Hurtownie Danych - laboratorium Lista 5

Analysis Services

Zadania do wykonania

Baza danych: AdventureWorks

Zad. 1. Wymiar czasowy

Przygotować wymiar czasowy: utworzyć i wypełnić danymi tabelę DIM_TIME. Tabela DIM_TIME powinna być tabelą zawierającą wymiar czasowy (klucze obce do tej tabeli znajdują się w tabeli faktów).

Tabela DIM TIME powinna zawierać następujące kolumny:

- PK_TIME (klucz główny liczba całkowita postaci yyyymmdd format taki sam jak kolumn OrderDate, ShipDate)
- Rok
- Kwartał
- Miesiąc
- Miesiąc słownie (wykorzystać tabelę pomocniczą z 12 rekordami dokonać odpowiedniego złączenia)
- Dzień tygodnia słownie (wykorzystać tabelę pomocniczą z 7 rekordami dokonać odpowiedniego złączenia)
- Dzień miesiąca

Rozwiązania:

```
Stworzenie tabeli pomocniczej z nazwami miesięcy w języku polskim (12 rekordów):
CREATE TABLE STROZIK.months_names (
    month_number INTEGER,
    month_name VARCHAR(20)
);
INSERT INTO STROZIK.months_names (month_number, month_name)
VALUES (1, 'Styczeń'),
        (2, 'Luty'),
        (3, 'Marzec'),
        (4, 'Kwiecień'),
        (5, 'Maj'),
        (6, 'Czerwiec'),
        (7, 'Lipiec'),
        (8, 'Sierpień'),
        (9, 'Wrzesień'),
        (10, 'Październik'),
        (11, 'Listopad'),
        (12, 'Grudzień');

SELECT * FROM STROZIK.months_names;
```

| | month_number | month_name |
|----|--------------|------------|
| 1 | 1 | Styczen |
| 2 | 2 | Luty |
| 3 | 3 | Marzec |
| 4 | 4 | Kwiecien |
| 5 | 5 | Maj |
| 6 | 6 | Czerwiec |
| 7 | 7 | Lipiec |
| 8 | 8 | Sierpien |
| 9 | 9 | Wrzesien |
| 10 | 10 | Pazdziemik |
| 11 | 11 | Listopad |
| 12 | 12 | Grudzien |

```
Stworzenie tabeli pomocniczej z nazwami dni tygodnia w języku polskim (7 rekordów):

CREATE TABLE STROZIK.weekday_names (
    weekday_number INTEGER,
    day_name VARCHAR(20)
);

INSERT INTO STROZIK.weekday_names (weekday_number, day_name)

VALUES (1, 'Poniedziałek'),
    (2, 'Wtorek'),
    (3, 'Środa'),
    (4, 'Czwartek'),
    (5, 'Piątek'),
    (6, 'Sobota'),
    (7, 'Niedziela');
```

SELECT * FROM STROZIK.weekday_names;

| | weekday_number | day_name |
|---|----------------|--------------|
| 1 | 1 | Poniedzialek |
| 2 | 2 | Wtorek |
| 3 | 3 | Sroda |
| 4 | 4 | Czwartek |
| 5 | 5 | Piatek |
| 6 | 6 | Sobota |
| 7 | 7 | Niedziela |

Stworzenie wymiaru czasu STROZIK.DIM_TIME:

```
CREATE TABLE STROZIK.DIM_TIME (
PK_TIME DATETIME PRIMARY KEY,
Rok INT,
Kwartal INT,
Miesiac INT,
Miesiac_slownie VARCHAR(20),
Dzien_tyg_slownie VARCHAR(20),
Dzien_miesiaca INT
);
```

Tutaj pewna uwaga dotycząca typu dla atrybutu i klucza głównego PK_TIME, dla bazy, na której aktualnie wykonywane są listy, atrybuty OrderDate, ShipDate z tabeli Sales.SalesOrder-Header posiadają typ DATETIME. To samo było wskazywane w raporcie dotyczącym listy 4. Zdecydowano, że dalej będziemy wykorzystywać ten typ w celu zachowania spójności między rozwiązaniami listy 4 i listy 5. Po krótkiej analizie stwierdzono, że nie wpłynie to znacząco na dalsze rozwiązania listy.

Wypełnienie danymi wymiaru czasowego odpowiednio złączając tabele pomocnicze:

```
INSERT INTO STROZIK.DIM_TIME
SELECT DISTINCT
    OrderDate AS PK_TIME,
    YEAR(OrderDate) AS Rok,
    DATEPART(Q, OrderDate) AS Kwartal,
    MONTH(OrderDate) AS Miesiac,
        mn.month_name AS Miesiac_slownie,
        wdn.day_name AS Dzien_tyg_slownie,
        DAY(OrderDate) AS Dzien_miesiaca
FROM STROZIK.FACT_SALES
    JOIN STROZIK.weekday_names wdn
    ON DATEPART(DW, OrderDate) = wdn.weekday_number
    JOIN STROZIK.months_names mn
    ON MONTH(OrderDate) = mn.month_number;

SELECT * FROM STROZIK.DIM TIME;
```

Filip Strózik 260377

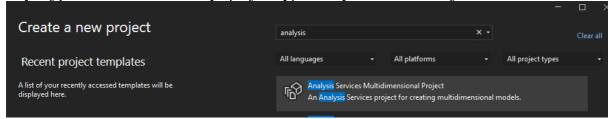
| | PK_TIME | Rok | Kwartal | Miesiac | Miesiac_slownie | Dzien_tyg_slownie | Dzien_miesiaca |
|----|-------------------------|------|---------|---------|-----------------|-------------------|----------------|
| 1 | 2011-05-31 00:00:00.000 | 2011 | 2 | 5 | Maj | Sroda | 31 |
| 2 | 2011-06-01 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Czwartek | 1 |
| 3 | 2011-06-02 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Piatek | 2 |
| 4 | 2011-06-03 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Sobota | 3 |
| 5 | 2011-06-04 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Niedziela | 4 |
| 6 | 2011-06-05 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Poniedzialek | 5 |
| 7 | 2011-06-06 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Wtorek | 6 |
| 8 | 2011-06-07 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Sroda | 7 |
| 9 | 2011-06-08 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Czwartek | 8 |
| 10 | 2011-06-09 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Piatek | 9 |
| 11 | 2011-06-10 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Sobota | 10 |
| 12 | 2011-06-11 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Niedziela | 11 |
| 13 | 2011-06-12 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Poniedzialek | 12 |
| 14 | 2011-06-13 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Wtorek | 13 |
| 15 | 2011-06-14 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Sroda | 14 |
| 16 | 2011-06-15 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Czwartek | 15 |
| 17 | 2011-06-16 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Piatek | 16 |
| 18 | 2011-06-17 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Sobota | 17 |
| 19 | 2011-06-18 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Niedziela | 18 |
| 20 | 2011-06-19 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Poniedzialek | 19 |
| 21 | 2011-06-20 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Wtorek | 20 |
| 22 | 2011-06-21 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Sroda | 21 |
| 23 | 2011-06-22 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Czwartek | 22 |
| 24 | 2011-06-23 00:00:00.000 | 2011 | 2 | 6 | Czerwiec | Piatek | 23 |

Na koniec należało wprowadzić więzły integralności:
ALTER TABLE STROZIK.FACT_SALES
ADD CONSTRAINT FK_TIMEID FOREIGN KEY (OrderDate) REFERENCES STROZIK.DIM_TIME(PK_TIME);

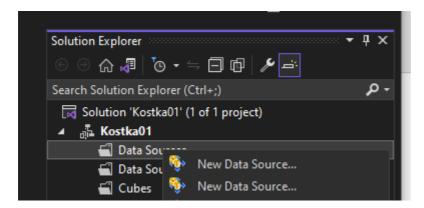
Zad. 2. Tworzenie kostki

Należy utworzyć projekt Analysis Services, w którym zostanie przygotowana kostka zawierająca utworzone wymiary (klienta, produktu, sprzedawcy i czasu) oraz tabelę faktów.

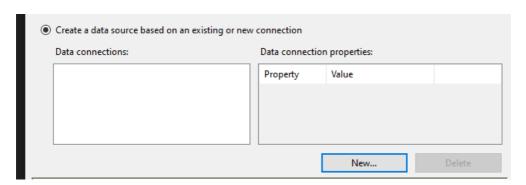
Używając Visual Studio utworzyć projekt typu Analysis Services Project



a) Dodać źródło danych, które będzie wskazywało na bazę danych, która przechowuje tabele faktów i wymiarów.



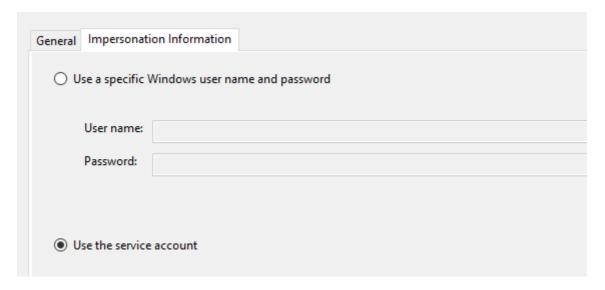
Utworzyć nowe połącznie do serwera bazy danych:



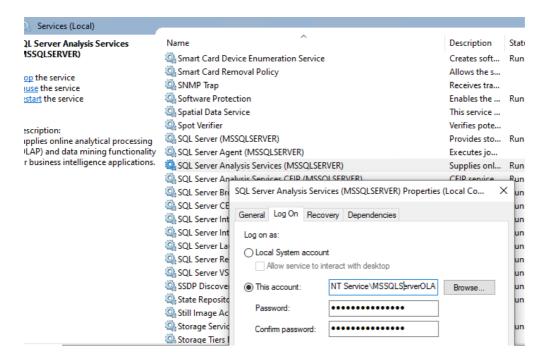
Podając nazwę serwera i baze danych, wktórej utworzone tabale wymiarów i faktów:



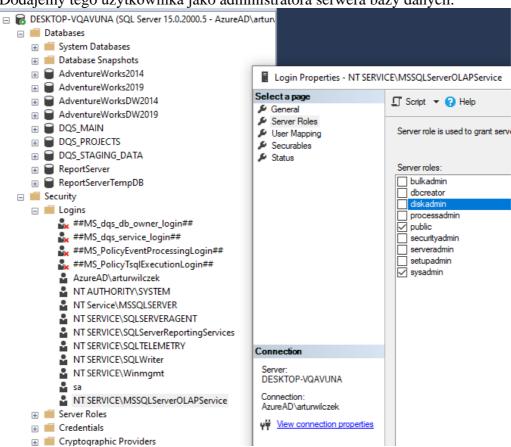
W zakładce Impersonation Information wybieramy "Use the service account":



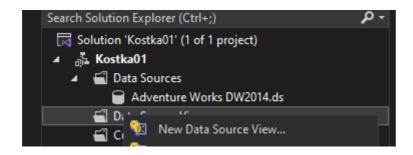
W ustawieniach usług sprawdzamy użytkownika na którym pracuje usługa Analysis Services:



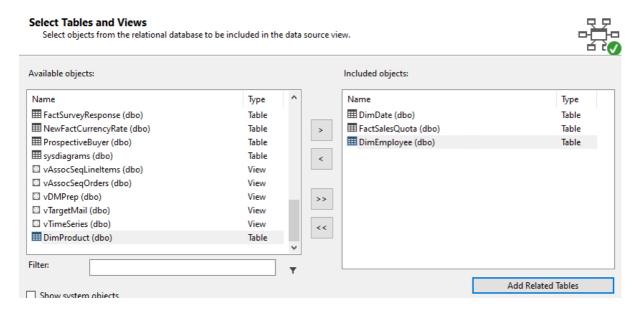
Dodajemy tego użytkownika jako administratora serwera bazy danych:



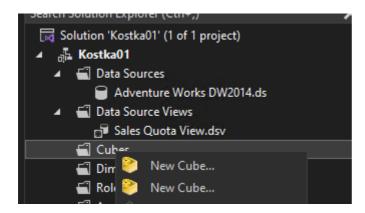
b) Utworzyć nowy widok źródła danych. Dodać wcześniej utworzone tabele.



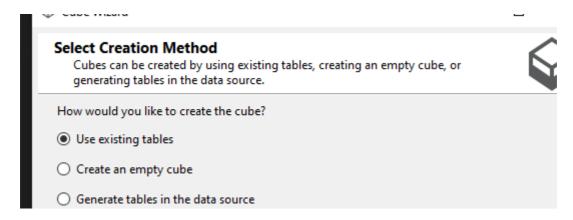
Dodać wcześniej utworzone tabele. Sprawdzić działanie funkcji "Add Related Tables":



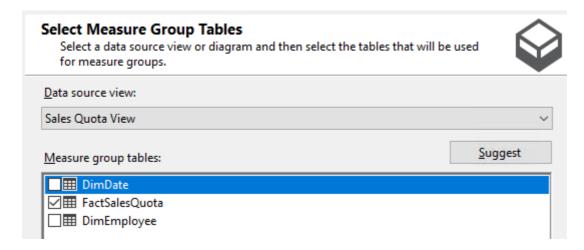
c) Utworzyć nową kostkę za pomocą asystenta



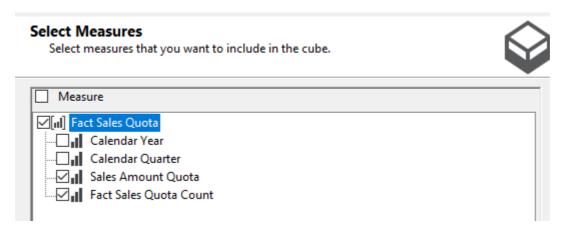
Wybrać utworzenie kostki na podstawie istniejących tabel:



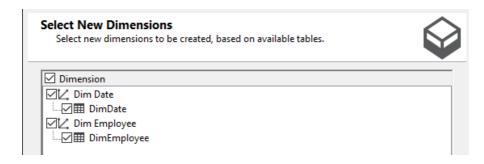
Wybrać, utworzony w poprzednim punkcie, widok źródła danych. Jako tabelę faktów (Measure group tables) wybrać FACT_SALES:



Na stronie dotyczącej miar wybrać OrderQty, UnitPriceDiscount, Line Total. Zastanowić się nad użytecznością wybranych miar. Dlaczego nie wszystkie atrybuty tabeli FACT_SALES mogą być użyte jako miary?

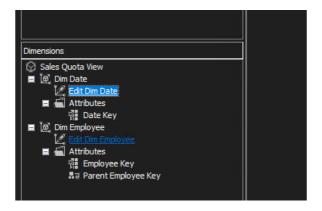


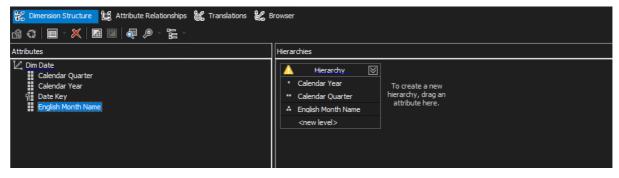
Na stronie dotyczącej wymiarów wybrać wszystkie tabele z przedrostkiem DIM.



Po utworzeniu kostki dokonać edycji wymiarów (Solution Explorer -> Dimensions lub zakładka Cube Structure -> Dimensions).

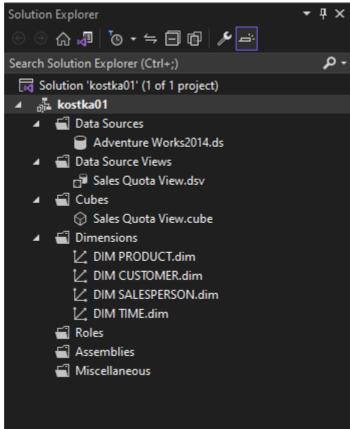
Dla każdego z wymiarów zdefiniować potrzebne atrybuty i hierarchie dla wymiaru czasu. Przykładowo wymiar produkt powinien zawierać: Nazwę, Cenę, Kolor, Podkategorię i Kategorię.



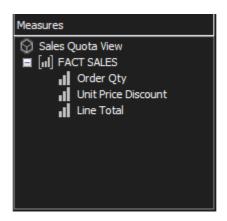


Rozwiązanie:

Rozwiązanie projektu:



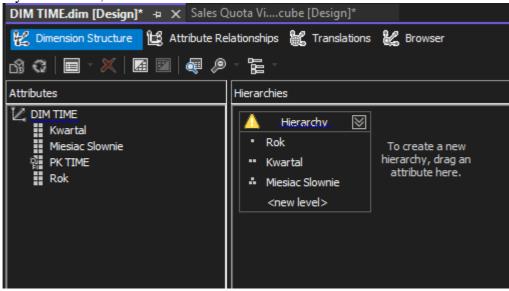
Miary kostki:

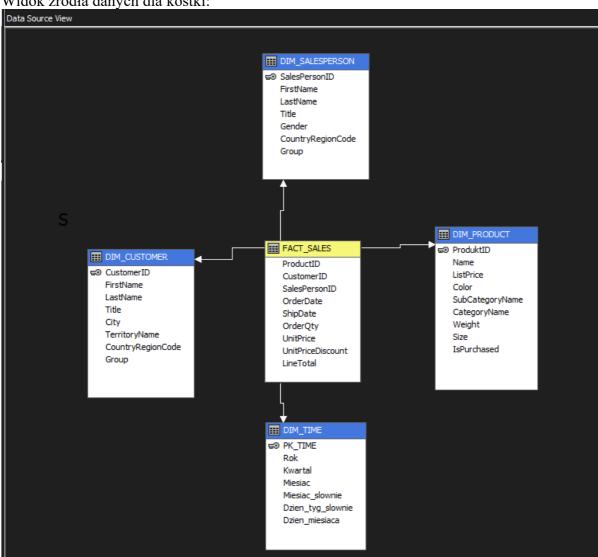


Wymiary kostki:



Edycja wymiaru czasu, hierarchia:



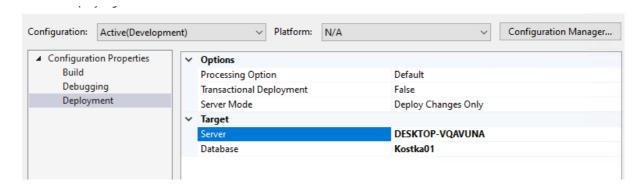


Widok źródła danych dla kostki:

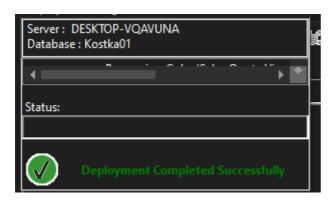
Tabela FACT SALES zawiera dane numeryczne, które mogą być użyte jako miary, takie jak ilość sprzedanego towaru, cena, wartość sprzedaży, itp. Natomiast atrybuty nie numeryczne, takie jak nazwa produktu, nazwa klienta, czy data sprzedaży, nie mogą być używane jako miary, ponieważ nie mają wartości numerycznych, które można agregować, sumować, czy obliczać średnie. Zamiast tego, atrybuty te mogą być wykorzystane jako wymiary, czyli jako kategorie, po których można grupować dane i wykonywać na nich analizy. Na przykład, można zgrupować sprzedaż według nazwy produktu, daty, czy nazwy klienta, aby uzyskać statystyki sprzedaży dla tych wymiarów.

Zad. 3. Uruchomienie kostki

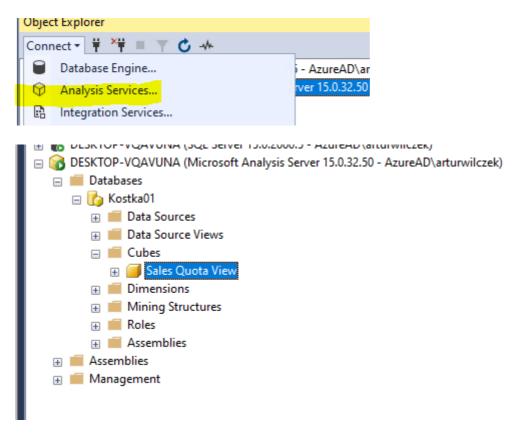
- a) Sprawdzić ustawienia projektu (Solution Explorer -> Nazwa projektu -> Properties).
- b) Sprawdzić ustawienia konfiguracji wdrożenia (Configuration properties -> Deployment -> Target) projektu - Server: [NAZWA], Database: Kostka01



c) Wdrożyć projekt (Menu Build -> Deploy).

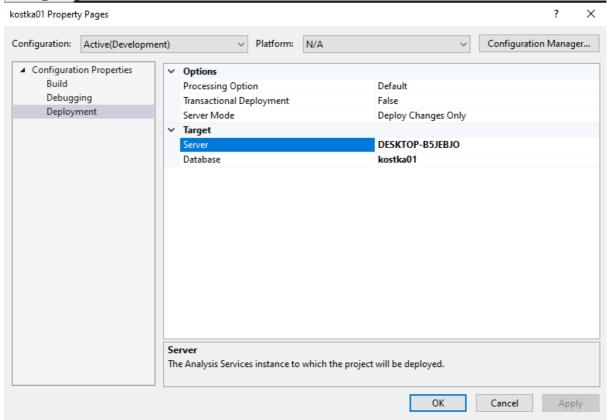


- d) Przetworzyć kostkę (Przycisk Process).
- e) Sprawdź dostępność kostki w Analysis Services (SQL Server Management Studio/New Connection):

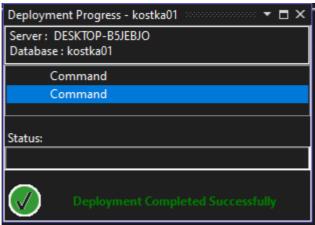


Rozwiązanie:

Konfiguracja wdrożenia kostki:

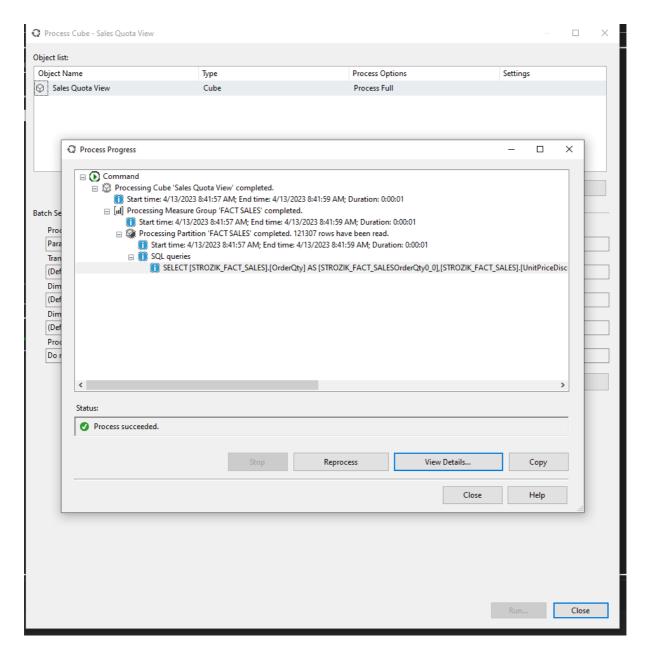


Status wdrożenia kostki:

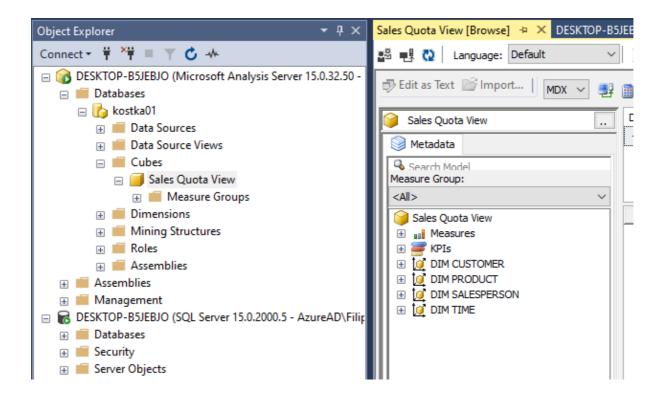


Przetworzenie kostki:

Filip Strózik 260377



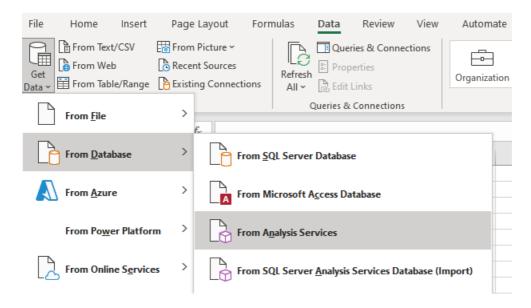
Dostępność kostki w SSMS:



Zad. 4. Przykładowe raporty

Połączyć się z programem MS Excel i przygotować 4 przykładowe raporty (tabele i wykresy przestawne), w których zostaną ujęte ciekawe zależności pomiędzy danymi.

Połączenie powinno być zbudowane do Analysis Services:

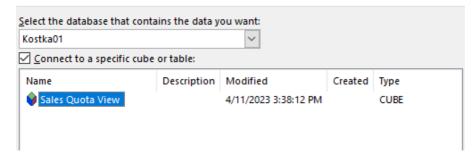


Data Connection Wizard

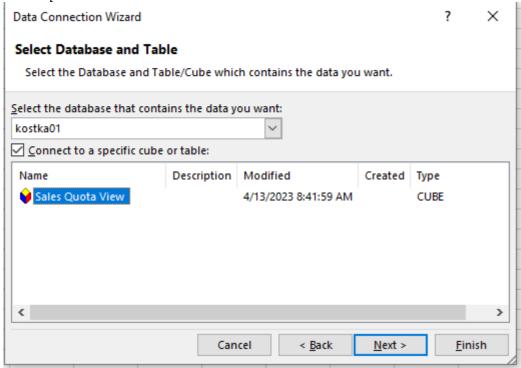


Select Database and Table

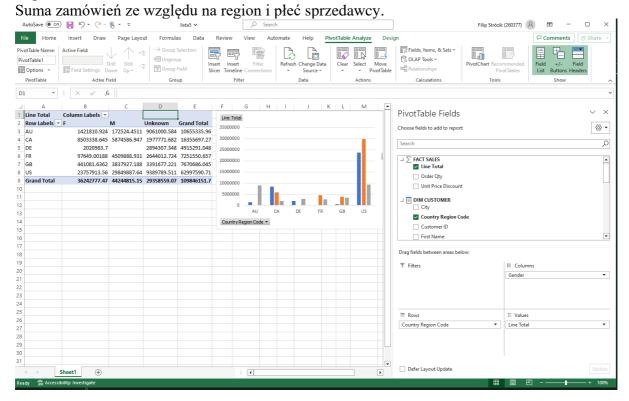
Select the Database and Table/Cube which contains the data you want.



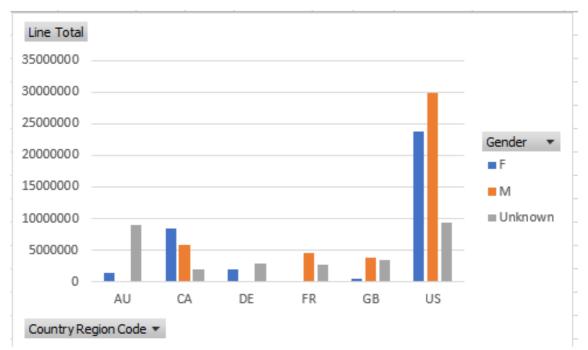
Rozwiązania:



Raport 1:

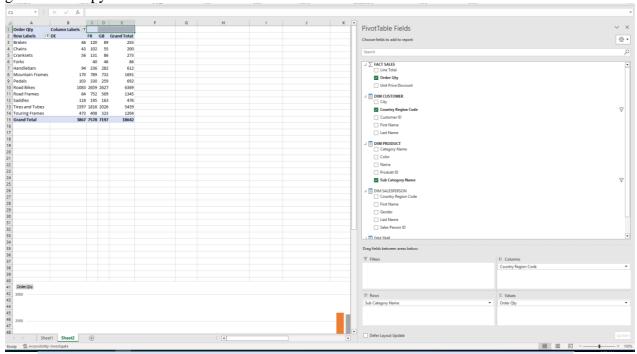


| Line Total Colu | umn Labels 🔻 | | | |
|-----------------|--------------|-------------|-------------|--------------------|
| Row Labels ▼ F | | М | Unknown | Grand Total |
| AU | 1421810.924 | 172524.4511 | 9061000.584 | 10655335.96 |
| CA | 8503338.645 | 5874586.947 | 1977771.682 | 16355697.27 |
| DE | 2020983.7 | | 2894307.348 | 4915291.048 |
| FR | 97649.00188 | 4509888.931 | 2644012.724 | 7251550.657 |
| GB | 441081.6362 | 3837927.188 | 3391677.221 | 7670686.045 |
| US | 23757913.56 | 29849887.64 | 9389789.511 | 62997590.71 |
| Grand Total | 36242777.47 | 44244815.15 | 29358559.07 | 109846151.7 |

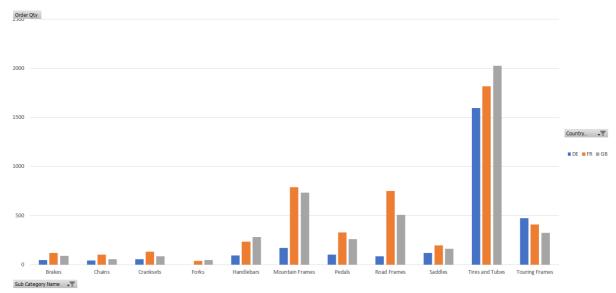


Widać, że kraje leżące na kontynencie Ameryka północna mają większą ilość jakichkolwiek zaznaczeń płci. W Kanadzie sumarycznie więcej sprzedały kobiety niż mężczyźni w sklepie. W Niemczech możemy znaleźć brak informacji o sumie sprzedanej przez mężczyzn w sklepie. Dostępnych sprzedawców dla sklepu jest dość mała liczba, dlatego możemy się domyśleć, że pracownika płci męskiej w regionie DE po prostu nie ma.

Raport 2: Zestawienie ilości dokonanych zamówień na części składające się na budowę roweru w regionach Europy.

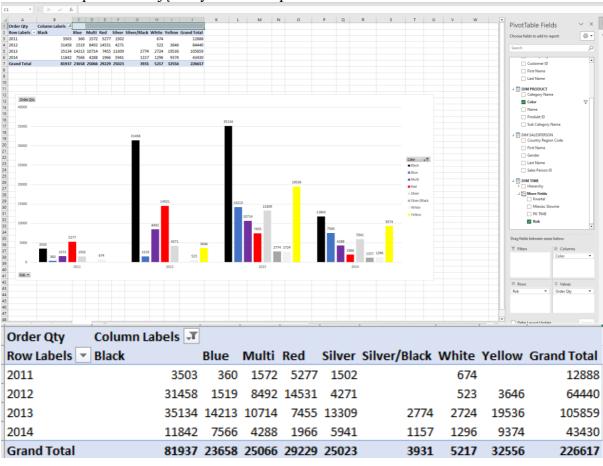


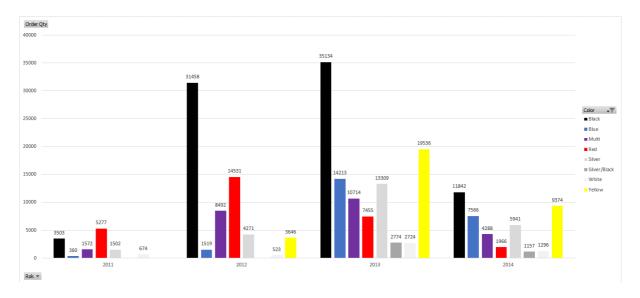
| Order Qty | Column Labels 📧 | | | |
|-----------------|-----------------|------|------|--------------------|
| Row Labels | DE | FR | GB | Grand Total |
| Brakes | 46 | 120 | 89 | 255 |
| Chains | 43 | 102 | 55 | 200 |
| Cranksets | 56 | 131 | 86 | 273 |
| Forks | | 40 | 46 | 86 |
| Handlebars | 94 | 236 | 282 | 612 |
| Mountain Frames | 170 | 789 | 732 | 1691 |
| Pedals | 103 | 330 | 259 | 692 |
| Road Frames | 84 | 752 | 509 | 1345 |
| Saddles | 118 | 195 | 163 | 476 |
| Tires and Tubes | 1597 | 1816 | 2026 | 5439 |
| Touring Frames | 473 | 408 | 323 | 1204 |
| Grand Total | 2784 | 4919 | 4570 | 12273 |



Oprócz oczywistego, że bardzo dużo jest zamówień na opony i dętki, to można zauważyć, że region niemiecki ma z reguły mniej zamówień w każdej części rowerowej, no chyba, że ramy do rowerów trekkingowych, co mogłoby być odpowiadające ukształtowaniu terenu w samych Niemczech. Od razu widać, że Francja ma najwięcej zamówień na ramy do rowerów szosowych, co być może jest spowodowane sławnym, kultowym wyścigiem kolarskim Tour de France. Równie dużo jest zamówień na ramy do rowerów górskich, co może być spowodowane ukształtowaniem terenu oraz mnogością ścieżek górskich.

Raport 3: Zestawienie przedstawiające wybór koloru przedmiotu wraz z latami.

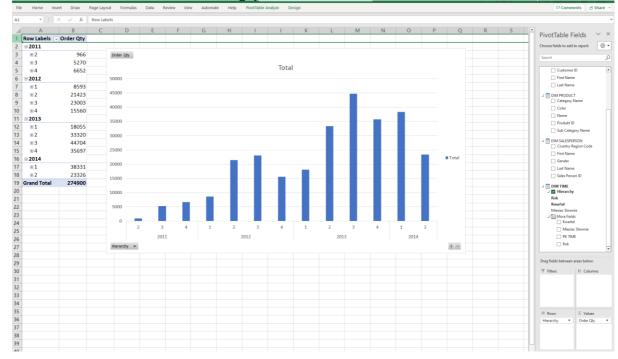




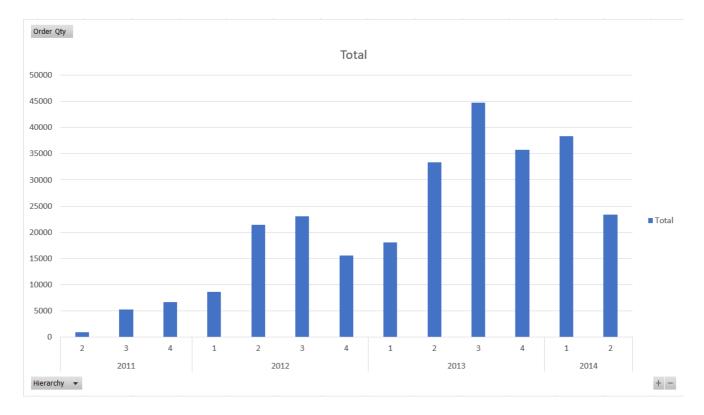
Widać, że dla początkowych lat czarny oraz czerwony, były głównymi wyborami klientów. Potem popularność koloru czerwonego spadała, a kolor żółty, niebieski i szarości zyskały. Co do wyborów klientów między rokiem 2013 i 2014 to tendencje się znacznie nie zmieniły tylko odnotowano ogólny spadek zamówień.

Raport 4:

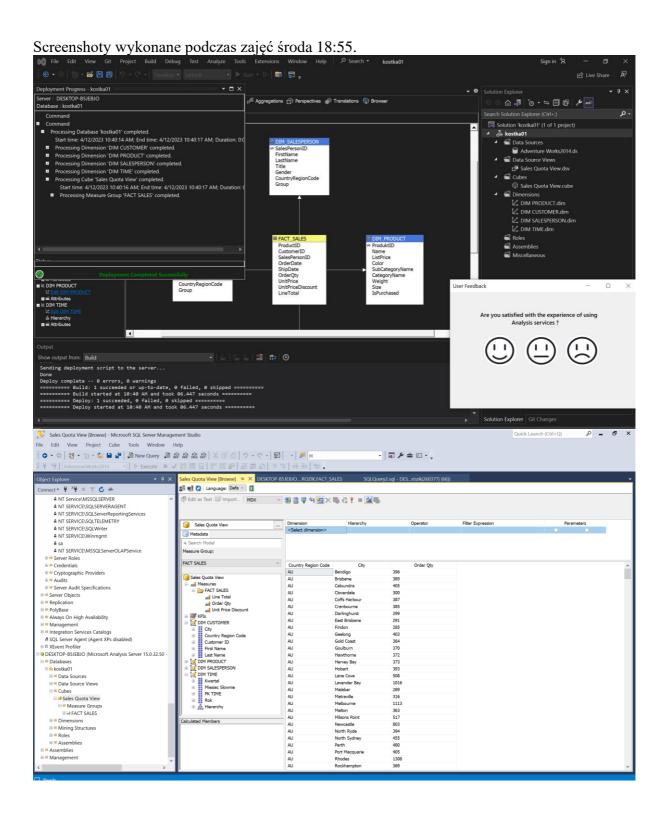
Zestawienie ilości zamówień ze względu na Rok, Kwartał.

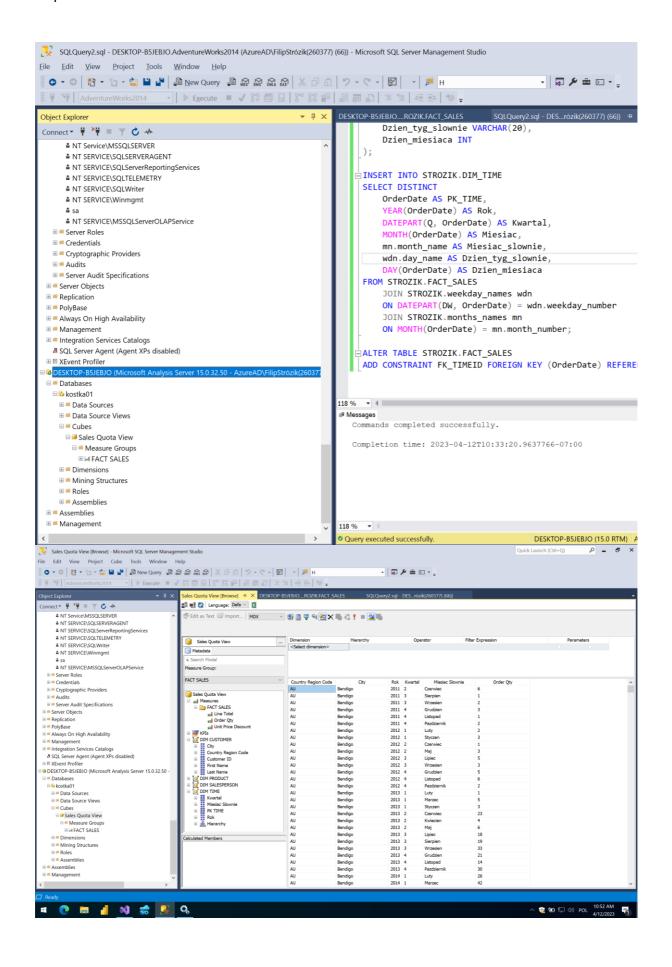


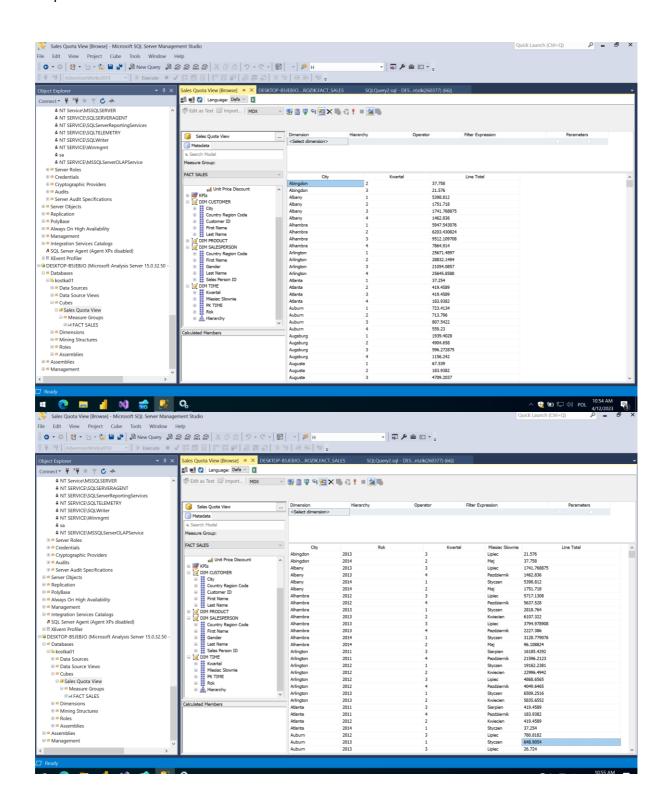
| Row Labels | Order Qty |
|--------------------|-----------|
| =2011 | |
| ± 2 | 966 |
| ⊕3 | 5270 |
| ±4 | 6652 |
| =2012 | |
| ⊞1 | 8593 |
| ⊞ 2 | 21423 |
| ⊞3 | 23003 |
| ±4 | 15560 |
| ■2013 | |
| ±1 | 18055 |
| ⊞ 2 | 33320 |
| ⊞3 | 44704 |
| ±4 | 35697 |
| =2014 | |
| ±1 | 38331 |
| ⊞ 2 | 23326 |
| Grand Total | 274900 |



Można zauważyć, że ilość zamówień do trzeciego kwartału 2013 roku miała tendencję wzrostową. Dość normlanym jest też, że kwartał drugi oraz trzeci czyli sezony rowerowe są kluczowymi kwartałami dla sklepu co widać na wykresie. Rok 2014 i dostępne tylko dwa pierwsze kwartały rozwiązują przyczynę wcześniejszych wniosków, że w roku 2014 sklep radził sobie gorzej. Okazuje się, że nie mamy dostępnych informacji o reszcie kwartałów dla tego roku. Fakt ten stawia historie sklepu w innym świetle.







Wnioski:

Wykonanie listy oraz przejście krok po kroku stanów z instrukcji pozwala na zaznajomienie się z procesem tworzenia kostki. Prawidłowe wykonanie poprzedniej listy pozwoliło na dość łatwą konfigurację kostki oraz jej uruchomienie.

Istnieje jedna wątpliwość co do typu atrybutów OrderDate i ShipDate dla Sales.SalesOrderHeader i wszystkich tabel wytworzonych podczas listy poprzedniej i aktualnej, które taki atrybut o takim typie jak DATETIME posiadają. Mimo, że w instrukcji była mowa typie INT przedstawiającej w pewien sposób informacje o dacie. Dość intuicyjnym pomysłem byłoby przerobienie typu na zwykły typ dat Date. Zaletą takiego rozwiązania byłaby możliwość pobierania części daty za pomocą operatora DATEPART(). W przypadku używania typu INT trzeba byłoby inaczej wydzielać interesujące nas informacje o dacie.

Tworzenie kostki pozwala na przetworzenie i przechowywanie danych w sposób zoptymalizowany pod kątem analizy. Kostki danych pozwalają na szybkie i skuteczne generowanie raportów i analiz biznesowych z dużych zbiorów danych. Warto zaznaczyć, że kostki są szczególnie przydatne dla dużych zbiorów danych, gdzie tradycyjne metody analizy, takie jak filtrowanie czy sortowanie, moga być bardzo czasochłonne.

Zdecydowano się wykonać część związaną z raportami przy pomocy Excel na samej maszynie wirtualnej. Było to spowodowane brakiem połączenia z lokalnego komputera do instancji serwisu. Być może było to spowodowane niepełną konfiguracją samego serwisu Analisys Services.

Same wrażenia z wykonywania raportów w Excel za pomocą połączenia się z usługą Analisys Services były lepsze niż wykonywanie raportów w Excel i Power BI przez zwykłe połączenie do bazy danych, ponieważ nie trzeba było pracować na wcześniej przygotowanych zapytaniach, połączenie do kostki pozwalało na większą elastyczność, a wcześniej zdefiniowane wymiary pozwalają lepiej dopasować dane do potrzeb raportu i uzyskać bardziej szczegółowe wyniki, co w połączeniu z zaawansowanymi funkcjami Excela pozwoliło na jeszcze bardziej rozbudowane analizy i wizualizacje danych.

Było to pierwsze "zetknięcie" się z kostką, mimo dość długiego procesu konfiguracji to końcowe efekty wydajnościowe oraz jakość danych były zdecydowanie lepsze niż przy korzystaniu z tradycyjnych baz danych. Kostka pozwalała na szybsze i bardziej efektywne wykonywanie zapytań, a zdefiniowane wymiary i hierarchie ułatwiały analizowanie danych z różnych perspektyw. Dzięki temu proces podejmowania decyzji był bardziej precyzyjny i oparty na rzetelnych informacjach. Mimo długiego procesu konfiguracji, zdecydowanie warto było zainwestować w stworzenie kostki danych.

Uwaga!

 Sprawozdanie, bez wniosków podsumowujących aspekt zagadnień analizowanych na zajęciach laboratoryjnych i zawartych w sprawozdaniu, jest automatycznie oceniane negatywnie!