

DOKUMENTACJA TECHNICZNA APLIKACJI

*„Transport miejski w Warszawie
analiza wybranych statystyk za lata 2018-2022”*

Marta Cieślak

Bartłomiej Olszewski

Rok akademicki 2022 / 2023



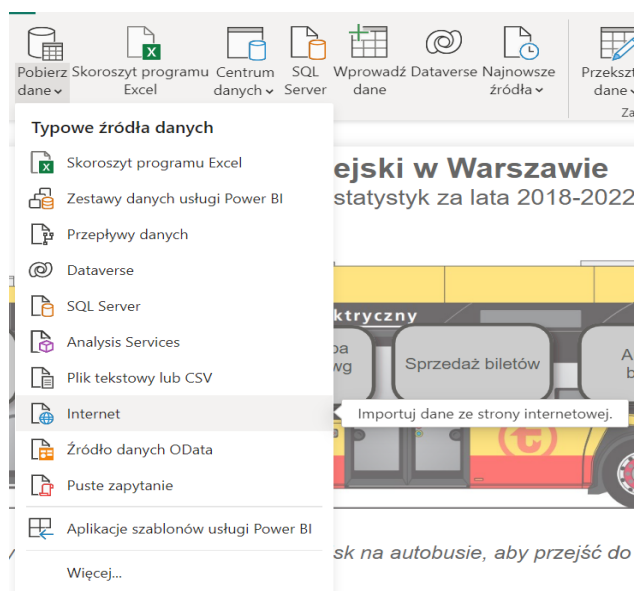
Spis treści

1. Źródła danych wykorzystane do przygotowania aplikacji	3
2. Spis tabel zawartych w aplikacji wraz z opisem najważniejszych przekształceń danych	5
3. Miary wykorzystane w aplikacji.....	8
4. Model danych w aplikacji	11

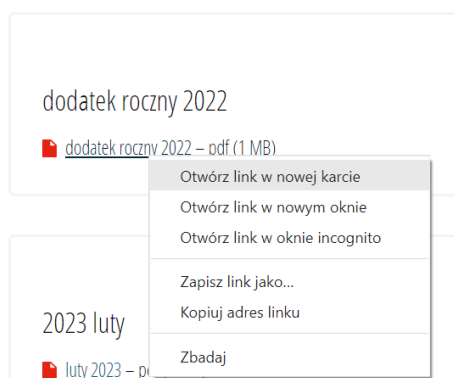
1. Źródła danych wykorzystane do przygotowania aplikacji

Aplikacja została przygotowana głównie w oparciu o dane statystyczne publikowane w informatorze statystycznym ZTM na stronie <https://www.ztm.waw.pl/statystyki/>. Dane w informatorze publikowane są w formacie pdf. W aplikacji wykorzystano dane z informatorów rocznych za lata 2018-2022.

1.1. Pobranie danych z wykorzystaniem konektora „Internet”



1.2. Przejście do strony internetowej ze źródłem danych



1.3. Zdefiniowanie źródła danych

13-BIULETYN-DODATEK-ROCN\ x +

← → ↺ <https://www.ztm.waw.pl/wp-content/uploads/2023/05/13-BIULETYN-DODATEK-ROCN-2022v2-korekta-przejazdow.pdf>

Z sieci Web

☒ Podstawowy ☐ Zaawansowane

Adres URL

1.4. Czyszczenie, transformacja i scalanie danych przy pomocy narzędzia Power Query (bardziej szczegółowy opis przekształceń w kolejnym rozdziale).

Nawigator

Opcje wyświetlania


- ☐ Table026 (Page 10)
- ☐ Table027 (Page 11)
- ☐ Table028 (Page 11)
- ☐ Table029 (Page 11)
- ☒ Table030 (Page 12)
- ☐ Table031 (Page 12)
- ☐ Table032 (Page 13)
- ☐ Table033 (Page 13)
- ☐ Table034 (Page 14)
- ☐ Table035 (Page 14)
- ☐ Table036 (Page 14)
- ☐ Table037 (Page 14)
- ☐ Page001
- ☐ Page002
- ☐ Page003
- ☐ Page004
- ☐ Page005
- ☐ Page006
- ☐ Page007

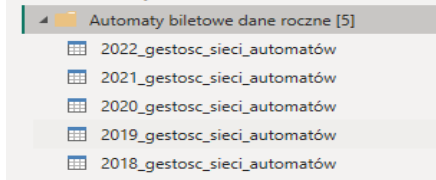
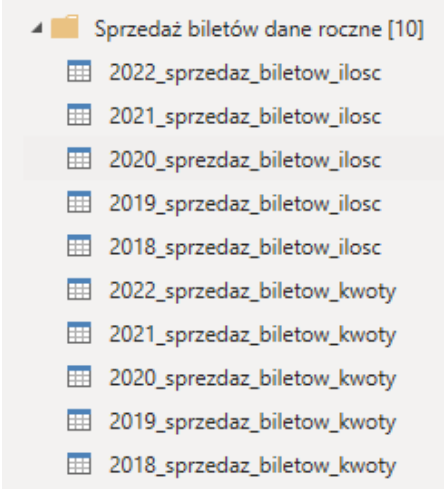
Table030 (Page 12)

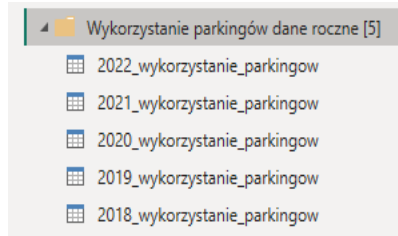
Column1	Column2	Średnie zapelnienie	Liczba zaparkowa
parking	liczba miejsc	w dni powszednie o godz. 10.00	łącznie w dni g
Metro Marymont	405	79,05%	80 359
Półczyńska	500	29,98%	37 631
Metro Wilanowska	280	85,11%	59 815
Metro Młociny	1010	59,87%	151 781
Anin SKM	83	13,33%	2 777
Metro Ursynów	166	45,46%	18 939
Metro Młociny II	72	97,56%	17 632
Al.Krakowska	415	56,49%	58 842
Wawer SKM	149	40,00%	14 961
Ursus Niedźwiadek	345	64,99%	56 280
Wawrzyszew	80	94,13%	18 900
Warszawa Stadion*	110	91,75%	25 332
Metro Stokłosy	393	34,47%	34 006
Metro Młociny III	157	74,21%	29 245
Metro Młociny IV	254	91,43%	58 289
P+R Żerań PKP**	235	6,64%	3 918
P+R Jeziorki PKP***	138	26,25%	9 091
Łącznie	4792		null 677 799
* parking funkcjonował do dnia 2022-11-30			
		null	null
	** P+R Żerań uruchomiony w lutym 2022, P+R Jeziorki uruchomiony w list		
		null	null

2. Spis tabel zawartych w aplikacji wraz z opisem najważniejszych przekształceń danych

Poniżej przedstawiono spis tabel wykorzystanych w widoku raportu w aplikacji wraz z opisem ich zawartości i najważniejszymi przekształceniami danych.

Tabela	Opis tabeli	Opis najważniejszych przekształceń
Daty	<p>Tabela kalendarzowa, utworzona w Power BI na potrzeby analiz czasowych wykorzystywanych w aplikacji.</p> <p><code>=CALENDAR(DATE(YEAR(MIN('Pasażerowie metro'[rok_miesiac_data])),1,1), date(year(max('Pasażerowie metro'[rok_miesiac_data])),12,31))</code></p> <p>Początek kalendarza to 1 stycznia pierwszego roku analizy (tu 01.01.2018).</p> <p>Początek kalendarza to 31 grudnia ostatniego roku z analizy (tu 31.12.2022).</p>	<ul style="list-style-type: none"> dodanie kolumn: Rok, Miesiąc, Dzień na potrzeby analizy czasowej w aplikacji
Matro data otwarcia stacji	<p>Tabela z datami otwarcia poszczególnych stacji metra.</p> <p>Źródło danych: tabela z listą stacji metra i datami ich otwarcia pobrana z Wikipedii (= Web.BrowserContents(https://pl.wikipedia.org/wiki/Lista_stacji_metra_w_Warszawie"))</p>	<ul style="list-style-type: none"> wyodrębnienie odpowiedniej tabeli ze źródła danych ujednolicenie kodów stacji metra, tak żeby odpowiadały kodom z innych źródeł danych (np. tym z informatora statystycznego ZTM) ujednolicenie nazw kolumn na potrzeby stworzenia modelu danych w aplikacji oznaczenie kolumny z datami otwarcia stacji typem danych „date”
Pasażerowie metro	<p>Tabela zawiera dane na temat liczby pasażerów metra według stacji w poszczególnych miesiącach w latach w 2018-2022 (identyfikatorem stacji w modelu jest kod stacji i nazwa stacji).</p> <p>Źródło danych: informator statystyczny ZTM.</p> <p>Zgodnie z informacjami w informatorze liczba pasażerów dla danej stacji odpowiada liczbie pasażerów wchodzących na daną stację metra.</p>	<ul style="list-style-type: none"> pobranie tabel z danymi z informatorów statystycznych ZTM dla poszczególnych lat  <ul style="list-style-type: none"> dodanie kolumny z identyfikatorem roku w poszczególnych tabelach scalenie tabel z poszczególnych lat w jedną zbiorczą tabelę „Pasażerowie

		<p>metro”, która została wykorzystana w widoku raportu</p> <ul style="list-style-type: none"> wydzielenie z nazwy stacji kodu stacji dodanie kolumny warunkowej na podstawie kodu stacji – wyróżnienie linii M1 i M2 przestawienie kolumn tak, żeby identyfikator miesiąca był w wierszu a nie w kolumnie ujednoczenie kodów stacji metra, tak żeby były spójne z kodami z innych źródeł danych
Sieć automatów	<p>Tabela z danymi dotyczącymi liczby automatów biletowych w podziale na dzielnice Warszawy oraz gęstością sieci automatów dla poszczególnych dzielnic (wyrażoną jako liczba automatów biletowych przypadająca na kilometr kwadratowy).</p> <p>Źródło danych: informator statystyczny ZTM.</p>	<ul style="list-style-type: none"> pobranie tabel z danymi z informatorów statystycznych ZTM dla poszczególnych lat  <ul style="list-style-type: none"> dodanie kolumny z identyfikatorem roku w poszczególnych tabelach scalenie tabel z poszczególnych lat w jedną zbiorczą tabelę „Sieć automatów”, która została wykorzystana w widoku raportu
Sprzedaż biletów	<p>Tabela z danymi liczbowymi dotyczącymi liczby sprzedanych biletów i wartością sprzedanych biletów.</p> <p>Źródło danych: informator statystyczny ZTM.</p>	<ul style="list-style-type: none"> pobranie tabel z danymi z informatorów statystycznych ZTM dla poszczególnych lat (oddzielne tabele z liczbą sprzedanych biletów wyrażoną w sztukach i oddzielne z wartością sprzedaży wyrażoną w złotych)  <ul style="list-style-type: none"> dodanie kolumny z identyfikatorem roku w tabelach z poszczególnych lat przestawienie kolumn tak, żeby informacje na temat strefy

		<p>obowiązywania biletu i typu biletu były w wierszach a nie w kolumnach</p> <ul style="list-style-type: none"> wyróżnienie z kolumny typu biletu i strefy, jako oddzielnych kolumn scalenie tabel z poszczególnych lat w dwie zbiorcze tabele „Sprzedaz_biletow_ilosc” i „Sprzedaz_biletow_wartosc” dodanie kolumny „ujęcie_sprzedaży” w obu zbiorczych tabelach – kolumna posłuży do rozróżnienia czy dane liczbowe dotyczą liczby sprzedanych biletów czy wartości sprzedanych biletów scalenie tabeli „Sprzedaz_biletow_ilosc” i „Sprzedaz_biletow_wartosc” w jedną tabelę „Sprzedaż biletów” doprecyzowanie nazw rodzajów biletów (bilet jednorazowy ważny w I strefie to bilet 75-minutowy, bilet jednorazowy ważny w 1 i 2 strefie to bilet 90-minutowy)
Stacje metra dane geograficzne	<p>Tabela z danymi geograficznymi dla poszczególnych stacji metra (długość i szerokość geograficzna).</p> <p>Tabela została stworzona w Excelu i zaimportowana do modelu danych w aplikacji.</p> <p>Źródło danych geograficznych: strony dedykowane poszczególnym stacjom metra w Wikipedii (Geo Hack), np.: GeoHack - Kabaty (stacja metra)</p>	<ul style="list-style-type: none"> nadanie kolumnie Latitude kategorii danych „Szerokość geograficzna” nadanie kolumnie Longitude kategorii danych „Długość geograficzna” ujednoczenie nazw kolumn na potrzeby stworzenia modelu danych w aplikacji
Wykorzystanie parkingów	<p>Tabela z informacjami o liczbie miejsc na poszczególnych parkingach „Parkuj i jedź” i ich wypełnieniu oraz o liczbie zaparkowanych samochodów na poszczególnych parkingach.</p> <p>Źródło danych: informator statystyczny ZTM.</p>	<ul style="list-style-type: none"> pobranie tabel z danymi z informatorów statystycznych ZTM dla poszczególnych lat  <ul style="list-style-type: none"> dodanie kolumny z identyfikatorem roku w poszczególnych tabelach scalenie tabel z poszczególnych lat w jedną zbiorczą tabelę „Wykorzystanie parkingów”, która została wykorzystana w widoku raportu

3. Miary wykorzystane w aplikacji

W poniższej tabeli, opisano miary, które zostały przygotowane w języku DAX na potrzeby wizualizacji wykorzystanych w aplikacji.

Miara	Kod	Opis działania
Liczba analizowanych miesięcy	= <code>DISTINCTCOUNT('Pasażerowie metro'[rok_miesiac_data])</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – raport szczegółowy dla stacji”.</p> <p>Miara wskazuje na liczbę miesięcy, które upłynęły od momentu otwarcia stacji do ostatniego miesiąca, który jest analizowany.</p> <p>Miara później jest wykorzystywana do wyliczenia średniej liczby pasażerów w miesiącu (niektóre stacje były otwierane w trakcie roku, przy wyliczaniu średniej liczby pasażerów na miesiąc w danym roku, nie zawsze brane jest 12 miesięcy tylko tyle ile działała dana stacja).</p>
Liczba pasażerów wybrany rok	= <code>CALCULATE(sum('Pasażerowie metro'[Liczba pasażerów]), Daty[Rok]=SELECTEDVALUE(Daty[Rok]))</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>W raporcie po wybraniu konkretnego roku pokazywana jest m. in. liczba pasażerów dla wybranego roku. Miara sumuje liczbę pasażerów z tabeli „Pasażerowie metro” dla wybranego przez użytkownika roku.</p>
Liczba stacji wybrany rok	= <code>CALCULATE(DISTINCTCOUNT('Pasażerowie metro'[Kod stacji]), Daty[Rok]=SELECTEDVALUE(Daty[Rok]))</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>W raporcie po wybraniu konkretnego roku pokazywana jest m. in. liczb stacji metra w wybranym roku. W mierze zliczane są unikalne kody stacji występujące w tabeli „Pasażerowie metro” dla wybranego przez użytkownika roku.</p>
Liczba stacji rok poprzedni	Liczba stacji rok poprzedni = <code>IF(SELECTEDVALUE(Daty[Rok])=2018,</code> <code>CALCULATE(DISTINCTCOUNT('Pasażerowie metro'[Kod stacji]), Daty[Rok]=SELECTEDVALUE(Daty[Rok])) ,</code> <code>CALCULATE(DISTINCTCOUNT('Pasażerowie metro'[Kod stacji]), Daty[Rok]=SELECTEDVALUE(Daty[Rok])-1))</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>W raporcie po wybraniu konkretnego roku pokazywany jest m. in. miernik, który pokazuje jak zmieniła się liczba stacji metra w wybranym przez użytkownika roku w stosunku do roku poprzedniego.</p>

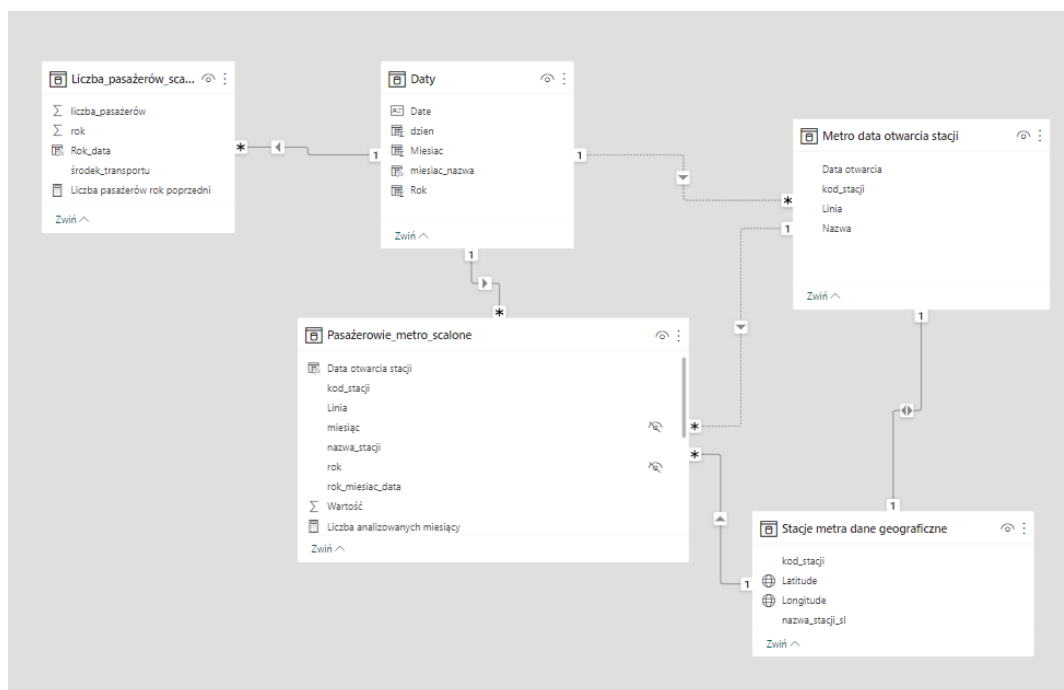
		<p>W mierze zliczane są unikalne kody stacji występujące w tabeli „Pasażerowie metro” dla jednego roku wcześniej niż wybrany przez użytkownika rok.</p> <p>Problematyczny był rok 2018, dla którego brak danych o liczbie stacji metra w poprzedzającym go roku (użyto funkcji IF, żeby miara obsługiwała ten wyjątek). Dla uproszczenia przyjęto, że liczba stacji w roku poprzedzającym 2018 rok była taka sama jak w 2018 roku.</p>
Link do informacji o stacji Wikipedia	<code>= "https://pl.wikipedia.org/wiki/" & IFERROR(REPLACE(SELECTEDVALUE('Pasażerowie metro'[Nazwa stacji]), FIND(" ", SELECTEDVALUE('Pasażerowie metro'[Nazwa stacji])), 1, "_"), SELECTEDVALUE('Pasażerowie metro'[Nazwa stacji]) & "_ (stacja_metra) "</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – raport szczegółowy dla stacji”.</p> <p>Miara generuje tekst, który później jest wykorzystywany jako link do artykułu na Wikipedii dotyczącego wybranej stacji metra.</p> <p>Żeby link poprawnie przekierowywał do Wikipedii konieczne było znalezienie w nazwie stacji wszystkich spacji i zamienienie ich na znak podkreślenia.</p> <p>Dla jednowyrazowych nazw stacji funkcja FIND zwracała błąd (nie znajdowała żadnej spacji w nazwie stacji), dlatego żeby treść linku generowała się również w takich przypadkach poprawnie, użyto dodatkowo funkcji IFERROR. Dla jednowyrazowych nazw stacji metra do treści linku wklejana jest nazwa stacji bez żadnych modyfikacji.</p>
Metro raport szczegółowy linia	<code>= IF(SELECTEDVALUE('Pasażerowie metro'[Nazwa stacji])="Świętokrzyska", "M1 M2", SELECTEDVALUE('Pasażerowie metro'[Linia]))</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – raport szczegółowy dla stacji”.</p> <p>Raport pokazuje m. in. do której linii należy wybrana stacja metra. Stacja Świętokrzyska jest jedyną stacją, która występuje w obu liniach metra. Żeby również dla tej stacji prawidłowo pokazywała się linia, konieczne było stworzenie miary i użycie funkcji IF, żeby obsłużyć ten wyjątek.</p>
Średnia liczba pasażerów w miesiącu	<code>= DIVIDE(sum('Pasażerowie metro'[Liczba pasażerów]), [Liczba analizowanych miesięcy])</code>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – raport szczegółowy dla stacji”.</p> <p>Miara sumuje liczbę pasażerów dla danej stacji w analizowanym okresie i dzieli ją przez liczbę miesięcy, przez które wybrana stacja działała w analizowanym okresie. Niektóre stacje były otwierane w trakcie analizowanego okresu, dlatego nie zawsze liczba pasażerów jest</p>

		<p>dzielona przez 60 (dla większości stacji liczba analizowanych miesięcy to 60 - 5 lat x 12 miesięcy). Maksymalny analizowany okres to 01.2018-12.2022.</p>
Średnia liczba pasażerów rok poprzedni	<p>= CALCULATE(sum('Pasażerowie metro'[Liczba pasażerów])/12, Daty[Rok]=SELECTEDVALUE(Daty[Rok])-1)</p>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>W raporcie po wybraniu konkretnego roku pokazywana jest m. in. wizualizacja <i>Kluczowy wskaźnik wydajności</i>, który pokazuje jak zmieniła się średnia miesięczna liczba pasażerów metra w wybranym przez użytkownika roku w stosunku do roku poprzedniego.</p> <p>Miara sumuje liczbę pasażerów w roku poprzedzającym wybrany przez użytkownika rok i dzieli ją przez 12 (liczba miesięcy w roku), żeby wyliczyć średnią miesięczną liczbę pasażerów.</p>
Średnia liczba pasażerów wybrany rok	<p>CALCULATE(sum('Pasażerowie metro'[Liczba pasażerów])/12, Daty[Rok]=SELECTEDVALUE(Daty[Rok]))</p>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>W raporcie po wybraniu konkretnego roku pokazywana jest m. in. wizualizacja <i>Kluczowy wskaźnik wydajności</i>, który pokazuje jak zmieniła się średnia miesięczna liczba pasażerów metra w wybranym przez użytkownika roku w stosunku do roku poprzedniego.</p> <p>Miara sumuje liczbę pasażerów w wybranym przez użytkownika roku i dzieli ją przez 12 (liczba miesięcy w roku), żeby wyliczyć średnią miesięczną liczbę pasażerów.</p>
Tytuł liczba pasażerów wybrany rok	<p>= "Liczba pasażerów w "& SELECTEDVALUE(Daty[Rok])& " roku"</p>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>Miara służy do dynamicznego wyznaczenia tytułu dla wizualizacji prezentującej liczbę pasażerów metra w wybranym przez użytkownika roku.</p>
Tytuł liczba stacji YoY	<p>= "Liczba stacji "& SELECTEDVALUE(Daty[Rok])& " vs." "& SELECTEDVALUE(Daty[Rok])-1 & " rok"</p>	<p>tabela: Pasażerowie metro</p> <p>Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”.</p> <p>Miara służy do dynamicznego wyznaczenia tytułu dla wizualizacji prezentującej liczbę stacji</p>

		metra w wybranym przez użytkownika roku w stosunku do roku poprzedniego.
Tytuł średnia miesięczna liczba pasażerów YoY	= "Średnia miesięczna liczba pasażerów "& SELECTEDVALUE(Daty[Rok])& " vs &SELECTEDVALUE(Daty[Rok])-1&" rok"	tabela: Pasażerowie metro Miara jest wykorzystywana w raporcie „Metro – liczba pasażerów rok do roku”. Miara służy do dynamicznego wyznaczenia tytułu dla wizualizacji prezentującej średnią miesięczną liczbę pasażerów dla wybranego przez użytkownika roku w stosunku do roku poprzedniego.

4. Model danych w aplikacji

Na potrzeby modelu danych w aplikacji utworzono relacje między poszczególnymi tabelami ze statystykami dotyczącymi metra. W modelu identyfikatorem konkretnej stacji metra jest kod_stacji.



Pozostałe tabele wykorzystane w aplikacji nie wymagały utworzenia relacji w modelu danych.