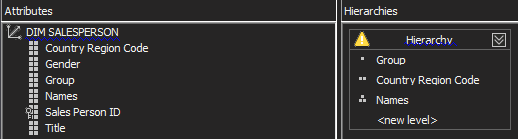
Hurtownie danych  
Laboratorium Czw 11:15  
  
Lista 6  
  
Kajetan Pynka 254495

**Zad 1**

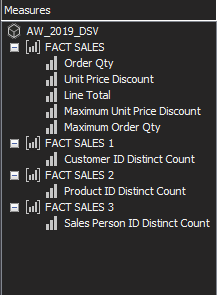
**a)**



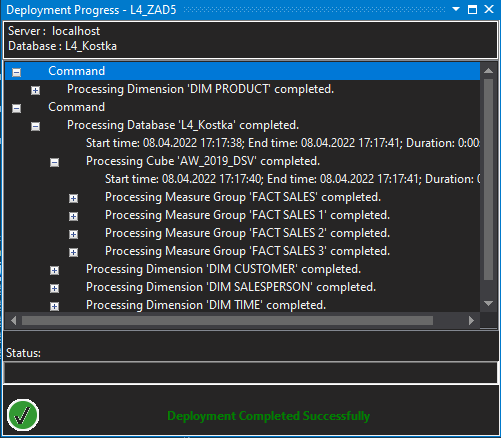
**b)**

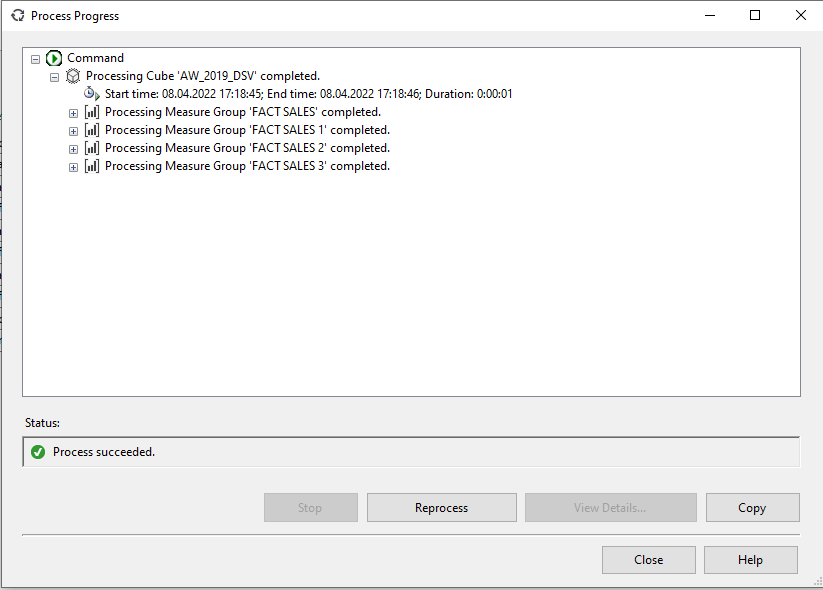


**c)**

****

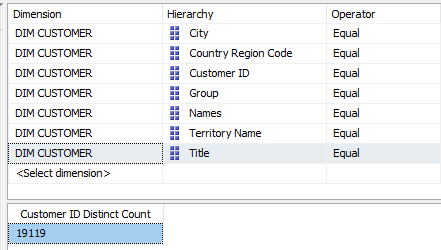
**d)**

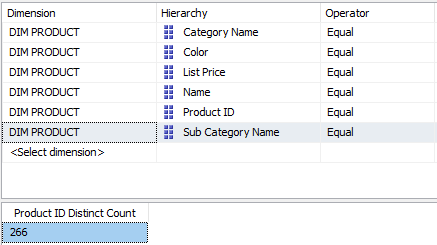


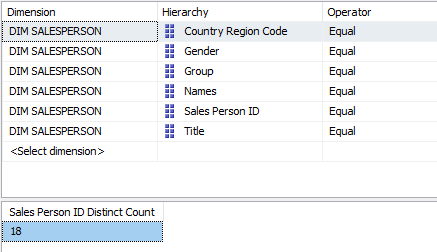


**Zad 2**

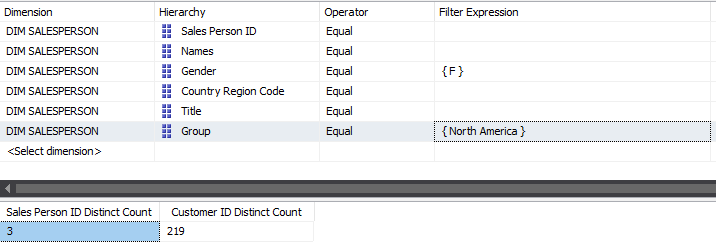
**a)**

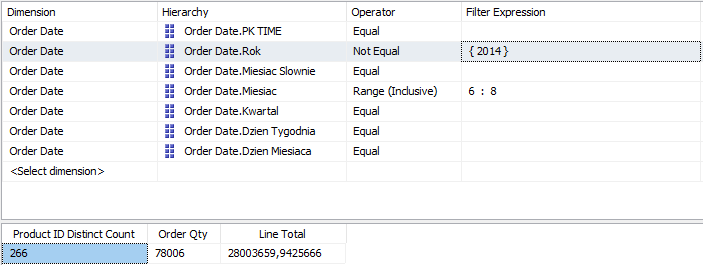
****

****

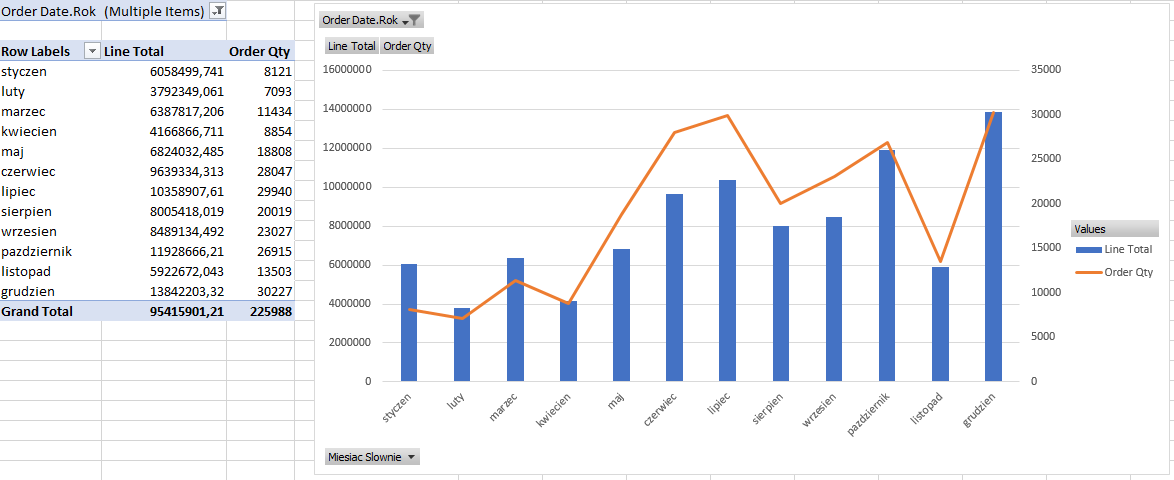
****

**b)**

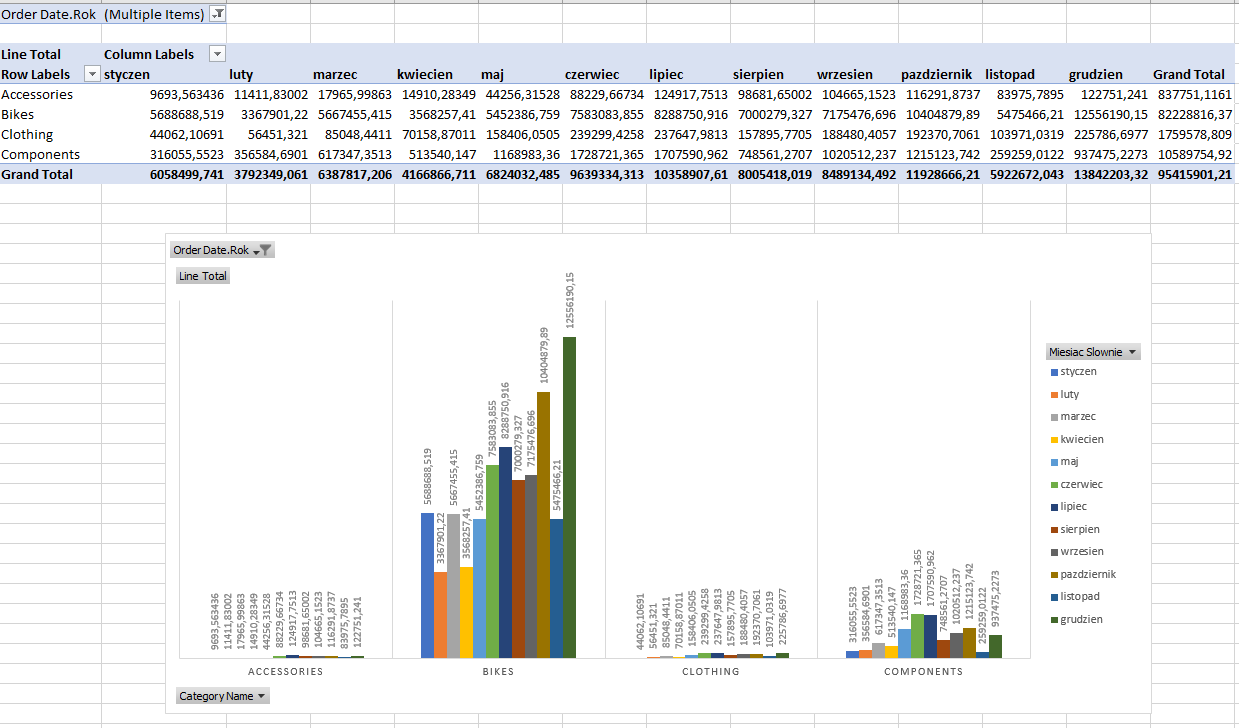
****

****

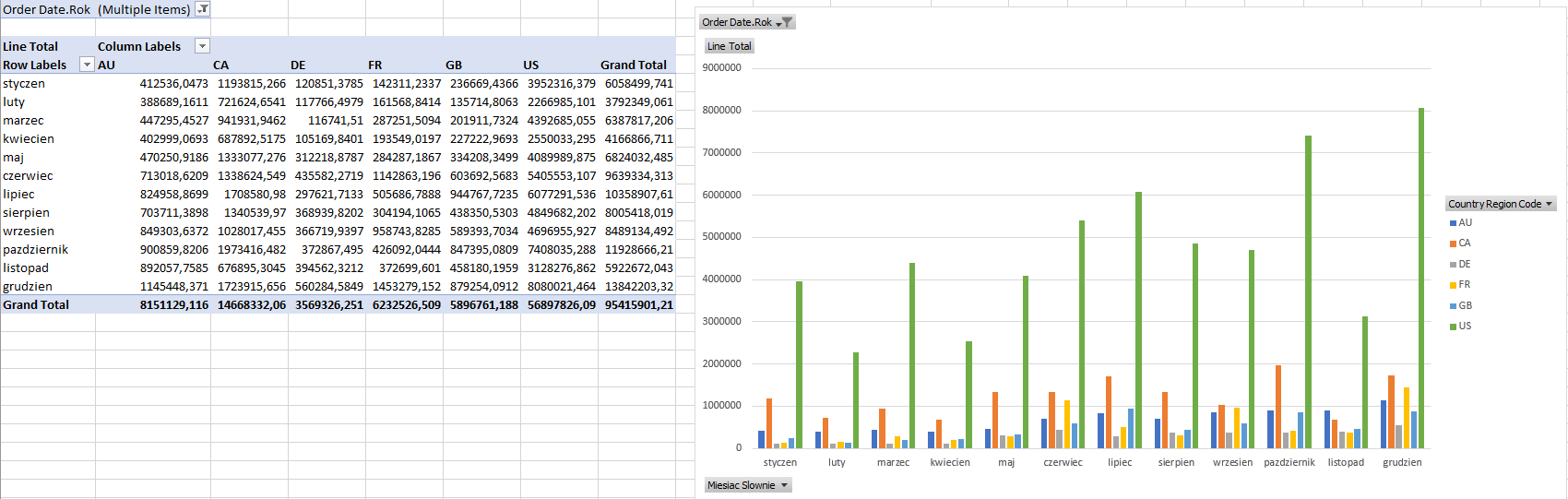
**c)**

****

**Wniosek:** Dane uwzględniają jedynie lata 2011-2013 (pominąłem rok 2014, ponieważ ostatnie transakcje kończą się na czerwcu przez co zaburza to trend dla drugiej połowy roku). Na podstawie zebranych danych możemy stwierdzić, że początek roku (od stycznia do kwietnia włącznie) jest najmniej przychodowy (w porównaniu z innymi miesiącami). Jedynie listopad wydaje się ponownie spadać do poziomu pierwszego kwartału.

****

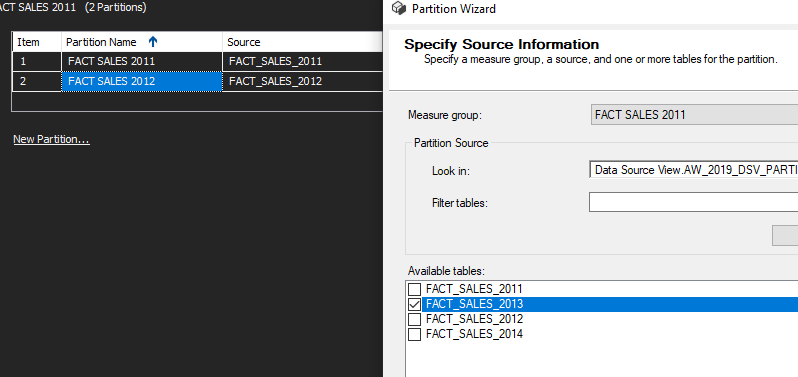
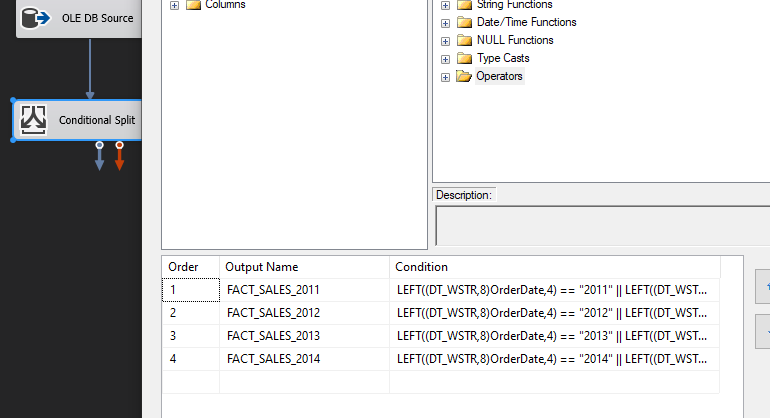
**Wniosek:** Analizując miesięczne przychody ze względu na kategorie możemy dostrzec, że dla każdej kategorii produktów ta sama tendencja jest zachowana (styczeń-kwiecień + listopad są najgorsze). Jeśli chodzi o najbardziej przychodową kategorię to widać różnice: dla rowerów jest to grudzień (prezenty świąteczne?), a przykładowo dla części czerwiec+lipiec (naprawy w sezonie rowerowym?).

****

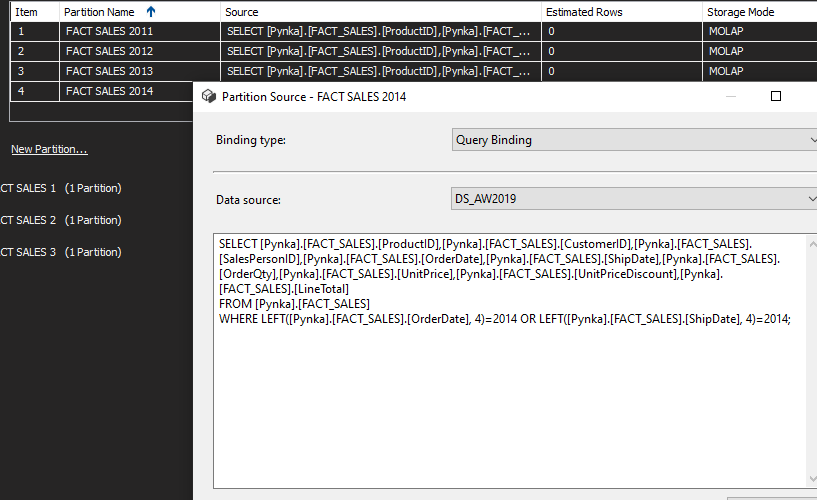
**Wniosek:** Analizując miesięczne przychody ze względu na kraje pochodzenia klientów możemy zauważyć silną korelację między słupkami Stanów Zjednoczonych a słupkami rowerów (z poprzedniego zrzutu ekranu). Oznacza to, że jednocześnie najbardziej dochodowe są rowery (ze wszystkich kategorii) a najwięcej pieniędzy wydają na nie ludzie ze Stanów Zjednoczonych. Dla pozostałych regionów ogólnie potwierdza się, że najgorszym jest okres styczeń-kwiecień.

**Zad 3**

**a)**

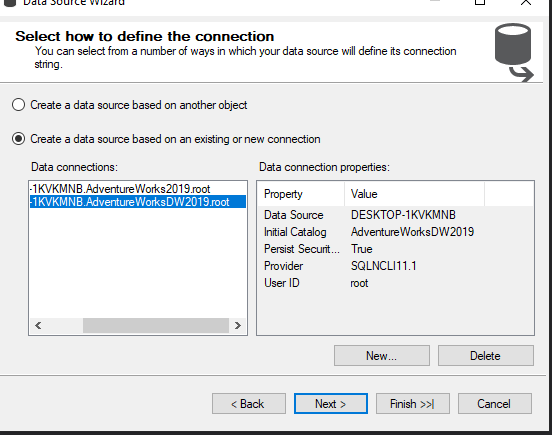
****

**b)**

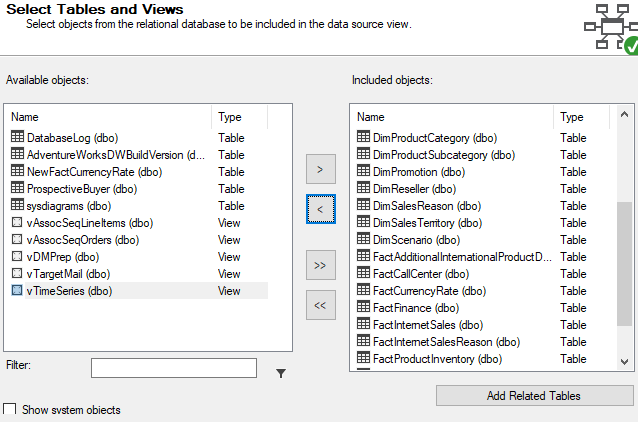
****

**Zad 4**

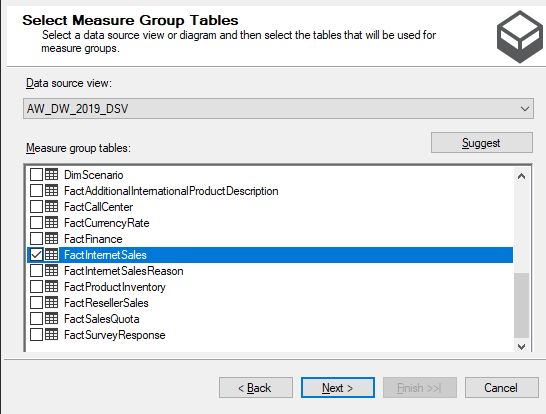
**a)**

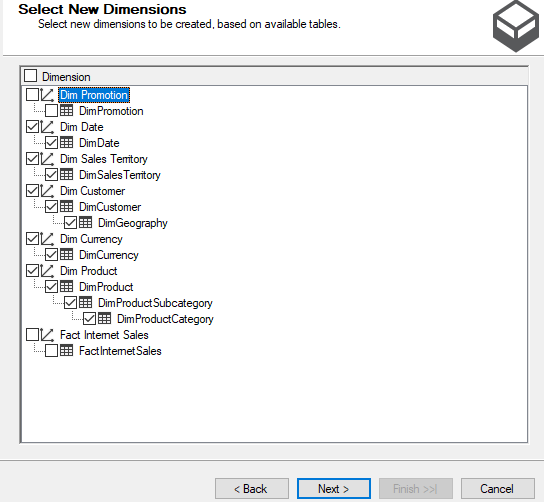
****

**b)**

****

**c)**

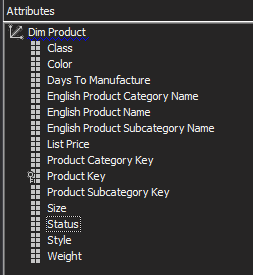
****

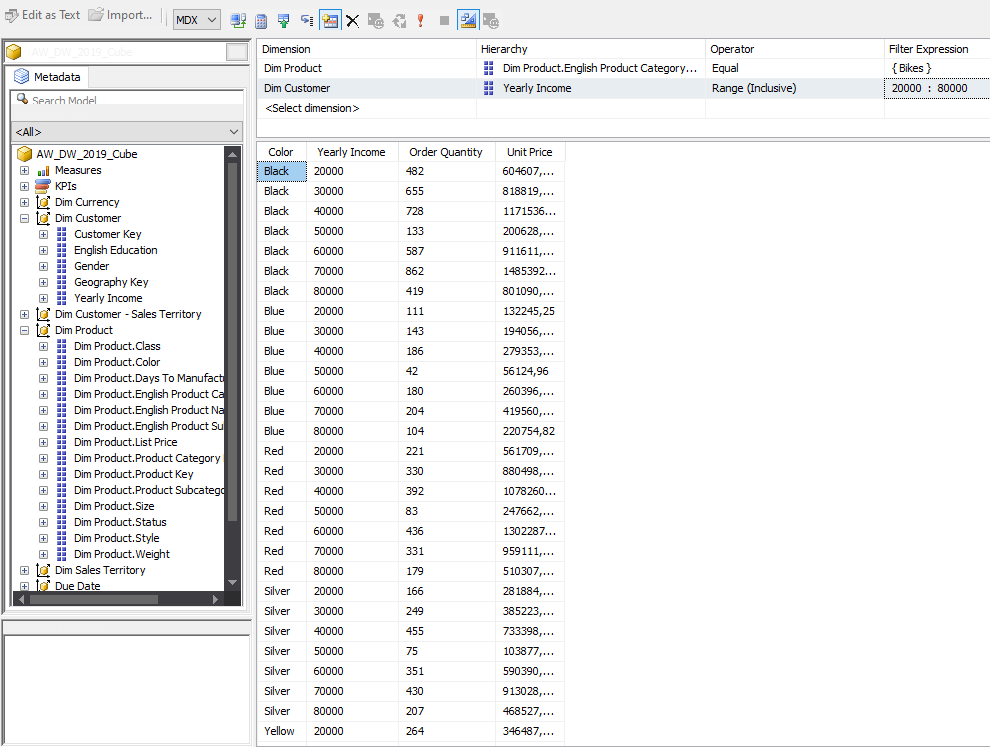
****

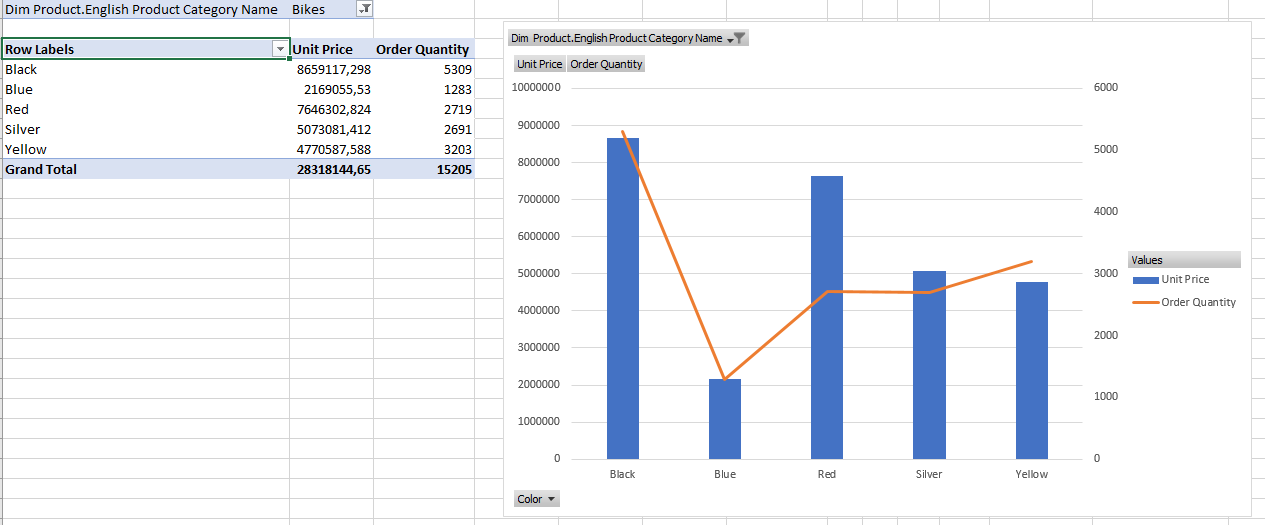
**d)**

Na moim sprzęcie nie zaobserwowałem znaczącej różnicy między mniejszą a większą liczbą wybranych wymiarów (tzn. nawet wybierając wszystkie wymiary kostka procesuje się w 1s lub mniej). Zakładam, że posiadając dostatecznie dużo danych i wymiarów zauważyłbym wzrost czasu w przypadku zwiększenia liczby wymiarów.

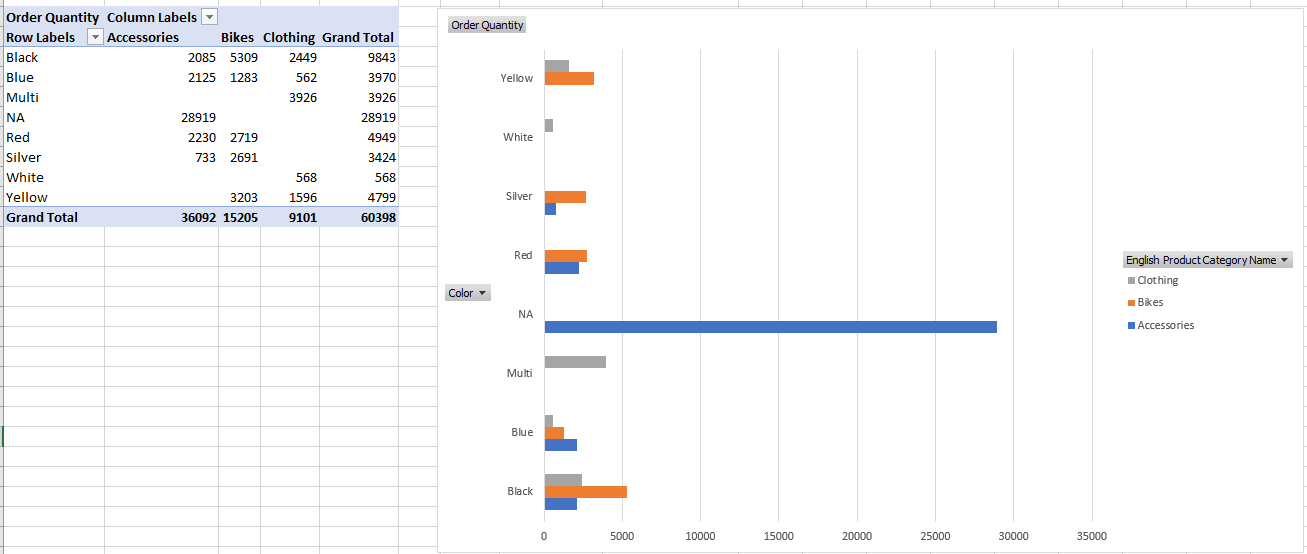
**e)**

****

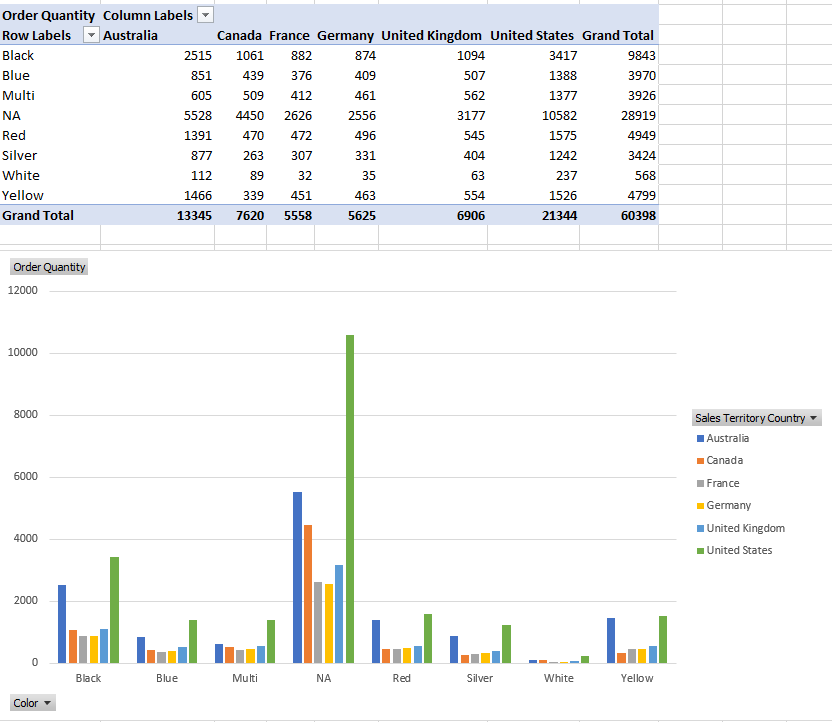
****

****

**Wniosek:** Najdroższe są rowery w kolorach czarnym oraz czerwonym, natomiast zdecydowanie najwięcej klienci zamawiali czarnych rowerów.

****

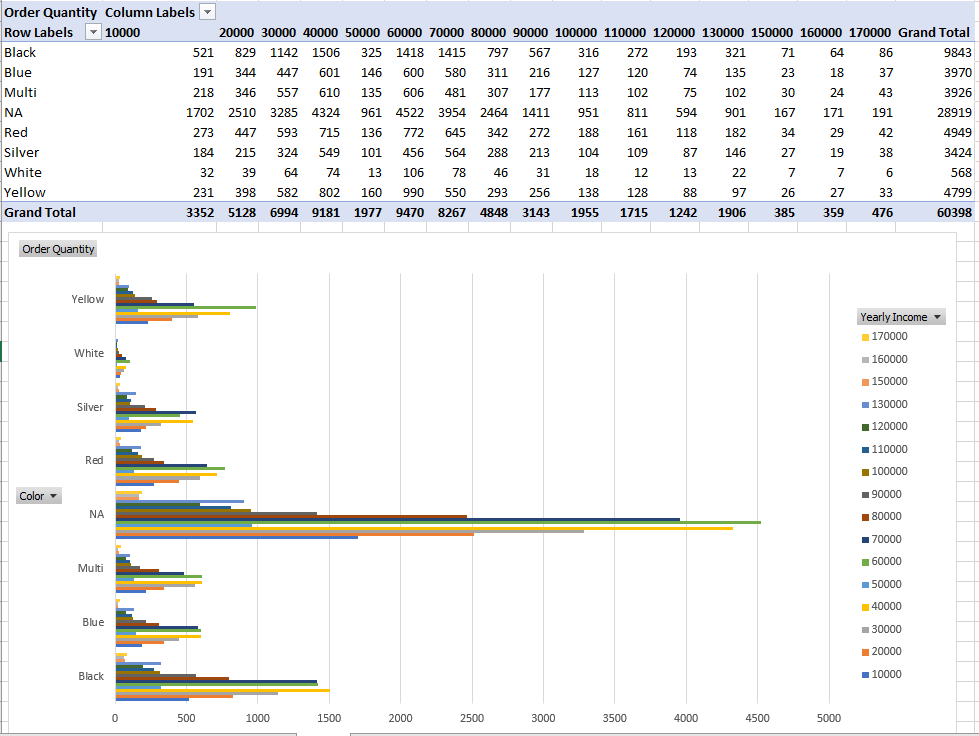
**Wniosek:** Patrząc ogólniej, dokonuję analizy liczby zamówionych przez klientów sztuk produktów ze względu na kolor oraz kategorię produktu. Okazuje się, że czarny kolor przeważa jedynie w przypadku rowerów, dla ubrań najpopularniejszy jest wariant „Multi”, natomiast dla akcesoriów zdecydowanie przeważają produkty bez podanego koloru (a wśród tych z określonym kolorem czarny jest tuż za czerwonym i niebieskim).

****

**Wniosek:** Niezależnie od kraju pochodzenia klienta, nadal przeważający jest kolor czarny (jeśli pominiemy masę zakupionych akcesoriów bez określonego koloru).

****

**Wniosek:** Niezależnie od płci klienta nadal dominującym kolorem ze względu na liczbę sprzedanych sztuk jest czarny.

****

**Wniosek:** Najwięcej sztuk produktów zamawiają klienci z rocznymi przychodami 40000 lub 60000 (zakładam ich lokalnej waluty). Niezależnie jednak od rocznych przychodów, kolor czarny wydaje się dominować w przypadku liczby kupowanych sztuk.

Ogólnie można stwierdzić, że ze względu na dane demograficzne jednoznacznie najczęściej kupowane są produkty w kolorze czarnym (niezależnie od płci, przychodów czy kraju pochodzenia klienta) lub bez określonego koloru, jeżeli bierzemy je pod uwagę (całość takowych stanowią akcesoria). Patrząc na kategorie produktów z asortymentu sklepowego to kolor czarny dominuje jedynie dla rowerów.

**Wnioski:**

* Jako że nie zawsze potrzebujemy wszystkich rekordów z tablicy faktów do analizy, to warto je podzielić na mniejsze „porcje” w celu zwiększenia wydajności procesu analizy danych. Możemy tego dokonać tworząc partycje, albo ze względu na już istniejący podział tabel faktów, albo niejako filtrując odpowiednie rekordy z jednej tablicy np. po roku.
* Narzędzia takie jak SSIS i SSAS w ramach Visual Studio pozwalają w bardzo przejrzysty i szybki sposób przygotować dane do analizy oraz dopasować wymiary / miary pod konkretne zagadnienia, które pragniemy przeanalizować.