

Hurtownie danych: Specyfikacja wymagań dla procesów transportu tramwajowego

1. Ogólny opis procesów biznesowych

1.1 Proces operacyjny przejazdów tramwajów

Proces operacyjny przejazdów tramwajów obejmuje systematyczne zarządzanie liniami tramwajowymi, rozkładami jazdy i dostarczaniem usług w czasie rzeczywistym. Tramwaje kursują na predefiniowanych trasach przez różne dzielnice Gdańska, zgodnie z ustalonymi rozkładami jazdy i przypisanymi motorniczymi. Każdy tramwaj działa w ramach określonego schematu usług zdefiniowanego datami kalendarza i regułami wyjątków. System monitoruje czasy przyjazdu na każdym przystanku, porównuje je z czasami zaplanowanymi i rejestruje wszelkie odchylenia. Gdy dochodzi do problemów operacyjnych, kierowcy zgłaszają awarie w systemie, co uruchamia interwencje serwisowe.

Metryki wydajności generowane przez ten proces:

- Miesięczna redukcja całkowitych opóźnień tramwajów o co najmniej 2% w stosunku do poprzedniego miesiąca
- Miesięczna redukcja całkowitej liczby awarii tramwajów o co najmniej 3% w stosunku do poprzedniego miesiąca
- Procent punktualności (mierzony jako odsetek przejazdów przybywających w ciągu 4 minut od czasu zaplanowanego)
- Średnie opóźnienie na przejazd w minutach

Problemy analityczne:

- Dlaczego opóźnienia się istotnie różnią między różnymi liniami tramwajowymi?
- Jakie czynniki (pogoda, godzina dnia, dzień tygodnia) mają największy wpływ na niezawodność usługi?

1.2 Proces obsługi serwisowej tramwajów

Gdy tramwaj doświadczy awarii podczas kursu, motorniczy zgłasza problem do systemu, a do miejsca zdarzenia wysyłany jest zespół serwisowy. Informacja o awarii jest rejestrowana w tabeli malfunctions, natomiast powiązane opóźnienie w tabeli stop_times.

Zespoły techniczne mogą usunąć usterkę na miejscu lub odholować tramwaj do pętli w celu naprawy.

Dane te są kluczowe do analizy niezawodności pojazdów i infrastruktury oraz do

identyfikacji typów awarii, które najczęściej powodują opóźnienia lub wymagają interwencji w zajezdni.

Metryki wydajności generowane przez ten proces:

- Procent awarii usunięty na miejscu w stosunku do napraw w pętli
- Średnie opóźnienie spowodowane awariami
- Częstość powtarzających się awarii na tej samej linii

Problemy analityczne:

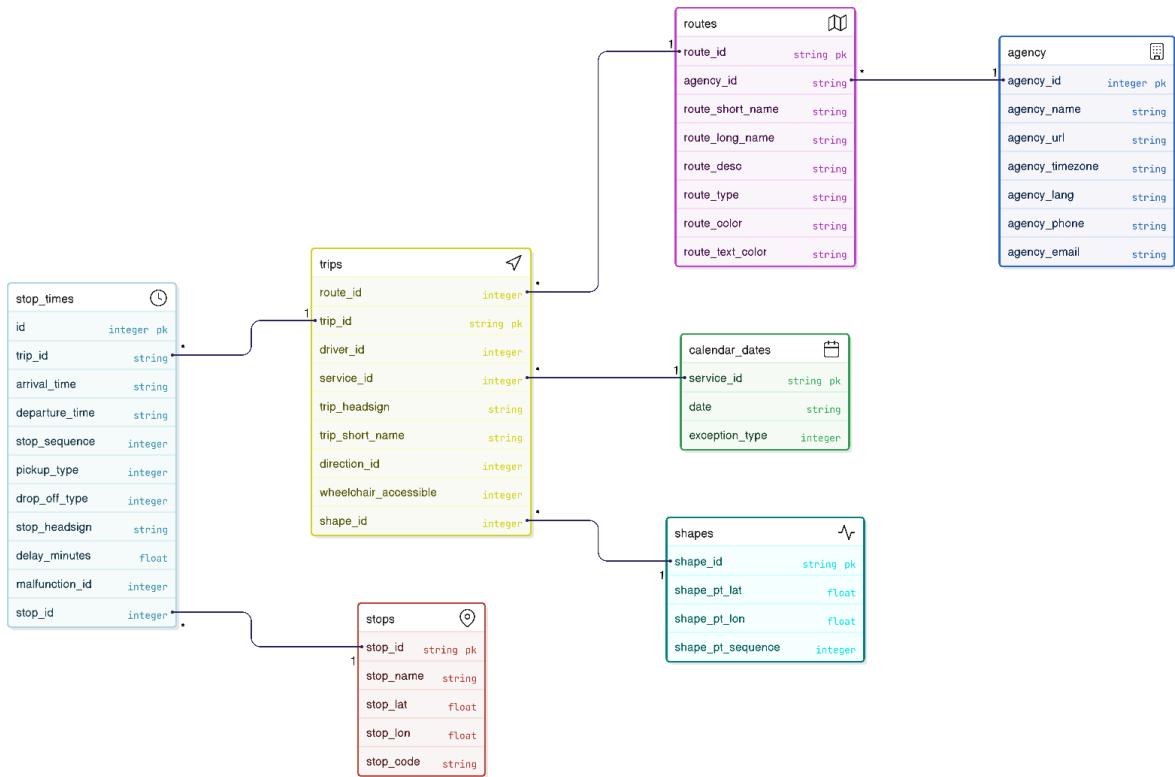
- Które linie tramwajowe mają największączęstość występowania awarii?
 - Czy określone typy awarii częściej wymagają naprawy w pętli?
-

2. Cele mierzalne

Cel 1: Osiągnąć zmniejszenie całkowitych opóźnień tramwajów z miesiąca na miesiąc o co najmniej 2%, mierzone jako suma wszystkich indywidualnych opóźnień przejazdów w całej sieci dla każdego miesiąca w porównaniu z miesiącem poprzednim, na podstawie danych z okresu październik 2024 - październik 2025.

Cel 2: Osiągnąć zmniejszenie całkowitej liczby awarii tramwajów z miesiąca na miesiąc o co najmniej 3%, mierzone jako liczba wszystkich zdarzeń awaryjnych zgłoszonych we wszystkich tramwajach dla każdego miesiąca w porównaniu z miesiącem poprzednim, na podstawie danych z ostatnich 12 miesięcy.

3. Struktury źródeł danych



3.1 Źródło danych relacyjne: Baza danych GTFS (PostgreSQL/MySQL)

Tabela: AGENCY

Atrybut	Typ	Opis	Klucz
agency_id	Integer	Unikatowy identyfikator operatora transportu	PK
agency_name	String	Nazwa firmy transportowej (ZTM)	
agency_url	String	Adres strony internetowej	
agency_timezone	String	Strefa czasowa obszaru obsługi	
agency_lang	String	Język podstawowy	
agency_phone	String	Numer telefonu kontaktowego	
agency_email	String	Adres e-mail	

Tabela: ROUTES

Atrybut	Typ	Opis	Klucz
route_id	Integer	Unikatowy identyfikator linii	PK
agency_id	Integer	Odniesienie do operatora	FK
route_short_name	String	Krótką nazwa linii (np. "3", "12")	
route_long_name	String	Pełna nazwa linii	
route_desc	String	Opis linii	
route_type	String	Typ usługi (tramwaj)	
route_color	String	Kolor wizualny linii	
route_text_color	String	Kolor tekstu dla wyświetlania	

Tabela: CALENDAR_DATES

Atrybut	Typ	Opis	Klucz
id	Integer	Syntetyczny klucz główny	PK
service_id	Integer	Identyfikator schematu usług	
date	String	Data (format YYYY-MM-DD)	
exception_type	Integer	1=usługa dodana, 2=usługa usunięta	

Tabela: TRIPS

Atrybut	Typ	Opis	Klucz
trip_id	String	Unikatowy identyfikator kursu	PK

driver_id	Integer	Unikatowy identyfikator kierującego pojazdem	FK
tram_id			
route_id	Integer	Odniesienie do linii	FK
service_id	Integer	Odniesienie do schematu usług	FK
trip_headsign	String	Kierunek/cel wyświetlany	
trip_short_name	String	Wewnętrzny identyfikator kursu	
direction_id	bool	Kierunek (0=w jedną stronę, 1=powrót)	
shape_id	Integer	Odniesienie do kształtu trasy	FK
wheelchair_accessible	Integer	Wskaźnik dostępności	

Tabela: STOPS

Atrybut	Typ	Opis	Klucz
stop_id	String	Unikatowy identyfikator przystanku	PK
stop_name	String	Nazwa przystanku	
stop_lat	Float	Szerokość geograficzna	
stop_lon	Float	Długość geograficzna	
stop_code	String	Alternatywny kod przystanku	

Tabela: STOP_TIMES

Atrybut	Typ	Opis	Klucz
id	String	Syntetyczny	PK

		klucz główny	
trip_id	String	Odniesienie do kursu	FK
arrival_time	String	Zaplanowany czas przyjazdu (HH:MM:SS)	
departure_time	String	Zaplanowany czas odjazdu (HH:MM:SS)	
stop_id	String	Odniesienie do przystanku	FK
stop_sequence	Integer	Kolejność na trasie	
pickup_type	Integer	Wysiadanie pasażerów dozwolone	
drop_off_type	Integer	Wiadanie pasażerów dozwolone	
stop_headsign	String	Cel specyficzny dla przystanku	
delay_minutes	Float	Opóźnienie w minutach	
malfunction_id	Integer	Klucz awarii	FK

3.2 Źródło danych nierelacyjne: Plik CSV z danymi o opóźnieniach

Plik 1: malfunctions.csv

Nazwa pliku: malfunctions.csv

Struktura pliku: malfunction_id,type,on_site,description

Kolumna	Typ	Opis	Format
malfunction_id	Integer	Unikatowy identyfikator awarii	Numeryczny (PK)
type	String	Typ usterki	Tekst (electrical, mechanical, brake, door, pantograph,

			itp.)
on_site	Boolean	Czy naprawa odbyta się na miejscu	0 lub 1 (0=pętla, 1=na przystanku)
description	String	Szczegółowy opis usterki	Tekst

Plik 2: drivers.csv

Nazwa pliku: drivers.csv

Struktura pliku: driver_id,driver_name,driver_surname,PESEL

Kolumna	Typ	Opis	Format
driver_id	Integer	Unikatowy identyfikator motorniczego	Numeryczny (PK)
driver_name	String	Imię motorniczego	Tekst
driver_surname	String	Nazwisko motorniczego	Tekst
PESEL	VARCHAR(12)	Numer PESEL motorniczego	11-cyfrowy numer

Strategia integracji danych:

- Plik malfunctions.csv zawiera informacje o tym, czy naprawa odbyta się na miejscu czy w pętli. Powiązany jest z tabelą stop_times poprzez pole malfunction_id
- Plik drivers.csv powiązany jest z tabelą TRIPS poprzez driver_id, umożliwiając analizę wpływu doświadczenia motorniczego na awaryjność

4. Problemy analityczne

Problem analityczny 1: Dlaczego opóźnienia się istotnie różnią między różnymi liniami tramwajowymi?

Opis: Zarząd ZTM musi zrozumieć pierwotne przyczyny opóźnień usługi w całej sieci tramwajowej. Obejmuje to identyfikację linii z największymi opóźnieniami, czasowe wzorce opóźnień oraz czynniki środowiskowe, które przyczyniają się do opóźnień.

Problem analityczny 2: Jak awarie i ich charakterystyka wpływają na opóźnienia w sieci tramwajowej?

Opis: Zarząd ZTM chce zrozumieć, w jaki sposób typy awarii tramwajów, ich lokalizacja oraz występowanie w czasie wpływają naczęstość i długość opóźnień. Celem jest określenie, które typy usterek mają największy wpływ na punktualność, oraz wskazanie linii i przystanków najbardziej podatnych na problemy techniczne.

5. Zapytania analityczne

Zapytania dla problemu analitycznego 1: "Dlaczego opóźnienia się istotnie różnią między różnymi liniami tramwajowymi?"

Zapytanie 1: Porównać średnie opóźnienie (w minutach) dla każdej linii tramwajowej w bieżącym miesiącu w stosunku do poprzedniego miesiąca, identyfikując, które linie się poprawiły, a które pogorszyły.

Zapytanie 2: Zidentyfikować dni tygodnia z największymi opóźnieniami (średnia wszystkich opóźnień w minutach) i porównać ten wzorzec między bieżącym miesiącem a poprzednim.

Zapytanie 3: Analizować wzorce opóźnień godzinowych w celu identyfikacji godzin szczytu dla każdej linii. Porównać opóźnienia w godzinach szczytu między dniami tygodnia a weekendami.

Zapytanie 4: Porównać średnie opóźnienia w minutach na konkretnych przystankach i sprawdzić czy częstotliwość opóźnień jest skorelowana z określonymi lokalizacjami.

Zapytanie 5: Sklasyfikować opóźnienia według ważności (niskie: 1-5 min, średnie: 6-15 min, wysokie: >15 min) dla każdej linii i obliczyć procentowy rozkład dla bieżącego miesiąca w stosunku do poprzedniego miesiąca.

Zapytania dla problemu analitycznego 2: "Jak awarie i ich charakterystyka wpływają na opóźnienia w sieci tramwajowej?"

Zapytanie 6: Dla każdej linii tramwajowej obliczyć średnie opóźnienie wynikające z awarii oraz porównać je z opóźnieniami niezwiązanymi z awariami.

Zapytanie 7: Zidentyfikować typy awarii, które najczęściej skutkują opóźnieniami przekraczającymi 10 minut i określić średnie opóźnienie spowodowane każdym typem usterek (wymaga danych z pliku stop_times i malfunctions).

Zapytanie 8: Dla każdego przystanku obliczyć średnie opóźnienie spowodowane awariami i liczbę wystąpień awarii.

Zapytanie 9: Porównać frekwencję opóźnień związanych z awariami elektrycznymi, mechanicznymi i innymi typami w celu określenia priorytetów serwisu.

Zapytanie 10: Dla każdej linii tramwajowej obliczyć miesięczny procent redukcji opóźnień i awarii w celu oceny postępu w osiąganiu celów redukcji o 2% i 3% (wymaga danych z tabel ROUTES, STOP_TIMES).

6. Zapytanie wymagające dodatkowego źródła danych

Zapytanie 11 (Wymagane dodatkowe źródło danych): Analizować wpływ warunków pogodowych naczęstość opóźnień według linii i typu awarii, porównując pogodę suchą, deszczową i śnieżną.

Wymagane dodatkowe dane: Historyczne dane pogodowe powiązane datą. Dane historyczne mogą pochodzić z publicznie dostępnych API meteorologicznych (np. OpenWeatherMap dane historyczne).

7. Zapytanie wymagające zmian w procesach biznesowych

Zapytanie 12 (Wymagana zmiana procesu): Określić zależność między doświadczeniem motorniczego a częstotliwością awarii tramwaju; zidentyfikować, czy bardziej doświadczeni kierowcy mają mniej awarii pojazdu i czy problemy specyficzne dla kierowcy korelują z określonymi liniami lub typami tramwajów.

Wymagane dodatkowe dane: Wzbogacone dane profilu motorniczego obejmujące:

- Staż pracy/doświadczenie (lata pracy na stanowisku)
- Kwalifikacje i specjalizacje kierowcy
- Historia poprzednich przypisanych pojazów
- Formalne raporty incydentów przypisane do działania kierowcy w stosunku do awarii mechanicznych

Dlaczego wymagana jest zmiana procesu biznesowego: Obecnie system nie różnicuje systematycznie między awariami spowodowanymi niepowodzeniami mechaniczno-pojazdu a awariami spowodowanymi błędem lub zaniedbaniem kierowcy. Aby odpowiedzieć na to zapytanie, proces serwisu musiałby być zmodyfikowany, aby wymagać od techników zaklasyfikowania każdej awarii jako „przypisanej pojazdu” lub „przypisanej kierowcy” z udokumentowanymi dowodami. Dodatkowo dane przypisania motorniczych muszą być wzbogacone o wskaźniki doświadczenia. Zmiany te wymagałyby:

- Zmodyfikowanego formularza raportowania napraw w systemie
- Nowego szkolenia dla pracowników serwisu dotyczącego procedur klasyfikacji
- Lepszej integracji danych HR pokazujących staż i doświadczenie motorniczych
- Możliwego wdrożenia protokołów dochodzenia incydentów w celu określenia pierwotnych przyczyn

8. Oczekiwane wyniki analityki

System BI umożliwi zarządowi ZTM:

- Monitorowanie postępu w osiąganiu celów miesięcznej redukcji opóźnień i awarii
- Identyfikację i priorytetyzację interwencji na problemowych liniach i pojazda
- Optymalizację harmonogramów serwisu na podstawie wzorców awarii
- Korelację wydajności operacyjnej z czynnikami środowiskowymi i kontekstowymi
- Prognozowanie niezawodności usług i planowanie inwestycji w pojemność
- Wsparcie podejmowania decyzji dotyczących optymalizacji tras i alokacji zasobów