

Wielowymiarowa Analiza Danych

Analiza Sprzedaży

Katarzyna Szostak

Wioleta Szuma

Rzeszów, 2024

1. Wstęp

Dane zostały pobrane ze strony <https://excelbianalytics.com>. Są to fikcyjne dane wygenerowane w celach testowych i edukacyjnych. Duża ilość rekordów sprawia, że dane te nadają się do analizy w zakresie Business Intelligence.

W realizowanym przez nas projekcie użyjemy danych Sprzedaży oraz Klientów.

Dane dotyczące sprzedaży liczą milion wierszy oraz prezentują się następująco:

Region	Country	Item.Type	Sales.Channel	Order.Priority	Order.Date	Order.ID	Ship.Date	Units.Sold	Unit.Price	Unit.Cost	Total.Revenue	Total.Cost	Total.Profit
1 Sub-Saharan Africa	South Africa	Fruits	Offline	M	7/27/2012	443368995	7/28/2012	1593	9.33	6.92	14862.69	11023.56	3839.13
2 Middle East and North Africa	Morocco	Clothes	Online	M	9/14/2013	667593514	10/19/2013	4611	109.28	35.84	503890.08	165258.24	338631.84
3 Australia and Oceania	Papua New Guinea	Meat	Offline	M	5/15/2015	940995585	6/4/2015	360	421.89	364.69	151880.40	131288.40	20592.00
4 Sub-Saharan Africa	Djibouti	Clothes	Offline	H	5/17/2017	880811536	7/2/2017	562	109.28	35.84	61415.36	20142.08	41273.28
5 Europe	Slovakia	Beverages	Offline	L	10/26/2016	174590194	12/4/2016	3973	47.45	31.79	188518.85	126301.67	62217.18
6 Asia	Sri Lanka	Fruits	Online	L	11/7/2011	830192887	12/18/2011	1379	9.33	6.92	12866.07	9542.68	3323.39
7 Sub-Saharan Africa	Seychelles	Beverages	Online	M	1/18/2013	425793445	2/16/2013	597	47.45	31.79	28327.65	18976.63	9349.02
8 Sub-Saharan Africa	Tanzania	Beverages	Online	L	11/30/2016	659878194	1/16/2017	1476	47.45	31.79	70036.20	46922.04	23114.16
9 Sub-Saharan Africa	Ghana	Office Supplies	Online	L	3/23/2017	601245963	4/15/2017	896	651.21	524.96	583484.16	470364.16	113120.00
10 Sub-Saharan Africa	Tanzania	Cosmetics	Offline	L	5/23/2016	739008080	5/24/2016	7768	437.20	263.33	3396169.60	2045547.44	1350622.16
11 Asia	Taiwan	Fruits	Offline	M	2/9/2014	732588374	2/23/2014	8034	9.33	6.92	74957.22	55595.28	19361.94
12 Middle East and North Africa	Algeria	Cosmetics	Online	M	2/18/2011	761723172	2/24/2011	9669	437.20	263.33	4227286.80	2546137.77	1681149.03
13 Asia	Singapore	Snacks	Online	C	1/28/2013	176461303	2/7/2013	7676	152.58	97.44	1171204.08	747949.44	423254.64
14 Australia and Oceania	Papua New Guinea	Clothes	Offline	L	6/20/2011	647164094	7/14/2011	9092	109.28	35.84	993573.76	325857.28	667716.48
15 Asia	Vietnam	Personal Care	Online	M	4/4/2010	314505374	5/6/2010	7984	81.73	56.67	652532.32	452453.28	200079.04
16 Sub-Saharan Africa	Uganda	Personal Care	Online	M	6/19/2014	539471471	7/21/2014	451	81.73	56.67	36860.23	25558.17	11302.06
17 Sub-Saharan Africa	Zimbabwe	Office Supplies	Offline	C	3/28/2011	953361213	4/8/2011	9623	651.21	524.96	6266593.83	5051690.08	1214903.75

Dane dotyczące klientów będą połączone z dwóch tabeli (obie mają po 500 tys. wierszy).

Pierwsza tabela:

Emp.ID	Name.Prefix	First.Name	Middle.Initial	Last.Name	Gender	E.Mail	Father's.Name	Mother's.Name	Mother's.Maiden.Name	Date.of.Birth	Time.of.Birth			
1	857211	Ms.	Hermia	J	Suhr	F	hermia.suhr@gmail.com	Todd Suhr	Cathrine Suhr	Hinojosa	9/4/1992	04:29:56 AM		
2	514341	Mr.	Antonio	Q	Joy	M	antoniojoy@yahoo.com	Clark Joy	Clarisa Joy	Gagliardi	12/24/1989	08:01:44 AM		
3	314598	Prof.	Sebastian	J	Moore	M	sebastian.moore@ibm.com	Everette Moore	Sol Moore	Paden	9/23/1980	01:42:17 AM		
4	987773	Mr.	Alec	S	Rittenhouse	M	alec.rittenhouse@gmail.com	Jeffery Rittenhouse	Maira Rittenhouse	Clanton	6/1/1974	07:39:19 AM		
5	690850	Mr.	Reggie	C	Doughty	M	reggie.doughty@gmail.com	Jerrod Doughty	Crista Doughty	Abron	11/27/1966	06:02:39 AM		
6	358257	Prof.	Elisha	B	Bottom	M	elisha.bottom@yahoo.com	Nicolas Bottom	Nu Bottom	Yuhas	9/23/1972	10:02:33 PM		
7	303899	Mr.	Danilo	R	Inwin	M	danilo.inwin@gmail.com	Ron Inwin	Angelina Inwin	McWilliams	12/22/1964	09:21:46 PM		
8	901163	Ms.	Madalene	D	Dierks	F	madalene.dierks@yahoo.com	Efen Dierks	Rebecca Dierks	Grosse	10/29/1957	03:51:08 AM		
Time.of.Birth	Age.in.Yrs.	Weight.in.Kgs.	Date.of.Joining	Quarter.of.Joining	Half.of.Joining	Year.of.Joining	Month.of.Joining	Month.Name.of.Joining	Short.Month	Day.of.Joining				
04:29:56 AM	24.91	57	9/9/2014	Q3	H2	2014	9	September	Sep	9				
08:01:44 AM	27.61	55	8/2/2011	Q3	H2	2011	8	August	Aug	2				
01:42:17 AM	36.87	62	4/1/2012	Q2	H1	2012	4	April	Apr	1				
07:39:19 AM	43.19	65	3/19/1998	Q1	H1	1998	3	March	Mar	19				
06:02:39 AM	50.70	61	1/4/2010	Q1	H1	2010	1	January	Jan	4				
10:02:33 PM	44.87	82	5/22/2015	Q2	H1	2015	5	May	May	22				
09:21:46 PM	52.63	52	2/14/1994	Q1	H1	1994	2	February	Feb	14				
03:51:08 AM	59.79	43	3/21/2011	Q1	H1	2011	3	March	Mar	21				
DOW.of.Joining	Short.DOW	Age.in.Company.Years.	Salary	Last...Hike	SSN	Phone.No.	Place.Name	County	City	State	Zip	Region	User.Name	Password
Tuesday	Tue	2.88	168991	12%	275-17-8844	479-539-4593	Peach Orchard	Clay	Peach Orchard	AR	72453	South	hjsuhr	o2%[-6wNIA
Tuesday	Tue	5.99	53504	30%	646-23-6213	229-234-6154	Rocky Ford	Screven	Rocky Ford	GA	30455	South	aqjoy	7_!%fEsaZB
Sunday	Sun	5.33	158859	3%	499-29-8139	212-231-9912	Antwerp	Jefferson	Antwerp	NY	13608	Northeast	sjmoore	Z0_dR^OQhQIF
Thursday	Thu	19.37	76105	2%	204-84-0943	229-873-6796	Milledgeville	Milledgeville	Milledgeville	GA	31059	South	asrittenhouse	7#7GLEd0
Monday	Mon	7.57	134436	8%	321-11-0416	314-677-4501	Springfield	Greene	Springfield	MO	65809	Midwest	rcdoughty	4~1+L^P^O
Friday	Fri	2.19	119237	22%	517-49-8835	206-233-8897	Dryden	Chelan	Dryden	WA	98821	West	ebottom	H50B-!tcBsEftc
Monday	Mon	23.47	115449	25%	605-87-8120	319-326-7935	Urbandale	Polk	Urbandale	IA	50323	Midwest	driwin	qd7XF[m5
Monday	Mon	6.36	104968	6%	050-02-8165	503-778-8049	Deer Island	Columbia	Deer Island	OR	97054	West	mdierks	4f_N7!%C!9aQ6&

Druga tabela:

Age	Attrition	BusinessTravel	DailyRate	Department	DistanceFromHome	Education	EducationField	EmployeeCount	EmployeeNumber	EnvironmentSatisfaction	Gender
43	No	Travel_Frequently	1417	Hardware	35	4	Marketing	1	1		4 Male
58	No	Travel_Rarely	968	Human Resources	20	3	Other	1	2		1 Female
34	Yes	Non-Travel	794	Research & Development	38	4	Human Resources	1	3		4 Male
42	Yes	Travel_Rarely	1385	Support	47	1	Marketing	1	4		3 Male

HourlyRate	JobInvolvement	JobLevel	JobRole	JobSatisfaction	MaritalStatus	MonthlyIncome	MonthlyRate	NumCompaniesWorked	Over18	OverTime
129	1	1	Human Resources	3	Single	46893	1359897		1 Y	No
53	1	5	Manufacturing Director	2	Divorced	23178	23178		8 Y	No
109	2	2	Manufacturing Director	3	Married	9008	189168		4 Y	Yes
78	2	5	Healthcare Representative	3	Married	36308	835084		3 Y	No

OverTime	PercentSalaryHike	PerformanceRating	RelationshipSatisfaction	StandardHours	StockOptionLevel	TotalWorkingYears
No	45	1	2	80	3	11
No	36	2	1	80	4	17
Yes	40	2	2	80	4	40
No	4	2	1	80	3	3

TrainingTimesLastYear	WorkLifeBalance	YearsAtCompany	YearsInCurrentRole	YearsSinceLastPromotion	YearsWithCurrManager
6	3	11	10	10	4
6	2	12	7	10	5
1	3	7	5	7	4
5	4	1	1	1	1
4	2	6	1	2	6

2. Przepływ danych

Do realizacji projektu posłużymy nam środowisko Oracle Analytics Desktop.

2.1. Załadowanie i przetworzenie danych

Nasze dane nie posiadają ID w danych, zatem w pierwszej kolejności musimy wykonać procesy na osobnych tabelach.

Załadujemy dane do tabeli sprzedaży:

←

Utwórz tabelę zbioru danych z: SalesRecordsM.csv

Nazwa

SalesRecordsM

Utworzono

W toku

Opis

Wysłano z "SalesRecordsM.csv".

Zmodyfikowano

W toku

Wysłany plik

SalesRecordsM.csv

Wybó...

Odsłowiezono

Nigdy

Z ogranicznikiem

Przecinek

Separator tysięcy

Przecinek





Separator dziesiętny

Okres

A	Region	A	Country	A	Item Type	A	Sales Channel	A	Order Priority	Order Date	Order ID	
	Sub-Saharan Africa		South Africa		Fruits		Offline		M	27.07.2012	443 368 995	28
	Middle East and North Africa		Morocco		Clothes		Online		M	14.09.2013	667 593 514	19
	Australia and Oceania		Papua New Guinea		Meat		Offline		M	15.05.2015	940 995 585	04
	Sub-Saharan Africa		Djibouti		Clothes		Offline		H	17.05.2017	880 811 536	02
	Europe		Slovakia		Beverages		Offline		L	26.10.2016	174 590 194	04
	Asia		Sri Lanka		Fruits		Online		L	07.11.2011	830 192 887	18
	Sub-Saharan Africa		Seychelles		Beverages		Online		M	18.01.2013	425 793 445	16
	Sub-Saharan Africa		Tanzania		Beverages		Online		L	30.11.2016	659 878 194	16
	Sub-Saharan Africa		Ghana		Office Supplies		Online		L	23.03.2017	601 245 965	15
	Sub-Saharan Africa		Tanzania		Office Supplies		Offline		L	27.05.2014	770 000 000	14

Anuluj OK

Załadowane dane prezentują się następująco:

Region	Country	Item Type	Sales Channel	Order Priority	Order Date	Order ID	Ship Date	Units Sold	Unit Price
Sub-Saharan Africa	Syria	Fruits	Offline	L					
Europe	Belarus	Baby Food	Online	M					
Asia	Madagascar	Vegetables		C					
Middle East and North Africa	Zimbabwe	Cosmetics		H					
Central America and the Caribbean	Andorra	Cereal							
Australia and Oceania	Bhutan	Office Supp...							
North America	Cameroon	Personal Care							
	Finland	Household							
	Israel	Beverages							
	Kiribati	Clothes							
A Region	A Country	A Item Type	A Sales Channel	A Order Priority	Order Date	# Order ID	Ship Date	# Units Sold	# Unit Price
North America	Mexico	Fruits	Online	M	21.06.2011	632 483 613	20.07.2011	4811	9.33
Europe	Cyprus	Cosmetics	Online	C	19.07.2017	984 329 354	19.07.2017	3785	437.20
Middle East and North Africa	Iran	Office Supplies	Online	C	20.11.2016	778 384 196	24.12.2016	36	651.21
Sub-Saharan Africa	South Sudan	Fruits	Online	C	16.06.2012	732 410 275	23.07.2012	5331	9.33
Middle East and North Africa	United Arab Emirates	Personal Care	Offline	C	18.08.2016	357 125 890	01.09.2016	342	81.73
Europe	Romania	Personal Care	Offline	L	27.06.2017	502 466 857	06.08.2017	9964	81.73
Asia	China	Personal Care	Offline	C	14.10.2010	445 153 844	16.11.2010	9770	81.73

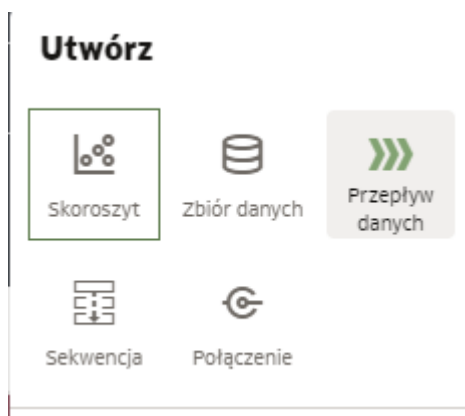
W kolumnie Country niepoprawne wartości stanowią 2%.

Poprawne wartości: Country	98%
Niepoprawne wartości: Country	2%
Sub-Saharan Africa	Syria
Middle East and North Africa	Belarus
Central America and the Caribbean	Madagascar
Australia and Oceania	Zimbabwe
North America	Andorra
	Bhutan
	Cameroon
	Finland
	Israel
	Kiribati

Używając filtra, otrzymujemy nazwy niepoprawnych nazw krajów, które później usuniemy z baz danych.

Country	Item Type	Sales Channel	Order Priority
Federated States of Micr...	Cosmetics		
Cote d'Ivoire	Household		
The Gambia	Cereal		
The Bahamas	Clothes		
Vatican City	Fruits		
	Meat		
	Office Supp...		
A Country	A Item Type	A Sales Channel	A Order Priority
Cote d'Ivoire	Baby Food	Offline	L
Federated States of Micro...	Snacks	Online	L
Federated States of Micro...	Household	Online	H
Cote d'Ivoire	Baby Food	Offline	L
The Gambia	Cereal	Online	L
Federated States of Micro...	Snacks	Online	H

Tworzymy nowy przepływ danych:



Ładujemy dane i wybieramy wszystkie kolumny:

Wad_proje...

Dodawanie danych - Wad_projekt

Zbiór danych: Wad_projekt Wybór...

Po ☐ Monitoruj o wybranie zbioru danych uruchomi...

Kolumny: Wszystkie (14) Wybór (14)

☒ Nazwa

☒ Region

☒ Country

☒ Item Type

☒ Sales Channel

☒ Order Priority

ab Region	ab Country	ab Item Type	ab Sales Channel	ab Order Priority	ff Order Date	99 Order ID
Sub-Saharan Africa	South Africa	Fruits	Offline	M	27.07.2012	443368995
Middle East and North Africa	Morocco	Clothes	Online	M	14.09.2013	667593514
Australia and Oceania	Papua New Guinea	Meat	Offline	M	15.05.2015	940995585

Priorytet zamówienia był dotychczas oznaczony literami, które mogą być problematyczne w zrozumieniu dla przeciętnego użytkownika, dlatego przekształcimy je do pełnych nazw.

Transformacja kolumny

Transformacja Order Priority

Nazwa

```
case
when "SalesRecordsM"."Order Priority" = 'M' then 'Medium'
when "SalesRecordsM"."Order Priority" = 'L' then 'Low'
when "SalesRecordsM"."Order Priority" = 'H' then 'High'
else 'Very High'
end
```

Weryfikuj


✓ Obliczenie zweryfikowane

ab Order Priority	ab OrderPriority
C	Medium
L	Medium
M	Medium
H	High
C	Low
C	Low
H	Medium
M	Low
L	Low
C	Low
C	Medium


Dodajemy kolumnę „ClientID” losowych ID klientów z zakresu od 1 do 500000, dzięki czemu będziemy mogli przyporządkować dane zamówienie do losowego klienta z tabeli klientów.


Nazwa


```
cast(floor(rand()*500000)+1 as int)
```


 Obliczenie zweryfikowane


99 CliendID
411 421
373 303
87 055
429 472
355 251
256 768
151 998
7493
45 702


 Przepływ danych "Nowy przepływ danych" ukończony.

 Wad_projekty

 Filtrowani...

 Transform... kolumny

 Dodawani... kolumn

 Zapisywan... danych

Zapisywanie zbioru danych

Zbiór danych

Tabela zbiorów danych

Opis

Zapisz dane w

Po uruchomieniu ☐ Monitoruj o określenie zbioru danych

Kolumny

Nazwa	Traktuj jako	Agregacja domyślna
Region	Atrybut ▼	
Country	Atrybut ▼	
Item Type	Atrybut ▼	
Sales Channel	Atrybut ▼	
OrderPriority	Atrybut ▼	

Pierwsza tabela klientów:

Przesyłamy pierwszą tabelę klientów:

←

Utwórz tabelę zbioru danych z: Customer_Records_500tys.csv

Nazwa

Customers1

Opis

Wysłano z "Customer_Records_500tys.csv".

Wysłany plik

Customer_Records_500tys.csv

Wybó...

Z ogranicznikiem

Przecinek

Separator tysięcy

Przecinek

Separator dziesiętny

Okres

Utworzono

W toku

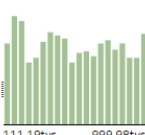
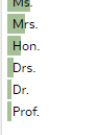

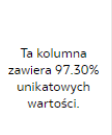
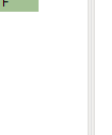
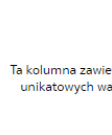
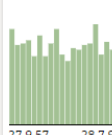
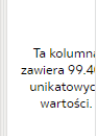
Zmodyfikowano

W toku

Odświeżono

Nigdy

Customers1

Emp ID	Name Prefix	First Name	Middle Initial	Last Name	Gender	E Mail	Date of Birth	Time of Birth
 111,19tys. 999,98tys.	 Mr. Ms. Mrs. Hon. Drs. Dr. Prof.	Francis Sean Alexander Angelo Claude Damien Darrick Eliseo Erin Giuseppe	 N C G H V W R S L A	 Ta kolumna zawiera 97.30% unikatowych wartości.	 M F	 Ta kolumna zawiera 100% unikatowych wartości.	 27.9.57 28.7.96	 Ta kolumna zawiera 99.4% unikatowych wartości.
# Emp ID	A Name Prefix	A First Name	A Middle Initial	A Last Name	A Gender	A E Mail	🕒 Date of Birth	A Time of Birth
209 797	Mr.	Abe	L	Devita	M	abe.devita@yahoo.com	02.09.1995	06:01:45 AM

Zauważamy niepoprawne wartości w Gender i innych kolumnach, lecz ich transformację zajmiemy się później w procesie ETL, aby nie popsuło nam się indeksowanie (przypisanie sprzedaży do

ID klienta od 1 do 500 tys.).

Poprawne wartości: gender 0%

Niepoprawne wartości: gender 100%

M

F

Place Name	County	City	State	Zip
Washington	Los Angeles	Washington	CA	14 ...
Birmingham	Jefferson	Birmingham	TX	14 ...
Miami	Washington	Miami	NY	19 ...
Fairfield	Franklin	Fairfield	PA	20 ...
Huntington	Dallas	Huntington	NC	28 ...
Los Angeles	District of ...	Los Angeles	FL	28 ...
Oklahoma ...	Lincoln	Oklahoma ...	VA	28 ...
Andover	Montgomery	Andover	IL	36 ...
Brooklyn	Clay	Brooklyn	OH	54 ...
Dallas	Cumberland	Dallas	MI	61 ...
A Place Name	A County	A City	A State	# Zip
Dalbo	Isanti	Dalbo	MN	55 017
Collettsville	Caldwell	Collettsville	NC	28 611
Clarkdale	Yavapai	Clarkdale	AZ	86 324
Bonita Springs	Lee	Bonita Springs	FL	34 133

Tworzymy nowy przepływ danych Customers1:

- Załadujemy dane:

Dodawanie danych - Customers1

Zbiór danych: Customers1

Wybór...

Po ☐ Monitoruj o wybranie zbioru danych uruchomi...

Kolumny: Wszystkie (37) Wybór (37)

☒ Nazwa

☒ Emp ID

☒ Name Prefix

☒ First Name

☒ Middle Initial

☒ Last Name

- Wybieramy kolumny:

Wybrano (14/34)



Usuń wszystkie

Usuń wybrane

Emp ID
Name Prefix
First Name
Last Name
Gender
E Mail
Date of Birth
Age in Yrs.
Weight in Kgs.
Date of Joining
Year of Joining
DOW of Joining
Region
Last % Hike

- Zmieniamy nazwy kolumn – **“Emp ID”** na **“Client ID”** oraz **“Last % Hike”** na **“Discount”**.
“Discount” będzie oznaczało procent zniżki na zakupy jaki posiada klient.

Customer...
Wybór kolumn
Zmiana na... kolumn

Zmiana nazw kolumn

Źródło	Zmień nazwę
Emp ID	Client ID
Name Prefix	Name Prefix
First Name	First Name
Last Name	Last Name
Gender	Gender
E Mail	E Mail
Date of Birth	Date of Birth
Age in Yrs.	Age in Yrs.
Weight in Kgs.	Weight in Kgs.
Date of Joining	Date of Joining
Year of Joining	Year of Joining
DOW of Joining	DOW of Joining
Region	Region
Last % Hike	Discount

- Dodajemy nową kolumnę klasyfikującą klientów ze względu na rok dołączenia. Klienci, którzy dołączyli poniżej 2015 roku nazwiemy “long-time”, a pozostali zostaną nazwani jako “new”.

Customer...
Wybór kolumn
Zmiana na... kolumn
Dodawani... kolumn

Dodawanie kolumn

Kolumna

Nowa kolumna1

Nazwa
Customer Type

```

case
when Year of Joining >= 2015 then 'new'
else 'long-time'
end

```

Weryfikuj
Obliczenie zweryfikowane

ab Customer Type
long-time
long-time
new
long-time
long-time

Transformacja kolumny

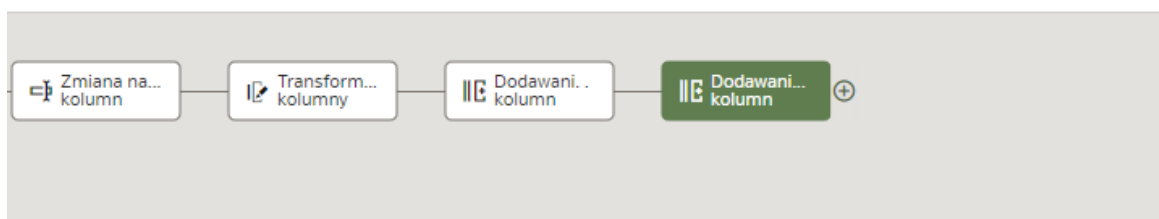
Transformacja Age in Yrs.

Nazwa

```
CAST((CURRENT_DATE - Date of Birth) / 365.25 AS INT)
```

99 Age in Yrs.
38
47
59
53
53

- Tworzymy nową kolumnę **Generation** klasyfikującą ze względu na generację. Do tego celu użyliśmy kolumny określającą wiek w latach.



Nazwa

Generation

```
CASE
  WHEN Age in Yrs. >=79 THEN 'Silent Generation'
  WHEN Age in Yrs. <79 AND Age in Yrs. >= 60 THEN 'Baby Boomers'
  WHEN Age in Yrs. <60 AND Age in Yrs. >= 44 THEN 'Generation X'
  WHEN Age in Yrs. <44 AND Age in Yrs. >=28 THEN 'Millennials'
  WHEN Age in Yrs. < 28 AND Age in Yrs. >=12 THEN 'Generation Z'
  WHEN Age in Yrs. <12 THEN 'Alpha Generation'
  ELSE 'Unknown Generation'
END
```

ab	Generation
	Millennials
	Millennials
	Generation Z
	Generation Z
	Millennials

Customer...
Wybór kolumn
Zmiana na... kolumn
Transform... kolumny
Dodawani... kolumn
Dodawani... kolumn
Zapisywan... danych

Zapisywanie zbioru danych

Zbiór danych
Tabela_Klienci1

Tabela zbiorów danych

Opis
Tabela zawiera wybrane 14 kolumn, zmienione kolumny, stransomrację oraz dodanie dwóch kolumn.

Zapisz dane w
Magazyn zbiorów danych

Po uruchomieniu
☐
Monituj o określenie zbioru danych

Nazwa	Traktuj jako	Agregacja domyślna
Client ID	Miara	Suma
Name Prefix	Atrybut	
First Name	Atrybut	
Last Name	Atrybut	
Gender	Atrybut	
E Mail	Atrybut	
Date of Birth	Atrybut	
Age in Yrs.	Miara	Suma
Weight in Kgs.	Miara	Suma
Date of Joining	Atrybut	
Year of Joining	Miara	Suma

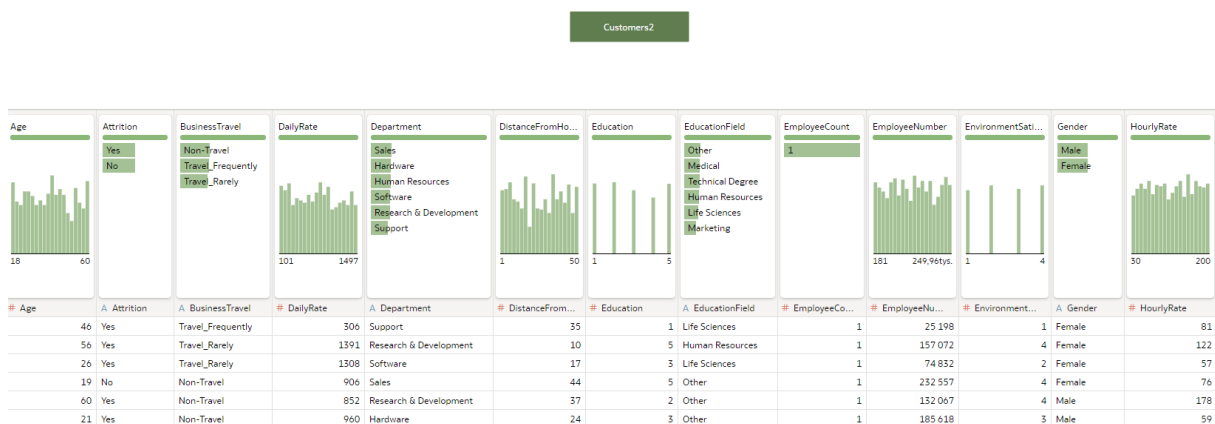
99 Client ID	ab Name Prefix	ab First Name	ab Last Name	ab Gender	ab E Mail	99 Date of Birth	99 Age in Yrs.	99
271310	Prof.	Aaron	Allums	M	aaron.allums@ibm.com	12.04.1986	38	78
175661	Dr.	Aaron	Almendarez	M	aaron.almendarez@hotmail.com	18.08.1976	47	58

Proces przepływu danych przebiegł pomyślnie.

✓
Przepływ danych "Customers_data_flow" ukończony.

Customer...
Wybór kolumn
Zmiana na... kolumn
Transform... kolumny
Dodawani... kolumn
Dodawani... kolumn
Zapisywan... danych

Tworzymy nowy zbiór danych – Klienci2



Over18 Y	OverTime Yes No	PercentSalaryHike 0 49	PerformanceRati... 1 4	RelationshipSatis... 1 4	StandardHours 80
A Over18	A OverTime	# PercentSalary...	# Performance...	# RelationshipS...	# StandardHours
Y	Yes	30	3	4	80
Y	Yes	43	3	2	80
Y	No	27	1	4	80
Y	No	34	3	3	80
Y	No	42	1	1	80
Y	No	27	1	4	80
Y	No	13	2	1	80
Y	Yes	44	3	4	80
Y	No	35	1	2	80

Jedynie w kolumnie Over18 pojawiają się problemy, lecz tym problemem zajmiemy się w późniejszym etapie.

Tworzymy nowy przepływ danych – Customers2

Dodawanie danych - Customers2

Zbiór danych
Opis

Customers2
Wybór...

Po ☐ Monitoruj o wybranie zbioru danych uruchomi...

Kolumny Wszystkie (35) Wybór (35)

☒ Nazwa

☒ Age

☒ Attrition

☒ BusinessTravel

☒ DailyRate

☒ Department

99 Age	ab Attrition	ab BusinessTravel	99 DailyRate	ab Department	99 DistanceFrom...	99 Education	ab EducationField
43	No	Travel_Frequently	1417	Hardware	35	4	Marketing
58	No	Travel_Rarely	968	Human Resources	20	3	Other
34	Yes	Non-Travel	794	Research & Development	38	4	Human Resources

Wybieramy 9 interesujących nas kolumn:

Wybór kolumn									
Szukaj <input type="text"/>					Dodaj wszystkie		Dodaj wybrane		Wybrano (9/35)
Age					=		EmployeeNumber		
DailyRate							Attrition		
Department							BusinessTravel		
EmployeeCount							DistanceFromHome		
Gender							Education		
99 EmployeeNu...	ab Attrition	ab BusinessTravel	99 DistanceFrom...	99 Education	ab EducationField	99 Environment...	ab MaritalStatus	99 YearsAtComp...	
1	No	Travel_Frequently	35	4	Marketing	4	Single	11	
2	No	Travel_Rarely	20	3	Other	1	Divorced	12	
3	Yes	Non-Travel	38	4	Human Resources	4	Married	7	

Zmieniamy nazwy kolumn:

- “EmployeeNumer” na “Client ID”
- “Attrition” na “Newsletter subscription”
- “DistanceFromHome” na “Distance from the Store”
- “Education” na “Client Category”
- “EnviromentSatisfaction” na “Service Satisfaction”
- “YearsAtCompany” na “Number of Orders”

Źródło	Zmień nazwę
EmployeeNumber	Client ID
Attrition	Newsletter Subscription
BusinessTravel	Business Travel
DistanceFromHome	Distance from the Store
Education	Client Category
EducationField	Education Field
EnvironmentSatisfaction	Service Satisfaction
MaritalStatus	Marital Status
YearsAtCompany	Number of Orders

W kolejnym kroku zmieniamy kolumnę "Client Category" na podstawie liczby zamówień. Jeśli klient dokonał więcej niż 30 zamówień ma on status "platinum". Jeśli natomiast mniej niż 10 jego status wskazuje "bronze"

Customer...X

Wybór kolumn

Zmiana na... kolumn

Transform... kolumny

Transformacja kolumny

Transformacja Client Category

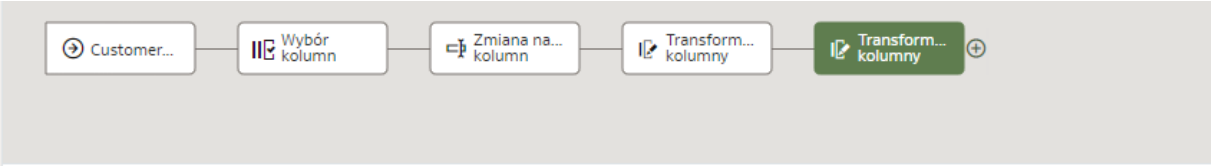
Nazwa Client Category

```
CASE
WHEN Number of Orders >= 30 then 'platinum'
WHEN Number of Orders < 30 and Number of Orders >= 20 then 'golden'
WHEN Number of Orders < 20 and Number of Orders >= 10 then 'sliver'
ELSE 'bronze'
END
```

Weryfikuj

Obliczenie zweryfikowane

ab	Client Category
	sliver
	sliver
	bronze
	bronze
	bronze
	bronze
	sliver



Transformacja kolumny.

Transformacja Service Satisfaction

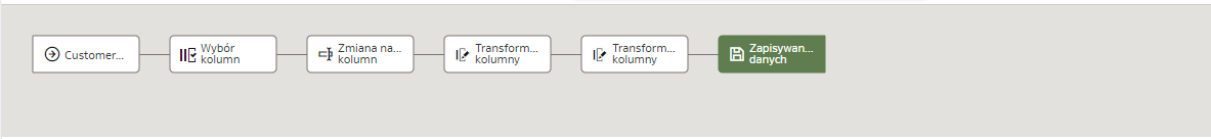
Nazwa Service Satisfaction

```
CASE
WHEN Service Satisfaction = 1 THEN 'Low'
WHEN Service Satisfaction = 2 THEN 'Medium'
WHEN Service Satisfaction = 3 THEN 'High'
ELSE 'Very High'
END
```

Weryfikuj Obliczenie zweryfikowane

ab Service Satisf...	\$
Very High	\$
Low	[
Very High]
High]
High	[
High	[
Low	[
Low]

Ukończono: Przepływ danych "Customers2".



Zapisywanie zbioru danych

Zbiór danych Customers_2

Tabela zbiorów danych

Opis

Zapisz dane w Magazyn zbiorów danych

Po uruchomieniu ☐ Monitoruj o określenie zbioru danych

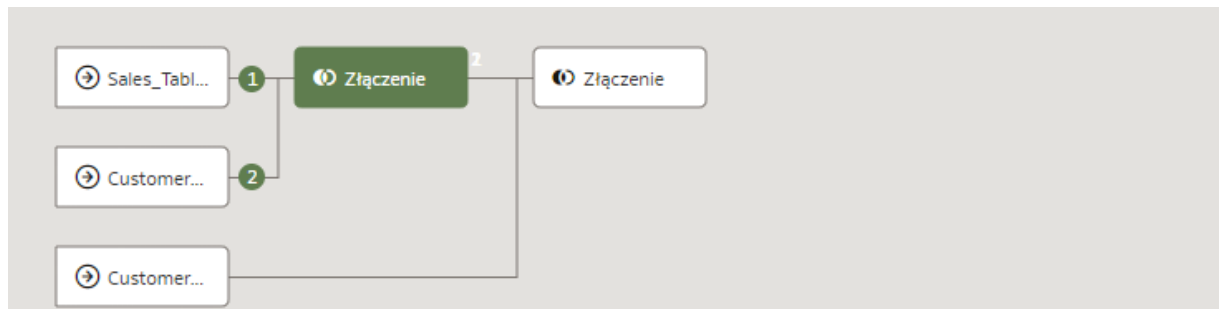
Nazwa	Traktuj jako	Agregacja domyślna
Client ID	Miara	Suma
Newsletter Subscription	Atrybut	
Business Travel	Atrybut	
Distance from the Store	Miara	Suma
Client Category	Atrybut	

99 Client ID	ab Newsletter Su...	ab Business Travel	99 Distance fro...	ab Client Category	ab Education Field	ab Service Satisf...	ab Marital Status	99 Number of Or...
1	No	Travel_Frequently	35	silver	Marketing	Very High	Single	11
2	No	Travel_Rarely	20	silver	Other	Low	Divorced	12
7	Yes	Not Travel	70	bronze	Business Development	Very High	Married	7

2.2. Łączenie danych

Customer_table1

Customers_2



Złączenie

Zachowaj wiersze

1 Wejście 1 Zgodne wiersze ▼

2 Wejście 2 Zgodne wiersze ▼

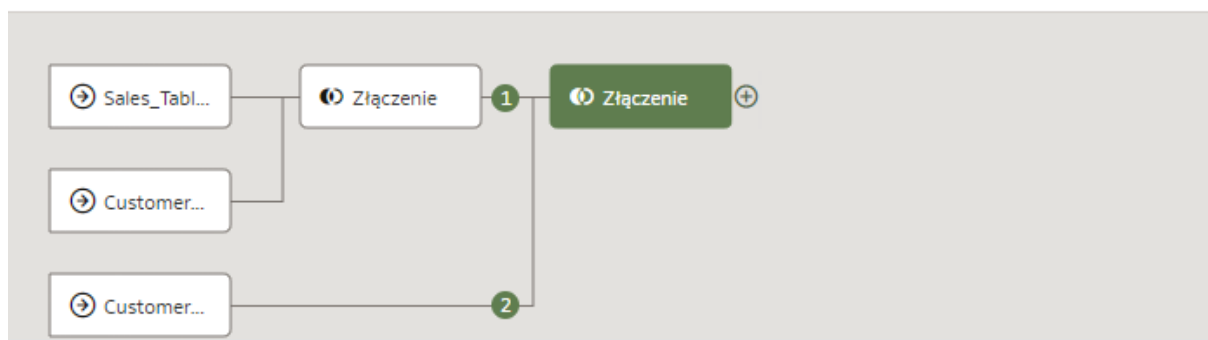
Uzgadnianie kolumn



Wejście 1	Operator	Wejście 2	
ClientID ▼	= ▼	Client ID ▼	🗑️

.....

ab Region	ab Country	ab Item Type	ab Sales Channel	ab OrderPriority
Sub-Saharan Africa	South Africa	Clothes	Online	Very High
Europe	Russia	Clothes	Offline	Very High
Middle East and North Africa	Syria	Vegetables	Offline	High



Złączenie

Zachowaj wiersze

1 Wejście 1 Zgodne wiersze ▼

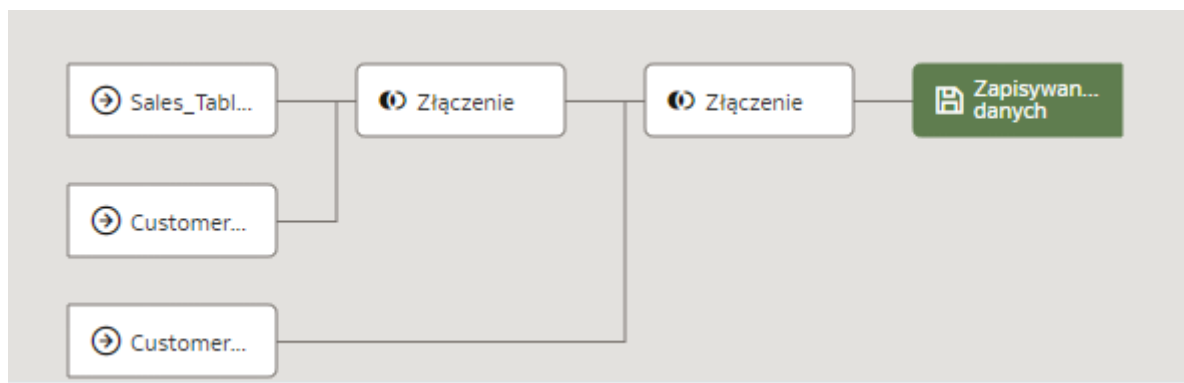
2 Wejście 2 Zgodne wiersze ▼

Uzgadnianie kolumn

Wejście 1	Operator	Wejście 2	
Client ID ▼	= ▼	Client ID ▼	🗑️

99 CliendID	99 Client ID	99 Client ID_1
244210	244210	244210
254616	254616	254616
243645	243645	243645
247918	247918	247918
243630	243630	243630
256249	256249	256249
243599	243599	243599
243599	243599	243599
246499	246499	246499
248986	248986	248986

Dzięki zgodności połączenia tabel według ID klientów, będziemy mogli stworzyć finalne procesy ETL do analizy sprzedaży.



2.3. Finalne przepływy ETL

Dodawanie danych

Utwórz zbiór danych

CSV

Customer_table1

CSV

Customers_2

CSV

Customers1

CSV

Customers2

CSV

Sales_Table

X

Sample Order Lines

X

Sample States

CSV

WAD_dane_zlaczone

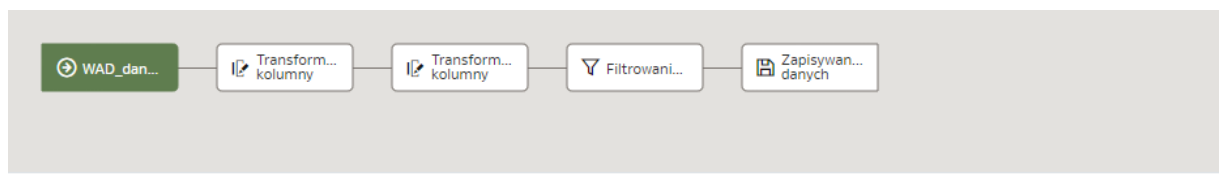
CSV

Wad_projekt

Anuluj

Dodaj

Wybieramy interesujące nas kolumny:



Dodawanie danych - WAD_dane_zlaczone

Zbiór danych: Wybór...

Po ☐ Monituj o wybranie zbioru danych uruchomi...

Kolumny: Wszystkie (40) Wybór (37) 🔍

- ☐ Nazwa
- ☒ Region
- ☒ Country
- ☒ Item Type
- ☒ Sales Channel
- ☒ OrderPriority

ab Region	ab Country	ab Item Type	ab Sales Channel	ab OrderPriority	📅 Order Date	🔢 Order ID	📅
Europe	Netherlands	Snacks	Online	Medium	22.08.2015	498503196	23.
Sub-Saharan Africa	Ghana	Personal Care	Offline	Very High	02.10.2010	141938483	12.
North America	Greenland	Fruits	Offline	High	08.01.2016	280631005	03.
Middle East and North Africa	Iran	Clothes	Online	High	19.05.2017	000000077	10.

Gender

☒ F

☐ M

A Gender

F

M

M

F

F

Weryfikuj ☒ Obliczenie zweryfikowane

Workflow: WAD_dan... -> Transform... kolumny (+)

Transformacja kolumny

Transformacja: Gender

Nazwa:

```
case
when Gender = 'F' then 'Female'
else 'Male'
end
```

WAD_dan...

Transform...
kolumny

Transform...
kolumny

+

Transformacja kolumny

Transformacja Region

Nazwa

Region

```
CASE
  WHEN Region LIKE '%America%' THEN 'America'
  WHEN Region LIKE '%Asia%' THEN 'Asia'
  WHEN Region LIKE '%Europe%' THEN 'Europe'
  WHEN Region LIKE '%Australia%' THEN 'Australia'
  WHEN Region LIKE '%Africa%' THEN 'Africa'
  ELSE Region
END
```

Weryfikuj

Obliczenie zweryfikowane

ab	Region
	Europe
	Africa
	Africa
	Australia
	Africa
	Australia
	Europe
	Australia

WAD_dan...X
Transform...kolumny
Transform...kolumny
Filtrowani...
Zapisywan...danych

Filtrowanie

Order Date

Order Date
Przedział dat
Zakres
1.01.2012 - 31.12.2016

WAD_dan...
Transform...kolumny
Transform...kolumny
Filtrowani...
Zapisywan...danych

Zapisywanie zbioru danych

Zbiór danych: WAD_dane_gotowe
Tabela zbiorów danych:
Opis:
Zapisz dane w: Magazyn zbiorów danych
Po uruchomieniu ☐ Monitoruj o określenie zbioru danych

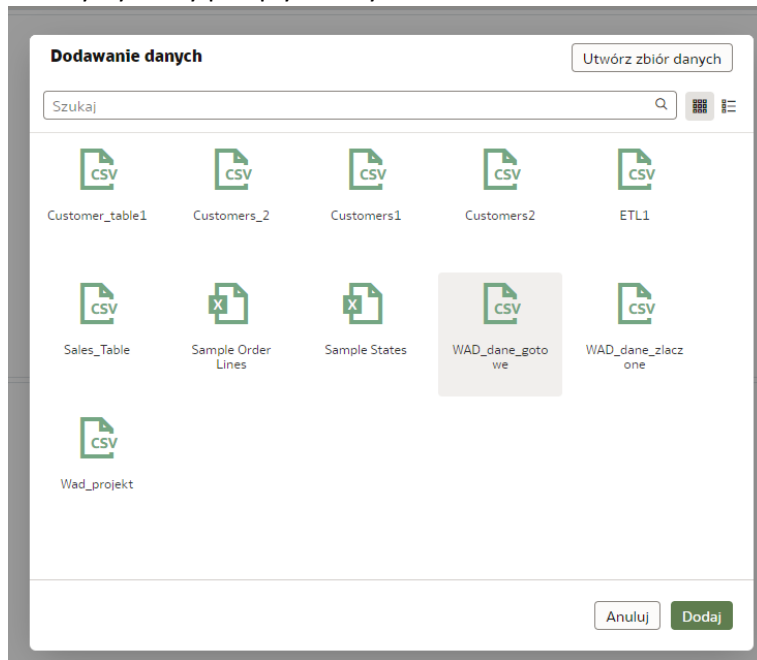
Kolumny

Nazwa	Traktuj jako	Agregacja domyślna
Continent	Atrybut	
Country	Atrybut	
Item Type	Atrybut	
Sales Channel	Atrybut	
Order Date	Atrybut	

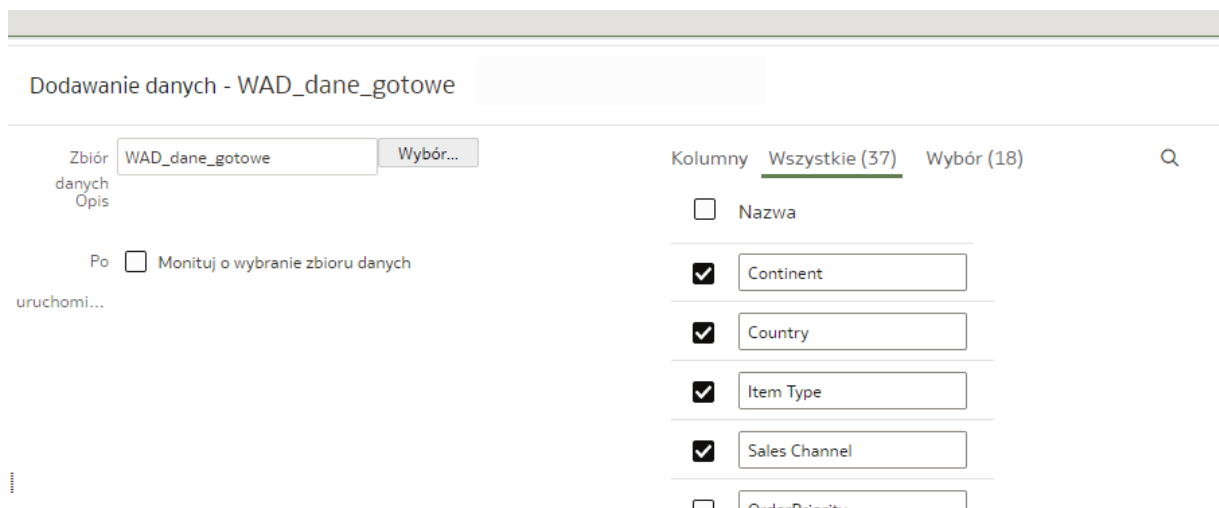
ab Continent	ab Country	ab Item Type	ab Sales Channel	Order Date	99 Units Sold	99 Total Revenue	99 Total Cost
Europe	Italy	Household	Offline	02.03.2015	8475	5663588.25	4259026.5

2.3.1. ETL - 1 Analiza tendencji sprzedaży

Tworzymy nowy przepływ danych:



Wybieramy interesujące nas kolumny:



Transformujemy kolumnę Distance from the Store

WAD_dan... Transform... kolumny

Transformacja kolumny

Transformacja Distance from the Store

Nazwa Distance from the Store

```
case
when Distance from the Store >=0 and Distance from the Store < 20 then 'short'
when Distance from the Store >=20 and Distance from the Store < 30 then 'medium'
else 'long'
end
```

Weryfikuj Obliczenie zweryfikowane

W filtrowaniu wykluczamy Rosję, Białoruś i Izrael:

WAD_dan... Transform... kolumny Filtrowani...

Filtrowanie

Country (4)

Country

Lista

Wyklucz ☒ Wartości Null ☐

north x [A]

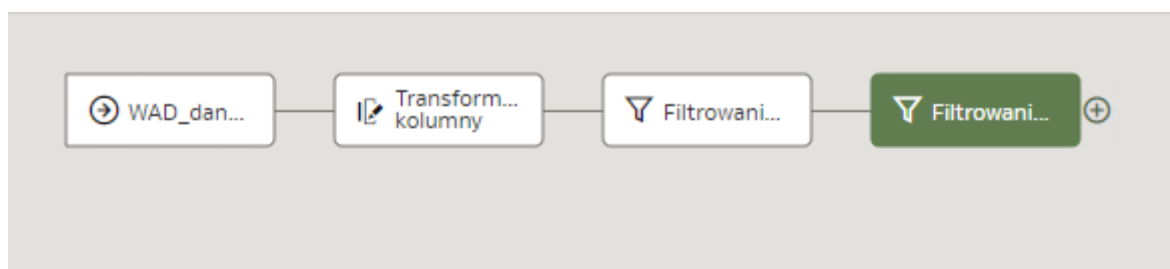
Wykluczenia (4/4)

- Belarus
- Israel
- North Korea
- Russia

Dodaj (0) Wyczyść (4)

Australia	Kiribati	Fruits	Online	annel	02	14	28	09	18	18
-----------	----------	--------	--------	-------	----	----	----	----	----	----

Ograniczamy datę zamówień jedynie do tych, które zostały zrealizowane w roku 2016.



Filtrowanie

☆ Order Date

Order Date

Przedział dat

Zakres ▼

1.01.2016



- 31.12.2016



Agregacja

Agregacja	Funkcja	Nazwa nowej kolumny	Grupuj wg
Total Revenue	Średnia ▼	Total Revenue Średnia	Continent
Units Sold	Suma ▼	Units Sold Suma	Country
Dodaj agregację			Item Type
			Sales Channel
			Order Date

ab Continentab Countrvab Item Tvoeab Sales Channelab Order Dateab Genderab Date of Joiningab Discountab Distance fro...ab Client Categorv99 Total Revenue...

WAD_dan...
Transform... kolumny
Filtrowani...
Filtrowani...
Agregacja
Zapisywan... danych

Zapisywanie zbioru danych

Zbiór danych: Analiza tendencji sprzedażowych-ETL1
Tabela zbiorów danych:
Opis:
Zapisz dane w: Magazyn zbiorów danych
Po uruchomieniu ☐ Monitoruj o określenie zbioru danych

Kolumny

Nazwa	Traktuj jako	Agregacja domyślna
Continent	Atrybut	
Country	Atrybut	
Item Type	Atrybut	
Sales Channel	Atrybut	
Order Date	Atrybut	

2.3.2. ETL 2 – Segmentacja klientów

Wybieramy odpowiednie kolumny:

WAD_dane_gotowe
Filtrowani...
Filtrowani...
Filtrowani...
Agregacja
Zapisywan... danych

Dodawanie danych - WAD_dane_gotowe

Zbiór danych: WAD_dane_gotowe
Opis:
Po ☐ Monitoruj o wybranie zbioru danych uruchomi...

Kolumny: Wszystkie (37) Wybór (23)

☐ Nazwa
☒ Continent
☒ Country
☒ Item Type
☒ Sales Channel
☒ OrderPriority

ab Continent	ab Country	ab Item Type	ab Sales Channel	ab OrderPriority	Order Date	Ship Date	Units Sold
Europe	Italy	Household	Offline	Very High	02.03.2015	20.03.2015	8475
Africa	Libya	Cereal	Online	Medium	14.10.2016	23.11.2016	7874
Africa	Nigeria	Meat	Offline	High	28.11.2014	17.01.2015	9989

Filtrujemy zakres dat zamówień od 01.01.2012 do 31.12.2016

WAD_dan...

Filtrowani...

Filtrowanie

Order Date

Order Date

Przedział dat

Zakres

1.01.2012

-

31.12.2016

Filtrujemy wiek klientów na przedział od 30 do 80.

WAD_dan...

Filtrowani...

Filtrowani...

Filtrowanie

Age in Yrs.

Age in Yrs.

Przedział

Zakres

Początek

30

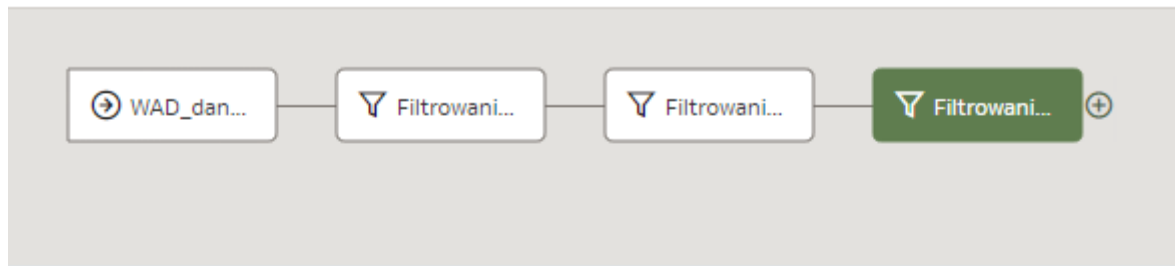
Min.

Koniec

80

Maks.

Pozostawiamy jedynie klientów, którzy są z nami od dawna.



Filtrowanie

Customer Type (1) +

Customer Type

Lista ☐ Wyklucz ☐ Wartości Null

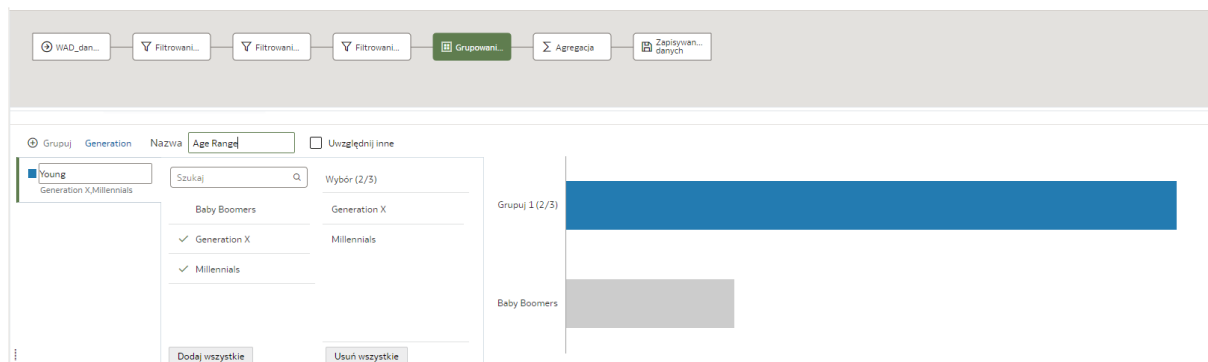
Szukaj [A]

new

Wybory (1/2)

long-time

Dokonujemy grupowania ze względu na pokolenie. Dzielimy klientów na młodych i na starszych.



Dokonujemy agregacji pięciu kolumn, aby określić cechy, styl życia oraz zadowolenie klientów oraz ich wpływu na zamówienia.

WAD_dan...
Filtrowani...
Filtrowani...
Filtrowani...
Grupowani...
Agregacja
Zapisywan... danych

Agregacja	Funkcja	Nazwa nowej kolumny	Grupuj wg
Units Sold	Suma	Units Sold Suma	Continent
Total Revenue	Średnia	Total Revenue Średnia	Country
Total Profit	Suma	Total Profit Suma	Item Type
Age in Yrs.	Średnia	Age in Yrs. Średnia	Sales Channel
Weight in Kgs.	Średnia	Weight in Kgs. Średnia	OrderPriority
			Order Date

Dodaj agregację

Zapisujemy zbiór danych:

WAD_dan...
Filtrowani...
Filtrowani...
Filtrowani...
Grupowani...
Agregacja
Zapisywan... danych

Zapisywanie zbioru danych

Zbiór danych
Segmentacja klientów - ETL2
Tabela zbiorów danych
Opis
Zapisz dane w
Magazyn zbiorów danych

Kolumny

Nazwa	Traktuj jako	Agregacja domyślna
Continent	Atrybut	
Country	Atrybut	
Item Type	Atrybut	
Sales Channel	Atrybut	
OrderPriority	Atrybut	

Po uruchomieniu
☐
Monituj o określenie zbioru danych

2.3.3. ETL 3 - Badanie wyników sprzedaży

WAD_dan...
Filtrowani...

Dodawanie danych - WAD_dane_gotowe

Zbiór danych
WAD_dane_gotowe
Wybór...
Opis

Po
☐
Monituj o wybranie zbioru danych
uruchomi...

Kolumny
Wszystkie (37)
Wybór (20)

☐ Nazwa
☒ Continent
☒ Country
☒ Item Type
☒ Sales Channel
☒ OrderPriority

ab	Continent	ab	Country	ab	Item Type	ab	Sales Channel	ab	OrderPriority	Order Date	99	Order ID	Ship Date
	Europe		Italy		Household		Offline		Very High	02.03.2015		382497870	20.03.2015
	Africa		Libya		Cereal		Online		Medium	14.10.2016		328833091	23.11.2016

WAD_dan...
Filtrowani...

Filtrowanie

Order Date

Order Date
Przedział dat

Zakres
1.01.2012
-
31.12.2016

WAD_dan...
Filtrowani...
Agregacja
Zapisywan... danych

Agregacja

Agregacja	Funkcja	Nazwa nowej kolumny	Grupuj wg
Units Sold	Suma	Units Sold Suma	Continent
Total Revenue	Średnia	Total Revenue Średnia	Country
Total Profit	Suma	Total Profit Suma	Item Type

Dodaj agregację

WAD_dan...
Filtrowani...
Agregacja
Zapisywan... danych

Zapisywanie zbioru danych

Zbiór danych
Badanie wyników sprzedaży - 3ETL

Tabela zbiorów danych

Opis

Zapisz dane w
Magazyn zbiorów danych

Po uruchomieniu
☐
Monituj o określenie zbioru danych

Kolumny	Nazwa	Traktuj jako	Agregacja domyślna
Continent		Atrybut	
Country		Atrybut	
Item Type		Atrybut	
Sales Channel		Atrybut	
OrderPriority		Atrybut	

2.3.4. ETL 4 - Badanie nastrojów konsumenckich

WAD_dan...

Filtrowani...

Dodawanie danych - WAD_dane_gotowe

Zbiór danych

Opis

WAD_dane_gotowe

Wybór...

Po

☐

Monituj o wybranie zbioru danych

ruchomi...

Kolumny

Wszystkie (37)

Wybór (25)

Q

☐

Nazwa

☒

Continent

☒

Country

☒

Item Type

☒

Sales Channel

☒

OrderPriority

Continent	ab Country	ab Item Type	ab Sales Channel	ab OrderPriority	Order Date	99 Order ID	Ship Date
Europe	Italy	Household	Offline	Very High	02.03.2015	382497870	20.03.2015
Africa	Libya	Cereal	Online	Medium	14.10.2016	328833091	23.11.2016
...

WAD_dan...

Filtrowani...

+

Filtrowanie

Ship Date

Ship Date

Przedział dat

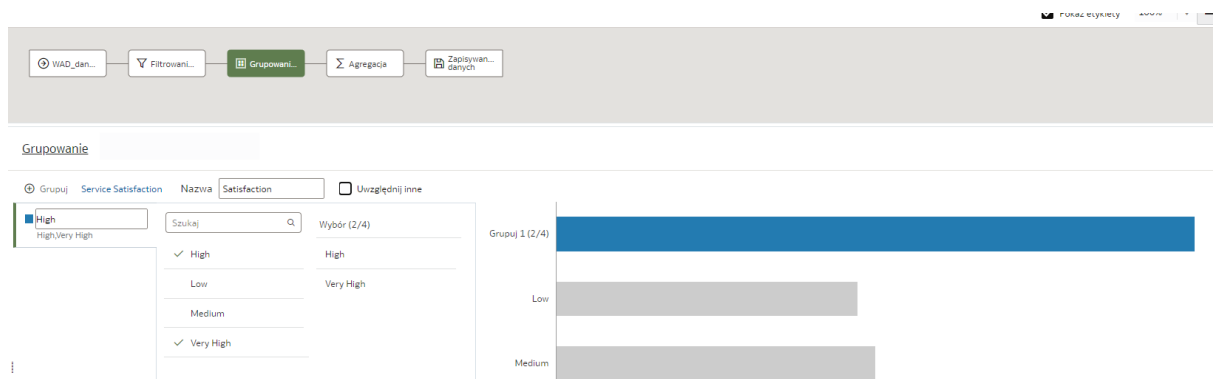
Zacznij od

▼

1.01.2015

📅

Dodajemy etap grupowania klientów ze względu na satysfakcję. Bardzo zadowoleni klienci i zadowoleni będą teraz jako po prostu zadowoleni.



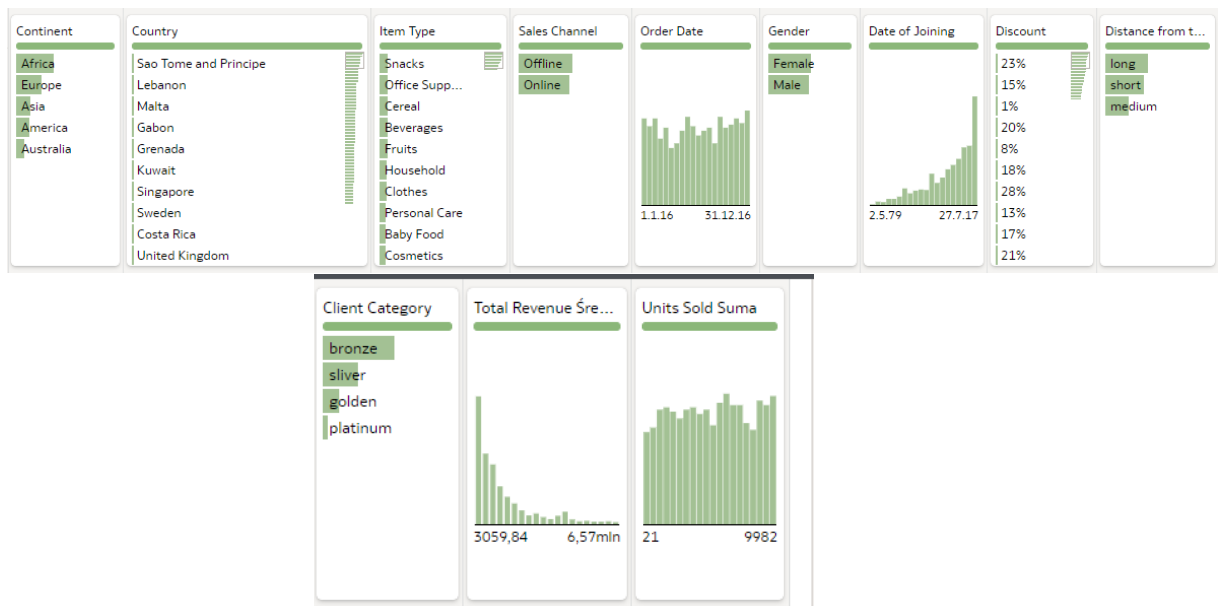
Dokonujemy agregacji:

Zapisujemy zbiór danych jako "Badanie nastrojów konsumenckich – ETL4"

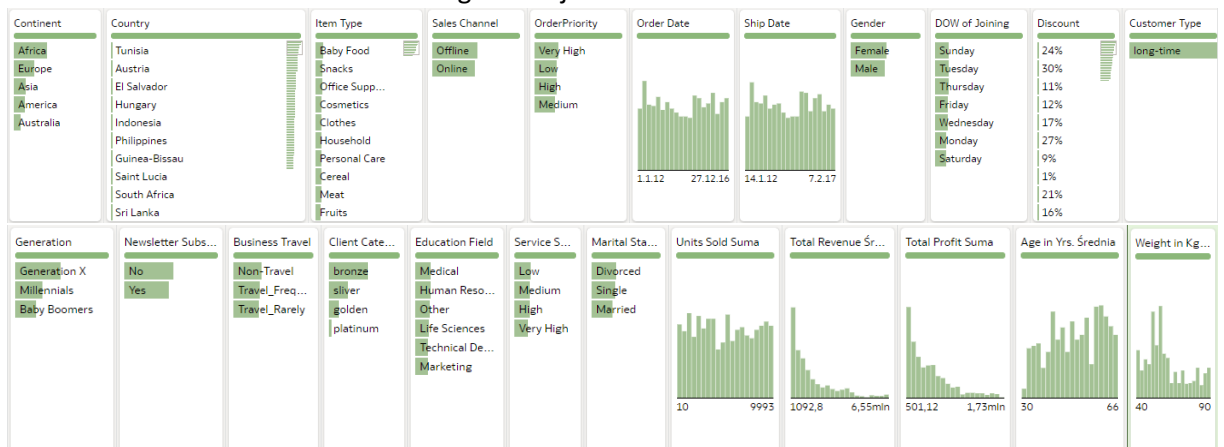
4. Weryfikacja i zapisanie plików do HD

4.1. Weryfikacja

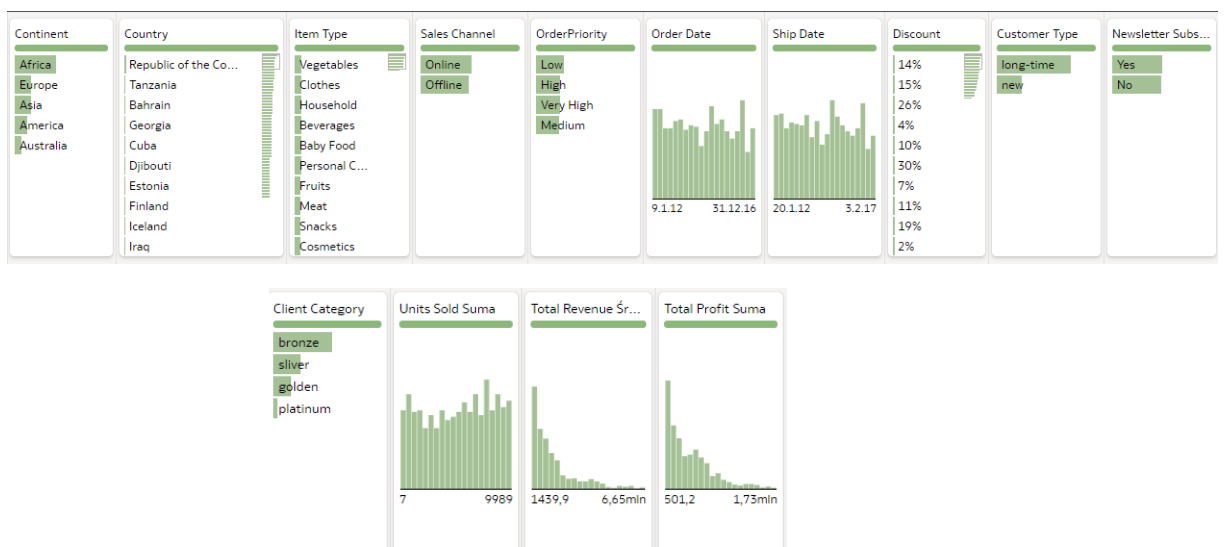
Analiza tendencji sprzedaży:



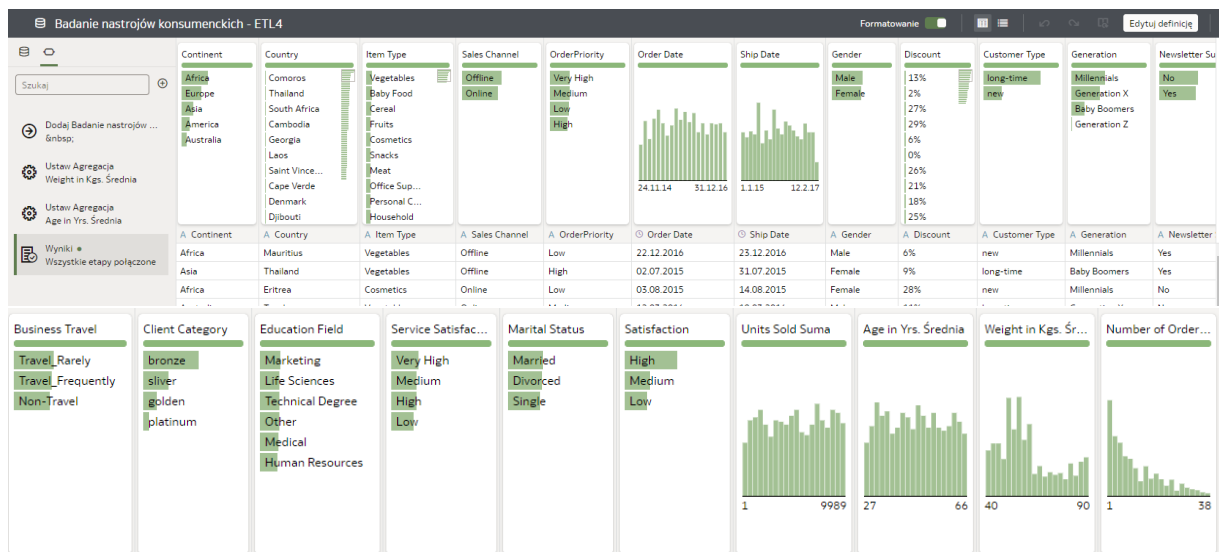
Segmentacja klientów:



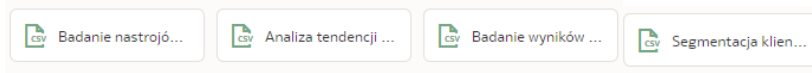
Badanie wyników sprzedaży



Badanie nastrojów konsumenckich:



4.2. Zapisanie plików



5. Scenariusze KPI



Analiza tendencji sprzedaży:

Proces ten skupia się na przetworzeniu danych potrzebnych do analizy sprzedaży według czynników takich jak kategoria klienta, kanał sprzedaży, odległość od sklepu, typ produktu, rabat czy położenie geograficzne. Opracowane scenariusze KPI są odpowiedzialne za dostarczenie cennych informacji, które mogą być użyte do podejmowania decyzji biznesowych i optymalizacji strategii sprzedaży.

Segmentacja klientów:

Proces ten pomaga w zrozumieniu, jak różne czynniki demograficzne, edukacyjne, geograficzne oraz związane z podróżami służbowymi wpływają na tendencje sprzedaży. Analiza tych danych pozwala na lepsze zrozumienie klientów oraz optymalizację strategii marketingowych i sprzedażowych.

Badanie wyników sprzedaży

Te KPI pomagają w zrozumieniu, jak różne produkty, rynki geograficzne i czynniki czasowe wpływają na wyniki sprzedaży i zyskowność firmy. Analiza tych danych umożliwia lepsze podejmowanie decyzji strategicznych, optymalizację oferty produktowej oraz planowanie działań marketingowych i sprzedażowych.

Badanie nastrojów konsumenckich

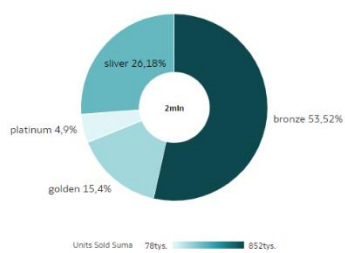
Proces ten pozwala przeanalizować, jak różne czynniki demograficzne, takie jak wiek, dziedzina edukacji oraz status cywilny, wpływają na poziom satysfakcji klientów. Analiza tych danych pozwala na lepszą identyfikację potrzeb klientów oraz optymalizację strategii marketingowych i sprzedażowych w celu zwiększenia zadowolenia klientów.

6. Wizualizacje

Analiza tendencji sprzedaży:

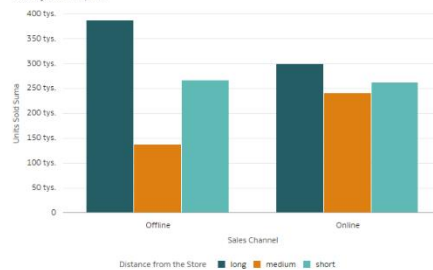
Analiza tendencji sprzedaży

Units Sold Suma wg Client Category, Units Sold Suma
Country: Czech Republic



Units Sold Suma wg Sales Channel, Distance from the Store

Country: Czech Republic



Units Sold Suma, Item Type

Country: Czech Republic

Item Type	Units Sold Suma
Baby Food	107 936
Beverages	125 162
Cereal	115 265
Clothes	185 146
Cosmetics	115 523
Fruits	198 154
Household	129 527
Meat	127 519
Office Supplies	136 919
Personal Care	107 701
Snacks	107 100
Vegetables	116 668

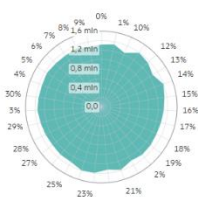
Units Sold Suma wg Country, Units Sold Suma

Filtr oparty na wyrażeniu

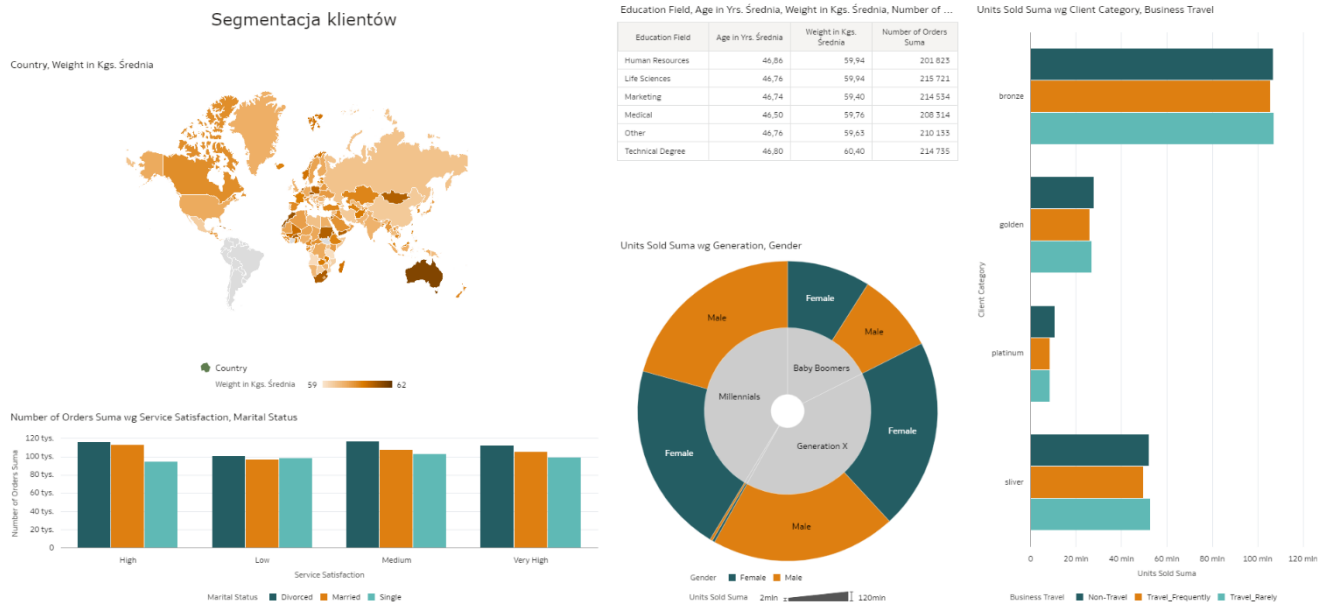


Total Revenue Średnia wg Discount

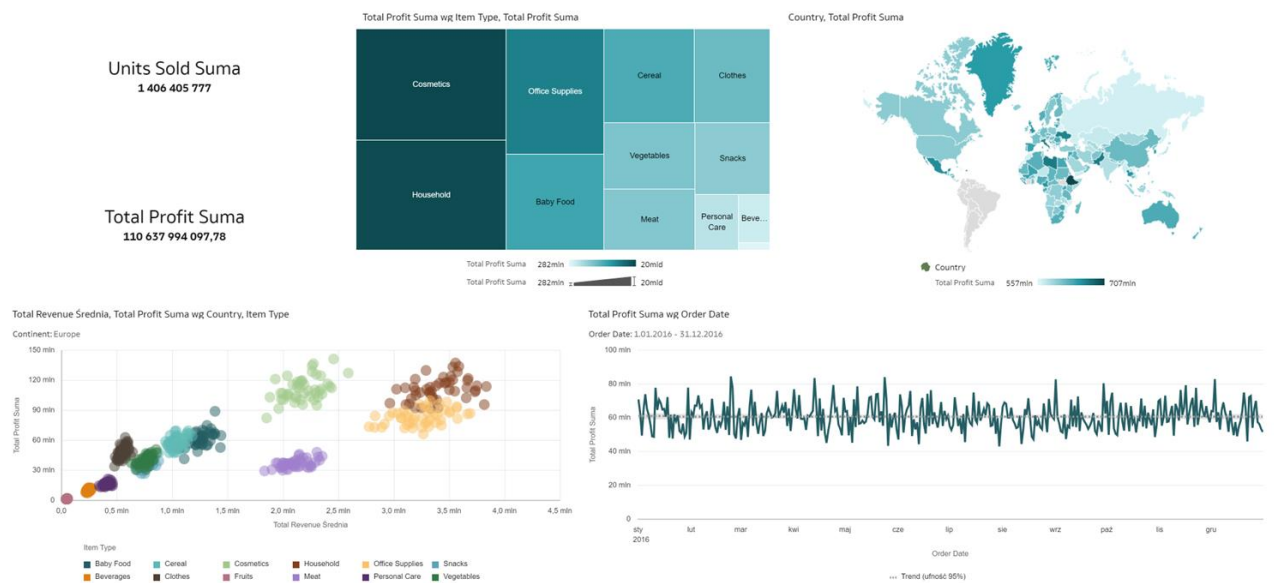
Country: Wszystkie



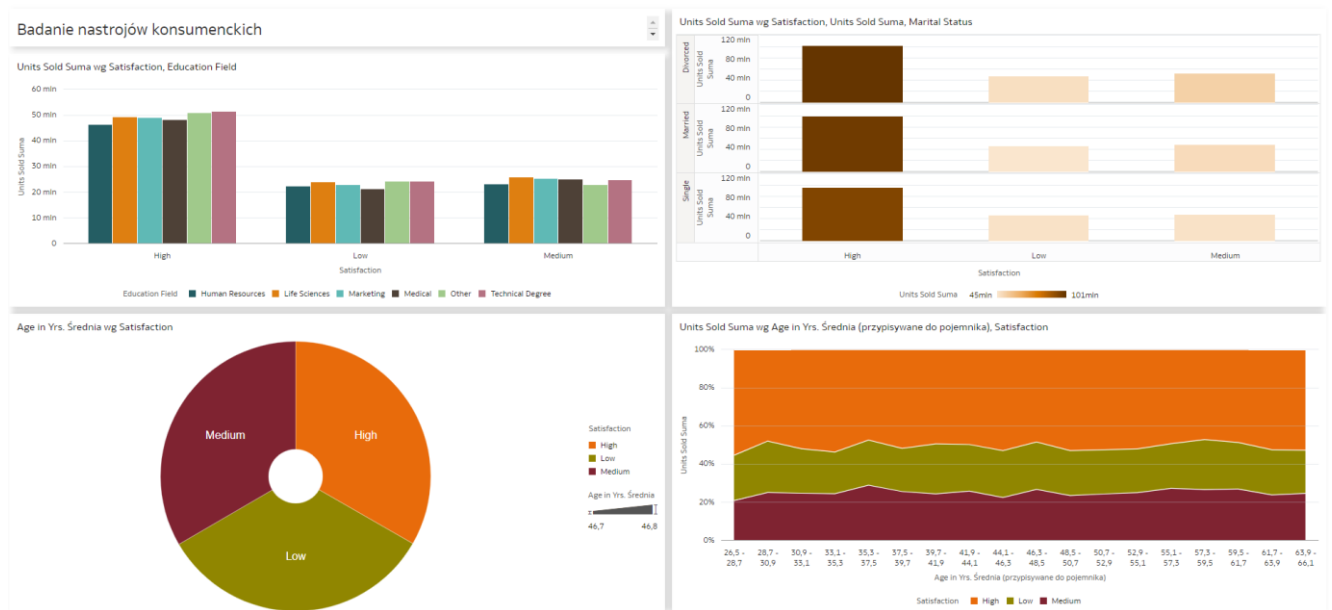
Segmentacja klientów:



Badanie wyników sprzedaży:



Badanie nastrojów konsumenckich:



8. Podsumowanie i wnioski

Projekt pozwolił nam na zrozumienie tematu wielowymiarowej analizy danych. Program Oracle Analytics Desktop okazał się dobrym wyborem w analizę dużych zbiorów informacji z powodu przejrzystego interfejsu. Pozwolił on na przejrzyste i intuicyjne dokonywanie procesów analizy danych dotyczących sprzedaży.