

Raport projektu Covid Data Warehouse z przedmiotu Hurtownie danych i systemy BI

Zuzanna Piróg, Jakub Rymarski

Cel biznesowy

Ostatnie lata uświadomiły ludzkości jak ważne jest zdrowie. Pandemia Covid-19 była dużym wyzwaniem dla całego świata. W zamkniętych mieszkaniach obserwowaliśmy w telewizorach globalny kryzys. Tragiczna sytuacja w szpitalach, coraz gorsze statystyki śmiertelności z powodu wirusa i przerażające egzekwowanie kwarantanny w Chinach pokazywały jak wielkim problemem jest pandemia. Trzeba było szybko i mądrze go rozwiązać. Po pewnym czasie, eksperci opracowali szczepionki, które miały być skutecznym orężem w walce z chorobą. Jako, że współczesny świat mierzy się z pandemią na tak szeroką skalę po raz pierwszy, istnieje wiele kontrowersji co do sposobów jej zwalczania, m.in. odnośnie skuteczności szczepionek.

Celem biznesowym naszego projektu jest zwiększenie skuteczności działań w walce z pandemią Covid-19, poprzez lepsze zrozumienie dynamiki rozprzestrzeniania się wirusa oraz efektywniejsze podejmowanie decyzji politycznych na poziomie lokalnym i globalnym. Kolejnym celem jest szybsze identyfikowanie obszarów o wysokim ryzyku, które skutkować będzie wprowadzaniem odpowiednich środków zaradczych. Projekt może również przyczynić się do dalszego rozwoju nauk medycznych poprzez analizę skuteczności stosowanych szczepionek oraz identyfikację czynników wpływających na przebieg choroby.

Stworzyliśmy hurtownię danych dotyczącą pandemii Covid-19. Analizy na podstawie zgromadzonych danych umożliwiły zobrazowanie przebiegu pandemii Covid-19 w różnych regionach świata. Dane i statystyki na temat choroby wzbogaciliśmy o różne wskaźniki ekonomiczne i gospodarcze, tak aby dostarczyć klientowi kompleksowe dane i możliwość tworzenia szczegółowych raportów. Nasze rozwiązanie przeznaczone jest dla lekarzy i epidemiologów, którzy będą dzięki niemu mogli analizować rozwój pandemii i opracowywać odpowiednie strategie działania na wypadek przyszłych kryzysów. Projekt może także okazać się pomocny także dla polityków. Na podstawie danych gospodarczych i ekonomicznych będą mogli oni przewidzieć ryzyko ciężkiego przebiegu pandemii w ich kraju.

Architektura systemu



Korzystaliśmy z MS SQL Management Studio w celu zarządzania bazą danych oraz jej modelowania. Ponadto w SSIS (SQL Server Integration Services) zaimplementowaliśmy operacje ETL. W celu wizualizacji danych skorzystaliśmy z programu Power BI, który posłużył nam do generowania raportów.

Opis zbiorów danych

Podstawowym źródłem danych była strona: Coronavirus (COVID-19) Vaccinations, Our World in Data, <https://ourworldindata.org/covid-vaccinations>, na której znajdują się dane odnośnie rozwoju pandemii Covid-19 w większości państw świata. Zawierają m.in informacje o nowych zachorowaniach, śmiertelności czy szczepieniach. Są aktualizowane codziennie, począwszy od 05.01.2020. Dostarczają także informacji na temat różnych wskaźników ekonomicznych i sytuacji służby zdrowia w państwie.

Dane miały 66 kolumn, dlatego tutaj wypiszę tylko te, z których korzystaliśmy: continent, location, date, new_cases, new_deaths, new_tests, people_vaccinated_per_hundred, people_fully_vaccinated_per_hundred, new_vaccinations, aged_65_older, gdp_per_capita, cardiovasc_death_rate, male_smokers, hospital_beds_per_thousand, life_expectancy, human_development_index, population

Dane geograficzne na potrzeby wymiaru lokalizacji pobraliśmy ze strony <https://gist.github.com/ofou/df09a6834a8421b4f376c875194915c9>

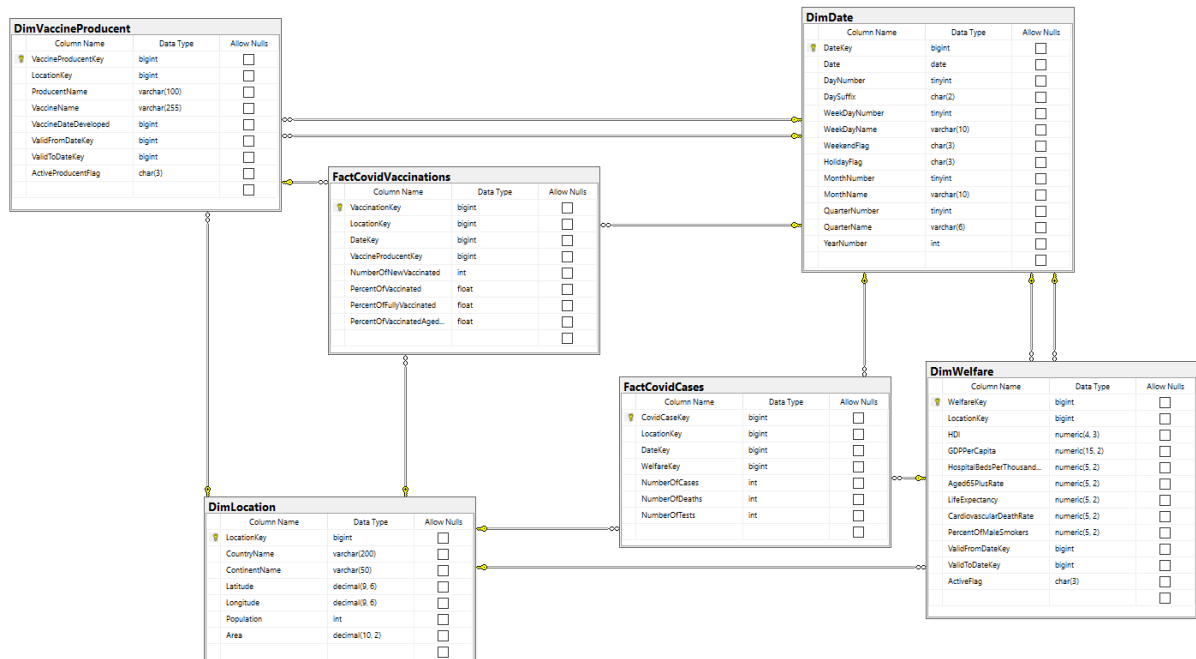
dodając informacje o kontynencie i powierzchni kraju.

Kolumny: CountryId, Country, Capital, Latitude, Longitude, Population, Continent, Area

Na potrzeby biznesowe i zapewnienia śledzenia zmian typu SCD2 w jednym z wymiarów stworzyliśmy własnoręcznie plik csv z wymyślonymi informacjami odnośnie producentów szczepionek.

Kolumny: Location, ProducentName, VaccineName, VaccineDateDeveloped

Schemat hurtowni danych



Nasz model hurtowni składa się z dwóch tabel faktowych i 4 tabeli wymiarów. Tabela faktowa FactCovidCases zawiera klucze obce do wymiarów lokalizacji, daty oraz czynników mierzących dobrobyt. Grupowania po wymiarach pozwalają analizować statystyki dotyczące przebiegu epidemii uwzględniając np. różne wskaźniki ekonomiczne i demograficzne poszczególnych państw w ustalonym okresie. Druga tabela faktowa gromadzi dane dotyczące szczepień. Powiązana jest ona z wymiarem lokalizacji, wymiarem daty i wymiarem producenta szczepionek.

Opis ETL

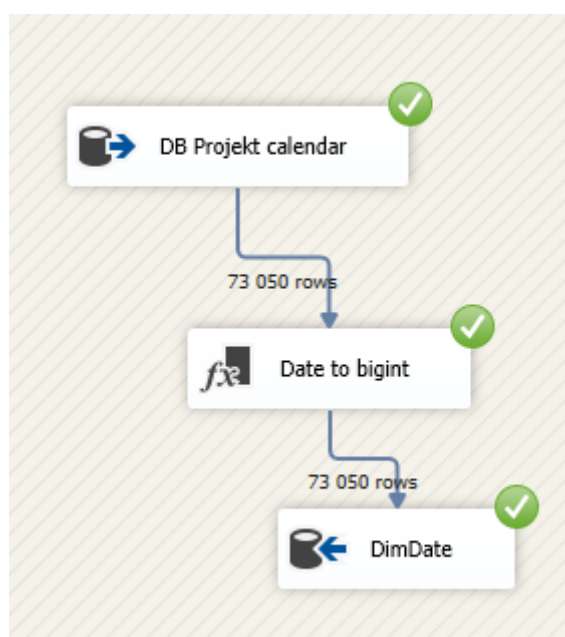
Tabele ze względu na wymagania biznesowe i powiązania między nimi ładowaliśmy w następującej kolejności:

Tabele wymiarów DimDate i DimLocation zostały załadowane tylko raz na początku przy tworzeniu hurtowni. Następnie załadowaliśmy dwie kolejne tabele wymiarowe DimWelfare i DimVaccineProducent. Miały one zapewnione śledzenia zmian typu SCD2. Odbędzie się ono w następujący sposób:

Po aktualizacji danych np. danych producenta, starszemu, dotychczas aktualnemu rekordowi, została ustawiona flaga ważności ActiveProducentFlag na "No". Polu ValidToDateKey została przypisana data, w której odnotowano zmianę. Nowy rekord dla tego dostawcy otrzymał flagę ActiveFlag "Yes" i odpowiednią datę w kolumnie ValidFromDateKey.

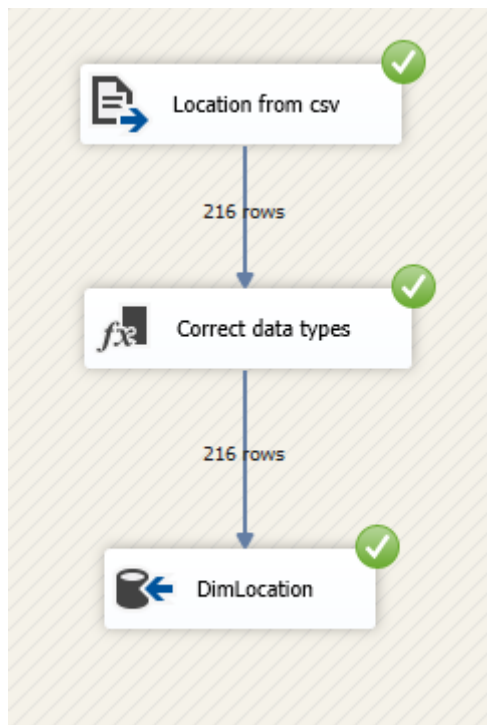
1. DimDate

Wymiar daty był ładowany raz na początku tworzenia hurtowni. Tworząc go skorzystaliśmy z szablonu i danych z udostępnionej na Teams tabeli Calendar. Na podstawie skryptu stworzyliśmy tabelę Calendar w pomocniczej bazie - Projekt. Wybraliśmy odpowiednie pola i za pomocą transformacji przekształciliśmy datę w formacie date na liczbę bigint zgodnie z hurtownianą konwencją, tak więc np. datę 2020-12-31 zapisywaliśmy jako liczbę 20201231.



2. DimLocation

Kolejny wymiar to wymiar lokalizacji. Ładowany on był również tylko raz przy tworzeniu hurtowni. Pobraliśmy dane z pliku csv zawierającego potrzebne dla nas informacje o państwach świata. Wybraliśmy interesujące nas kolumny i ustawiliśmy odpowiednie formaty dla danych kolumn.



3. DimVaccineProducent

Wymiar producenta szczepionki miał mieć zapewnioną możliwość śledzenia zmian typu SCD2. Na potrzeby biznesowe nazwa producenta była niezmienna. W danym momencie na rynku mogła być dostępna tylko jedna szczepionka danego producenta. Tak więc gdy producent zdecydował się udostępnić nową szczepionkę, trzeba było archiwizować informacje o jego dotychczasowej szczepionce i wprowadzić do hurtowni informację o tej nowej, aktualnej. Dodatkowo założyliśmy, że zmieniać może się także lokalizacja producenta.

- Vaccine producent from csv

Dane pobieramy z csv, która częściowo zawiera wymyślone dane. Nie udało nam się znaleźć dokładnych informacji, który pacjent był szczepiony daną szczepionką, dlatego przyjęliśmy, że w jednym państwie wszystkich szczepiono preparatem jednej firmy. Szczepionki do krajów przydzieliliśmy losowo. Dane w pliku wyglądały np. tak: United States,Pfizer,ExtraVac,2020-12-11

- Data type conversion

Tutaj ustawiamy odpowiednie typy danych:

Location

(DT_STR,200,1250) Location

ProducentName

(DT_STR,100,1250) ProducentName

VaccineDateDeveloped

(DT_I8)(SUBSTRING((DT_WSTR,10)VaccineDateDeveloped,1,4) +
SUBSTRING((DT_WSTR,10)VaccineDateDeveloped,6,2) +
SUBSTRING((DT_WSTR,10)VaccineDateDeveloped,9,2))

VaccineName

(DT_STR,255,1250)VaccineName

- Wrong data types

W pliku tym zbieramy ewentualne błędy przy konwersji danych

- Lookup for location

Jako, że Location było nazwą państwa a w tabeli wymiarowej przechowujemy informację odnośnie lokalizacji jako odpowiedni klucz obcy LocationKey w formacie bigint. Lookup ten służy nam więc do znajdowania LocationKey dla nazwy kraju (CountryName) odpowiadającej naszej Location z pliku csv.

- Lookup producent name

Ten lookup z kolei jest już częścią logiki SCD2. Szukamy w nim, którzy producenci są już zapisani w hurtowni.

- Lookup changes

W tym kroku wybieramy te rekordy, które powinny zostać zarchiwizowane

- Multicast

Rozdzielamy dane na insert i update zgodnie z SCD2.

- Union all

Łączymy nowe i zmodyfikowane rekordy

- Rows to be updated

W tym kroku ustawiamy odpowiednio ValidToDateKey i ActiveProducentFlag dla nieaktualnych już rekordów.

```

UPDATE DimVaccineProducent
SET [ValidToDateKey] = CONVERT(BIGINT, CONVERT(VARCHAR(8),
GETDATE(), 112)),
    [ActiveProducentFlag] = 'No'
WHERE [ProducentName] = ? and [ActiveProducentFlag] = 'Yes'

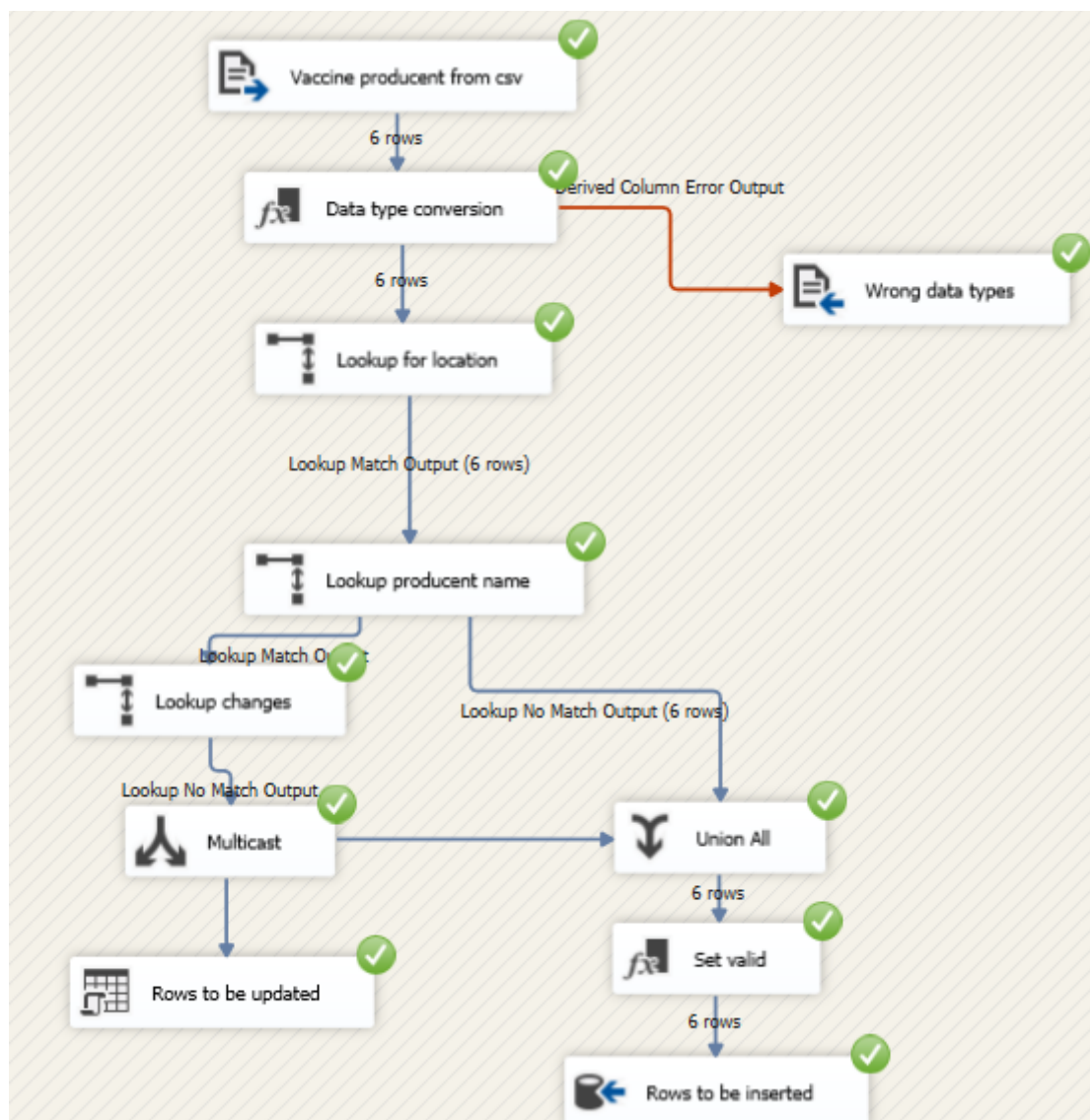
```

- Set valid

Dodajemy kolumny ValidFromDateKey, ValidToDateKey, ActiveProducentFlag dla insertowanych aktualnych rekordów.

- Rows to be inserted

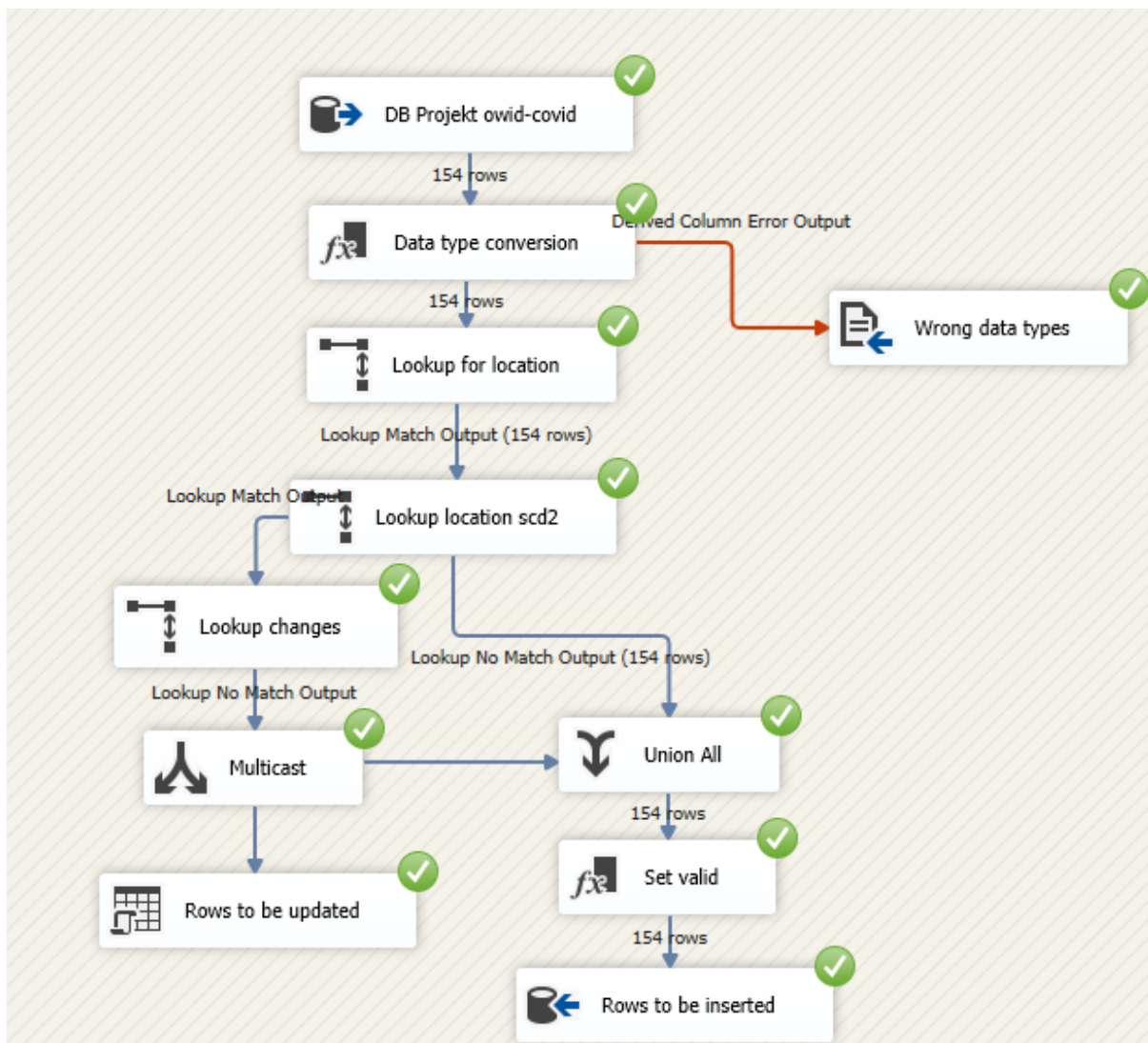
Wstawiamy aktualne nowe rekordy do hurtowni. Tu warto wspomnieć, że w hurtowni ustawiliśmy samoczynnie inkrementowany klucz zastępczy VaccineProducentKey.



4. DimWelfare

W wymiarze dobrobytu, w którym przechowywane są statystyki związane ze służbą zdrowia i wskaźniki ekonomiczne, także zagwarantowaliśmy możliwość śledzenia zmian typu SCD2, aby móc monitorować np. przyrost liczby łóżek w szpitalach. Proces ETL dla tej tabeli przebiegał podobnie jak dla tabeli DimVaccineProducent. Tym razem jednak pobieraliśmy dane z pomocniczej bazy Projekt z tabeli owid-covid-data.

Tak jak poprzednio musieliśmy znaleźć odpowiedni LocationKey aby móc połączyć naszą tabelę z wymiarem lokalizacji. Wymiar lokalizacji pełnił rolę identyfikatora rekordu. Za pomocą lookupa sprawdzaliśmy czy dana lokalizacja znajduje się już w hurtowni. Zmieniane i archiwizowane zgodnie z założeniami SCD2 były statystyki opisujące dobrobyt.



5. FactCovidCases

- DW Projekt owid-covid

Tabela faktowa dostarczająca danych o liczbie zachorowań, śmierci i testów na covid była ładowana w zaprezentowany poniżej sposób. Dane pobieramy z tabeli owid-covid z pomocniczej bazy - Projekt.

- Data type transformation

Przekształcamy odpowiednio typy danych w kolumnach, datę na bigint i stringi new_cases, new_deaths na int. Jako, że pole new_tests w wejściowej tabeli mogło być puste, przyjęliśmy na potrzeby biznesowe, że w takim przypadku, możemy uznać wartość pustą jako brak testów tego dnia. Zamieniamy więc puste stringi na liczbę zero.

```
(DT_I4)(DT_R8)(LEN(TRIM(new_tests1)) == 0 ? "0" : REPLACE(new_tests1, ".", ","))
```

- Lookup date

Ten lookup służy do sprawdzenia czy dana data występuje w wymiarze DimDate

- Lookup location

Podobnie w tym lookupie sprawdzamy czy istnieje dana lokalizacja w tabeli DimLocation

- Lookup welfare

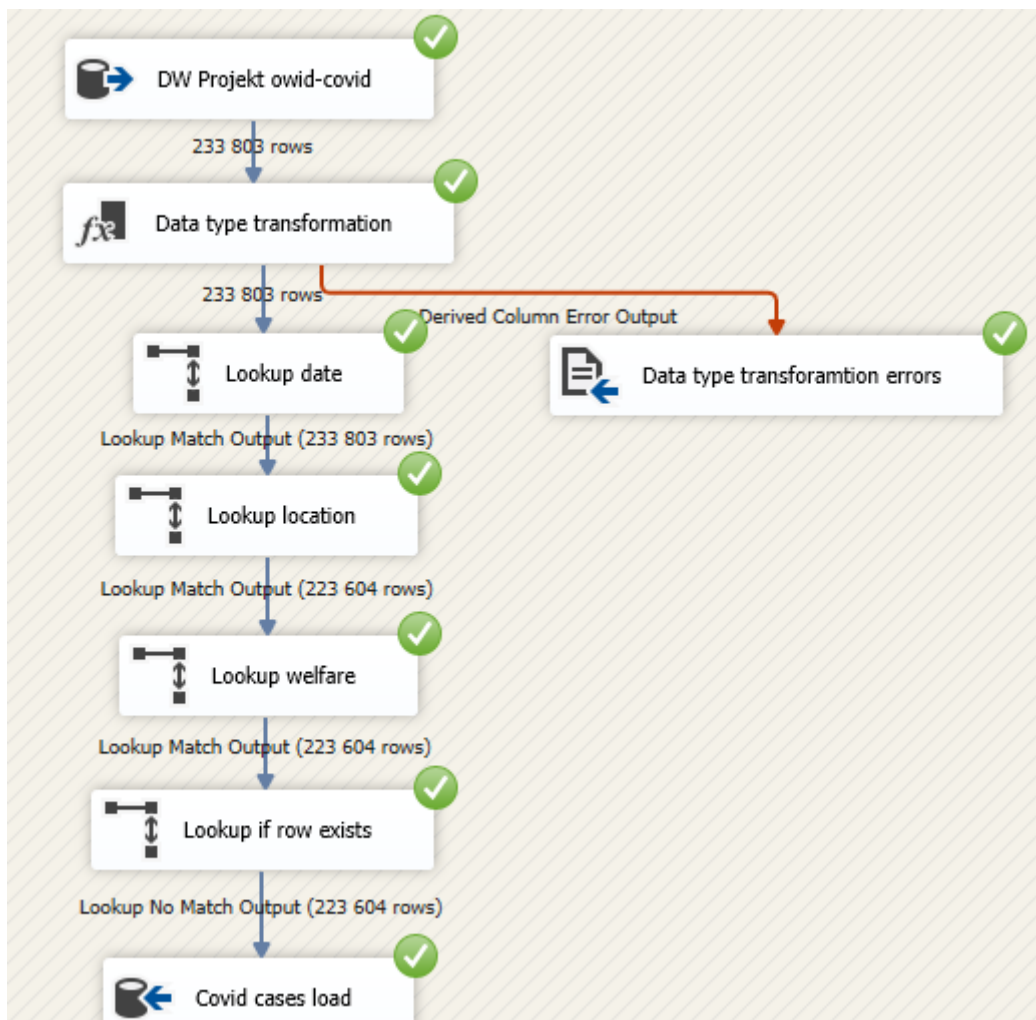
Sprawdzamy czy istnieje dany rekord w DimWelfare

- Lookup if row exists

Tutaj już sprawdzamy czy dany rekord identyfikowany jednoznacznie po LocationKey i DateKey istnieje już w hurtowni.

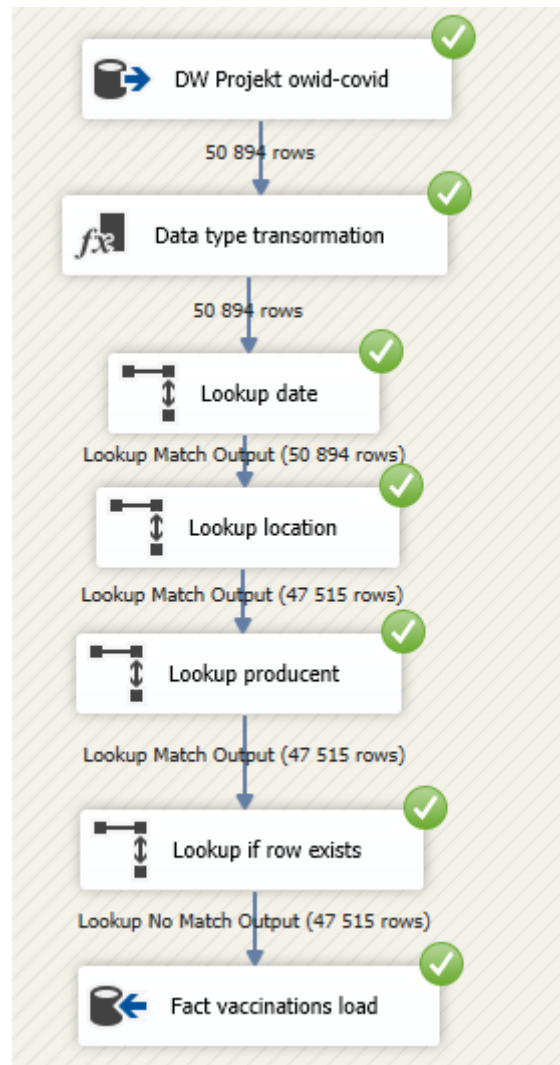
-Covid cases load

Wstawiamy dane to tabeli faktowej FactCovidCases



6. FactCovidVaccinations

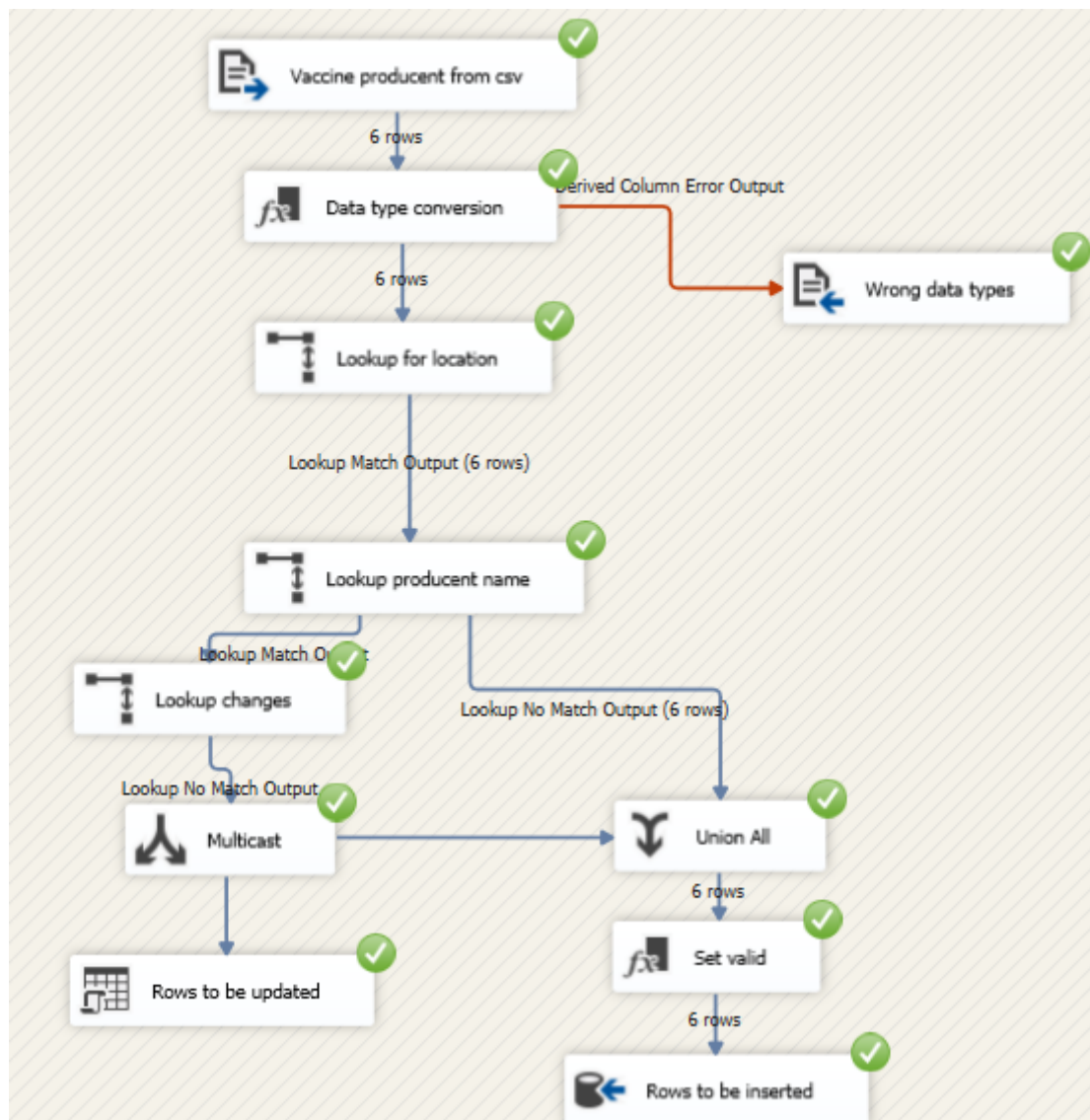
W tej tabeli faktowej przechowujemy informacje związane ze szczepieniami. Proces ETL został zaprojektowany podobnie jak w poprzednim przypadku. Mieliśmy oczywiście inne tabele wymiarowe. Podobnie jak w poprzednim wypadku ładowaliśmy dane z pomocniczej bazy. Tym razem jednak okazało się, że statystyki dotyczące szczepień prowadzi mniej państw, dlatego przefiltrowaliśmy odpowiednio rekordy, żeby nie uwzględniać państw dla których w ogóle nie było żadnych danych odnośnie szczepień. Jedno z używanych pól w tabeli owid-covid - new_vaccinations mogło przyjmować puste stringi, dlatego w transformacji kolumn zamieniliśmy takie puste pola na zera przyjmując, że w tym wypadku brak danych można uznać za brak szczepień tego dnia.



Testy ETL

DimVaccineProducent

W tabeli DimVaccineProducent zaimplementowano śledzenie zmian typu SCD 2. Pola ValidFromDateKey, ValidToDateKey, ActiveProducentFlag odpowiadały za wskazywanie aktualnego rekordu danego producenta i przechowywanie ich starszych informacji dotyczących dawnych lokalizacji i nazw szczepionek. Po pierwszym załadowaniu tabeli sytuacja przedstawiała się w następujący sposób:



```
select * from DimVaccineProducent
```

VaccineProducentKey	LocationKey	ProducentName	VaccineName	VaccineDateDeveloped	ValidFromDateKey	ValidToDateKey	ActiveProducentFlag
1	207	Pfizer	ExtraVac	20201211	20240608	99991231	Yes
2	206	Moderna	Spikevax	20210131	20240608	99991231	Yes
3	206	AstraZeneca	Health	20210215	20240608	99991231	Yes
4	206	Johnson	Janssen	20210311	20240608	99991231	Yes
5	159	Sinopharm	CorV	20201230	20240608	99991231	Yes
6	42	Sinovac	CoronaVac	20211230	20240608	99991231	Yes

ETL przebiegł poprawnie, a każdy producent otrzymał flagę aktywności “Yes”. Pola ValidFromDateKey ustawiamy domyślnie na datę załadowania danych do hurtowni, co widać na załączonym obrazku. Oczywiście jako, że szczepionki zostały opracowane już w 2021 roku, moglibyśmy ustawić ręcznie w pierwszym ładowaniu tabeli wcześniejsze daty ValidFromDateKey, żeby zasymulować realne warunki, ale zakładamy, że nie o to chodziło w projekcie, a w wizualizacjach też nie wykorzystywaliśmy tych dat.

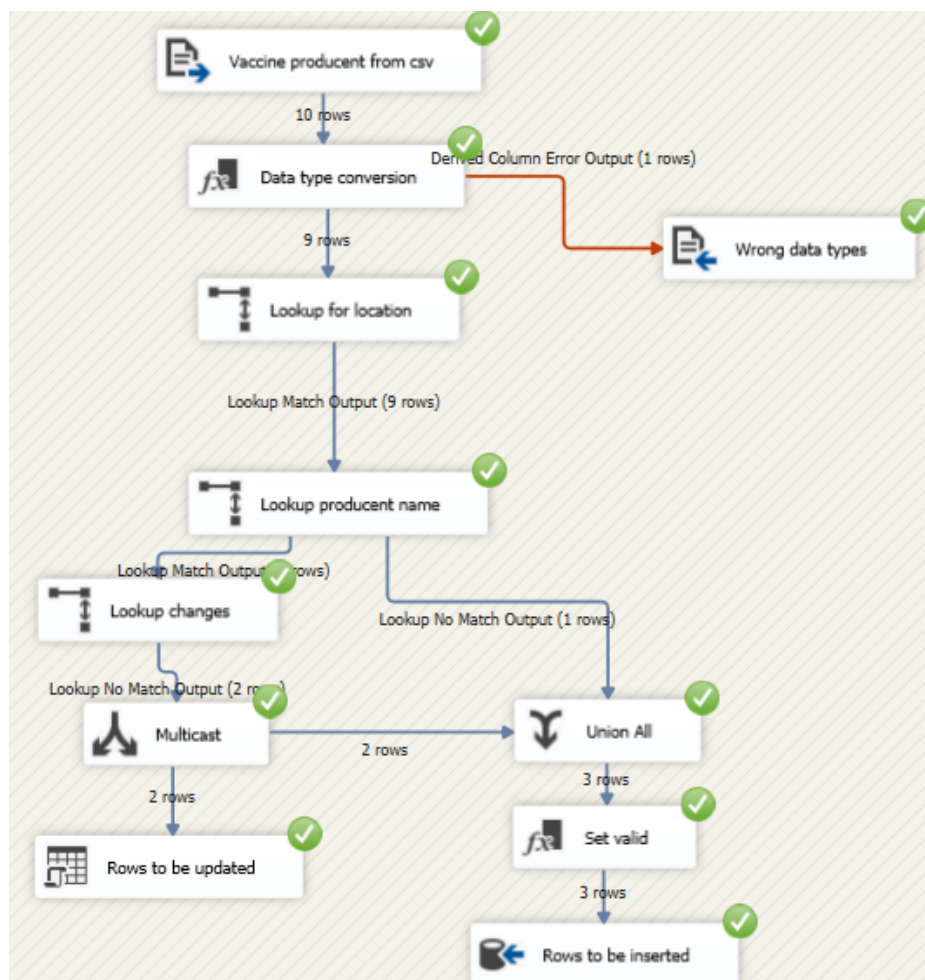
Następnie dodaliśmy do naszego pliku csv trzy nowe linijki postaci:

```
United States,Pfizer,ExtraVacPlus,2023-12-11
United Kingdom,Moderna,SecondEditionSuperModVaccine,2024-02-31
Poland,RatujmyNasLab,Szczepionka123,Mayy-2024|
```

Pfizer i Moderna opracowali nowe szczepionki, więc teraz to one będą tymi aktualnymi. Oczekujemy, że starsze rekordy dla tych producentów zostaną odpowiednio zarchiwizowane.

Nowy kanadyjski producent także udostępnił swoją szczepionkę. W hurtowni nie było jeszcze informacji o tym producencie, dlatego powinien zostać dodany nowy rekord.

Dodatkowo młoda aspirująca polska firma opracowała swoją szczepionkę, jednak niedoświadczone sekretarki wpisały niepoprawną datę wynalezienia szczepionki. Ten rekord powinien być uznany jako błąd a sekretarki muszą go najpierw naprawić przed załadowaniem rekordu do bazy

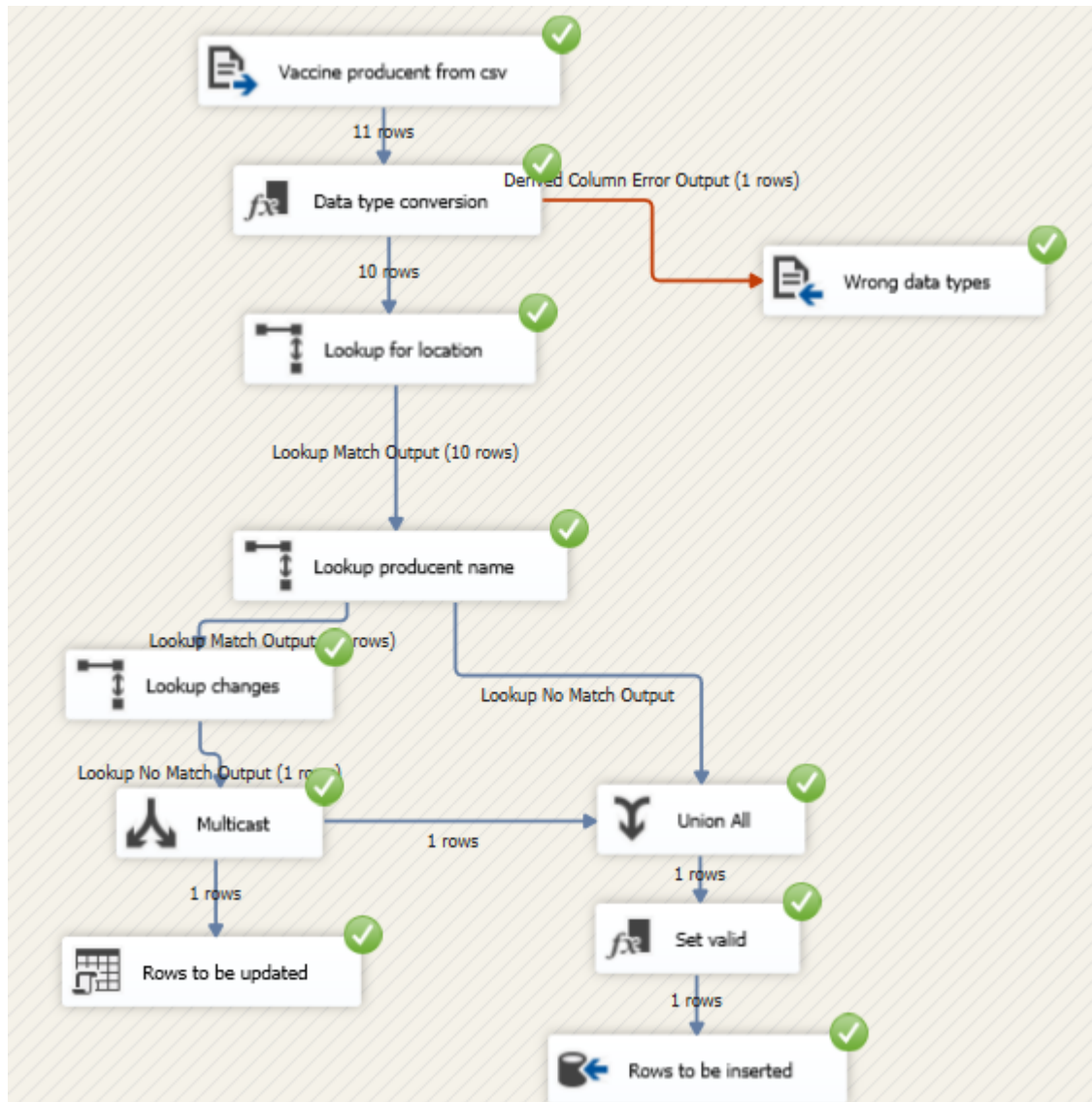


VaccineProducentKey	LocationKey	ProducentName	VaccineName	VaccineDateDeveloped	ValidFromDateKey	ValidToDateKey	ActiveProducentFlag
1	207	Pfizer	ExtraVac	20201211	20240608	20240608	No
2	206	Moderna	Spikevax	20210131	20240608	20240608	No
3	206	AstraZeneca	Health	20210215	20240608	99991231	Yes
4	206	Johnson	Janssen	20210311	20240608	99991231	Yes
5	159	Sinopharm	CorV	20201230	20240608	99991231	Yes
6	42	Sinovac	CoronaVac	20211230	20240608	99991231	Yes
7	37	CompanyMed	VaccineNational1	20240212	20240608	99991231	Yes
8	207	Pfizer	ExtraVacPlus	20231211	20240608	99991231	Yes
9	206	Moderna	SecondEditionSuperModVaccine	20240231	20240608	99991231	Yes

Widzimy, że ETL przebiegł tak jak byśmy tego oczekiwali, jeden rekord został polskiego producenta został sklasyfikowany jako błąd przy próbie konwertowania daty. Starsze rekordy dla Pfizera i Moderny - producentów którzy istnieli już w bazie - zostały odpowiednio zmodyfikowane, a ich nowe rekordy i nowy rekord kanadyjskiej firmy zostały dodane do hurtowni z aktywną flagą.

Możemy pójść jeszcze dalej i sprawdzić czy dla kolejnych aktualizacji szczepionki danego producenta np. Pfizera proces przebiegnie poprawnie. W tym celu dodaliśmy do pliku csv nową linijkę:

United States,Pfizer,ExtraVacPlus2,2024-01-11



```

select * from DimVaccineProducent
where ProducentName='Pfizer'

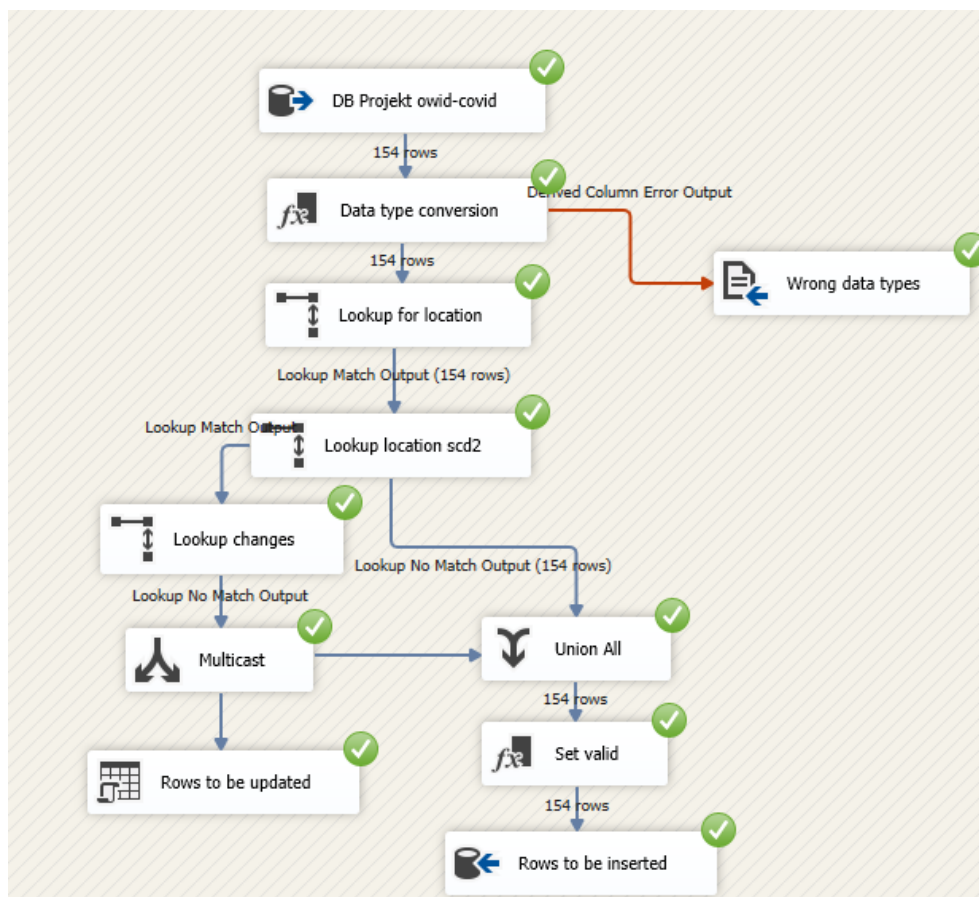
```

VaccineProducentKey	LocationKey	ProducentName	VaccineName	VaccineDateDeveloped	ValidFromDateKey	ValidToDateKey	ActiveProducentFlag
1	207	Pfizer	ExtraVac	20201211	20240608	20240608	No
8	207	Pfizer	ExtraVacPlus	20231211	20240608	20240609	No
10	207	Pfizer	ExtraVacPlus2	20240111	20240609	99991231	Yes

DimWelfare

Sprawdźmy poprawność działania mechanizmu śledzenia zmian SCD2 w wymiarze DimWelfare. Załadowaliśmy najpierw informacje o statystykach ekonomicznych z pomocniczej bazy danych Projekt z tabeli owid-covid-data. Następnie zmodyfikowaliśmy w niej 4 rekordy, zmieniając liczbę w kolumnie PercentOfMaleSmokers dla 4 wybranych krajów.

Pierwsze ładowanie danych:

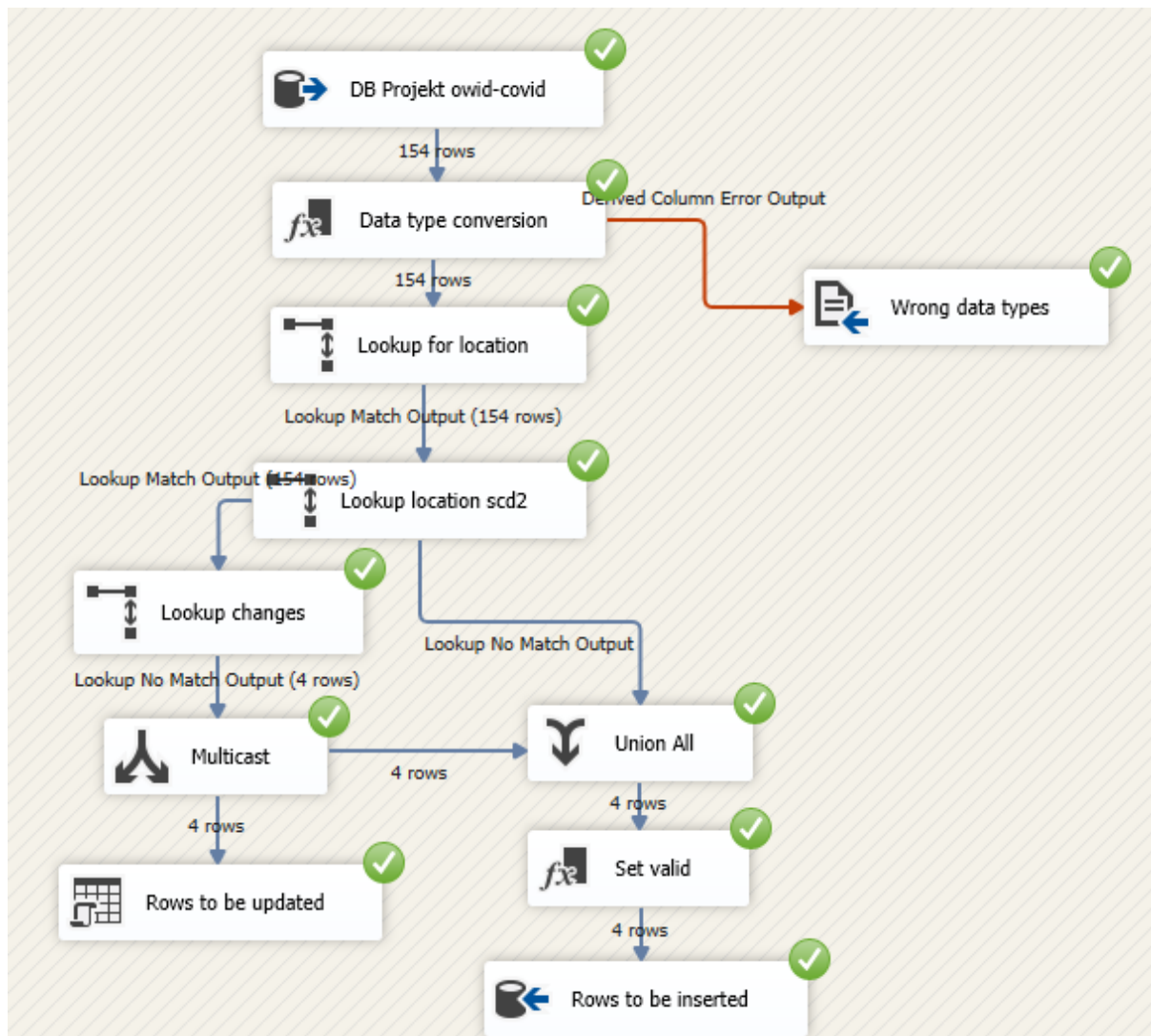


```
select COUNT(*) from DimWelfare
```

(No column name)

154

Drugie ładowanie danych:



```
select COUNT(*) from DimWelfare
```

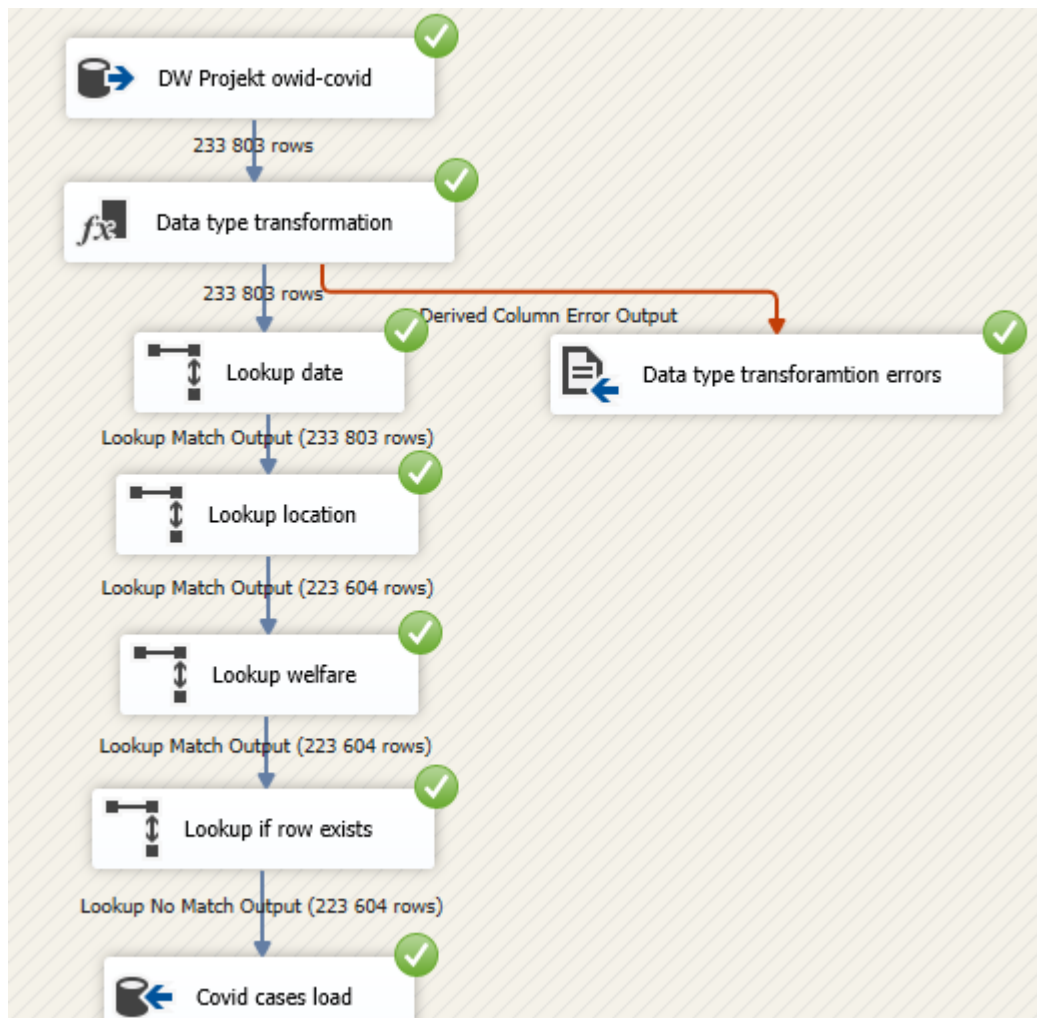
(No column name)
158

Widać, że ETL działa poprawnie

FactCovidCases

Sprawdzimy, czy ładowanie danych do tabeli faktowej wygląda poprawnie. Za pierwszym razem załadowaliśmy dane dla dat do końca 2023 roku, a następnie także dane z pierwszego tygodnia 2024 roku.

Pierwsze załadowanie do dnia 2023.12.31:



Po załadowaniu danych do dnia 2024.01.08:



```
select Count(*) from FactCovidCases  
where DateKey>20231231
```

(No column name)
1200

Lookupy wykryły rekordy, które już były w hurtowni. Wstawiono do tabeli tylko nowe dane z pierwszego tygodnia 2024 roku.

Dla tej tabeli faktycznym wartym uwagi jest także sprawdzenie, czy puste pola w kolumnie NumberOfTests zostały zamienione na liczbę 0. Przykładowo rekord dla Malty z dnia 2024-01-07 nie miał wypełnionego pola new_tests:

```
select location, date, new_cases, new_deaths, new_tests from [owid-covid-data]
where date='2024-01-07' and location='Malta'
```

location	date	new_cases	new_deaths	new_tests
Malta	2024-01-07	110.0	2.0	

Sprawdźmy także czy po załadowaniu do hurtowni mamy w polu NumberOfTests liczbę 0.

```
select LocationKey, CountryName from DimLocation where CountryName='Malta'
```

LocationKey	CountryName
127	Malta

```
select * from FactCovidCases
where DateKey=20240107 and LocationKey=127
```

CovidCaseKey	LocationKey	DateKey	WelfareKey	NumberOfCases	NumberOfDeaths	NumberOfTests
224591	127	20240107	62	110	2	0

Widzimy, że transformacja kolumny przebiegła poprawnie.

Dla tej tabeli możemy wykonać przykładowe zapytanie, sprawdzając czy połączenia z innymi tabelami są dobrze utworzone. Załóżmy, że klient biznesowy poprosił nas o informacje odnośnie liczby nowych zachorowań na covid w marcu 2023 roku.

```
select SUM(NumberOfCases) from FactCovidCases f
join DimLocation l on f.LocationKey=l.LocationKey
where l.CountryName='Poland' and f.DateKey>=20230301 and f.DateKey<=20230401
group by f.LocationKey
```

(No column name)
73177

Wszystko się zgadza, klient jest zadowolony.

Raporty

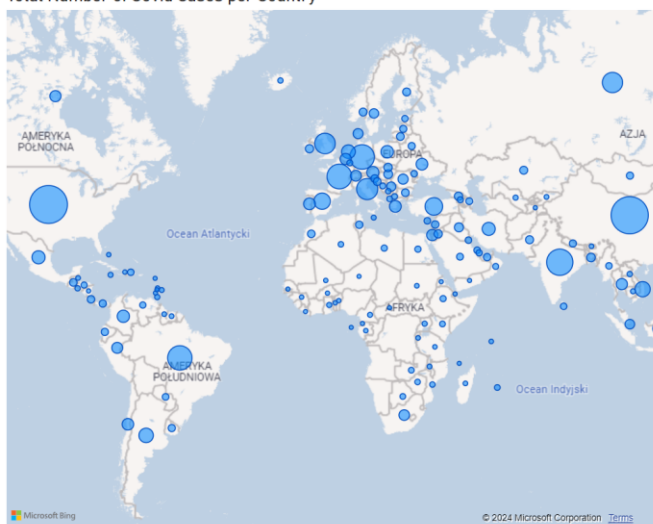
Raporty tworzone są przy pomocy narzędzia PowerBI, które umożliwia utrzymywanie aktywnego połączenia do bazy danych i generowanie interaktywnych raportów. Każdy raport składa się z analiz zgromadzonych na 3 podstronach.

Covid Cases Around the World

Total Covid Cases Total Number of Deaths Total Number of Tests

762M 7M 5bn

Total Number of Covid Cases per Country



Country

All

Year

2019

2025

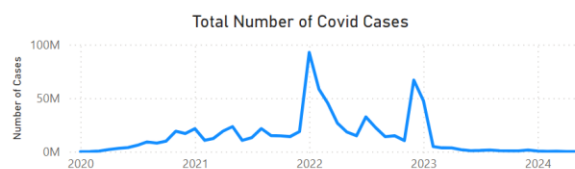
ContinentName

Covid Cases

Deaths

Tests

Asia	283622051	1519621	1646370232
Europe	263546905	2152831	2033095142
France	38997490	167985	278234000
Germany	38437756	174979	0
Italy	26722507	197017	224719845
United Kingdom	24927820	232112	502482746
Russia	24169359	402870	156015815
Turkey	17004714	101419	163164533
Spain	13980340	121852	93162168
Netherlands	8636310	22986	30687346
Poland	6662732	120711	0
Austria	6082347	22534	188768459
Portugal	5643747	28156	42825406
Greece	5633940	39011	48561395
Ukraine	5531611	109920	13377389
Belgium	4861861	34339	34315605
Switzerland	4453795	14188	21277734
Total	762020683	6858869	4978456750



Pierwsza strona dokumentu dotyczy ogólnego zagadnienia związanego z przypadkami zachorowań na Covid. Dzięki drop-down list oraz suwaka możliwe jest zawężenie odczytywanych informacji. W skład wizualizacji wchodzi m.in. mapa reprezentująca łączną liczbę zachorowań w danym kraju, wykres liniowy umożliwiający obserwację zmian tego trendu w czasie, a także tabela zawierająca szczegółowe dane dla każdego kraju.

Na podstawie raportu jesteśmy w stanie śledzić zmiany liczby zachorowań w czasie. Możliwość zawężenia obszaru umożliwia łatwą identyfikację potencjalnych terenów epidemiologicznych i szybką reakcję w odpowiedzi na te zmiany.

Testy:

Zobaczmy czy wykonując odpowiednie zapytanie w hurtowni otrzymamy taką tabelkę jak w prawym górnym rogu:

```

select l.CountryName, SUM(NumberOfCases) as total_cases,
SUM(NumberOfDeaths) as total_deaths,
SUM(NumberOfTests) as total_tests from FactCovidCases c
join DimLocation l on l.LocationKey=c.LocationKey
where l.ContinentName='Europe'
group by l.CountryName, l.ContinentName
order by total_cases desc

```

	CountryName	total_cases	total_deaths	total_tests
1	France	38997490	167985	278234000
2	Germany	38437756	174979	0
3	Italy	26722507	197017	224719845
4	United Kingdom	24927820	232112	502482746
5	Russia	24169359	402870	156015815
6	Turkey	17004714	101419	163164533
7	Spain	13980340	121852	93162168
8	Netherlands	8636310	22986	30687346
9	Poland	6662732	120711	0
10	Austria	6082347	22534	188768459
11	Portugal	5643747	28156	42825406

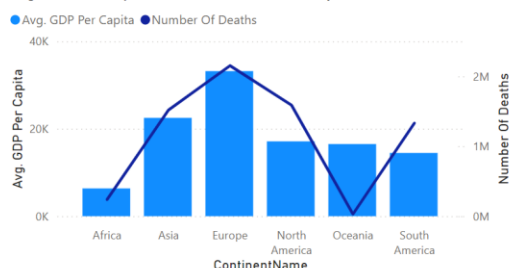
Wyniki się zgadzają. Na tym przykładzie można także zauważyć, że Polska i Niemcy nie udostępniały danych o testach i tak jak to przyjęliśmy na potrzeby biznesowe, w odpowiedniej kolumnie mamy 0.

Wellfare and Covid Cases

73,71
Average of LifeExpectancy

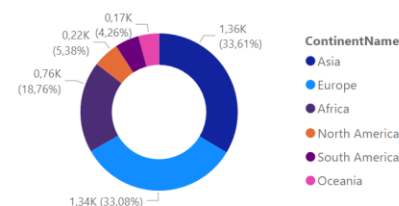
0,74
Average of HDI

Avg. GDP Per Capita and Number Of Deaths by ContinentName

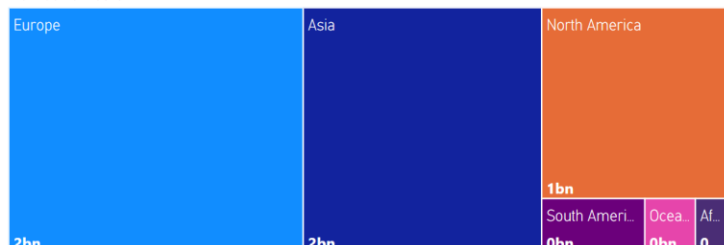


Country: All Year: 2019 2025

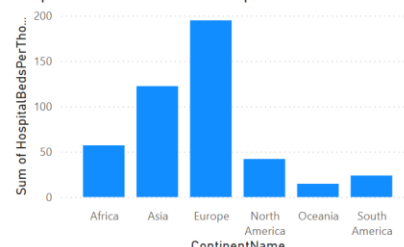
Male smokers by ContinentName



Number of tests

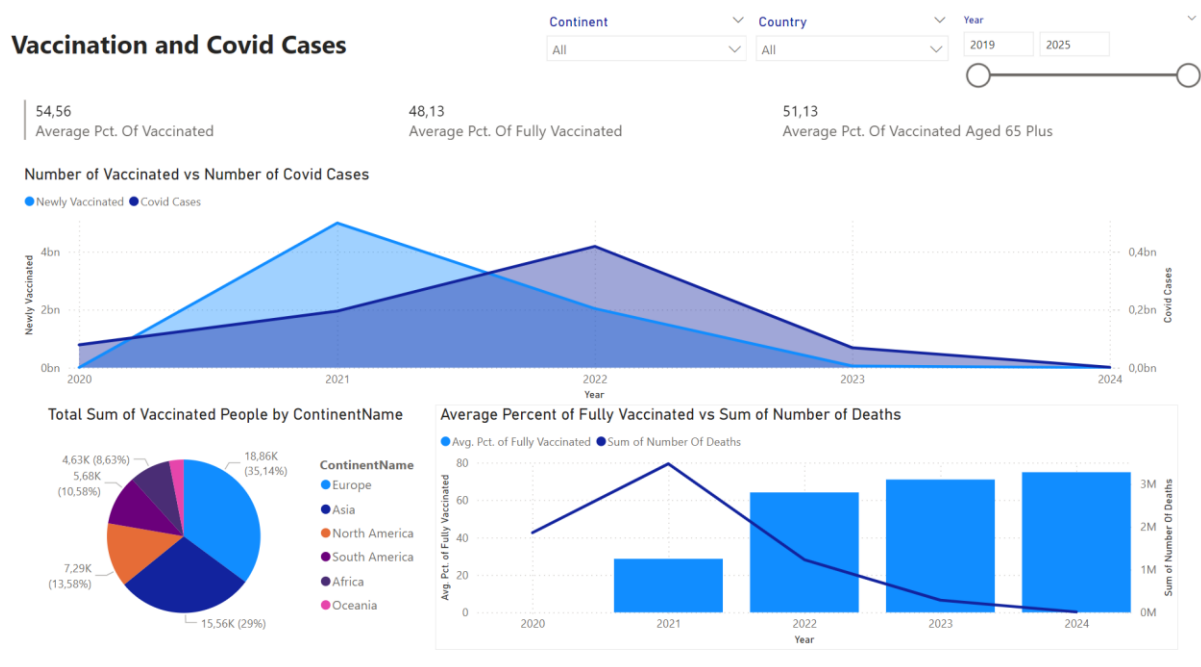


Hospital Beds Per Thousand People



Kolejna strona raportu dotyczy czynników ekonomicznych każdego kraju a także kontynentów. Ponownie mamy możliwość filtrowania interesujących nas obszarów oraz zakresu czasu. W skład wizualizacji wchodzi m.in. treemap reprezentujący łączną liczbę testów wykonanych na Covid, wykres kołowy informujący o odsetku osób palących w danych częściach świata, wykres słupkowy oraz liniowy, zestawiający średnią wartość współczynnika GDP oraz łączną liczbę zgonów z powodu zachorowań na Corona Virus.

Dzięki wizualizacjom mamy możliwość w łatwy sposób połączyć wiele czynników dotyczących dobrobytu w danym kraju lub kontynencie. Ułatwia to wyciągnięcie wniosków na temat jakości życia i służby zdrowia w tych rejonach. Jesteśmy w stanie wywnioskować czy w danym kraju istnieje zapotrzebowanie na dotacje w celu poprawienia warunków osób chorujących, czy istnieje konieczność dostarczenia dodatkowych testów na obecność Corona Virus oraz jaki jest odsetek śmiertelności w danym obszarze.



Przedmiotem zainteresowania ostatniej strony raportu jest wpływ szczepienia na liczbę zachorowań na Covid oraz liczbę osób zmarłych w wyniku tej choroby. Nietrudno jest zauważyć, że wraz z czasem i wzrostem średniej wartości odsetka ludzi w pełni zaszczepionych maleje liczba zgonów. Łatwo również wywnioskować, że wysoka łączna liczba osób zaszczepionych miała wpływ na niską liczbę wystąpień choroby, zaś malejąca liczba osób które poddawały się szczepieniom odzwierciedla się we wzroście przypadków zachorowań. Filtry umieszczone na stronie, ponownie przydadzą się w celu pozyskania informacji szczegółowych na temat konkretnego obszaru.

We wszystkich przedstawionych wizualizacjach zastosowano hierarchie związane z wymiarem czasu i lokalizacji. Kliknięcie na kolumnę reprezentującą dany rok przedstawiłaby informacje z poszczególnych miesięcy tego roku, zaś kliknięcie na kontynent pokazuje dane konkretnych krajów tego obszaru.

Podsumowując, przedstawione raporty pozwalają w skuteczny sposób zidentyfikować potencjalne problemy krajów objętych epidemią Covid-19. Poszczególne podstrony stanowią skuteczny wgląd w sytuację ekonomiczną na świecie w świetle zachorowań na Corona Virus a także ogólny wpływ szczepień i ich skuteczność w walce z chorobą.

Testy:

Sprawdźmy czy wykres w prawym dolnym rogu, przedstawiający liczbę przypadków śmiertelnych i średnią liczbę osób w pełni zaszczepionych w kolejnych latach jest stworzony zgodnie z danymi z hurtowni.


```

select Year(CAST(CAST(c.DateKey as varchar(8)) as date)) as comparsion_year,
SUM(c.NumberOfDeaths) as total_deaths,
ROUND(AVG(v.PercentOfFullyVaccinated), 2) as average_percent_of_fully_vaccinated from FactCovidCases c
join DimDate d on d.DateKey=c.DateKey
join FactCovidVaccinations v on v.DateKey=d.DateKey
group by Year(CAST(CAST(c.DateKey as varchar(8)) as date))
order by comparsion_year

```

	comparsion_year	total_deaths	average_percent_of_fully_vaccinated
1	2020	570693	0
2	2021	225657157	28,61
3	2022	88120986	64,08
4	2023	9667371	70,96
5	2024	49377	75,08

Statystyki obliczone na podstawie zapytania wydają się być poprawne. Po najechaniu na odpowiednie pola w interaktywnym raporcie można się przekonać, że liczby dokładnie się pokrywają.