



POLITECHNIKA RZESZOWSKA

im. Ignacego Łukasiewicza

WYDZIAŁ MATEMATYKI I FIZYKI STOSOWANEJ

**Projekt 3-3.5**

**WIELOWYMIAROWA ANALIZA DANYCH  
GLOBAL FASHION RETAIL SALES ANALYSIS**

Filip Kosiorowski, Andrii Kotovych,

173161, 173163

Inżynieria i Analiza Danych Rok 3

Rzeszów 2025

Spis treści	
Opis danych .....	4
Zbiór danych .....	4
Informacja .....	6
Tworzenie bazy .....	7
Tabela na błędy .....	9
Projekt w Visual Studio .....	9
Przesyłanie danych .....	9
Utworzenie bazy danych nr2/czystej bazy .....	18
Utworzenie procesu ETL do oczyszczenia danych .....	18
Tworzenie procesów ETL .....	25
Time & Sales (etl_TimeAndSales.dtsx) .....	25
Employee Performance (etl_EmployeePerformance.dtsx) .....	30
Store Sales (etl_StoreSales.dtsx) .....	32
Discount Effectiveness (DiscountEffectiveness.dtsx) .....	33
TxByHour (etl_TxByHour.dtsx) .....	36
Return Rate (etl_ReturnRate.dtsx) .....	40
FactAvgBasketByCity (etl_FactAvgBasketByCity.dtsx) .....	44
Wizualizacja .....	49
Canva nr 1 Trend i Sezonowość sprzedaży .....	49
Wykres nr 1 .....	49
Wykres nr 2 .....	54
Wykres nr 3 .....	55
Wykres nr 4 .....	56
Canva nr 2 Wydajność pracowników .....	57
Wykre nr 1 .....	57
Wykres nr 2 .....	57
Wykres nr 3 .....	58
Wykres nr 4 .....	59
Canva nr 3 Statystyki sklepów i krajów .....	60
Wykres nr 1 .....	60
Wykres nr 2 .....	61
Wykres nr 3 .....	62
Wykres nr 4 .....	63
Canva nr 4 Analiza klientów według wieku i płci .....	64
Wykres nr 1 .....	64
Wykres nr 2 .....	65
Wykres nr 3 .....	65

Wykres nr 4.....	66
Podsumowanie.....	66

# Opis danych

Global Fashion Retail Stores Dataset to **syntetyczny**, dwuletni zbiór danych symulujący operacje międzynarodowej sieci odzieżowej. Zawiera:

4+ miliony rekordów sprzedaży

35 sklepów z 7 krajów z dokładną lokalizacją tych sklepów.

# Zbiór danych

W pobranym pakiecie danych można znaleźć 6 Tabel

- transactions.csv – każdy wiersz to jedna pozycja na paragonie: Invoice ID, Line, Customer ID, Product ID, Size, Unit Price, Quantity, Date, Discount, Line Total, Store ID, Employee ID, Currency, SKU, Transaction Type, Payment Method, Invoice Total.

```
Invoice ID,Line,Customer ID,Product ID,Size,Color,Unit Price,Quantity,Date,Discount,Line Total,Store ID,Employee ID,Currency,Currency Symbol,SKU,Transaction Type,Payment Method,Invoice Total
INV-US-001-03558761,1,47162,485,M,,80,5,1,2023-01-01 15:42:00,0.0,80.5,1,7,USD,$,MASU485-M-,Sale,Cash,126.7
INV-US-001-03558761,2,47162,2779,G,,31.5,1,2023-01-01 15:42:00,0.4,18.9,1,7,USD,$,CHC02779-G-,Sale,Cash,126.7
INV-US-001-03558761,3,47162,64,M,NEUTRAL,45.5,1,2023-01-01 15:42:00,0.4,27.3,1,7,USD,$,MAC064-M-NEUTRAL,Sale,Cash,126.7
INV-US-001-03558762,1,10142,131,M,BLUE,70.0,1,2023-01-01 20:04:00,0.4,42.0,1,6,USD,$,FEC0131-M-BLUE,Sale,Cash,77.0
INV-US-001-03558762,2,10142,716,L,WHITE,26.0,1,2023-01-01 20:04:00,0.0,26.0,1,6,USD,$,MAT-716-L-WHITE,Sale,Cash,77.0
INV-US-001-03558762,3,10142,644,,9,0,1,2023-01-01 20:04:00,0.0,9.0,1,6,USD,$,MAAC644--.,Sale,Cash,77.0
INV-US-001-03558763,1,8286,1395,M,,95.5,1,2023-01-01 20:47:00,0.0,95.5,1,6,USD,$,MASU1395-M-,Sale,Cash,95.5
INV-US-001-03558764,1,1968,1208,M,SILVER,75.0,1,2023-01-01 08:06:00,0.4,45.0,1,9,USD,$,MAC01208-M-SILVER,Sale,Credit Card,72.3
INV-US-001-03558764,2,1968,2639,M,WHITE,45.5,1,2023-01-01 08:06:00,0.4,27.3,1,9,USD,$,MASW2639-M-WHITE,Sale,Credit Card,72.3
INV-US-001-03558765,1,26536,1822,M,,32.0,1,2023-01-01 17:25:00,0.4,19.2,1,12,USD,$,FESW1822-M-,Sale,Credit Card,19.2
INV-US-001-03558766,1,18139,2147,M,,16.0,1,2023-01-01 10:57:00,0.0,16.0,1,12,USD,$,MASH2147-M-,Sale,Credit Card,16.0
INV-US-001-03558767,1,14906,2924,XXL,NEUTRAL,54.0,1,2023-01-01 20:54:00,0.4,32.4,1,13,USD,$,MAC02924-XXL-NEUTRAL,Sale,Credit Card,32.4
INV-US-001-03558768,1,64777,951,M,,17.0,1,2023-01-01 19:08:00,0.0,17.0,1,7,USD,$,MASH951-M-,Sale,Credit Card,17.0
INV-US-001-03558769,1,47155,1770,M,,34.0,1,2023-01-01 08:26:00,0.4,20.4,1,13,USD,$,FESW1770-M-,Sale,Credit Card,20.4
INV-US-001-03558770,1,840,1884,XL,,58.0,1,2023-01-01 16:09:00,0.4,34.8,1,11,USD,$,MAC01884-XL-,Sale,Credit Card,34.8
RET-US-001-03558764,1,1968,1208,M,SILVER,75.0,1,2023-01-01 00:00:00,0.0,-45.0,1,9,USD,$,MAC01208-M-SILVER,Return,Credit Card,-72.3
RET-US-001-03558764,2,1968,2639,M,WHITE,45.5,1,2023-01-01 00:00:00,0.0,-27.3,1,9,USD,$,MASW2639-M-WHITE,Return,Credit Card,-72.3
INV-US-001-03558771,1,34586,636,M,,70.0,1,2023-01-01 11:12:00,0.4,42.0,1,6,USD,$,MAC0636-M-,Sale,Credit Card,42.0
INV-US-001-03558772,1,2597,235,M,GREEN,35.0,1,2023-01-01 20:27:00,0.4,21.0,1,13,USD,$,FEC0235-M-GREEN,Sale,Cash,21.0
INV-US-001-03558773,1,62769,2768,L,RED,61.0,1,2023-01-01 19:37:00,0.4,36.6,1,12,USD,$,MAC02768-L-RED,Sale,Cash,122.6
INV-US-001-03558773,2,62769,264,L,,39.0,1,2023-01-01 19:37:00,0.0,39.0,1,12,USD,$,FESH264-L-,Sale,Cash,122.6
INV-US-001-03558774,1,13166,105,M,YELLOW,45.0,3,2023-01-01 14:43:00,0.4,81.0,1,9,USD,$,FEC0105-M-YELLOW,Sale,Cash,81.0
INV-US-001-03558775,1,77051,1431,M,WHITE,70.5,1,2023-01-01 19:59:00,0.4,42.3,1,13,USD,$,FEC01431-M-WHITE,Sale,Credit Card,42.3
INV-US-001-03558776,1,109863,2929,L,,87.0,1,2023-01-01 20:05:00,0.0,87.0,1,6,USD,$,MASU2929-L-,Sale,Cash,87.0
INV-US-001-03558777,1,58395,2135,M,BLUE,32.0,1,2023-01-01 17:59:00,0.0,32.0,1,8,USD,$,FEDR2135-M-BLUE,Sale,Credit Card,32.0
```

- customers.csv – dane demograficzne klientów: Customer ID, Name, Email, Telephone, City, Country, Gender, Date Of Birth, Job Title.

```
Customer ID,Name,Email,Telephone,City,Country,Gender,Date Of Birth,Job Title
1,Tyler Garcia,tyler.garcia@fake_gmail.com,922.970.2265x47563,New York,United States,M,2003-07-15,
2,Joshua Miller,joshua.miller@fake_gmail.com,+1-958-729-6169,New York,United States,M,2000-06-16,Records manager
3,Alison Marshall DDS,alison.marshall.dds@fake_hotmail.com,+1-645-567-0876x5409,New York,United States,F,2003-07-22,
4,Jeffery Acosta,jeffery.acosta@fake_yahoo.com,212.336.0912x84994,New York,United States,M,1996-11-12,Proofreader
5,Ashley Sanders,ashley.sanders@fake_hotmail.com,7814535781,New York,United States,F,1998-02-10,Exercise physiologist
6,Steven Chavez,steven.chavez@fake_yahoo.com,469.238.0828,New York,United States,M,1988-01-06,Technical brewer
7,Steven Snyder,steven.snyder@fake_gmail.com,+1-294-543-6438,New York,United States,M,2000-10-27,Agricultural engineer
8,Paul Walker,paul.walker@fake_hotmail.com,205.335.5074x1759,New York,United States,M,2007-03-03,
9,Edwin Bryant,edwin.bryant@fake_hotmail.com,+1-436-359-4105x40651,New York,United States,M,1999-07-19,"Biochemist, clinical"
10,Michelle Bird,michelle.bird@fake_gmail.com,+1-325-528-1530x02284,New York,United States,F,2003-03-03,
11,Natalie Espinoza DDS,natalie.espinoza.dds@fake_yahoo.com,325-379-6960x933,New York,United States,F,1970-09-08,Public relations account executive
12,Autumn Green,autumn.green@fake_gmail.com,(841)850-4720x0041,New York,United States,F,2001-06-24,
13,Jennifer Henderson,jennifer.henderson@fake_hotmail.com,(859)776-6705x615,New York,United States,F,1999-06-03,"Engineer, control and instrumentation"
14,Jennifer Pearson,jennifer.pearson@fake_yahoo.com,354.559.8541,New York,United States,F,2006-06-23,
15,Corey Lopez,corey.lopez@fake_hotmail.com,7359516571,New York,United States,M,2000-06-27,Records manager
16,Patrick Parker,patrick.parker@fake_hotmail.com,420.510.1425x994,New York,United States,M,1999-12-10,Psychiatrist
17,Jason Wong,jason.wong@fake_yahoo.com,925.672.7219x347,New York,United States,M,1977-09-17,Legal secretary
18,Jasmine Miller,jasmine.miller@fake_hotmail.com,9492316332,New York,United States,F,2005-06-05,
19,Regina Edwards,regina.edwards@fake_yahoo.com,927-209-4313x3292,New York,United States,F,1995-01-23,"Pilot, airline"
20,Ann Hall,ann.hall@fake_hotmail.com,375.997.2367,New York,United States,F,2005-07-08,
21,Bobby Brown,bobby.brown@fake_gmail.com,001-856-231-8439,New York,United States,M,2000-07-25,Public relations officer
22,Jason Mcmillan,jason.mcmillan@fake_hotmail.com,001-406-598-5588x37147,New York,United States,M,1976-09-13,"Nurse, learning disability"
23,Stacey Cohen,stacey.cohen@fake_yahoo.com,775-224-3651x07873,New York,United States,F,1989-12-09,Database administrator
24,Sara Bennett,sara.bennett@fake_hotmail.com,(815)795-1809x90572,New York,United States,F,1999-04-08,"Surveyor, rural practice"
25,Justin Bowman,justin.bowman@fake_gmail.com,+1-902-249-7728x7570,New York,United States,M,1992-06-07,"Geologist, engineering"
26,Connor Baxter,connor.baxter@fake_hotmail.com,807.359.5296x021,New York,United States,M,1993-07-08,Nutritional therapist
27,Margaret Alexander,margaret.alexander@fake_gmail.com,606-285-3439,New York,United States,F,2003-06-14,
28,Jasmine Ramsey,jasmine.ramsey@fake_hotmail.com,+1-260-221-9212x05707,New York,United States,F,2002-09-25,
29,Steven Foster,steven.foster@fake_gmail.com,(440)513-5542x5342,New York,United States,M,2000-02-13,"Optician, dispensing"
30,Wayne Brown,wayne.brown@fake_gmail.com,+1-412-720-4529x07235,New York,United States,M,2004-04-11,
31,Craig Velazquez,craig.velazquez@fake_hotmail.com,220-542-6649,New York,United States,M,2004-10-05,
32,Christopher Mills,christopher.mills@fake_hotmail.com,6003249289,New York,United States,M,2002-09-20,
33,Jonathan Martinez,jonathan.martinez@fake_yahoo.com,834.638.8138x152,New York,United States,M,2003-10-18,
```

- employees.csv – pracownicy sklepu: Employee ID, Store ID, Name, Position.

```
Employee ID,Store ID,Name,Position
1,1,Stephen Johnson,Store Manager
2,1,Rebecca Myers,Assistant Manager
3,1,Katherine Buchanan,Cashier
4,1,Jessica Hicks,Stock Clerk
5,1,Ryan Gross,Sales Associate
6,1,Jeffery Carlson,Sales Associate
7,1,Melissa Wilson,Sales Associate
8,1,Edward Hicks,Sales Associate
9,1,Shari Jordan,Sales Associate
10,1,Mrs. Melissa Caldwell DDS,Sales Associate
11,1,Regina Nelson,Sales Associate
12,1,Kevin Dawson,Sales Associate
13,1,Nicole Murphy,Sales Associate
14,2,Susan Brown,Store Manager
15,2,Sean Miller,Assistant Manager
16,2,Christopher Dominguez,Cashier
17,2,Lisa Mason,Stock Clerk
18,2,David Pitts,Sales Associate
```

- stores.csv – informacje o sklepach: Store ID, Country, City, Store Name, Number of Employees, ZIP Code, Latitude, Longitude.

```
Store ID,Country,City,Store Name,Number of Employees,ZIP Code,Latitude,Longitude
1,United States,New York,Store New York,10,10001,40.7128,-74.006
2,United States,Los Angeles,Store Los Angeles,8,90001,34.0522,-118.2437
3,United States,Chicago,Store Chicago,9,60601,41.8781,-87.6298
4,United States,Houston,Store Houston,10,77001,29.7604,-95.3698
5,United States,Phoenix,Store Phoenix,9,85001,33.4484,-112.074
6,中国,上海,Store 上海,8,200000,31.2304,121.4737
7,中国,北京,Store 北京,10,100000,39.9042,116.4074
8,中国,广州,Store 广州,10,510000,23.1291,113.2644
9,中国,深圳,Store 深圳,9,518000,22.5429,114.0596
10,中国,重庆,Store 重庆,10,400000,29.5638,106.5505
11,Deutschland,Berlin,Store Berlin,9,10115,52.5125,13.3903
12,Deutschland,Hamburg,Store Hamburg,10,20095,53.5503,9.992
13,Deutschland,München,Store München,9,80331,48.1371,11.5763
14,Deutschland,Köln,Store Köln,8,50667,50.938,6.9578
15,Deutschland,Frankfurt am Main,Store Frankfurt am Main,10,60311,50.1145,8.6785
16,United Kingdom,London,Store London,7,EC2P 2E,51.5072,-0.1275
17,United Kingdom,Birmingham,Store Birmingham,9,B1 1AA,52.4862,-1.8904
18,United Kingdom,Glasgow,Store Glasgow,7,G1 1AA,55.858,-4.259
19,United Kingdom,Liverpool,Store Liverpool,9,L1 1AA,53.4084,-2.9916
20,United Kingdom,Bristol,Store Bristol,7,BS1 1AA,51.4545,-2.5879
21,France,Paris,Store Paris,7,75000,48.8656,2.343
```

- products.csv – katalog produktów z wielojęzycznymi opisami (Description PT/DE/FR/ES/EN/ZH), Category, Sub Category, Color, Sizes, Production Cost.

```
Product ID,Category,Sub Category,Description PT,Description DE,Description FR,Description ES,Description EN,Description ZH,Color,Sizes,Production Cost
1,Feminine,Coats and Blazers,Esportivo Veludo Verde Com Botões,Sport Samt Sport Mit Knöpfen,Sports Velvet Sports Avec Des Boutons,Deportes De Terciopelo Con Botones,Sports Velvet Sports With Buttons,运动天鹅绒运动与按钮,,S|M|L|XL,10.73
2,Feminine,Sweaters and Knitwear,Luxuoso Denim Rosa Com Botões,Luxuriöser Rosa Jeans Mit Knöpfen,Léopard De Denim Rose Avec Boutons,Denim Rosa Lujoso Con Botones,Luxurious Pink Denim With Buttons,豪华的粉红色牛仔布和纽扣,PINK,S|M|L|XL,19.55
3,Feminine,Dresses and Jumpsuits,Retrô Tricot Preto Estampado,Black Tricot Gedruckter Tricot,Tricot Imprimé En Tricot Noir,Tricot Negro Tricot Impreso,Black Tricot Printed Tricot,黑色三角形印刷三角形,BLACK,S|M|L|XL,25.59
4,Feminine,Shirts and Blouses,Blusa De Algodão Básica,Basis -Baumwollbluse,Chemisier En Coton De Base,Blusa De Algodón,Basic Cotton Blouse,基本的棉衬衫,,S|M|L|XL,27.62
5,Feminine,T-shirts and Tops,T-Shirt Básica De Algodão,Basis-Baumwoll-T-Shirt,T-Shirt En Coton De Base,Camiseta Básica De Algodón,Basic Cotton T-Shirt,基本棉T恤,,S|M|L,11.69
```

- discounts.csv – promocje: okres (Start, End), Discount (kod), Description, Category, Sub Category.

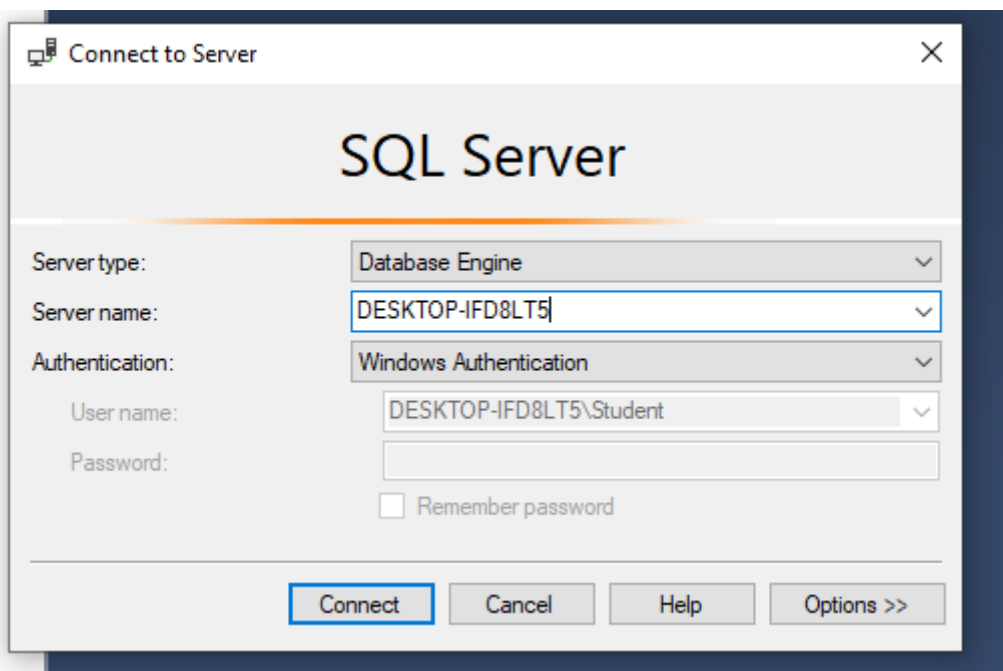
```
Start,End,Discont,Description,Category,Sub Category
2020-01-01,2020-01-10,0.4,40% discount during our New Year Winter Sale,Feminine,Coats and Blazers
2020-01-01,2020-01-10,0.4,40% discount during our New Year Winter Sale,Feminine,Sweaters and Knitwear
2020-01-01,2020-01-10,0.4,40% discount during our New Year Winter Sale,Masculine,Coats and Blazers
2020-01-01,2020-01-10,0.4,40% discount during our New Year Winter Sale,Masculine,Sweaters and Sweatshirts
2020-01-01,2020-01-10,0.4,40% discount during our New Year Winter Sale,Children,Coats
2020-01-01,2020-01-10,0.4,40% discount during our New Year Winter Sale,Children,Sweaters
2020-03-15,2020-03-31,0.35,35% discount during our Early Spring Collection Refresh,Feminine,Dresses and Jumpsuits
2020-03-15,2020-03-31,0.35,35% discount during our Early Spring Collection Refresh,Feminine,Shirts and Blouses
2020-03-15,2020-03-31,0.35,35% discount during our Early Spring Collection Refresh,Masculine,T-shirts and Polos
2020-03-15,2020-03-31,0.35,35% discount during our Early Spring Collection Refresh,Masculine,Shirts
2020-03-15,2020-03-31,0.35,35% discount during our Early Spring Collection Refresh,Children,"Girl and Boy (1-5 years, 6-14 years)"
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Feminine,T-shirts and Tops
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Feminine,Skirts and Shorts
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Feminine,Dresses and Jumpsuits
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Masculine,T-shirts and Polos
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Masculine,Pants and Jeans
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Children,"Girl and Boy (1-5 years, 6-14 years)"
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Children,Pajamas
2020-05-01,2020-05-15,0.25,25% discount during our Mid-Spring Refresh Sale,Children,Accessories
```

## Informacja

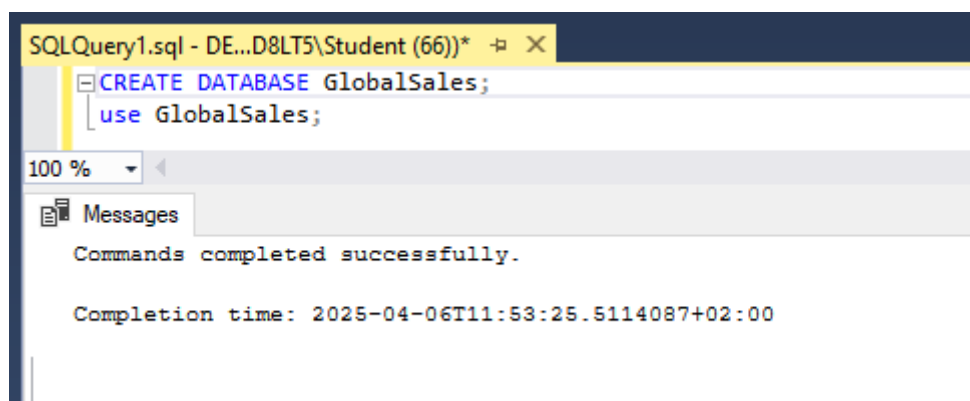
Projekt został zrealizowany za pomocą maszyny wirtualnej z OS Windows gdzie były zainstalowane programy: SQL Server Management Studio 19, Visual Studio 2017 i Tableau. Podczas trwania projektu występowały momenty, podczas których program Visual Studio 2017 crashował się co dodatkowo utrudniało pracę na projektem.

# Tworzenie bazy

Pierwszy krok to utworzenie bazy i nawiązanie połączenia z SMS.



Kolejnym krokiem było utworzenie bazy danych GlobalSales do której później były importowane dane z plików .csv



## Tworzenie nowego Użytkownika GlobalSales User

**Login - New**

Select a page: General, Server Roles, User Mapping, Securables, Status

Script Help

Login name: GlobalSales User Search...

☐ Windows authentication

☒ SQL Server authentication

Password: .....

Confirm password: .....

☐ Specify old password

Old password: .....

☒ Enforce password policy

☒ Enforce password expiration

☒ User must change password at next login

☐ Mapped to certificate

☐ Mapped to asymmetric key

☐ Map to Credential

Mapped Credentials: Credential

Default database: GlobalSales

Default language: <default>

Progress: Ready

OK Cancel

## Nadajemy role nowemu użytkownikowi

**Login - New**

Select a page: General, Server Roles, User Mapping, Securables, Status

Script Help

Users mapped to this login:

Map	Database	User	Default Schema
<input checked="" type="checkbox"/>	GlobalSales	GlobalSales User	
<input type="checkbox"/>	master		
<input type="checkbox"/>	model		
<input type="checkbox"/>	msdb		
<input type="checkbox"/>	ReportServer-173163		
<input type="checkbox"/>	ReportServer-173163Te...		
<input type="checkbox"/>	tempdb		

☐ Guest account enabled for: GlobalSales

Database role membership for: GlobalSales

- ☐ db\_accessadmin
- ☐ db\_backupoperator
- ☐ db\_datareader
- ☐ db\_datawriter
- ☐ db\_dtladmin
- ☐ db\_denydatareader
- ☐ db\_denydatawriter
- ☒ db\_owner
- ☐ db\_securityadmin
- ☒ public

Progress: Ready

OK Cancel



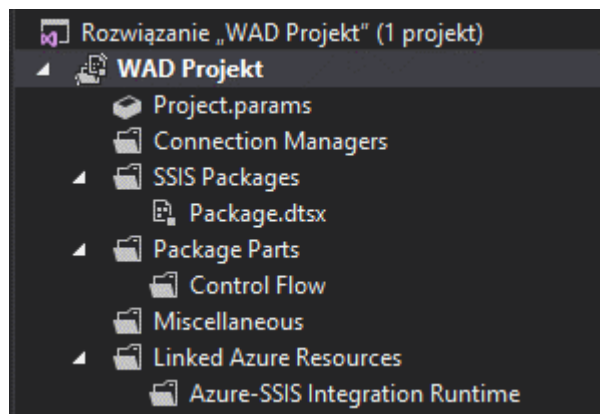
# Tabela na błędy

Tworzymy tabelę, do której w kolejnych częściach projektu będziemy wysyłać informacje o błędach.

```
SQLQuery1.sql - DE...D8LT5\Student (66)* - X
use [GlobalSales]
create table [dbo].[Errors](
[Flat File Source Error Output Column] [varchar](max) Null,
[ErrorCode] [int] null,
[ErrorColumn] [int] null)
```

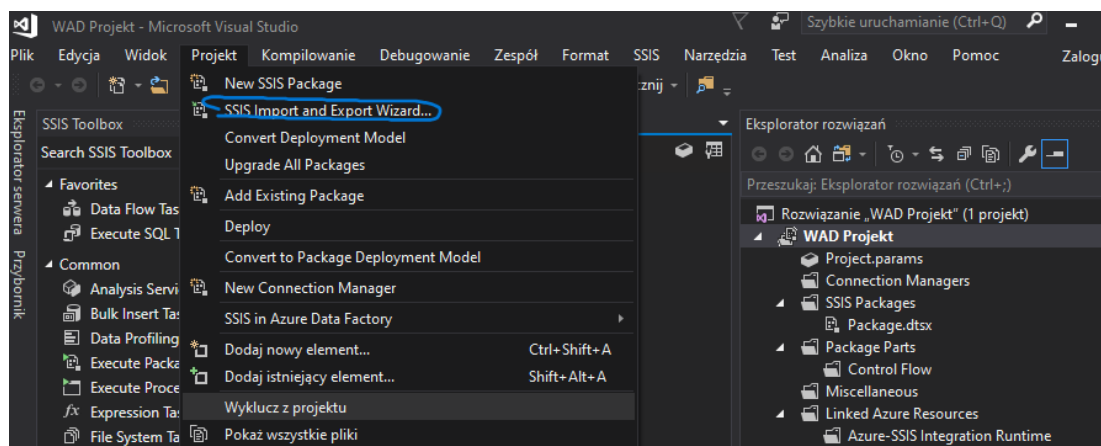
## Projekt w Visual Studio

Kolejnym etapem projektu było utworzenie nowego projektu o tytule: „Projekt WAD” w programie Visual Studio. W tym programie wykonaliśmy importowanie danych do utworzonej wcześniej bazy „GlobalSales”, oczyszczenie ich wraz z sortowaniem i utworzenie ETL.



## Przesyłanie danych

Wykonanie importu danych do bazy przebiegało w następujący sposób. Na panelu górnym wybieraliśmy zakładkę Projekt, a z niej SSIS Import and Export Wizard.

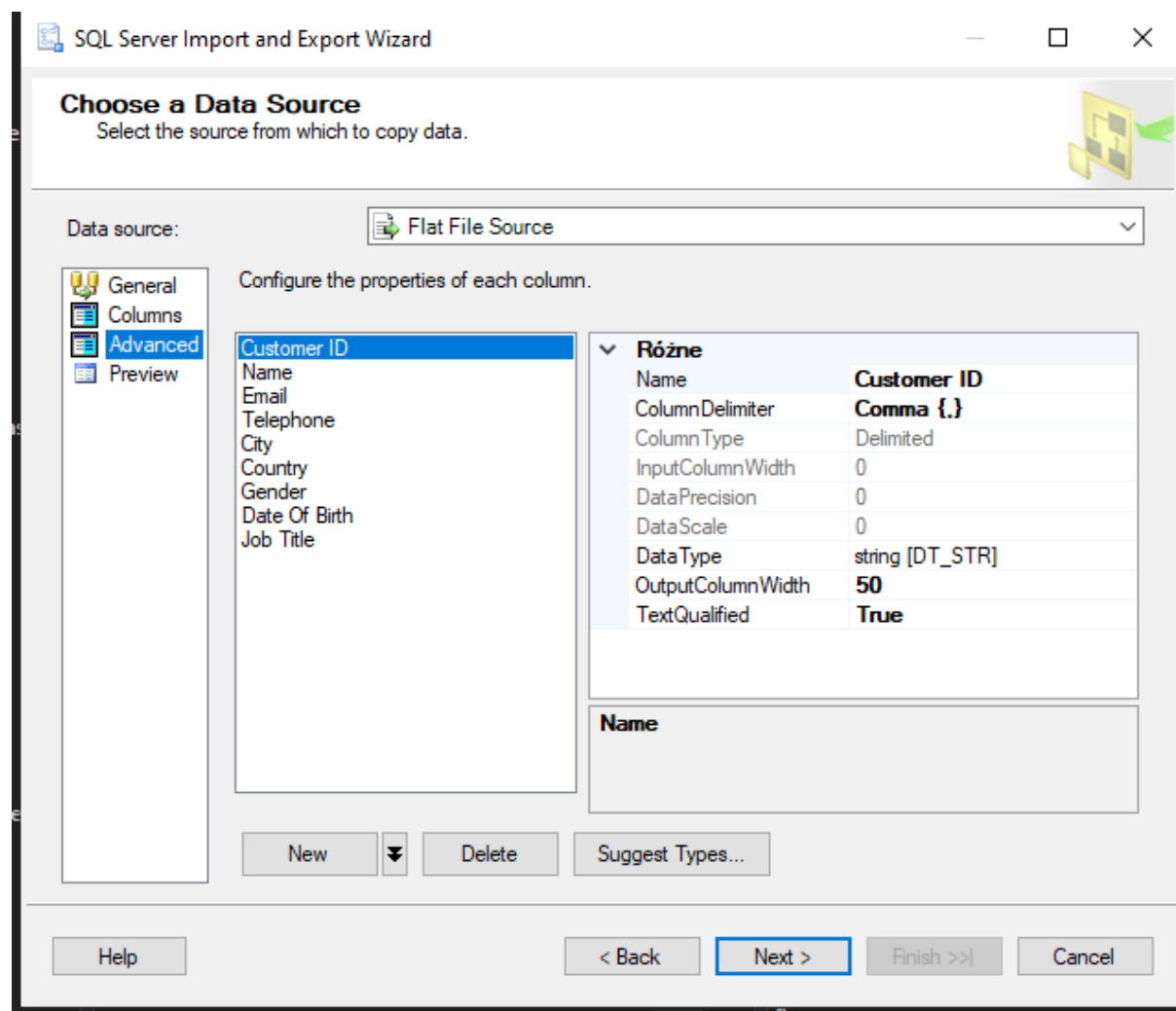


Po otwarciu się okna z listy wybieraliśmy opcję Flat File Source.

Następnie klikamy w Browse... i wybieramy jakiś plik o rozszerzeniu .csv chcemy wgrać do naszej bazy.

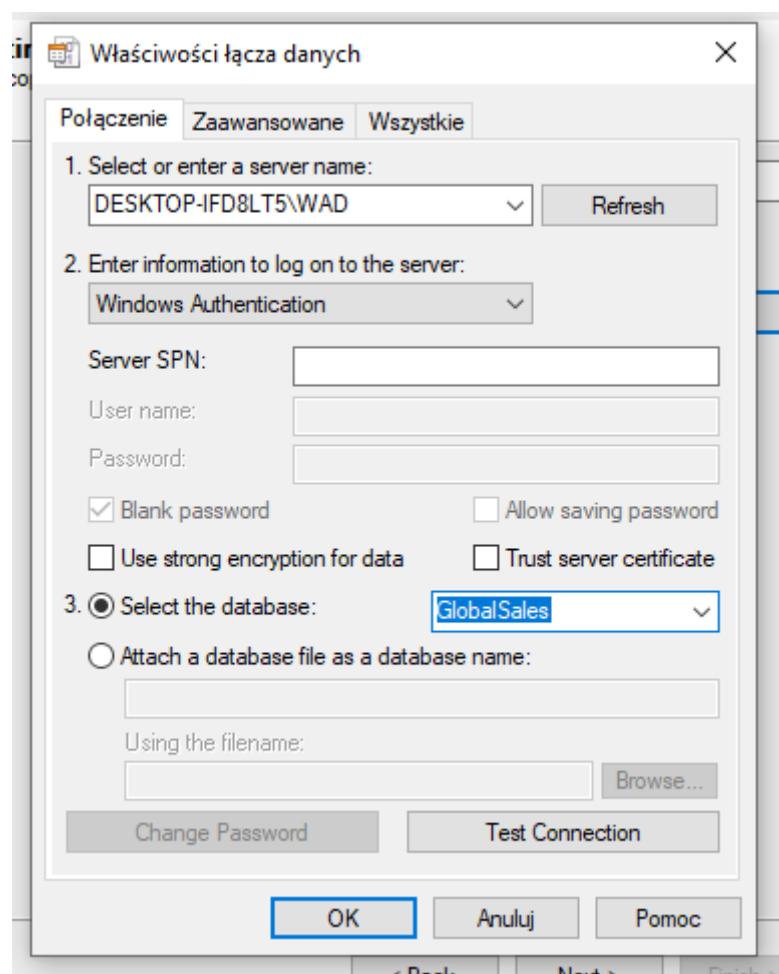
Po wybraniu pliku w kolumnie znajdującej się po prawej stronie przechodzimy do zakładki Advanced i Suggest Types....

W tym miejscu odznaczamy zaznaczone automatycznie 2 opcje Suggest the smallest integer data type i Suggest the smallest real data type i zatwierdzamy wykonaną operację.



Po przejściu do kolejnej części kreatora zmieniamy destynację czyli miejsce, do którego chcemy zaimportować nasze dane na Microsoft OLE DB Driver for SQL Server, a następnie przechodzimy do Properties....

W właściwościach łączymy się z naszą bazą danych, którą utworzyliśmy wcześniej. Początkowo uzupełniamy nazwę serwera, z którym się łączymy. W naszym przypadku było to: DESKTOP-IFD8LT5\WAD. W drugim punkcie po rozwinięciu paska wybieramy Windows Authentication. Wtedy w trzeciej części możemy wybrać nazwę naszej bazy danych, do której zostaną zaimportowane dane.

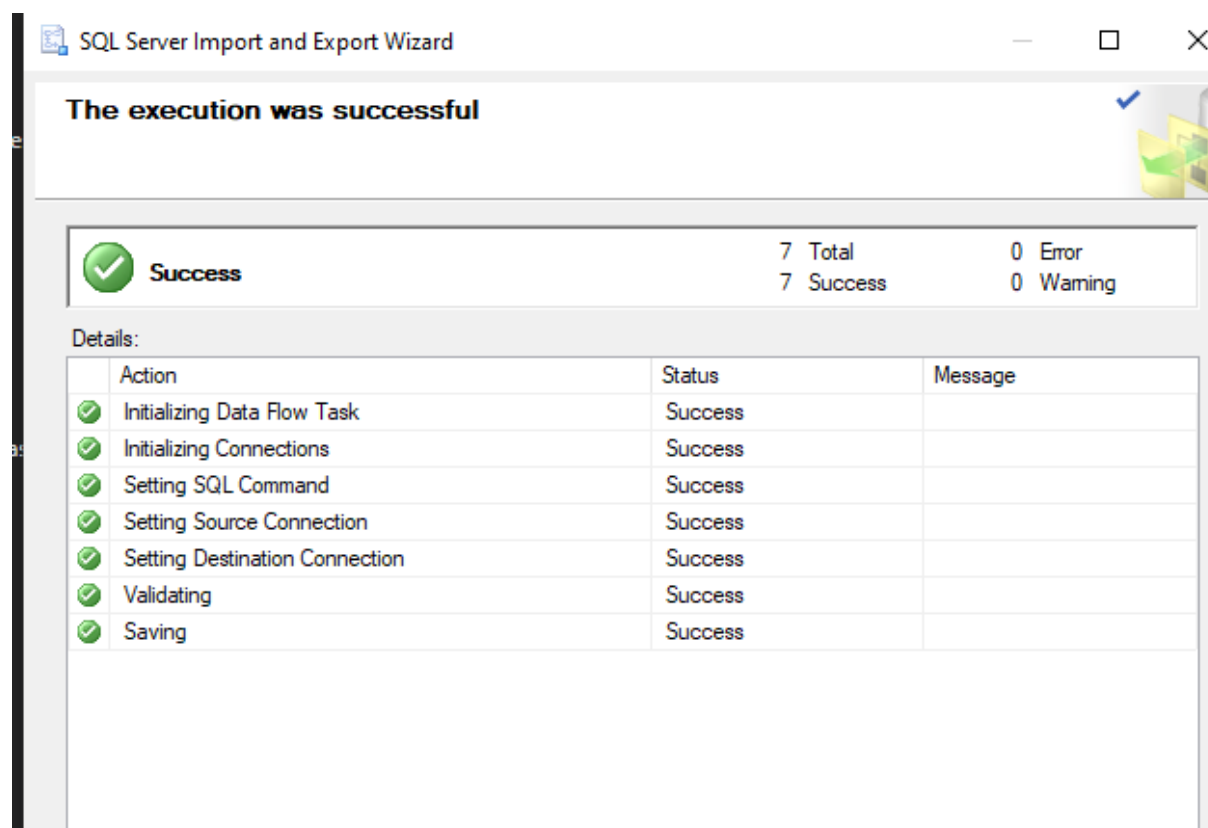


Aby upewnić się, że wszystko zostało uzupełnione poprawnie dokonujemy testu połączenia, którego rezultatem powinna być pozytywna informacja o połączeniu. Po otrzymaniu jej zatwierdzamy wszystkie zmiany i przechodzimy do kolejnej części kreatora

Przechodzimy do Edit Mappings....

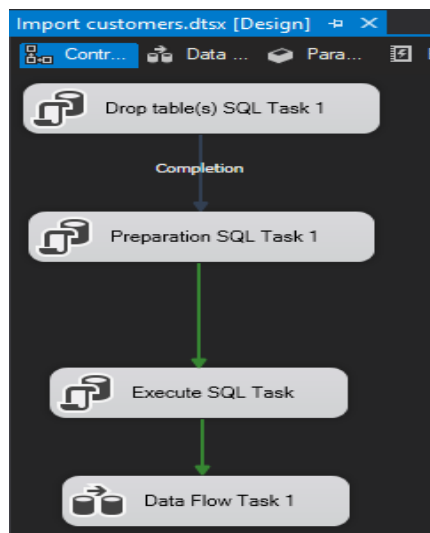
Edytując mapowanie danych zaznaczamy Drop and re-create destination table, a potem zatwierdzamy dokonaną zmianę.

Zatwierdzamy wszystko i kończymy tworzenie nowej tabeli. Poprawny wynik jaki powinniśmy otrzymać przedstawiony jest na poniższym zrzucie ekranu.



Po zamknięciu okna pojawia się nam utworzony przez nas proces ETL. Będąc w zakładce Control Flow powinniśmy widzieć proces.

Ze znajdującego się po lewej stronie Toolbox`a dodajemy Execute SQL Task i podpinamy go do naszego procesu w następujący sposób.



Podwójnym kliknięciem na niego wchodzimy do jego właściwości i rozpoczynamy konfigurację. W oknie, które się nam pojawi przy opcji Connection wybieramy DestinationConnectionOLEDB, a przy opcji SQLStatement wpisujemy frazę DELETE FROM Errors. Zatwierdzamy wykonane zmiany.

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

General  
Parameter Mapping  
Result Set  
Expressions

<b>General</b>	
Name	Execute SQL Task
Description	Execute SQL Task
<b>Options</b>	
TimeOut	0
CodePage	1250
TypeConversionMode	Allowed
<b>Result Set</b>	
ResultSet	None
<b>SQL Statement</b>	
ConnectionType	OLE DB
Connection	DestinationConnectionOLEDB
SQLSourceType	Direct input
SQLStatement	delete from Errors
IsQueryStoredProcedure	False
BypassPrepare	True

**BypassPrepare**  
Indicates whether the task should prepare the query before executing it.

Browse... Build Query... Parse Query

OK Cancel Help

Sprawdzamy czy poprawnie jest skonfigurowany kafelek Drop table(s) SQL Task 1. Jego poprawna konfiguracja przedstawia się w następujący sposób:

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

**General**  
 Parameter Mapping  
 Result Set  
 Expressions

<b>General</b>	
Name	Drop table(s) SQL Task 1
Description	
<b>Options</b>	
TimeOut	0
CodePage	1250
TypeConversionMode	Allowed
<b>Result Set</b>	
ResultSet	None
<b>SQL Statement</b>	
ConnectionType	OLE DB
Connection	DestinationConnectionOLEDB
SQLSourceType	Direct input
SQLStatement	drop table [dbo].[customers]GO
IsQueryStoredProcedure	False
BypassPrepare	True

**Name**  
 Specifies the name of the task.

Browse... Build Query... Parse Query

OK Cancel Help

Tak samo sprawdzamy konfiguracje kafelka Preparation SQL Task 1.

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

**General**  
 Parameter Mapping  
 Result Set  
 Expressions

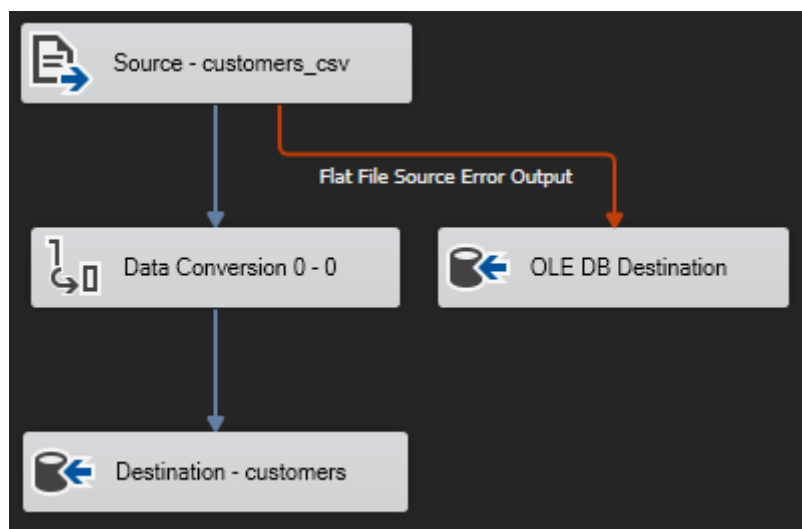
<b>General</b>	
Name	Preparation SQL Task 1
Description	
<b>Options</b>	
TimeOut	0
CodePage	1250
TypeConversionMode	Allowed
<b>Result Set</b>	
ResultSet	None
<b>SQL Statement</b>	
ConnectionType	OLE DB
Connection	DestinationConnectionOLEDB
SQLSourceType	Direct input
SQLStatement	CREATE TABLE [dbo].[customers] ([Customer ID
IsQueryStoredProcedure	False
BypassPrepare	True

**Name**  
 Specifies the name of the task.

Browse... Build Query... Parse Query

OK Cancel Help

Następnie przechodzimy do Data Flow i z ToolBox'a przeciągamy na pole projektu kafelek OLE DB Destination. Do istniejącego już wcześniej schematu stworzonego przez kreator podłączamy go w następująco.



Gdy podłączymy kafelki automatycznie otwiera nam się okno do edycji. W tym miejscu dla kolumny Error i Truncation dla wszystkich kolumn zmieniamy zawartość z Fail component na Redirect row i klikamy Apply w celu zatwierdzania dokonanych zmian.

Configure Error Output

Specify how row-level errors are handled by this component. You can handle errors in the row, or truncation errors in columns. Errors can fail the component, or they can be ignored, or they can be redirected to an error output.

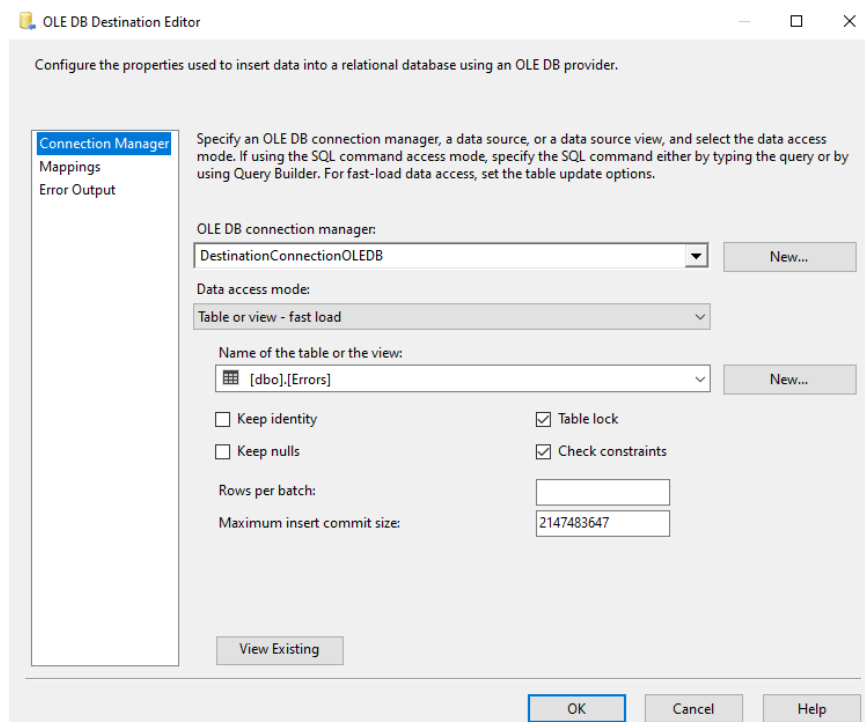
Input or Output	Column	Error	Truncation	Description
Flat File Sourc...				
	Customer ID	Fail component	Fail component	Conversion
	Name	Fail component	Fail component	Conversion
	Email	Fail component	Fail component	Conversion
	Telephone	Fail component	Fail component	Conversion
	City	Fail component	Fail component	Conversion
	Country	Fail component	Fail component	Conversion
	Gender	Fail component	Fail component	Conversion
	Date Of Birth	Fail component	Fail component	Conversion
	Job Title	Fail component	Fail component	Conversion

Set this value to selected cells: Redirect row

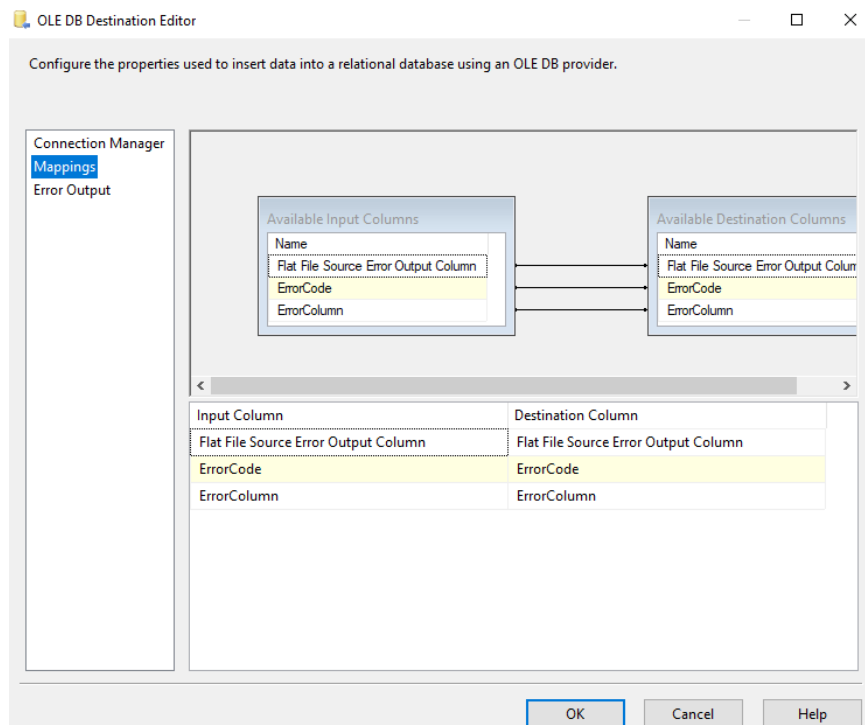
Apply

OK Cancel

Klikamy następnie na kafelek by go edytować. Wybieramy tam gdzie mają być wysłane błędy czyli wcześniej utworzoną tabelę Errors.

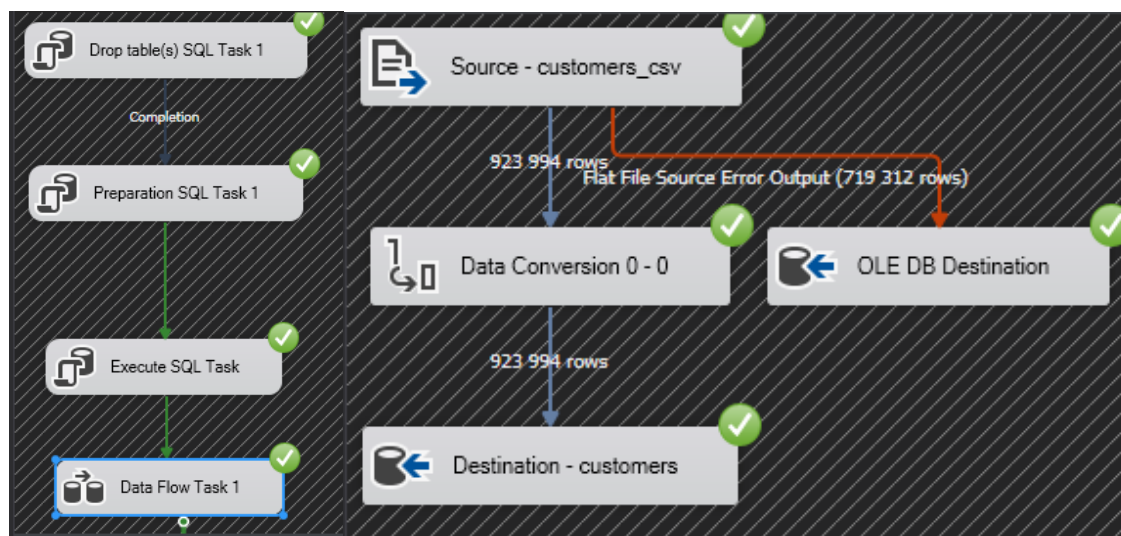


Potem przechodzimy do zakładki Mappings znajdującej się w kolumnie po lewej stronie i sprawdzamy czy wszystko jest poprawnie połączone. Zatwierdzamy dokonane zmiany.

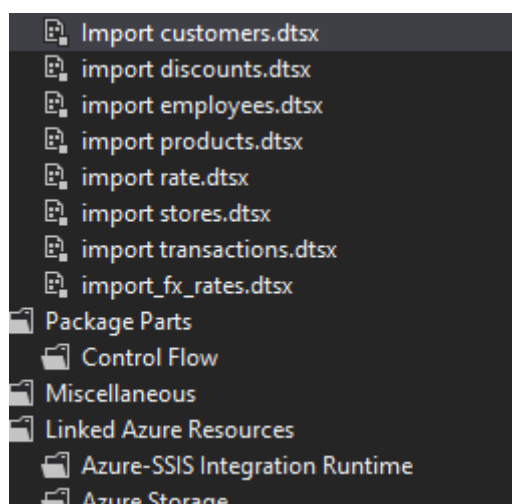




Po ukończeniu powyższej części proces jest poprawnie skonfigurowany i gotowy do uruchomienia. Uruchamiamy go i przesyłamy dane do naszej bazy. Na poniższych zrzutach ekranu przedstawione jest udane przesłanie danych. Pierwszy zrzut jest z zakładki Control Flow, a drugi z Data Flow.



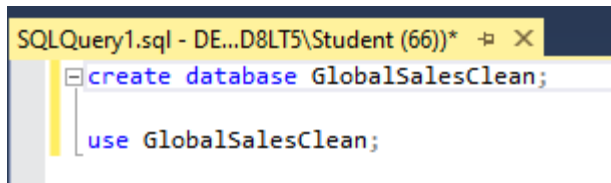
Powyższe kroki znajdujące się w rozdziale Przesyłanie danych powtórzyliśmy dla wszystkich danych. Po ich wykonaniu projekt przedstawia się tak:



Import każdej tabeli zakończył się powodzeniem.

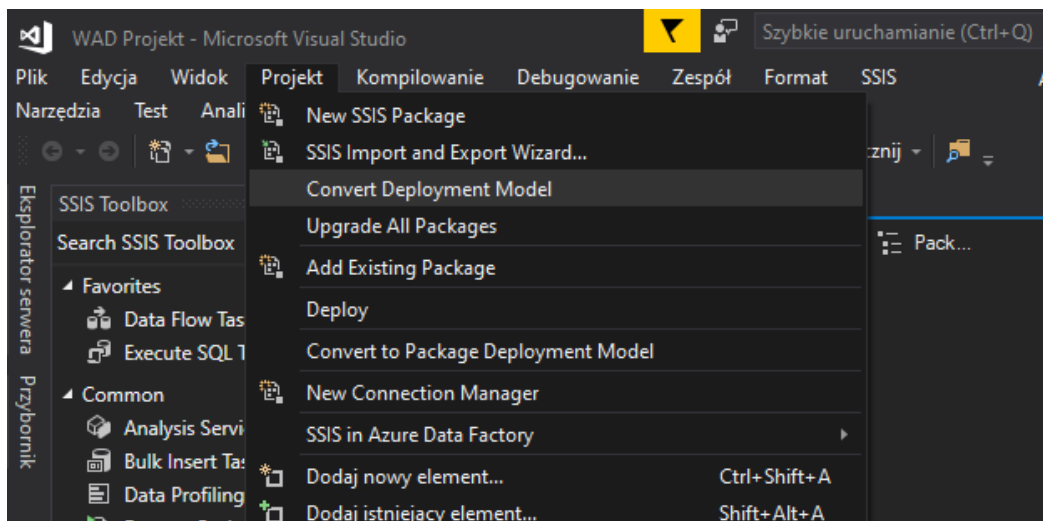
# Utworzenie bazy danych nr2/czystej bazy

Zostaje utworzona nowa baza danych o nazwie GlobalSalesClean. Będą do niej wysyłane później oczyszczone dane. W późniejszej części projektu dane z tej bazy będą wykorzystywane w ETL i wizualizacji.

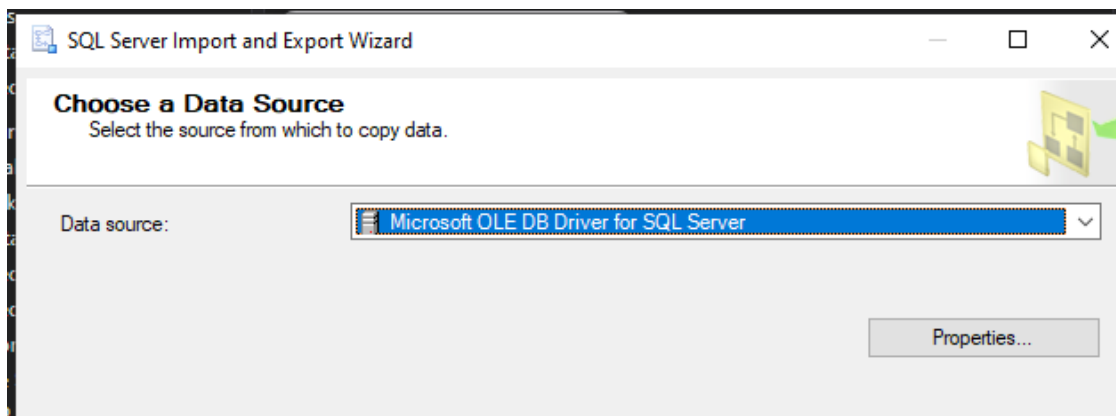


# Utworzenie procesu ETL do oczyszczenia danych

Z górnego panelu wybieramy Projekt i SSIS Import and Export Wizard....

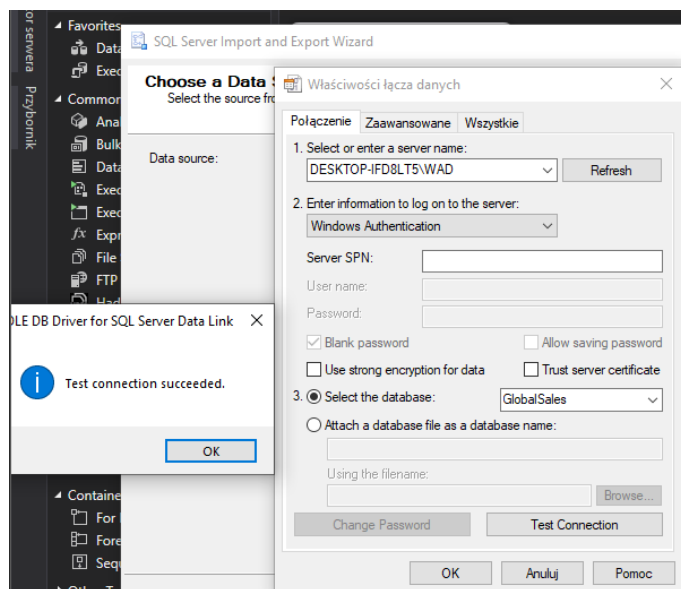


W pierwszej części kreatora wybieramy opcje Microsoft OLE DB Driver for SQL Server i przechodzimy do Properties....



Tak samo jak w przypadku importowania danych w właściwościach łączymy się z bazą GlobalSales, którą utworzyliśmy wcześniej. W pierwszej kolejności uzupełniamy nazwę serwera, z którym się łączymy. W naszym przypadku było to: DESKTOP-IFD8LT5\WAD. W drugim punkcie po rozwinięciu paska wybieramy Windows Authentication. W części trzeciej wybieramy bazę GlobalSales.

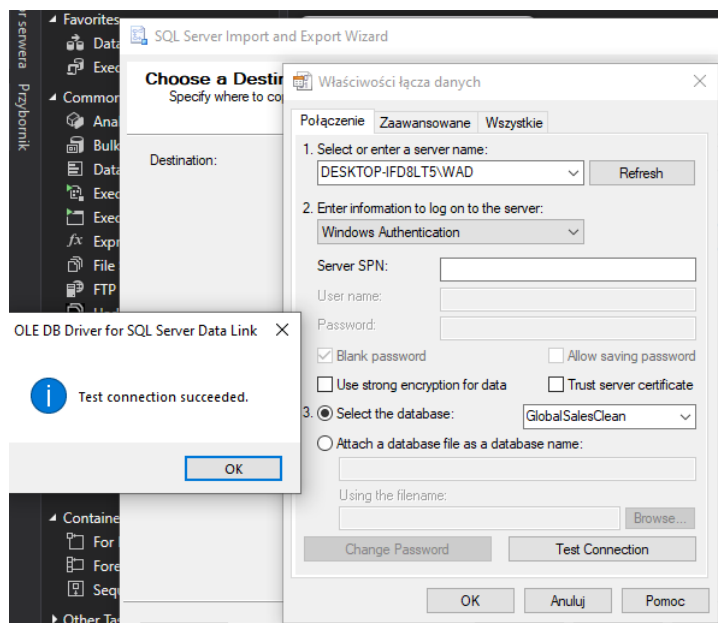
Wykonujemy test połączenia.



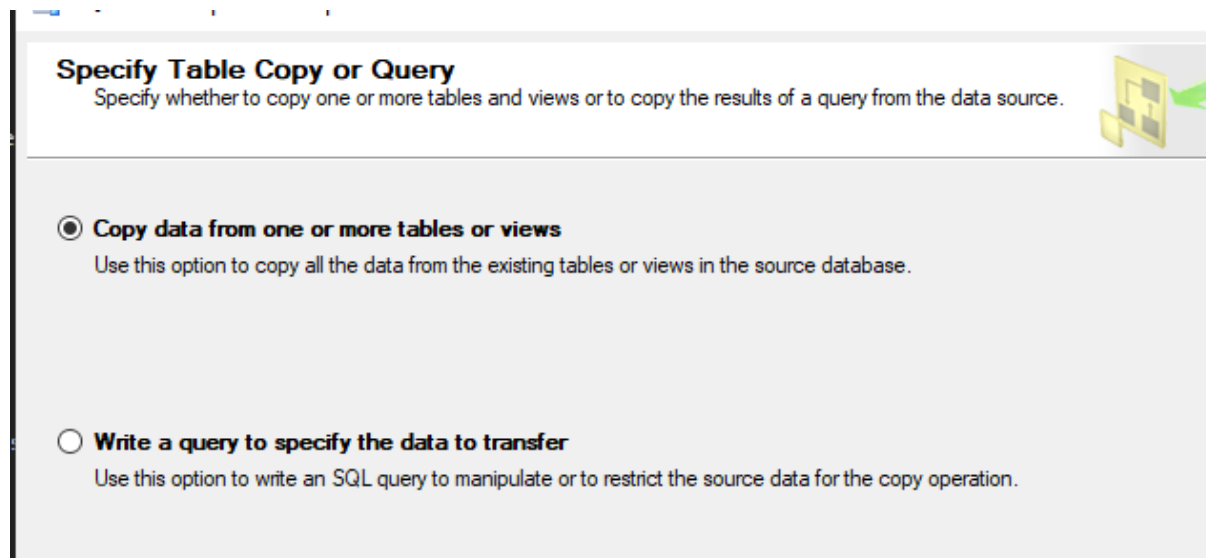
Przechodząc dalej z rozsuwanego paska ponownie wybieramy opcję Microsoft OLE DB Driver for SQL Server i przechodzimy do Properties....

W właściwościach wpisujemy tak samo nazwę serwera i w drugim punkcie zmieniamy na Windows Authentication. Zaś w trzeciej części wybieramy nowo utworzoną bazę MarketBasketClean jako naszą bazę docelową.

Ponownie wykonujemy test połączenia.



W kolejnej części edytora wybieramy Copy data from one or more tables or views i przechodzimy dalej.

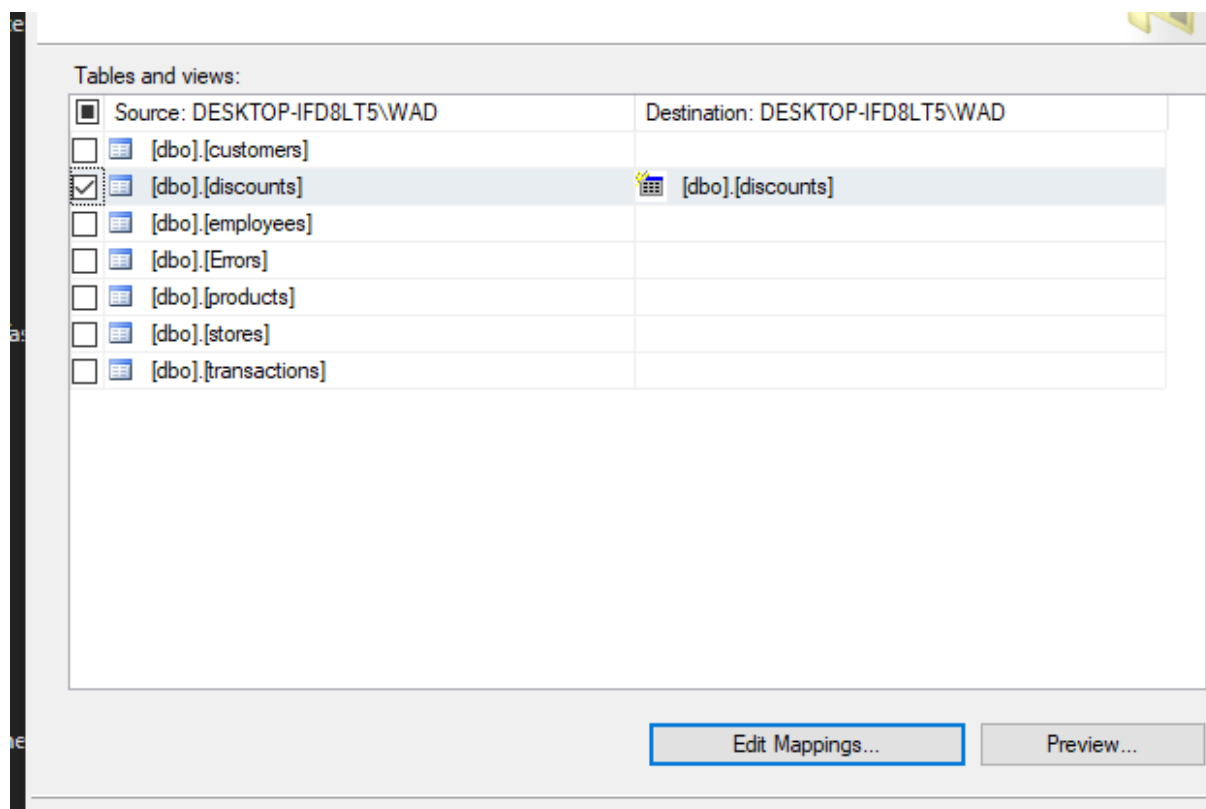


**Specify Table Copy or Query**  
Specify whether to copy one or more tables and views or to copy the results of a query from the data source.

☒ **Copy data from one or more tables or views**  
Use this option to copy all the data from the existing tables or views in the source database.

☐ **Write a query to specify the data to transfer**  
Use this option to write an SQL query to manipulate or to restrict the source data for the copy operation.

Zatwierdzamy dane i przechodzimy do kolejnej części. Wybieramy w niej tabele, którą chcemy oczyścić.



Tables and views:

Source: DESKTOP-IFD8LT5\\WAD	Destination: DESKTOP-IFD8LT5\\WAD
<input type="checkbox"/> [dbo].[customers]	
<input checked="" type="checkbox"/> [dbo].[discounts]	[dbo].[discounts]
<input type="checkbox"/> [dbo].[employees]	
<input type="checkbox"/> [dbo].[Errors]	
<input type="checkbox"/> [dbo].[products]	
<input type="checkbox"/> [dbo].[stores]	
<input type="checkbox"/> [dbo].[transactions]	

Edit Mappings... Preview...

Przechodzimy do Edit mappings.... Zaznaczamy tam opcje Drop and re-create destination table. Zatwierdzamy edycje i przechodzimy do dalszej części kreatora.

Column Mappings

Source: [dbo].[discounts]  
Destination: [dbo].[discounts]

☒ Create destination table Edit SQL...  
☐ Delete rows in destination table ☒ Drop and re-create destination table  
☐ Append rows to the destination table ☐ Enable identity insert

Mappings:

Source	Destination	Type	Nullable	Size	Precision	Scale
Start	Start	datetime	<input checked="" type="checkbox"/>			
End	End	datetime	<input checked="" type="checkbox"/>			
Discont	Discont	varchar	<input checked="" type="checkbox"/>	4		
Description	Description	varchar	<input checked="" type="checkbox"/>	55		
Category	Category	varchar	<input checked="" type="checkbox"/>	9		
Sub Category	Sub Category	varchar	<input checked="" type="checkbox"/>	24		

Source column: Start datetime

OK Cancel

Klikamy Finish i otrzymujemy informację o udanym stworzeniu procesu.

SQL Server Import and Export Wizard

**The execution was successful**

**Success** 7 Total 0 Error  
7 Success 0 Warning

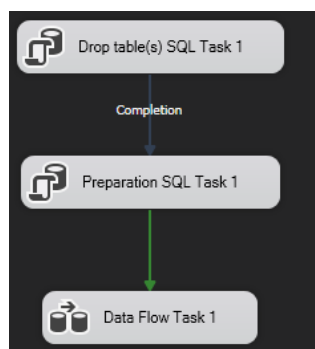
Details:

Action	Status	Message
✓ Initializing Data Flow Task	Success	
✓ Initializing Connections	Success	
✓ Setting SQL Command	Success	
✓ Setting Source Connection	Success	
✓ Setting Destination Connection	Success	
✓ Validating	Success	
✓ Saving	Success	

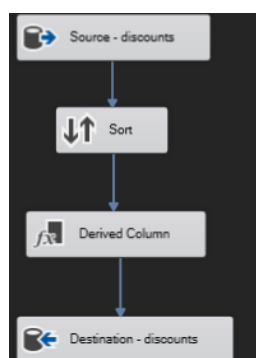
Stop Report

Close

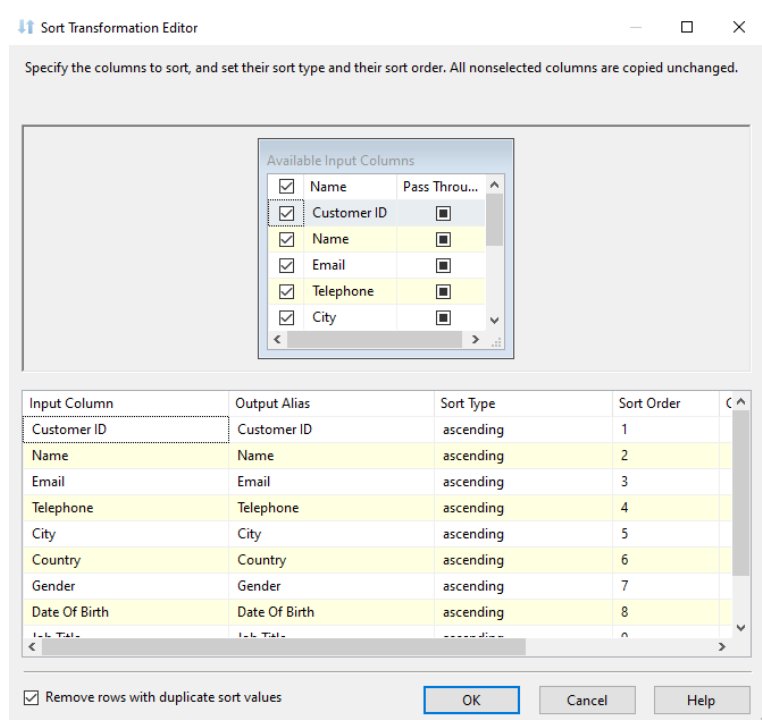
Otrzymujemy schemat naszego procesu. W Control Flow wygląda następująco:



Następnie z ToolBox'a przeciągamy kafelki Sort i Derived Column i podłączamy je jak na przedstawionym zrzucie ekranu.



Dwuklikiem na kafelek Sort odpalamy właściwości i rozpoczynamy jego konfigurację. Upewniamy się że mamy zaznaczone wszystkie kolumny, a następnie zaznaczamy opcje Remove rows with duplicate sort values. Zatwierdzamy konfigurację.



Kolejnym krokiem jest konfiguracja kafelka Derived Column. Dwuklikiem wchodzimy do niego następnie rozwijamy folder Columns i przeciągamy niżej kolumnę, którą chcemy oczyścić z pustych wartości. Następnie w sekcji Expression wpisujemy komendę. Wykonujemy dla każdej kolumny. Zatwierdzamy konfigurację.

Derived Column Transformation Editor

Specify the expressions used to create new column values, and indicate whether the values update existing columns or populate new columns.

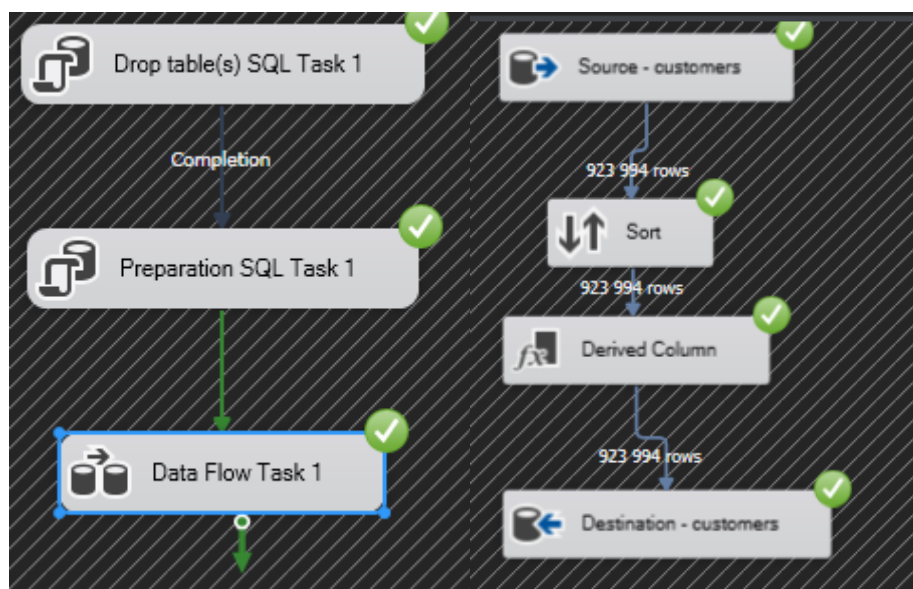
Variables and Parameters
Columns
Customer ID
Name
Email
Telephone
City
Country
Gender
Date Of Birth
Job Title

Mathematical Functions
String Functions
Date/Time Functions
NULL Functions
Type Casts
Operators

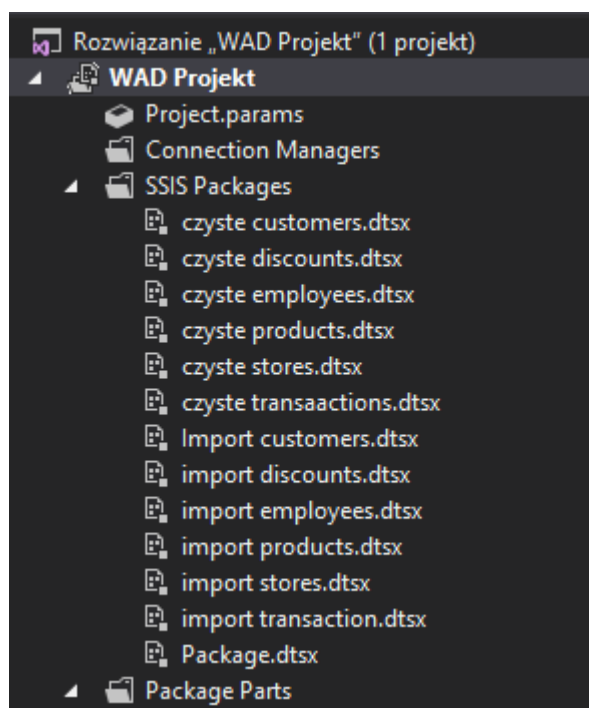
Description:

Derived Column Name	Derived Column	Expression	Data Type	Length	Precision	Scale	Code Page
Customer ID	Replace 'Cust...	ISNULL([Customer ID]) ? NUL...	eight-byte signed integer [DT_I8]				
Name	Replace 'Name'	([Name] == "") ? NULL(DT_WS...	string [DT_STR]	22			1250 (AN
Email	Replace 'Email'	([Email] == "") ? NULL(DT_WS...	string [DT_STR]	37			1250 (AN
Telephone	Replace 'Telep...	([Telephone] == "") ? NULL(DT...	string [DT_STR]	22			1250 (AN
City	Replace 'City'	([City] == "") ? NULL(DT_WST...	string [DT_STR]	8			1250 (AN
Country	Replace 'Coun...	([Country] == "") ? NULL(DT_...	string [DT_STR]	13			1250 (AN
Gender	Replace 'Gend...	([Gender] == "") ? NULL(DT_W...	string [DT_STR]	1			1250 (AN
Date Of Birth	Replace 'Date ...	ISNULL([Date Of Birth]) ? NUL...	database timestamp [DT_DBTIMESTAMP]				
Job Title	Replace 'Job T...	([Job Title] == "") ? NULL(DT_...	string [DT_STR]	43			1250 (AN

Utworzony przez nas proces ETL jest gotowy do uruchomienia. Poniżej zostają przedstawione pozytywne wyniki jego działania. Pierwszy zrzut jest z zakładki Control Flow, a drugi z Data Flow.



Wszystkie czynności przedstawione w tym rozdziale zostały wykonane dla wszystkich tabel. W tym momencie projekt wygląda jak poniżej:





# Tworzenie procesów ETL

W tej części wykonamy ETL w programie Visual Studio 2017. Zadaniem procesów ETL jest wyciągnięcie z naszej bazy potrzebnych danych z konkretnych kolumn. Następnie wykonywana jest na nich sortowanie czy agregacja jak i inne czynności, które są nam potrzebne dla wartości, które są dla nas interesujące.

Każdy ETL można wykonać na kilka sposobów, chcieliśmy pokazać różne rozwiązania takie jak: użycie Execute SQL Task (z WITH), Data Flow Task (z użyciem dużej ilości kafelków), Data Flow Task (z rozbudowanymi kwerendami w OLE DB Source co upraszczało schemat ETL).

## Time & Sales (etl\_TimeAndSales.dtsx)

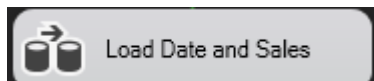
Co robi: Jednym pakietem SSIS w Data Flow (Multicast → Sort → Derived Column → Aggregate) oraz/lub w Execute SQL Task (CTE) tworzymy:

DimDate – wymiar kalendarza z kolumnami DateKey, FullDate, Year, Month, DayOfWeek, Quarter,

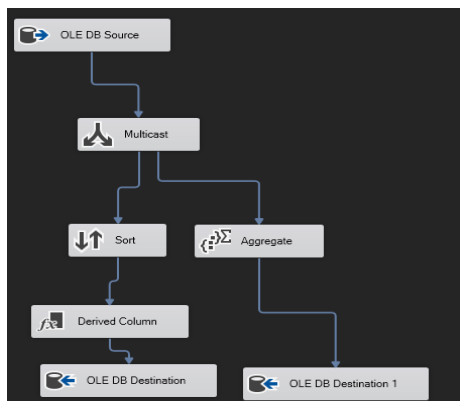
FactSalesDaily – fakt dziennej sprzedaży (FullDate, TotalSales).

Tworzymy nowy projekt poprzez wybranie z górnego paska opcji Projekt i następnie New SSIS Package.

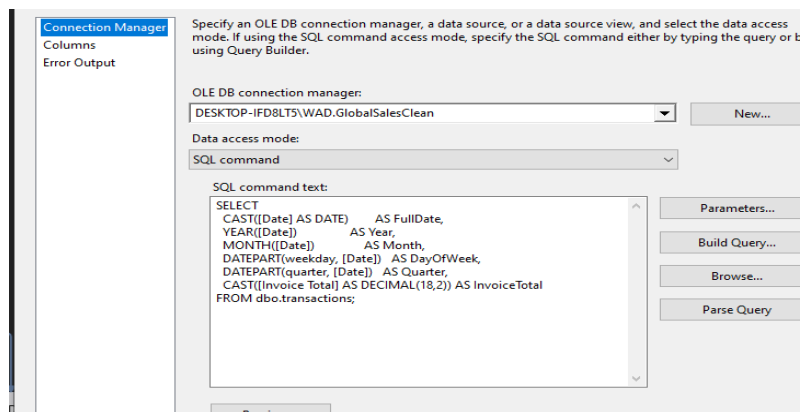
Na pole operacyjne z BoxTool'a przeciągamy kafelek Data Flow Task i przechodzimy do zakładki Data Flow.



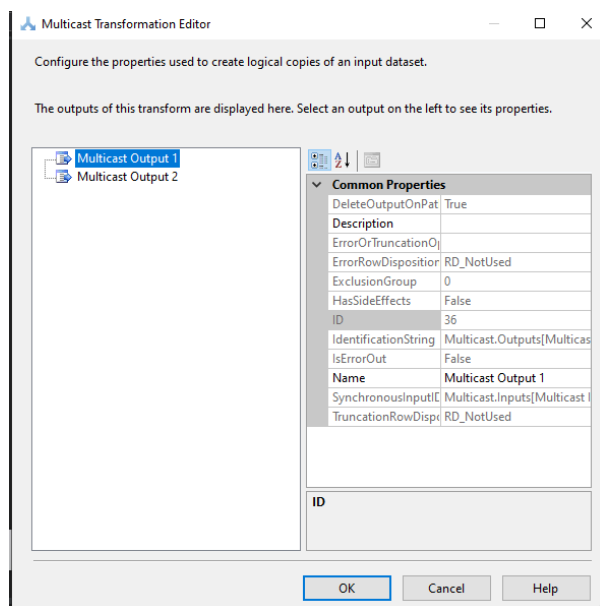
Następnie tworzymy tam przedstawiony poniżej schemat. Należy jednak pamiętać, że nie łączymy od razu wszystkich kafelków.



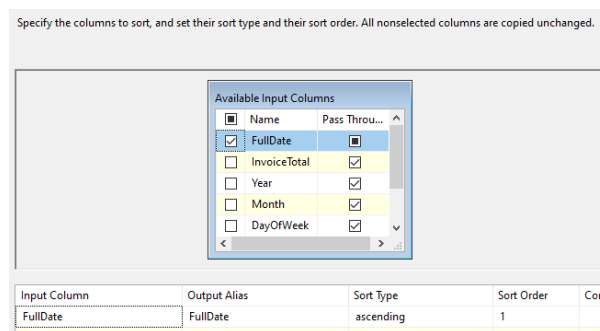
Poprawna konfiguracja OLE DB Source przedstawia się w następujący sposób.



Konfiguracja Multicast, duplikuje strumień danych na **dwa równoległe** wyjścia.



Konfiguracja Sort Sortuje po kolumnie FullDate i usuwa duplikaty (Remove rows with duplicate sort values), dzięki czemu zostaje dokładnie jedna unikalna data.



Konfiguracja Aggregate, grupa po FullDate i sumuje pole Invoice Total, żeby wyliczyć łączną wartość sprzedaży na dany dzień.

Σ Aggregate Transformation Editor

Aggregations

Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Available Input Column...

☒ Name

☐ (\*)

☒ FullDate

☒ InvoiceTotal

☐ Year

☐ Month

Input Column	Output Alias	Operation	Comparison
FullDate	FullDate	Group by	
InvoiceTotal	TotalSales	Sum	

<

>

OK

Cancel

Help

Konfiguracja Derived Column, tworzy z tej daty wszystkie potrzebne atrybuty wymiaru:

DateKey = YEAR(FullDate)\*10000 + MONTH(FullDate)\*100 + DAY(FullDate)

Year = YEAR(FullDate)

Month = MONTH(FullDate)

DayOfWeek = DATEPART(weekday, FullDate)

Quarter = DATEPART(quarter, FullDate)

Derived Column Name	Derived Column	Expression	Data Type	Length
DateKey	<add as new column>	YEAR(FullDate) * 10000 + MONTH(FullDate) * 100 ...	four-byte signed integ...	Le

Konfiguracja OLE DB Destination, wrzuca wynik do tabeli **DimDate**

Konfiguracja OLE DB Destination 1, wrzuca wynik do tabeli **FactSalesDaily**

OLE DB Destination Editor

Configure the properties used to insert data into a relational database using an OLE DB provider.

**Connection Manager**  
Mappings  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:  
 DESKTOP-IFD8LT5\WAD.GlobalSalesClean New...

Data access mode:  
 Table or view - fast load

Name of the table or the view:  
 [dbo].[FactSalesDaily] New...

☐ Keep identity ☒ Table lock  
☐ Keep nulls ☒ Check constraints

Rows per batch:

Maximum insert commit size:

View Existing

OK Cancel Help

## Wyniki DimDate

Object Explorer

Connect - [Server] [Database] [Table] [View] [External Resource]

DESKTOP-IFD8LT5\WAD (SQL Server 15.0.2000.5 - D)

Databases

System Databases

Database Snapshots

GlobalSales

GlobalSalesClean

Database Diagrams

Tables

System Tables

FileTables

External Tables

Graph Tables

dbo.CurrencySalesSummary

dbo.customers

dbo.DimDate

Columns

Keys

Constraints

Triggers

Indexes

Statistics

dbo.DiscountEffectiveness

dbo.discounts

dbo.EmployeePerformance

dbo.employees

dbo.ExchangeRates

dbo.FactAvgBasketByCity

dbo.FactReturnRate

dbo.FactSalesDaily

dbo.FactStoreSales

dbo.FactTransactionsByHour

dbo.fx\_data\_clean

dbo.products

dbo.stores

dbo.transactions

Views

External Resources

SQLQuery1.sql - DE...D8LT5(Student (58))

```
SELECT TOP (1000) [DateKey]
, [FullDate]
, [Year]
, [Month]
, [DayOfWeek]
, [Quarter]
FROM [GlobalSalesClean].[dbo].[DimDate]
```

100 %

Results Messages

	DateKey	FullDate	Year	Month	DayOfWeek	Quarter
1	20230101	2023-01-01	2023	1	1	1
2	20230102	2023-01-02	2023	1	2	1
3	20230103	2023-01-03	2023	1	3	1
4	20230104	2023-01-04	2023	1	4	1
5	20230105	2023-01-05	2023	1	5	1
6	20230106	2023-01-06	2023	1	6	1
7	20230107	2023-01-07	2023	1	7	1
8	20230108	2023-01-08	2023	1	1	1
9	20230109	2023-01-09	2023	1	2	1
10	20230110	2023-01-10	2023	1	3	1
11	20230111	2023-01-11	2023	1	4	1
12	20230112	2023-01-12	2023	1	5	1
13	20230113	2023-01-13	2023	1	6	1
14	20230114	2023-01-14	2023	1	7	1
15	20230115	2023-01-15	2023	1	1	1
16	20230116	2023-01-16	2023	1	2	1
17	20230117	2023-01-17	2023	1	3	1
18	20230118	2023-01-18	2023	1	4	1
19	20230119	2023-01-19	2023	1	5	1
20	20230120	2023-01-20	2023	1	6	1
21	20230121	2023-01-21	2023	1	7	1
22	20230122	2023-01-22	2023	1	1	1
23	20230123	2023-01-23	2023	1	2	1
24	20230124	2023-01-24	2023	1	3	1
25	20230125	2023-01-25	2023	1	4	1
26	20230126	2023-01-26	2023	1	5	1

## Wyniki FactSaleDaily

SQLQuery2.sql - DE...D8LTS\Student (61)

```

SELECT TOP (1000) [FullDate]
, [TotalSales]
FROM [GlobalSalesClean].[dbo].[FactSalesDaily]

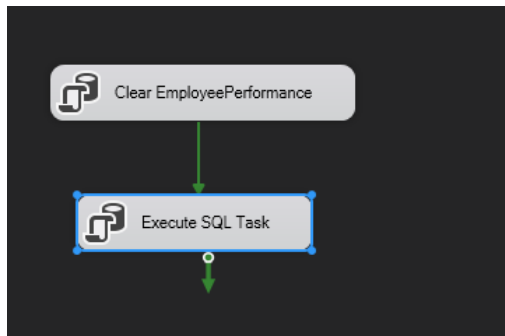
```

100 %

	FullDate	TotalSales
1	2023-01-01	903238.45
2	2023-01-02	217145.44
3	2023-01-03	183533.33
4	2023-01-04	195137.21
5	2023-01-05	185506.72
6	2023-01-06	183455.25
7	2023-01-07	367607.11
8	2023-01-08	383033.72
9	2023-01-09	169861.29
10	2023-01-10	184798.03
11	2023-01-11	138864.65
12	2023-01-12	163002.20

## Employee Performance (etl\_EmployeePerformance.dtsx)

– Co robi: Wykorzystując pojedynczy Execute SQL Task z CTE (WITH SalesCTE ..., EmployeeAgg ...) agregujemy TransactionsClean po EmployeeID (suma InvoiceTotal, liczba transakcji) i dołączamy EmployeesClean, a następnie TRUNCATE+INSERT do tabeli EmployeePerformance(EmployeeID, Name, Position, TotalSales, TransactionCount).



„Clear EmployeePerformance” (Execute SQL Task)

Co robi: usuwa wszystkie dotychczasowe wiersze z tabeli docelowej EmployeePerformance, aby każdy przebieg procesu ładował dane od nowa (bez duplikatów).

Execute SQL Task Editor

Configure the properties required to run SQL statements and stored procedures using the selected connection.

**General**

Name: Clear EmployeePerformance

Description: Execute SQL Task

**Options**

TimeOut: 0

CodePage: 1250

TypeConversionMode: Allowed

**Result Set**

ResultSet: None

**SQL Statement**

ConnectionType: OLE DB

Connection: DESKTOP-IFD8LTS\WAD.GlobalSalesClean

SQLSourceType: Direct input

SQLStatement: TRUNCATE TABLE dbo.EmployeePerformance;

IsQueryStoredProcedure: False

BypassPrepare: False

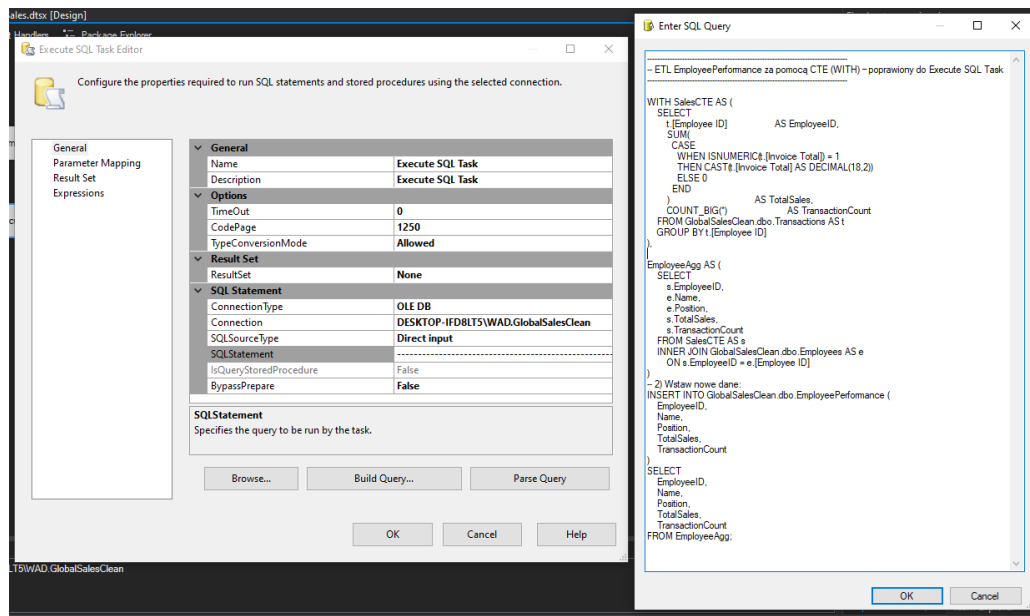
Name: Specifies the name of the task.

Browse... Build Query... Parse Query

OK Cancel Help

## „Load EmployeePerformance” (Execute SQL Task)

Co robi: jednym blokiem T-SQL (CTE + INSERT) agreguje dane z transakcji i wstawia je do EmployeePerformance.



## Wyniki EmployeePerformance

SQLQuery3.sql - DE...D8LT5\Student (62)) SQLQuery2.sql - DE...D8LT5\Student (61))

```

SELECT TOP (1000) [EmployeeID]
, [Name]
, [Position]
, [TotalSales]
, [TransactionCount]
FROM [GlobalSalesClean].[dbo].[EmployeePerformance]
  
```

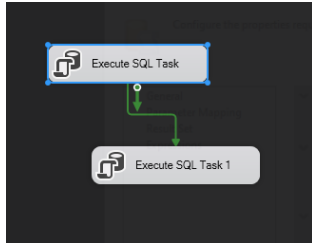
100 %

Results Messages

	EmployeeID	Name	Position	TotalSales	TransactionCount
1	5	Ryan Gross	Sales Associate	4019054.30	43748
2	6	Jeffery Carlson	Sales Associate	4031832.82	44428
3	7	Melissa Wilson	Sales Associate	4114263.54	44017
4	8	Edward Hicks	Sales Associate	4119978.62	44528
5	9	Shari Jordan	Sales Associate	4000387.48	43766
6	10	Mrs. Melissa Caldwell DDS	Sales Associate	4001504.40	43138
7	11	Regina Nelson	Sales Associate	4173457.07	44133
8	12	Kevin Dawson	Sales Associate	4022404.76	44056
9	13	Nicole Murphy	Sales Associate	4041029.45	43915
10	18	David Pitts	Sales Associate	5022135.47	54503
11	19	Lance Schultz	Sales Associate	5077100.62	54922
12	20	Caroline Hall	Sales Associate	4973984.37	53754
13	21	Heather Wilson	Sales Associate	4942808.30	54102
14	22	Allison Young	Sales Associate	4958086.47	54153
15	23	Kathleen Thomas	Sales Associate	5087826.56	54452
16	24	Carl Walters	Sales Associate	4931266.68	53878

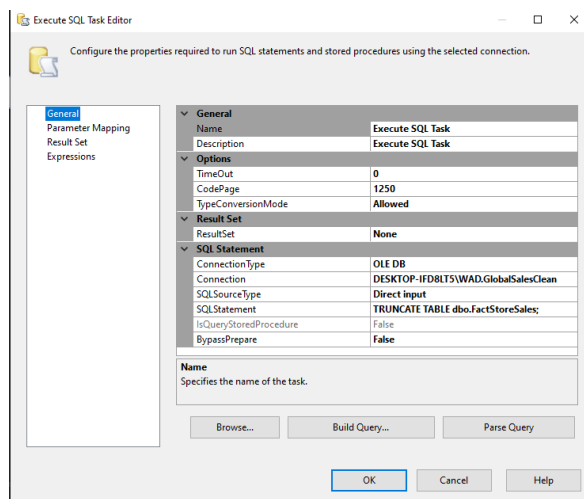
## Store Sales (etl\_StoreSales.dtsx)

– Co robi: Również jednym Execute SQL Task/CTE (WITH StoreCTE ...) agregujemy TransactionsClean po StoreID (suma InvoiceTotal, liczba transakcji) i wstawiamy do FactStoreSales(StoreID, TotalSales, TxCount).



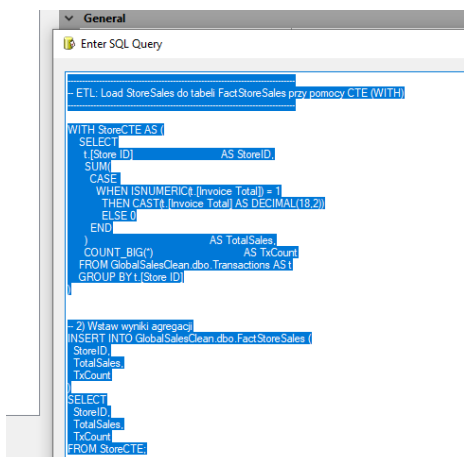
### Execute SQL Task

Wykonuje TRUNCATE TABLE GlobalSalesClean.dbo.FactStoreSales; Czyści starą zawartość faktu sprzedaży sklepów, żeby przy każdym uruchomieniu dane były ładowane od nowa.



### Execute SQL Task 1

Używa CTE (WITH StoreCTE AS (...)) do agregacji transakcji z tabeli TransactionsClean:StoreID – klucz sklepu. TotalSales – suma wartości paragonów (rzuconych na DECIMAL(18,2) z uwzględnieniem tylko tych wierszy, gdzie Invoice Total jest numeryczne). TxCount – liczba transakcji (COUNT\_BIG(\*)). Następnie INSERT INTO GlobalSalesClean.dbo.FactStoreSales wybiera z StoreCTE i wstawia te trzy kolumny do tabeli faktów.





## Wyniki SalesStore

SQLQuery4.sql - DE...D8LT5\Student (57)) SQLQuery3.sql - DE...D8LT5\Student (62))

```

SELECT TOP (1000) [StoreID]
, [TotalSales]
, [TxCount]
FROM [GlobalSalesClean].[dbo].[FactStoreSales]

```

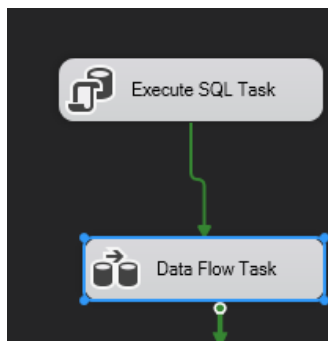
100 %

	StoreID	TotalSales	TxCount
1	1	36523912.44	395729
2	2	34993208.47	379764
3	3	18258820.11	196557
4	4	23758551.25	256042
5	5	15492174.85	167795
6	6	3516519.35	7893
7	7	2566884.95	5644
8	8	3544817.25	7789
9	9	2967433.00	6419
10	10	1838247.35	4024
11	11	17518801.11	204163
12	12	10371919.71	120881
13	13	7166692.50	84253
14	14	8001131.59	93958
15	15	4802106.38	55969
16	16	10041153.74	160477

## Discount Effectiveness (DiscountEffectiveness.dtsx)

Pakiet DiscountEffectiveness.dtsx składa się z dwóch etapów w Control Flow oraz jednego prostego Data Flow, w obliczamy kluczowe metryki dla transakcji z uwzględnieniem tego, czy były objęte rabatem.

Control Flow

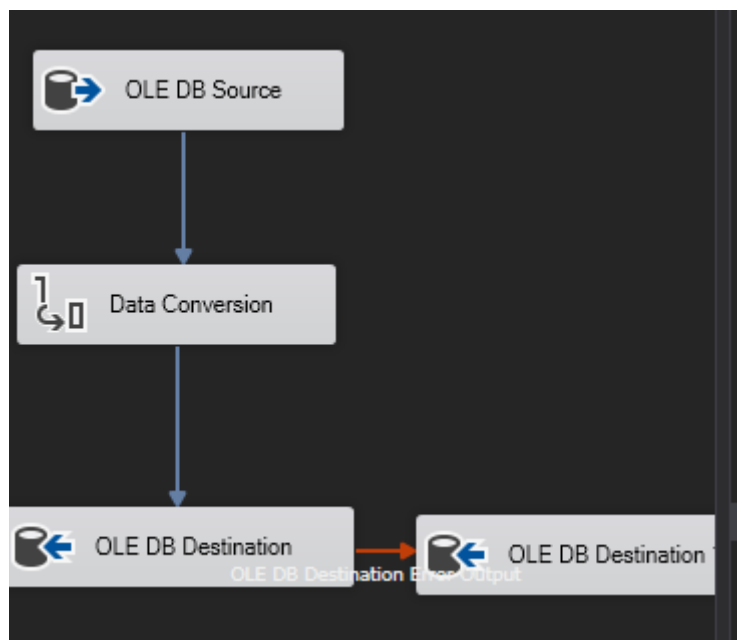


Execute SQL Task czyści tabelę docelową, aby każdy przebieg ETL wstawiał dane od nowa

General	
Name	Execute SQL Task
Description	Execute SQL Task
Options	
TimeOut	0
CodePage	1250
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	DESKTOP-IFD8LT5\WAD.GlobalSalesClean
SQLSourceType	Direct input
SQLStatement	TRUNCATE TABLE dbo.DiscountEffectiveness;
IsQueryStoredProcedure	False
BypassPrepare	False

## Data Flow Task

Po jego uruchomieniu przechodzimy do zakładki Data Flow



### OLE DB Source

Co robi: w locie tworzy flagę (DiscountFlag) = 'WithDiscount' lub 'NoDiscount', a następnie dla każdej z tych dwóch grup liczy: sumę wartości transakcji (TotalSales), średnią wartość paragonu (AverageSale), liczbę transakcji (TransactionCount).

OLE DB connection manager:  
DESKTOP-IFD8LT5\WAD.GlobalSalesClean

Data access mode:  
SQL command

SQL command text:

```
SELECT
CASE
  WHEN TRY_CAST([Discount] AS DECIMAL(18,2)) = 0
    OR [Discount] IS NULL
  THEN 'NoDiscount'
  ELSE 'WithDiscount'
END AS DiscountFlag,
SUM(TRY_CAST([Invoice Total] AS DECIMAL(18,2))) AS TotalSales,
AVG(TRY_CAST([Invoice Total] AS DECIMAL(18,2))) AS AverageSale,
COUNT(*) AS TransactionCount
FROM dbo.transactions
GROUP BY
CASE
  WHEN TRY_CAST([Discount] AS DECIMAL(18,2)) = 0
    OR [Discount] IS NULL
  THEN 'NoDiscount'
  ELSE 'WithDiscount'
END;
```

Parameters...  
Build Query...  
Browse...  
Parse Query

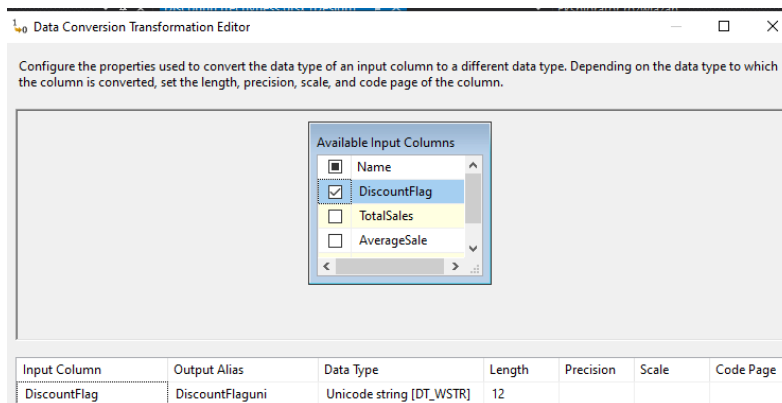
## Data Conversion

Input Column: DiscountFlag

Output Alias: DiscountFlaguni

Data Type: Unicode string [DT\_WSTR], długość 12

Co robi: konwertuje tekstową flagę na Unicode, by pasowała do docelowej kolumny NVARCHAR w tabeli DiscountEffectiveness.



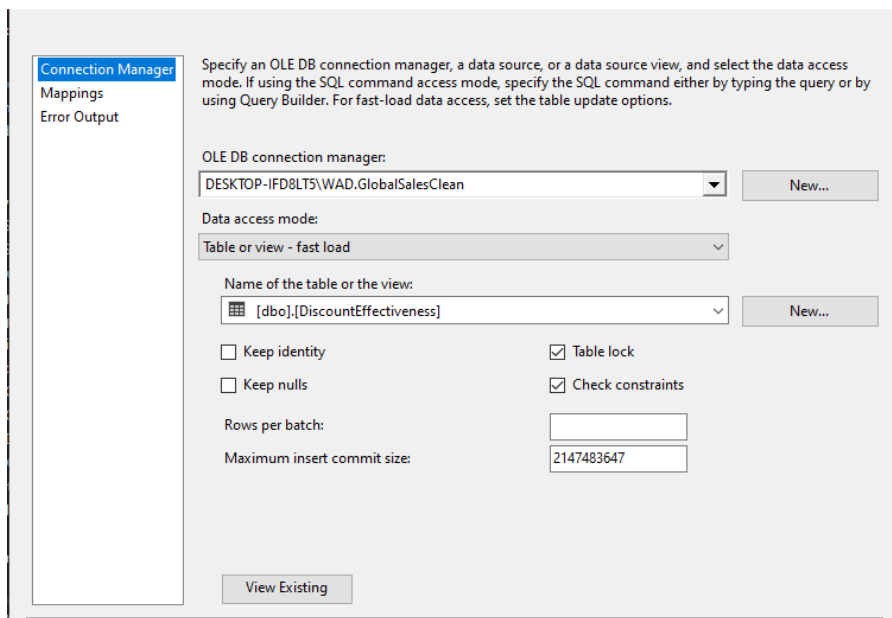
## OLE DB Destination

Connection Manager: GlobalSalesClean

Data access mode: Table or view – fast load

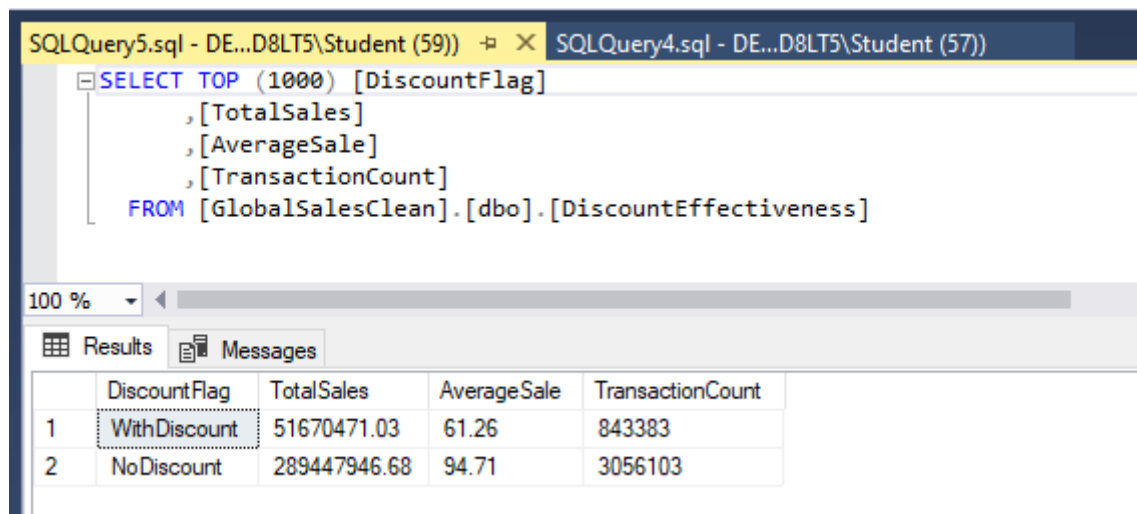
Name of the table or the view: dbo.DiscountEffectiveness

Mapowanie zadziała automatycznie.



Error Output: przekierowuje ewentualne wiersze z błędami (np. niezgodne długości, problemy z rzutowaniem) do drugiego OLE DB Destination, który ładuje je do tabeli dbo.Errors w bazie GlobalSales.

## Wyniki DiscountEffectivness



The screenshot shows a SQL query window with the following query:

```
SELECT TOP (1000) [DiscountFlag]
, [TotalSales]
, [AverageSale]
, [TransactionCount]
FROM [GlobalSalesClean].[dbo].[DiscountEffectiveness]
```

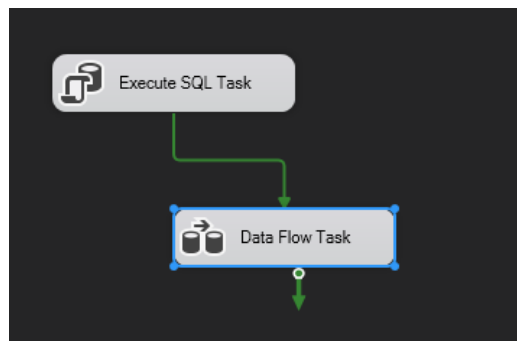
The results are displayed in a table with the following columns: DiscountFlag, TotalSales, AverageSale, and TransactionCount.

	DiscountFlag	TotalSales	AverageSale	TransactionCount
1	WithDiscount	51670471.03	61.26	843383
2	NoDiscount	289447946.68	94.71	3056103

## TxByHour (etl\_TxByHour.dtsx)

Pakiet etl\_TxByHour.dtsx zbiera dane o transakcjach zagregowane według daty i godziny, po czym ładuje wynik do tabeli faktu FactTransactionsByHour. Całość składa się z dwóch kroków w Control Flow oraz prostego Data Flow:

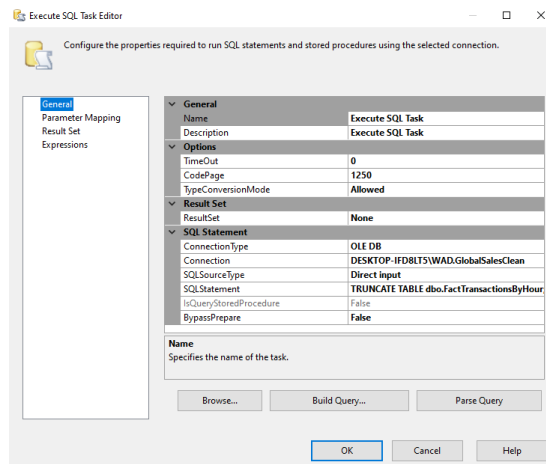
### Control Flow



Execute SQL Task czyści tabelę faktu, aby przy każdym uruchomieniu ładować dane od zera.

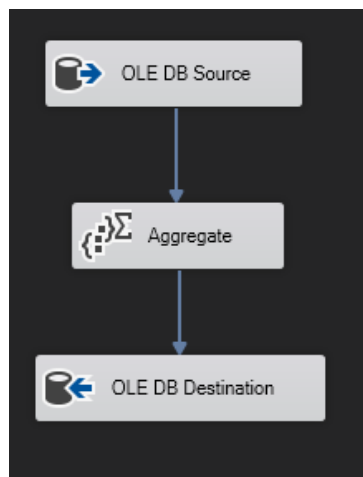
SQLStatement:

TRUNCATE TABLE GlobalSalesClean.dbo.FactTransactionsByHour;



Data Flow Task („Load Transactions By Hour”)

Pod spodem znajduje się strumień Data Flow, który wykonuje wszystkie etapy transformacji.



## OLE DB Source

Connection Manager: GlobalSalesClean

Data access mode: SQL command

TxDate – data transakcji bez czasu,

TxHour – godzina (0–23) wyciągnięta z daty.

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:  
DESKTOP-IFD8LTS\WAD.GlobalSalesClean New...

Data access mode:  
SQL command

SQL command text:

```
SELECT  
  CAST([Date] AS DATE)      AS TxDate,  
  DATEPART(HOUR, [Date])   AS TxHour  
FROM dbo.transactions;
```

Parameters...  
Build Query...  
Browse...  
Parse Query

## Aggregate

Group By:

TxDate → Output Alias: TxDate

TxHour → Output Alias: TxHour

Aggregation:

COUNT(\*) → Output Alias: TxCount, zlicza, ile transakcji było w każdym połączeniu (TxDate, TxHour).

Available Input Columns

Input Column	Output Alias	Operation	Compa
TxDate	TxDate	Group by	
TxHour	TxHour	Group by	
(*)	TxCount	Count all	

## OLE DB Destination

Connection Manager: GlobalSalesClean

Data access mode: Table or view – fast load

Table: dbo.FactTransactionsByHour

**Connection Manager**  
Mappings  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:  
 DESKTOP-IFD8LT5\WAD.GlobalSalesClean New...

Data access mode:  
 Table or view - fast load

Name of the table or the view:  
 [dbo].[FactTransactionsByHour] New...

☐ Keep identity ☒ Table lock  
☐ Keep nulls ☒ Check constraints

Rows per batch:

Maximum insert commit size:

## Mappings:

Input Column	Destination Column
TxDate	TxDate
TxHour	TxHour
TxCount	TxCount

## Wynik TxByHour

SQLQuery1.sql - DE...D8LT5\Student (58)

```

SELECT TOP (1000) [TxDate]
, [TxHour]
, [TxCount]
FROM [GlobalSalesClean].[dbo].[FactTransactionsByHour]
  
```

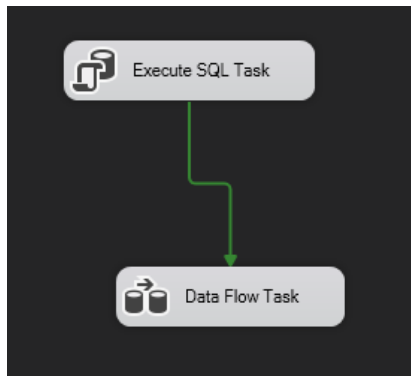
100 %

Results Messages

	TxDate	TxHour	TxCount
1	2023-01-01	0	672
2	2023-01-01	8	699
3	2023-01-01	9	693
4	2023-01-01	10	643
5	2023-01-01	11	696
6	2023-01-01	12	647
7	2023-01-01	13	693
8	2023-01-01	14	696
9	2023-01-01	15	721
10	2023-01-01	16	1628

## Return Rate (etl\_ReturnRate.dtsx)

Pakiet etl\_ReturnRate.dtsx liczy współczynnik zwrotów (stosunek liczby transakcji zwrotowych do wszystkich transakcji) i ładuje wynik do tabeli faktu FactReturnRate. Składa się z dwóch kroków w Control Flow oraz czterech transformacji w Data Flow:



Control Flow

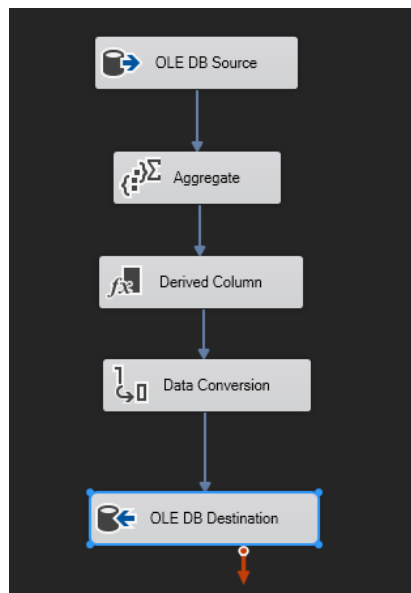
Execute SQL Task SQLStatement:

TRUNCATE TABLE GlobalSalesClean.dbo.FactReturnRate; Czyści starą zawartość tabeli faktu zwrotów.

General	
Name	Execute SQL Task
Description	Execute SQL Task
Options	
Timeout	0
CodePage	1250
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	DESKTOP-IFD8LT5\WAD.GlobalSalesClean
SQLSourceType	Direct input
SQLStatement	TRUNCATE TABLE dbo.FactReturnRate;
IsQueryStoredProcedure	False
BypassPrepare	False



Data Flow Task uruchamia poniższy proces Data Flow.



OLE DB Source

Data access mode: SQL command – Dla każdej transakcji zwraca 1 (zwrot) lub 0 (sprzedaż).

**Connection Manager**  
Columns  
Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:  
DESKTOP-IFD8LT5\WAD.GlobalSalesClean New...

Data access mode:  
SQL command

SQL command text:  

```
SELECT
CASE WHEN CAST([Invoice Total] AS DECIMAL(18,2)) < 0 THEN 1
ELSE 0 END AS IsReturn
FROM dbo.transactions;
```

Parameters...  
Build Query...  
Browse...  
Parse Query

## Aggregate

Group by: (brak kluczy) → zwraca pojedynczy wiersz.

Operations:

SUM(IsReturn) → ReturnCount

COUNT(\*) → TotalCount

Aggregations Advanced

Configure the properties used to perform group by operations and to calculate aggregate values. Optionally, apply comparison options to the operation. To configure multiple group by operations, click Advanced.

Advanced

Available Input Co...

- ☒ Name
- ☒ (\*)
- ☒ IsReturn

Input Column	Output Alias	Operation	Compa
IsReturn	ReturnCount	Sum	
(*)	TotalCount	Count all	

## Derived Column

Tworzy nowe pole ReturnRate\_Dbl – surowy współczynnik jako zmiennoprzecinkowa.

Variables and Parameters

Columns

Mathematical Functions

String Functions

Date/Time Functions

NULL Functions

Type Casts

Operators

Description:

Derived Column Name	Derived Column	Expression	Data Type	Le
ReturnRate_Dbl	<add as new column>	(DT_R8)ReturnCount / (DT_R8)TotalCount	double-precision float ...	

### Data Conversion

Konwertuje ReturnRate\_Dbl na typ numeric [DT\_NUMERIC] (precision = 5, scale = 4) → ReturnRate\_dec.

Configure the properties used to convert the data type of an input column to a different data type. Depending on the data type to which the column is converted, set the length, precision, scale, and code page of the column.

Available Input Columns

☒ Name  
☐ ReturnCount  
☐ TotalCount

Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
ReturnRate_Dbl	ReturnRate_dec	numeric [DT_NUMERIC]		5	4	

### OLE DB Destination

Table or view – fast load → dbo.FactReturnRate

Connection Manager

Mappings

Error Output

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:  
DESKTOP-IFD8LTS\WAD.GlobalSalesClean New...

Data access mode:  
Table or view - fast load

Name of the table or the view:  
[dbo].[FactReturnRate] New...

☐ Keep identity ☒ Table lock  
☐ Keep nulls ☒ Check constraints

Rows per batch:

Maximum insert commit size:

Mapping:

Available Input Col...

Available Destin...

ReturnCount

TotalCount

ReturnRate\_Dbl

ReturnCount

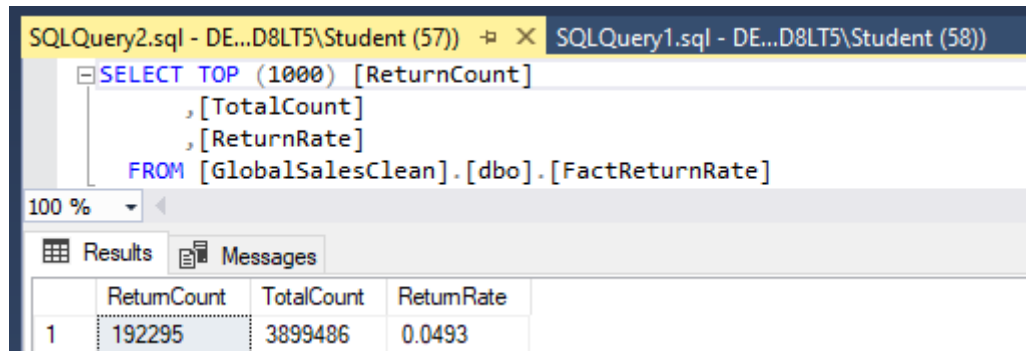
TotalCount

ReturnRate

Input Column	Destination Column
ReturnCount	ReturnCount
TotalCount	TotalCount
ReturnRate_dec	ReturnRate

Wyniki ReturnRate

W efekcie w tabeli FactReturnRate mamy jeden rekord z trzema kolumnami



The screenshot shows a SQL query window with the following text:

```
SELECT TOP (1000) [ReturnCount]
, [TotalCount]
, [ReturnRate]
FROM [GlobalSalesClean].[dbo].[FactReturnRate]
```

Below the query, the 'Results' tab is active, displaying a single row of data:

	ReturnCount	TotalCount	ReturnRate
1	192295	3899486	0.0493

## FactAvgBasketByCity (etl\_FactAvgBasketByCity.dtsx)

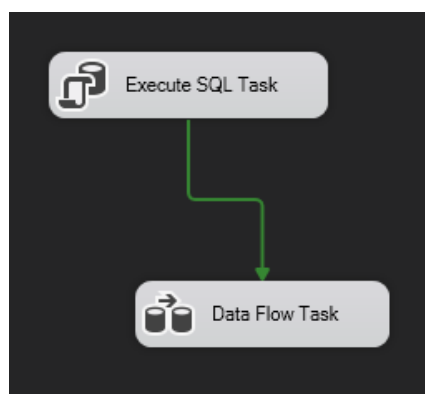
Pakiet etl\_FactAvgBasketByCity.dtsx buduje tabelę faktu FactAvgBasketByCity, w której dla każdego miasta policzona jest:

AvgBasket – średnia wartość koszyka (średnia wartość paragonu)

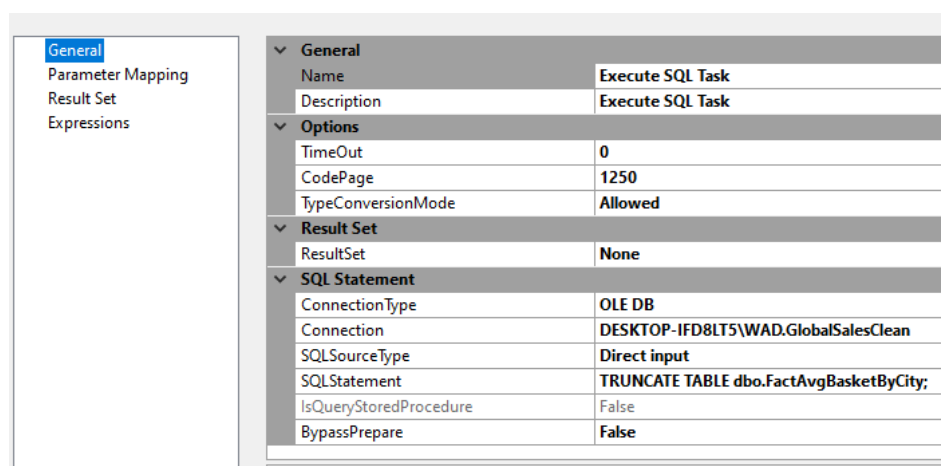
TxCount – liczba transakcji

Cały proces składa się z dwóch zadań w Control Flow oraz jednego Data Flow:

Control Flow



Execute SQL Task – opróżnia tabelę faktu, żeby uniknąć duplikatów i ładować świeże dane.

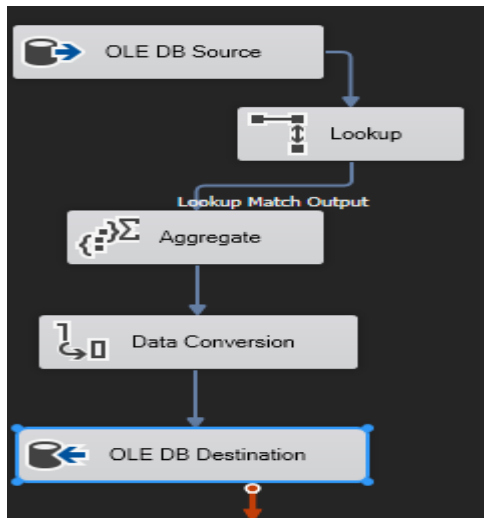


The screenshot shows the 'General' tab of the 'Execute SQL Task' configuration. The configuration is as follows:

General	
Name	Execute SQL Task
Description	Execute SQL Task
Options	
TimeOut	0
CodePage	1250
TypeConversionMode	Allowed
Result Set	
ResultSet	None
SQL Statement	
ConnectionType	OLE DB
Connection	DESKTOP-IFD8LT5\WAD.GlobalSalesClean
SQLSourceType	Direct input
SQLStatement	TRUNCATE TABLE dbo.FactAvgBasketByCity;
IsQueryStoredProcedure	False
BypassPrepare	False

Data Flow Task – Po zakończeniu Execute SQL Task uruchamia się Data Flow, w którym przebieg jest następujący.

Data Flow



OLE DB Source

Connection Manager: WAD.GlobalSalesClean

Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder.

OLE DB connection manager:  
DESKTOP-IFD8LT5\WAD.GlobalSalesClean New...

Data access mode:  
SQL command

SQL command text:  

```
SELECT
  [Customer ID],
  CAST([Invoice Total] AS DECIMAL(18,2)) AS Sale
FROM dbo.transactions;
```

Parameters...  
Build Query...  
Browse...  
Parse Query

Preview...

Pobiera z TransactionsClean dwie kolumny: identyfikator klienta oraz wartość paragonu (rzutowaną na DECIMAL).

## Lookup

Connection Manager: WAD.GlobalSalesClean

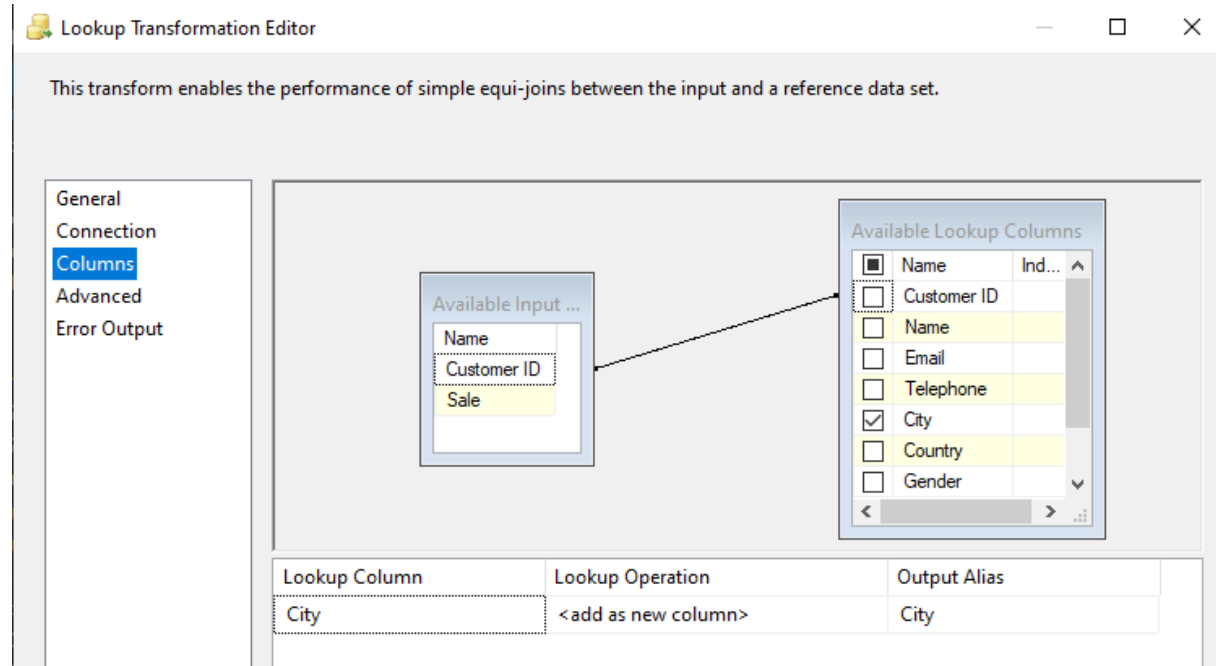
Reference table: dbo.Customers

Join key:

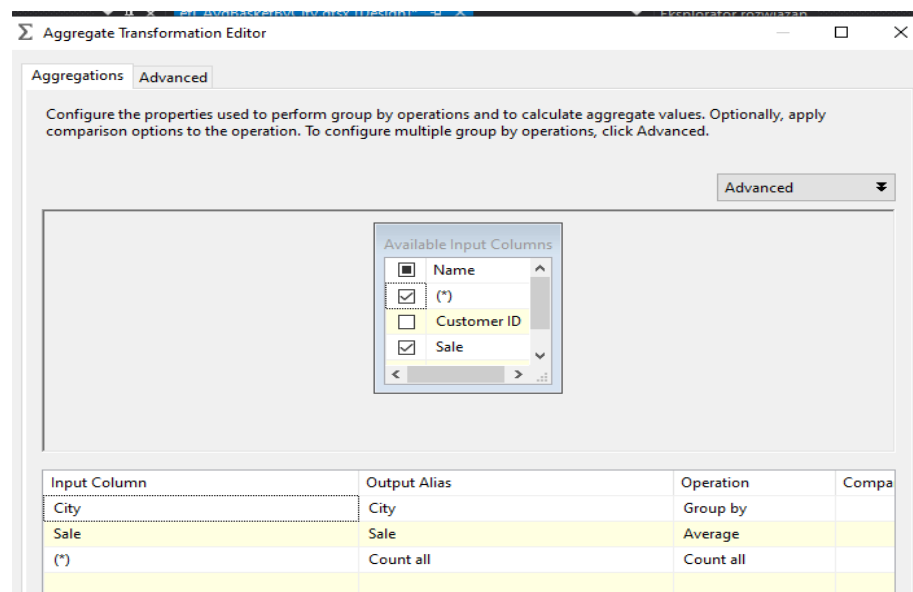
Input Column: Customer ID

Lookup Column: CustomerID

Dołącza do każdego wiersza transakcji nazwę miasta, w którym klient robił zakupy.



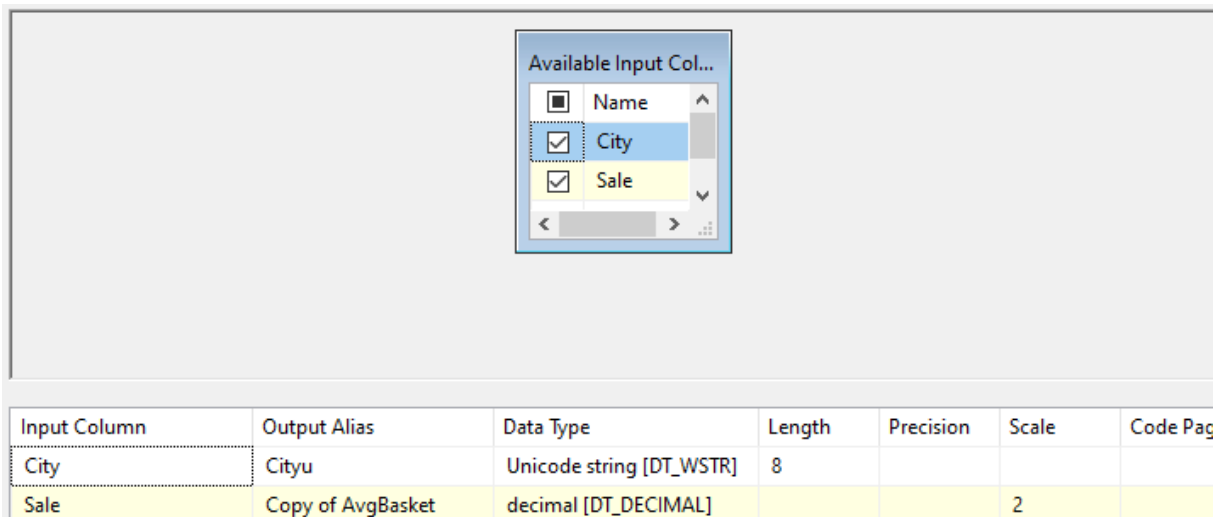
Aggregate - dla każdego miasta oblicza średnią wartość koszyka i łączną liczbę transakcji.



Data Conversion - konwertuje wyjściowe kolumny do typów zgodnych ze schematem tabeli docelowej.

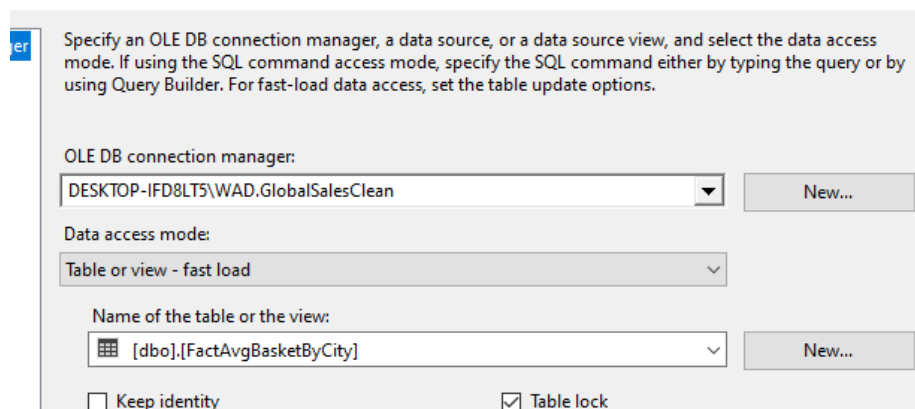
Input Column: City → Output Alias: Cityu → Data Type: Unicode string (DT\_WSTR), Length = 50

Input Column: AvgBasket → Output Alias: AvgBasket\_dec → Data Type: Decimal (DT\_DECIMAL), Precision = 18, Scale = 2



Input Column	Output Alias	Data Type	Length	Precision	Scale	Code Page
City	Cityu	Unicode string [DT_WSTR]	8			
Sale	Copy of AvgBasket	decimal [DT_DECIMAL]			2	

OLE DB Destination - Ładuje przetworzone w Data Flow dane do tabeli faktu FactAvgBasketByCity.



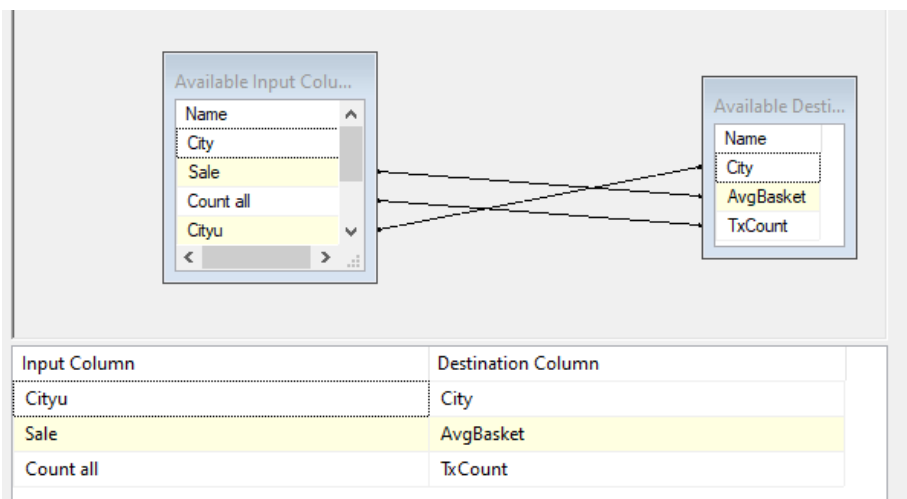
Specify an OLE DB connection manager, a data source, or a data source view, and select the data access mode. If using the SQL command access mode, specify the SQL command either by typing the query or by using Query Builder. For fast-load data access, set the table update options.

OLE DB connection manager:

Data access mode:

Name of the table or the view:

☐ Keep identity ☒ Table lock



Input Column	Destination Column
Cityu	City
Sale	AvgBasket
Count all	TxCount

Wyniki FactAvgBasketByCity

```

SELECT TOP (1000) [City]
, [AvgBasket]
, [TxCount]
FROM [GlobalSalesClean].[dbo].[FactAvgBasketByCity]

```

100 %



Results



Messages

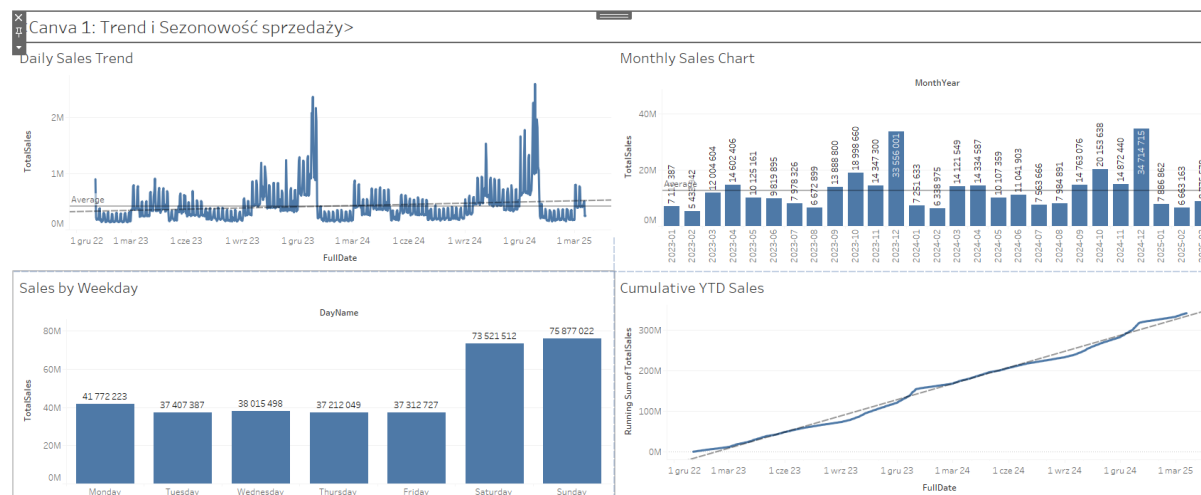
	City	AvgBasket	TxCount
301	Oullins	82.10	1118
302	Ovar	84.42	280
303	Oviedo	70.43	7
304	PÃ¶rols	69.47	80
305	Paiporta	86.00	510
306	Palmela	86.95	156
307	Paredes	79.04	29
308	Paris	85.11	58860
309	Pasadena	89.27	5978
310	Patema	80.57	529
311	Paterson	97.60	6077
312	Pearland	90.19	4454
313	Penacova	87.98	665
314	Pessac	95.20	43
315	Phoenix	91.80	93925
316	Planegg	84.35	1041
317	Pollensa	94.87	72
318	Pombal	88.92	1107
319	Portillo	63.89	20
320	Porto	85.05	56420
321	Potsdam	84.28	2408
322	Pulheim	81.02	1121
323	Quincy	83.21	93
324	Ratingen	94.43	133
325	Redmond	102.72	247
326	Renton	84.31	262
327	RezÃ©	85.79	123



# Wizualizacja

W programie Tableau dokonaliśmy wizualizacji danych.

## Canva nr 1 Trend i Sezonowość sprzedaży



## Wykres nr 1

Opis wykresu

Tytuł: Daily Sales Trend

Typ wizualizacji: Continuous Line Chart

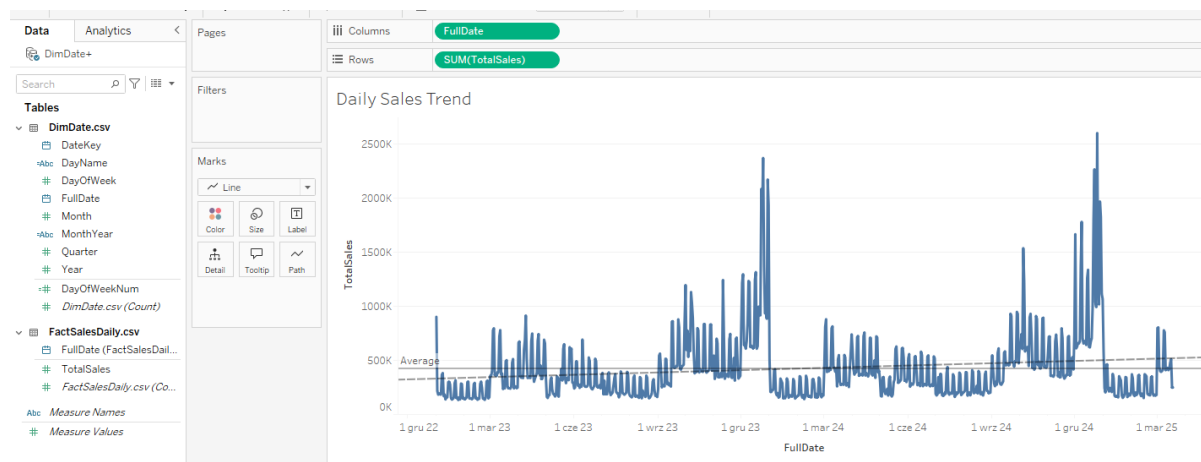
Co pokazuje:

Oś X: kolejne daty (FullDate) w zakresie od grudnia 2022 do marca 2025.

Oś Y: dzienna suma wartości sprzedaży (SUM(TotalSales)).

Szara przerywana linia to linia średniej (Average) wartości sprzedaży.

Czarna linia (z kropkami) to trend regresji liniowej (Trend Line), pokazujący ogólny kierunek zmian.



Dane źródłowe

Wczytałeś do Tableau dwa pliki CSV (ekstrakty):

DimDate.csv (DimDate)

Pola: FullDate (Date), Year, Month, DayOfWeek, Quarter ...

FactSalesDaily.csv (FactSalesDaily)

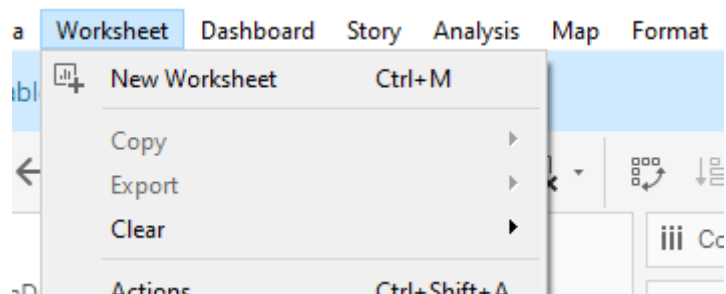
Pola: FullDate (Date), TotalSales (Decimal)

Na karcie Data Source połączyłeś je po kolumnie FullDate.

Budowa wykresu – krok po kroku

Nowy arkusz

Kliknij ikonę New Worksheet → nazwij Daily Sales Trend.



## Columns

Przeciągnij wymiar FullDate (z FactSalesDaily lub DimDate) na półkę Columns.

Kliknij zieloną pigułkę FullDate, wybierz:

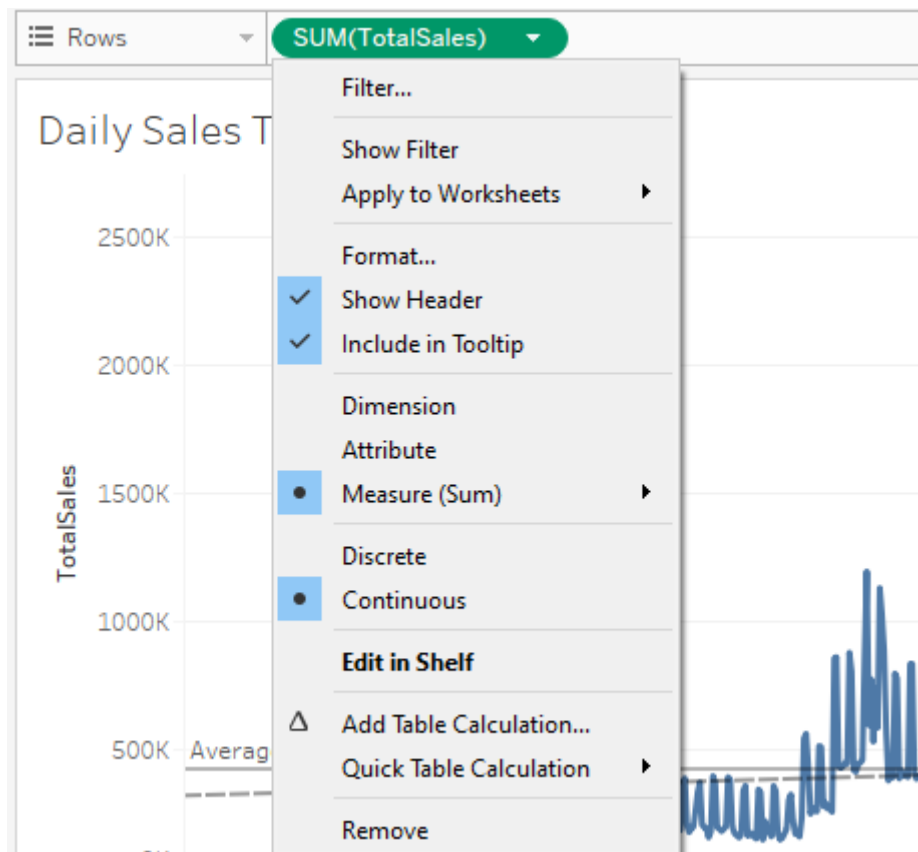
Exact Date (żeby uwzględnić każdy dzień).

The screenshot shows the Tableau Desktop interface. On the left, the 'Data' pane displays two tables: 'DimDate.csv' and 'FactSalesDaily.csv'. The 'Columns' shelf at the top right contains the 'FullDate' dimension. A context menu is open for 'FullDate', showing options like 'Filter...', 'Show Filter', 'Apply to Worksheets', 'Format...', 'Show Header', 'Include in Tooltip', 'Show Missing Values', 'Extend Date Range', 'Standard Gregorian', 'ISO-8601 Week-Based', 'Year', 'Quarter', 'Month', 'Day', 'More', 'Exact Date', 'Attribute', and 'Measure'. The 'Exact Date' option is selected. The main view shows a line chart titled 'Daily Sales T' with 'TotalSales' on the y-axis and dates on the x-axis. The status bar at the bottom indicates '808 marks', '1 row by 1 column', and 'SUM(TotalSales): 341 118 418'.

Rows

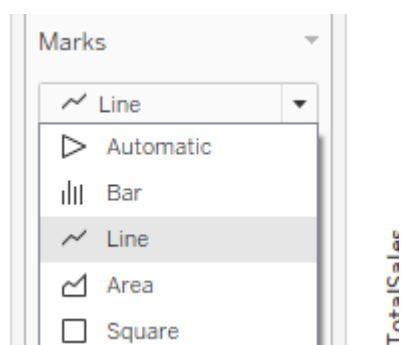
Przeciągnij miarę TotalSales na półkę Rows.

Tableau automatycznie agreguje SUM(TotalSales).



Marks

W panelu Marks wybierz Line z listy (zamiast Automatic).



### Reference Line – Average

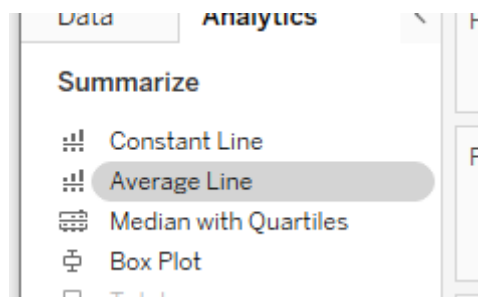
Przejdź do zakładki Analytics (obok Data).

Rozwiń Average Line, przeciągnij na wykres → upuść na Table.

W ustawieniach linii:

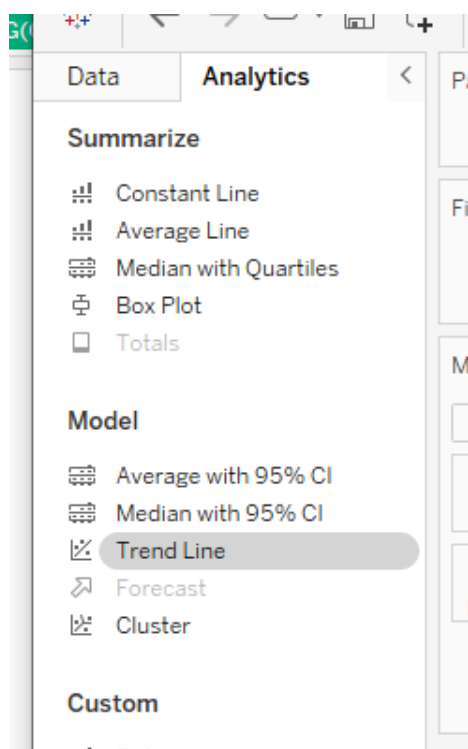
Label: Value (wyświetli „Average = 512 345”).

Styl: np. przerywana, kolor szary.



### Trend Line

W Analytics → rozwiń Trend Line → przeciągnij Linear na wykres (ponownie na Table).



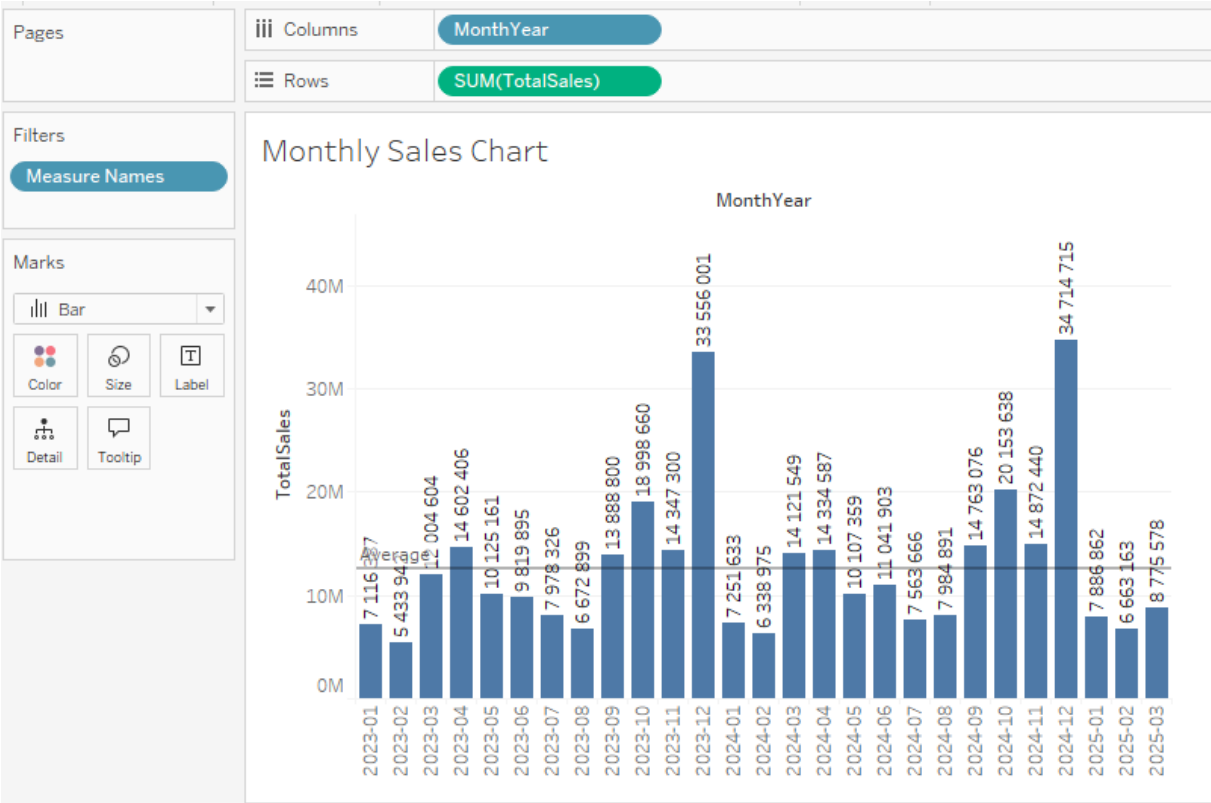
Dlaczego ważne: Umożliwia identyfikację krótkoterminowych pików i dołków w aktywności klientów oraz analizę sezonowości i nietypowych zdarzeń.

# Wykres nr 2

## Monthly Sales Chart

Co pokazuje: Słupki sumy sprzedaży w każdym miesiącu (grudzień 2022 – marzec 2025).

Dlaczego ważne: Ułatwia porównanie obrotów miesiąc do miesiąca i śledzenie długoterminowego trendu sezonowego.

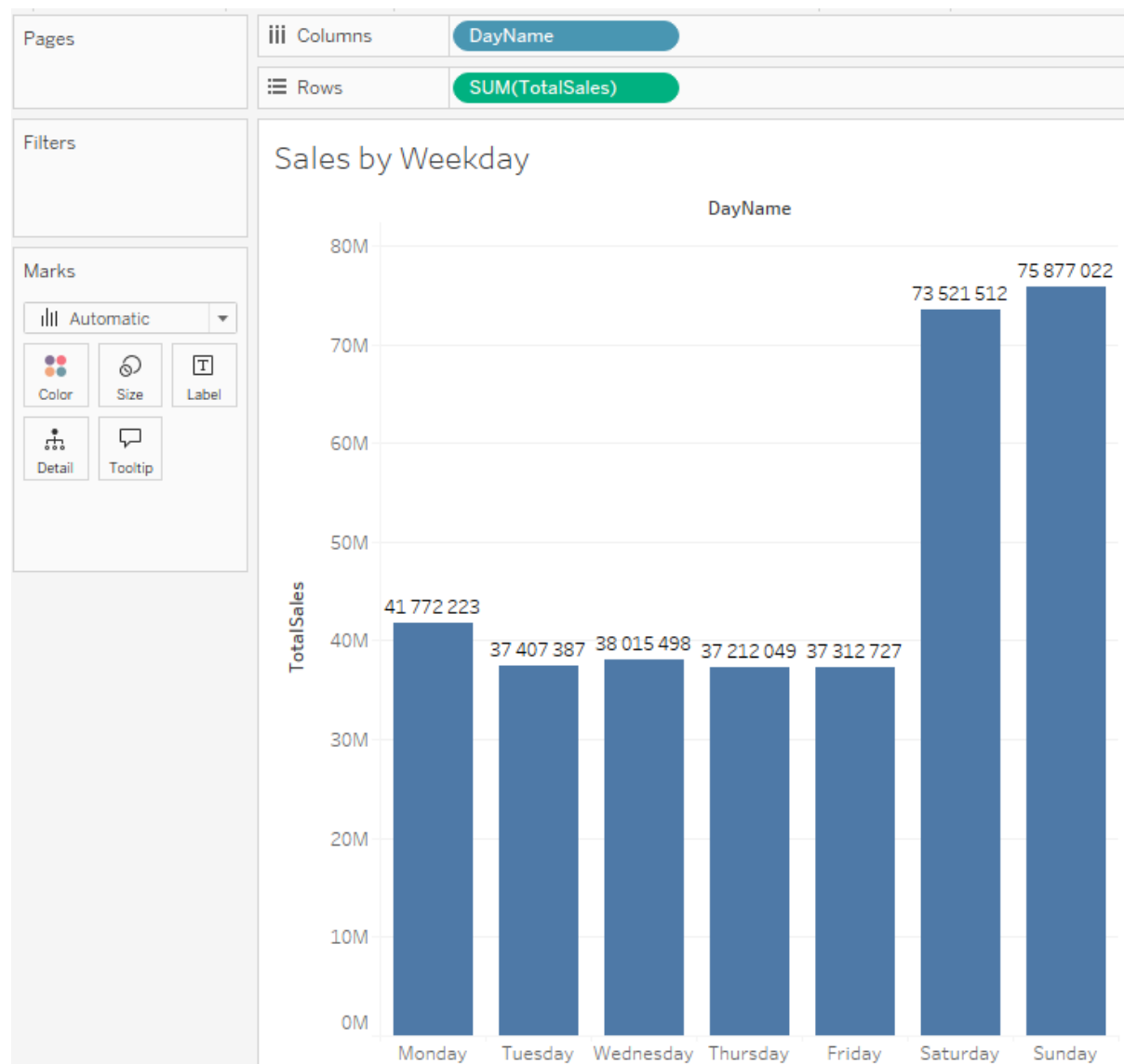


### Wykres nr 3

#### Sales by Weekday

Co pokazuje: Łączną wartość sprzedaży rozbity na dni tygodnia (poniedziałek–niedziela).

Dlaczego ważne: Pokazuje, w które dni tygodnia klienci najczęściej robią zakupy, co pozwala optymalizować personel i godziny otwarcia.

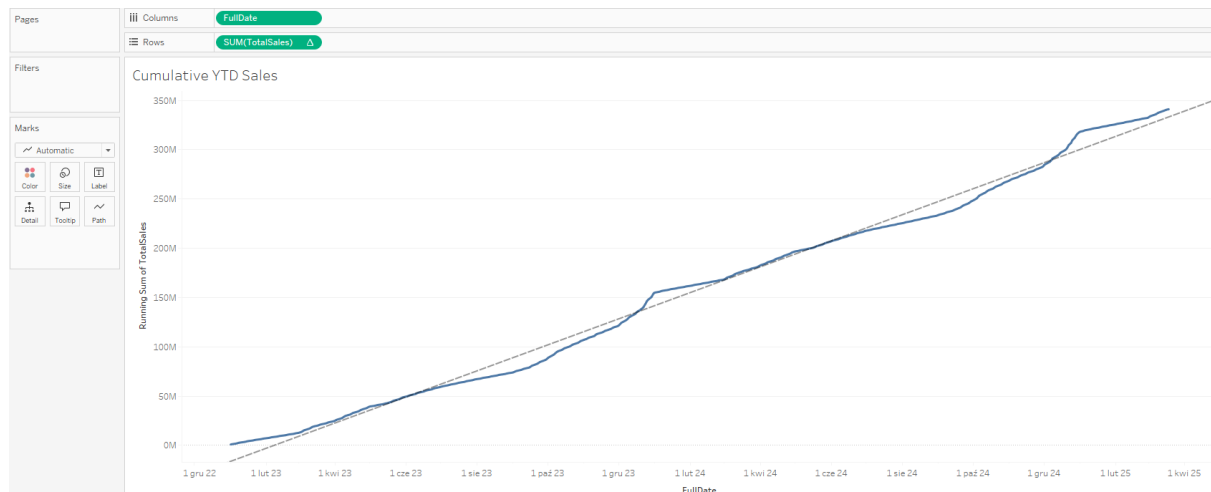


## Wykres nr 4

### Cumulative YTD Sales

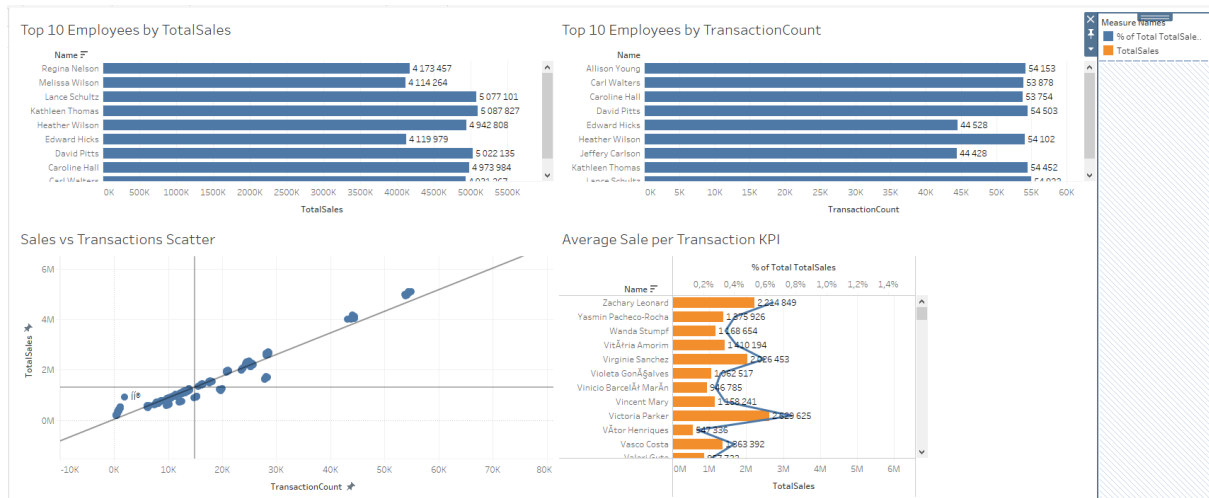
Co pokazuje: Narastającą sumę sprzedaży od początku roku do każdej daty (running total), z linią idealnego, „prostoliniowego” wzrostu.

Dlaczego ważne: Pozwala ocenić, jak rzeczywisty przebieg sprzedaży wypada na tle idealnego równomiernego wzrostu oraz wykryć przyspieszenia i spowolnienia w trakcie roku.





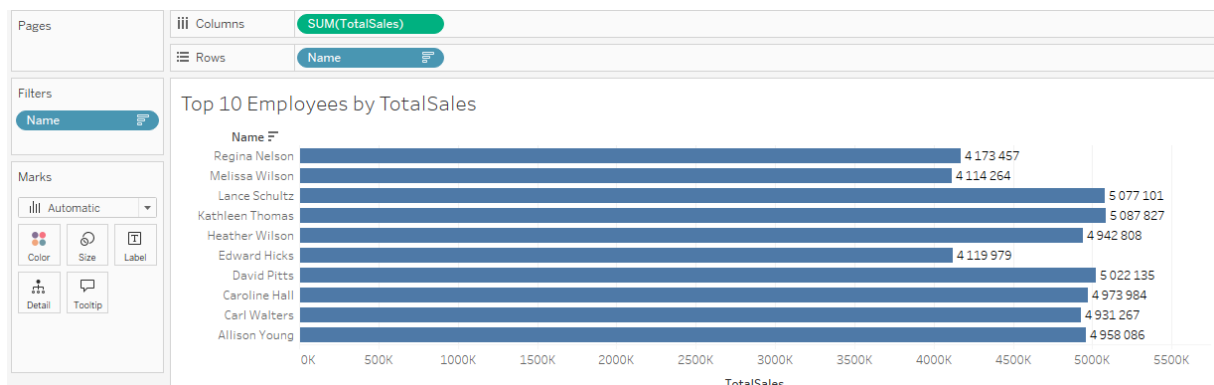
## Canva nr 2 Wydajność pracowników



## Wykre nr 1

### Top 10 Employees by TotalSales

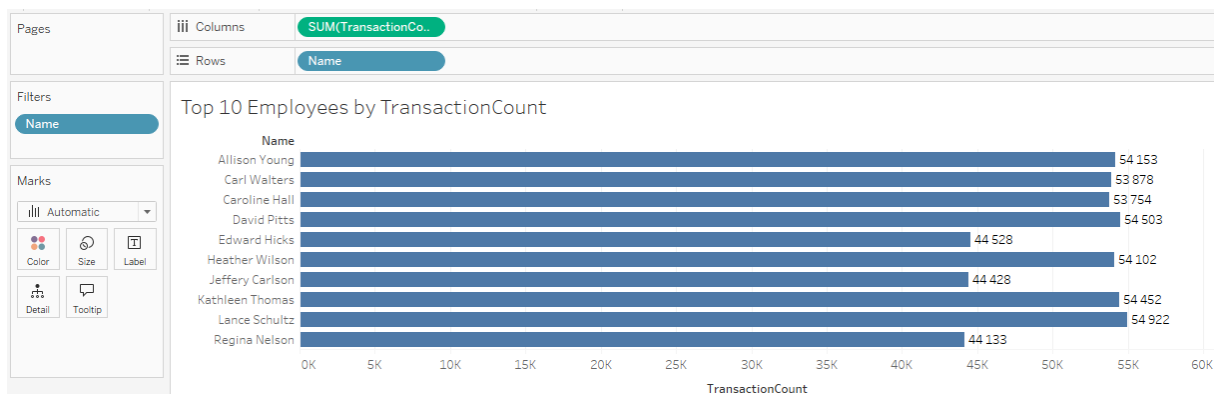
Co pokazuje: poziomy wykres słupkowy z dziesięcioma pracownikami o najwyższej łącznej wartości sprzedaży. Pokazuje, kto wygenerował największy przychód dla firmy.



## Wykres nr 2

### Top 10 Employees by TransactionCount

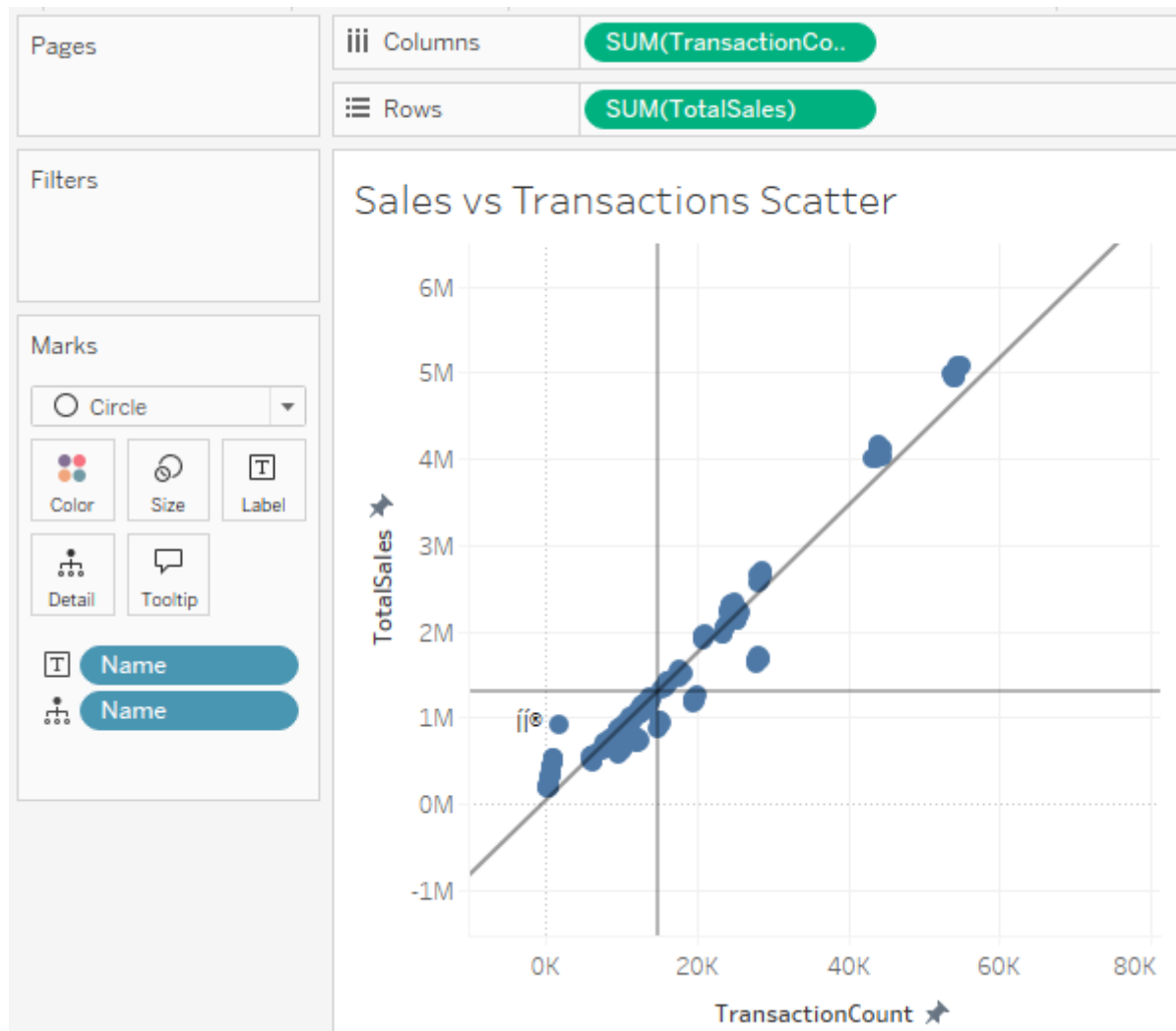
Co pokazuje: poziomy wykres słupkowy przedstawiający dziesięciu pracowników z największą liczbą transakcji. To prosty sposób, by zobaczyć najbardziej aktywnych sprzedawców.



### Wykres nr 3

#### Sales vs Transactions Scatter

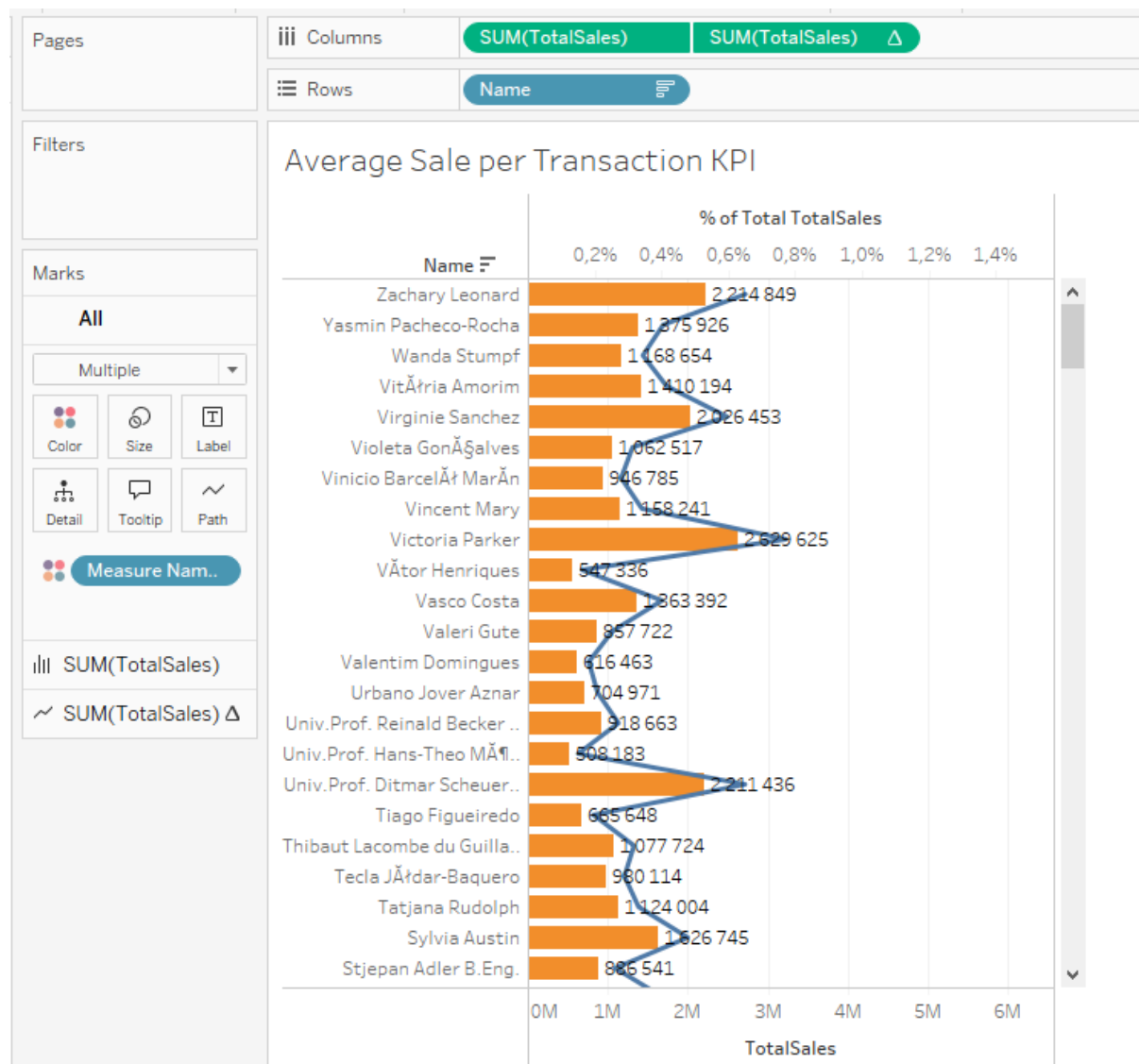
Co pokazuje: punktowy rozrzut pracowników według liczby transakcji (oś X) i łącznej wartości sprzedaży (oś Y), z poziomą i pionową linią średnich oraz linią trendu regresji. Ułatwia identyfikację osób o wysokiej efektywności (wysokie X i Y) oraz tych, którzy często sprzedają, ale nisko marżowo, i odwrotnie.



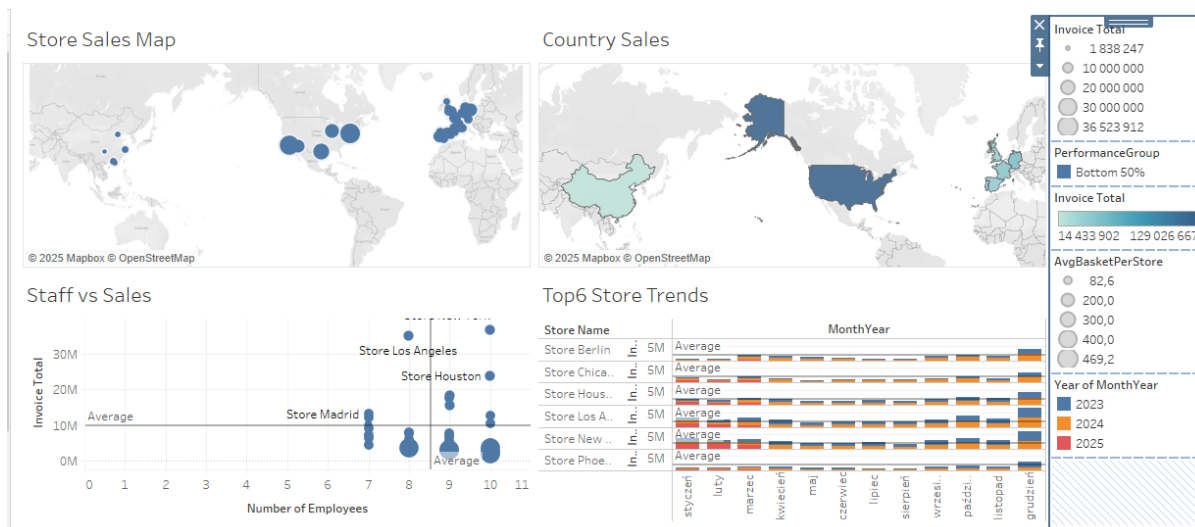
## Wykres nr 4

### Average Sale per Transaction KPI

Co pokazuje: dla każdego pracownika słupki z łączną wartością sprzedaży oraz nakładającą się linię Pareto (kumulatywny % udziału w całości obrotu). Pozwala zobaczyć, którzy sprzedawcy generują największe pojedyncze transakcje i jaki procent całkowitej sprzedaży odpowiadają najlepsi 20%.



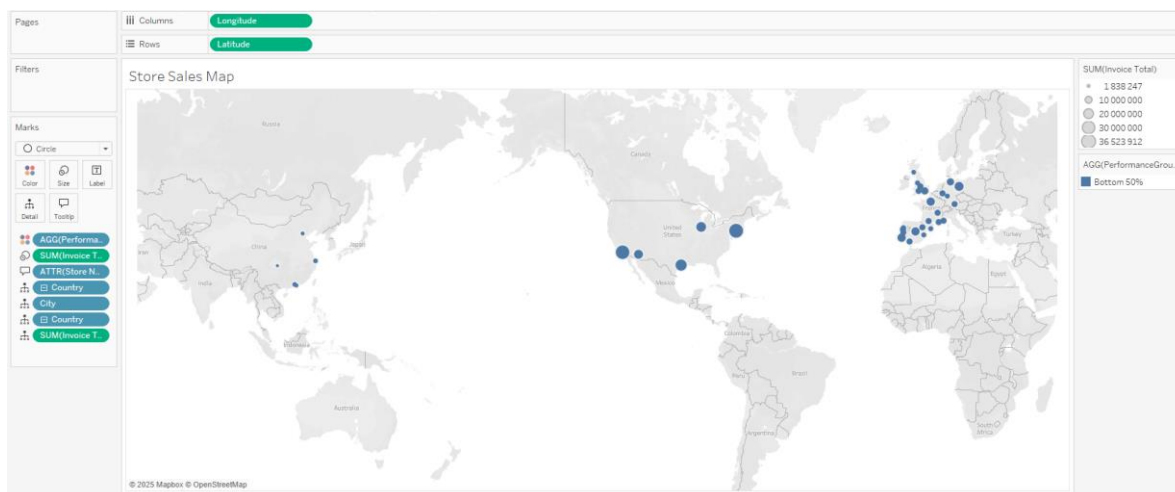
## Canva nr 3 Statystyki sklepów i krajów



### Wykres nr 1

### Store Sales Map (Bubble Map)

- Co pokazuje: Punkty na mapie według współrzędnych sklepu (Latitude, Longitude), wielkość bąbla = sprzedaż, kolor = top 50% vs bottom 50% sklepów.
- Dlaczego: Wizualizuje rozkład logistyczny sklepów i ich udział w ogólnej sprzedaży, wskazując kluczowe lokalizacje.

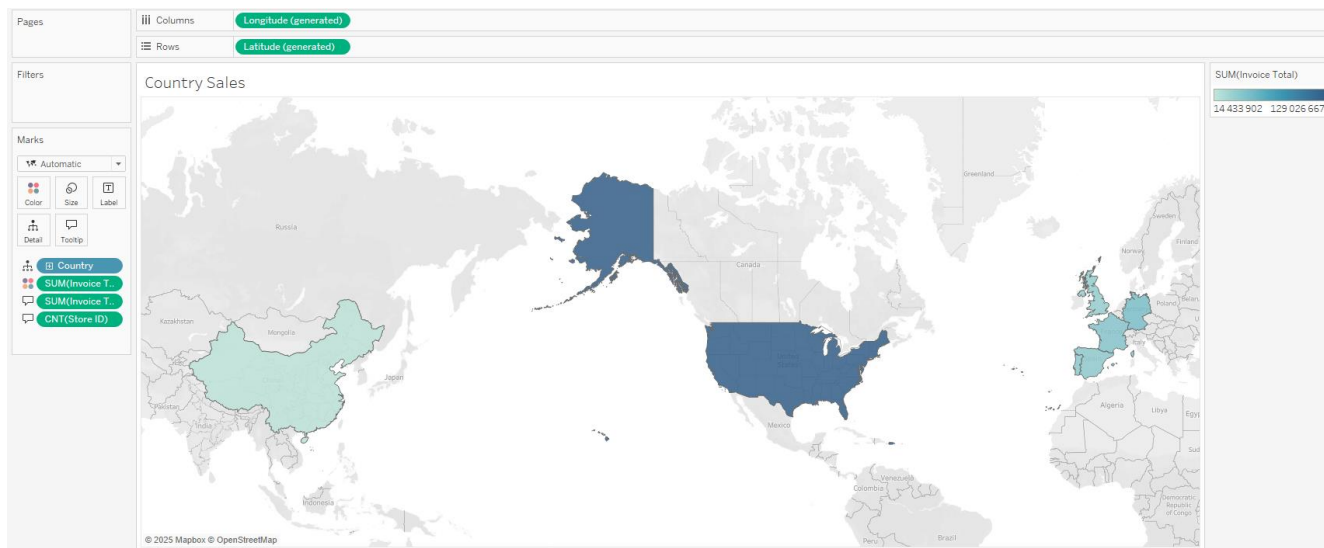


## Wykres nr 2

### Country Sales (Choropleth Map)

Co pokazuje: Kraje pokolorowane gradacją od jasnego do ciemnego według łącznej sprzedaży (Sum(Invoice Total)), z liczbą sklepów w tooltipie.

Dlaczego: Pozwala ocenić geograficzną dystrybucję przychodów i wielkość rynku w poszczególnych państwach.



### Wykres nr 3

#### Staff vs Sales

Co pokazuje: Scatter plot, w którym każdy punkt to sklep:

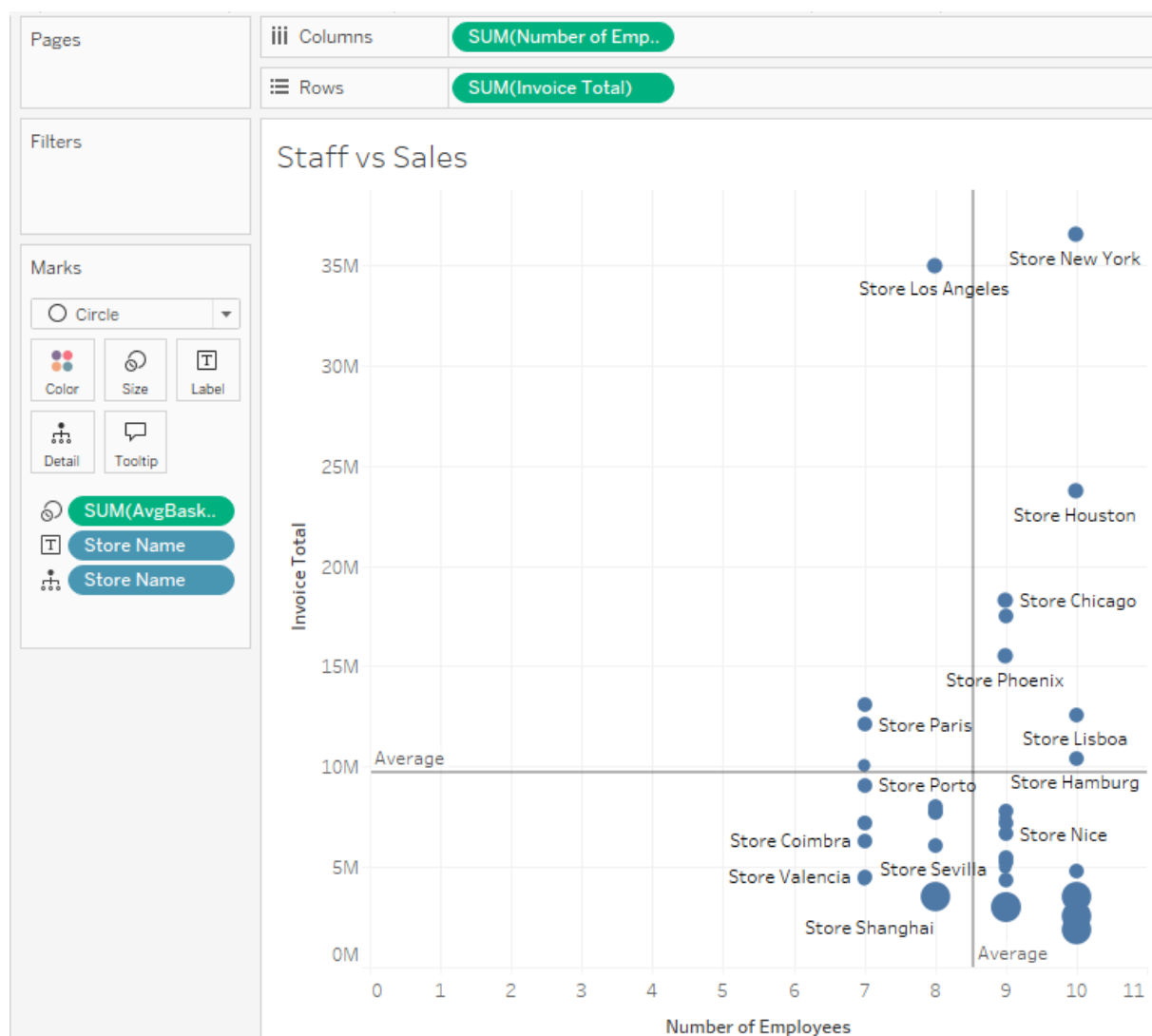
Oś X = liczba pracowników (Number of Employees),

Oś Y = całkowita sprzedaż (Sum(Invoice Total)),

rozmiar punktu = średnia wartość koszyka,

pionowa i pozioma linia = średnie.

Dlaczego: Łatwo wyłapać sklepy, w których zatrudnienie jest nisko- lub wysokowydajne, oraz te, które łączą wysoką sprzedaż z mniejszym zespołem.

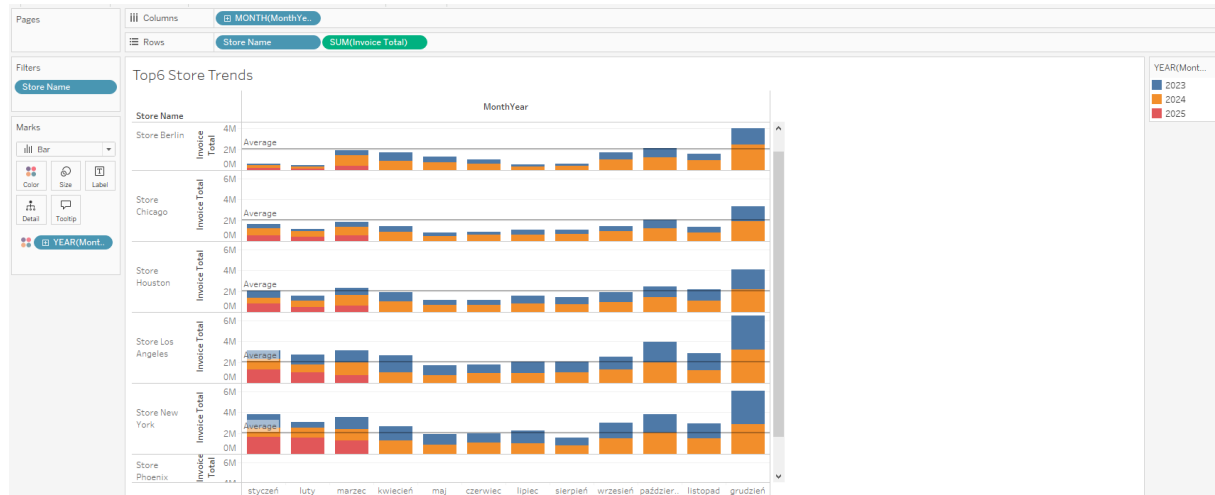


## Wykres nr 4

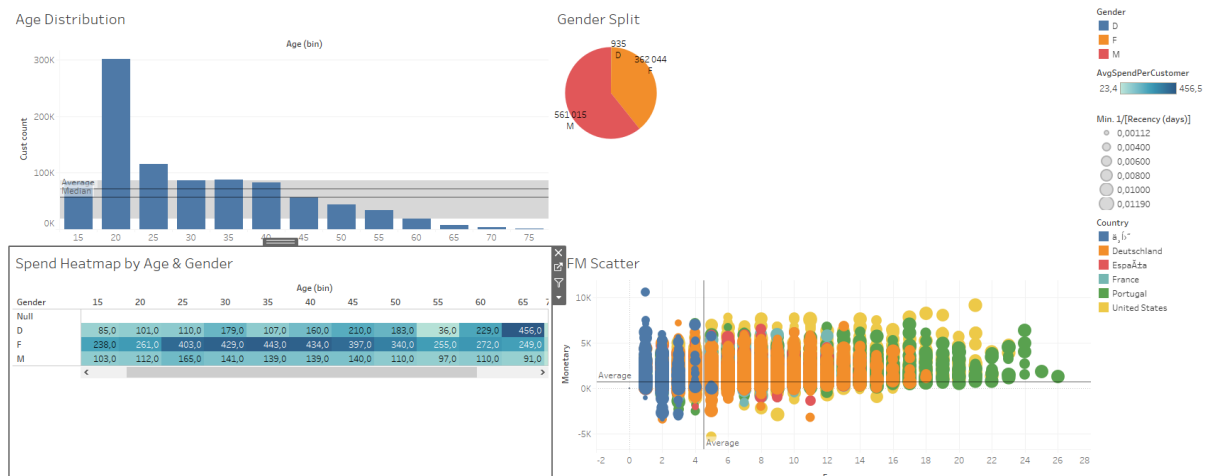
### Top 6 Store Trends (Small Multiples)

Co pokazuje: Miesięczne słupki sprzedaży (Invoice Total) dla sześciu najlepszych sklepów, rozłożone w pionowej siatce (każdy w osobnym panelu).

Dlaczego: Umożliwia szybkie porównanie dynamiki sprzedaży rok-do-roku i sezonowości między różnymi lokalizacjami.



# Canva nr 4 Analiza klientów według wieku i płci

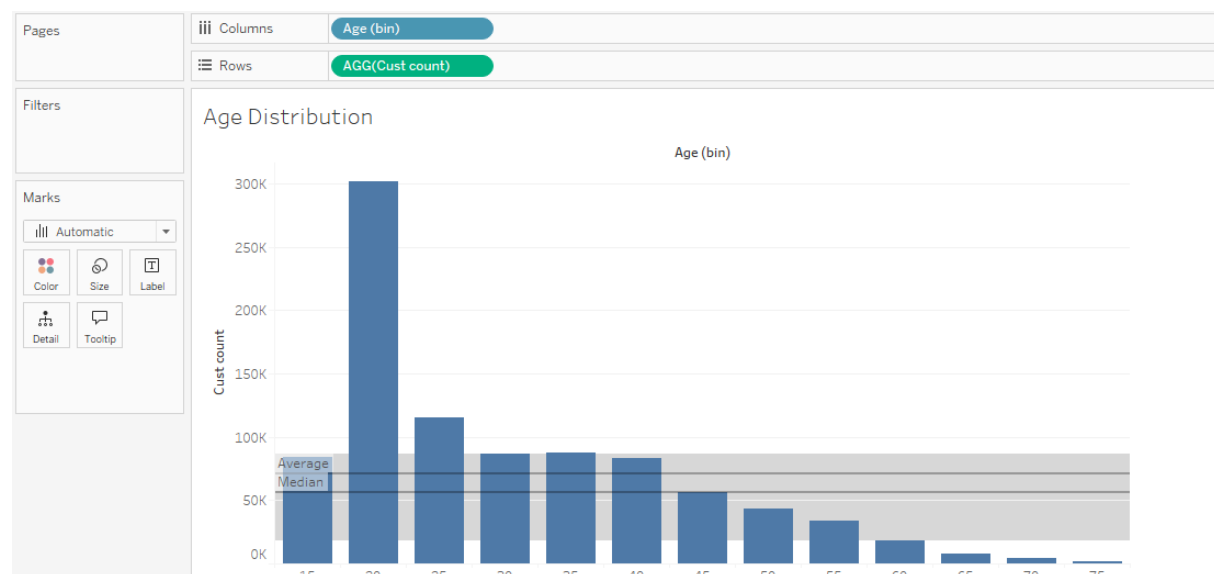


## Wykres nr 1

### Age Distribution

Co pokazuje: histogram liczby klientów grupowanych co 5 lat (bins), z widoczną średnią i medianą rozkładu wieku.

Po co: pozwala szybko ocenić dominujące grupy wiekowe i ogólną strukturę demograficzną bazy klientów.



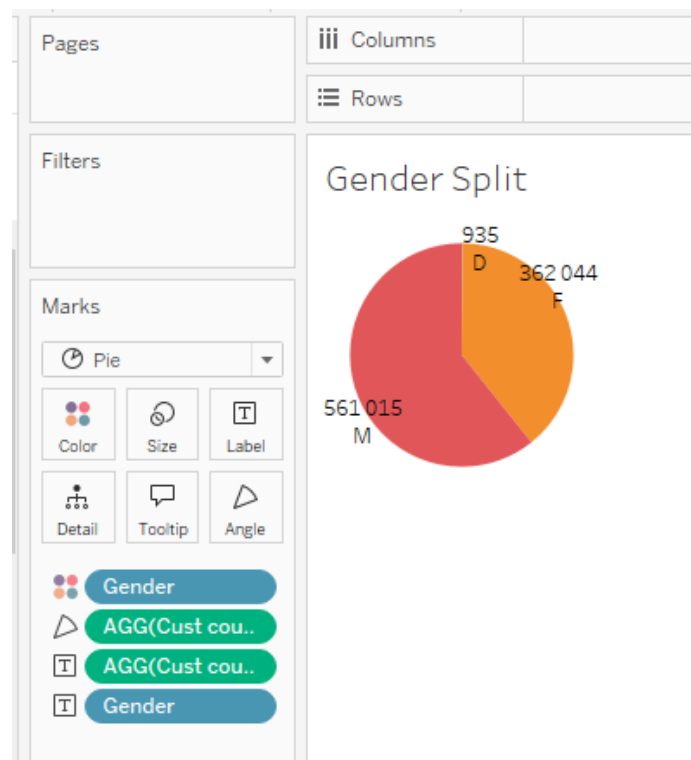


## Wykres nr 2

### Gender Split

Co pokazuje: wykres kołowy udziału klientów wg płci (M, F, Null) w całej bazie.

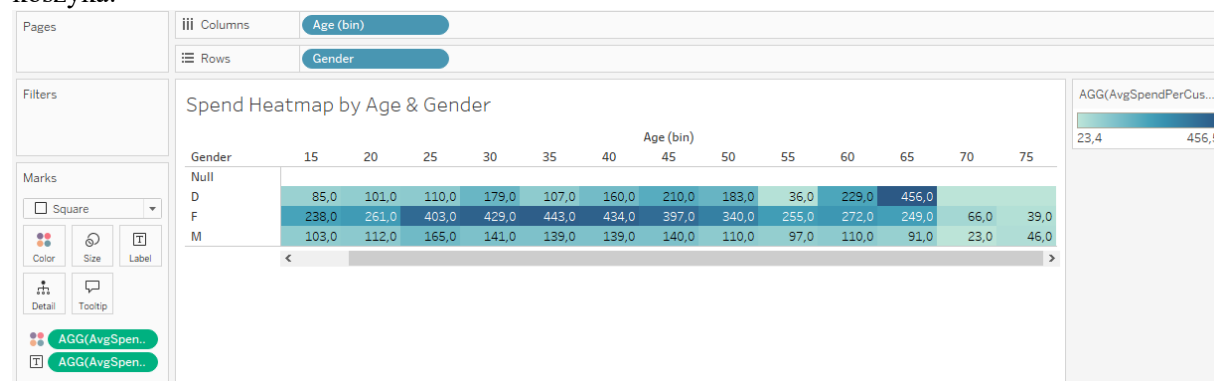
Po co: obrazuje strukturę płciową klienteli i ewentualne luki („Unknown”), co może być przydatne przy targetowaniu kampanii.



## Wykres nr 3

### Spend Heatmap by Age & Gender

Co pokazuje: średnie wydatki na klienta (AvgSpendPerCustomer) w macierzy: osie to grupa wiekowa (bins) i płeć (M/F/Null). Po co: identyfikuje, które połączenia wieku i płci generują największą wartość koszyka.

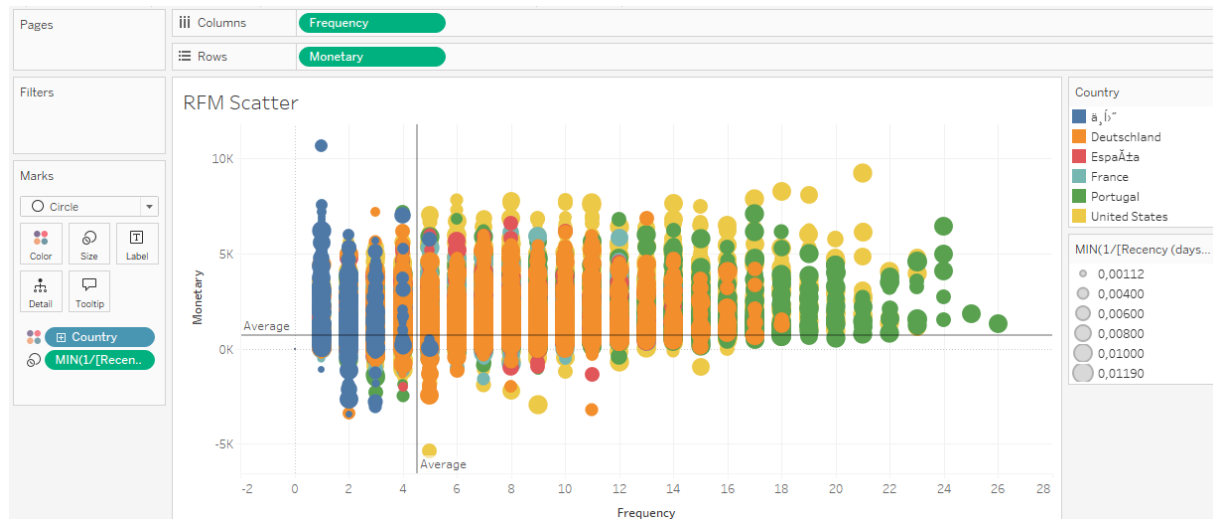


## Wykres nr 4

### RFM Scatter

Co pokazuje: każdy punkt to klient, oś X = liczba transakcji (Frequency), oś Y = łączna wartość zakupów (Monetary), rozmiar bąbla = odwrotność liczby dni od ostatniego zakupu (Recency), kolor = kraj. Średnie linie dzielą wykres na ćwiartki.

Po co: pozwala wychwycić „Champions” (dużo, wartościowo, niedawno) vs klienci o niskiej aktywności lub wysokiej recencji.



## Podsumowanie

Zadanie projektowe na bazie „Global Fashion Retail Stores” zostało zrealizowane pomyślnie. Dzięki niemu opanowaliśmy:

SQL Server Management Studio 19 – import i modelowanie danych, definiowanie tabel faktów i wymiarów, pisanie zapytań SQL oraz tworzenie i weryfikacja ekstraktów.

Visual Studio 2017 / SSIS – budowę procesów ETL: czyszczenie plików CSV, agregacje za pomocą CTE i komponentów Data Flow, obsługę błędów, konwersje typów, szybkiego ładowania (fast load) do tabel hurtowni.

Tableau Desktop Public Edition – łączenie danych (ekstrakty), definiowanie relacji, zaawansowanych wizualizacji (linie trendu, pareto, mapy, small multiples, RFM scatter, heatmaps).

Na warstwie raportowej zbudowano cztery dashboardy – trend sprzedaży, wydajność pracowników, porównanie sklepów i segmentację klientów – z 16 wykresami, które pozwalają na wielowymiarową analizę i interaktywne filtrowanie.

Podczas pracy na maszynie wirtualnej zdarzały się przeciążenia przy operacjach na dużych tabelach. Aby utrzymać stabilność, zwiększyliśmy zasoby VM (więcej rdzeni i pamięci RAM), zoptymalizowaliśmy pakiety SSIS (zastąpiłem ciężkie Data Flow jednobatchowym ELT w Execute SQL Task) oraz w Tableau użyliśmy lokalnych ekstraktów, co pozwoliło na płynne wykonanie wszystkich procesów.