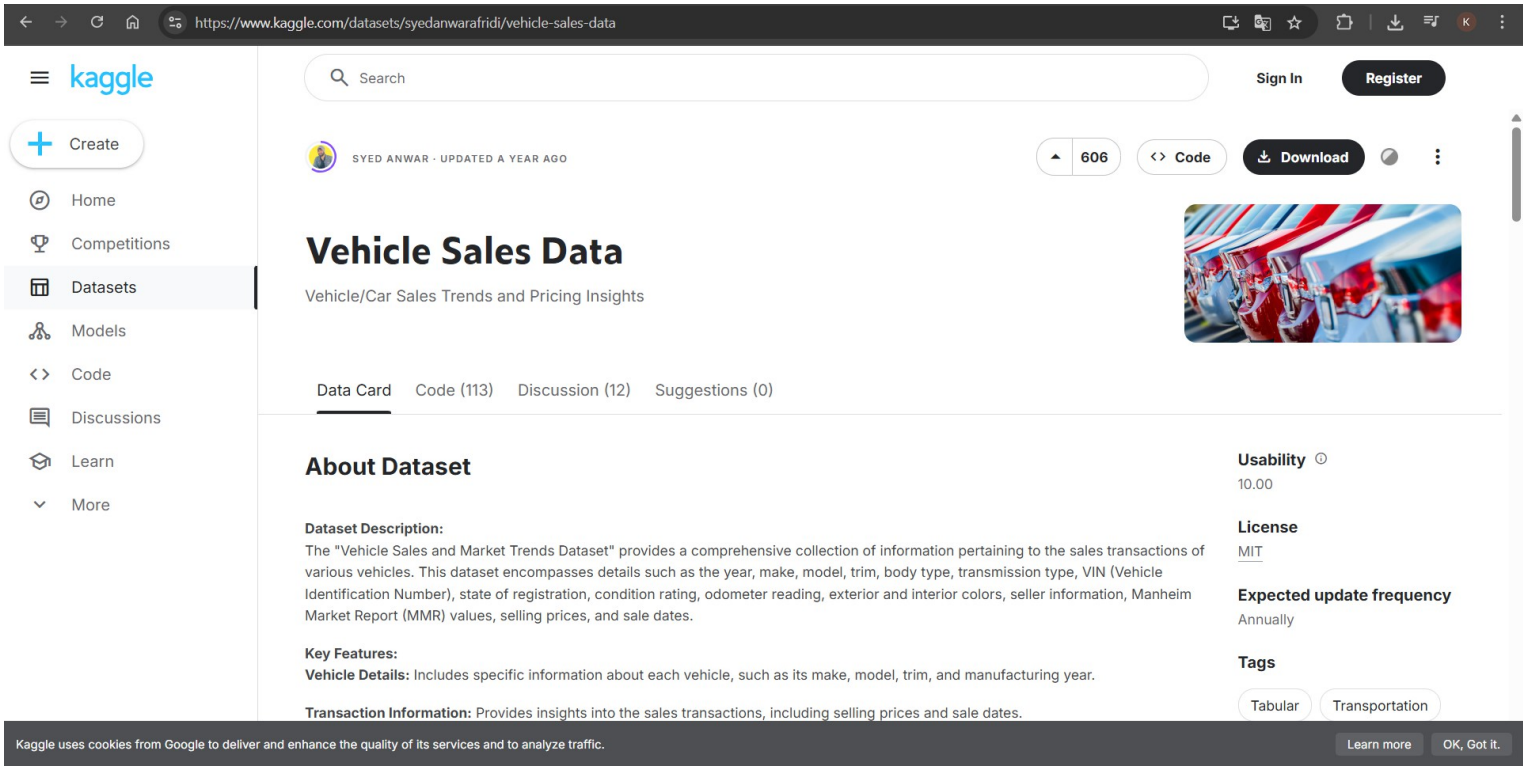
Temat zadania:
Analiza sprzedaży produktów w sklepie internetowym.

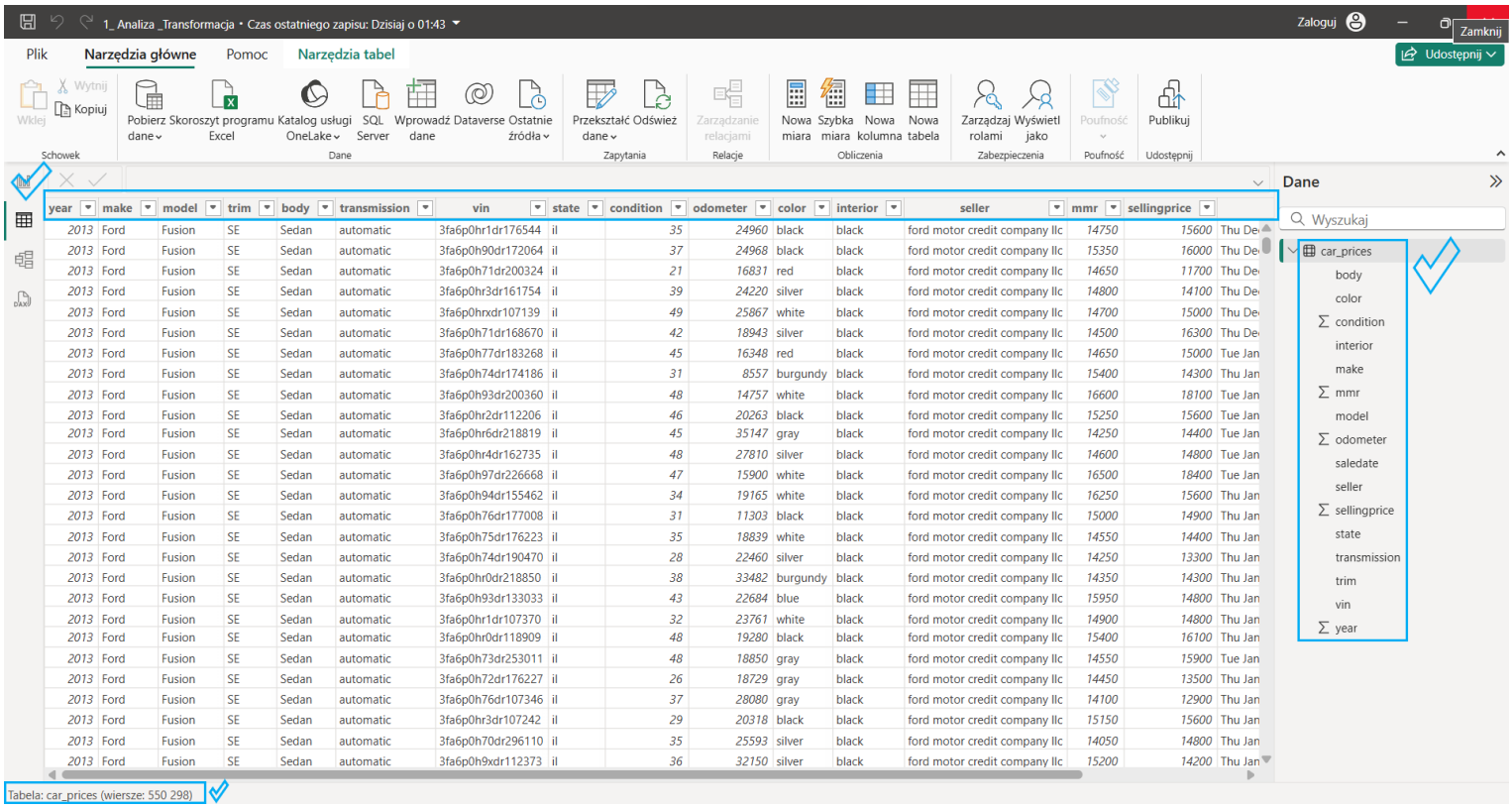
1. Ekstrakcja [2 pkt]:

- *Dane dotyczące sprzedaży są przechowywane w bazie danych sklepu internetowego:*

- *Wyeksportuj dane sprzedaży do pliku* CSV lub txt [*znajdź na portalu* https://**www.kaggle.com**/ *interesujący zbiór danych oparty o pliki* CSV]



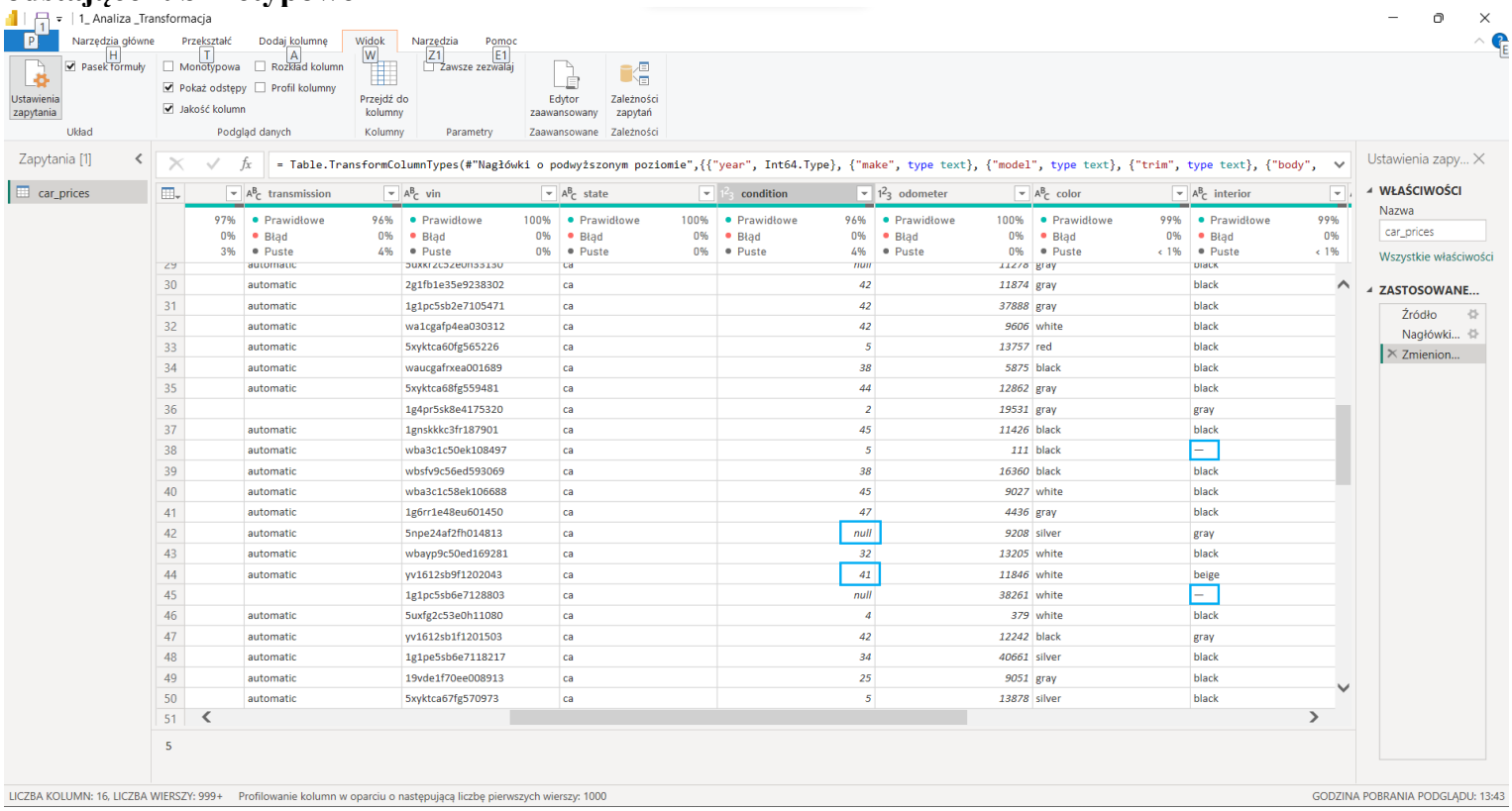
- *Zaprezentuj surowe dane wejściowe* (**strukturę pliku, liczbę wierszy, nazwy i typy kolumn**). *Wykonaj szybki przegląd* (**Preview, Column Quality**), *wskaż kolumny zawierające* **duplikaty**, **wartości odstające lub nietypowe**



Liczba wierszy: W dolnym lewym rogu widać: Tabela: car_prices	**Nazwy kolumn:** Widoczne w nagłówkach tabeli: year, make, model,

(wiersze: 550 298) – pokazuje liczbę rekordów.	trim, itd.
Typy danych:W panelu po prawej stronie (Dane) obok niektórych kolumn są symbole:	
∑ – **liczby** (np. condition, odometer, mmr, sellingprice)	**tekstowe** (np. make, model, state, vin, seller, interior)

Wykonaj szybki przegląd (Preview, Column Quality), **wskaż kolumny zawierające duplikaty, wartości odstające lub nietypowe**



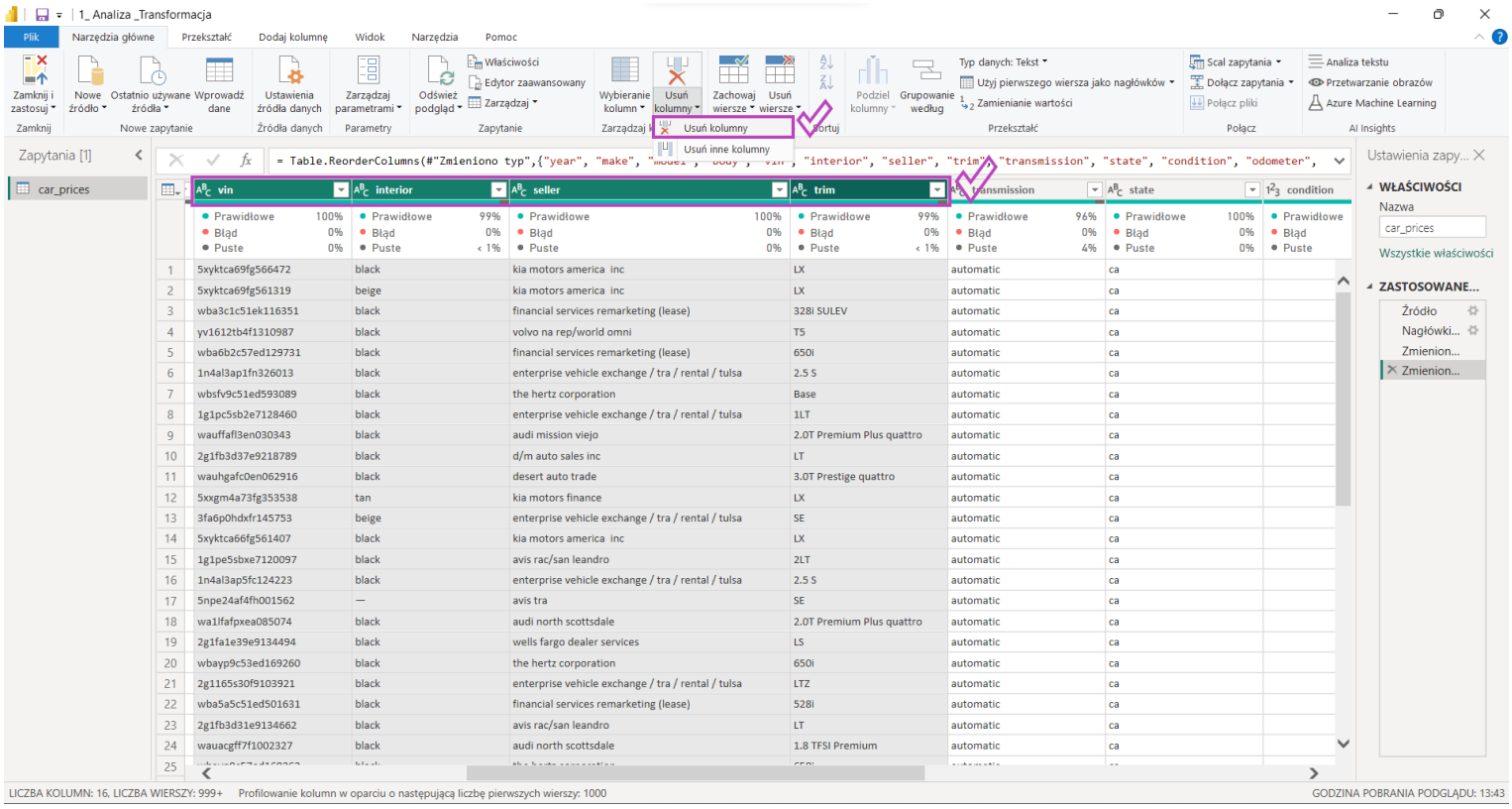
vin – unikalny identyfikator pojazdu – powinien być unikalny, ale zawiera duplikaty.	**seller** – wiele powtarzających się wartości, co jest oczekiwane (np. „ford motor credit company llc" pojawia się wielokrotnie)
condition i odometer – zawierają pojedyncze null	**mmr** – zawiera braki danych
Odometer: Wartości typu 10, 200000, itp. mogą być odstające względem średniej	**Condition**: Jeśli zakres np. od 1 do 50, a pojawi się 0 albo 100, będą to wartości odstające
Sellingprice: Zbyt niska cena np. 1, 100, może oznaczać błąd (np. nieprawidłowo wprowadzony rekord).	**Year**: Pojazdy z np. roku 1900 lub 2050 to wartości nietypowe – w tym zbiorze powinno być raczej 2000–2017.

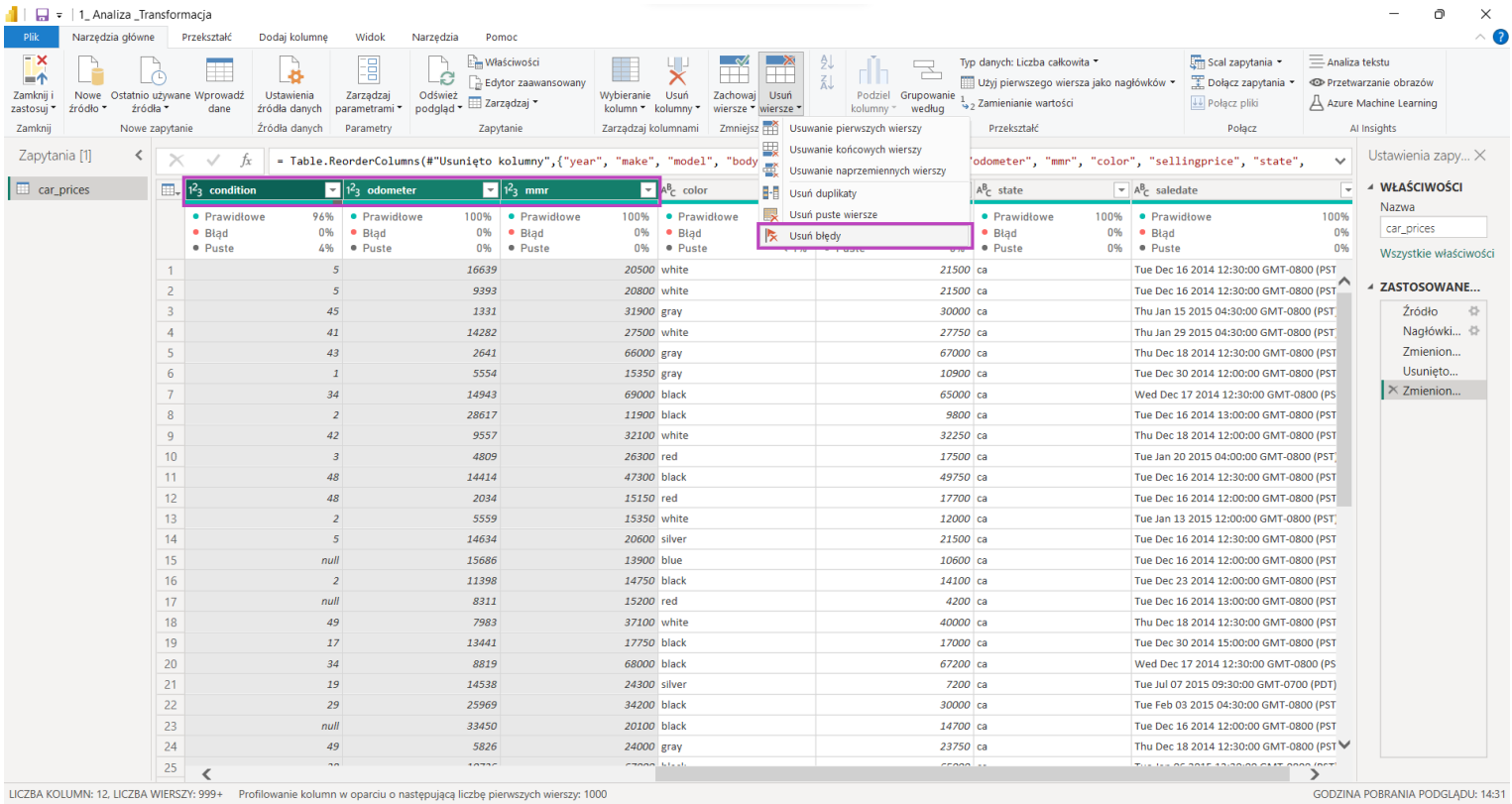
2. Transformacja [9 pkt]:

Wczytaj dane z pliku CSV do narzędzia ETL (np. Power Query w Power BI).
Przeprowadź następujące transformacje danych:

A). Usuń niepotrzebne kolumny i wiersze, Nulle i Errory [3 pkt].

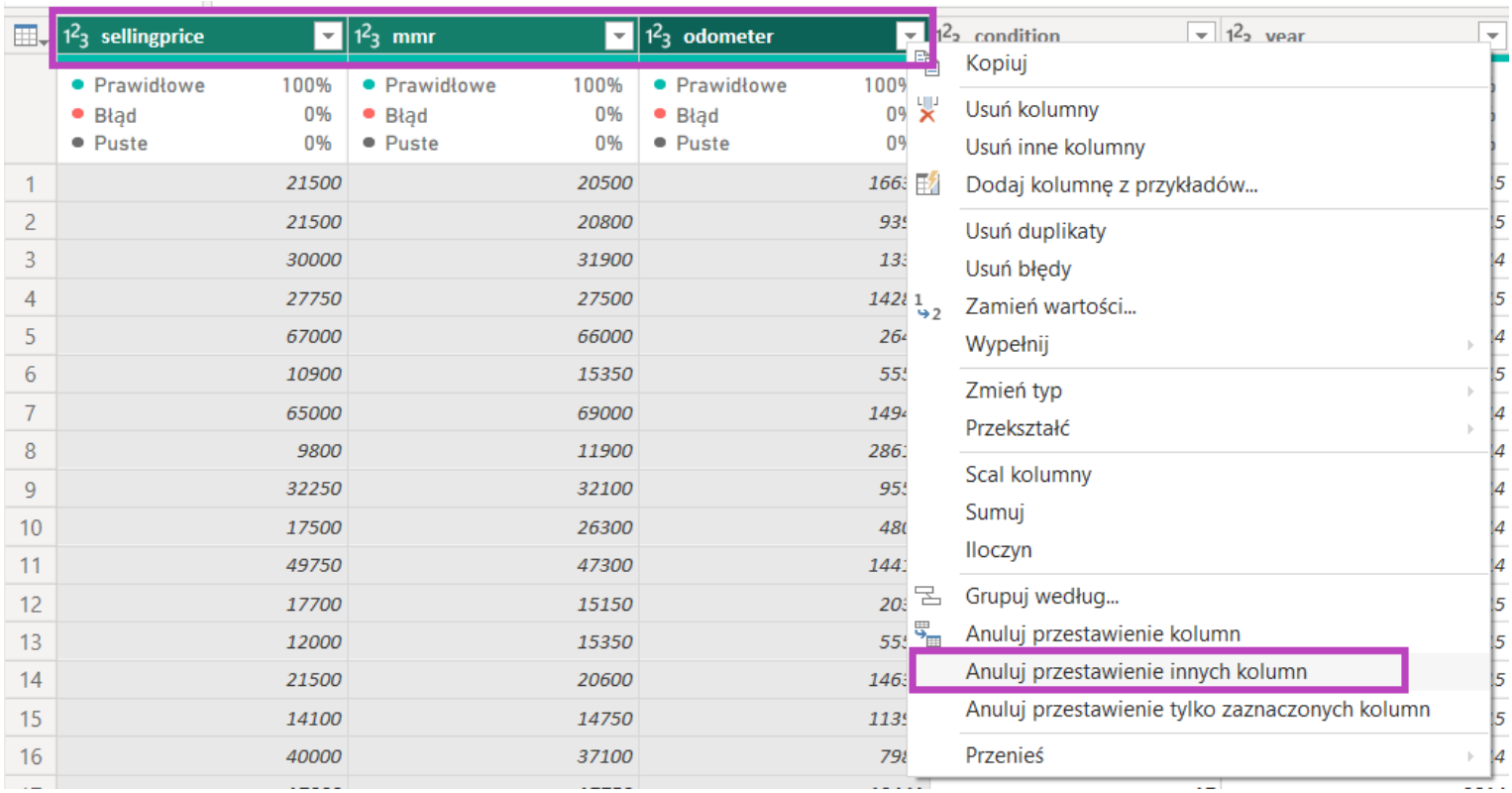
Kolumna	**Powód usunięcia**
vin	*Zawiera długie, unikalne identyfikatory – trudne do analizy.*
interior	*Często pokrywa się z kolorem zewnętrznym.*
seller	*Duplikujące się wartości, nie wnosi analitycznej wartości.*
trim	*Szczegół techniczny mało użyteczny dla ogólnej analizy.*

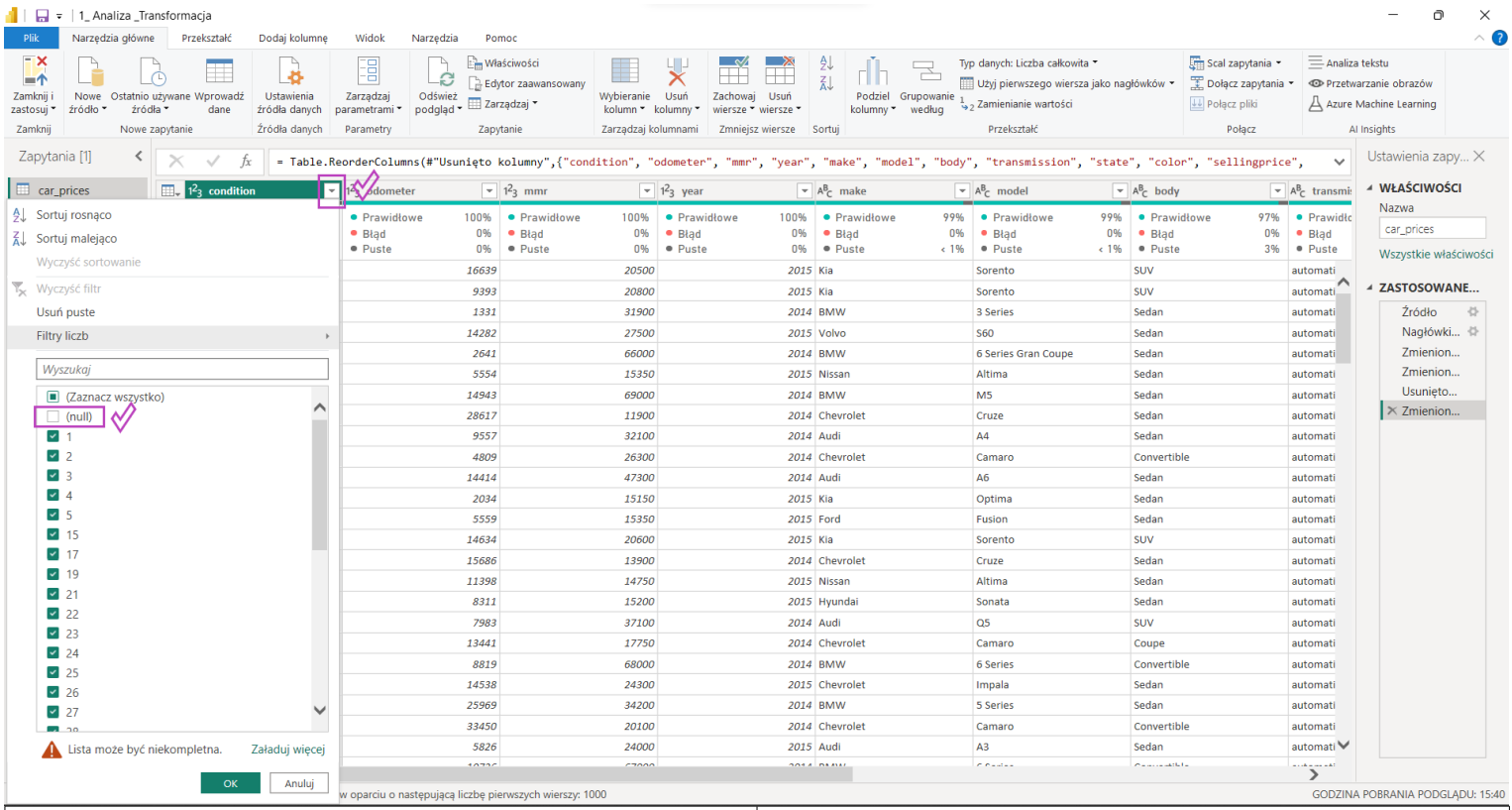




Usunięcie wierszy z pustymi wartościami (null)	condition, odometer, mmr

B). Wykonaj operacje : pivot lub unpivot [1 pkt]



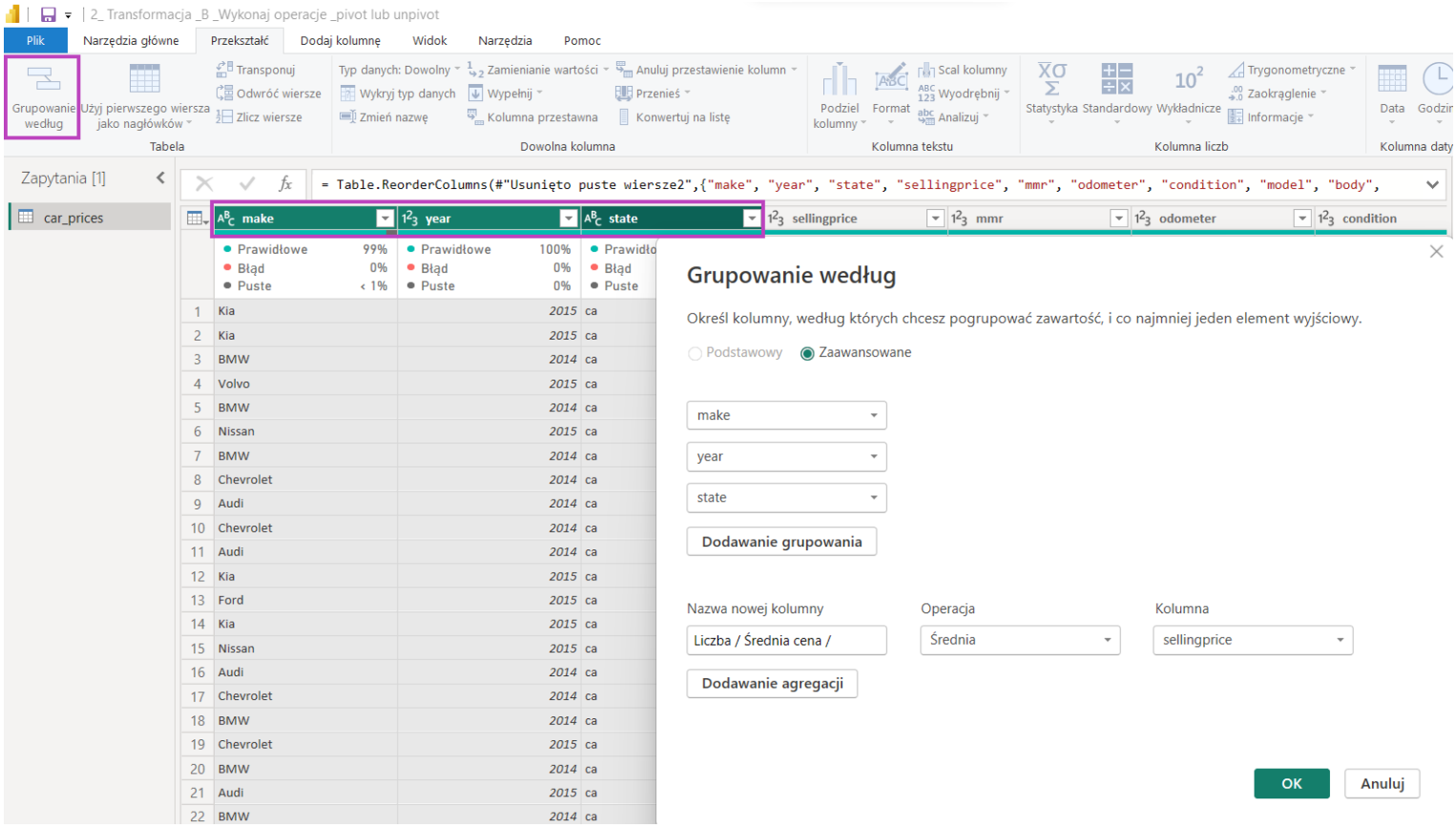


Struktura unpivot Dane są „rozciągnięte w dół” — czyli typowy rezultat Unpivot.	Każdy wiersz reprezentuje pojedynczą właściwość jednego samochodu.
Lewa kolumna: **Atrybut** – zawiera nazwy dawnych kolumn (**condition, year, make, model, body,**	Prawa kolumna: Wartość – zawiera wartości z tych kolumn (np. 5, 2015, Kia, Sorento, itd.).

transmission, state, color, saledate).	Każda para Atrybut – Wartość tworzy nowy wiersz.

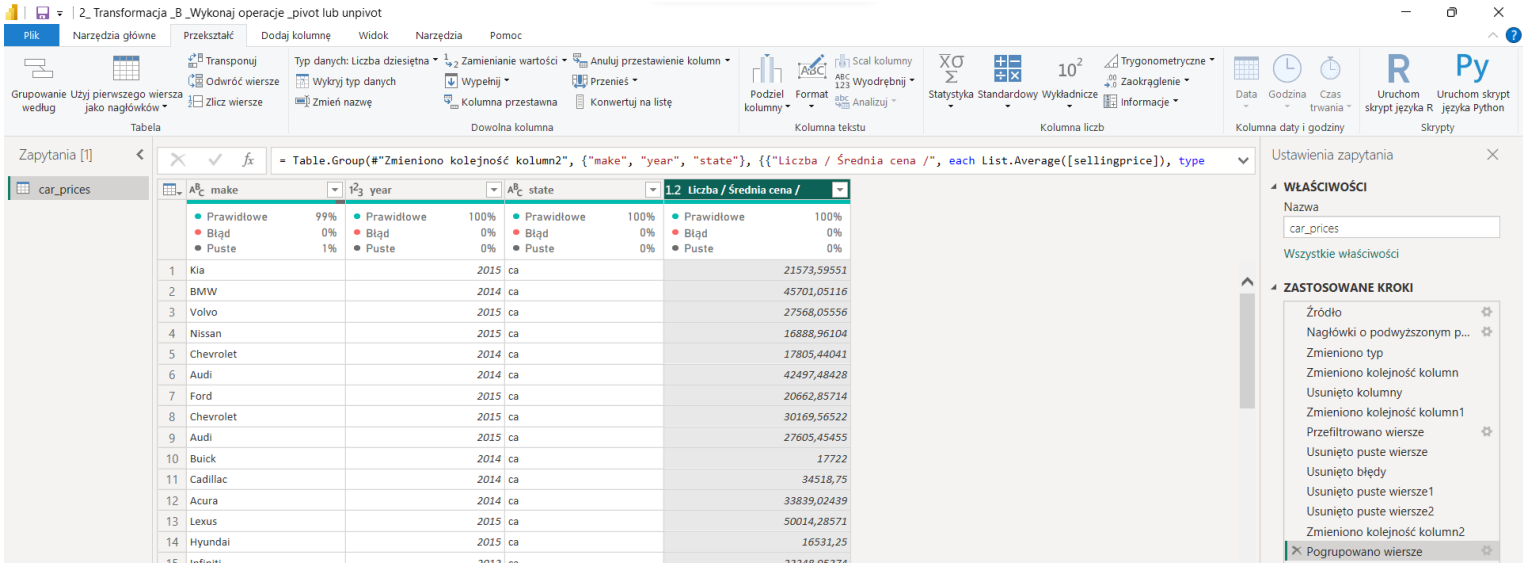
Atrybut	Wartość
Prawidłowe 100%	Prawidłowe 100%
Błąd 0%	Błąd 0%
Puste 0%	Puste 0%
condition	5
year	2015
make	Kia
model	Sorento
body	SUV
transmission	automatic
state	ca
color	white
saledate	Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
condition	5
year	2015
make	Kia
model	Sorento
body	SUV
transmission	automatic
state	ca
color	white
saledate	Tue Dec 16 2014 12:30:00 GMT-0800 (PST)
condition	45
year	2014

C). Wykonaj operacje : grupowanie [1 pkt]



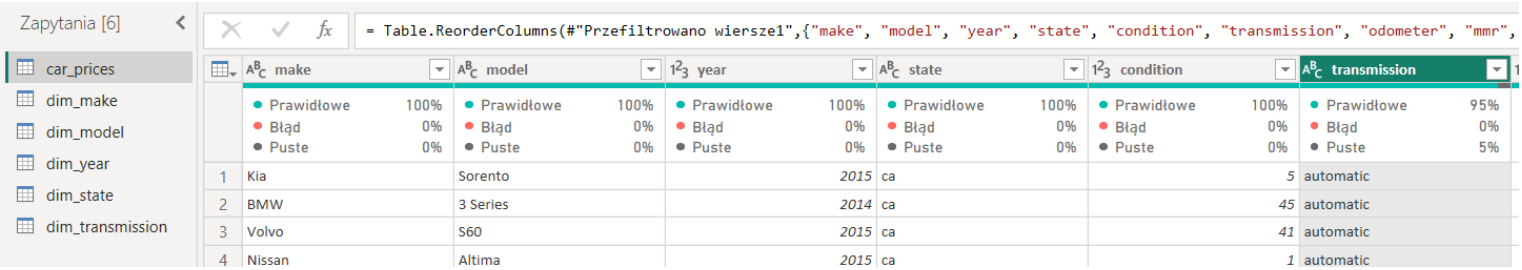
Nazwa kolumny grupującej (Grupuj według)	make (marka samochodu)
Nowa kolumna	Liczba / Średnia cena

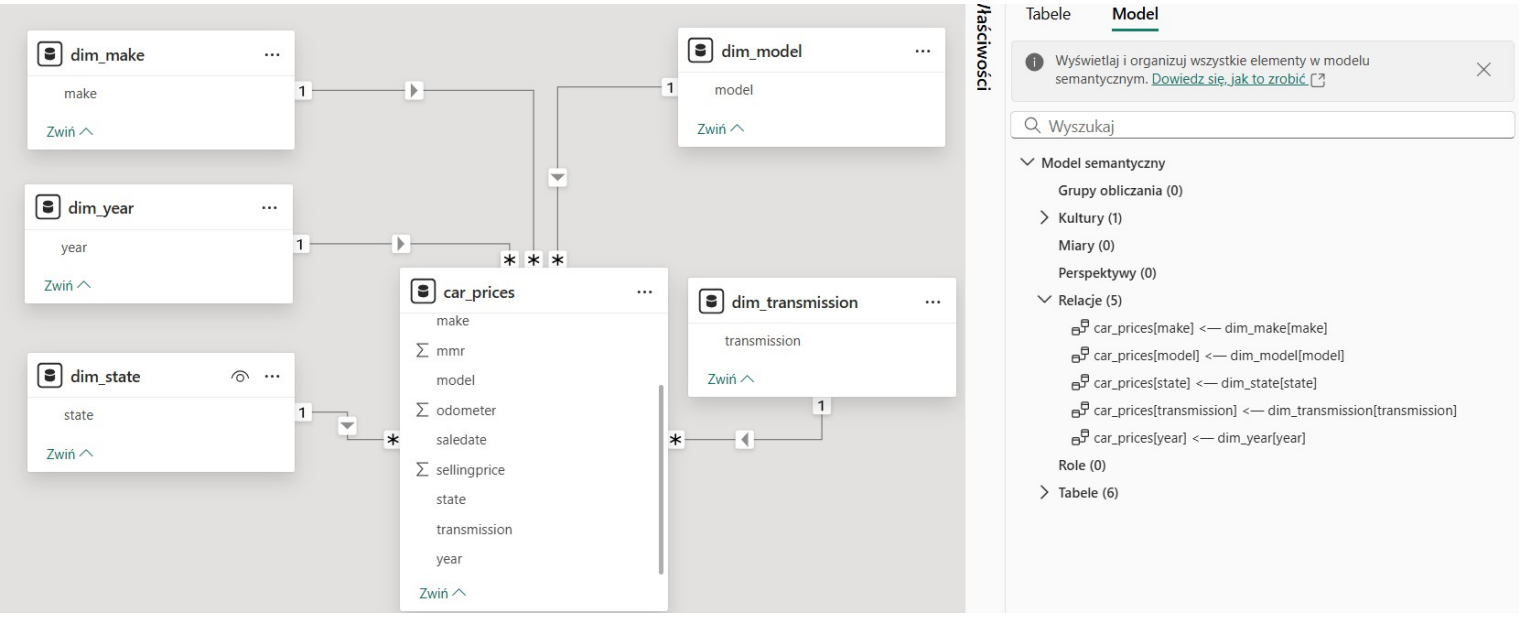
Operacja	Średnia (Average)
Kolumna:	sellingprice



D). Koniecznie stwórz tabelę danych/faktów i minimum 5 tabel wymiarów/opisowe/słownikowe

prawy przycisk myszy na	car_prices
Wybierz	Odwołanie
W nowej tabeli	car_prices (2) → dim_make





Klucze główne (PK) i obce (FK) w Twoim modelu:

1. Tabela faktów: car_prices

To tabela faktów i **zawiera klucze obce** do tabel wymiarów.

Klucze obce w car_prices:

- make → klucz obcy do dim_make[make]
- model → klucz obcy do dim_model[model]
- year → klucz obcy do dim_year[year]
- state → klucz obcy do dim_state[state]
- transmission → klucz obcy do dim_transmission[transmission]

Zidentyfikowane relacje JOIN w modelu

Wszystkie widoczne relacje mają:

- **symbol "1"** po jednej stronie (np. dim_make)
- oraz **"∗" (gwiazdkę)** po stronie drugiej (np. car_prices)

LEFT JOIN, RIGHT JOIN, FULL OUTER JOIN?

W Power BI **domyślnie** używany jest **INNER JOIN** przez relacje 1:N. Jeśli chcesz:

- **LEFT JOIN** – musisz użyć np. funkcji DAX RELATED() (czyli pobranie danych z tabeli podrzędnej nawet jeśli nie ma dopasowania – np. wyświetlanie wszystkich z car_prices, nawet jeśli brak powiązania w dim_model).
- **RIGHT JOIN / FULL OUTER JOIN** – Power BI **nie wspiera tego bezpośrednio**. Musisz użyć:
  - MERGE w Power Query z opcją Full Outer
  - lub skonstruować złożony DAX (UNION, EXCEPT, INTERSECT)

Z Twojego modelu:

- **wszystkie relacje to INNER JOIN (1:N)** – wykluczają wiersze bez dopasowania.
- **brak LEFT, RIGHT, FULL OUTER JOIN** – ale możesz je zasymulować w Power Query lub DAX.

Tabele wymiarów (dimension tables):

Każda z tych tabel zawiera **klucz główny (PK)**, który jest jednoelementowym atrybutem, wykorzystywanym jako klucz obcy w car_prices.

Tabela wymiarów	Klucz główny	Powiązanie w car_prices (klucz obcy)
dim_make	make	car_prices[make]
dim_model	model	car_prices[model]
dim_year	year	car_prices[year]
dim_state	state	car_prices[state]
dim_transmission	transmission	car_prices[transmission]

Tabela car_prices ma 5 kluczy obcych: make, model, year, state, transmission
Każda tabela wymiarów (dim_*) zawiera klucz główny, który jest połączony z odpowiednim kluczem obcym w car_prices.

E). Oblicz wartości zamówienia [1 pkt].

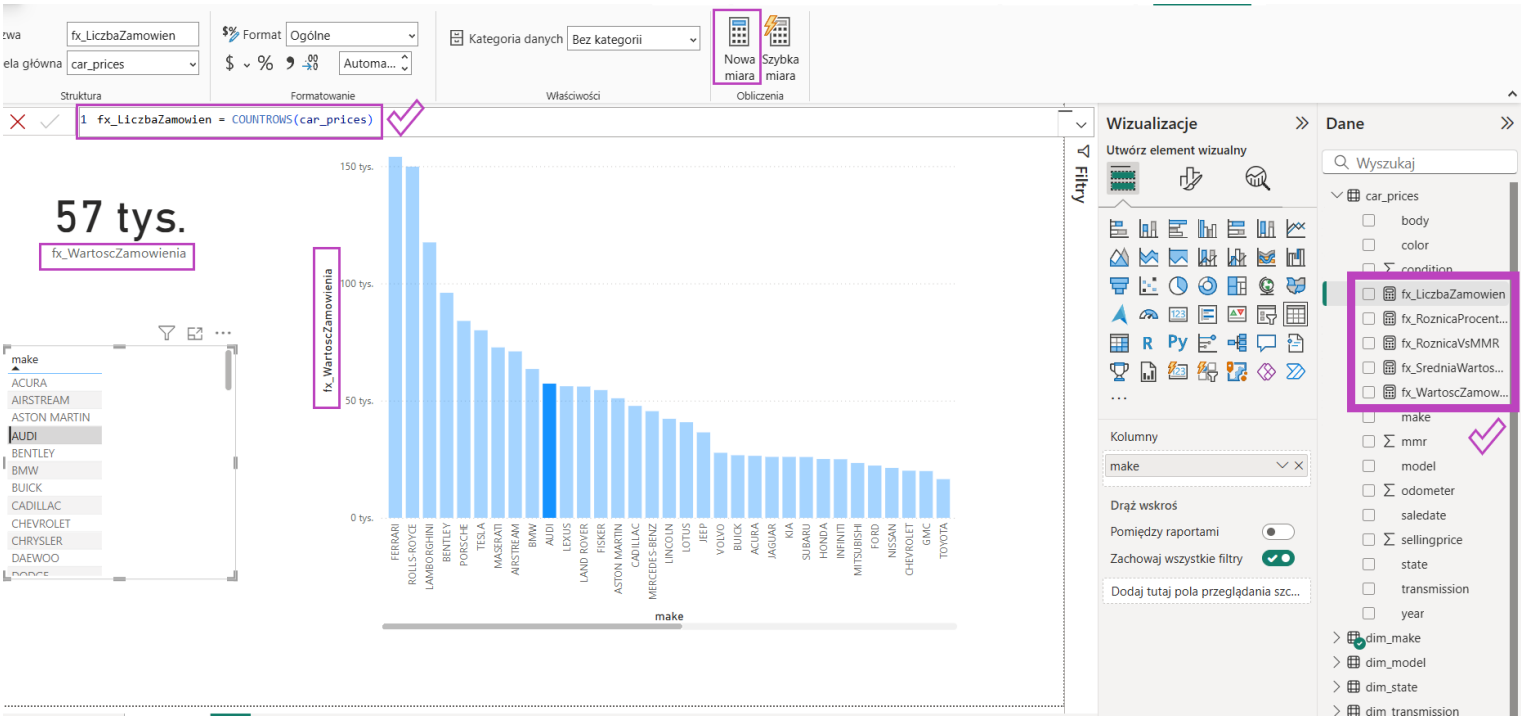
W zakładce 'Narzędzia główne' klik -> **'Nowa miara'**

fx_WartoscZamowienia = SUM(car_prices[sellingprice])	`1 fx_WartoscZamowienia = SUM(car_prices[sellingprice])`
fx_SredniaWartoscZamowienia =	`1 fx_SredniaWartoscZamowienia = AVERAGE(car_prices[sellingprice])`

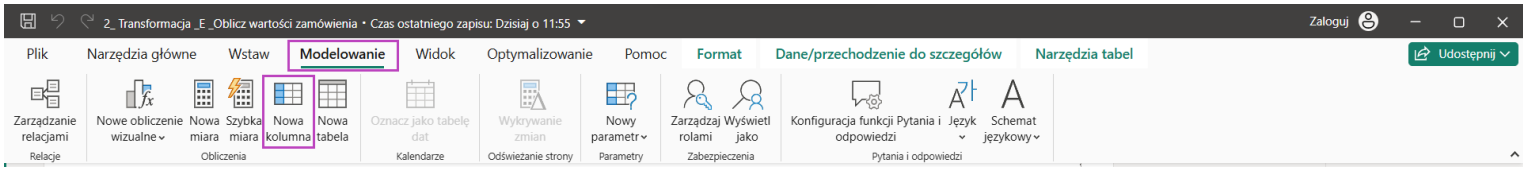
AVERAGE(car_prices[sellingprice])	
fx_RoznicaVsMMR = SUM(car_prices[sellingprice]) - SUM(car_prices[mmr])	`fx_RoznicaVsMMR = SUM(car_prices[sellingprice]) - SUM(car_prices[mmr])`

fx_RoznicaProcentowaVsMMR = DIVIDE(SUM(car_prices[sellingprice]) - SUM(car_prices[mmr]), SUM(car_prices[mmr]))	`fx_RoznicaProcentowaVsMMR = DIVIDE(`
	`  SUM(car_prices[sellingprice]) - SUM(car_prices[mmr]),`
	`  SUM(car_prices[mmr])`
	`)`

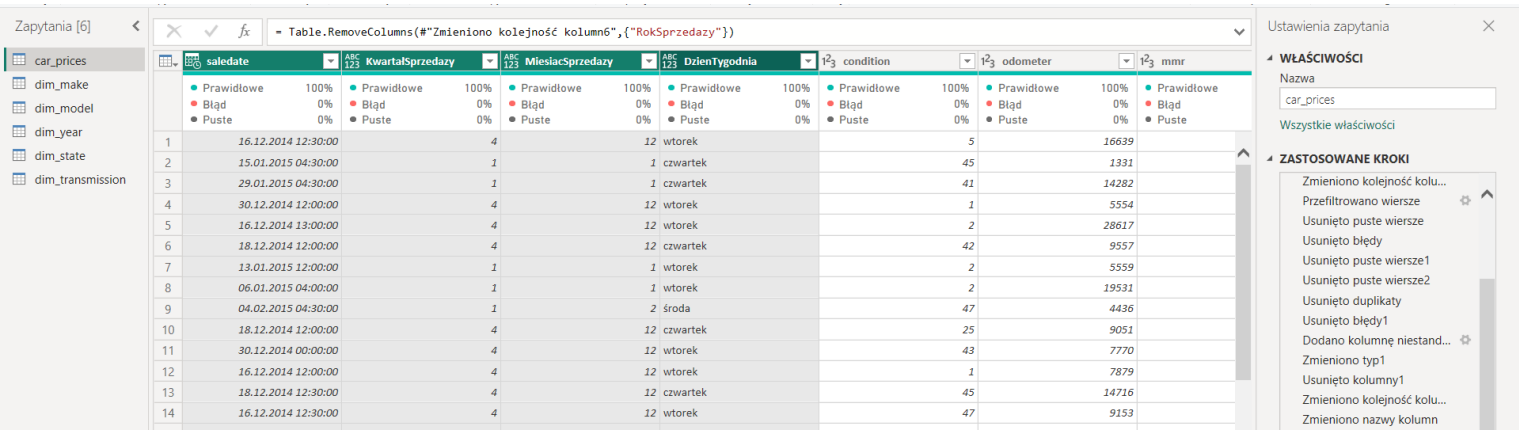
fx_LiczbaZamowien = COUNTROWS(car_prices)	`fx_LiczbaZamowien = COUNTROWS(car_prices)`
	fx_LiczbaZamowien fx_RoznicaProcent... fx_RoznicaVsMMR fx_SredniaWartos... fx_WartoscZamow...



F).Dodaj kolumny pomocnicze, takie jak rok, miesiąc, dzień tygodnia itp., które ułatwią analizę danych [1 pkt].

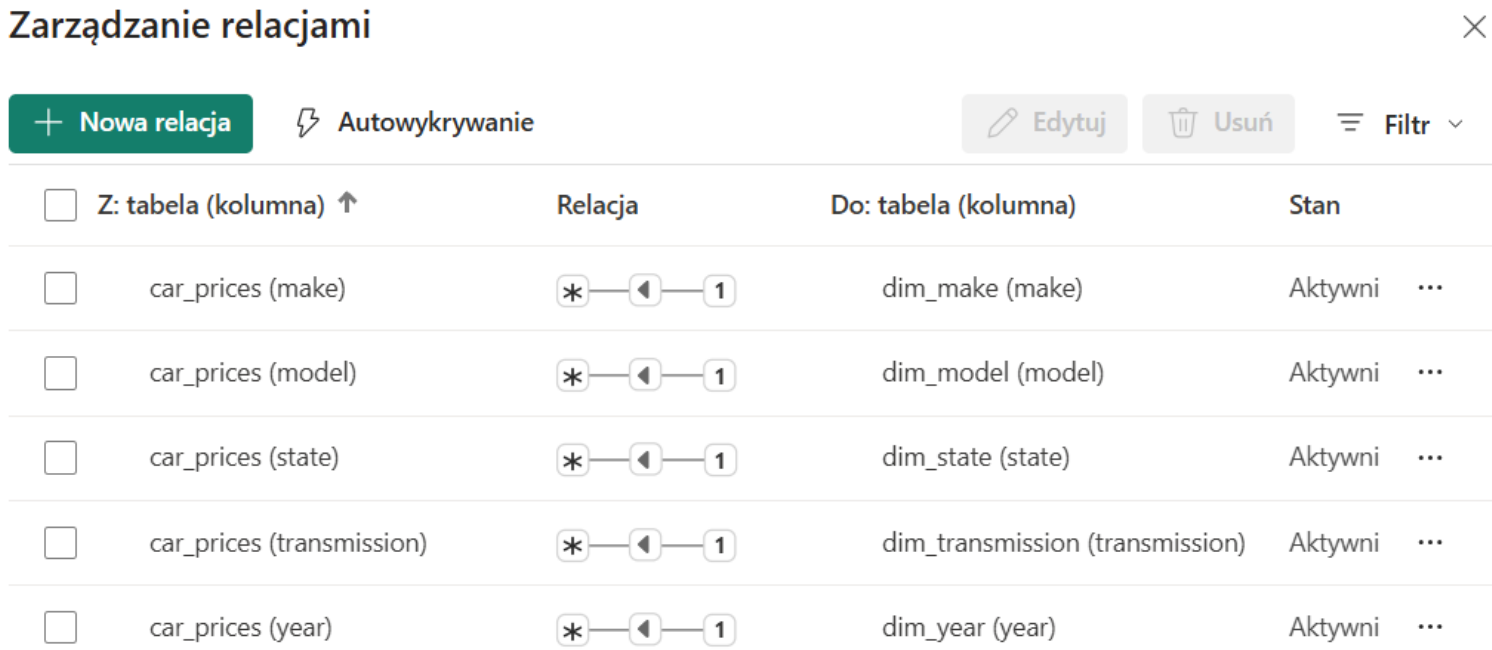


KwartałSprzedazy =	Date.QuarterOfYear([saledate])
MiesiacSprzedazy =	Date.Month([saledate])
DzienTygodnia =	Date.ToText(Date.From([saledate]), "dddd", "pl-PL")



3. Relacje i wizualizacja [10 pkt]

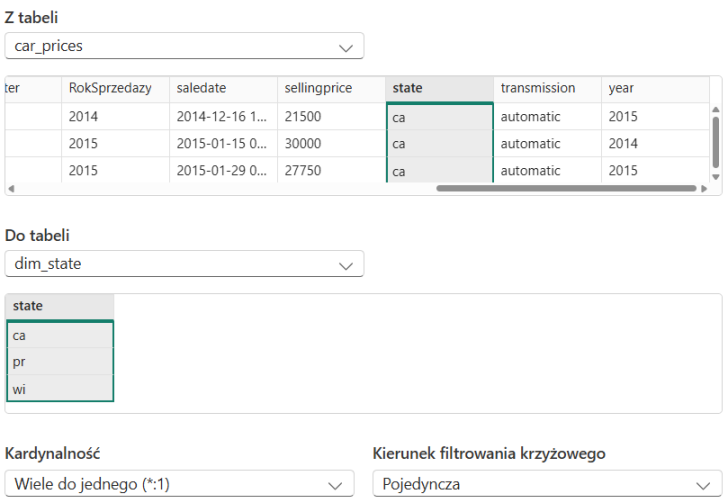
Załaduj przetworzone dane do raportu w Power BI.
Stwórz relacje między tabelami, aby uzyskać pełne informacje o sprzedanych produktach [4pkt]. Omów rodzaj połączenia i kierunek przepływu danych.



- Rodzaj relacji: Wiele do jednego (*:1)
- Źródło danych (tabela faktów): car_prices
- Cel danych (tabela wymiarów): np. dim_make

- W car_prices wiele rekordów (np. samochodów) może mieć tę samą markę (make), model (model), rok (year) itd.

- W tabeli wymiarów (dim_*) każda wartość występuje tylko raz (unikalna lista np. marek, modeli, lat itp.).



- Kierunek propagacji filtra: z tabeli wymiarów do tabeli faktów

- Oznacza to, że wybór/filtr w np. dim_model powoduje ograniczenie rekordów tylko w car_prices, ale nie odwrotnie.

- Taki kierunek jest typowy i zalecany w układzie modelu gwiazdy, ponieważ:

- Tabela wymiarów to źródło „opisu” (np. nazwy marek, modele),

- Tabela faktów to źródło „liczb” (np. sprzedaż, przebieg).

Z tabeli

car_prices

ter	RokSprzedazy	saledate	sellingprice	state	transmission	year
	2014	2014-12-16 1...	21500	ca	automatic	2015
	2015	2015-01-15 0...	30000	ca	automatic	2014
	2015	2015-01-29 0...	27750	ca	automatic	2015

Do tabeli

dim_transmission

transmission
automatic
manual

Kardynalność: Wiele do jednego (*:1)

Kierunek filtrowania krzyżowego: Pojedyncza

Z tabeli

car_prices

model	NazwaMiesiaca	odometer	RokSprzedazy	saledate	sellingprice	state
Sorento	grudzień	16639	2014	2014-12-16 1...	21500	ca
3 Series	styczeń	1331	2015	2015-01-15 0...	30000	ca
S60	styczeń	14282	2015	2015-01-29 0...	27750	ca

Do tabeli

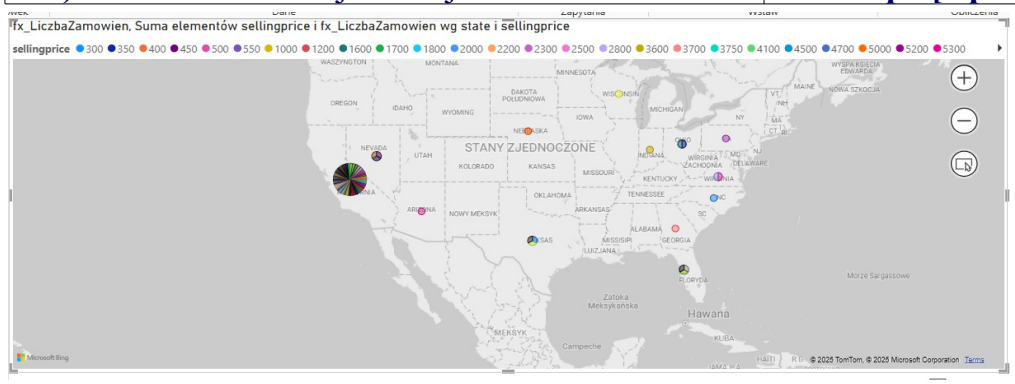
dim_model

model
SORENTO
3 SERIES
S60

Kardynalność: Wiele do jednego (*:1)

Kierunek filtrowania krzyżowego: Pojedyncza

3. A) Stwórz wizualizacje takie jak: **ArcGIS for Power BI/ mapa [1 pkt],**



Lokalizacja
state

Legenda
sellingprice

Szerokość geograficzna
Dodaj pola danych tutaj

Długość geograficzna
Dodaj pola danych tutaj

Rozmiar bańki
fx_LiczbaZamowien

Etykietki
Suma elementów sellingprice
fx_LiczbaZamowien

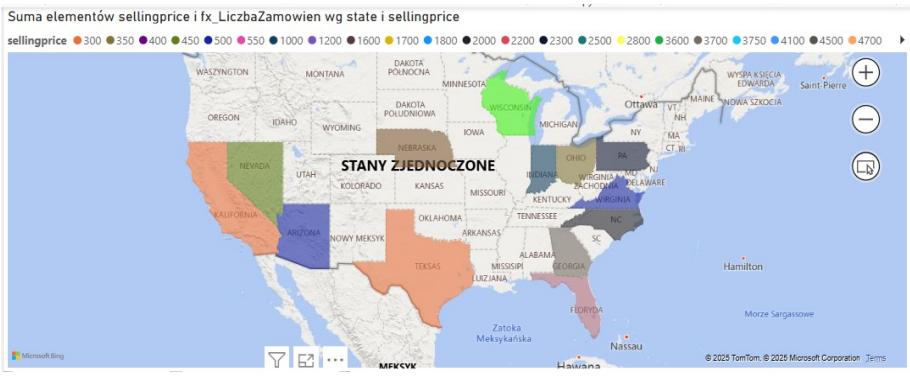
Interaktywna mapa, która:
pokazuje, gdzie sprzedano najwięcej aut,
ile sztuk sprzedano w danym stanie,
jaką wartość miała sprzedaż w danym regionie

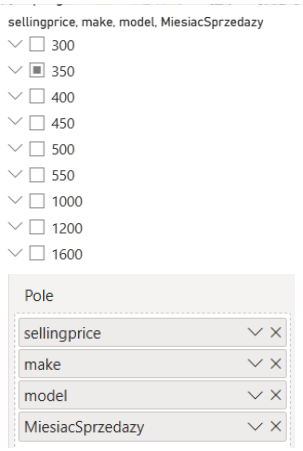
NazwaMiesiaca, make
- grudzień
- luty
- maj
- styczeń

Pole
NazwaMiesiaca
make

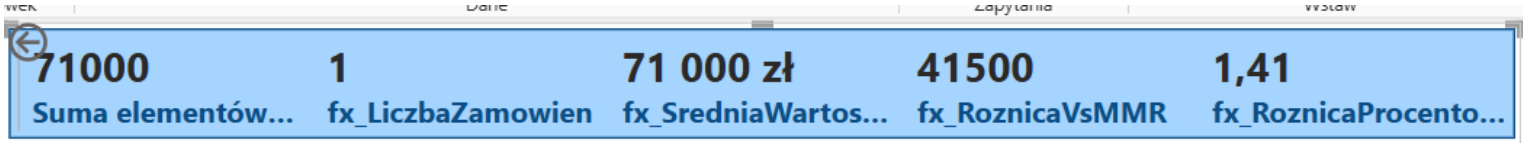
Drąż wskroś
Pomiędzy raportami
Zachowaj wszystkie filtry
Dodaj tutaj pola przegląda...

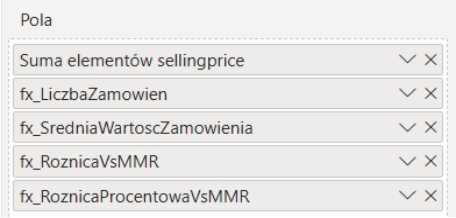
3. B). Stwórz wizualizacje takie jak: **kartogram [1 pkt],**

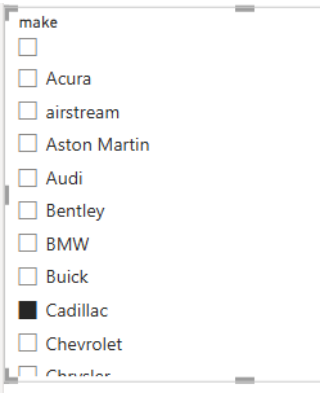




3. C). Stwórz wizualizacje takie jak: **karta z wierszami [1 pkt]**

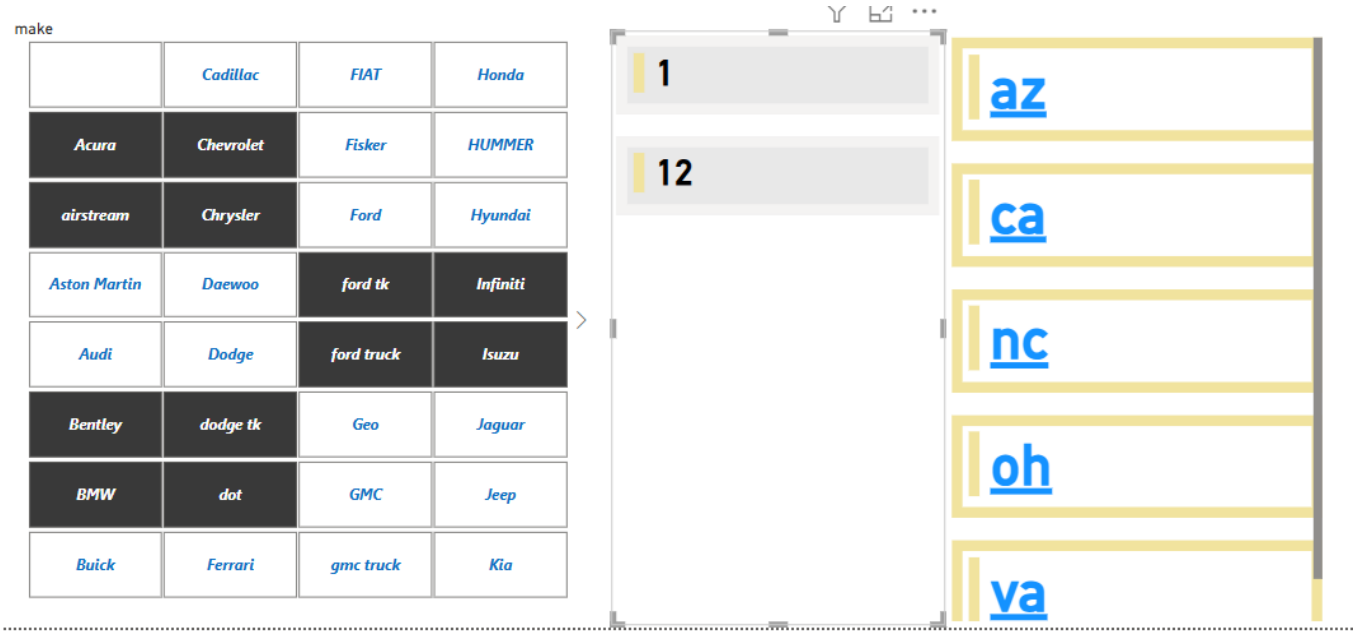




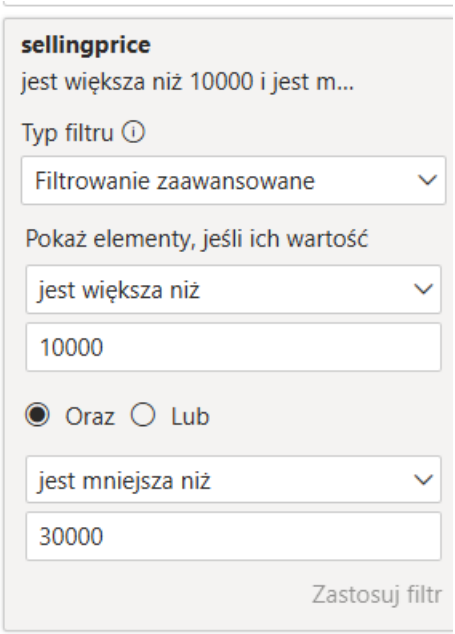


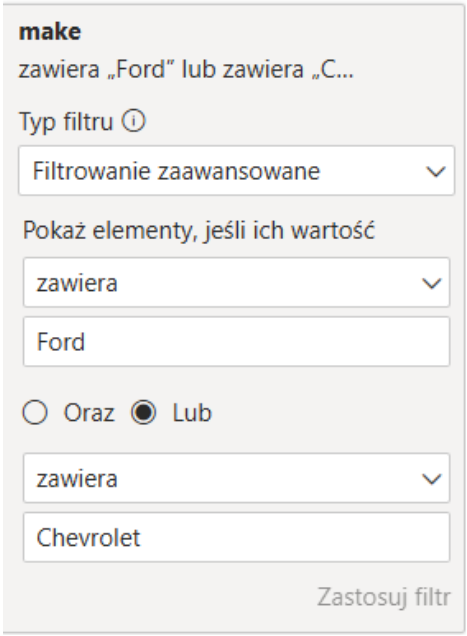
fx_LiczbaZamowien –ilość wierszy/zamówień
Suma elementów sellingprice – łączna wartość sprzedaży
fx_ŚredniaWartoscZamowienia – przeciągnij z listy po prawej
fx_RoznicaVsMMR – różnica ceny sprzedaży względem wartości rynkowej
fx_RoznicaProcentowaVsMMR – ta sama różnica w %

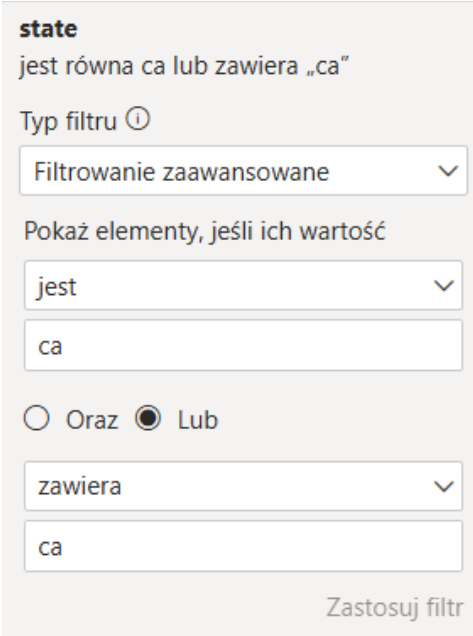
3. D). Stwórz wizualizacje takie jak: **fragmentator [1 pkt]**



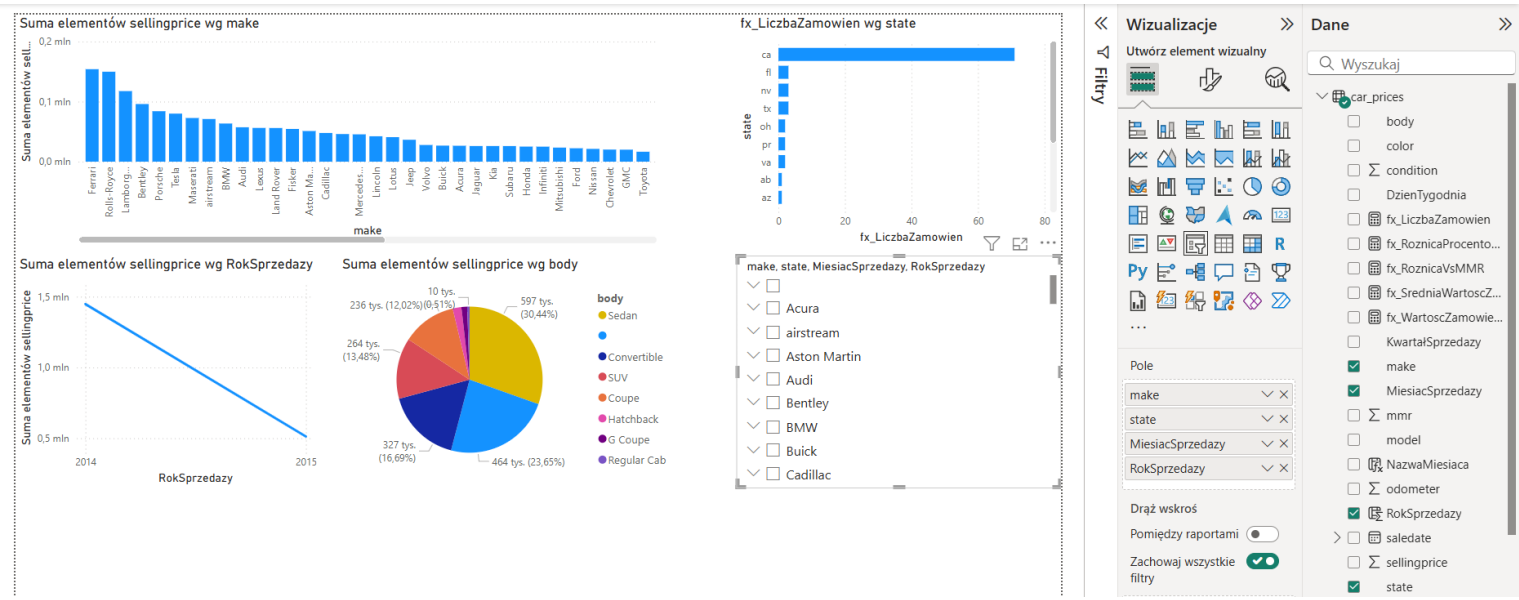
3. E). Stwórz wizualizacje takie jak: **filtry [1 pkt]**







3. F). Stwórz wizualizacje takie jak: **wykresy [1 pkt]**

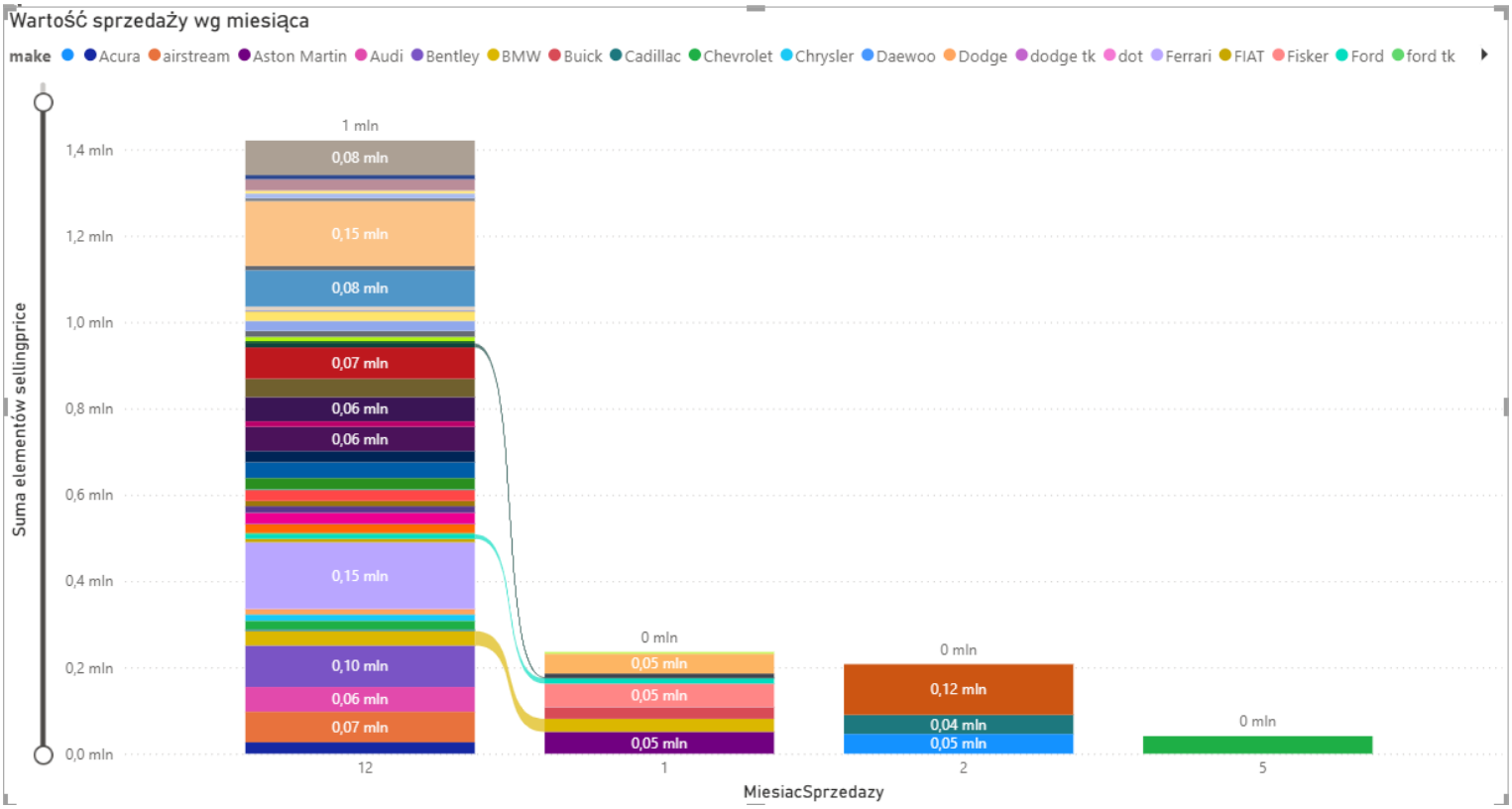


Wykresy prezentują:
- Kto najwięcej sprzedał (producent, stan)
- W którym czasie wystąpiła największa sprzedaż
- Jaki typ nadwozia generuje największe przychody
- Ile było zamówień w poszczególnych kategoriach

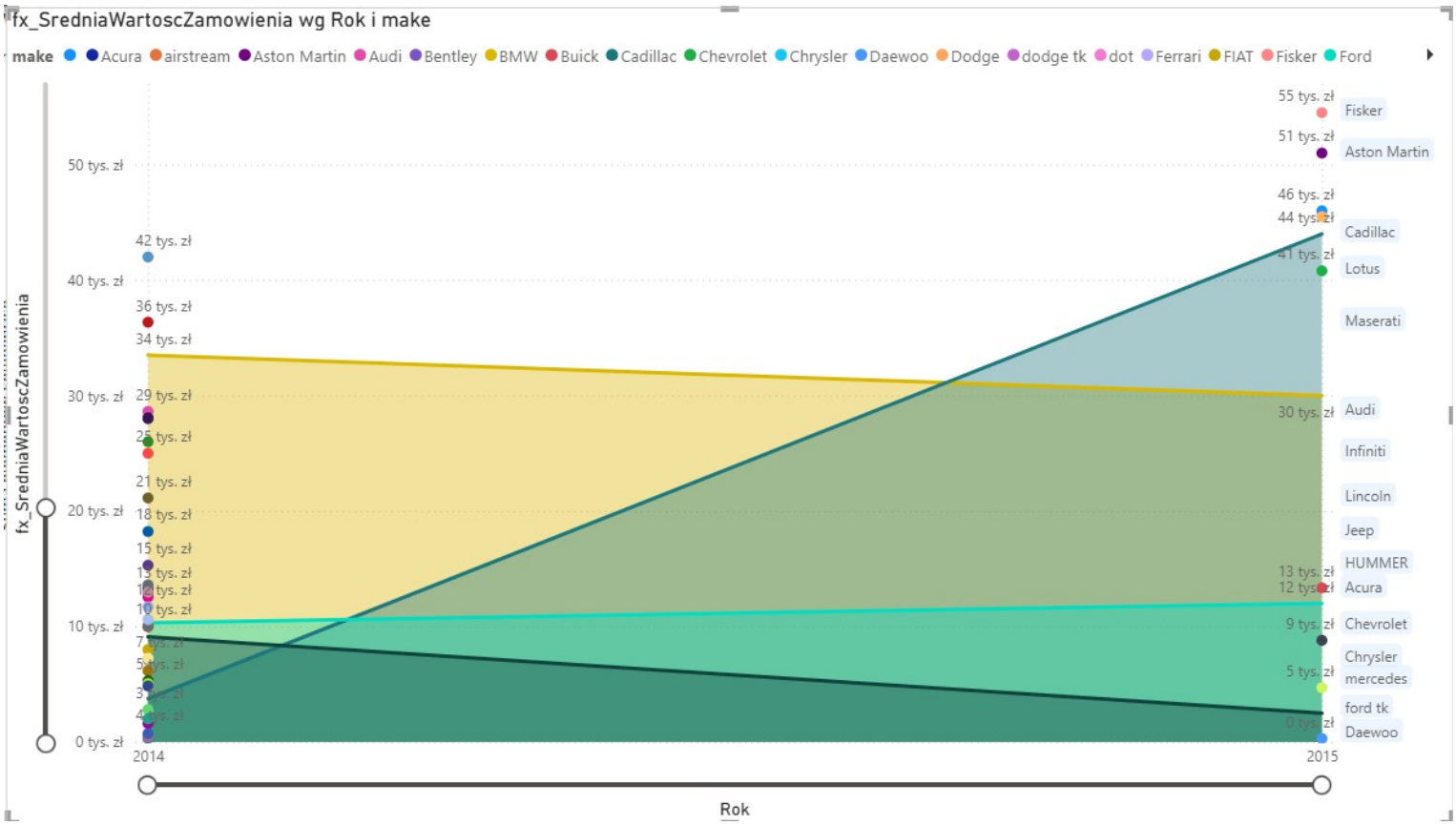
3. G). Stwórz wizualizacje takie jak: **wykresy [1 pkt]**

Rozmiar prostokąta odzwierciedla sumę wartości (np. łączną sprzedaż) dla danej grupy.

3. H). Możesz dodać inne, wybrane przez siebie wizualizacje **[2 pkt]**

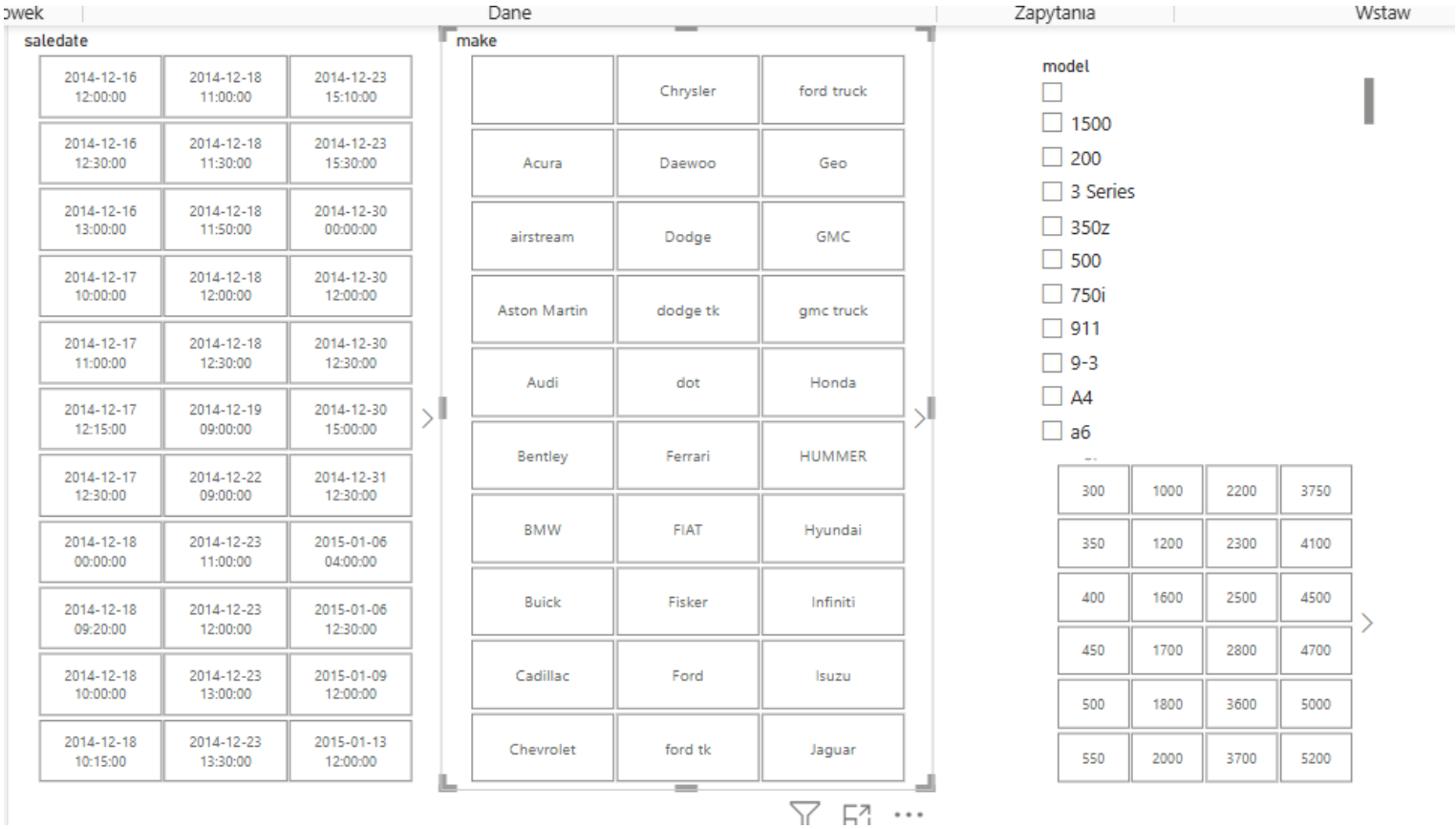


Skumulowany wykres kolumnowy
Porównanie wartości sprzedaży (sellingprice) w czasie — wg miesięcy
Legenda: make l – aby rozbić sprzedaż wg marek



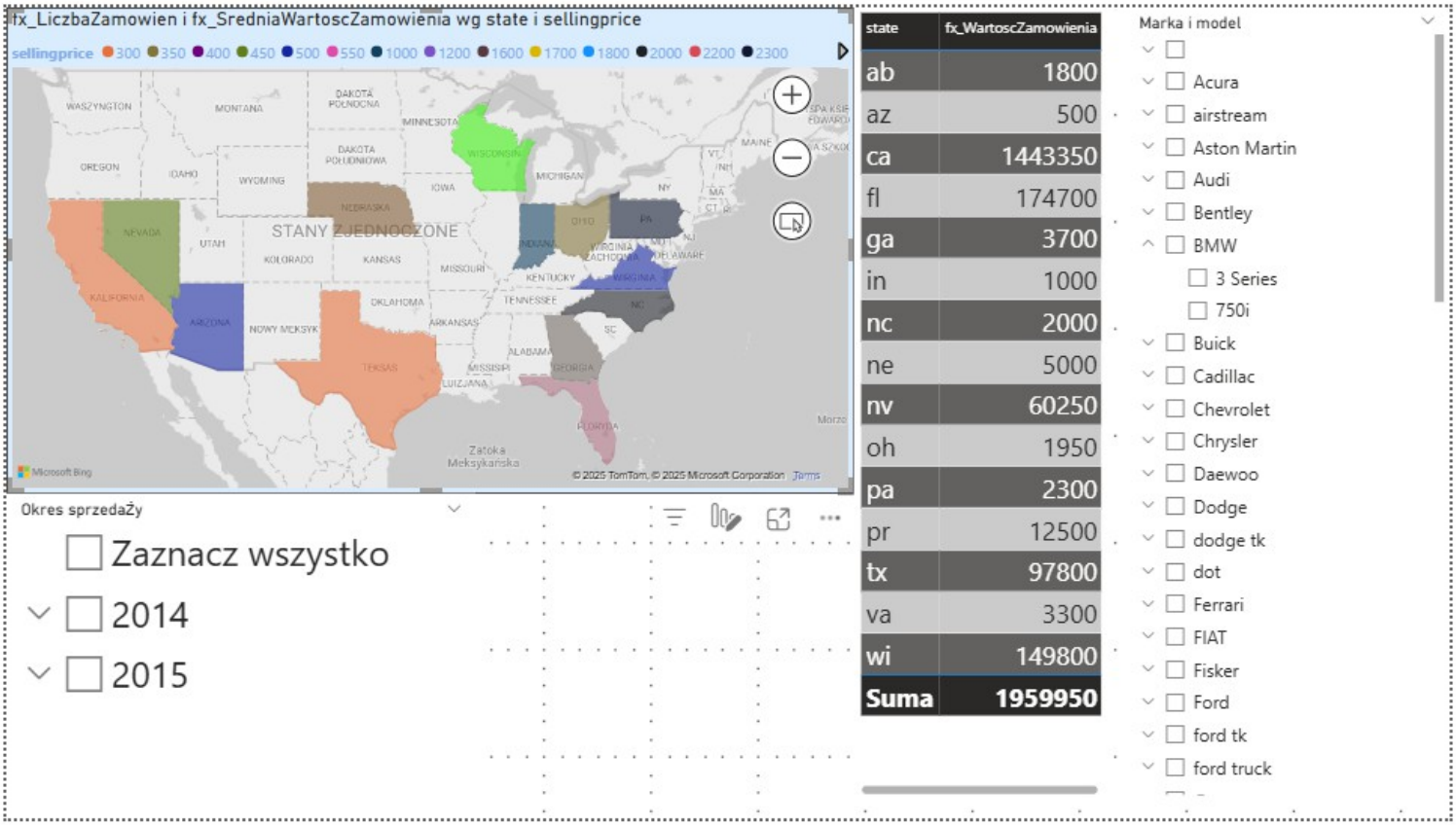
Wykres liniowy
Analiza trendu sprzedaży w czasie

3. I). Dodaj filtry, które umożliwią interaktywną analizę danych, np. filtry na datę, kategorię produktu itp. [1 pkt].



1. Filtr na datę sprzedaży (saledate)
2. Filtr na kategorię produktu – np. make (marka)
3. Filtr na model
4. Filtr przedziału wartości (sellingprice)

W raporcie przedstaw np:
- najlepiej i najgorzej sprzedawane produkty w miastach/regionach; w których kwartałach/miesiącach
- najlepszy/najgorszy sprzedawca, co sprzedał i za jaki zysk wygenerował
- które wymiary (np. kategoria, miasto) najsilniej wpływają na wzrost sprzedaży
- wymusić w raporcie możliwość drill-down (hierarchia danych) z poziomu rocznego do miesięcznego, tygodniowego
- przygotuj prosty storyboard z 2–3 zakładkami (np. Overview → Regiony → Produkty), wykorzystaj przyciski do nawigacji między nimi.

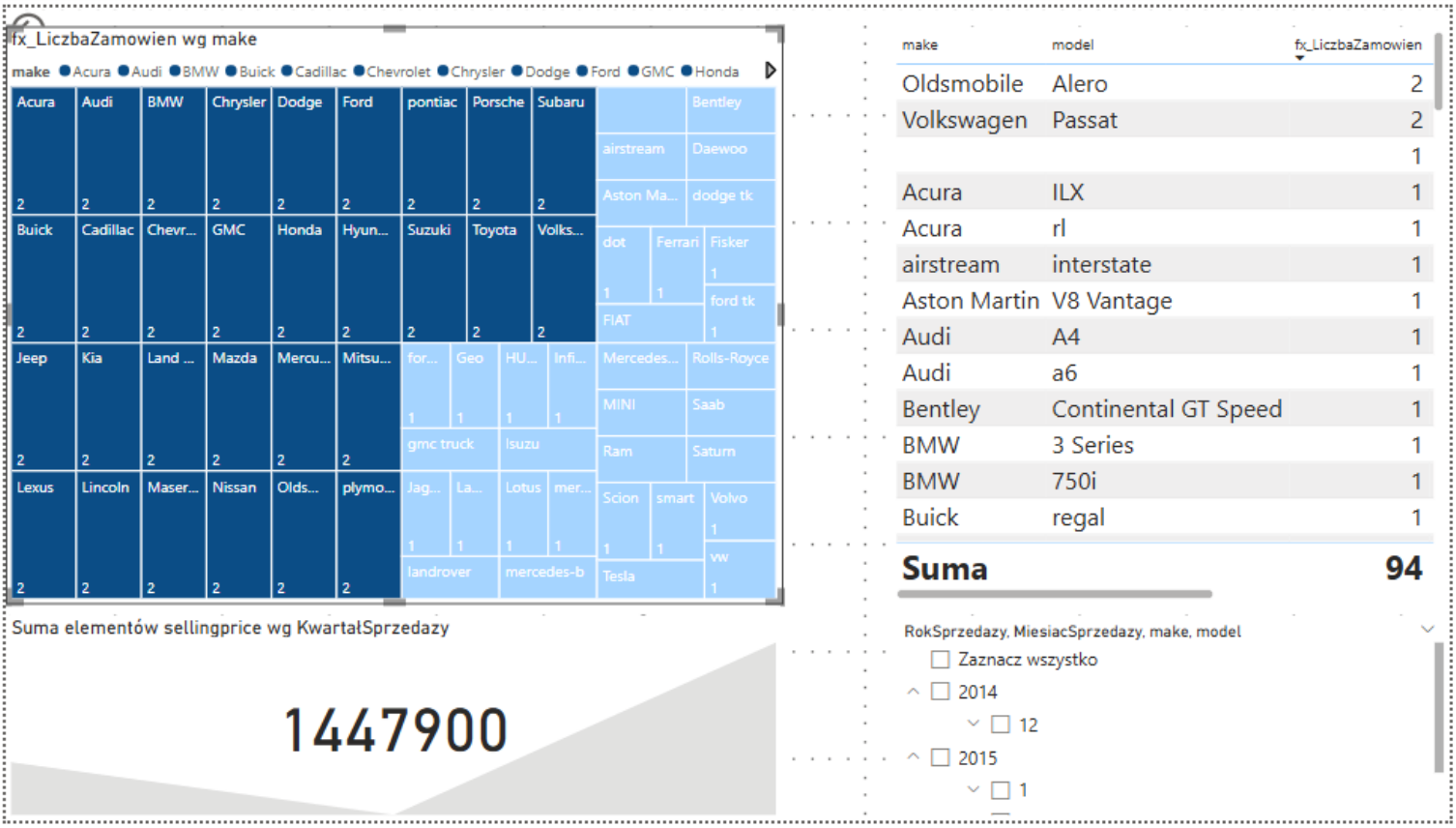


Zakładka „Regiony” – analiza geograficzna sprzedaży
Kartogram lub mapa ArcGIS:
Pole lokalizacji: state (ze słownika dim_state).
Rozmiar bańki / kolor: Suma sellingprice.

Tooltipy: fx_LiczbaZamowien, fx_SredniaWartoscZamowienia.
Tabela lub słupkowy wykres porównujący regiony:

Oś X: state, Y: fx_WartoscZamowienia.
Koloruj stan wg wartości sprzedaży (najlepsze/najgorsze stany).

Fragmentatory:
Rok, kwartał, miesiąc (RokSprzedazy, KwartalSprzedazy, MiesiacSprzedazy).
Marka (make), model (model) – dla zawężenia do konkretnych grup.



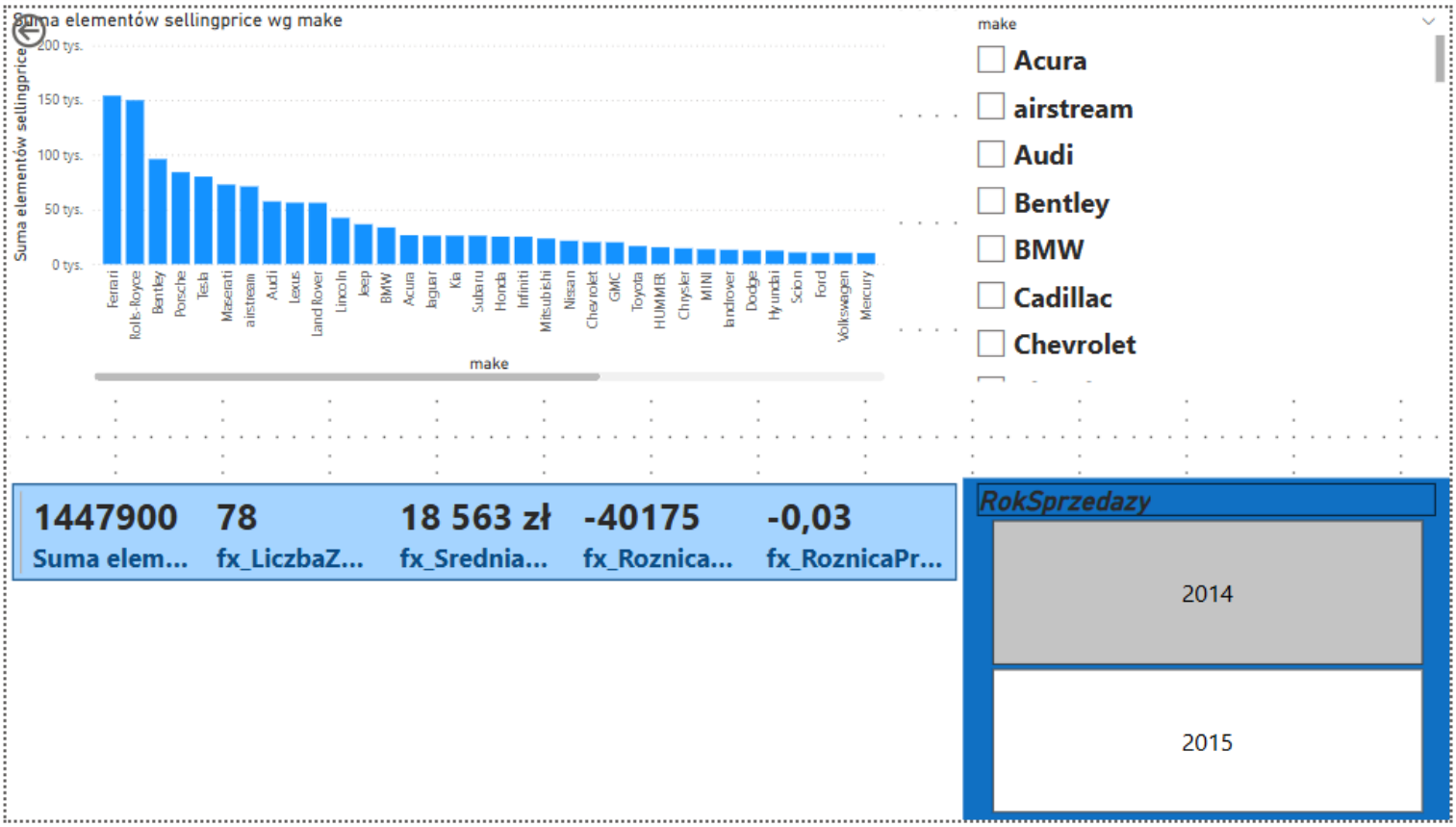
Treemap (drzewo prostokątne): pokazuje fx_LiczbaZamowien według make – bardzo dobrze do identyfikacji najlepiej sprzedających się marek.

Tabela modeli: przedstawia make, model, fx_LiczbaZamowien – idealna do szczegółów.

KPI (liczba 1 447 900): pokazuje Suma elementów sellingprice – działa jako wskaźnik ogólny.

Drugi KPI (94): liczba zamówień (fx_LiczbaZamowien) – świetnie jako metryka pomocnicza.

Fragmentatory: według RokSprzedazy, MiesiacSprzedazy, make, model – umożliwiają bardzo dobre filtrowanie.



1. Suma elementów sellingprice

Jest to **łączna wartość sprzedaży** (czyli przychód brutto) ze wszystkich zarejestrowanych zamówień w zestawie danych. Oznacza to, że produkty o różnej marce i modelu zostały sprzedane łącznie za **prawie 2 miliony jednostek waluty (np. PLN)**.

Ten wskaźnik pokazuje całkowity zasięg finansowy działalności.

2. fx_LiczbaZamowien

Wskazuje łączną **liczbę zamówień** (czyli przypadków sprzedaży). Każde zamówienie odpowiada jednej transakcji – niezależnie od wartości sprzedaży.

Może być użyteczny do obliczania średniej wartości zamówienia lub porównania intensywności sprzedaży między regionami

3. fx_SredniaWartoscZamowienia

Pokazuje średnią wartość sprzedaży na jedno zamówienie.
Jest to wyliczenie: Suma(sellingprice) / LiczbaZamowien, czyli przeciętny koszt pojedynczej transakcji.

Pozwala ocenić wartość przeciętnego klienta — przydatne do planowania działań marketingowych i rabatowych.

4. fx_RoznicaVsMMR

To różnica między ceną sprzedaży a wartością referencyjną MMR (Market Mean Retail).
Ujemna wartość oznacza, że pojazdy były sprzedawane poniżej ceny rynkowej.

Może wskazywać na wyprzedaże, nadwyżki magazynowe lub agresyjną politykę cenową.

5. fx_RoznicaProcentowaVsMMR

Wskaźnik pokazuje średnią procentową różnicę względem ceny MMR.
Wartość -0,04 oznacza, że sprzedaż była średnio o 4% niższa niż referencyjna cena rynkowa.

Pokazuje efektywność cenową – przydatny wskaźnik rentowności i atrakcyjności oferty.

https://app.powerbi.com/groups/me/lineage?actions=highlight%2Cimpact&artifactId=777e9a9e-593c-451e-b7bf-c4d5c21f8e22&src=Desktop&experience=power-bi