



HURTOWNIE DANYCH

Lista 2 – Podstawy SQL: PIVOT i CASE



ALEKSANDER STEPANIUK

NR. INDEKSU: 272644

Politechnika Wrocławska, Informatyka Stosowana

Rozwiązania:

Zadanie 1.

1.

Results													
	Year	1	2	3	4	5	6	7	8	9	10	11	12
1	2011	NULL	NULL	NULL	NULL	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	NULL	NULL	NULL	NULL	NULL	NULL

	Year	Month	Customers
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

Query executed successfully. ZAWO-LAPTOP (16.0 RTM)

--a) zapytanie z użyciem polecenia pivot

```
SELECT * FROM
(
  SELECT
    YEAR(OrderDate) AS Year,
    MONTH(OrderDate) AS Month,
    COUNT(DISTINCT CustomerID) AS Customers
  FROM Sales.SalesOrderHeader
  GROUP BY YEAR(OrderDate), MONTH(OrderDate)
) AS SourceTable
PIVOT
(
  SUM(Customers)
  FOR Month IN ([1], [2], [3], [4], [5], [6], [7], [8], [9], [10], [11], [12])
) AS PivotTable;
```

--b) zapytanie bez użycia polecenia pivot

```
SELECT
  YEAR(OrderDate) AS Year,
  MONTH(OrderDate) AS Month,
  COUNT(DISTINCT CustomerID) AS Customers
FROM Sales.SalesOrderHeader
GROUP BY YEAR(OrderDate), MONTH(OrderDate)
ORDER BY Year, Month;
```

2.

100 %					
Results Messages					
	SalesPerson	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

```
SELECT * FROM
(
  SELECT
    P.FirstName + ' ' + P.LastName AS SalesPerson,
    YEAR(SOH.OrderDate) AS Year,
    COUNT(SOH.SalesOrderID) AS Orders
  FROM Sales.SalesOrderHeader SOH
    JOIN Sales.SalesPerson SP ON SOH.SalesPersonID = SP.BusinessEntityID
    JOIN Person.Person P ON SP.BusinessEntityID = P.BusinessEntityID
  GROUP BY P.FirstName, P.LastName, YEAR(SOH.OrderDate)
) AS SourceTable
PIVOT
(
  SUM(Orders)
  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
```

3.

100 %

Results Messages

	Year	Month	Day	TotalSum	Products
1	2011	5	31	8094970.21	47
2	2011	6	1	15394.33	4
3	2011	6	2	16588.46	4
4	2011	6	3	7907.98	2
5	2011	6	4	16588.46	4
6	2011	6	5	15815.95	3
7	2011	6	6	8680.48	3
8	2011	6	7	8680.48	2
9	2011	6	8	23105.31	5
10	2011	6	9	11664.97	3
11	2011	6	10	15815.95	3
12	2011	6	11	15618.95	4
13	2011	6	12	7907.98	2
14	2011	6	13	27677.92	4
15	2011	6	14	12409.84	4
16	2011	6	15	15815.95	2
17	2011	6	16	15618.95	4
18	2011	6	17	7683.35	2
19	2011	6	18	28253.42	5
20	2011	6	19	12437.47	4
21	2011	6	20	15815.95	2
22	2011	6	21	42256.63	7
23	2011	6	22	16588.46	3

✓ Query executed successfully.

```

SELECT * FROM
(
  SELECT
    YEAR(OrderDate) AS Year,
    MONTH(OrderDate) AS Month,
    DAY(OrderDate) AS Day,
    CAST(SUM(TotalDue) AS DECIMAL(16,2)) AS TotalSum,
    COUNT(DISTINCT ProductID) AS Products
  FROM Sales.SalesOrderHeader SOH
    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID = SOD.SalesOrderID
  GROUP BY YEAR(OrderDate), MONTH(OrderDate), DAY(OrderDate)
) AS SourceTable
ORDER BY Year, Month, Day;

```

4.

100 %

Results Messages

	MonthName	DayName	TotalSum	ProductsCount
1	January	Sunday	57648378.74	134
2	January	Monday	45974239.20	151
3	January	Tuesday	916988.44	110
4	January	Wednesday	76253745.15	157
5	January	Thursday	1053981.70	106
6	January	Friday	823123.02	105
7	January	Saturday	707836.50	100
8	February	Sunday	855245.55	107
9	February	Monday	788985.19	100
10	February	Tuesday	760788.71	103
11	February	Wednesday	17915861.62	139
12	February	Thursday	81580280.80	172
13	February	Friday	825068.79	117
14	February	Saturday	677383.71	105

```

SELECT
    CASE MONTH(OrderDate)
        WHEN 1 THEN 'January'
        WHEN 2 THEN 'February'
        WHEN 3 THEN 'March'
        WHEN 4 THEN 'April'
        WHEN 5 THEN 'May'
        WHEN 6 THEN 'June'
        WHEN 7 THEN 'July'
        WHEN 8 THEN 'August'
        WHEN 9 THEN 'September'
        WHEN 10 THEN 'October'
        WHEN 11 THEN 'November'
        WHEN 12 THEN 'December'
    END MonthName,
    CASE DATEPART(WEEKDAY, 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 DayName,
    CAST(SUM(TotalDue) AS DECIMAL(16,2)) AS TotalSum,
    COUNT(DISTINCT ProductID) AS ProductsCount
FROM Sales.SalesOrderHeader SOH
JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID = SOD.SalesOrderID
GROUP BY MONTH(OrderDate), DATEPART(WEEKDAY, OrderDate)
ORDER BY MONTH(OrderDate), DATEPART(WEEKDAY, OrderDate);

```

5.

100 %

Results Messages

	Imie	Nazwisko	LiczbaTransakcji	SumaKwot	KolorKarty
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	Ackerman	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
19	Carlos	Adams	1	40.84	NULL
20	Charles	Adams	1	37.55	NULL
21	Chloe	Adams	1	1238.14	NULL
22	Connor	Adams	2	3191.86	Silver
23	Courtney	Adams	1	2646.44	NULL

✓ Query executed successfully.

100 %

Results Messages

	KolorKarty	Liczba kart
1	Gold	224
2	Platinum	426
3	Silver	14290
4	NULL	23298

```
CREATE TABLE KartyLojalnoscione (
    Imie NVARCHAR(50),
    Nazwisko NVARCHAR(50),
    LiczbaTransakcji INT,
    SumaKwot DECIMAL(16,2),
    KolorKarty NVARCHAR(50)
);

DECLARE @Srednia DECIMAL(16, 2);
SET @Srednia = (SELECT AVG(TotalDue) FROM Sales.SalesOrderHeader);
```

```

INSERT INTO KartyLojalnoscione (Imie, Nazwisko, LiczbaTransakcji, SumaKwot, KolorKarty)
SELECT
    Imie P.FirstName,
    Nazwisko P.LastName,
    LiczbaTransakcji T.liczba,
    SumaKwot T.kwota,
    CASE
        WHEN T.liczba ≥ 4
            AND T.duzeZamowienia ≥ 2
            AND EXISTS (
                SELECT 1
                FROM Sales.SalesOrderHeader SOH
                    JOIN Sales.SalesOrderDetail SOD ON SOH.SalesOrderID = SOD.SalesOrderID
                    JOIN Production.Product PP ON SOD.ProductID = PP.ProductID
                    JOIN Production.ProductSubcategory PSC ON PP.ProductSubcategoryID = PSC.ProductSubcategoryID
                    JOIN Production.ProductCategory PC ON PSC.ProductCategoryID = PC.ProductCategoryID
                WHERE SOH.CustomerID = T.CustomerID
                GROUP BY SOH.SalesOrderID
                HAVING COUNT(DISTINCT PC.ProductCategoryID) =
                    (SELECT COUNT(*) FROM Production.ProductCategory)
            ) THEN 'Platinum'
        WHEN T.liczba ≥ 4 AND T.duzeZamowienia ≥ 2 THEN 'Gold'
        WHEN T.liczba ≥ 2 THEN 'Silver'
        ELSE NULL
    END
FROM
    (
        SELECT
            C.CustomerID,
            COUNT(SOH.SalesOrderID) AS liczba,
            SUM(SOH.TotalDue) AS kwota,
            SUM(CASE WHEN SOH.TotalDue > 2.5 * @Srednia THEN 1 ELSE 0 END) AS duzeZamowienia
        FROM Sales.Customer C
            JOIN Sales.SalesOrderHeader SOH ON C.CustomerID = SOH.CustomerID
        WHERE C.PersonID IS NOT NULL -- pomijamy firmy które nie mają danych osobowych w tabeli Person.Person
        GROUP BY C.CustomerID
    ) AS T
    JOIN Sales.Customer C ON T.CustomerID = C.CustomerID
    JOIN Person.Person P ON C.PersonID = P.BusinessEntityID;

```

```

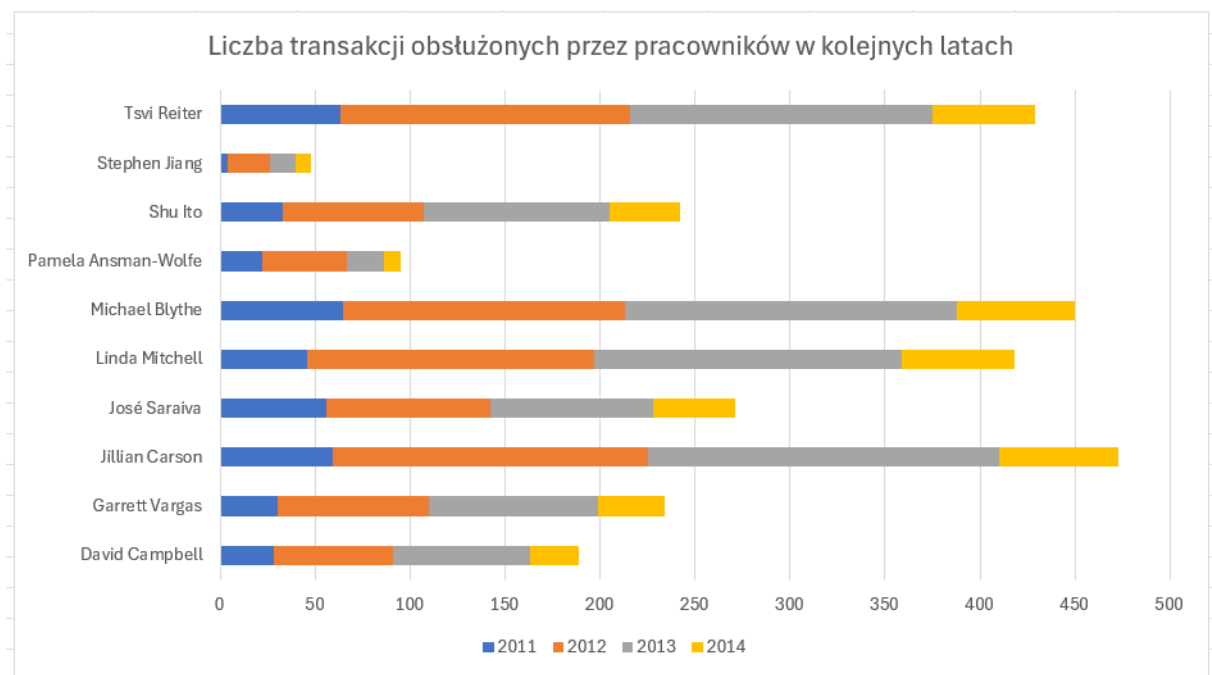
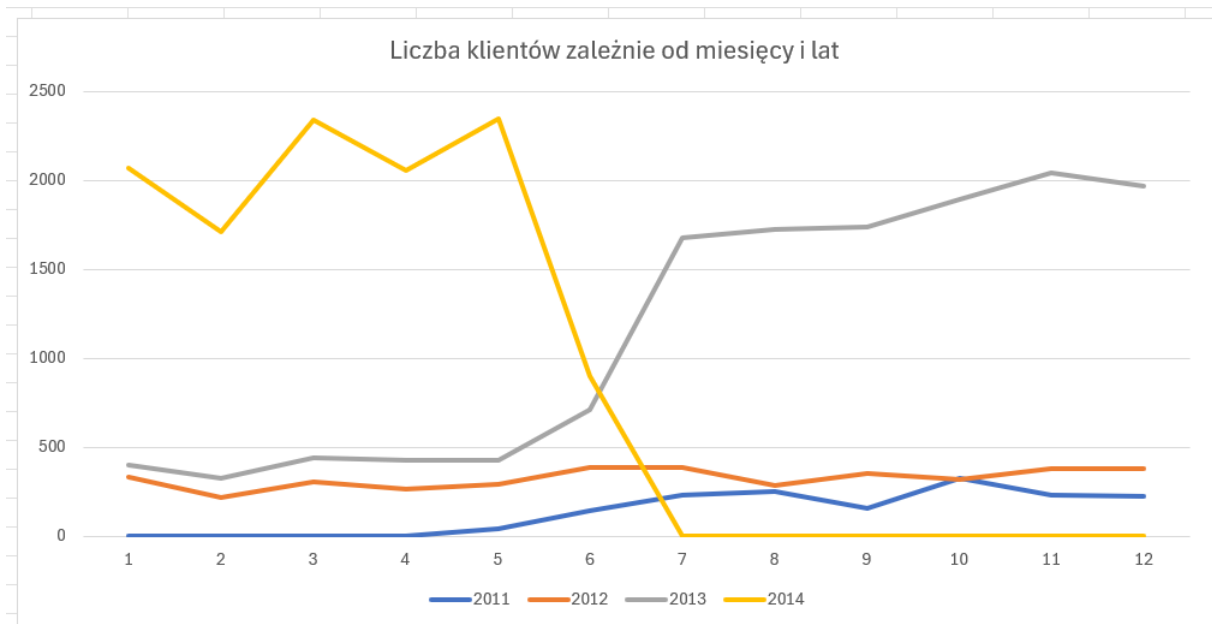
-- test1
SELECT * FROM KartyLojalnoscione;

-- test2
SELECT KolorKarty, COUNT(*) as "Liczba kart"
FROM KartyLojalnoscione
GROUP BY KolorKarty
ORDER BY 2;

```

Zadanie 2.

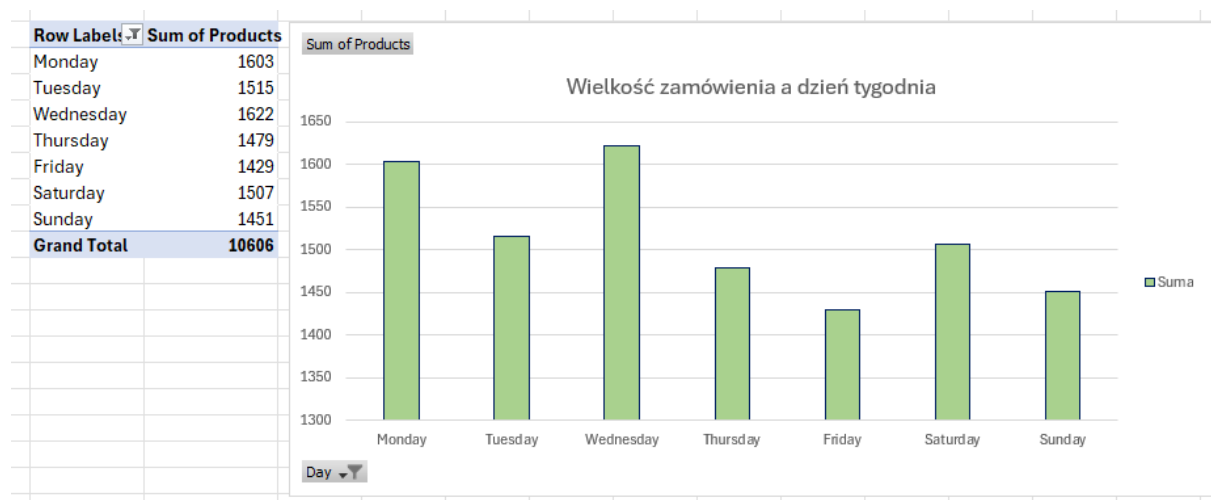
1. Excel



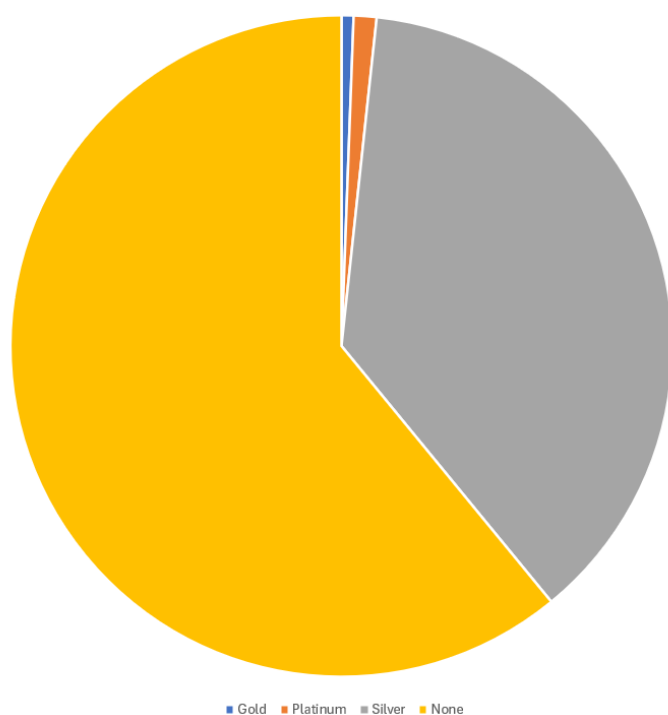
Kwota zamówienia, a miesiąc



Wielkość zamówienia, a dzień tygodnia

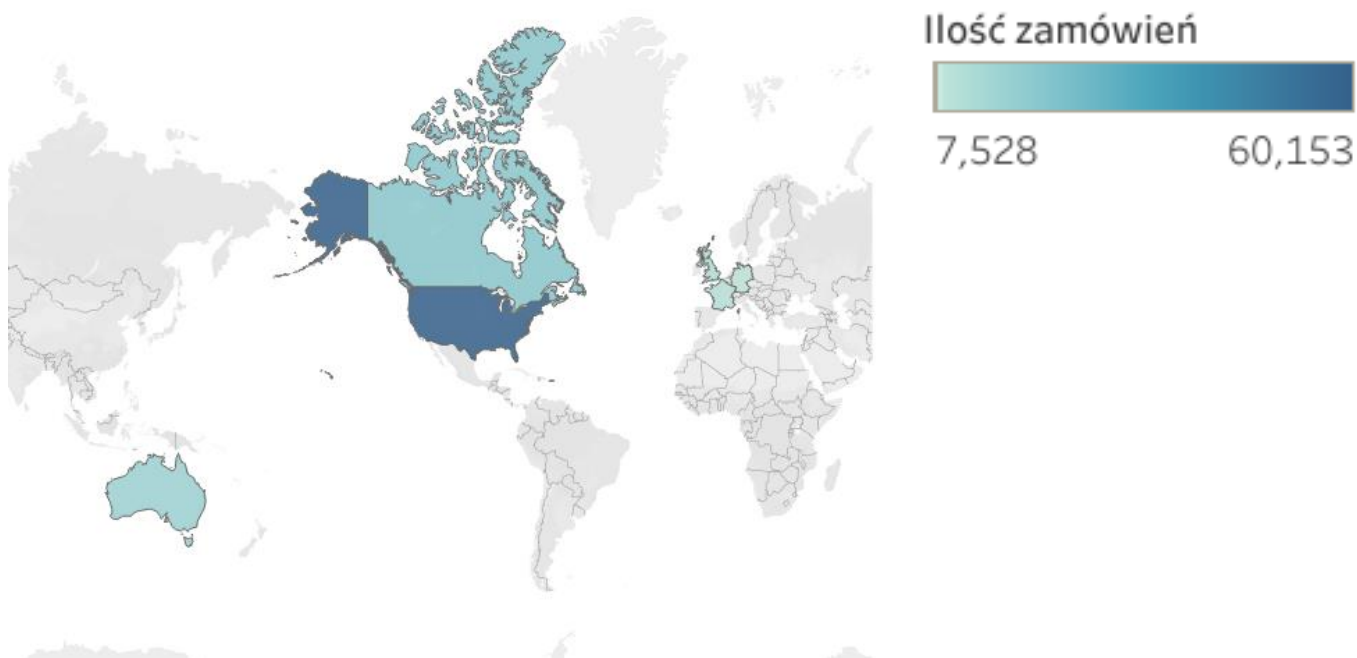


Liczba kart

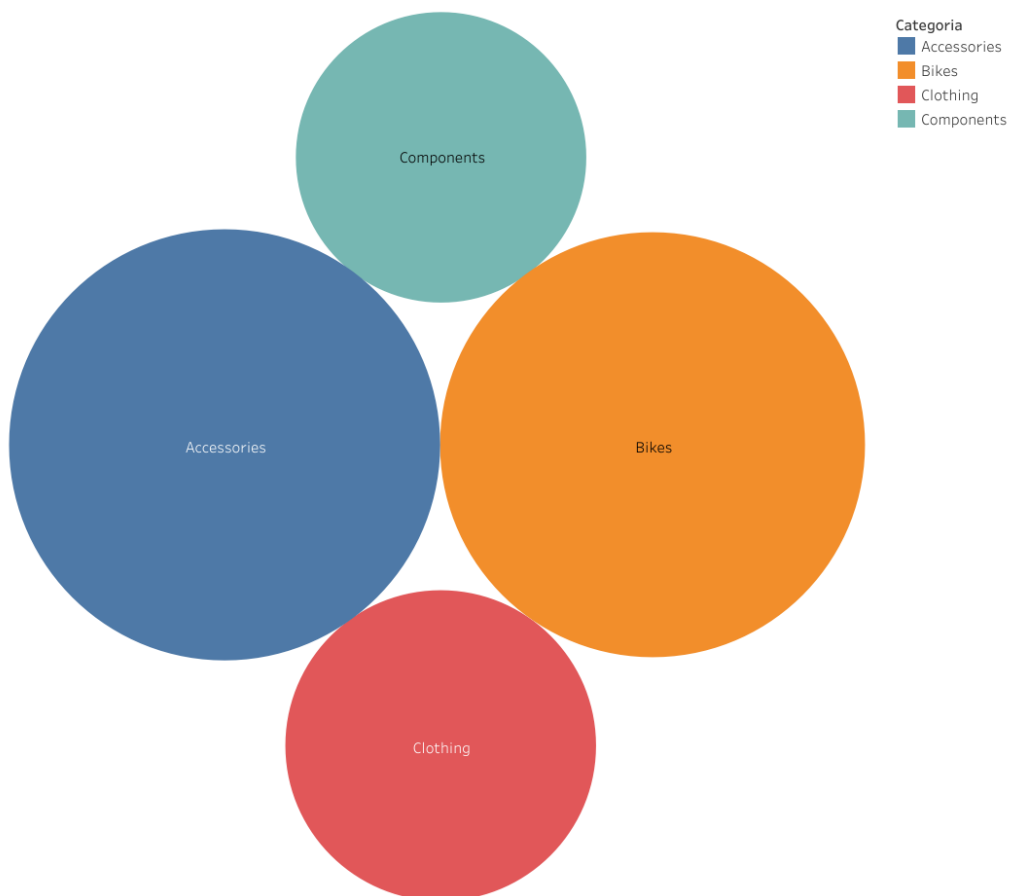


2. Tableau

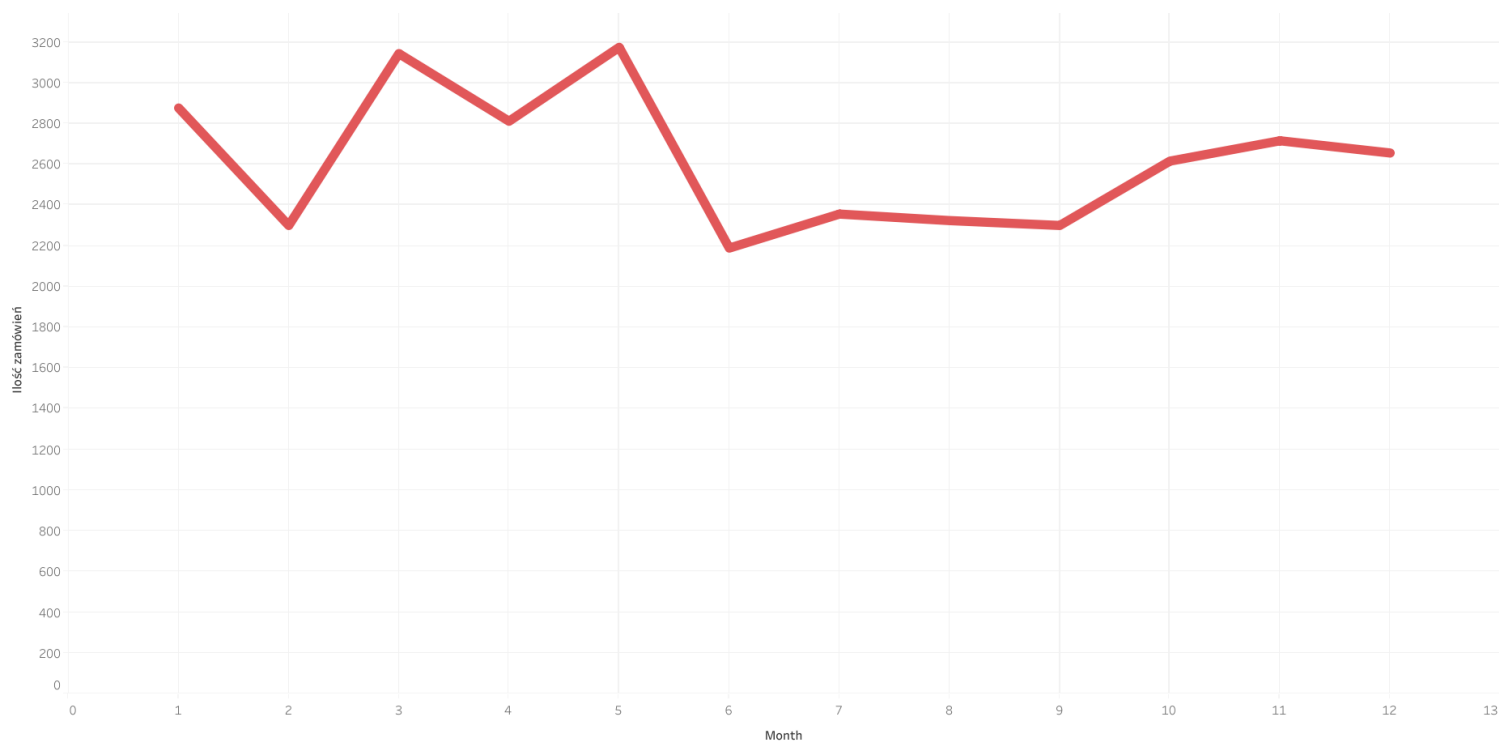
Ilość zamówień w poszczególnych regionach:



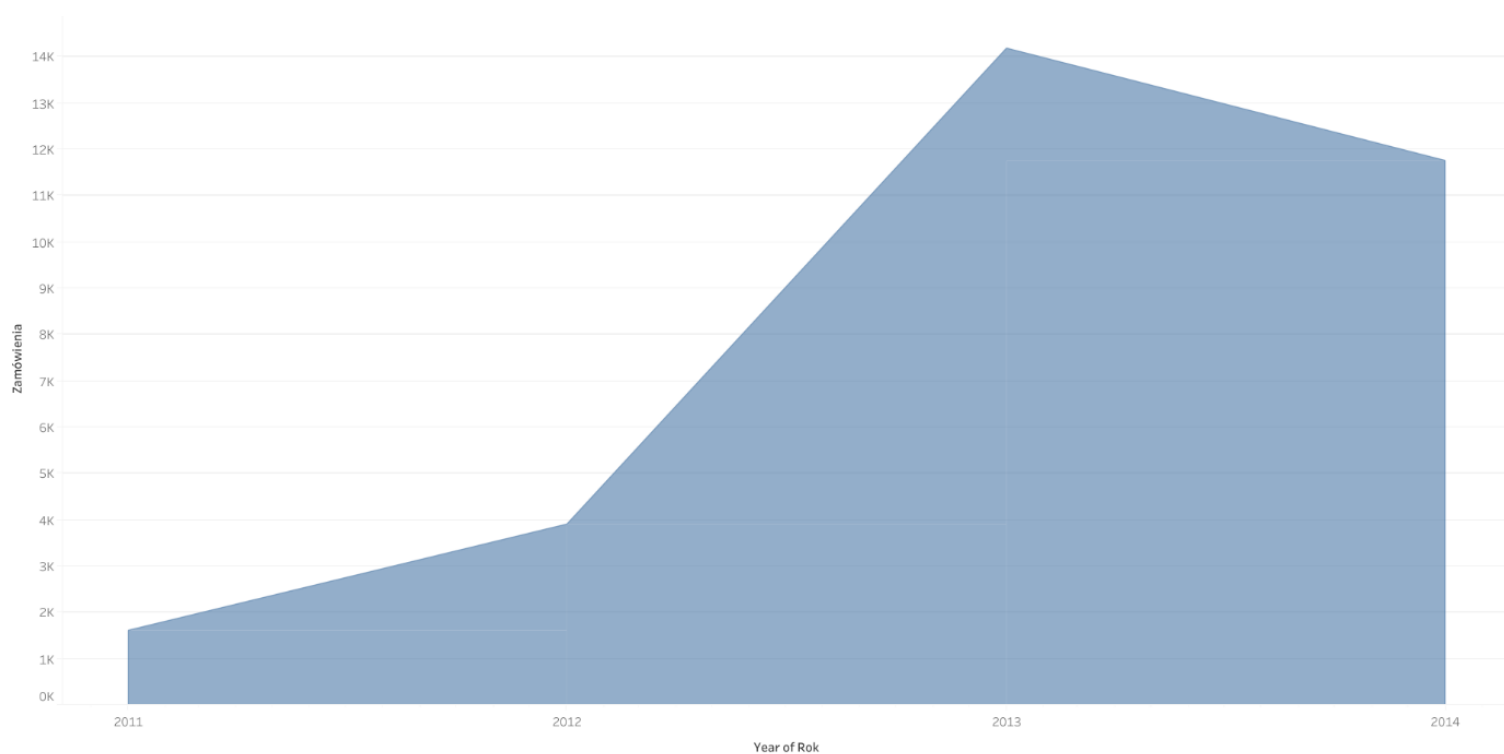
Wykres kołowy, przedstawiający ilość zamówień w poszczególnych kategoriach:



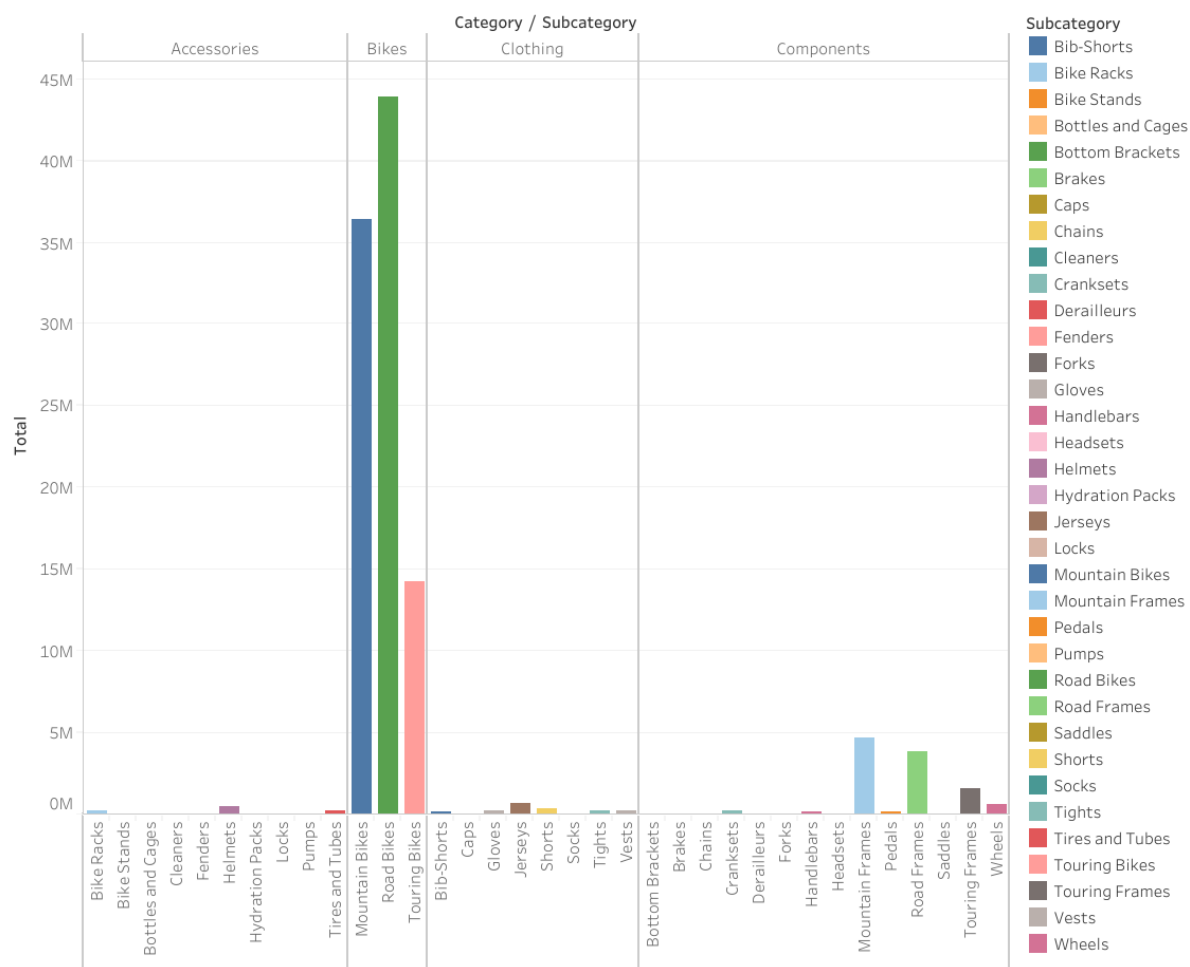
Wykres liniowy pokazujący liczbę zamówień w zależności od miesiąca:



Wykres liniowy pokazujący ilość zamówień w poszczególnych latach.



Wykres obrazujący wartość sprzedanych produktów ze względu na kategorie i podkategorie:



Wnioski:

Zapytania PIVOT są o wiele bardziej czytelne (mniejsza ilość rekordów, łatwiej czytelna dla ludzkiego oka) i proste do dalszego przetwarzania (np. wyświetlenia ich jako wykres).

Widać wyraźnie, że w roku 2011 liczba różnych klientów jest bardzo niska, co sugerowało by, że wtedy firma rozpoczęła swoją działalność rynkową. Bardzo duży pik w ilości klientów widać w połowie roku 2013, kiedy ich liczba wzrasta bardzo dynamicznie i do końca roku osiąga 2000 różnych klientów miesięcznie.

Widoczne jest także, że miesiące cieplejsze (maj, czerwiec, lipiec) są najbardziej korzystne dla sprzedaży (może to wynikać z tego, że jest to sklep rowerowy i ludzie chcą kupować sprzęt na ładną pogodę która wtedy statystycznie następuje).

Spośród wszystkich 17 sprzedawców, jedynie 10 z nich pracowało w firmie przez wszystkie 4 lata. Jest to relatywnie duża liczba i może sugerować że warunki pracy w firmie są wystarczająco dobre dla większości pracowników, aby pozostać w niej zatrudnionym przez 4 lata. Stephen Jiang oraz Pamela Ansman-Wolfe mają wyjątkowo małą ilość obsłużonych transakcji – może to sugerować, że albo region w którym pracują nie jest dość obfity w klientów lub to sami pracownicy mogą robić coś nie tak. Julian Carson, Tsvi Reiter, Micheal Blythe oraz Linda Mitchell mają bardzo dobre wyniki ilości obsłużonych transakcji, więc ich region jest bardzo obfity w klientów i/lub bardzo dobrze spełniają oni swoje obowiązki.

Około 2/3 klientów (prawie 24000 osób) nie posiada żadnej karty lojalnościowej. Blisko 1/4 (ponad 14000) z nich posiada srebrną kartę. Lekko ponad 200 osób posiada kartę złotą, natomiast ponad 400 posiada kartę platynową. Rozkład ten jest bardzo daleki od rozkładu naturalnego, więc można się zastanowić nad zmianą reguł przydzielania kart lojalnościowych, tak aby lepiej nagradzać przywiązanie do firmy.

Największe obroty widać na przełomie wiosny i lata (od marca do lipca), najmniejsze jesienią i zimą. Lata 2011 oraz 2014 nie mają pełnych danych, więc nagły spadek w roku 2014 do zera jest jedynie pozorny.

Zaskakującym wnioskiem jest to, że w poniedziałki i środy sprzedawane jest najwięcej produktów (ilościowo), więcej nawet niż w weekendy. Nie jest to jednak duża różnica, ponieważ wykres nie zaczyna się od poziomu zera. Drugim ciekawym wnioskiem jest fakt, że pomimo ogólnego trendu większej sprzedawalności w czasie wiosenno letnim – kwiecień pozostaje miesiącem z widocznie niższą sprzedażą niż marzec czy maj. Wynika to jednak z faktu, że na początku każdego miesiąca (a czasami na samym końcu poprzedniego miesiąca) jest bardzo widoczny wzrost zarówno w liczbie zamówień jak i w ich kwocie. Może to być związane jakimiś regulacjami faktur albo innego rodzaju biurokracją. Kwoty te są bardzo duże i „wyginają” końcowy wykres na rzecz sąsiednich miesięcy.

Najwięcej zamówień pochodzi z USA. Najgorszym regionem jest Europa, czyli Niemcy i Francja.

Największa wartość sprzedanych produktów to wszystkie podkategorie rowerów. Road i Mountain Bikes mają największą sprzedaż, a Tourist Bike ma nieco niższą. Wśród komponentów widać również nieco wyższe kwoty dla ram rowerowych (Mountain Frame, Road Frame) niż pozostałe komponenty, ponieważ są one droższe niż inne komponenty.