Hurtownie Danych - laboratorium Lista 2

Zadanie. 1. Ekstrakcja danych

1. Utworzyć zestawienie, które dla poszczególnych miesięcy i lat przedstawi informację o liczbie różnych klientów. Przygotuj zapytanie z i bez użycia polecenia **pivot**.

Bez użycia pivot:

```
ESELECT YEAR(soh.OrderDate) AS Year, MONTH(soh.OrderDate) AS Month, COUNT(DISTINCT soh.CustomerID) AS UniqueCustomersCount FROM Sales.SalesOrderHeader soh GROUP BY YEAR(soh.OrderDate), MONTH(soh.OrderDate)

ORDER BY 1, 2;
```

	Year	Month	UniqueCustomersCount
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

Z użyciem pivot:

	Year	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

2. Utworzyć zestawienie zawierające w wierszach imiona i nazwiska sprzedawców, a w kolumnach kolejne lata. Wartością będzie liczba obsłużonych transakcji. Wyświetlić tylko tych sprzedawców, którzy pracowali przez wszystkie 4 lata.

```
WITH SalesPersonTransactions AS (
    SELECT
        CONCAT(sp.FirstName, ' ', sp.LastName) AS Name,
        YEAR(soh.OrderDate) AS Year,
        COUNT(soh.SalesOrderID) AS TransactionCount
    FROM
        Sales.SalesOrderHeader soh
        HumanResources.Employee e ON soh.SalesPersonID = e.BusinessEntityID
        Person.Person sp ON e.BusinessEntityID = sp.BusinessEntityID
    GROUP BY
        sp.FirstName, sp.LastName, YEAR(soh.OrderDate), soh.SalesPersonID
SELECT
FROM
   SalesPersonTransactions
PIVOT (
    SUM(TransactionCount)
    FOR Year in ([2011], [2012], [2013], [2014])
    ) AS PivotTable
WHERE (
    [2011] IS NOT NULL AND
    [2012] IS NOT NULL AND
    [2013] IS NOT NULL AND
    [2014] IS NOT NULL
ORDER BY
    1:
```

	Name	2011	2012	2013	2014
1	David Campbell	28	63	72	26
2	Garrett Vargas	30	80	89	35
3	Jillian Carson	59	166	185	63
4	José Saraiva	56	86	86	43
5	Linda Mitchell	46	151	162	59
6	Michael Blythe	65	148	175	62
7	Pamela Ansman-Wolfe	22	45	19	9
8	Shu Ito	33	74	98	37
9	Stephen Jiang	4	22	14	8
10	Tsvi Reiter	63	153	159	54

3. Zdefiniować zapytanie wyznaczające sumę kwot sprzedaży towarów oraz liczbę różnych produktów w zamówieniach w poszczególnych latach, miesiącach, dniach.

```
SELECT
    YEAR(soh.OrderDate) AS Year,
    MONTH(soh.OrderDate) AS Month,
    DAY(soh.OrderDate) AS Day,
    CAST(SUM(sod.LineTotal) AS DECIMAL(9,2)) AS SalesSum, --nie mozemy sumowac po TotalDue z powodu powtorzen
    COUNT(DISTINCT sod.ProductID) AS ProductsCount
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;
```

	Year	Month	Day	SalesSum	ProductsCount
1	2011	5	31	503805.92	47
2	2011	6	1	13931.52	4
3	2011	6	2	15012.18	4
4	2011	6	3	7156.54	2
5	2011	6	4	15012.18	4
6	2011	6	5	14313.08	3
7	2011	6	6	7855.64	3
8	2011	6	7	7855.64	2
9	2011	6	8	20909.78	5
10	2011	6	9	10556.53	3
11	2011	6	10	14313.08	3
12	2011	6	11	14134.80	4
13	2011	6	12	7156.54	2
14	2011	6	13	25047.89	4
15	2011	6	14	11230.63	4
16	2011	6	15	14313.08	2
17	2011	6	16	14134.80	4
18	2011	6	17	6953.26	2
19	2011	6	18	25568.71	5
20	2011	6	19	11255.63	4

4. Wykorzystując polecenie CASE przygotować podsumowania do zestawienia z poprzedniego zadania tak, aby sumowane były kwoty zamówień oraz obliczana liczba różnych produktów dla poszczególnych miesięcy i dni tygodnia.

```
MONTH(soh.OrderDate) AS Month,
   CASE DATEPART(dw, soh.OrderDate)
       WHEN 1 THEN 'Sunday'
        WHEN 2 THEN 'Monday'
        WHEN 3 THEN 'Tuesday'
        WHEN 4 THEN 'Wednesday'
        WHEN 5 THEN 'Thursday'
        WHEN 6 THEN 'Friday'
        WHEN 7 THEN 'Saturday'
    END AS WeekDay,
    CAST(SUM(sod.LineTotal) AS DECIMAL(12,2)) AS SalesSum,
    COUNT(DISTINCT sod.ProductID) AS ProductCount
FROM
    Sales.SalesOrderHeader soh
JOIN
    Sales.SalesOrderDetail sod ON soh.SalesOrderID = sod.SalesOrderID
   MONTH(soh.OrderDate), DATEPART(dw, soh.OrderDate)
ORDER BY 1, 2;
```

	Month	WeekDay	SalesSum	ProductCount
1	1	Friday	348563.53	105
2	1	Monday	1934519.17	151
3	1	Saturday	296672.54	100
4	1	Sunday	3747550.25	134
5	1	Thursday	460434.10	106
6	1	Tuesday	423922.06	110
7	1	Wednesday	3136656.04	157
8	2	Friday	397978.50	117
9	2	Monday	334244.74	100
10	2	Saturday	302611.30	105
11	2	Sunday	383880.80	107
12	2	Thursday	2190222.54	172
13	2	Tuesday	327070.93	103
14	2	Wednesday	1194065.29	139
15	3	Friday	2686369.58	147
16	3	Monday	3760316.56	184
17	3	Saturday	5555826.71	229
18	3	Sunday	474329.92	113
19	3	Thursday	406354.45	112

- 5. Przygotować zestawienie, w którym dla wybranych klientów przygotujemy kartę lojalnościową:
 - a. srebrną, jeśli klient wykonał co najmniej 2 transakcje w sklepie;
 - złotą, jeśli wykonał co najmniej 4 transakcje w sklepie, w tym co najmniej 2 transakcje, których łączna kwota przekraczała 250% średniej wartości zamówień w bazie;
 - c. platynową, jeśli klient spełniał warunki otrzymania karty złotej oraz w co najmniej jednej transakcji kupił jednocześnie produkty ze wszystkich kategorii.

Schemat wynikowej tabeli powinien wyglądać następująco:

KartyLojalnosciowe(Imie, Nazwisko, Liczba transakcji, Łączna kwota transakcji, Kolor karty)

```
⊔WITH AllCategoriesOrders AS (
       {\tt SELECT} \ \ {\tt soh.SalesOrderID}, \ \ {\tt soh.CustomerID}
       FROM Sales SalesOrderHeader soh
       JOIN Sales.SalesOrderDetail sod ON soh.SalesOrderID = sod.SalesOrderID
       JOIN Production.Product p ON sod.ProductID = p.ProductID
       {\tt JOIN \ ProductSubcategory \ sc \ ON \ p.ProductSubcategoryID = sc.ProductSubcategoryID} \\
       GROUP BY soh.SalesOrderID, soh.CustomerID
HAVING COUNT(DISTINCT sc.ProductCategoryID) = 4
   AboveAverageOrders AS (
       SELECT SalesOrderID, CustomerID,
                WHEN TotalDue > 2.5 * (SELECT AVG(TotalDue) FROM Sales.SalesOrderHeader) THEN 1
                ELSE 0
           END AS AboveAvg
       FROM Sales.SalesOrderHeader
   CustomerOrderCount AS (
       SELECT CustomerID,
               COUNT(SalesOrderID) AS OrderCount,
               SUM(AboveAvg) AS AboveAvgOrderCount
       FROM AboveAverageOrders
       GROUP BY CustomerID
) SELECT p.FirstName, p.LastName, coc.OrderCount, CAST(SUM(soh.TotalDue) AS DECIMAL(9,2)) AS TransactionSum,
       CASE
            WHEN coc.OrderCount >= 4 AND coc.AboveAvgOrderCount >= 2 AND soh.CustomerID IN (SELECT CustomerID FROM AllCategoriesOrders) THEN 'Platinum'
            WHEN coc.OrderCount >= 4 AND coc.AboveAvgOrderCount >= 2 THEN 'Gold
            WHEN coc.OrderCount >= 2 THEN 'Silver'
           ELSE NULL
           END AS CardColor
       FROM
       Sales.SalesOrderHeader soh
       JOIN CustomerOrderCount coc ON soh.CustomerID = coc.CustomerID
JOIN Sales.Customer c ON soh.CustomerID = c.CustomerID
JOIN Person.Person p ON c.PersonID = p.BusinessEntityID
       GROUP BY soh.CustomerID, p.FirstName, p.LastName, coc.OrderCount, coc.AboveAvgOrderCount;
```

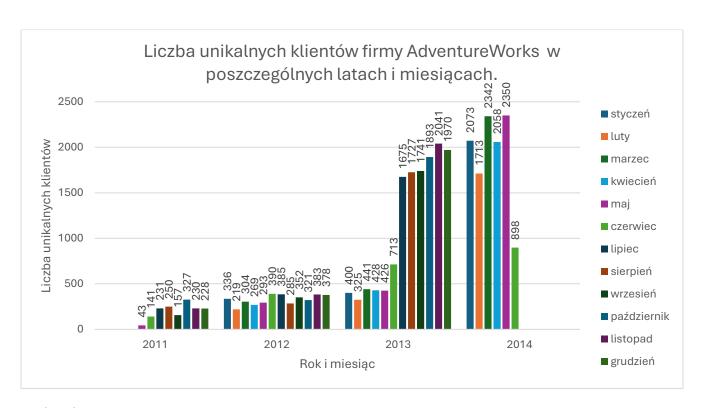
	FirstName	LastName	OrderCount	TransactionSum	CardColor
1	Catherine	Abel	4	127379.79	Platinum
2	Kim	Abercrombie	12	584949.13	Platinum
3	Humberto	Acevedo	11	77585.20	Gold
4	Gustavo	Achong	7	147804.92	Platinum
5	Pilar	Ackeman	4	249804.87	Platinum
6	Aaron	Adams	1	130.35	NULL
7	Adam	Adams	1	156.89	NULL
8	Alex	Adams	1	1918.26	NULL
9	Alexandra	Adams	1	3953.99	NULL
10	Allison	Adams	2	1770.73	Silver
11	Amanda	Adams	2	148.58	Silver
12	Amber	Adams	2	2211.60	Silver
13	Andrea	Adams	1	2630.42	NULL
14	Angel	Adams	1	142.51	NULL
15	Bailey	Adams	2	5037.17	Silver
16	Ben	Adams	2	4793.00	Silver
17	Blake	Adams	1	1918.26	NULL
18	Carla	Adams	4	98273.55	Gold

Zadanie 2. Analiza danych

 Przedstaw wyniki zadania 1 w postaci tabel i wykresów przestawnych w programie MS Excel. Zinterpretuj wyniki.

Liczba różnych klientów w poszczególnych latach i miesiącach.

Sum of CustomersCount	Column Labels 🔻												
Row Labels	styczeń	luty	marz	kwie	maj	czerwiec	lipied	sierp	wrzes	paźdz	listop	grudz	Grand Total
2011					43	141	231	250	157	327	230	228	1607
2012	336	219	304	269	293	390	385	285	352	321	383	378	3915
2013	400	325	441	428	426	713	1675	1727	1741	1893	2041	1970	13780
2014	2073	1713	2342	2058	2350	898							11434
Grand Total	2809	2257	3087	2755	3112	2142	2291	2262	2250	2541	2654	2576	30736

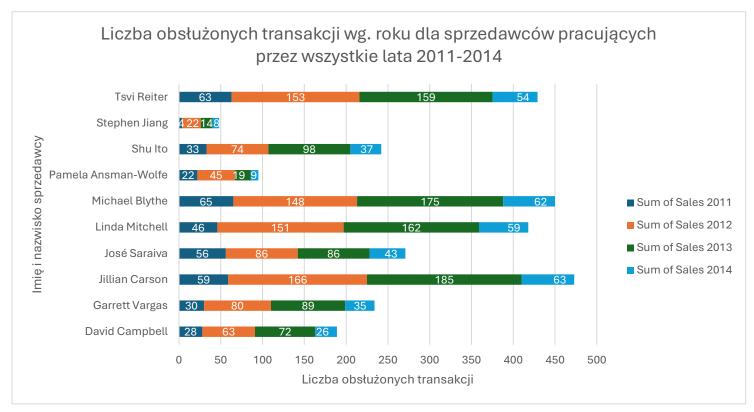


- Pierwsza transakcja firmy odbyła się w maju 2011 roku, a ostatnia odnotowana w czerwcu 2014 roku.
- Okres roku nie ma dużego wpływu na liczbę unikalnych klientów.
- Firma zwiększyła swoją liczbę unikalnych miesięcznych klientów do ponad 200% w lipcu 2013 roku.

• Najwięcej (2350) unikalnych klientów odnotowano w maju 2014 roku, a najmniej (43) w maju 2011 roku.

Liczba transakcji obsłużonych przez pracowników pracujących przez wszystkie lata działania firmy.

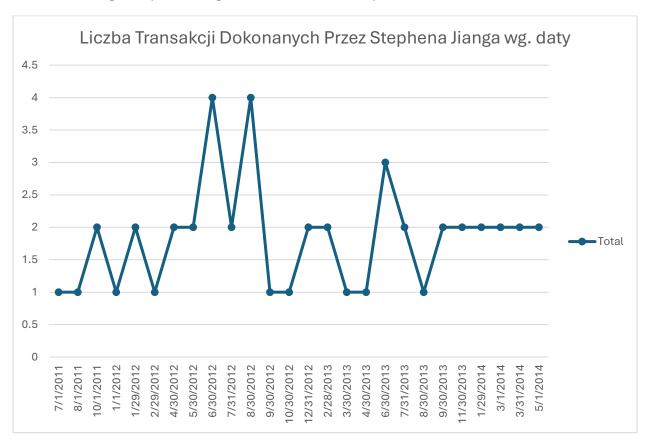
Row Labels	Sum of Sales 2011	Sum of Sales 2012	Sum of Sales 2013	Sum of Sales 2014
David Campbell	28	63	72	26
Garrett Vargas	30	80	89	35
Jillian Carson	59	166	185	63
José Saraiva	56	86	86	43
Linda Mitchell	46	151	162	59
Michael Blythe	65	148	175	62
Pamela Ansman-Wolfe	22	45	19	9
Shu Ito	33	74	98	37
Stephen Jiang	4	22	14	8
Tsvi Reiter	63	153	159	54
Grand Total	406	988	1059	396



Wnioski:

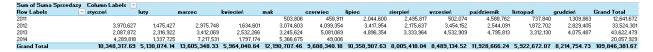
 Sprzedawcą, który dokonał największej liczby transakcji oraz pracuje przez wszystkie cztery lata jest Jillian Carson, najmniej transakcji dokonanych posiadał Stephen Jiang.

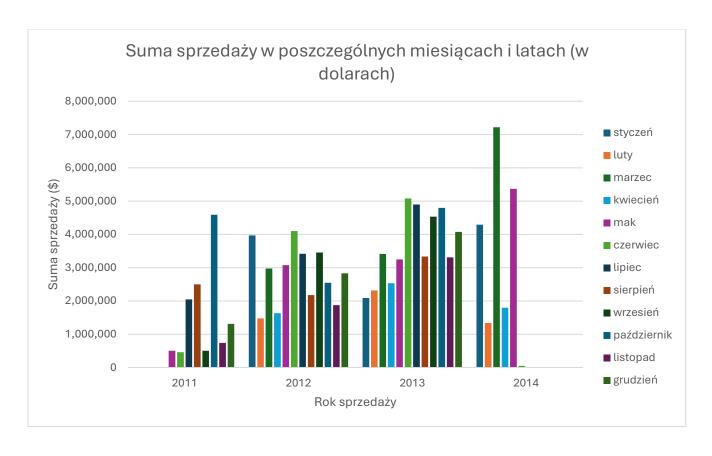
- Najwięcej obsłużonych transakcji w pojedynczym roku: 185 Jillian Carson 2013
- Tylko 10 sprzedawców pracowało przez wszystkie 4 lata działania firmy.
- Dlaczego Stephen Jiang ma tak mało transakcji?



- wszystkie transakcje Stephena Jianga odbywają się na początku miesiąca (1. Dzień) lub pod koniec (28-31. Dzień).
- Stephen Jiang sprzedaje produkty w dużych ilościach (wysokie ProductQty na transakcję)
- Prawdopodobnie oznacza to, że Stephen Jiang jest pracownikiem odpowiedzialnym za obsługiwanie zamówień inwentaryzacyjnych, zamawianych przez sklepy – a nie klientów indywidualnych.

Suma kwot sprzedaży oraz liczba unikalnych produktów wg. Lat i miesięcy.

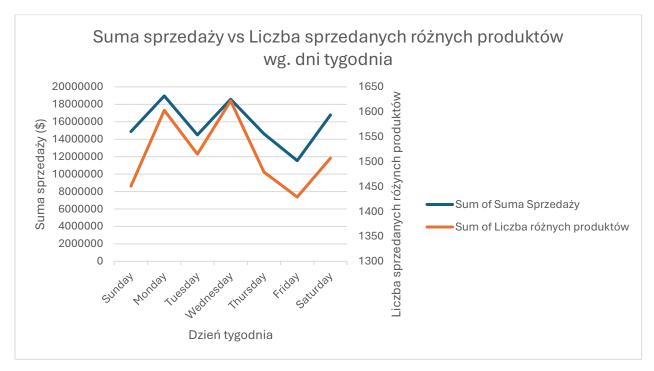




- Najlepsza sprzedaż odnotowywana jest w sezonie od maja do lipca, a najgorsza od listopada do lutego możemy po tym wywnioskować, że większość rynków sprzedaży znajduje się na półkuli północnej.
- W 2014 roku sprzedaż była bardzo nieregularna w lutym oraz kwietniu odniesiono bardzo niskie wyniki.
- Najlepszym miesiącem dla firmy był marzec 2014.

Suma kwot sprzedaży oraz Liczba różnych produktów vs dni tygodnia

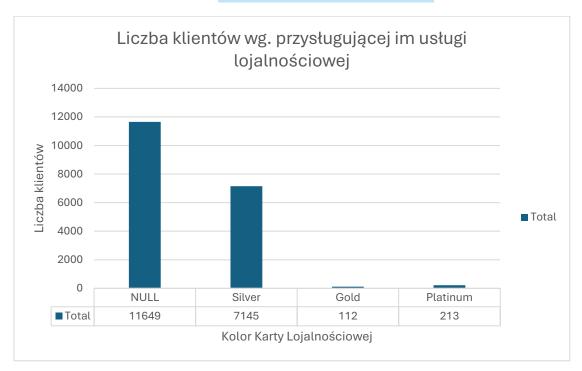
Row Labels ▼	Sum of Suma Sprzedaży	Sum of Liczba różnych produktów
Sunday	14873490.6	1451
Monday	18963526.25	1603
Tuesday	14484790.86	1515
Wednesday	18575137.89	1622
Thursday	14616544.77	1479
Friday	11539568.92	1429
Saturday	16793322.12	1507
Grand Total	109846381.4	10606



- W środy uzyskano największą sumę sprzedaży oraz największą sprzedaną liczbę różnych produktów.
- Istnieje korelacja (niekoniecznie wynikowość) pomiędzy liczbą unikalnych produktów sprzedanych w danym dniu tygodnia oraz sumą sprzedaży.
- Najmniejszą sumę sprzedaży oraz liczbę unikalnych produktów uzyskano w piątek.

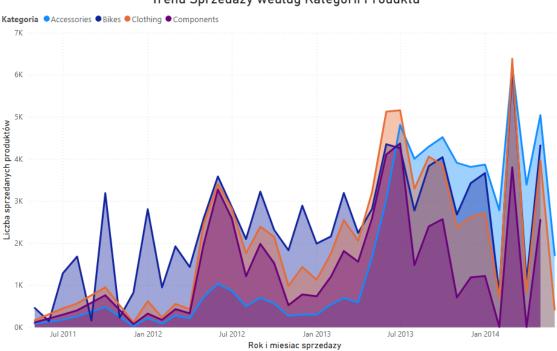
Karty lojalnościowe

CardColor	# of Clients
NULL	11649
Silver	7145
Gold	112
Platinum	213
Grand Total	19119



- 19119 Klientów dokonało jakiekolwiek zamówienie w sklepie.
- 61% z nich (11649) nigdy nie dokonało drugiego zamówienia.
- W pozostałych 39%:
 - o 95.6% (7145) przysługuje karta Srebrna
 - 1.5% (112) przysługuje karta złota (wykonali co najmniej 4 transakcje w sklepie oraz co najmniej 2 których kwota przekraczała 250% średniej wartości zamówienia)
 - 2.9% (213) spełniło założenia karty złotej oraz dokonało co najmniej 1 zamówienia na produkty z każdej kategorii.
 - Wśród ludzi którzy spełnili warunki karty złotej, większość z nich dokonała również zamówienia na produkty z każdej kategorii w jednym zamówieniu.

- 2. Przygotuj 5 dodatkowych tabel/wykresów, które pokażą ciekawe zależności w bazie AdventureWorks przy użyciu Power BI lub Tableau. Przedstaw wnioski biznesowe wynikające z tych zestawień.
- a. Trend sprzedaży w czasie według kategorii produktu



Trend Sprzedazy według Kategorii Produktu

Wnioski:

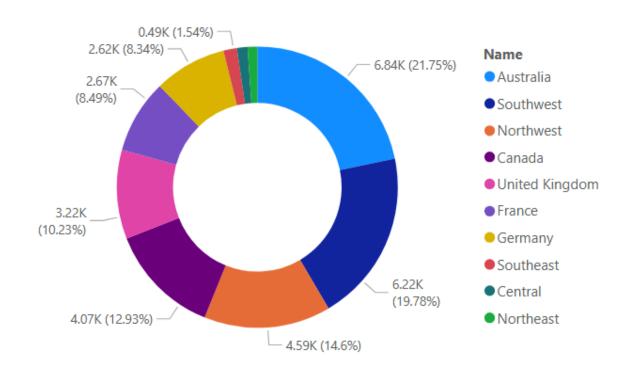
- Na początku działalności firmy najpopularniejsze były rowery.
- Sprzedaż akcesoriów bardzo wzrosła w okresie kwiecień 2013 lipiec 2013. W ostatnich miesiącach rejestru transakcji to akcesoria były najpopularniejszym zakupem.
- Liczba sprzedanych produktów bardzo wacha się z miesiąca na miesiąc najniższe sprzedaże odnotowywane są w okresie zimowym listopad luty.

Cele biznesowe:

 Optymalizacja Zarządzania Zasobami – Identyfikacja miesięcy o niskiej sprzedaży może pomóc w efektywniejszym zarządzaniu budżetem firmy. • Analiza skuteczności działań marketingowych – Zestawienie danych o sprzedaży z czasem kampanii marketingowych może pomóc w odnalezieniu działających wzorców kampanii i zwiększeniu efektywności działań marketingowych.

b. Liczba sprzedaży wg. Terytorium

Liczba sprzedazy według terytorium



Wnioski:

 Najwięcej sprzedaży odbyło się na terytoriach o niskiej gęstości zaludnienia (takich jak Australia, Stany Zjednoczone czy Kanada) oraz terytoriach korzystających z języka angielskiego (Australia, Stany Zjednoczone, Kanada).

Cele biznesowe:

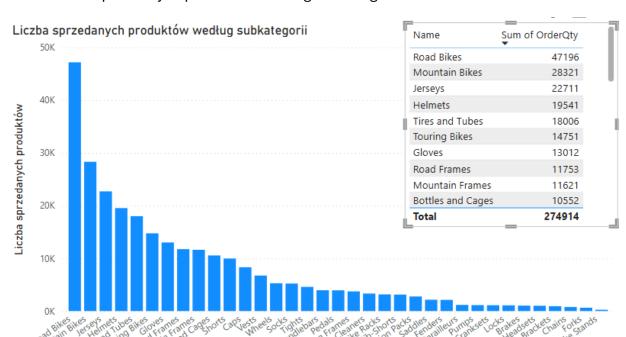
- Zwiększenie sprzedaży na terytoriach nieanglojęzycznych poprawa marketingu w językach regionalnych (niemiecki i francuski) mogłaby pomóc zwiększyć sprzedaż na terytoriach nieanglojęzycznych.
- Zwiększenie sprzedaży na terytoriach o dużej gęstości zaludnienia –
 dostosowanie oferty do terytoriów bardziej zurbanizowanych (na przykład
 wprowadzenie większego wyboru rowerów miejskich lub marketing rowerów
 miejskich) mogłoby zwiększyć sprzedaże w terytoriach takich jak Francja,
 Niemcy, Wielka Brytania lub Northeast USA.

c. Średnia wartość zamówienia wg. Sprzedawcy

BusinessEntityID	FirstName	LastName	Average of TotalDue ▼
280	Pamela	Ansman-Wolfe	\$39,455.2223
281	Shu	Ito	\$29,998.2144
290	Ranjit	Varkey Chudukatil	\$29,074.1555
276	Linda	Mitchell	\$27,978.5145
289	Jae	Pak	\$27,543.4625
274	Stephen	Jiang	\$25,748.6343
282	José	Saraiva	\$24,662.4969
277	Jillian	Carson	\$23,979.6742
275	Michael	Blythe	\$23,278.5935
283	David	Campbell	\$22,263.9926
287	Amy	Alberts	\$21,190.1915
279	Tsvi	Reiter	\$18,848.6566
284	Tete	Mensa-Annan	\$18,629.4027
278	Garrett	Vargas	\$17,390.6932
288	Rachel	Valdez	\$15,864.5626
286	Lynn	Tsoflias	\$14,737.9949
285	Syed	Abbas	\$12,220.549
Total			\$3,915.9951

Wnioski biznesowe:

- Identyfikacja kluczowych sprzedawców sprzedawcy z wysoką średnią wartością zamówienia generują większy przychód na jedno zamówienie, co sugeruje, że mogą być lepsi w sprzedaży lepszych produktów lub zestawów produktów. Takich sprzedawców można uznać za kluczowych.
- Potencjał do szkoleń Sprzedawców o najwyższych średnich wartościach zamówienia można wykorzystać do szkolenia innych sprzedawców w ramach upsellingu.
- Motywacja sprzedawców Nagrodzenie najlepszych sprzedawców premią może spowodować większą motywację wśród pracowników oraz zwiększenie zysków firmy.



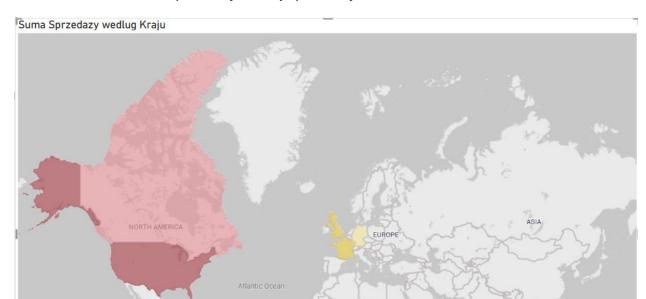
d. Liczba sprzedanych produktów według subkategorii

Wnioski biznesowe:

 Identyfikacja najlepiej sprzedających się typów produktów – może pomóc w optymalnym zarządzaniu zaopatrzeniem lub zatrudnianiu specjalistów dotyczących subkategorii.

Subkategoria

© 2025 TomTom, © 2025 Microsoft Corporation, © OpenStreetMap Terr



e. Suma wartości sprzedaży vs Kraj sprzedaży

Wnioski:

Microsoft Bing

Największe wartości sprzedaży firma zyskuje w państwach ameryki północnej.

SOUTH AMERICA

Cele biznesowe:

0-6,000,000\$

24,000,000\$ +

6,000,000\$ -12,000,000\$

12,000,000\$ -24,000,000\$

- Optymalizacja zarządzania zasobami największe wartości sprzedaży odnoszone są w państwach Ameryki Północnej. Firma powinna wziąć to pod uwagę w trakcie rozdysponowania towaru do magazynów w różnych częściach świata.
- Analiza rynków analiza strategii przeprowadzanych na różnych rynkach może pozwolić firmie na odnalezienie działających wzorców i zwiększenie sprzedaży na słabszych rynkach.
- Rozszerzenie globalnych rynków Analiza charakterystyki rynków silnych może pomóc w odpowiednim dobraniu potencjalnych nowych rynków.

Wnioski:

Wnioski z analizy danych:

1. Działalność firmy

- a. Dane sprzedażowe wskazują na dynamiczny rozwój firmy od 2011 do 2014 roku, ze znacznym wzrostem liczby klientów.
- b. Wzrost unikalnych miesięcznych klientów do ponad 200% w lipcu 2013 roku może sugerować udaną kampanię marketingową lub sezonowy popyt.
- c. Dni tygodnia wpływają na sprzedaż największe przychody odnotowano w środy, a najmniejsze w piątki.
- d. Sezonowość sprzedaży jest wyraźnie widoczna najlepsze wyniki osiągane są latem dla półkuli północnej, a najsłabsze w okresie zimowym.

2. Wydajność sprzedawców

- a. Tylko 10 sprzedawców pracowało przez wszystkie lata działalności firmy
- b. Najlepszym sprzedawcą pod względem liczby transakcji była Jillian Carson.
- c. Istnieją znaczące różnice w średniej wartości zamówienia pomiędzy sprzedawcami, co sugeruje potencjał do szkoleń z technik sprzedażowych.

3. Analiza klientów

- a. Większość klientów (61%) dokonała tylko jednego zakupu, co sugeruje potrzebę działań mających na celu poprawę retencji klientów.
- b. Jedynie 1,5% klientów zdobyło złotą kartę, a 2,9% platynową, co oznacza, że stosunkowo niewielka liczba klientów dokonuje zakupów na dużą skalę.

4. Analiza regionów

- a. Największa liczba transakcji pochodzi z krajów anglojęzycznych (USA, Kanada, Australia), co sugeruje, że firma powinna skupić się na poprawie marketingu w językach regionalnych.
- b. Różnice w sprzedaży między rynkami sugerują potrzebę analizy strategii cenowych i marketingowych dla poszczególnych lokalizacji.

Wnioski z wykorzystanych narzędzi:

- 1. Zapytania PIVOT oraz CASE pomogły w efektywnym agregowaniu oraz wydobywaniu danych niedostępnych bezpośrednio w bazie danych.
- 2. Tabele przestawne w programie Excel pozwalają na łatwą i szybką analizę zależności między poszczególnymi kolumnami tabeli.
- 3. PowerBI pozwala na szybką i dokładną analizę danych zawartych w bazie danych dzięki integracji z MS SQL Server.
- 4. PowerBI ułatwia wydobywanie danych z bazy danych, bez potrzeby własnoręcznego pisania zapytań SQL.
- 5. Relacje w PowerBI mogą zostać wygenerowane automatycznie, nie zawsze jednak działają poprawnie. Podczas generowania wykresu dla średniej wartości transakcji według sprzedawcy natrafiłem na problem:
 - a. Relacja 1...* między Sprzedawcą a SalesOrderHeader nie działała poprawnie z powodu występujących rekordów SalesOrderHeader, które nie miały przypisanego sprzedawcy (SalesPersonID = NULL). Aby relacja działała poprawnie, musiałem zmienić liczność relacji na wiele do wielu (*...*).