

BAZY DANYCH

WYKŁAD V

HURTOWNIE DANYCH CD.

MOTYWACJA

Cele:

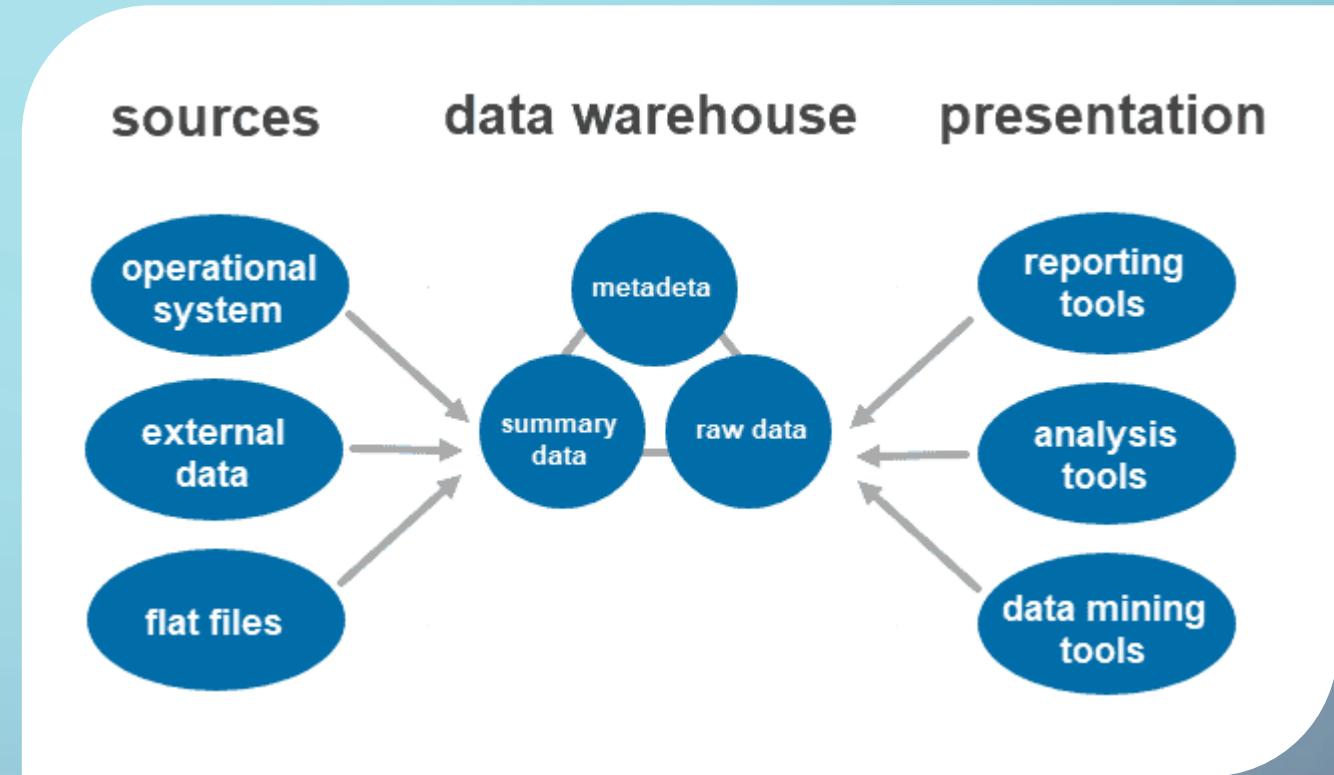
- analiza wzrostu sprzedaży,
- zwiększenie udziału firmy w rynku,

Jak to robić:

- analiza zachowania klientów,
- wskazanie słabych punktów firmy,
- analiza trendów.

Problemy techniczne:

- Dane (czym dysponujemy)?
- Jakie dane powinniśmy gromadzić w hurtowni?
 - Wszystkie?
 - Tylko to, co niezbędne?
- Czy wszystkie dane powinny być widoczne?



MODEL

Przypomnienie:

Centralna hurtownia danych projektowana jest jak zwykła baza danych i zawiera fakty opisane przez wymiary i określające wartość miar.

- Fakt - pojedyncze zdarzenie będące podstawą analiz (np. sprzedaż). Fakty opisane są przez wymiary i miary.
- Wymiar - cecha opisująca dany fakt, pozwalająca powiązać go z innymi pojęciami modelu przedsiębiorstwa (np. klient, data, miejsce, produkt). Wymiary są opisane atrybutami.
- Atrybut - cecha wymiaru, przechowująca dodatkowe informacje na temat faktu (np. wymiar data może mieć atrybuty: miesiąc, kwartał, rok; wymiar klient może mieć atrybuty: nazwisko, region).
- Miara - wartość liczbową przyporządkowana do danego faktu (np. wartość sprzedaży, liczba sztuk).

ROLAP

Relacyjny model przetwarzania analitycznego, w którym dane są przechowywane jak w relacyjnej bazie danych. W modelu ROLAP dane znajdują się z przodu użytkownika w wielowymiarowej formie. Aby wyświetlić dane, w widoku wielowymiarowym tworzona jest semantyczna warstwa metadanych, która odwzorowuje wymiary na tabele relacyjne. Metadane obsługuje również agregację danych.

Problemy:

- Zidentyfikowanie faktów
- Zidentyfikowanie kluczowych wymiarów
- Zaprojektowanie tabel faktów
- Zaprojektowanie tabel wymiarów

ROLAP – ZIDENTYFIKOWANIE FAKTÓW

Dla tego modelu przetwarzania analitycznego należy wskazać kluczowe typy transakcji w systemie produkcyjnym, realizujące kluczowe akcje/operacje w obszarze działania przedsiębiorstwa.

Przykłady:

- handel: transakcje sprzedaży
- bankowość: kursy walut, operacje na rachunkach
- giełda: wahania kursów akcji, operacje giełdowe
- ubezpieczenia: wykupienie polisy, zmiana warunków polisy, zgłoszenie szkody, wypłacenie odszkodowania
- telekomunikacja: zrealizowanie rozmowy przez abonenta, podłączenie telefonu, zawarcie umowy, zmiana abonamentu, płatności za abonament
- opieka zdrowotna: przyjęcie pacjenta do szpitala, forma leczenia, wynik leczenia

ROLAP – ZIDENTYFIKOWANIE WYMIARÓW

Należy określić kontekst dla analizy faktów:

- handel:
 - analiza sprzedaży w poszczególnych miastach i założonych okresach czasowych
- bankowość:
 - wahania kursów walut w poszczególnych dniach
 - analiza przyrostu ilości nowych rachunków w poszczególnych miesiącach z podziałem na rodzaje rachunków
- giełda:
 - wahania kursów akcji firm w danym przedziale czasowym
 - Ilość zawartych transakcji kupna lub sprzedaży w jednostce czasu i łączne kwoty tych operacji
- ubezpieczenia:
 - analiza przyrostu lub spadku ilości polis poszczególnych rodzajów w miastach w poszczególnych miesiącach
- telekomunikacja:
 - analiza rozkładu czasu rozmów poszczególnych klientów w czasie doby

MOLAP

MOLAP to wielowymiarowy model przetwarzania analitycznego online . Dane wykorzystywane do analizy przechowywane są w wyspecjalizowanych wielowymiarowych bazach danych (MDDB). Dane przetwarzane przez MOLAP są często wstępnie przeliczane oraz składowane w wielowymiarowej kostce, która implementowana jest przeważnie jako wielowymiarowa tablica.

Komórki lub moduły danych w tych wielowymiarowych bazach danych zawierają wstępnie obliczone i prefabrykowane dane. Własne systemy oprogramowania tworzą te wstępnie obliczone i sfabrykowane dane, podczas gdy dane są ładowane do MDDB z głównych baz danych.

MOLAP

Zalety MOLAP:

- duża wydajność zapytań dzięki optymalnej technice przechowywania danych, wielowymiarowemu indeksowaniu oraz cache'owaniu,
- zniwelowany rozmiar danych na dysku w porównaniu z rozwiązaniami relacyjnymi,
- automatyczne przetwarzanie agregatów wyższego rzędu,
- niewielkie rozmiary dla małej liczby wymiarów,
- model tablicowy zapewnia naturalne indeksowanie danych,
- efektywny dostęp do danych.

Wady MOLAP:

- Są przypadki w których wstępne przetwarzanie danych może być zadaniem złożonym obliczeniowo, zwłaszcza dla dużych porcji danych. Efekt ten niwelowany jest poprzez przyrostowe przetwarzanie jedynie zmienionych danych.
- problemy z wybieraniem danych o dużej liczbie wymiarów (np. kilka milionów elementów),
- redundancja danych.

ROLAP A MOLAP

- ROLAP to Relational Online Analytical Processing a MOLAP to Multidimensional Online Analytical Processing.
- Dane w modelu ROLAP i MOLAP są przechowywane w magazynie głównym. Dane ROLAP są pobierane bezpośrednio z głównego magazynu, podczas gdy dane MOLAP są pobierane z firmowych baz danych MDDB.
- W ROLAP dane są przechowywane w postaci tabel relacyjnych, ale w danych MOLAP są przechowywane w postaci wielowymiarowej tablicy złożonej z kostek danych.
- ROLAP zajmuje się dużymi wolumenami danych, podczas gdy MOLAP zajmuje się ograniczonymi zbiorami danych przechowywanymi w MDDB.
- Silniki ROLAP wykorzystują złożone zapytania SQL do pobierania danych z hurtowni danych. Jednak silnik MOLAP tworzy prefabrykowane i wstępnie obliczone datacuby, aby przedstawić wielowymiarowy widok danych dla użytkownika i zarządzać niespójnością danych w kostkach danych, MOLAP wykorzystuje technologię macierzy Sparse.
- Silnik ROLAP dynamicznie tworzy wielowymiarowy widok danych, podczas gdy MOLAP statycznie przechowuje wielowymiarowy widok danych w zastrzeżonych bazach danych MDDB, aby użytkownik mógł je przeglądać stamtąd.
- Ponieważ ROLAP dynamicznie tworzy wielowymiarowy widok danych, jest wolniejszy od MOLAP, który nie marnuje czasu na tworzenie wielowymiarowego widoku danych.

HOLAP

HOLAP (Hybrid OLAP) – stara się łączyć zalety ROLAP i MOLAP. Wykorzystane są tutaj jednocześnie wielowymiarowe (dane zagregowane) oraz relacyjne (dane elementarne, bez dublowania) bazy danych. Zapytania szczegółowe są wykonywane wolniej, ponieważ działają na strukturze relacyjnej (ROLAP). Dane w HOLAP mają zwykle mniejszy rozmiar niż dane w MOLAP. Struktury HOLAP są przeznaczone dla potrzeb szybkiego dostępu do agregacji bazujących na dużych zbiorach danych.

Przykładami produktów komercyjnych wykorzystujących technologie przechowywania HOLAP są między innymi : Microsoft Analysis Services, Oracle Database OLAP Option, MicroStrategy. Również programy serwerowe bazują na HOLAP np. Microsoft SQL Server 7.0 OLAP.

OLAP – PROJEKTOWANIE TABELI FAKTÓW

a) Analiza poziomu szczegółowości informacji (ma to wpływ na rozmiar tabeli faktów)

- rejestrowanie kwoty zakupu pojedynczego produktu
- rejestrowanie sumarycznej kwoty zakupu całego koszyka
- rejestrowanie sumarycznej kwoty zakupu w tygodniu

b) Horyzont czasowy danych

- jak długo przechowywać informacje na najwyższym poziomie szczegółowości?
- opracowanie strategii agregowania danych starszych
 - raporty roczne: najczęściej wystarczająq aggregaty sumującce fakty z dokładnością do tygodnia
 - raporty agregującce dane sprzed kilku lat: najczęściej wystarczająq aggregaty sumującce fakty z dokładnością do miesiąca lub roku

OLAP – PROJEKTOWANIE TABELI FAKTÓW

- c) Właściwy zbiór atrybutów tabeli faktów
- d) Usunięcie zbędnych atrybutów (wpływ na rozmiar tabeli faktów)
 - czy atrybut wnosi nową/niezbędną wiedzę o fakcie?
 - czy wartość atrybutu można wyliczyć?
- e) Minimalizacja rozmiarów atrybutów
 - przykład: telekomunikacja
 - tabela wymiaru Abonenci zawiera dużą liczbę, abonentów, każdy abonent dzwoni średnio kilka razy dziennie. W przypadku rocznego horyzontu tabeli faktów ilość danych jest bardzo duża.

TROCHĘ PRAKTYKI

AdventureWorks Database (plik AdventureWorksDW2019.bak) to przykładowy produkt firmy Microsoft dla bazy danych przetwarzania transakcji online. Baza danych AdventureWorks obsługuje fikcyjną, międzynarodową firmę produkcyjną o nazwie Adventure Works Cycles.

Wymagania:

- MSSQL
- Visual Studio

TROCHĘ PRAKTYKI

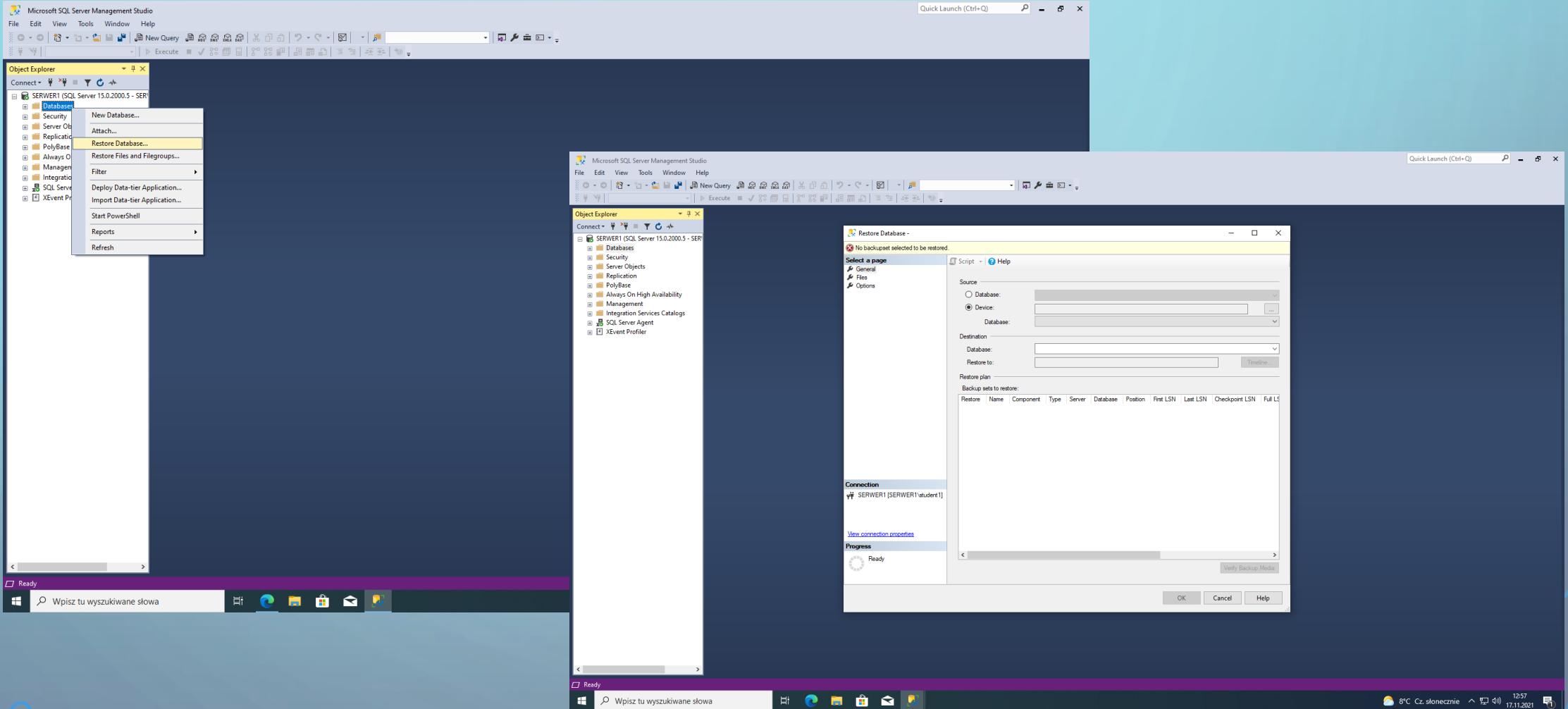
Usługi Analysis Services zapewniają różne podejścia do tworzenia modelu semantycznego Business Intelligence:

- Rozwiązania tabelaryczne wykorzystując modelowanie relacyjne do konstruowania tabel i relacji na potrzeby modelowania danych oraz mechanizm analizy w pamięci do przechowywania i obliczania danych.
- Rozwiązania wielowymiarowe i data mining wykorzystując modelowanie OLAP do konstrukcji (kostek i wymiarów) oraz pamięci masowej MOLAP, ROLAP lub HOLAP.

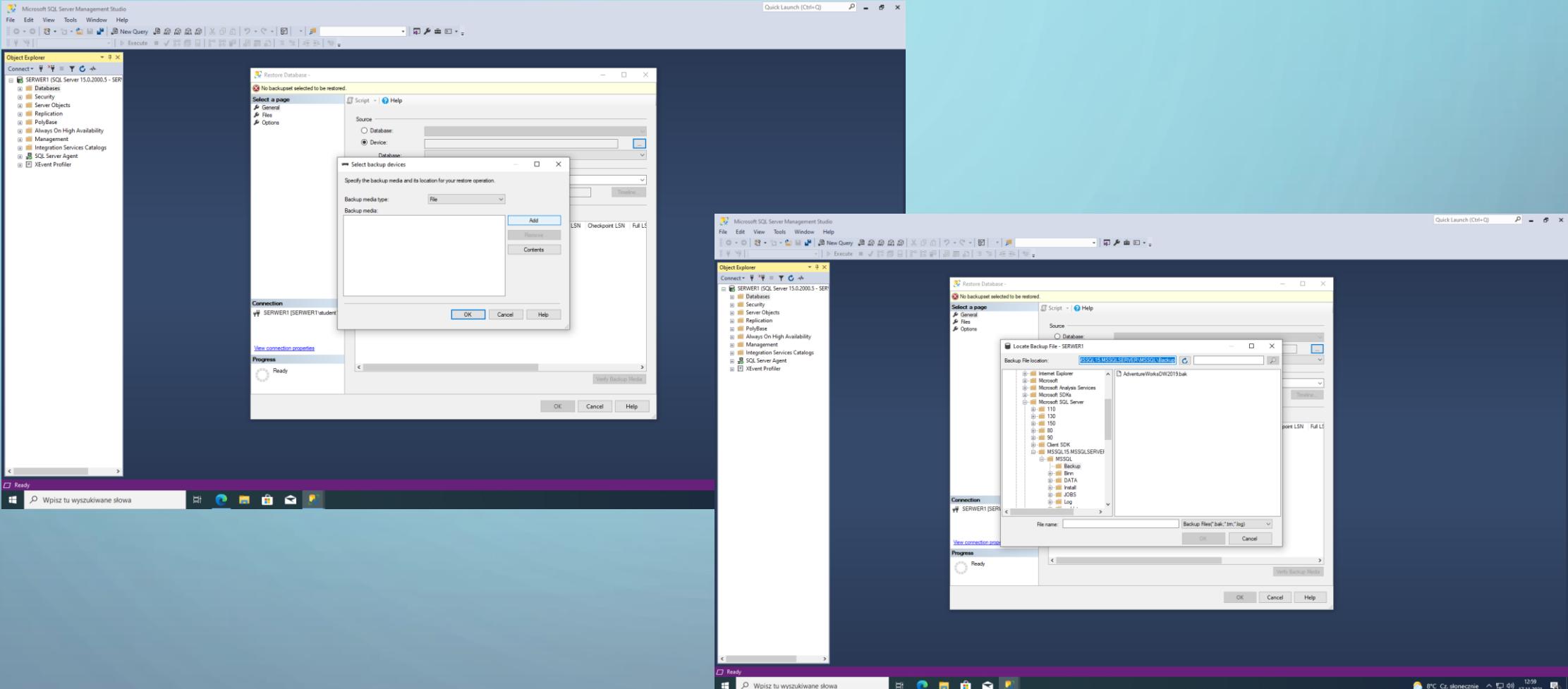
TROCHĘ PRAKTYKI

PRZYWRACANIE BAZY DANYCH

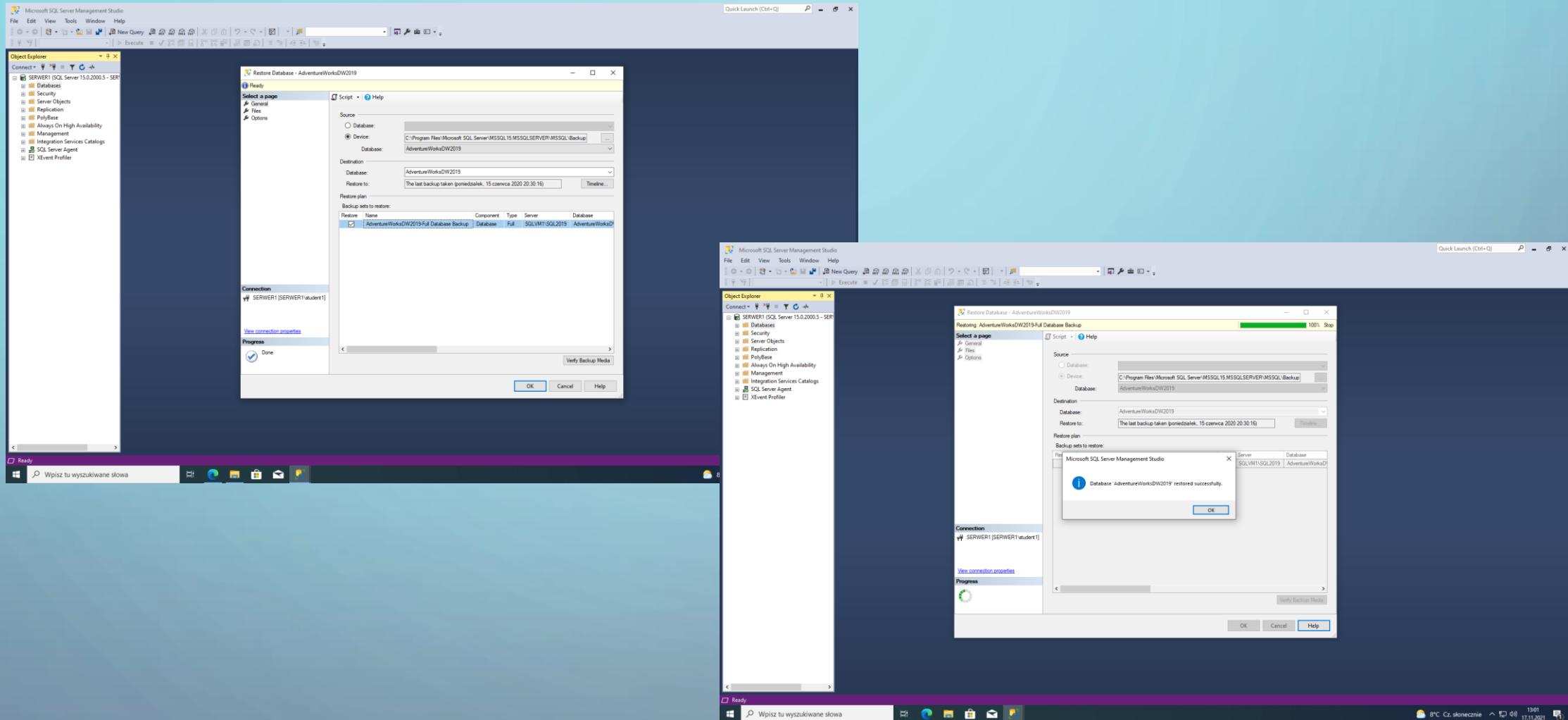
TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface. The title bar reads "SQLQuery2.sql - SERWER1.AdventureWorksDW2019 (SERWER1\student1 (57)) - Microsoft SQL Server Management Studio". The left pane is the Object Explorer, showing the database structure for "SERWER1 (SQL Server 15.0.2000.5) - AdventureWorksDW2019". The right pane displays the results of a query:

```
***** Script for SelectTopNRows command from SSMS *****/
SELECT TOP (1000) [CustomerKey]
,[GeographyKey]
,[CustomerAlternateKey]
,[Title]
,[FirstName]
,[MiddleName]
,[LastName]
,[NameStyle]
,[BirthDate]
,[MaritalStatus]
,[Suffix]
,[Gender]
,[EmailAddress]
,[YearlyIncome]
,[TotalChildren]
,[NumberChildrenAtHome]
,[EnglishEducation]
,[SpanishEducation]
,[FrenchEducation]
,[EnglishOccupation]
,[SpanishOccupation]
,[FrenchOccupation]
,[HouseOwnerFlag]
,[NumberCarsOwned]
,[AddressLine1]
```

The results grid shows data for 1000 rows from the "Customer" table. The columns include CustomerKey, GeographyKey, CustomerAlternateKey, Title, FirstName, MiddleName, LastName, NameStyle, BirthDate, MaritalStatus, Suffix, Gender, EmailAddress, YearlyIncome, TotalChildren, NumberChildrenAtHome, EnglishEducation, SpanishEducation, and FrenchOccupation. The first few rows of data are:

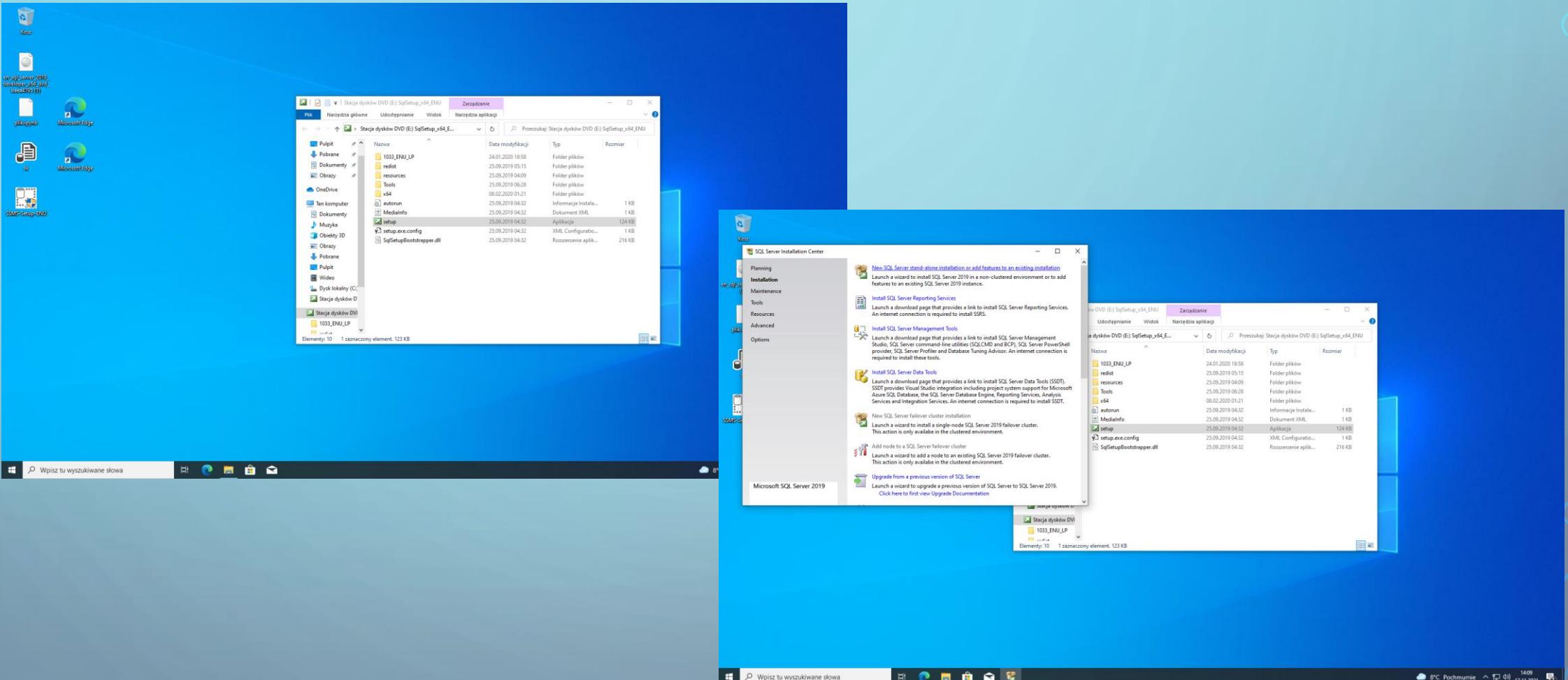
CustomerKey	GeographyKey	CustomerAlternateKey	Title	FirstName	MiddleName	LastName	NameStyle	BirthDate	MaritalStatus	Suffix	Gender	EmailAddress	YearlyIncome	TotalChildren	NumberChildrenAtHome	EnglishEducation	SpanishEducation	FrenchOccupation
983	11982	359	AW00011982	NULL	Alexandra	NULL	Foster	0	1962-10-02	S	NULL	F	alexandra37@adventure-works.com	60000.00	2	0	High School	Educación secundaria
984	11983	298	AW00011983	NULL	Cheryl	NULL	Ortega	0	1968-02-24	M	NULL	F	chery24@adventure-works.com	70000.00	3	2	Partial College	Estudios universitarios (en curso)
985	11984	59	AW00011984	NULL	Alex	NULL	Campbell	0	1962-12-04	M	NULL	M	alex35@adventure-works.com	80000.00	2	1	Bachelors	Licenciatura
986	11985	536	AW00011985	NULL	Melanie	A	Hughes	0	1963-05-18	M	NULL	F	melanie27@adventure-works.com	80000.00	2	1	Bachelors	Licenciatura
987	11986	34	AW00011986	NULL	Max	M	Alvarez	0	1969-01-20	S	NULL	M	max5@adventure-works.com	70000.00	0	0	Bachelors	Licenciatura
988	11987	38	AW00011987	NULL	Diane	D	Vazquez	0	1973-05-26	M	NULL	F	diane19@adventure-works.com	90000.00	4	4	Bachelors	Licenciatura
989	11988	4	AW00011988	NULL	Kathryn	NULL	Chapman	0	1968-02-17	S	NULL	F	kathryn14@adventure-works.com	70000.00	1	0	Partial College	Estudios universitarios (en curso)
990	11989	24	AW00011989	NULL	Kara	NULL	Anand	0	1984-03-20	S	NULL	F	kara20@adventure-works.com	70000.00	1	0	Partial College	Estudios universitarios (en curso)
991	11990	10	AW00011990	NULL	Joanna	J	Vazquez	0	1968-01-12	S	NULL	F	joanna13@adventure-works.com	70000.00	1	0	Partial College	Estudios universitarios (en curso)
992	11991	9	AW00011991	NULL	Frederick	NULL	Martinez	0	1968-02-07	M	NULL	M	frederick14@adventure-works.com	60000.00	1	0	Bachelors	Licenciatura
993	11992	20	AW00011992	NULL	Tonya	NULL	Chande	0	1972-06-09	S	NULL	F	tonya14@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)
994	11993	4	AW00011993	NULL	Rosa	I	Wang	0	1977-09-12	S	NULL	F	rosa1@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)
995	11994	27	AW00011994	NULL	Leah	NULL	Hu	0	1977-10-18	S	NULL	F	leah16@adventure-works.com	110000.00	0	0	Graduate Degree	Estudios de postgrado
996	11995	39	AW00011995	NULL	Kelvin	A	Carson	0	1965-12-12	M	NULL	M	kelvin11@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)
997	11996	21	AW00011996	NULL	Veronica	NULL	Sirin	0	1966-01-30	S	NULL	F	veronica9@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)
998	11997	2	AW00011997	NULL	Kristina	NULL	Kapoor	0	1971-04-04	M	NULL	F	kristina1@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)
999	11998	24	AW00011998	NULL	Donna	NULL	Sharma	0	1966-02-07	M	NULL	F	donna9@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)
1000	11999	10	AW00011999	NULL	Johnny	NULL	Shan	0	1965-12-03	M	NULL	M	johnny11@adventure-works.com	60000.00	1	0	Partial College	Estudios universitarios (en curso)

At the bottom, a message says "Query executed successfully." and the status bar shows "SERWER1 (15.0 RTM) | SERWER1\student1 (57) | AdventureWorksDW2019 | 00:00:00 | 1000 rows". The system tray at the bottom right shows the date and time as "13:03 8°C Cz. słonecznie 17.11.2021".

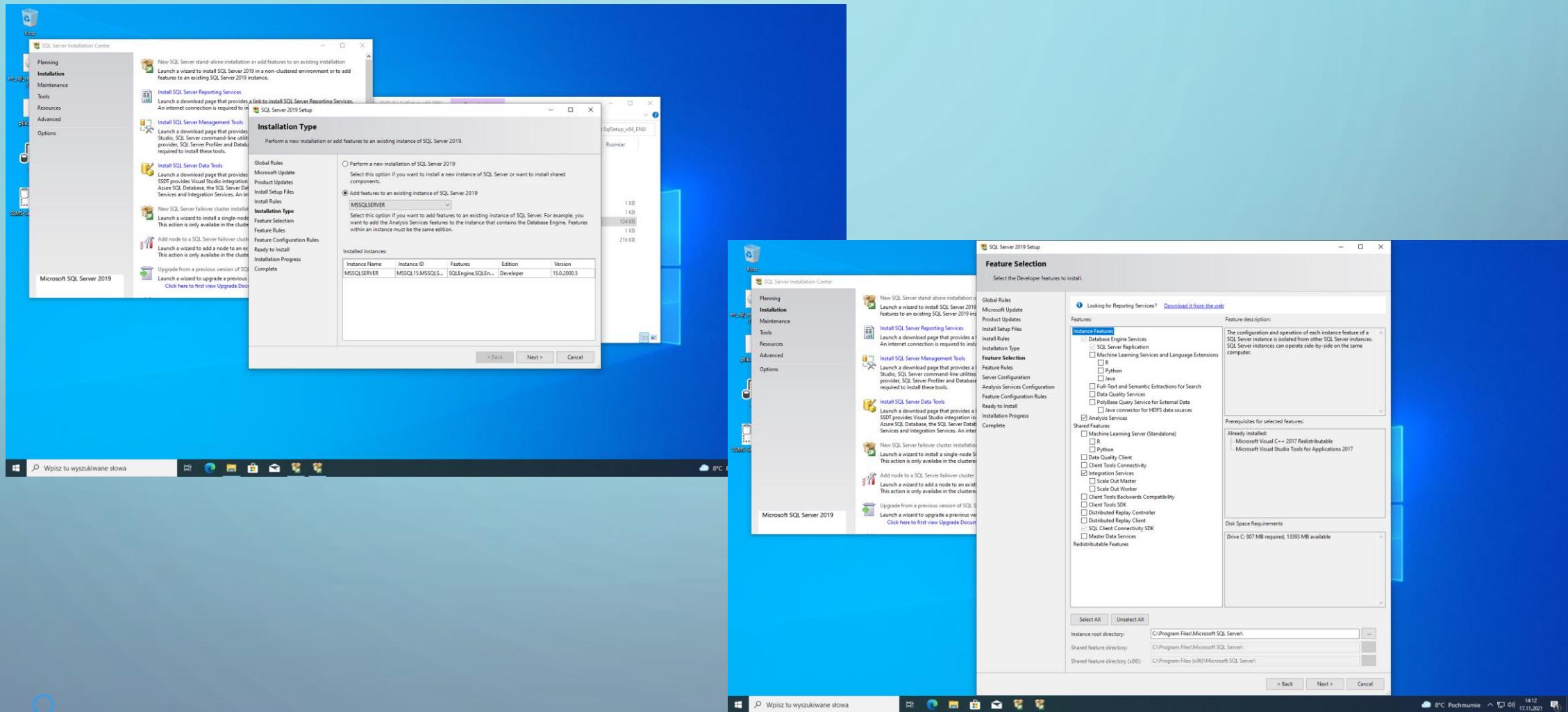
TROCHĘ PRAKTYKI

KONFIGURACJA MSSQL (INSTALACJA NARZĘDZI)

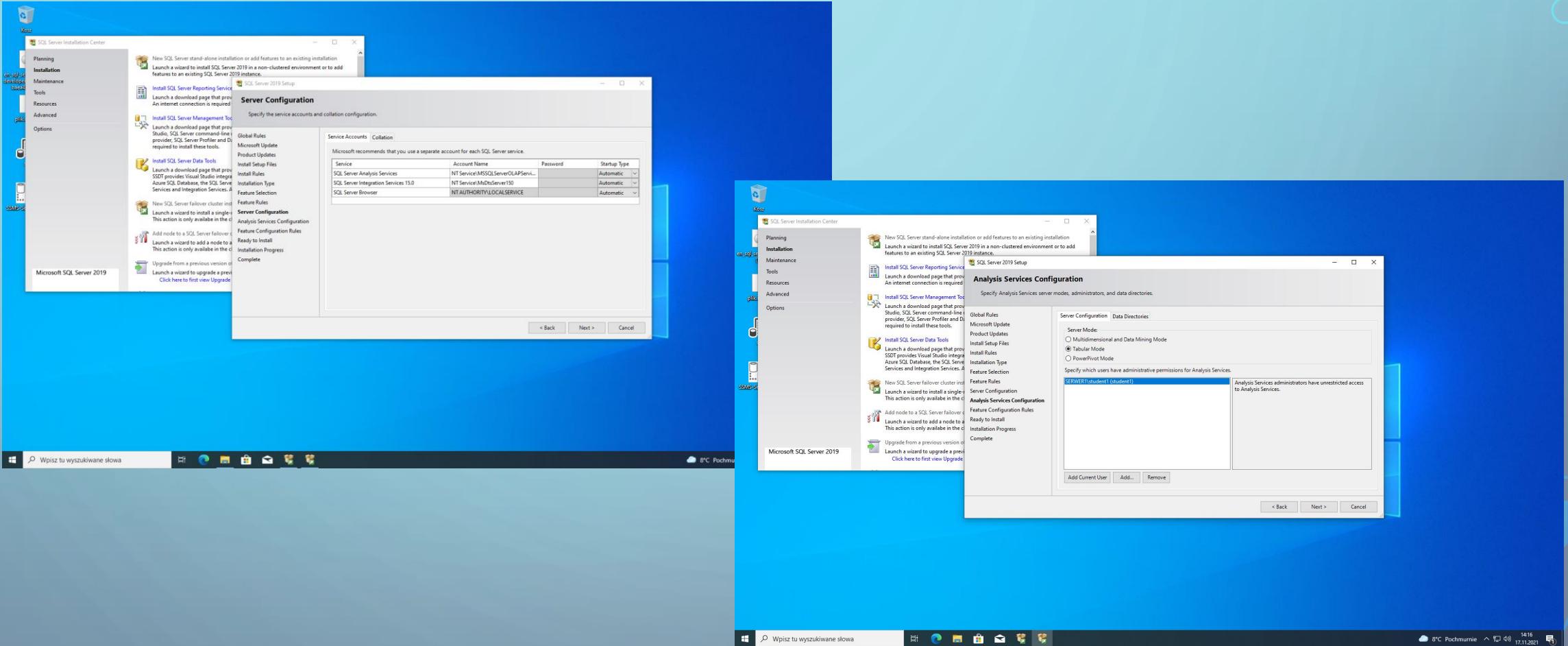
TROCHĘ PRAKTYKI



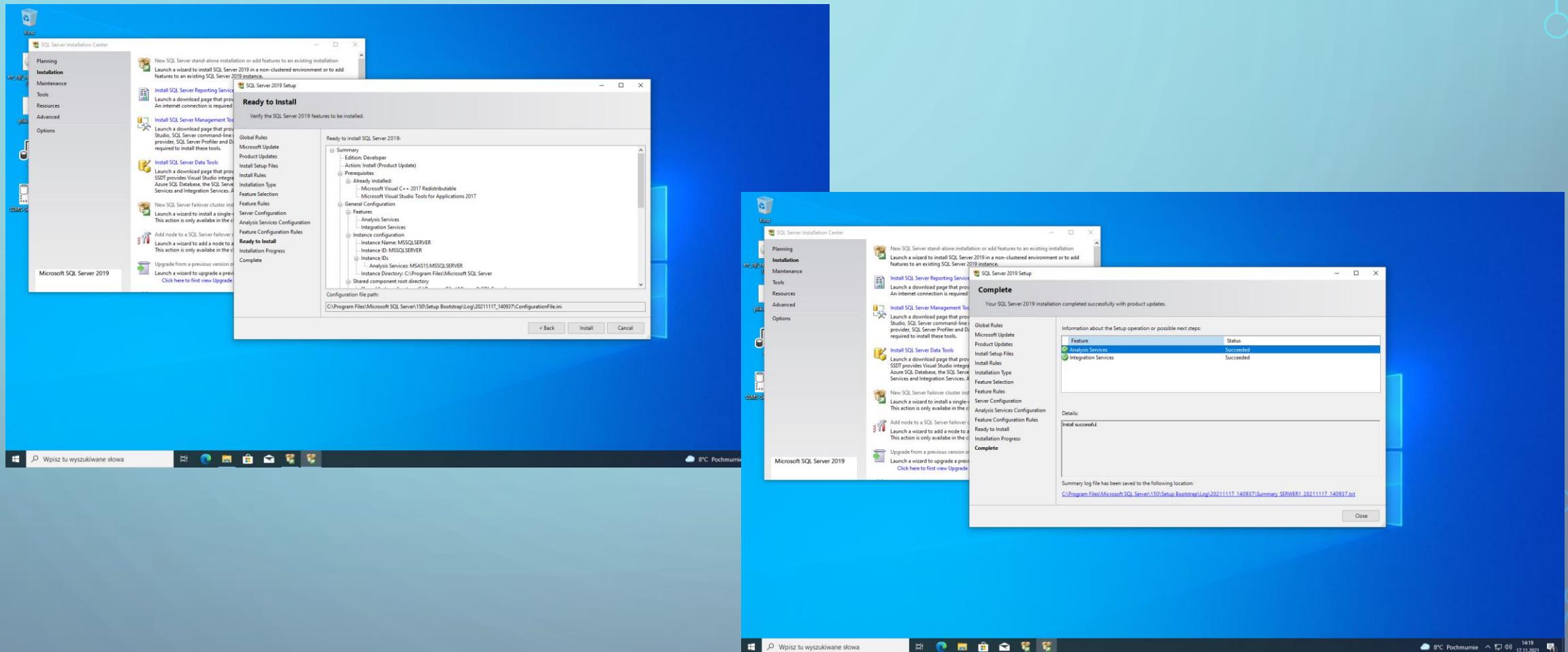
TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI



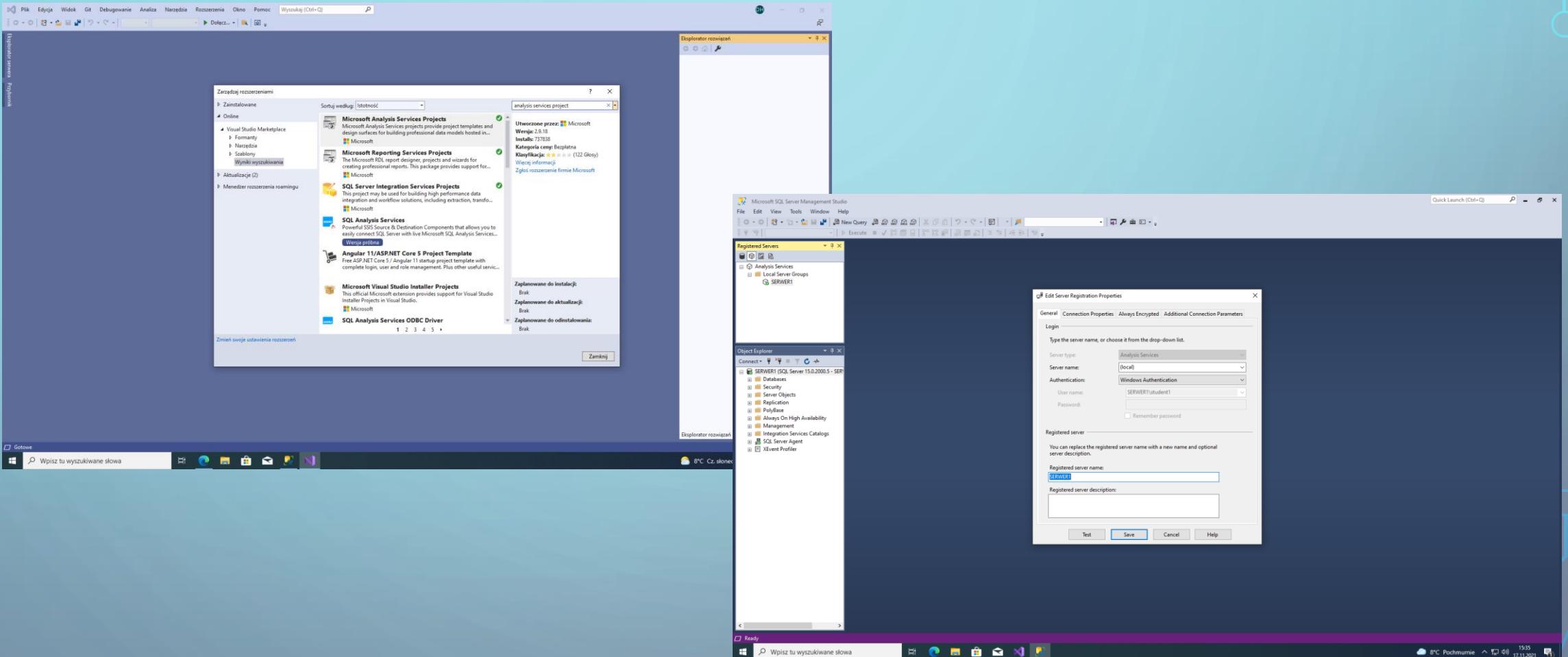
TROCHĘ PRAKTYKI



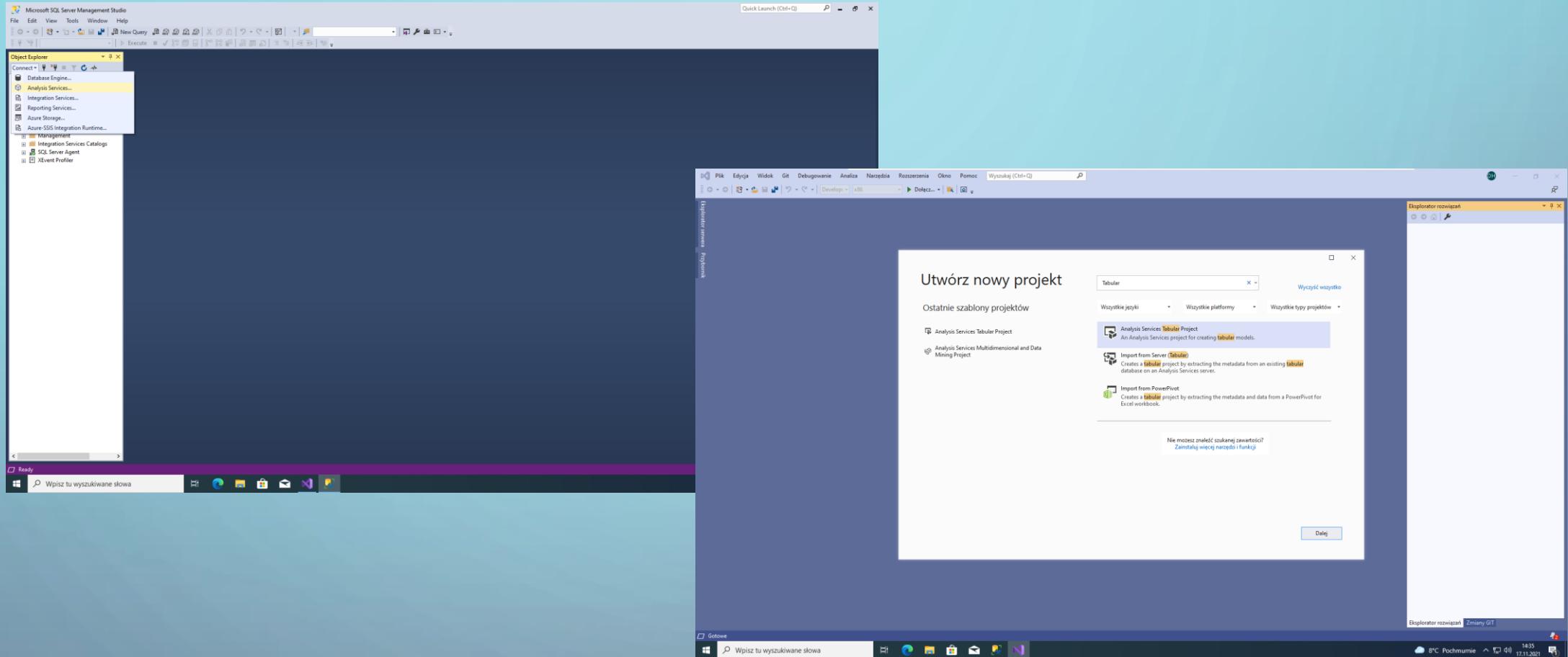
TROCHĘ PRAKTYKI

PROJEKT TABELARYCZNY

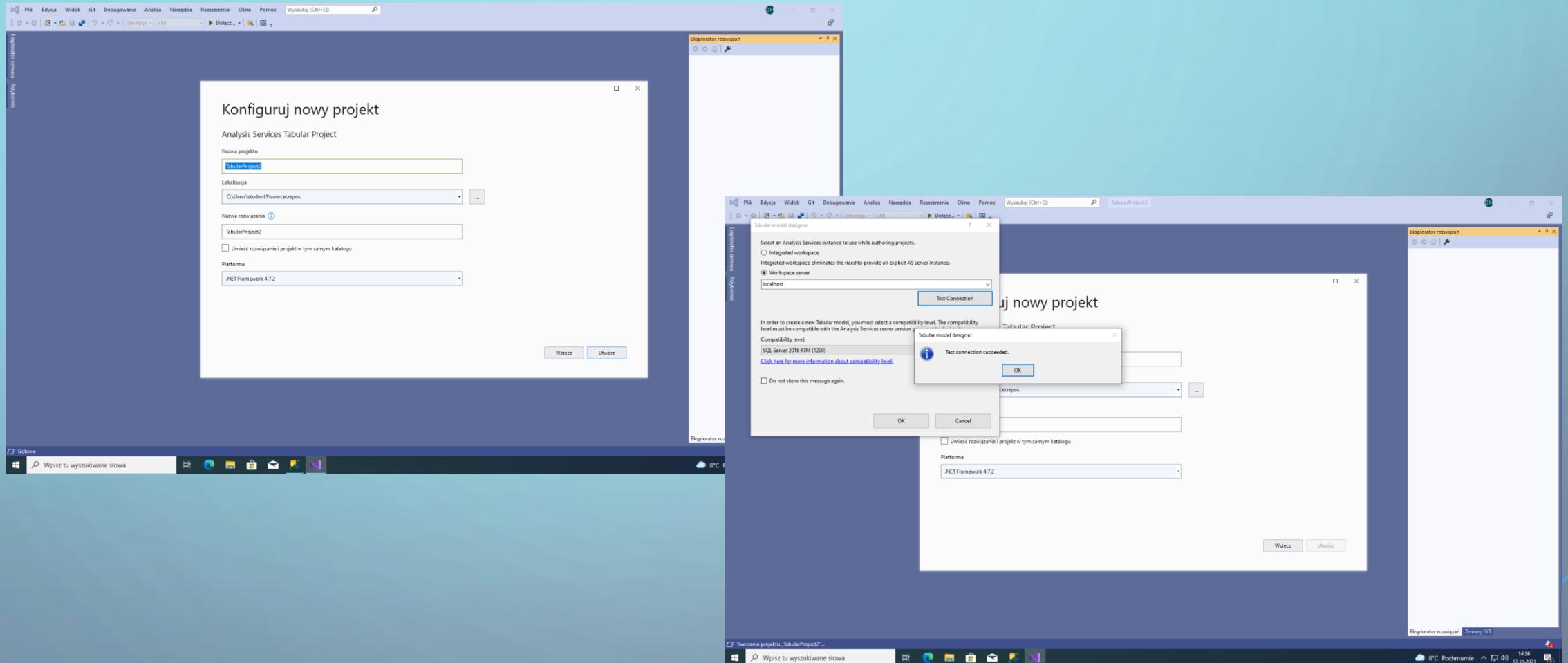
TROCHĘ PRAKTYKI



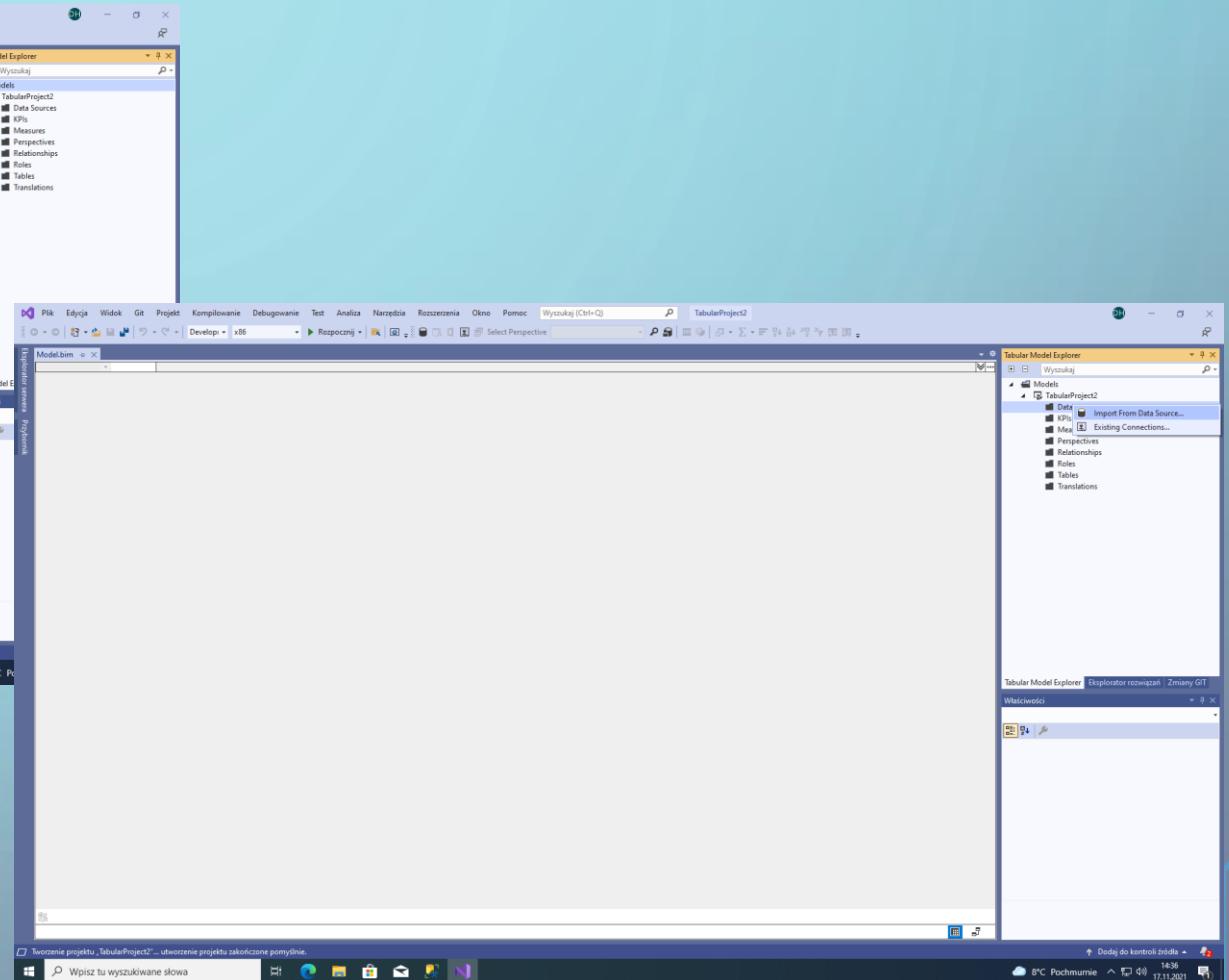
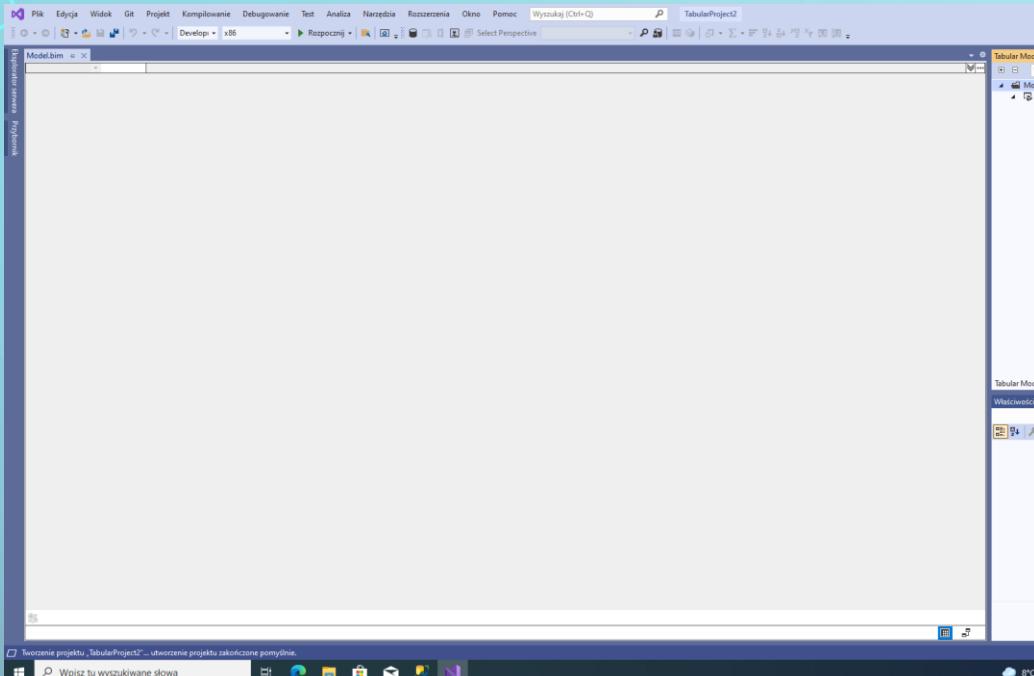
TROCHĘ PRAKTYKI



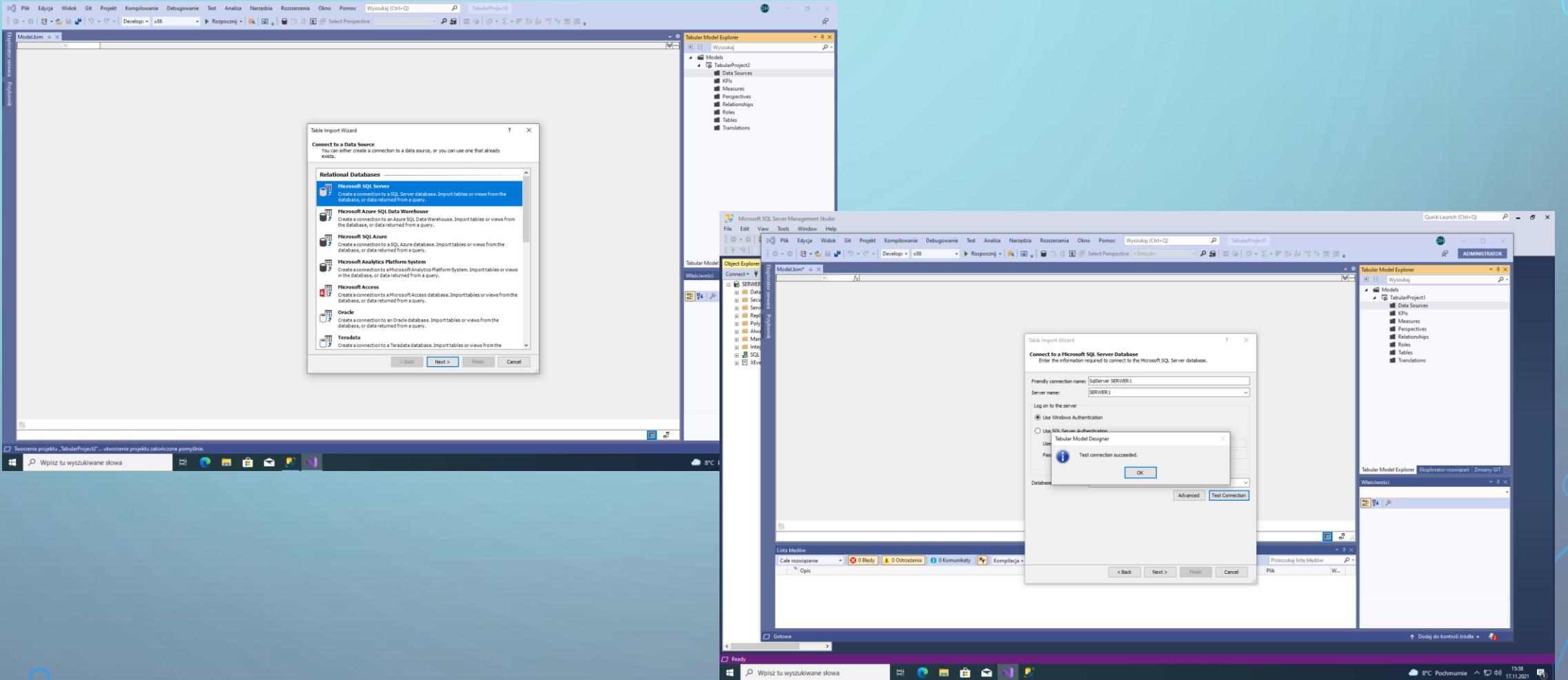
TROCHĘ PRAKTYKI



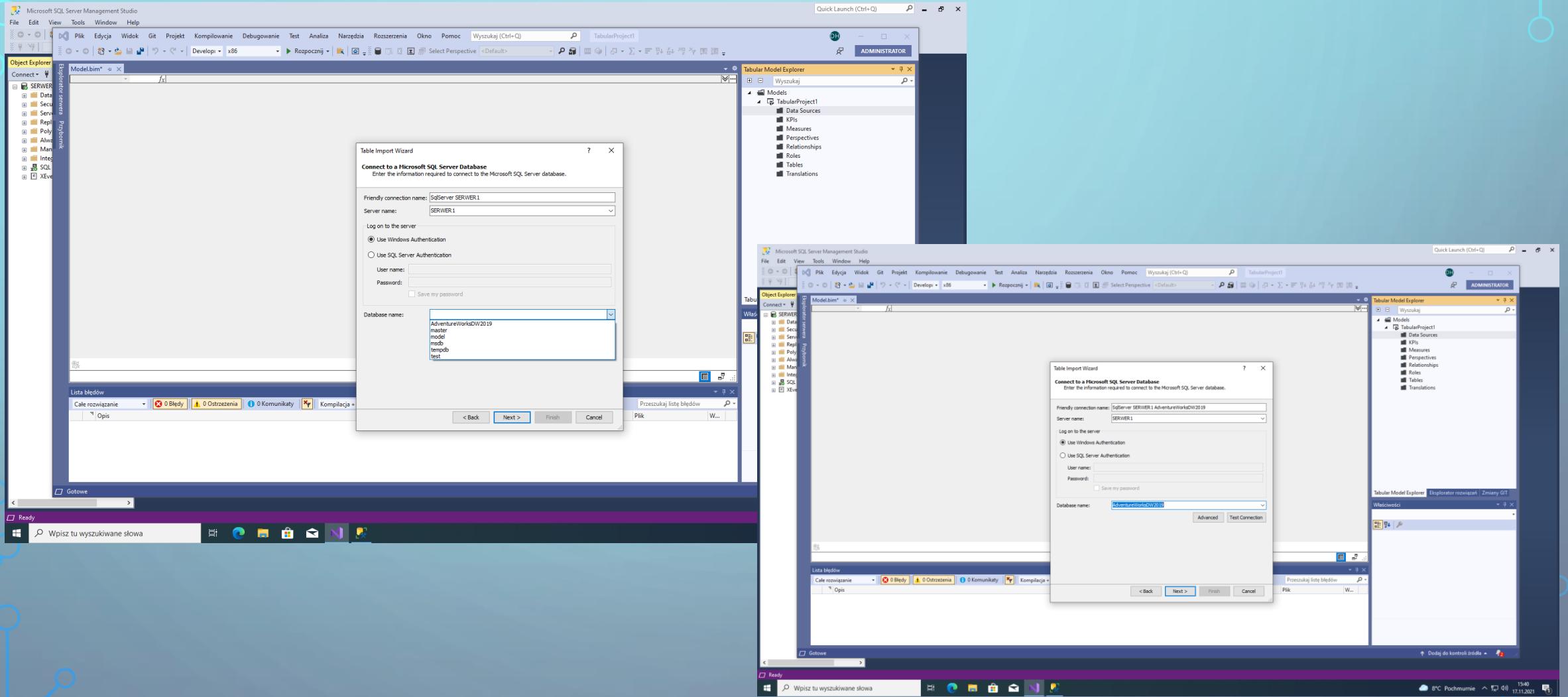
TROCHĘ PRAKTYKI



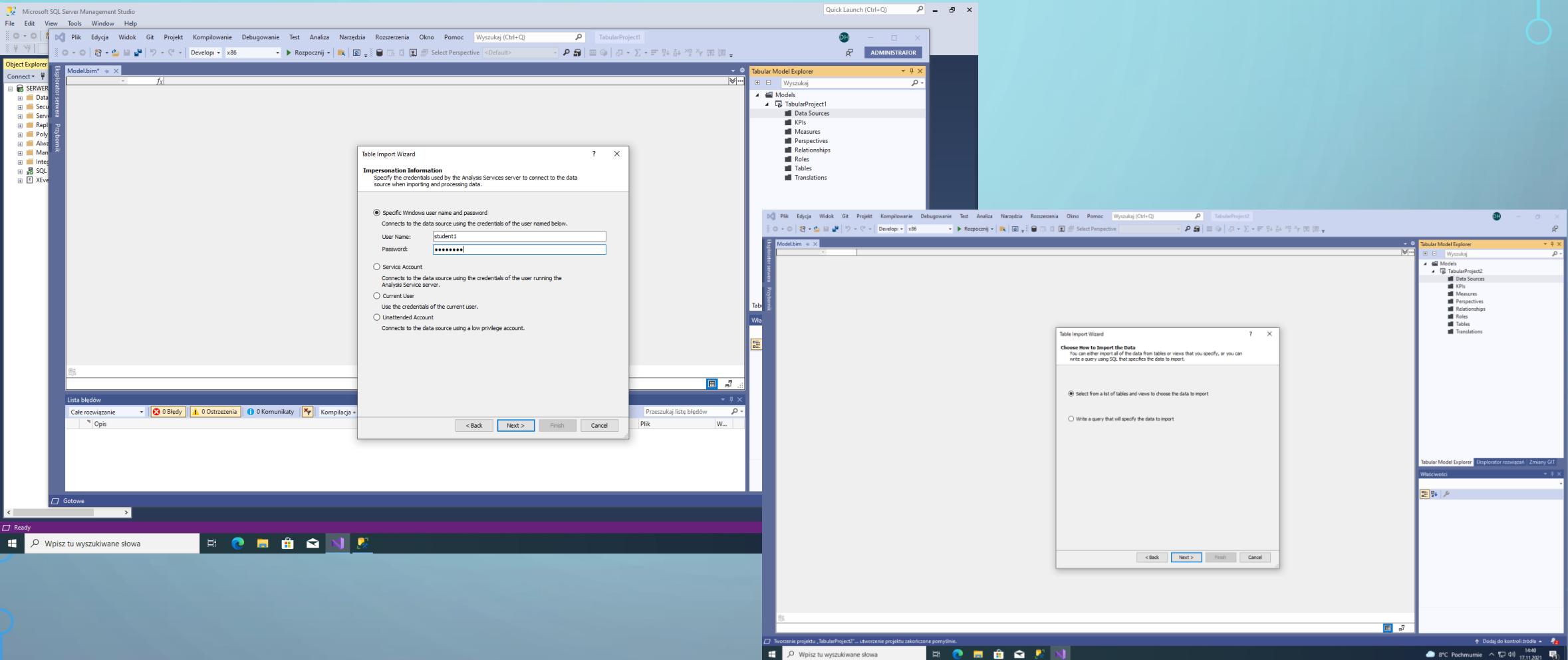
TROCHĘ PRAKTYKI



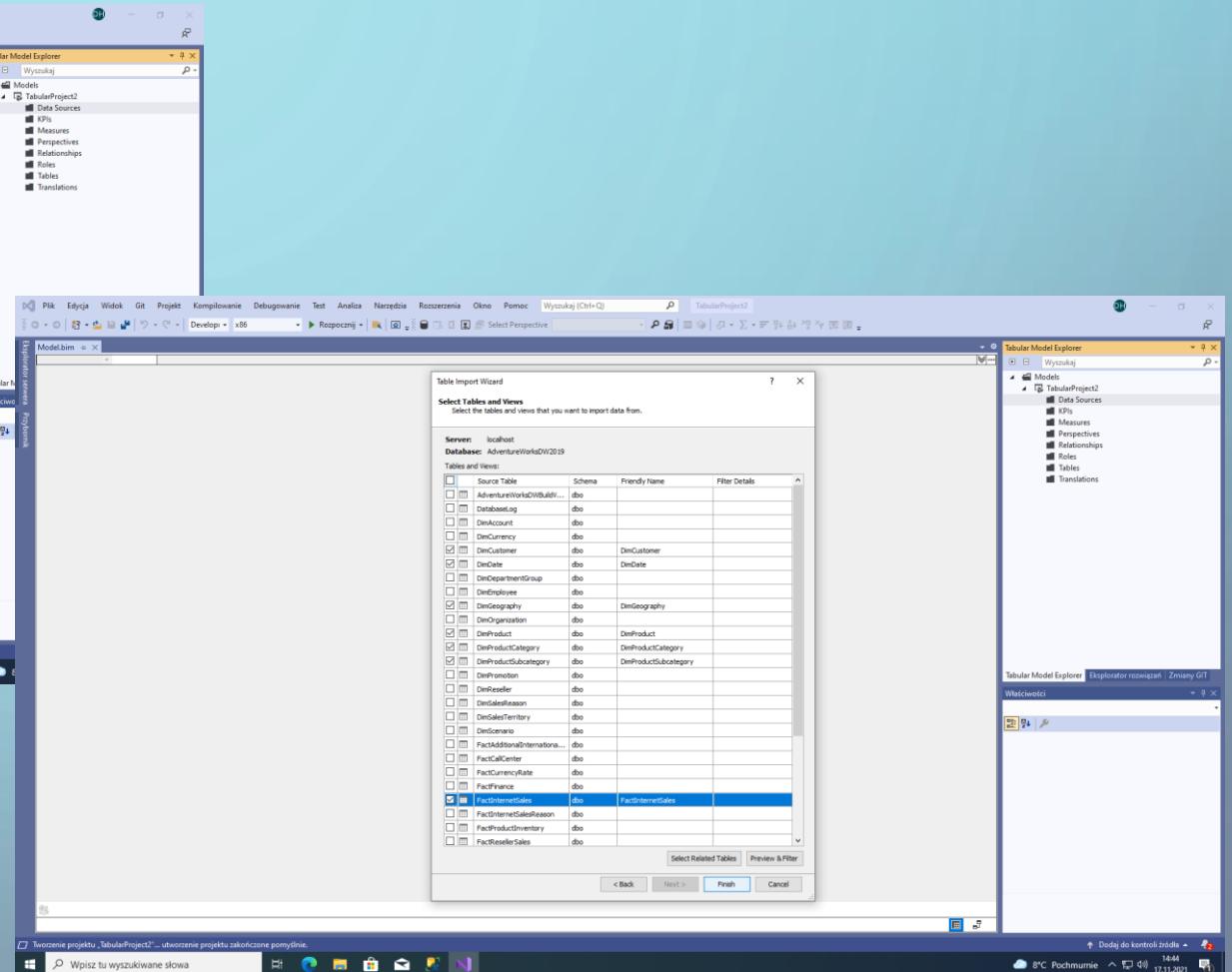
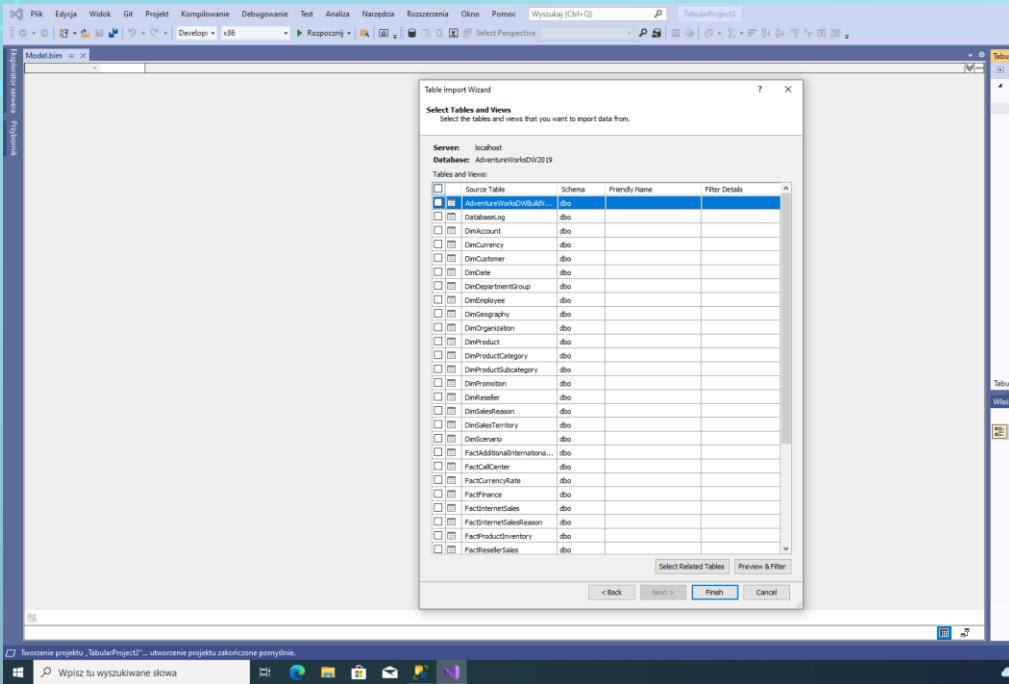
TROCHĘ PRAKTYKI



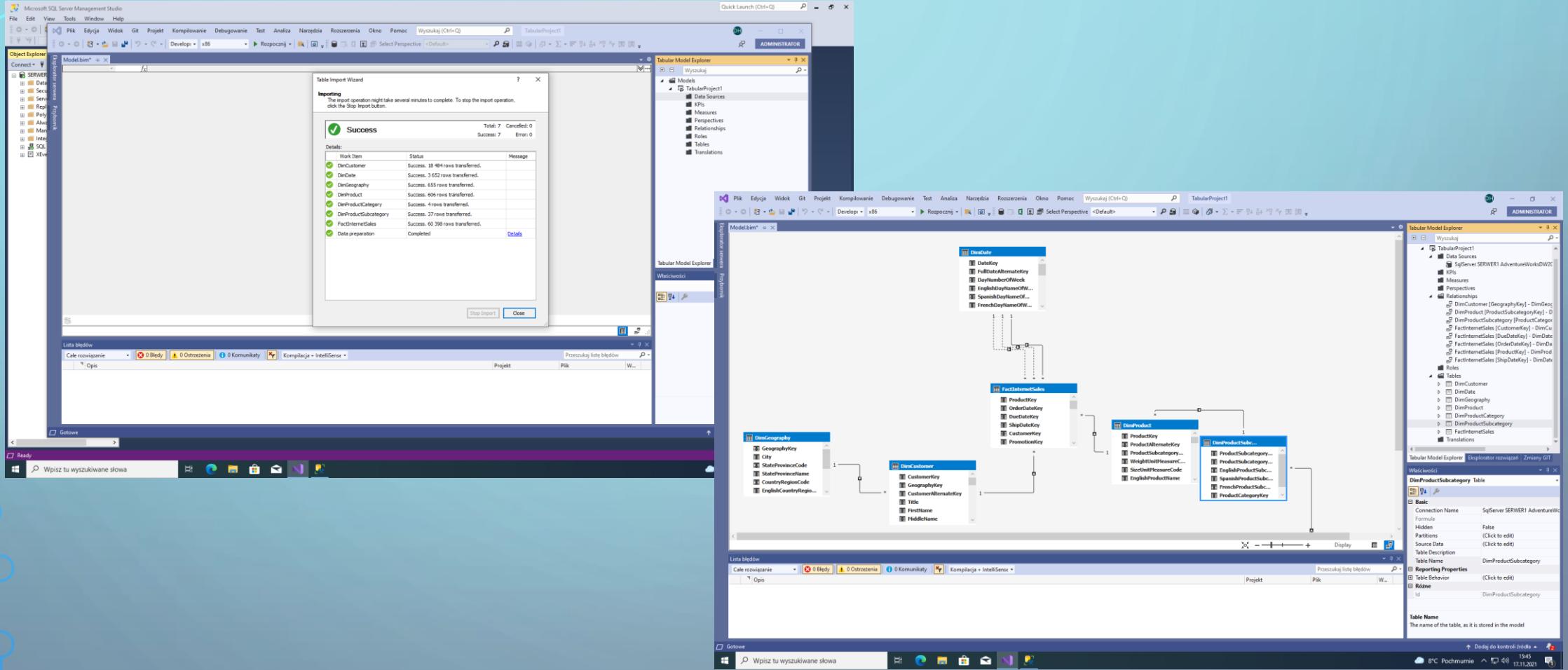
TROCHĘ PRAKTYKI



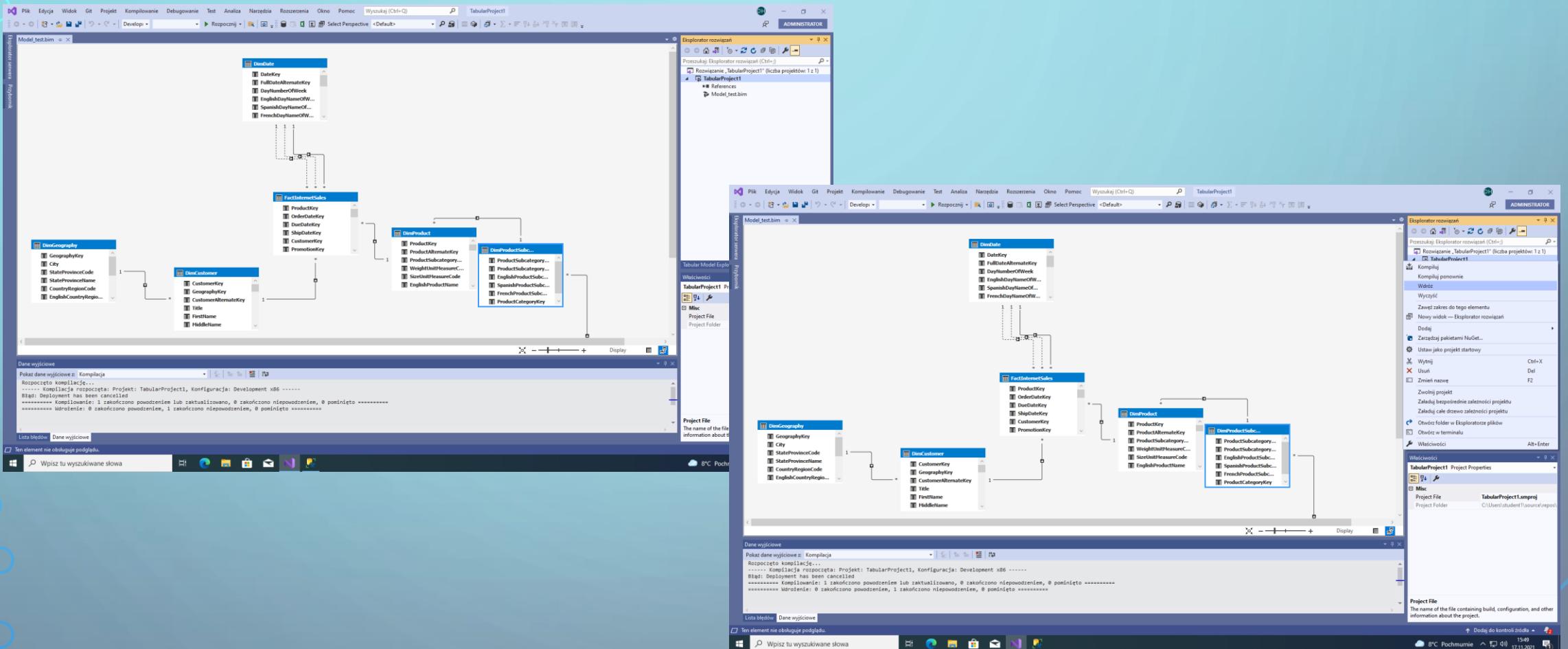
TROCHĘ PRAKTYKI



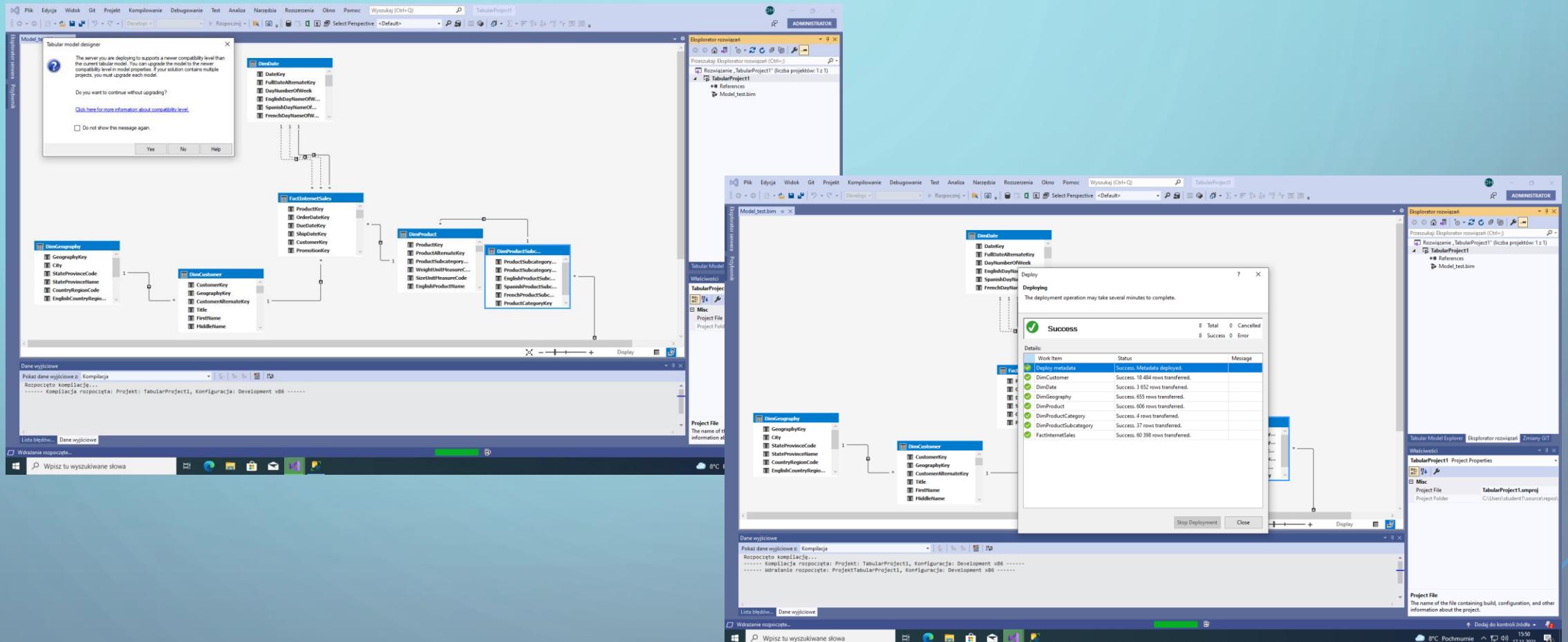
TROCHĘ PRAKTYKI



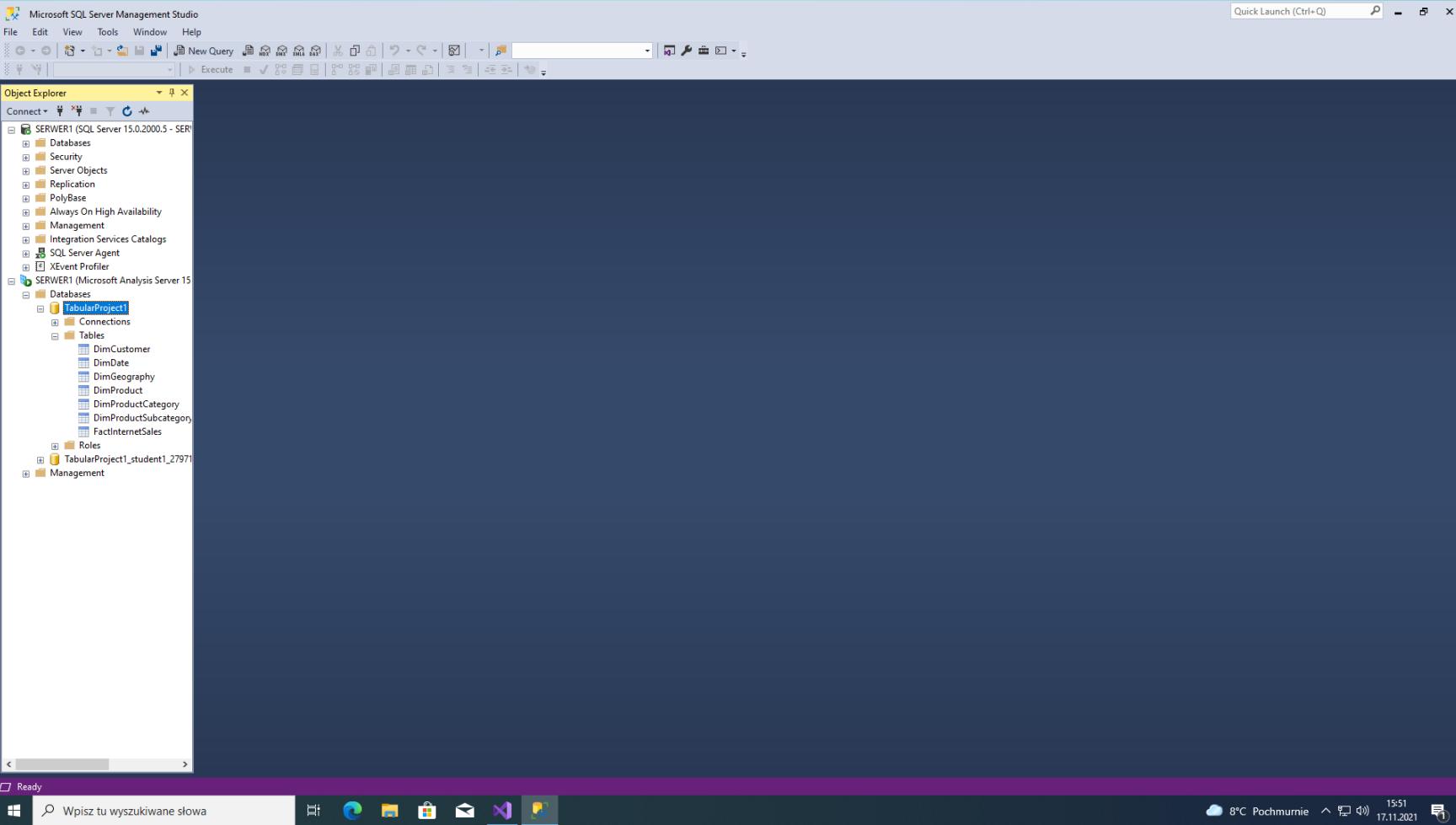
TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI



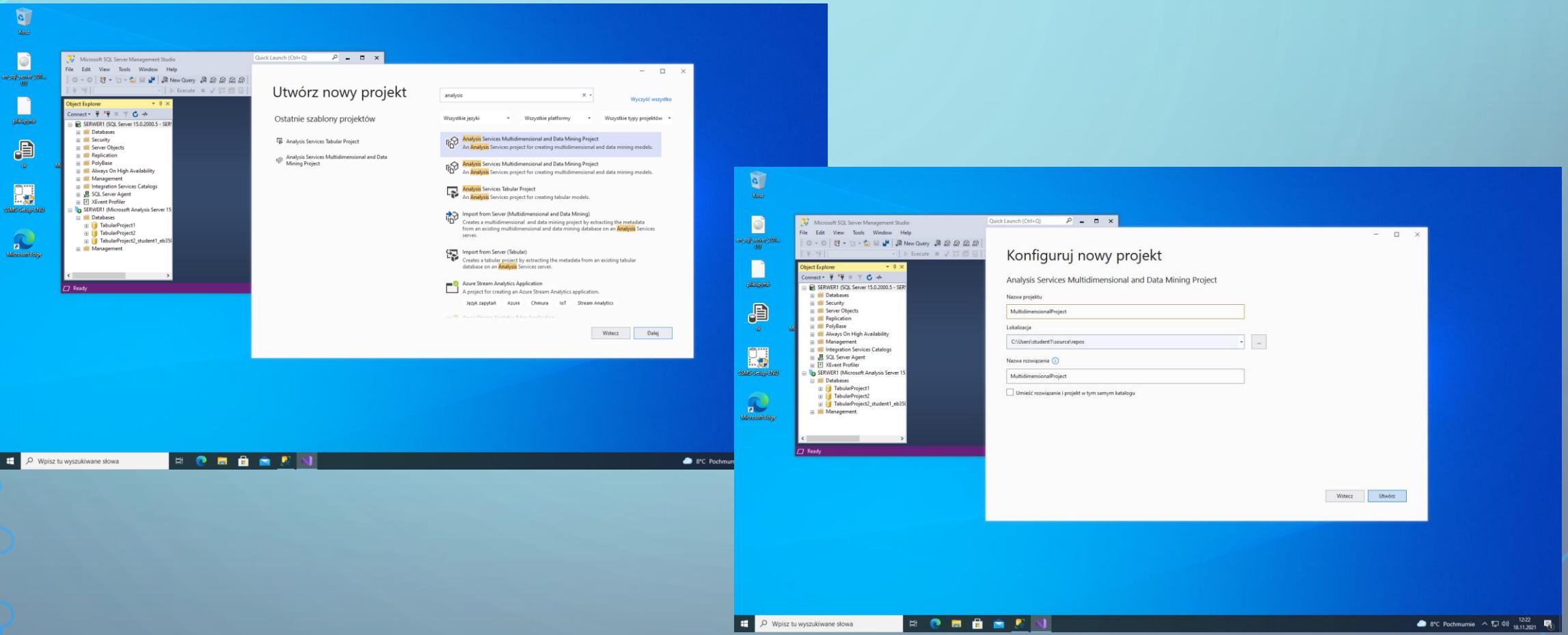
TROCHĘ PRAKTYKI



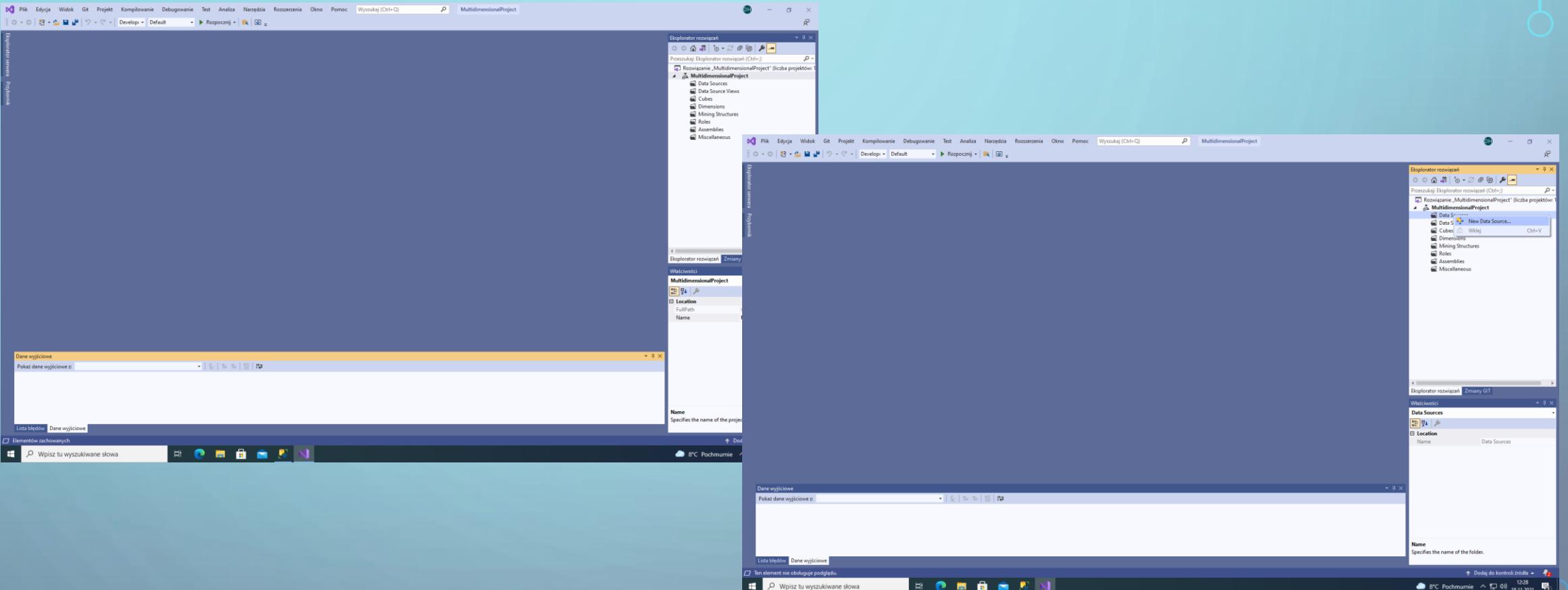
TROCHĘ PRAKTYKI

PROJEKT WIELOWYMIAROWY

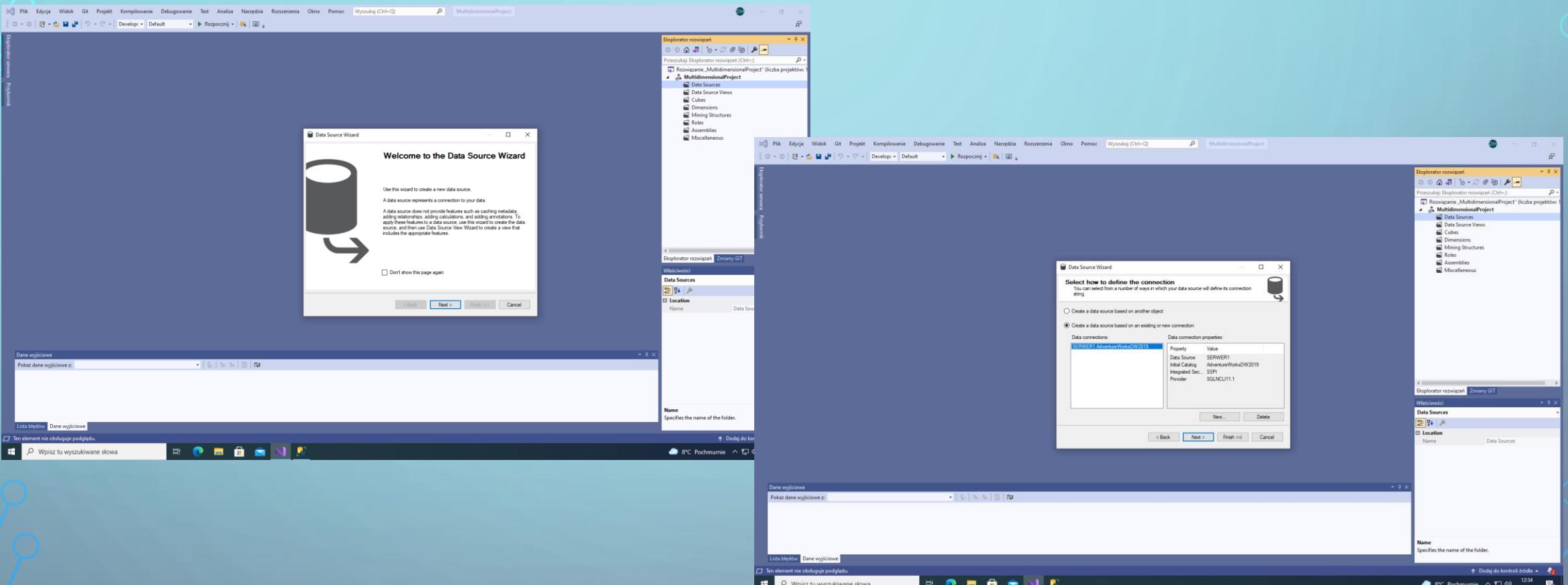
TROCHĘ PRAKTYKI



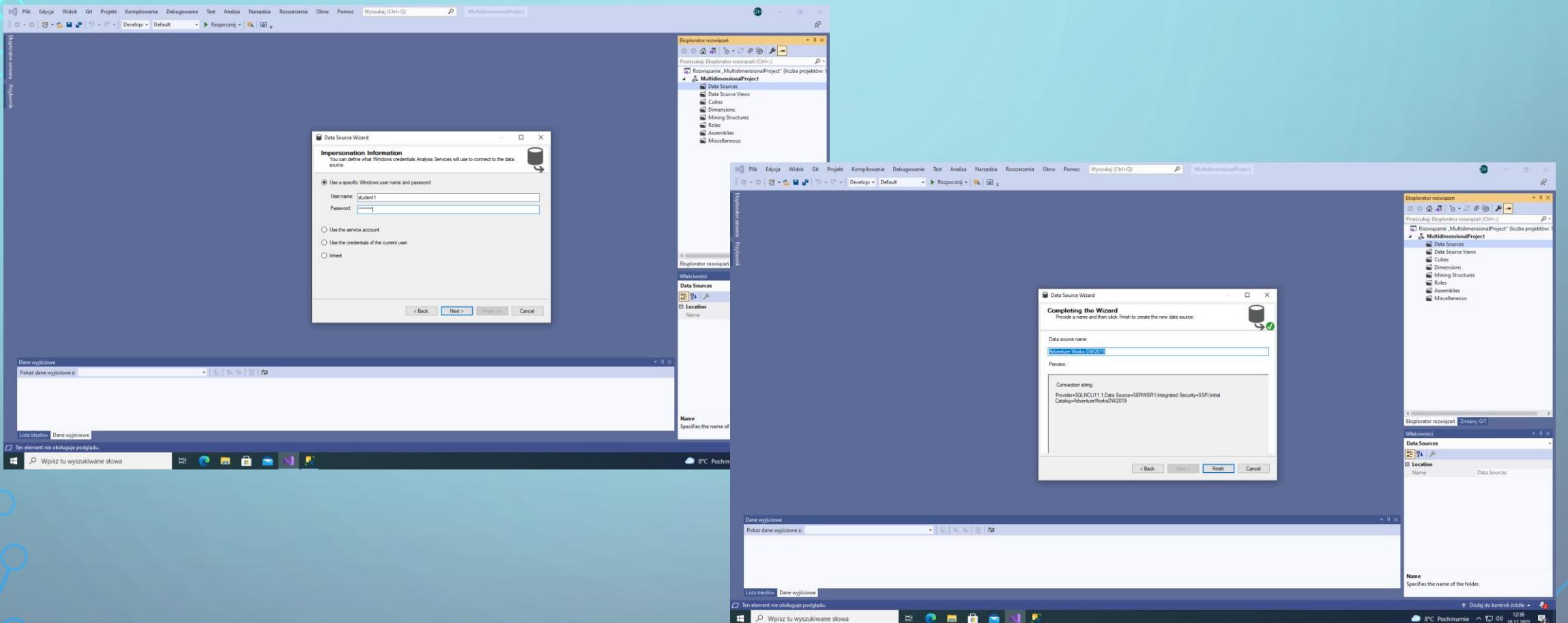
TROCHĘ PRAKTYKI



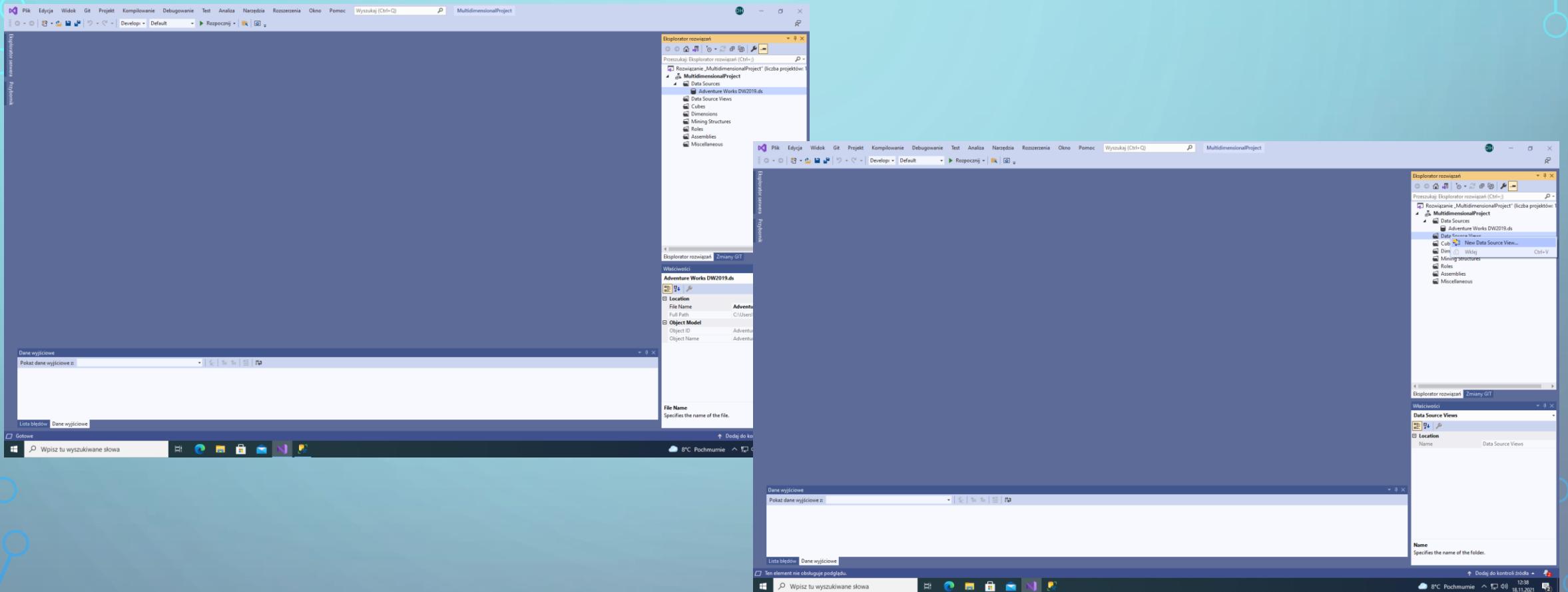
TROCHĘ PRAKTYKI



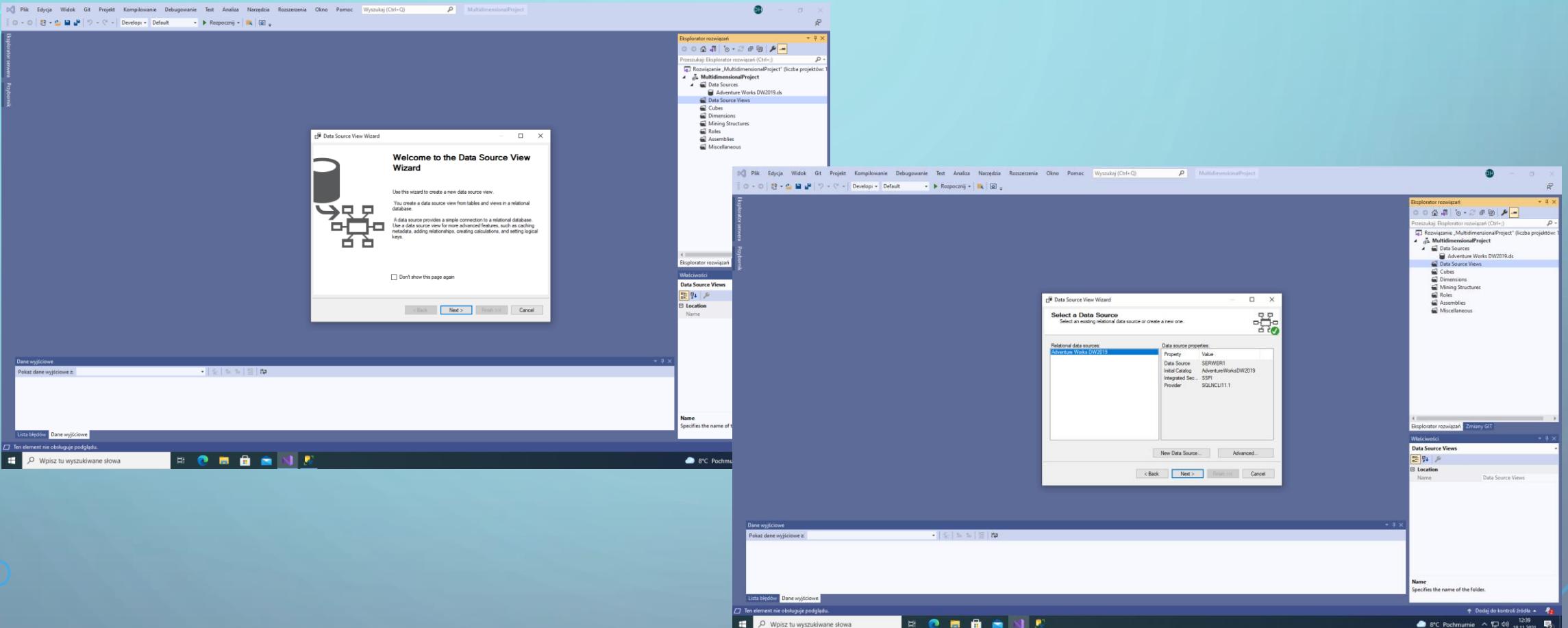
TROCHĘ PRAKTYKI



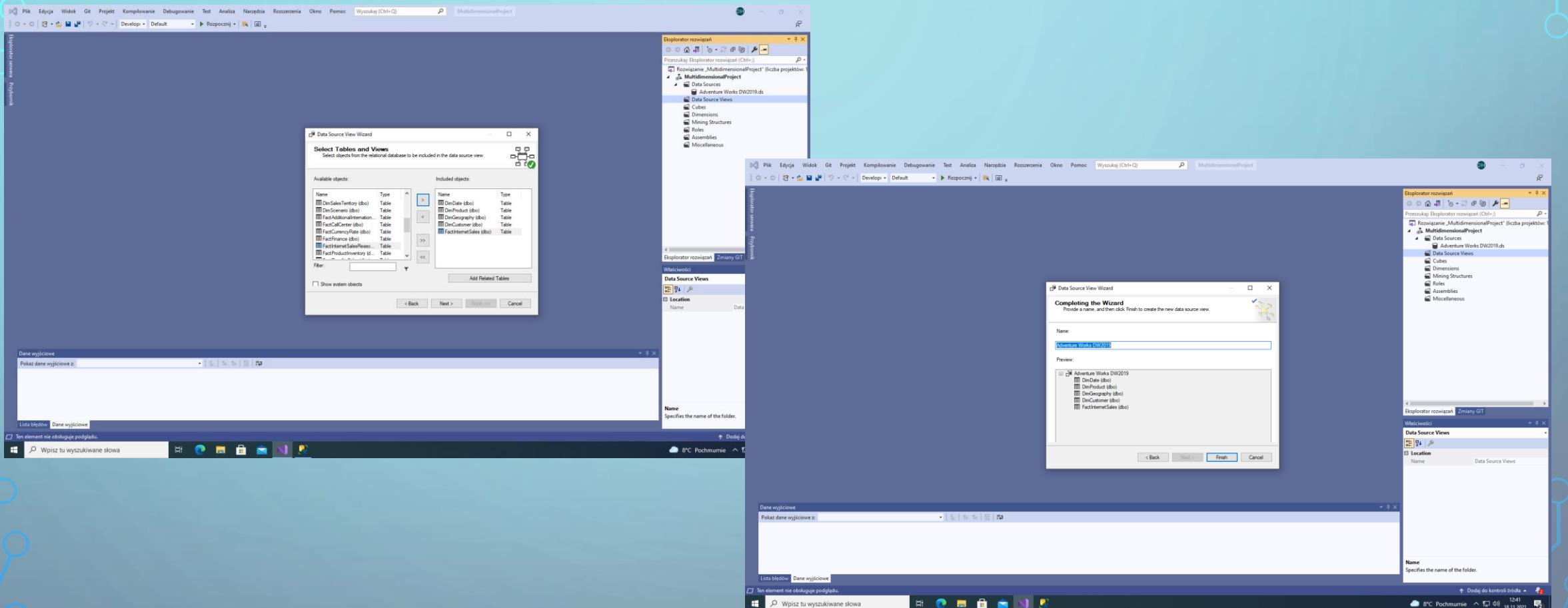
TROCHĘ PRAKTYKI



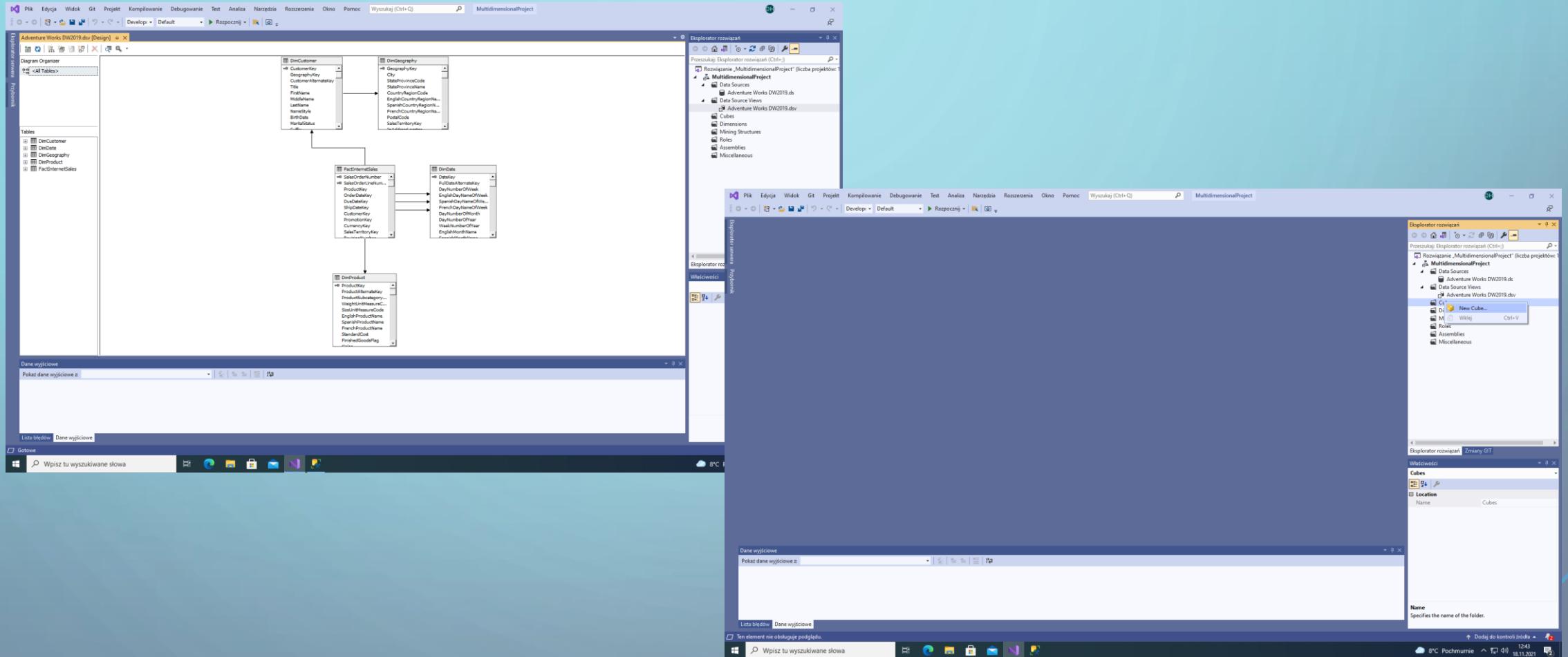
TROCHĘ PRAKTYKI



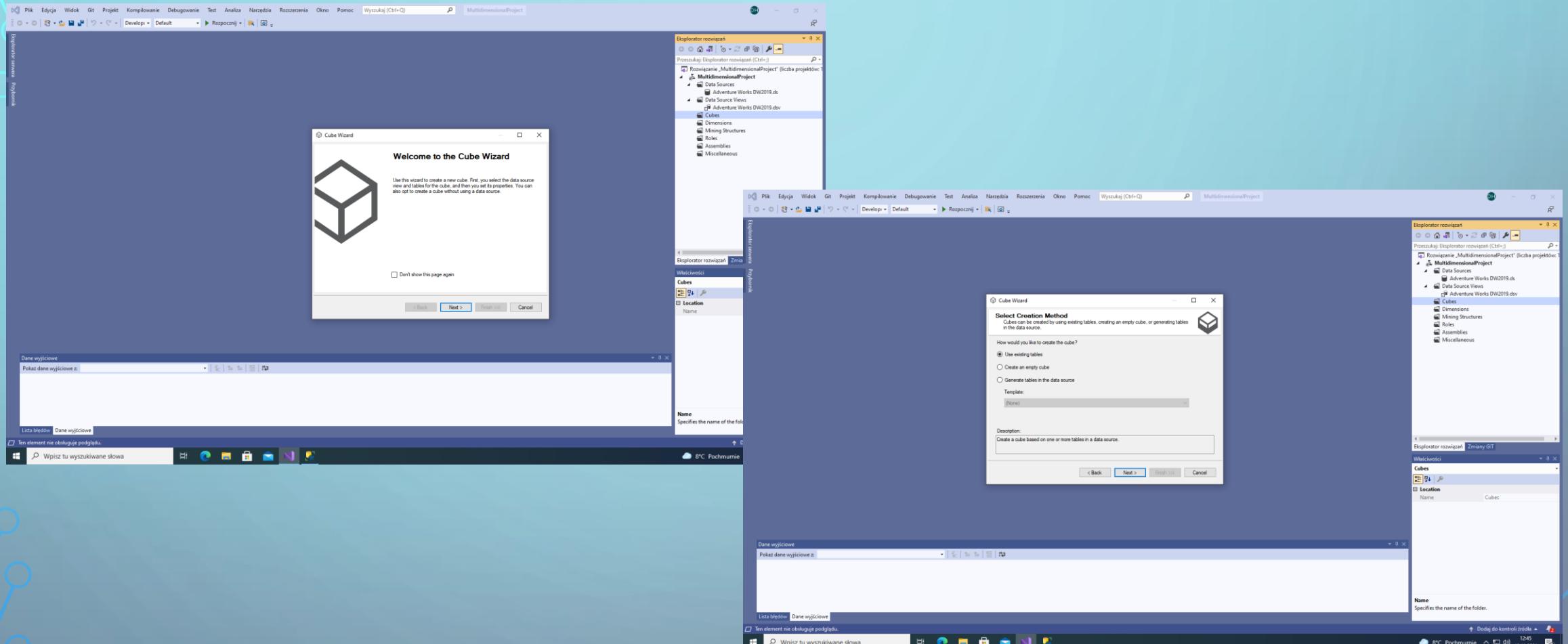
TROCHĘ PRAKTYKI



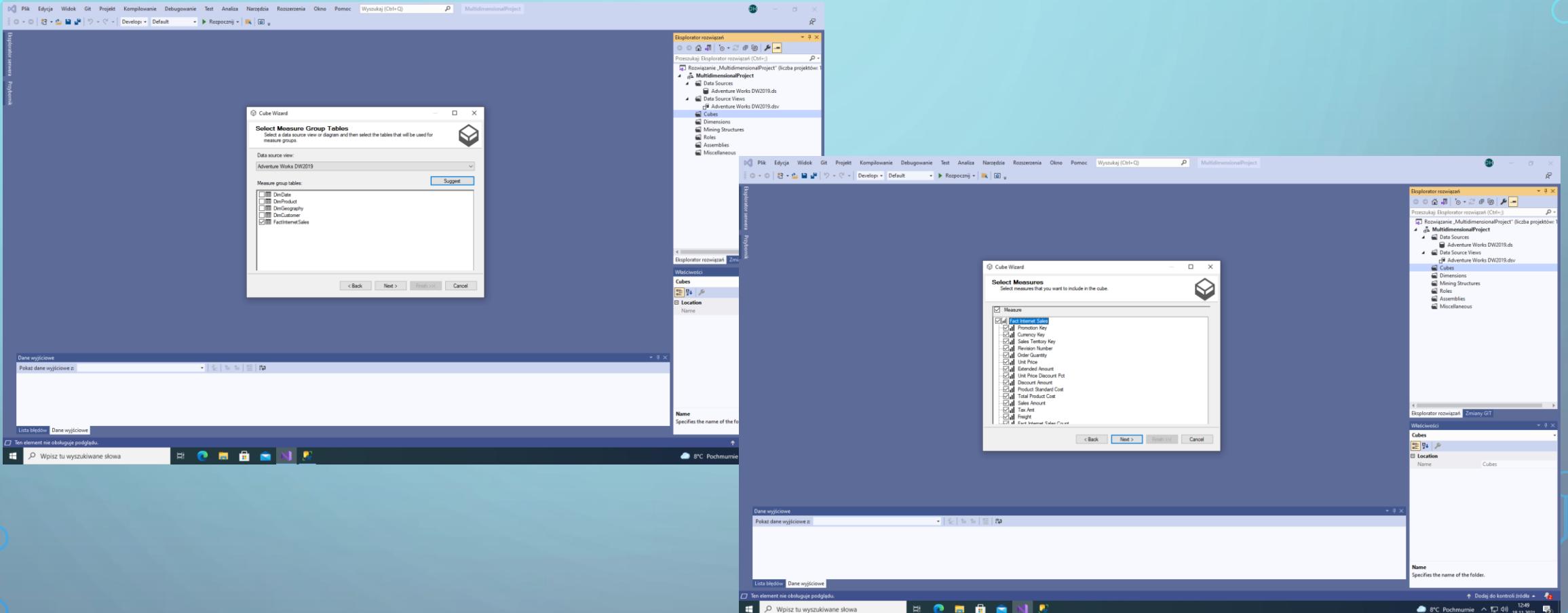
TROCHĘ PRAKTYKI



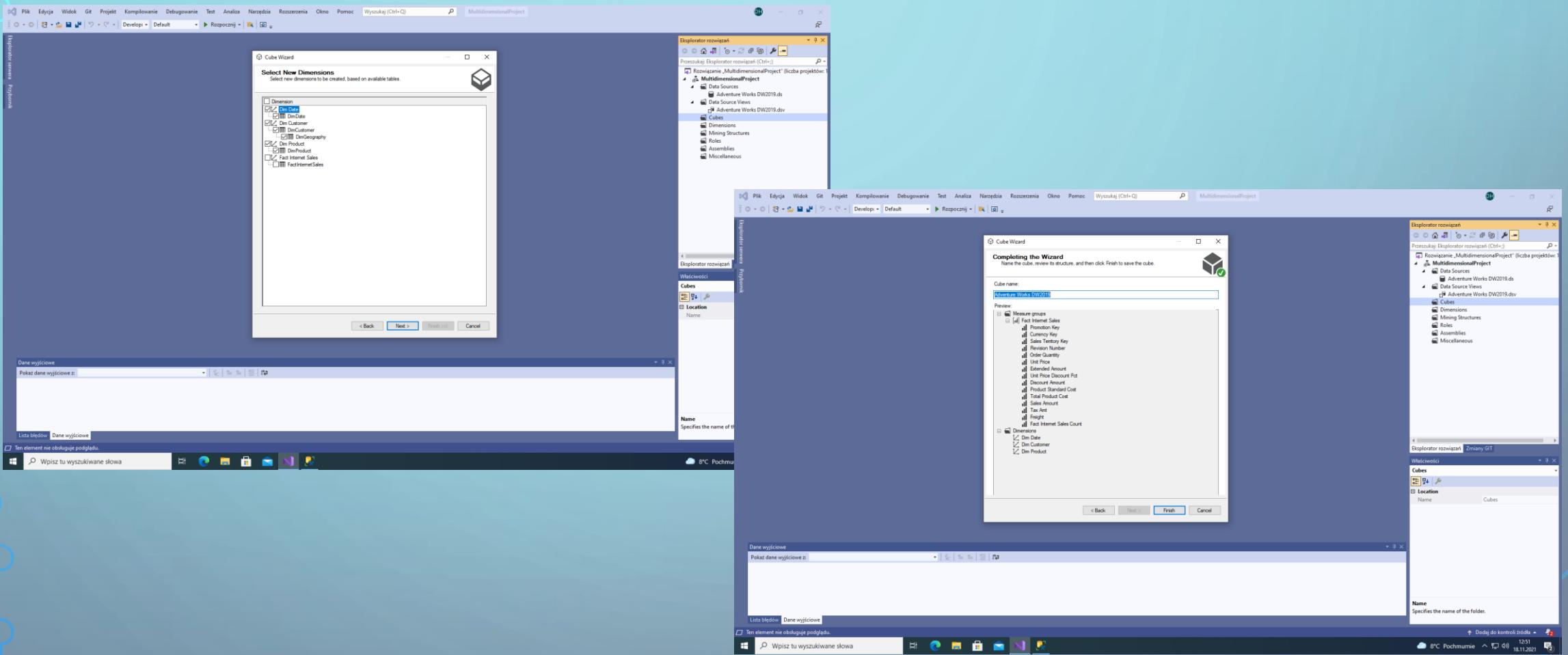
TROCHĘ PRAKTYKI



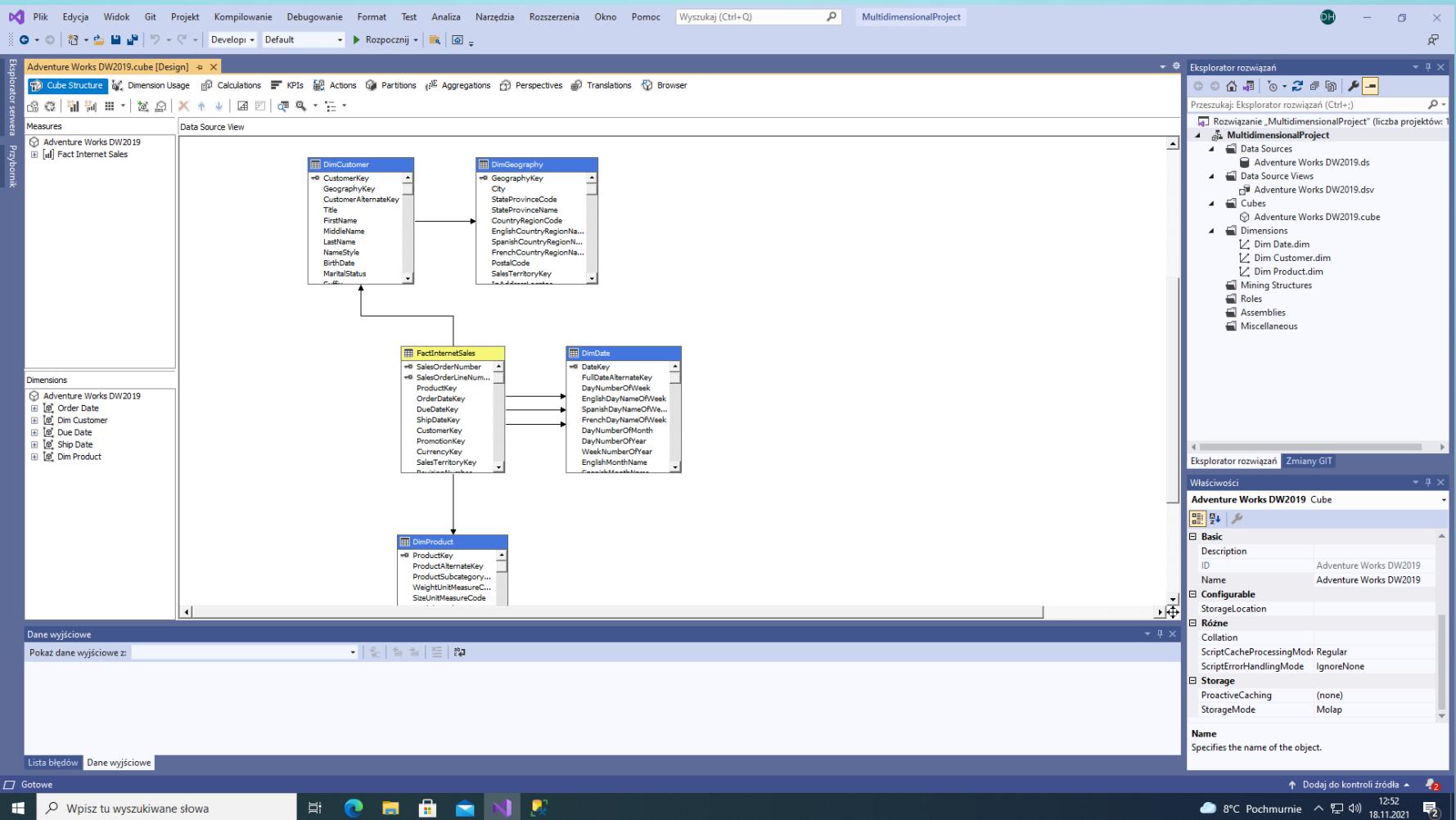
TROCHĘ PRAKTYKI



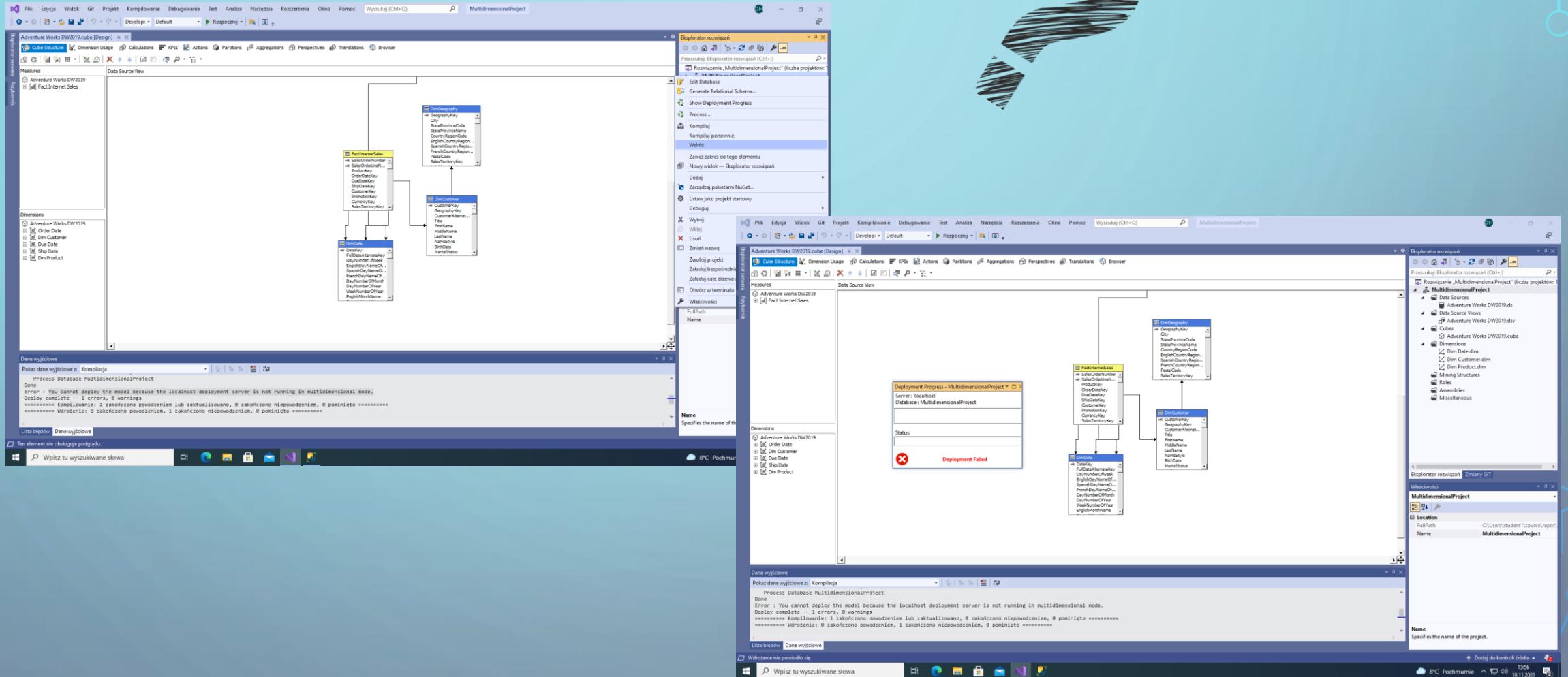
TROCHĘ PRAKTYKI



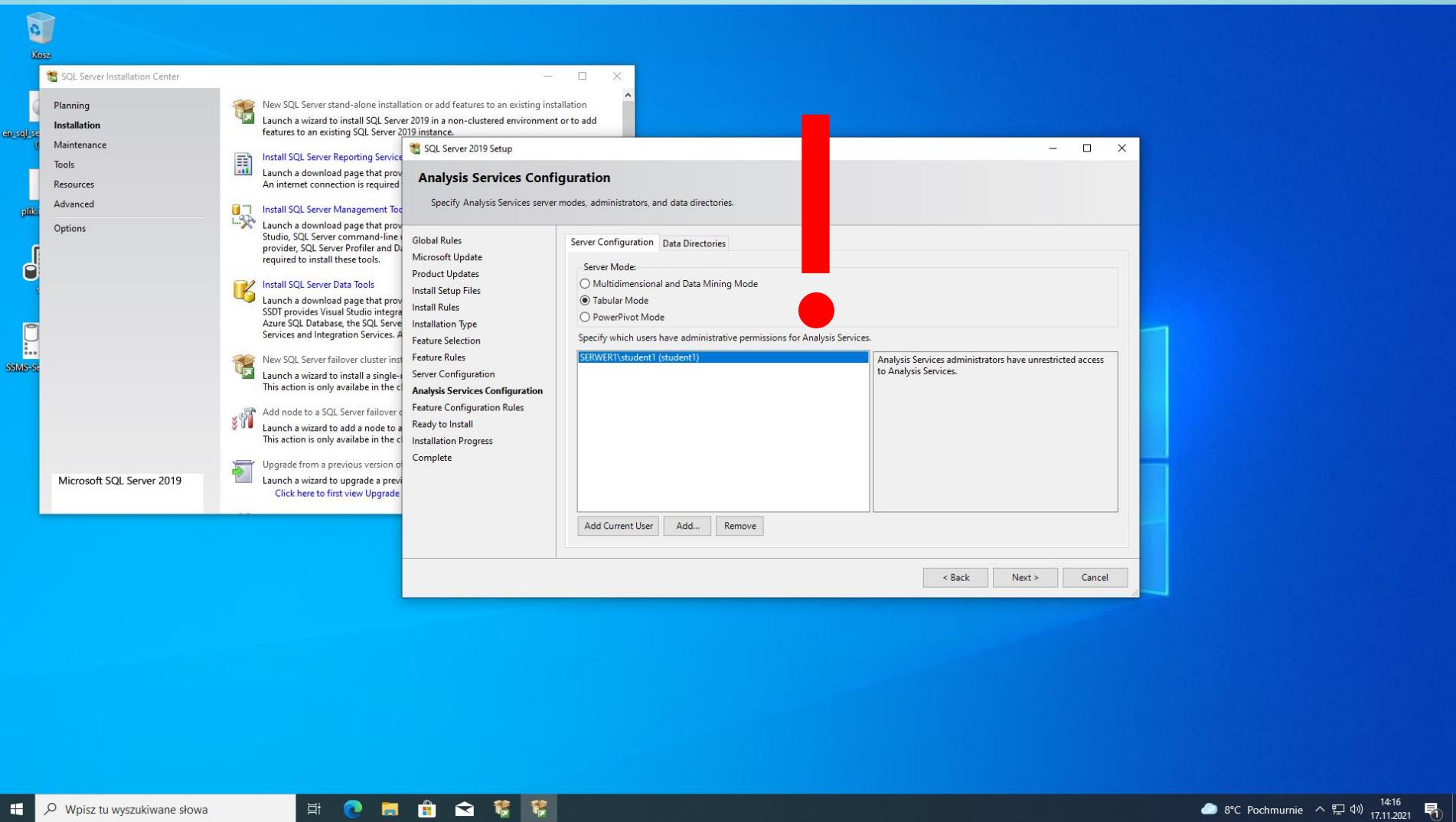
TROCHĘ PRAKTYKI



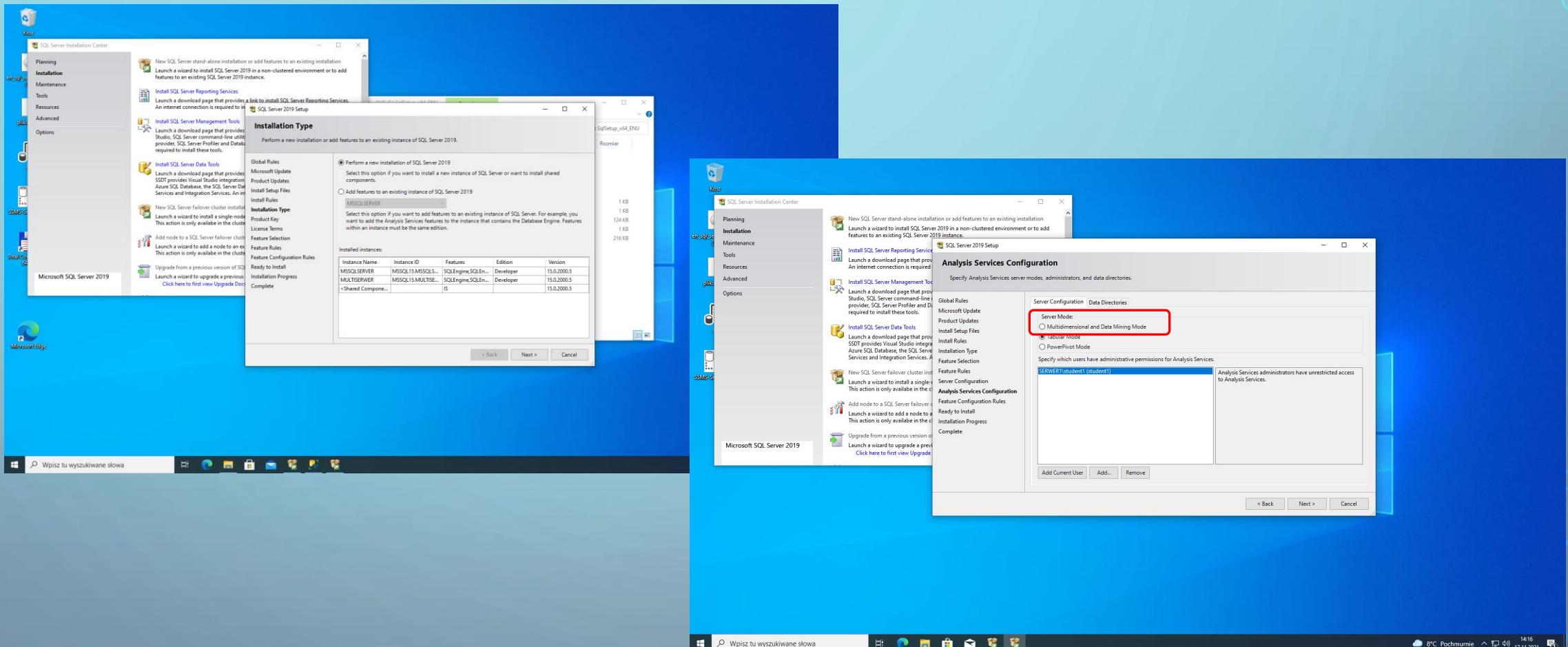
TROCHĘ PRAKTYKI



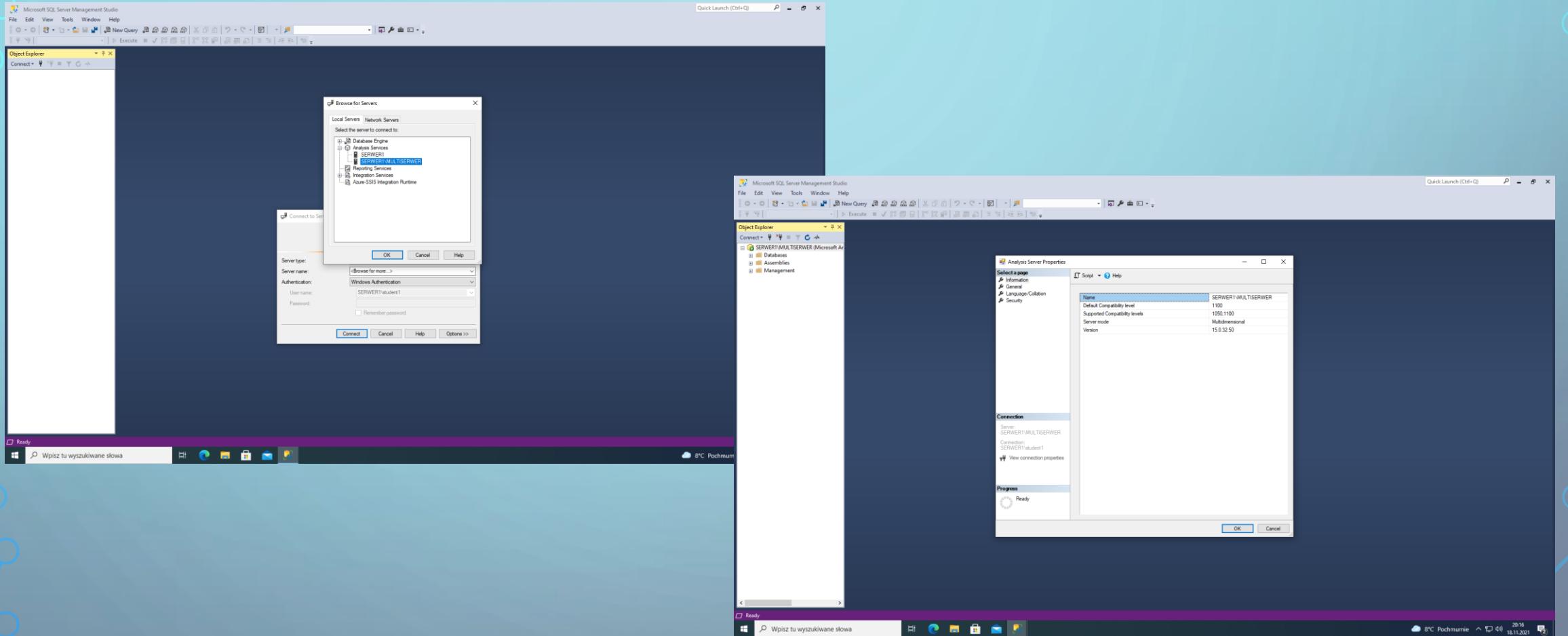
TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI



TROCHĘ PRAKTYKI

The screenshot shows the Microsoft SQL Server Management Studio (SSMS) interface with several windows open:

- Main Window:** Displays the "Adventure Works DW2019.cube [Design]" cube structure. It includes a "Measures" pane with "Fact Internet Sales", a "Dimensions" pane listing "Adventure Works DW2019" (Order Date, Dim Customer, Due Date, Ship Date, Dim Product), and a "Data Source View" pane showing relationships between dimensions like DimProductCategory, DimCustomer, DimGeography, and DimDate.
- Deployment Progress Window:** Shows the deployment status for "MultidimensionalProject". It lists completed tasks:
 - Processing Cube 'Adventure Works DW2019' completed.
 - Processing Measure Group 'Fact Internet Sales' completed.
 - Processing Dimension 'Dim Customer' completed.
 - Processing Dimension 'Dim Date' completed.
 - Processing Dimension 'Dim Product' completed.A green checkmark indicates "Deployment Completed Successfully".
- Object Explorer:** Shows the database structure under "SERWER1\MULTISERVER (Microsoft) \ MultidimensionalProject".
- MDXQuery1.mdx - SERWER1\MULTISERVER (SERWER1\student1)* - Microsoft Query:** A query window with the following content:

```
File Edit View Project Tools Window Help
...
MDXQuery1.mdx --SERWER1\student1*
Cube:
No cubes were found.
Metadata Functions
Search Model
Measure Group:
Error loading metadata: No cubes were found.
100 %
Messages
Executing the query ...
Query (1, 8).Parser: The following Run complete
100 %
Query completed w... SERWER1\MULTISERVER | SERWER1\student1 | 00:00:00
Ready
```
- Taskbar:** Shows the Windows taskbar with various pinned icons (File Explorer, Edge, Mail, etc.) and the SSMS icon.
- System Tray:** Shows the system tray with icons for battery level (8°C), weather (Pochmurnie), time (20:33), and date (18.11.2021).

TROCHĘ PRAKTYKI

The screenshot shows the Microsoft Analysis Services Management Studio interface. The main window displays the 'Adventure Works DW2019.cube [Design]' tab. On the left, the 'Adventure Works DW2019' dimension is selected in the 'Metadata' pane, showing its structure with measures like Fact Internet Sales, Product Key, and Sales Amount. The central area shows a table of sales data:

Customer Key	Product Key	Sales Amount
11000	214	34,99
11000	344	3399,99
11000	353	2319,99
11000	485	21,98
11000	488	53,99
11000	530	4,99
11000	541	28,99
11000	573	2384,07
11001	217	34,99
11001	225	8,99
11001	350	3374,99
11001	353	2319,99
11001	477	9,98
11001	478	9,99
11001	479	8,99
11001	485	21,98
11001	491	53,99
11001	604	539,99
11002	222	34,99
11002	346	3399,99
11002	359	2294,99
11002	561	2384,07
11003	225	8,99
11003	346	3399,99
11003	361	2294,99

The bottom status bar shows deployment progress:

```
Done
Sending deployment script to the server...
Done
Deploy complete -- 0 errors, 0 warnings
===== Komplikowanie: 1 zakonczeno powodzeniem lub zaktualizowano, 0 zakonczeno niepowodzeniem, 0 pominięto =====
===== Wdrożenie: 1 zakonczeno powodzeniem, 0 zakonczeno niepowodzeniem, 0 pominięto =====
```

The right side of the screen shows the 'Eksplorator rozwiązań' (Solution Explorer) and 'Właściwości' (Properties) panes.

DZIĘKUJĘ ZA UWAGĘ