Analiza rozkładów jazdy autobusów

Projekt w ramach przedmiotu: Hurtownie Danych i Systemy Business Intelligence

Piotr Bielecki, Laura Hoang

Maj 2023

1 Wprowadzenie

Niniejsza dokumentacja przedstawia szczegóły projektu dotyczącego analizy rozkładów jazdy autobusów. Projekt ma na celu zgłębienie i zrozumienie danych dotyczących planowanych przyjazdów i odjazdów autobusów, a także przystanków obsługiwanych przez system transportu miejskiego. Analiza tych danych pozwoli na identyfikację obszarów wymagających ulepszeń oraz optymalizację tras i usprawnienie funkcjonowania transportu publicznego.

Główne cele projektu to:

1. Zrozumienie aktualnego układu rozkładów jazdy autobusów i przystanków w systemie transportu miejskiego.
2. Analiza danych w celu odkrycia wzorców i trendów dotyczących czasu podróży, częstotliwości kursów oraz obszarów o największym obciążeniu.
3. Zaproponowanie optymalnych zmian w rozkładach jazdy autobusów, takich jak dostosowanie częstotliwości kursów oraz zmiana tras.
4. Przygotowanie raportów i wizualizacji prezentujących wyniki analizy, które będą użyteczne dla interesariuszy i decydentów w dziedzinie transportu publicznego.

Realizacja tego projektu może przynieść szereg korzyści dla operatora systemu transportu miejskiego oraz dla pasażerów korzystających z usług transportowych. Główne z nich to:

1. **Poprawa jakości usług transportowych** - optymalizacja rozkładów jazdy autobusów może przyczynić się do zwiększenia punktualności i regularności kursów.
2. **Skrócenie czasu podróży** - zminimalizowanie opóźnień dzięki lepszemu zrozumieniu i wykorzystaniu danych dotyczących czasu przejazdu i obciążenia tras.
3. **Zwiększenie wydajności i efektywności systemu** - identyfikacja obszarów wymagających optymalizacji i wprowadzenie odpowiednich zmian w rozkładach jazdy.

4. **Udoskonalenie planowania i zwiększenie komfortu podróży** - dostarczenie pasażerom dokładnych informacji o rozkładach jazdy i czasie oczekiwania na przystankach.

5. **Zredukowanie negatywnego wpływu na środowisko** - optymalizacja tras i częstotliwości kursów przyczyni się do zmniejszenia zatłoczenia ulic i emisji spalin.

6. **Zredukowanie negatywnego wpływu na środowisko** - optymalizacja tras i częstotliwości kursów przyczyni się do zmniejszenia zatłoczenia ulic i emisji spalin.

7. **Zwiększenie efektywności operacyjnej** - zoptymalizowanie planowania zasobów, takich jak ilość autobusów i kierowców potrzebnych do obsługi danej trasy.

Zakres projektu obejmuje następujące etapy:

1. **Zebranie danych**: Pobranie danych dotyczących rozkładów jazdy autobusów, współrzędnych przystanków oraz tras pojazdów z API Warszawskiego Transportu Publicznego.

2. **Przetwarzanie i czyszczenie danych**: Przetworzenie i wstępna analiza pobranych danych, odpowiednia obsługa błędów takich jak brakujących wartości i duplikaty, ewentualna modyfikacja danych.

3. **Analiza danych**: Wykorzystanie technik analizy danych, takich jak eksploracja danych, statystyka, aby odkryć wzorce, zależności i trendy dotyczące rozkładów jazdy autobusów.

4. **Optymalizacja rozkładów jazdy**: Na podstawie wyników analizy zaproponowanie optymalnych zmian w rozkładach jazdy autobusów, takich jak dostosowanie częstotliwości kursów, zmiana tras itp.

5. **Przygotowanie raportów i wizualizacji**: Przygotowanie czytelnych raportów, wykresów i wizualizacji przedstawiających wyniki analizy, które będą pomocne dla decydentów i zainteresowanych interesariuszy.

6. **Prezentacja i wdrożenie**: Prezentacja wyników analizy i zaproponowanych zmian dla zainteresowanych interesariuszy oraz przygotowanie materiałów do wdrożenia optymalizacji rozkładów jazdy.

Dokumentacja projektu zawiera szczegółowe informacje dotyczące każdego etapu, opis użytych narzędzi, metodyk i uzyskanych wyników.

Kody źródłowe do niniejszego projektu znajdują się na repozytorium.

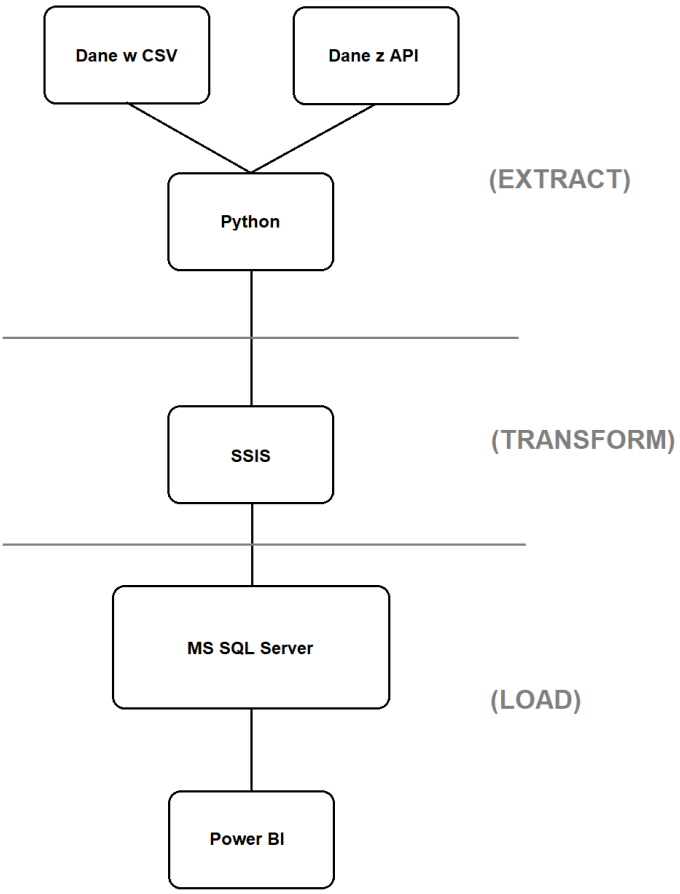
2 Opis zbiorów danych

Dane zostały pozyskane za pomocą API Warszawskiego, które są udostępnione przez Miasto Stołeczne Warszawa, na licencji Creative Commons Attribution. Wszystkie dane są przechowywane w formacie danych JSON. Informacje aktualizowane są jeden raz na dobę.

Analiza tych danych zostanie wykonana na zbiorach dotyczących transportu miejskiego, na które składają się zbiory danych dotyczące:

- rozkładów jazdy,
- współrzędnych przystanków,
- tras pojazdów.

3 Architektura: opis procesu ETL



Rysunek 1: Diagram architektury

W niniejszej sekcji opisany zostanie cały proces ETL. Zmiany które potencjalnie można wdrożyć dopisane są na diagramach na kolor czerwony.

3.1 Etap "Extract":

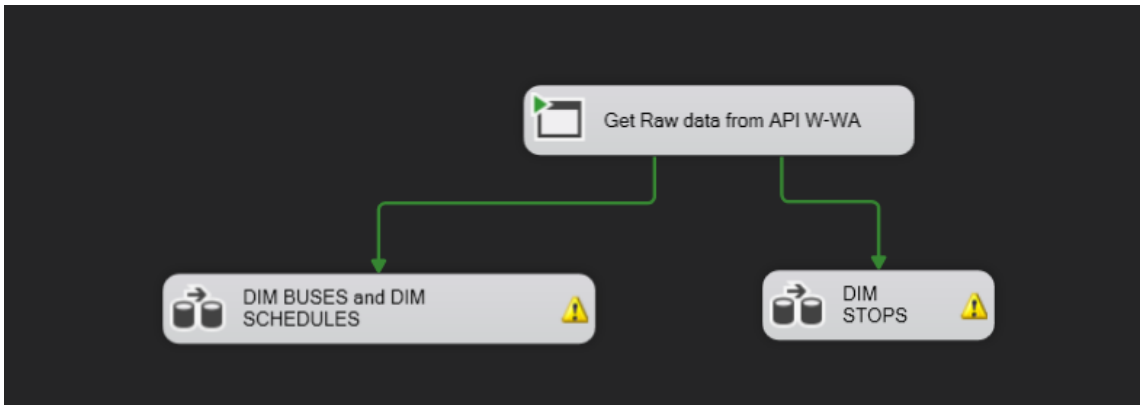
Na tym etapie dane zostają pozyskane z API Warszawskiego za pomocą kodu w Pythonie. Opierając się na nazwach przystanków z ramki danych z pliku CSV, zostają wyciągnięte po kolei: dane o nazwach przystanków i ich kodach, dane o liniach autobusowych na danych przystankach, a na koniec dane o rozkładach jazdy autobusów na danych przystankach. Na tym etapie kod w Pythonie jest w stanie sprawnie wyekstraktować dane z plików JSON. Skrypty *.py powiązane zostały z resztą procesu ETL w narzędziu SSIS.

3.2 Etap "Transform & Load":

Na tym etapie dane zostają zmodyfikowane jak na zrzutach ekranu z narzędzia SSIS poniżej, aby odpowiadały zaprojektowanemu modelowi. Warto dodać, że po prezentacji rozwiązania pojawiły się słuszne uwagi, których przykładową implementację opisujemy na wspomnianych zrzutach ekranu (czerwone fragmenty). Na koniec procesu ETL dane zostają załadowane do odpowiednich tabel w hurtowni danych.

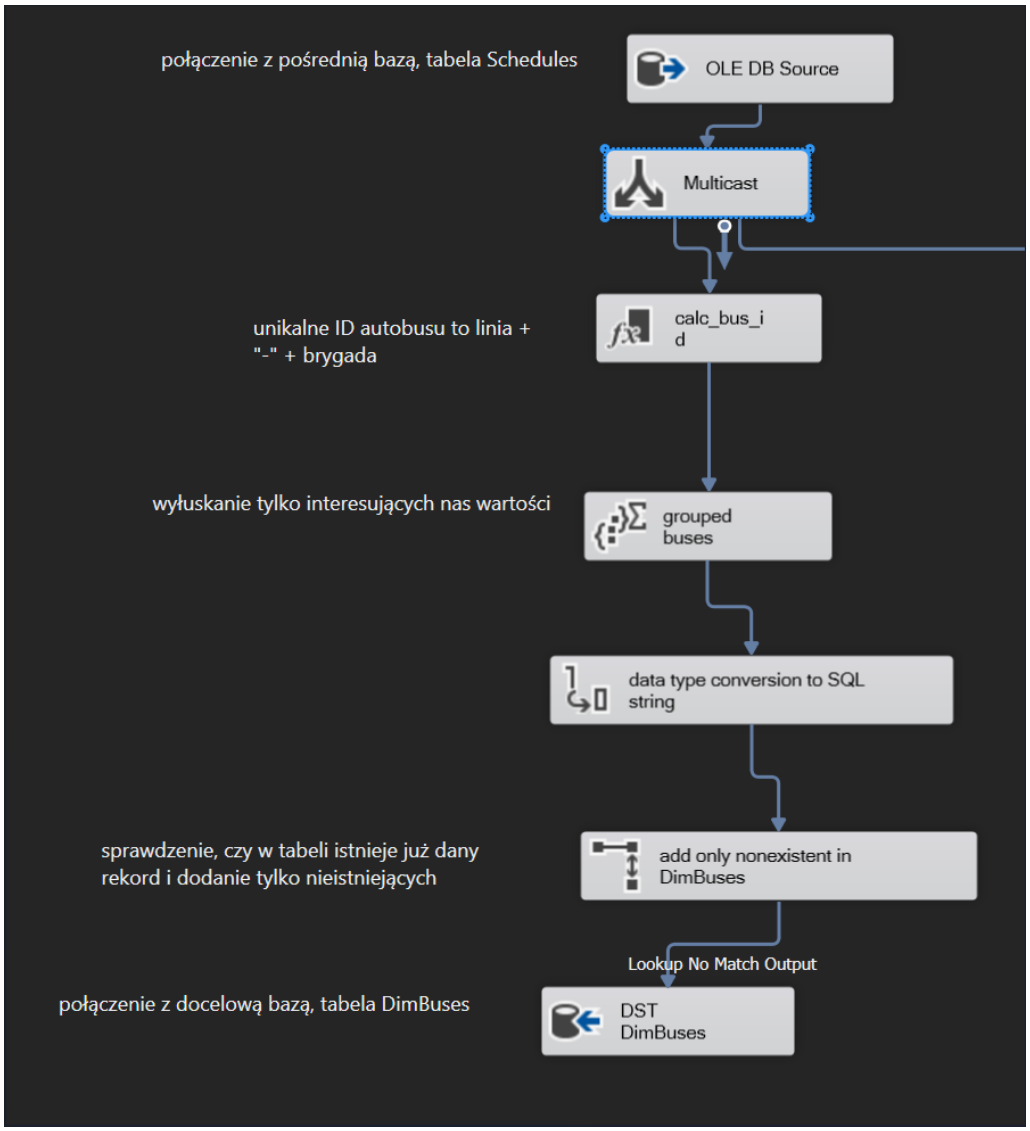
3.2.1 Wymiary

Przepływ kontroli procesu ETL dla wymiarów wygląda następująco:



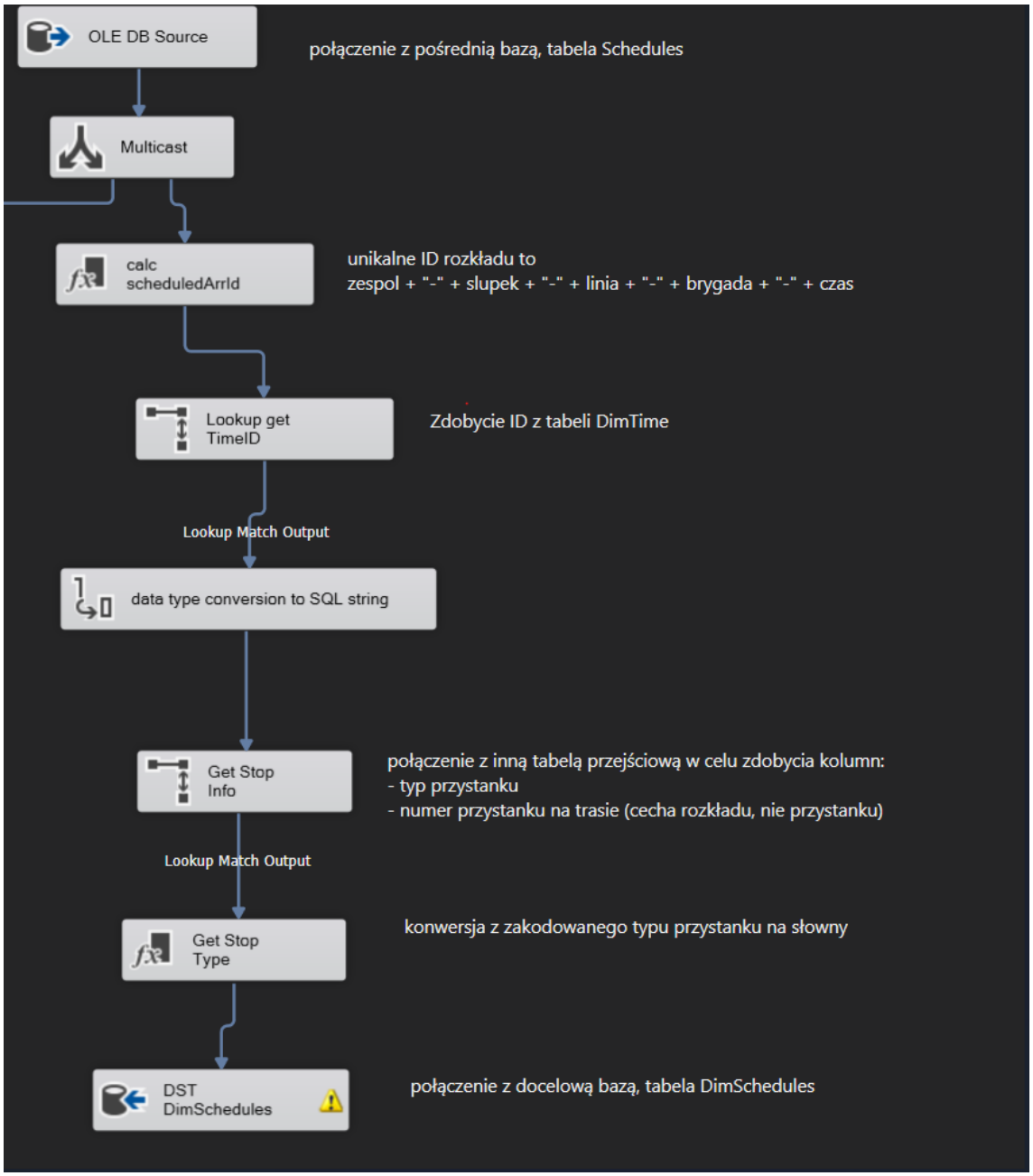
Rysunek 2: Control Flow dla całości procesu ETL dla wymiarów

Wymiar autobusów Wymiar autobusów jest budowany z tych samych informacji co wymiar rozkładów, stąd na początku element multicast.



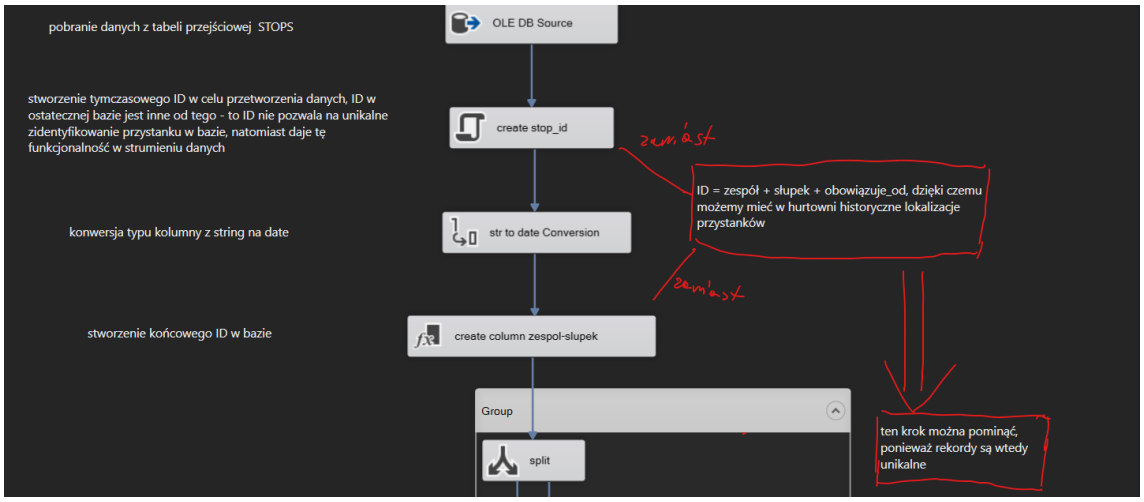
Rysunek 3: Data Flow dla wymiaru autobusów

Wymiar rozkładów Pojedynczym wpisem w tabeli wymiaru rozkładów nie jest cały rozkład jazdy, a pojedynczy wpis w taki rozkład jazdy (konkretna godzina, konkretna linia autobusu, konkretny przystanek)

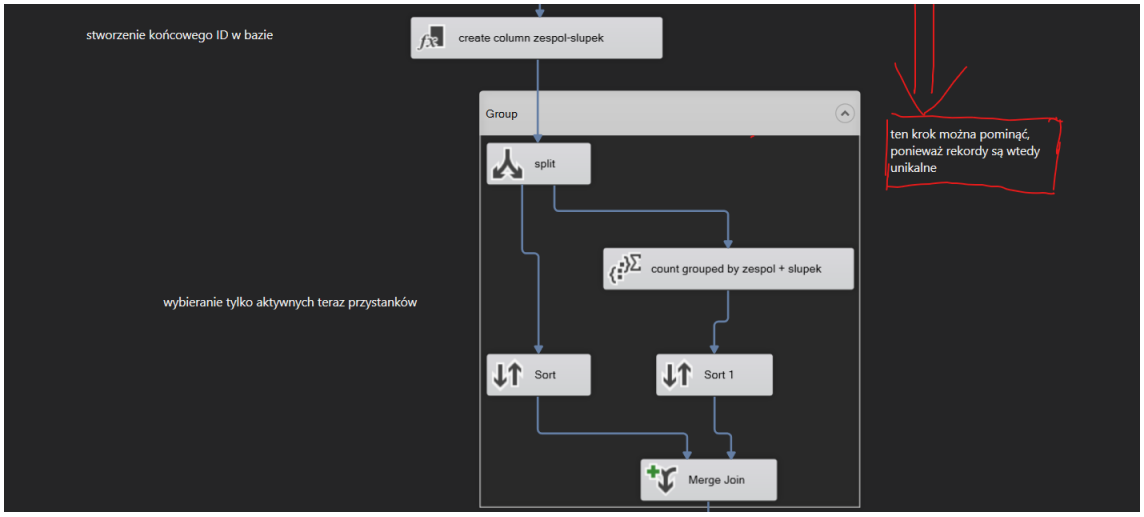


Rysunek 4: Data Flow dla wymiaru rozkładów

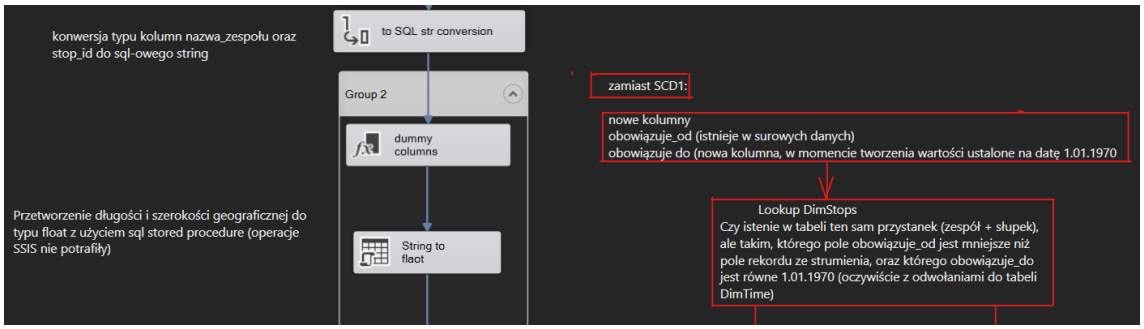
Wymiar przystanków Pozycje przystanków ulegają zmianie, postanowiliśmy potraktować ten wymiar jako SCD.



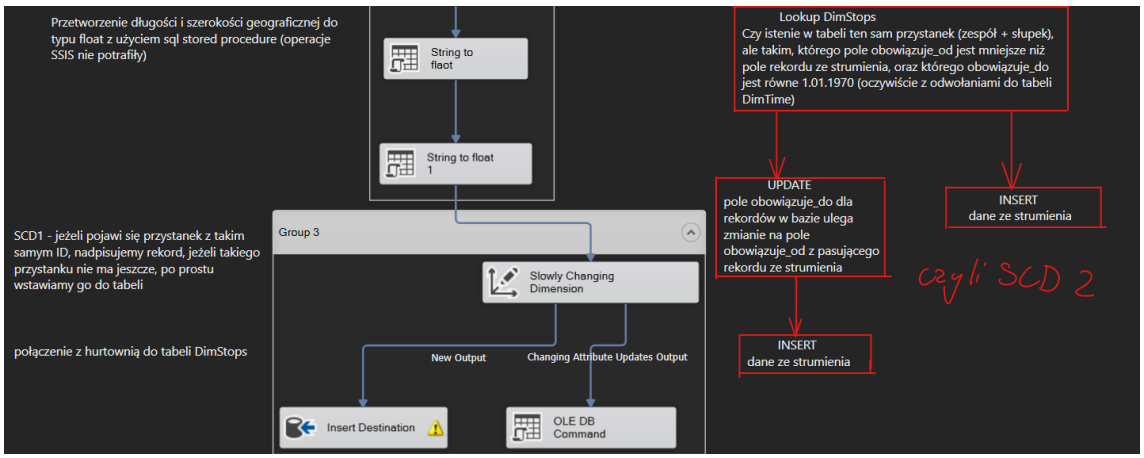
Rysunek 5: Data Flow dla wymiaru rozkładów cz.1



Rysunek 6: Data Flow dla wymiaru rozkładów cz.2



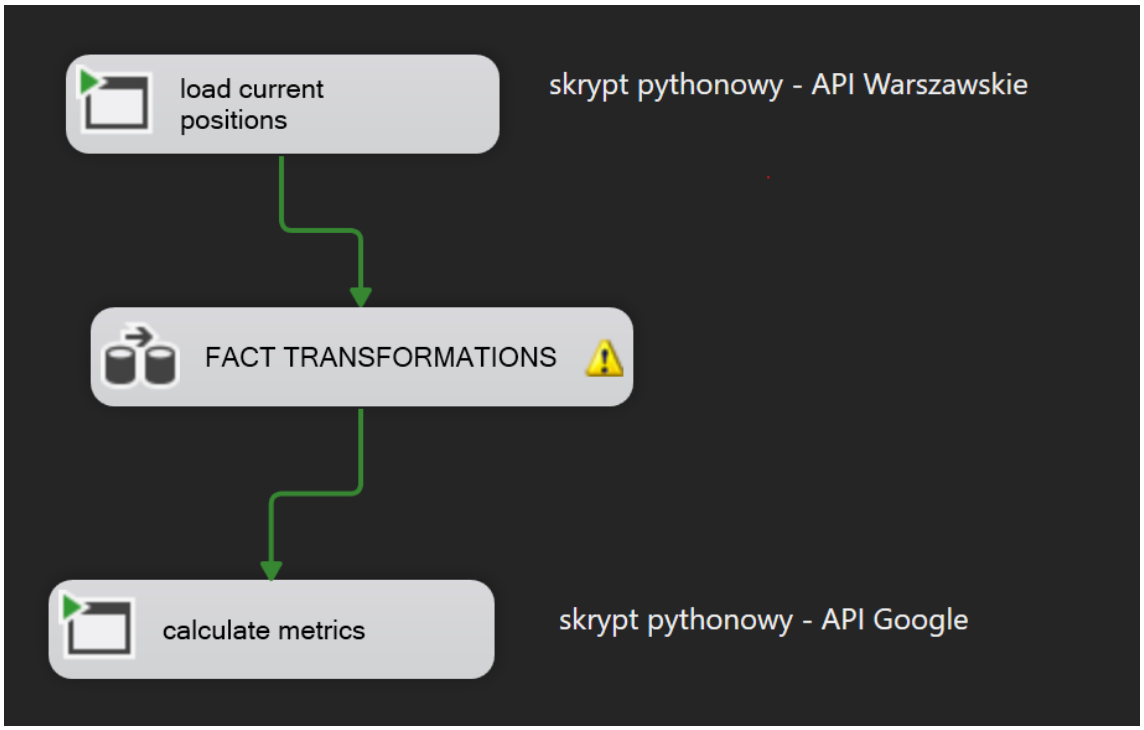
Rysunek 7: Data Flow dla wymiaru rozkładów cz.3



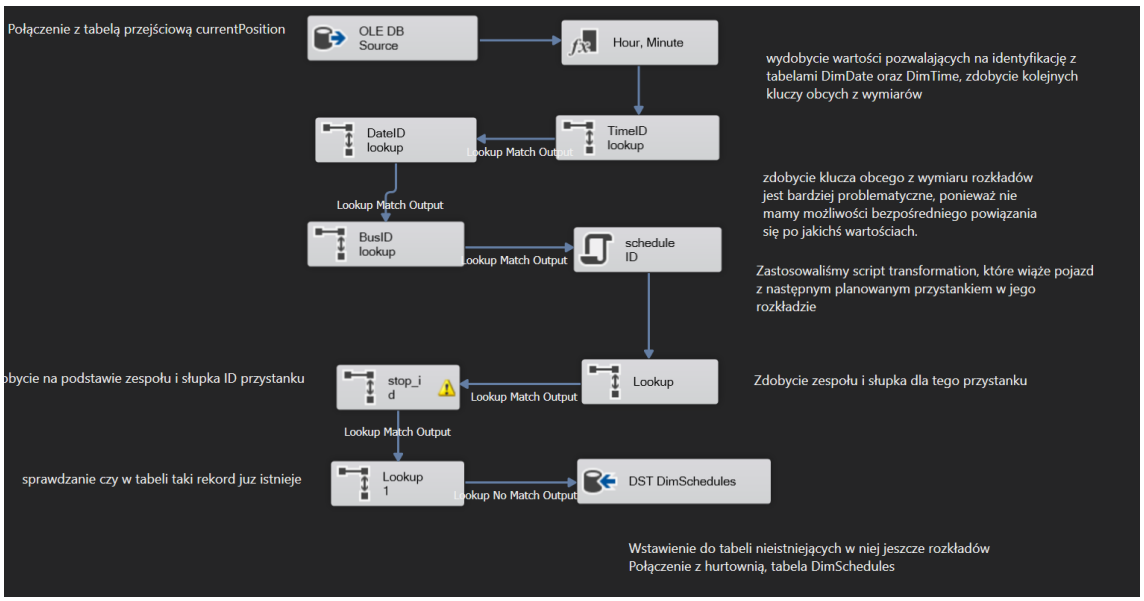
Rysunek 8: Data Flow dla wymiaru rozkładów cz.4

3.2.2 Fakt

Tabela faktowa jest tabelą transakcyjną. Istniał pomysł na stosowanie tabeli accumulated snapshot, natomiast przejazd autobusu po całej jego trasie nie jest na tyle precyzyjnym procesem do śledzenia (chociażby różna liczba przystanków na trasach), a także nie byłby tak przystępny w tworzeniu raportów, które mieliśmy w planie.



Rysunek 9: Control Flow dla procesu ETL dla tabeli faktowej



Rysunek 10: Data Flow dla faktu

3.3 Etap "Load":

Na koniec przetworzone dane zostają załadowane do hurtowni danych.

4 Model fizyczny hurtowni danych

Tabele z API

- StopCode - tabela zawiera słownik zespołów przystanków do ich numeru ID
  - **zespol_id** - numer identyfikacyjny zespołu przystanków
  - **nazwa** - nazwa zespołu przystanków
- StopInfo - tabela zawierająca linie autobusowe, w których rozkładzie znajduje się dany przystanek
  - **zespol_id**
  - **slupek_id** - numer przystanku w zespole
  - **linia** - numer linii autobusowej
- Routes - tabela tras pojazdów
  - **slupek_id**
  - **kod** - kod identyfikacyjny rozkładu jazdy (jedna linia autobusowa może mieć w bazie wiele rozkładów jazdy)
  - **enroute_nr** - numer przystanku na trasie (począwszy od 1)
  - **odleglosc** - odległość przystanku od początku trasy
  - **ulica_id** - numer identyfikacyjny ulicy, przy której znajduje się przystanek
  - **zespol_id**
  - **typ** - typ przystanku autobusowego na linii:

    - "0": "przelotowy"
    - "1": "stały"
    - "2": "na żądanie"
    - "3": "krańcowy"
    - "4": "dla wysiadających"
    - "5": "dla wsiadających"
    - "6": "zajezdnia"
    - "7": "techniczny"
    - "8": "postojowy"
  - **slupek_id**

- CurrentPosition
  - **linia**
  - **VehicleNumber** - numer identyfikacyjny pojazdu
  - **Brigade** - numer brygady, do której należy dany pojazd. Brygada to przyporządkowanie pojazdu do kursów w rozkładzie. Brygadą określa się kolejny autobus na danej linii.
  - **Lon** - szerokość geograficzna pojazdu
  - **Lat** - Długość geograficzna pojazdu
  - **Time** - Czas ostatniej aktualizacji pozycji

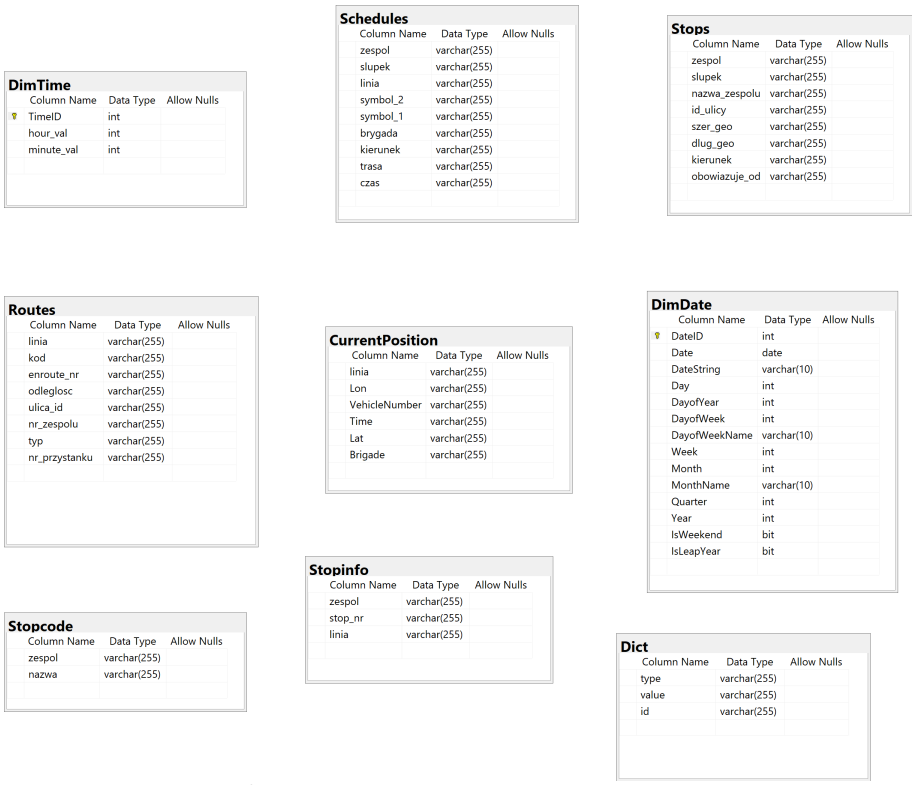
- Schedule - tabela rozkładów jazdy
  - **zespol_id**
  - **slupek_id**
  - **linia**
  - **brygada**
  - **kierunek** - kierunek linii autobusowej
  - **trasa** - zestaw skrótów (np. TP-RZS) oznaczający trasę autobusu (w tym przypadku P+R al.Krakowska - rondo Zesłańców Syberyjskich)
  - **czas** - czas przyjazdu autobusu na przystanek [HH:MM:SS]

- Stops - tabela przystanków, zaweiera lokalizację
  - **zespol_id**
  - **slupek_id**
  - **nazwa**
  - **ulica_id**
  - **Lat** - szerokość geograficzna przystanku
  - **Lon** - Długość geograficzna przystanku
  - **kierunek** - kierunek jazdy z przystanku
  - **obowiazuje_od**

- Dictionary - słownik id pewnych obiektów

– **type** - typ obiektu:
  * ulice
  * miejsca
  * typy przystanków
– **nazwa** - nazwa obiektu np (typ: ulica, nazwa: Galla Anonima)
– **id** - identyfikator obiektu

Model fizyczny hurtowni danych



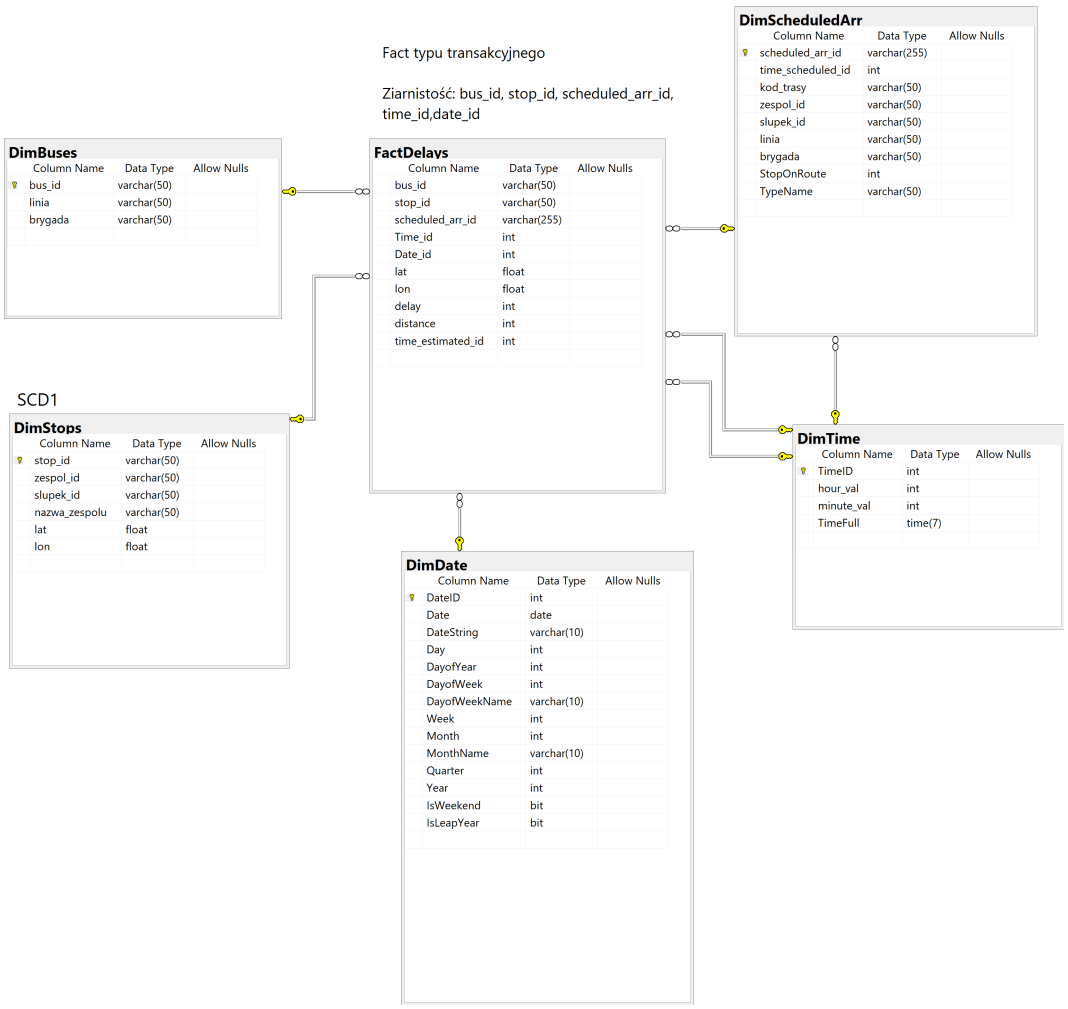
Rysunek 11: Oryginalne tabele w hurtowni danych.

W wcześniej opisanym procesie ekstrakcji danych, techniką web scrapping, otrzymane zostają wyżej opisane tabelki: Schedules, Stops, Routes, StopCode, StopInfo, Dict oraz CurrentPosition. Tabele DimTime oraz DimDate zostały sztucznie wygenerowane za pomocą skryptu SQL-owego. Schemat początkowych tabel pokazany jest powyżej [11].

Tabela faktów

Ostateczny model ma architekturę gwiazdy. Stanowi on podstawę analizy danych dotyczących transportu miejskiego. Model ten składa się z jednej tabeli faktów, **FactDelays**, która zawiera miarki: opóźnienie "delay", dystans do przebycia przez pojazd "distance", oraz estymowany czas przyjazdu.

Wprowadzona miara "delay" jest istotnym wskaźnikiem opóźnień na poszczególnych przystankach. Dzięki tej miarze możliwe będzie monitorowanie, analiza i raportowanie opóźnień na przystankach, co pozwoli na lepsze zarządzanie i optymalizację rozkładów jazdy oraz identyfikację problematycznych obszarów w systemie transportu miejskiego.



Rysunek 12: Model fizyczny hurtowni danych.

Wymiary

Zaprojektowany model składa się z 5 wymiarów:

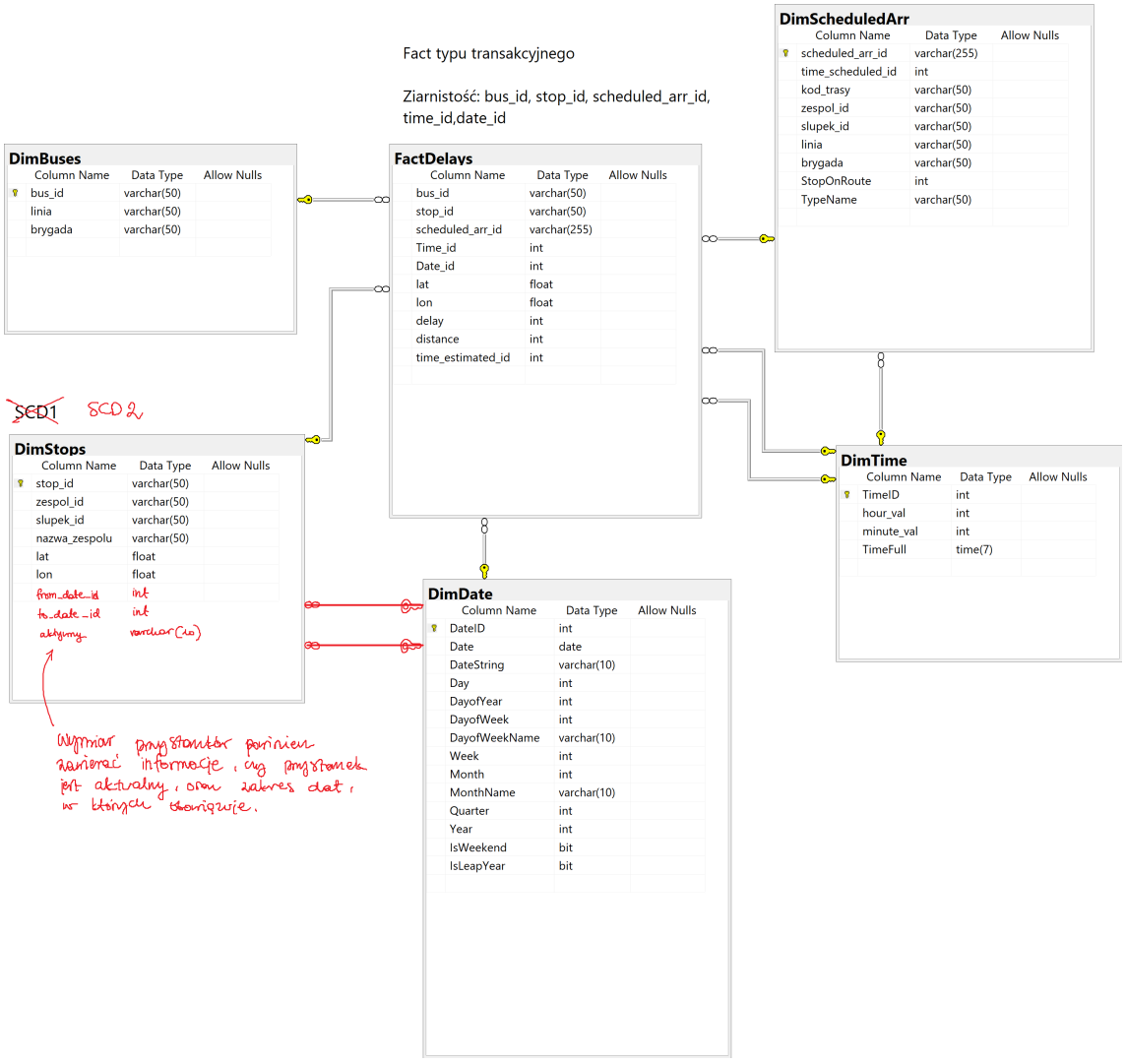
- **DimStops**, opisujący każdy przystanek, uwzględniając numer przystanku, np. Marszałkowska 01 i Marszałkowska 02. Zawiera informacje o lokalizacji danego przystanku.
- **DimBuses**, opisujący każdy autobus z danych. Identyfikuje każdy z nich jako odrębny obiekt, a nie jedynie np. "autobus linii 189".
- **DimScheduledArr**, przechowuje informacje o przystanku, lecz względem całej trasy. Wiersze identyfikuje para informacji: o autobusie oraz o przystanku. Zawiera informacje o kodzie trasy danego przejazdu autobusu na przystanek, oczekiwana godzina dotarcia autobusu na ten przystanek oraz typ tego przystanku na danej trasie.
- **DimDate**, opisujący każdy dzień w roku, w tej chwili z zakresu lat 2000-2030.
- **DimTime**, opisujący każdą minutę i godzinę w dobie.

Wszystkie wymiary i tabela faktowa pokazane są na diagramie [12].

Po wnioskach wyciągniętych po prezentacji projektu na zajęciach można rozwinąć i poprawić dotychczasowe rozwiązanie. Jedną z istotnych zmian jakie potencjalnie można wprowadzić, co zostało opisane przy procesie ETL, jest zmiana zmienności wymiaru DimStops z typu SCD1 na SCD2. To się również wiąże z dodaniem dodatkowych kolumn tabeli: "from_date_id" - identyfikator daty od której obowiązuje, "to_date_id" - identyfikator daty do której obowiązuje, "aktywny" - czy przystanek aktualnie obowiązuje.

W przypadku pojawienia się zmian, takich jak zmiana lokalizacji przystanku, to zachowujemy dane historyczne - wiersz ze starą lokalizacją pozostaje w wymiarze ale z odznaczoną zmienną "aktywny" i odpowiednio zmodyfikowaną wartością "to_date_id". Natomiast nowy wiersz ma nowe informacje na temat przystanku i ma zaznaczoną zmienną "aktywny". Dla aktywnych przystanków, zmienna "to_date_id" jest pusta.

Tak zmodyfikowany model jest pokazany poniżej [13].



Rysunek 13: Model fizyczny hurtowni danych, z poprawkami.

5 Planowane raporty

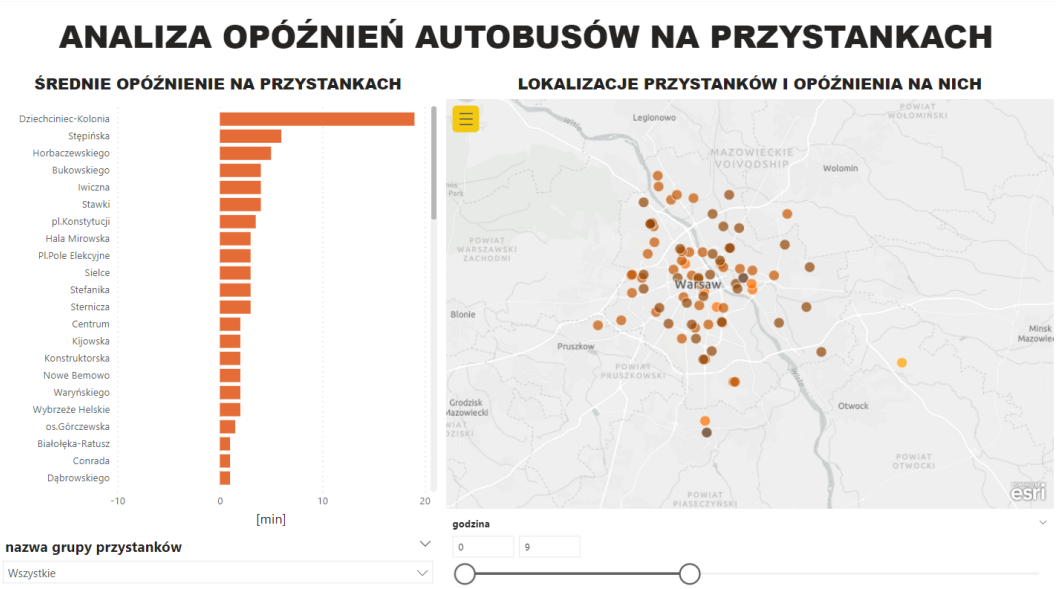
Dane dotyczące transportu miejskiego, takie jak rozkłady jazdy, informacje o przystankach, trasach, opóźnieniach, mogą być użyteczne w analizach dotyczących efektywności, wydajności i planowania transportu publicznego w Warszawie.

Część z raportowaniem została wykonana w narzędziu Power BI. Przykładowe pytania, na które można odpowiedzieć, to: Jakie są najbardziej obciążone przystanki autobusowe w różnych godzinach dnia? Jak wygląda punktualność środków transportu w różnych porach dnia? W których obszarach i o jakiej porze dochodzi do największych opóźnień?

Planowane raporty mają na celu odpowiedzieć na takie pytania i ulepszyć prognozę czasu i trasy podróży dla użytkowników środków transportu miejskiego.

Oto wizualizacje, które zostały zawarte w raportach:

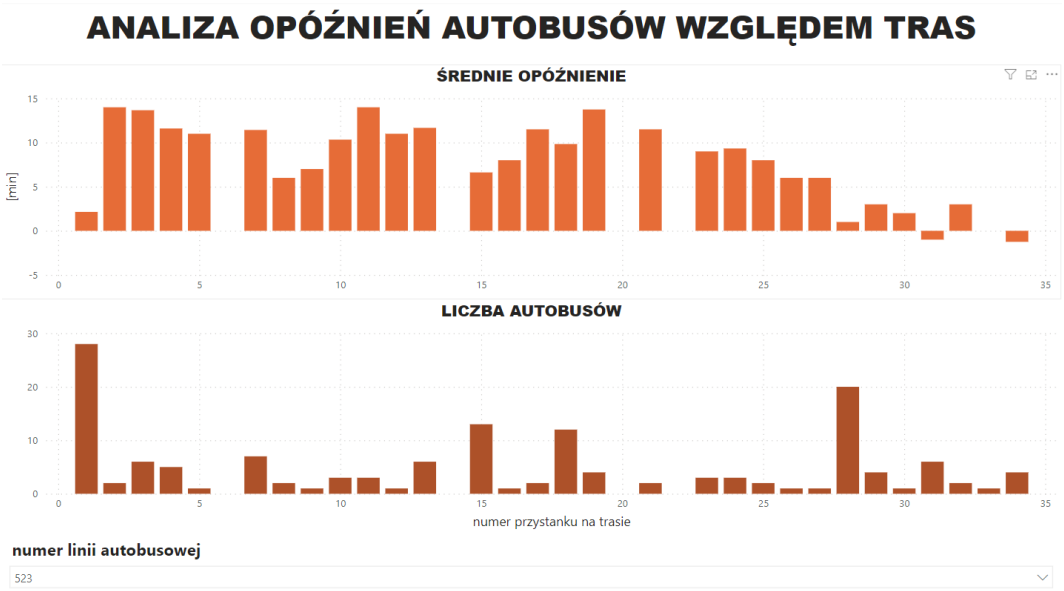
1. **Opóźnienia na przystankach** - ma ona na celu pokazanie obszarów na mapie, w których występują różnice między planowanym, a rzeczywistym czasem przyjazdu autobusów. Pierwsza wizualizacja pokazuje wykres słupkowy, która opisuje średnie opóźnienie pod daną grupą przystanków. Druga wizualizacja pokazuje przystanki, które są zaznaczone w odpowiednich miejscach na mapie, a te z opóźnieniami na trasie zakolorowane są jaśniejszym kolorem, w przeciwnym wypadku na odwrót.



Rysunek 14

Na tej stronie dostępne są 2 filtry: po nazwie grupy przystanków, jako rozwijana lista, oraz po godzinie w ciągu doby, w postaci suwaka. Oba leża u dołu strony.

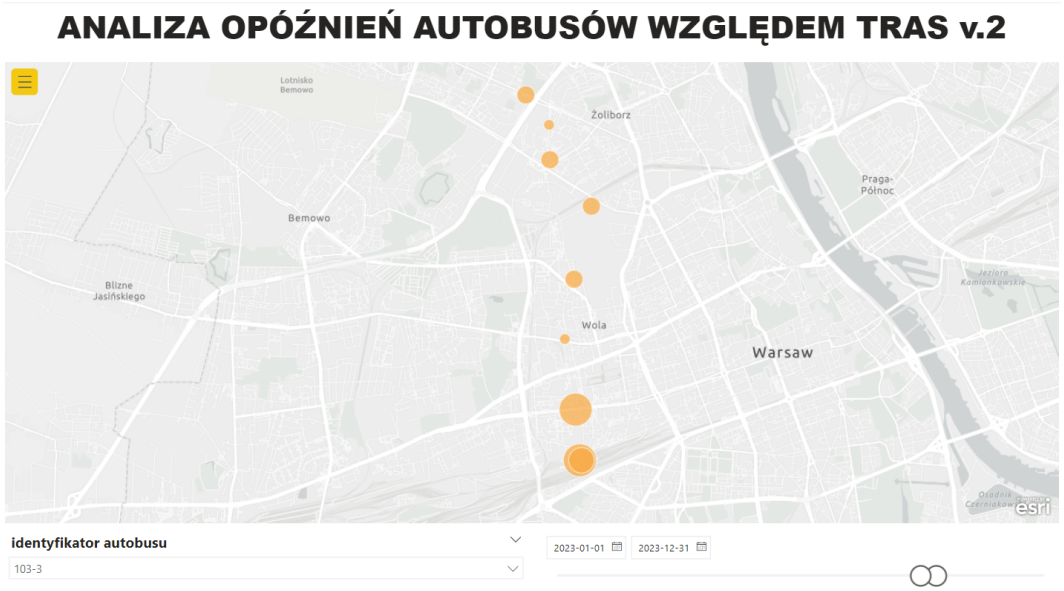
2. **Opóźnienia autobusów względem tras** - ma wizualizować średnie czasy opóźnienia na przystankach danej trasy oraz zagęszczenie położenia autobusów na trasie. Obie te rzeczy zostały zwizualizowane za pomocą wykresów słupkowych. Wykres składa się z osi poziomej, oznaczającej numer przystanku na trasie, i pionowej, czas opóźnienia podawany w minutach. Pod tym wykresem znajduje się filtr po numerze linii autobusowej.



Rysunek 15

Po prezentacji wyników, pojawiła się możliwość rozwoju tej analizy, o filtr czasu, co pozwoli na badanie "przepustowości" tras autobusowych w różnych okresach czasowych.

Oprócz powyższej [15] wizualizacji, do tej analizy wykonana została również wizualizacja położenia badanych autobusów, co jest pokazane na poniższej stronie raportu [16]. Wielkości punktów na mapie określają wartość średniego spóźnienia.

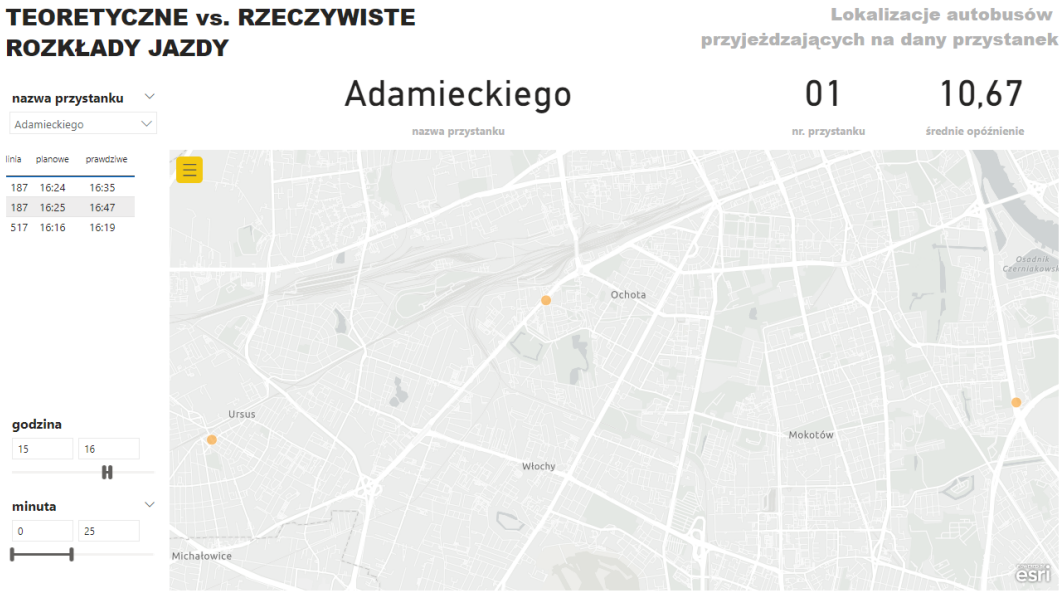


Rysunek 16

3. **Porównanie teoretycznych i "rzeczywistych" rozkładów jazdy z perspektywy przystanku** - w formie tabeli z 3 kolumnami: linia autobusowa, teoretyczny i estymowany czas przyjazdu pojazdu na dany przystanek. Tabela ma przypominać papierowe rozkłady jazdy, które znajdują się na przystanku, ale z dodatkową kolumną, która mówi nam o estymowanym czasie przyjazdu.

   Oprócz tego, na tej stronie znajduje się również mapa z lokalizacjami autobusów, które mają przyjechać na dany przystanek. Powyżej mapy znajduje się nazwa i numer przystanku, jak i średnie opóźnienie na dany przystanek.

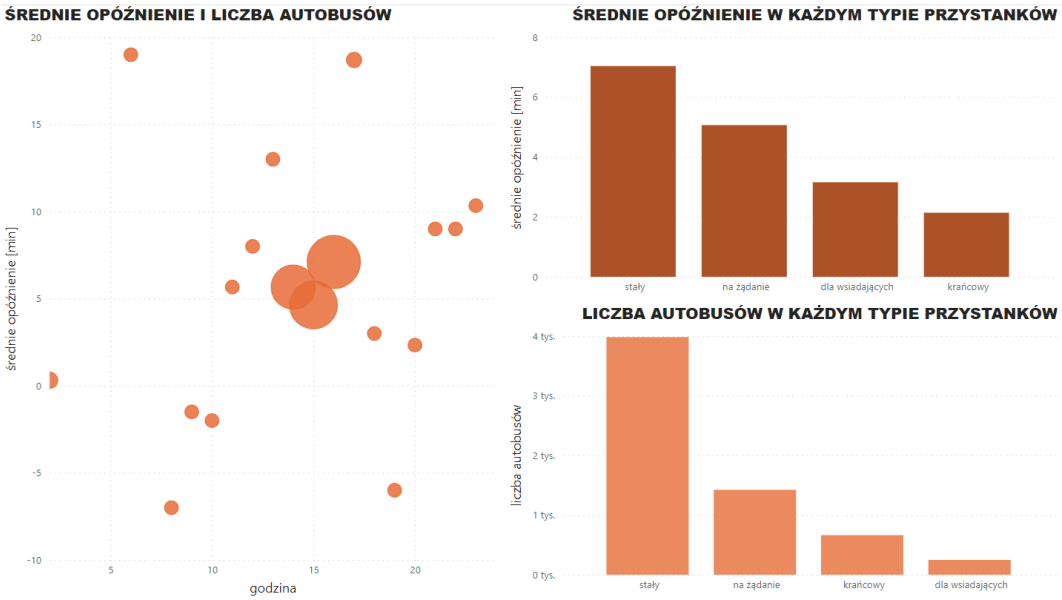
W lewym dolnym rogu znajdują się suwaki z możliwością filtrowania konkretnej godziny.



Rysunek 17

4. **Analiza opóźnienia w dziedzinie czasu** - pierwszy wykres opisuje zależność czasu, 24 godzin w ciągu dnia, do średniej wartości opóźnienia z tej godziny. Wielkości punktów określają liczbę pojedynczych przejazdów autobusów. Po prawo znajdują się 2 wykresy słupkowe, które wyliczają średnie opóźnienie i liczbę autobusów, z podziałem na typy przystanków na które przejazd się tyczy.

   Z pomocą tej analizy można zbadać godziny szczytu, co może pomóc przy optymalizacji liczby przejazdów w danych godzinach i na konkretnych typach przystanków.



Rysunek 18

6 Testowanie

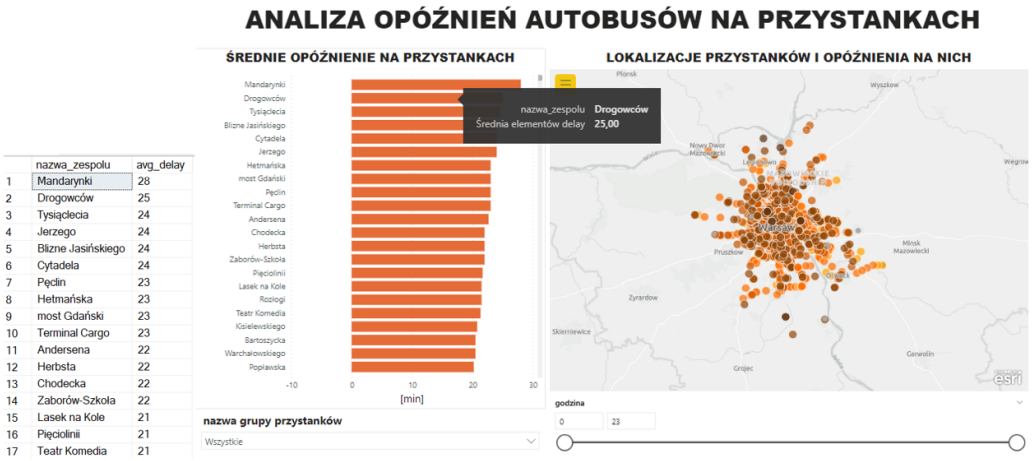
Testowanie obejmuje wszystkie kluczowe aspekty projektu, takie jak poprawność przetwarzania danych, zgodność z wymaganiami biznesowymi, prawidłowość generowanych raportów, wydajność systemu, odporność na błędy i niezawodność. Testy przeprowadzane są na wszystkich etapach projektu, aby upewnić się, że cały system działa zgodnie z oczekiwaniami.

Wyniki testów są monitorowanie, aby zapewnić, że wszelkie napotkane problemy zostaną rozwiązane przed wdrożeniem systemu.

6.1 Testowanie wyników raportów

Skrypt do testowania wyników raportów znajduje się w pliku WALIDACJA.sql. Poniżej zaprezentowane są wyniki otrzymane ze skryptu obok wyników z raportu.

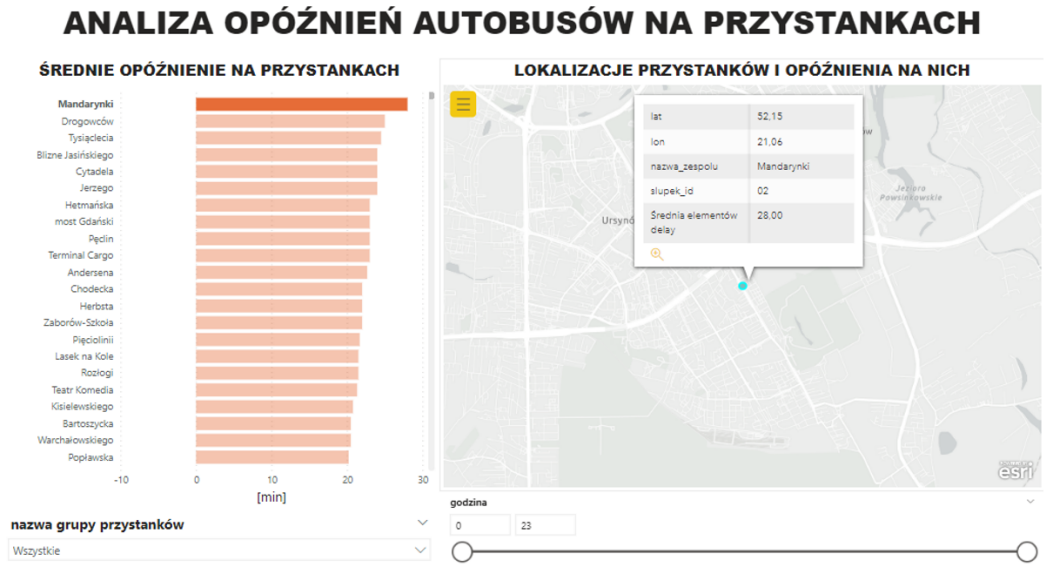
Test 1



	nazwa_zespolu	avg_delay
1	Mandarynki	28
2	Drogowców	25
3	Tysiąclecia	24
4	Jerzego	24
5	Blizne Jasińskiego	24
6	Cytadela	24
7	Pęclin	23
8	Hetmańska	23
9	most Gdański	23
10	Terminal Cargo	23
11	Andersena	22
12	Herbsta	22
13	Chodecka	22
14	Zaborów-Szkoła	22
15	Lasek na Kole	21
16	Pięciolinii	21
17	Teatr Komedia	21

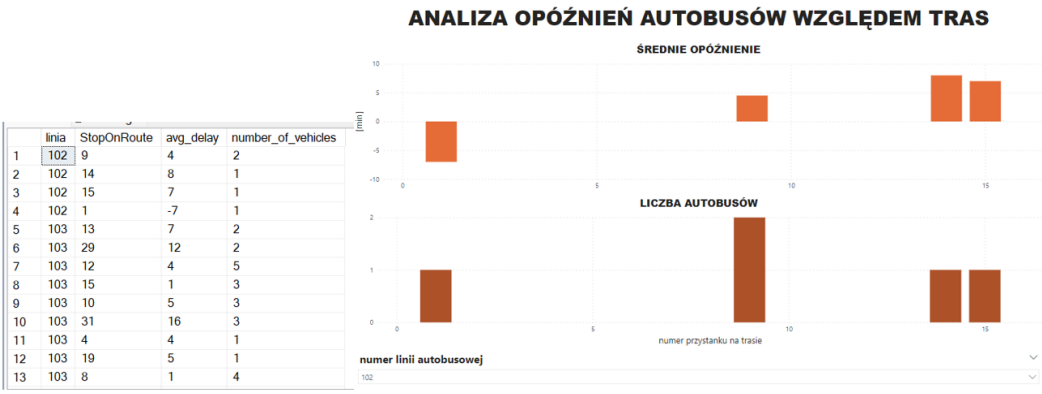
Rysunek 19

	stop_id	nazwa_zespolu	lon	lat	delay
1	3190-02	Mandarynki	21,063123	52,145823	28



Rysunek 20

Test 2



	linia	StopOnRoute	avg_delay	number_of_vehicles
1	102	9	4	2
2	102	14	8	1
3	102	15	7	1
4	102	1	-7	1
5	103	13	7	2
6	103	29	12	2
7	103	12	4	5
8	103	15	1	3
9	103	10	5	3
10	103	31	16	3
11	103	4	4	1
12	103	19	5	1
13	103	8	1	4

Rysunek 21

Test 3

	nazwa_zespolu	linia	TimeFull	TimeFull
1	Abrahama	123	14:42:00.0000000	14:53:00.0000000
2	Adamieckiego	187	14:42:00.0000000	15:03:00.0000000
3	Adamieckiego	187	15:40:00.0000000	15:53:00.0000000
4	Adamieckiego	187	16:26:00.0000000	16:35:00.0000000
5	Adamieckiego	187	16:26:00.0000000	16:47:00.0000000
6	Adamieckiego	187	16:41:00.0000000	16:43:00.0000000
7	Adamieckiego	191	14:03:00.0000000	14:04:00.0000000
8	Adamieckiego	191	15:33:00.0000000	15:33:00.0000000
9	Adamieckiego	207	15:39:00.0000000	15:41:00.0000000
10	Adamieckiego	207	16:39:00.0000000	16:41:00.0000000
11	Adamieckiego	517	15:32:00.0000000	15:38:00.0000000
12	Adamieckiego	517	16:17:00.0000000	16:19:00.0000000
13	Adampolska	111	14:32:00.0000000	14:34:00.0000000

nazwa przystanku

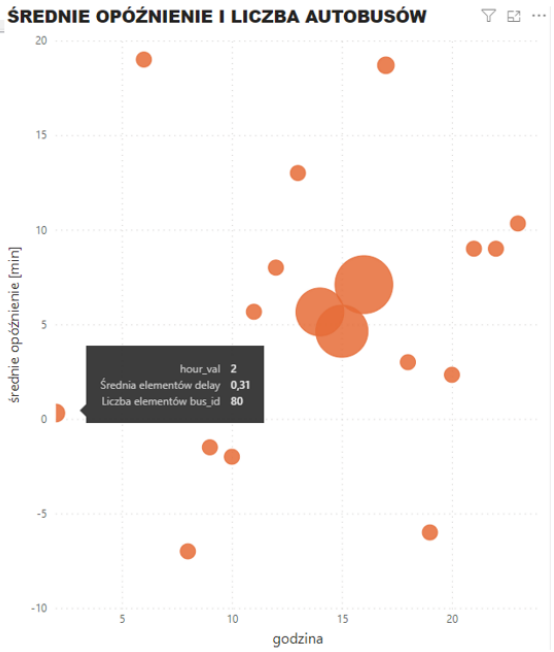
Adamieckiego

linia	planowe	prawdziwe
187	14:41	15:03
187	15:38	15:53
187	16:24	16:35
187	16:37	16:43
187	16:25	16:47
191	14:02	14:04
191	15:32	15:33
207	15:38	15:41
207	16:38	16:41
517	15:31	15:38
517	16:16	16:19

Rysunek 22

Test 4

	hour_val	avg_delay	number_of_vehicles
1	2	0	80
2	6	19	1
3	8	-7	1
4	9	-1	2
5	10	11	2
6	11	-4	2
7	12	8	1
8	13	7	2
9	14	6	1732
10	15	5	2768
11	16	5	1704
12	17	3	1
13	18	-6	1
14	19	0	1
15	20	7	4
16	21	8	4
17	22	11	4
18	23	8	5



Rysunek 23

Z powyższych wyników można zaobserwować, że porównane wyniki się zgadzają.

7 Podsumowanie i wnioski

Projekt został zrealizowany z sukcesem. Jego głównym celem było stworzenie modelu biznesowego umożliwiającego analizę opóźnień autobusów na przystankach oraz generowanie raportów z wynikami analizy.

W ramach projektu udało się zidentyfikować i pobrać niezbędne dane z API Warszawskiego, obejmujące rozkłady jazdy, współrzędne przystanków oraz trasy pojazdów. Dane te zostały przetworzone i załadowane do bazy danych MS SQL Server, co umożliwiło ich efektywne przetwarzanie i analizę.

Ostatecznie najlepiej sprawdził się model danych oparty na schemacie gwiazdy, gdzie tabelą faktową była tabela "FactDelays" przechowująca informacje o opóźnieniach autobusów na przystankach.

W trakcie projektu zostały osiągnięte główne cele, takie jak analiza opóźnień autobusów, generowanie raportów dotyczących średnich opóźnień w różnych godzinach i dniach tygodnia oraz identyfikacja najbardziej obciążonych tras i przystanków.

Wnioski z projektu są następujące:

1. Analiza opóźnień autobusów jest istotnym narzędziem do oceny jakości transportu publicznego i identyfikacji obszarów wymagających poprawy.

2. Raporty dotyczące średnich opóźnień w różnych godzinach i dniach tygodnia mogą dostarczyć cennych informacji dla zarządzania transportem i planowania rozkładów jazdy.

3. Identyfikacja najbardziej obciążonych tras i przystanków może pomóc w podejmowaniu decyzji dotyczących alokacji zasobów i optymalizacji sieci transportowej.

4. Analiza dostarczonych danych może również pomóc przy problemie wykrywania godzin szczytów w mieście.

Wszystkie uwzględnione potencjalne zmiany i ścieżki rozwoju projektu, takie jak dodatkowe dane do analizy i modyfikacje modelu fizycznego, zostały szczegółowo opisane w dokumentacji projektu. W ten sposób istnieje możliwość dalszego rozwijania projektu i dostosowywania go do ewentualnych zmian i wymagań biznesowych w przyszłości.

8 Podział pracy

1. Odnalezienie zródeł danych: Piotr Bielecki
2. Projekt modelu fizycznego: Laura Hoang, Piotr Bielecki
3. Proces ETL "Extract": Piotr Bielecki
4. Proces ETL "Transform & Load": Laura Hoang, Piotr Bielecki
5. Tworzenie raportów: Laura Hoang