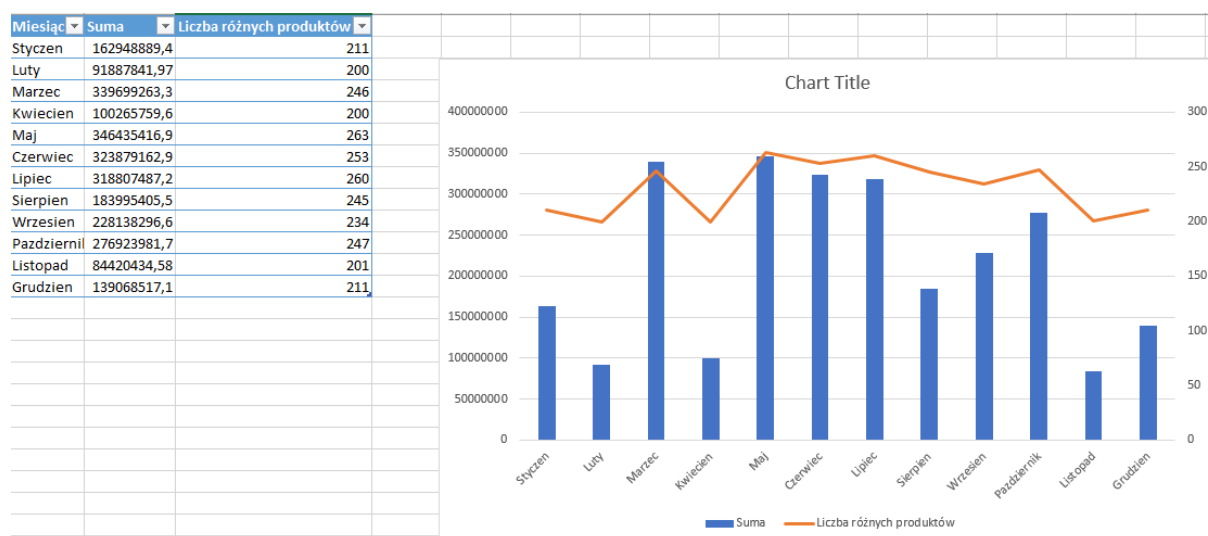
Wniosek: Ogólnie sprzedawcy z upływem lat zaczęli sprzedawać coraz więcej produktów. Różnice pomiędzy nimi wynikają prawdopodobnie z regionu, w którym pracują.

Zad 2.1.4



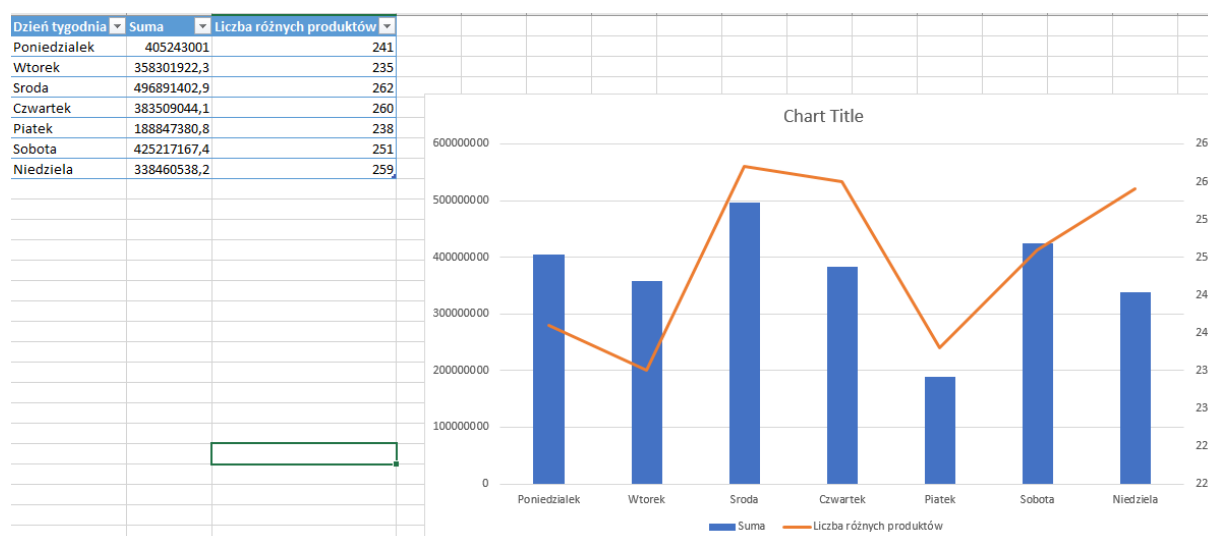
Wniosek: Najwięcej pieniędzy na różne produkty klienci przeznaczają w okresie wakacyjnym. Liczba sztuk sprzedawanych różnych produktów wydaje się być natomiast najwyższa na wiosnę.

Zad 2.1.5



Wniosek: Biorąc pod uwagę wszystkie lata to suma wydawanych przez klientów pieniędzy nadal wydaje się być najwyższa w lecie. Liczba różnych 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.

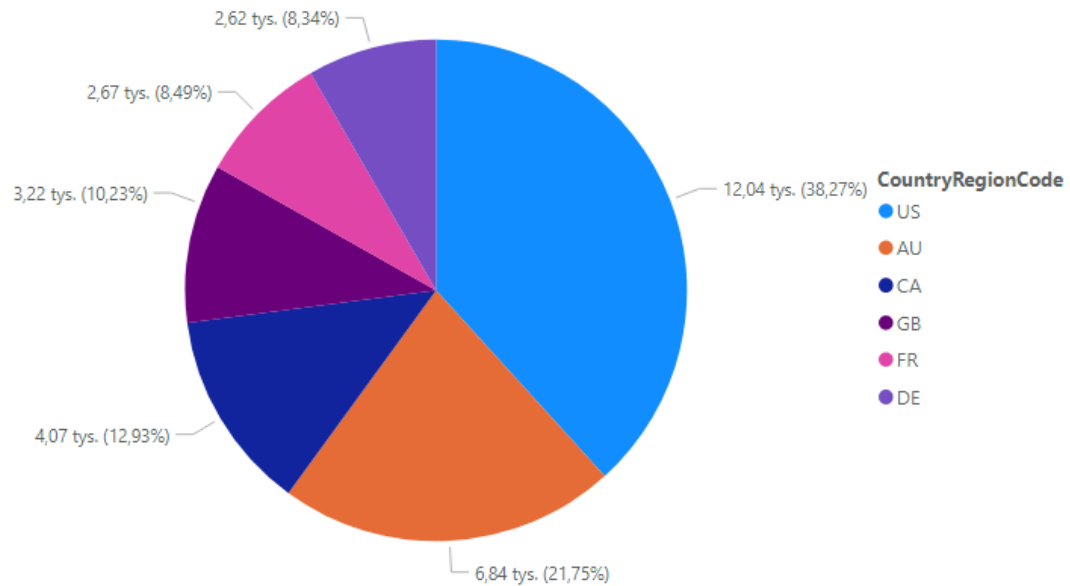
Zad 2.1.6



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

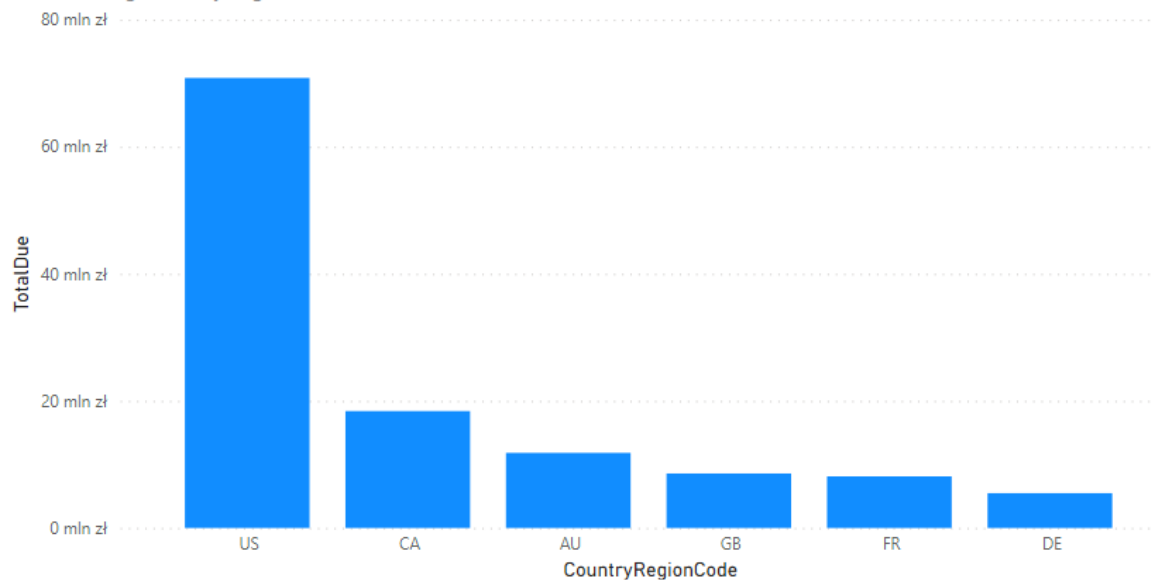
Liczba elementów SalesOrderID wg CountryRegionCode



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

Zad 2.2.2

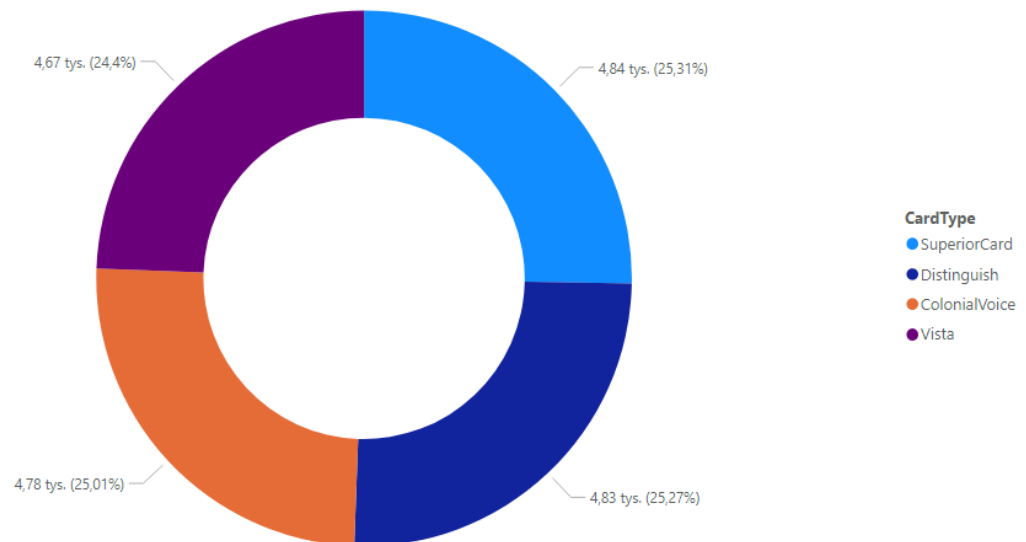
TotalDue wg CountryRegionCode



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

Zad 2.2.3

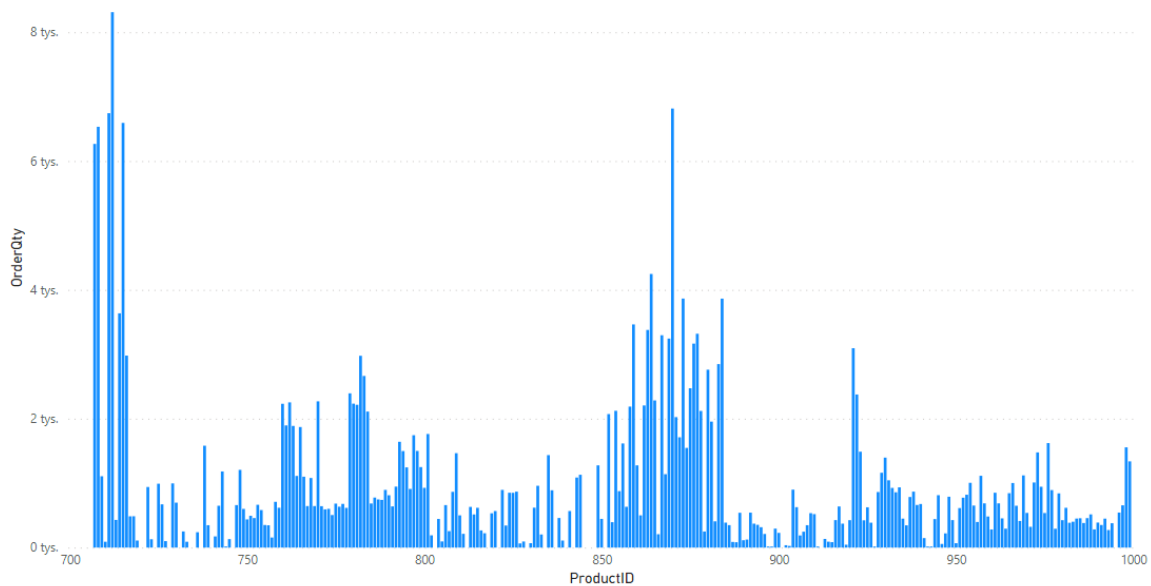
CreditCardID wg CardType



Wniosek: Wszystkie rodzaje kart kredytowych są równie popularne.

Zad 2.2.4

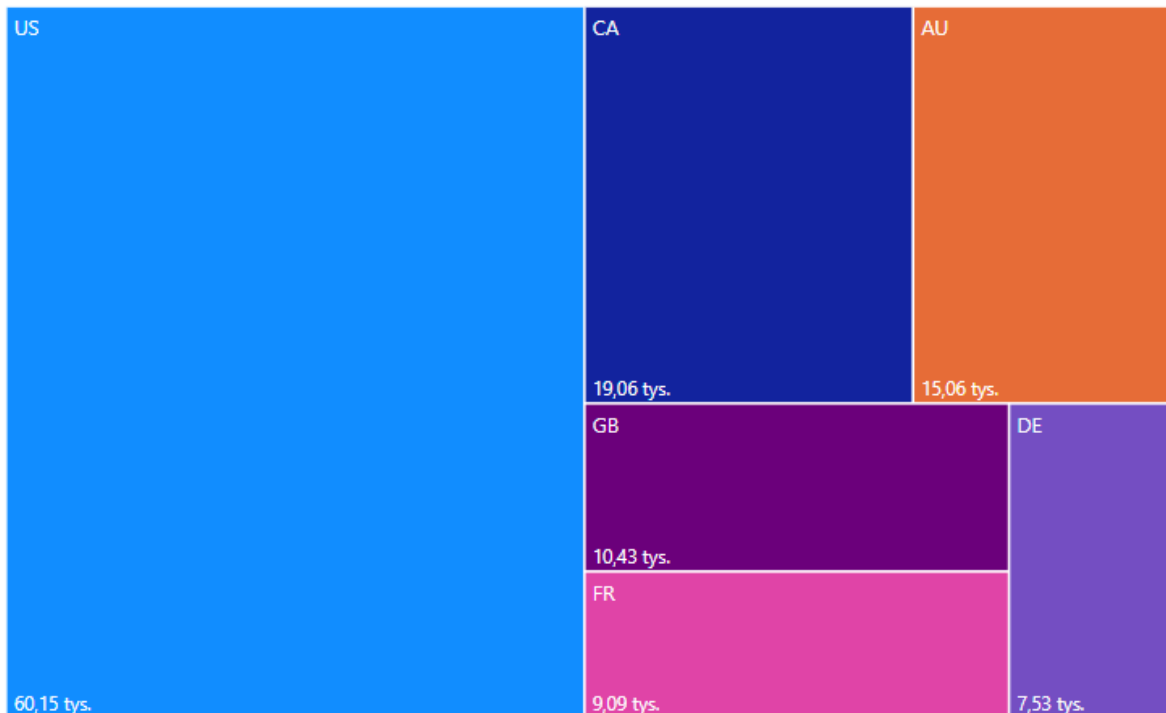
OrderQty wg ProductID



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

Zad 2.2.5

ProductID wg CountryRegionCode



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

Zad 3.1

Flat File Source

ADO NET Destination

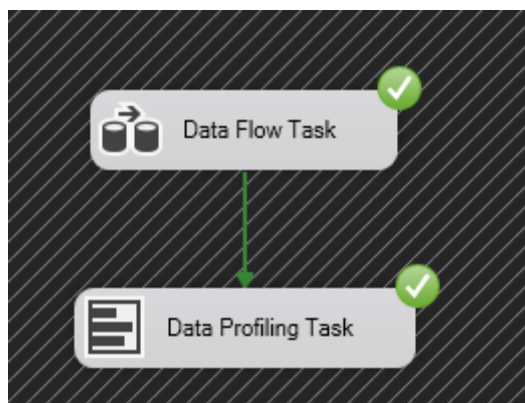
2_Z3 L2_Z3.xml dane_1

Connection Manager
[Mappings](#)
 Error Output

Available Input Columns

- Name
- Data zamówienia
- Kraj odbiorcy
- Cena jednostkowa
- Ilość
- Nazwa produktu
- Nazwisko
- Stanowisko

| Input Column | Destination Column |
|------------------|--------------------|
| Data zamówienia | Data zamówienia |
| Kraj odbiorcy | Kraj odbiorcy |
| Cena jednostkowa | Cena jednostkowa |
| Ilość | Ilość |
| Nazwa produktu | Nazwa produktu |
| Nazwisko | Nazwisko |
| Stanowisko | Stanowisko |
| Region | Region |



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

Zad 3.2

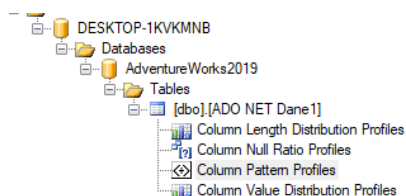
| Column | Null Count | Null Percentage |
|------------------|------------|-----------------|
| Cena jednostkowa | 0 | 0.0000 % |
| Data zamówienia | 0 | 0.0000 % |
| Ilość | 0 | 0.0000 % |
| Kraj odbiorcy | 0 | 0.0000 % |
| Nazwa produktu | 0 | 0.0000 % |
| Nazwisko | 0 | 0.0000 % |
| Region | 1704 | 26.3573 % |
| Stanowisko | 0 | 0.0000 % |

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

| Column | Minimum Length | Maximum Length | Ignore Leading Spaces |
|------------------|----------------|----------------|--------------------------|
| Cena jednostkowa | 7 | 9 | <input type="checkbox"/> |
| Data zamówienia | 10 | 10 | <input type="checkbox"/> |
| Ilość | 1 | 3 | <input type="checkbox"/> |
| Kraj odbiorcy | 3 | 15 | <input type="checkbox"/> |
| Nazwa produktu | 4 | 32 | <input type="checkbox"/> |
| Nazwisko | 4 | 9 | <input type="checkbox"/> |
| Region | 2 | 2 | <input type="checkbox"/> |
| Stanowisko | 22 | 33 | <input type="checkbox"/> |

| Length | Count | Percentage |
|--------|-------|------------|
| 9 | 1284 | 19.8608 % |
| 6 | 1227 | 18.9791 % |
| 7 | 2295 | 35.4988 % |
| 4 | 528 | 8.1671 % |
| 8 | 1131 | 17.4942 % |

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



| Column | Approximate |
|------------------|-------------|
| Cena jednostkowa | Approximate |
| Data zamówienia | Approximate |
| Ilość | Approximate |
| Kraj odbiorcy | Approximate |
| Nazwa produktu | Approximate |
| Nazwisko | Approximate |
| Region | Exact |
| Stanowisko | Approximate |

| Pattern Distribution - Nazwa produktu | | | Encrypted Connection | 1000 Rows |
|---------------------------------------|-----------------------------|--------------------|----------------------|-----------|
| No. | Pattern | Pattern Percentage | | |
| 1 | \p(L)+ \p(L)+ | 43 | | |
| 2 | \p(L)+ | 22 | | |
| 3 | \p(L)+ \p(L)+ \p(L)+ | 14 | | |
| 4 | \p(L)+ \p(L)+ \p(L)+ \p(L)+ | 7 | | |
| 5 | \p(L)+ \p(L)+S (\p(L)+)+ | 5 | | |
| 6 | (\p(L)+)+ | 1 | | |

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

| Column Value Distribution Profiles - [dbo].[ADO NET Dane1] | |
|--|---------------------------|
| Column | Number Of Distinct Values |
| Nazwa produktu | 77 |
| Nazwisko | 9 |
| Region | 1 |
| Stanowisko | 4 |

| Frequent Value Distribution (0.1000 %) - Stanowisko | | | | Encrypted Connection | 1000 Rows |
|---|-------|------------|--|----------------------|-----------|
| Value | Count | Percentage | | | |
| Przedstawiciel handlowy | 4611 | 71.3225 % | | | |
| Wiceprezes ds. sprzedaży | 723 | 11.1833 % | | | |
| Dyrektor ds. sprzedaży | 351 | 5.4292 % | | | |
| Koordinator sprzedaży we... | 780 | 12.0650 % | | | |

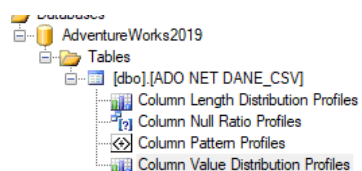
Wniosek: Większość pracowników jest na stanowisku przedstawiciela handlowego.

| Profiles (Table View) | | Column Null Ratio Profiles - [dbo].[ADO NET DANE_CSV] | |
|--|--|---|------------|
| <ul style="list-style-type: none"> Data Sources <ul style="list-style-type: none"> DESKTOP-1KVKMNB <ul style="list-style-type: none"> Databases <ul style="list-style-type: none"> AdventureWorks2019 <ul style="list-style-type: none"> Tables <ul style="list-style-type: none"> [dbo].[ADO NET DANE_CSV] <ul style="list-style-type: none"> Column Length Distribution Profiles Column Null Ratio Profiles Column Pattern Profiles Column Value Distribution Profiles | | Column | Null Count |
| | | GMT | 0 |
| | | London Mean Background Nitri... | 0 |
| | | London Mean Background Nitro... | 0 |
| | | London Mean Background Oxid... | 0 |
| | | London Mean Background Ozo... | 0 |
| | | London Mean Background PM1... | 0 |
| | | London Mean Background PM2... | 0 |
| | | London Mean Background Sulp... | 0 |
| | | London Mean Roadside Nitric O... | 0 |
| | | London Mean Roadside Nitroge... | 0 |
| | | London Mean Roadside Oxides... | 0 |
| | | London Mean Roadside Ozone... | 0 |
| | | London Mean Roadside PM10... | 0 |
| | | London Mean Roadside PM2 5... | 0 |
| | | London Mean Roadside Sulphu... | 0 |
| | | Month (text) | 0 |

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

| | | | | | |
|--|--|---|-------|------------|--------------------------|
| <ul style="list-style-type: none"> AdventureWorks2019 <ul style="list-style-type: none"> Tables <ul style="list-style-type: none"> [dbo].[ADO NET DANE_CSV] <ul style="list-style-type: none"> Column Length Distribution Profiles Column Null Ratio Profiles Column Pattern Profiles Column Value Distribution Profiles | | London Mean Background... | 0 | 18 | <input type="checkbox"/> |
| | | London Mean Background... | 2 | 18 | <input type="checkbox"/> |
| | | London Mean Background... | 0 | 18 | <input type="checkbox"/> |
| | | London Mean Background... | 2 | 18 | <input type="checkbox"/> |
| | | London Mean Background... | 2 | 18 | <input type="checkbox"/> |
| | | London Mean Background... | 0 | 18 | <input type="checkbox"/> |
| | | London Mean Background... | 1 | 18 | <input type="checkbox"/> |
| | | London Mean Roadside Nit... | 0 | 18 | <input type="checkbox"/> |
| | | London Mean Roadside Nit... | 2 | 18 | <input type="checkbox"/> |
| | | Length Distribution - London Mean Background Oxides of Nitrogen (ug m3) 1000 Rows | | | |
| | | Length | Count | Percentage | |
| | | 0 | 1152 | 18.1818 % | |
| | | 6 | 18 | 0.2841 % | |
| | | 18 | 1638 | 25.8523 % | |
| | | 4 | 38 | 0.5997 % | |
| | | 5 | 216 | 3.4091 % | |
| | | 16 | 254 | 4.0088 % | |
| Successfully loaded data profile from C:\Users\HDDVirtual\Desktop\lab\lista2\zad3_dane2.xml ... | | | | | |

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



| | |
|----------------------------|------|
| London Mean Background.. | 29/5 |
| London Mean Background.. | 2897 |
| London Mean Background.. | 2573 |
| London Mean Roadside Nit.. | 2577 |

| Frequent Value Distribution (0.1000 %) - London Mean Background Sulphur Dioxide (ug m3) | | | | |
|---|-------|----|------------|----------|
| Value | Count | | Percentage | |
| 2.9322580645161294 | | 10 | | 0.1578 % |
| 2.3548387096774195 | | 8 | | 0.1263 % |
| 2.032258064516129 | | 8 | | 0.1263 % |
| 2.8 | | 10 | | 0.1578 % |
| 4.935483870967742 | | 8 | | 0.1263 % |
| 2.7838709677419358 | | 8 | | 0.1263 % |
| 3.1333333333333333 | | 8 | | 0.1263 % |
| 2.4838709677419355 | | 12 | | 0.1894 % |
| 2.8387096774193545 | | 8 | | 0.1263 % |
| 2.709677419354838 | | 8 | | 0.1263 % |
| 3.466666666666667 | | 8 | | 0.1263 % |
| 2.580645161290323 | | 8 | | 0.1263 % |

Successfully loaded data profile from C:\Users\HDVirtual\Desktop\lab\lista2\zad3_dane2.xml ...

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, Tableau 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)