

Specyfikacja wymagań dla procesu biznesowego “Przewóz pasażerski”

Paweł Kuszczak, Jan Kalwasiński

1. Ogólny opis procesu biznesowego

- a. Pasażer kupuje bilet na wybrane połączenie za pośrednictwem zewnętrznych partnerów zajmujących się sprzedażą biletów. Pasażer wybiera połączenie, określa liczbę biletów oraz klasę i dokonuje zakupu. Na podstawie otrzymanych danych firma “Polskie Koleje” realizuje transport pasażerów.

Wzrost miesięcznej liczby obsłużonych pasażerów na poziomie minimum 0.5%

Utrzymanie miesięcznego średniego obłożenia miejsc w pociągach na poziomie co najmniej 60%.

- b. Typowe zapytania

- Podaj, które przejazdy miały obłożenie mniejsze niż 60%.
- Porównaj dni tygodnia pod względem ilości zakupionych biletów dla danej trasy.
- Porównaj obłożenie pociągów jeżdżących o różnych godzinach tego samego dnia tą samą trasą.
- Porównaj liczbę sprzedanych biletów na dany kurs z tym samym z poprzedniego miesiąca.
- Ile biletów zostało sprzedanych w poszczególnych kategoriach (np. normalne, ulgowe, dziecięce).
- Jakie są najpopularniejsze trasy w różnych porach roku.

- c. Dane

Wszystkie dane o rezerwacjach biletów są pozyskiwane z systemu sprzedaży biletów partnerów. Partnerzy rejestrują transakcję w systemie, który zapisuje szczegóły dotyczące daty, godziny, trasy, typu pociągu, klasy i ulgi biletu oraz liczby sprzedanych biletów na dany kurs. Dodatkowo posiadamy plik CSV zawierający szczegółowe dane posiadanych pociągów.

2. Struktury źródeł danych

Nazwa tabeli	Atrybut	Typ atrybutu	Opis
Połączenie	Połączenie ID	int	PK - ID
	NazwaPołączenia	varchar(30)	Słowna nazwa połączenia np. Gdańsk -> Warszawa itp.
Przystanek	PrzystanekID	int	PK - ID
	PołączenieID	int	PK – FK (Połączenie)
	NazwaPrzystanku	varchar(30)	Słowna nazwa przystanku
	KolejnoscNaTrasie	int	Który to z kolei przystanek danego połączenia
Kurs	KursID	int	PK - ID
	Data	DATETIME	Dzień i godzina odjazdu ze stacji początkowej
	PociągID	int	ID pociągu FK (dane z pliku)
	PołączenieID	int	FK (Połączenie)
Bilet	BiletID	int	PK - ID
	PrzystanekPoczID	int	Przystanek początkowy FK (Przystanek)
	PrzystanekKonID	int	Przystanek końcowy FK (Przystanek)
	TypWagonu	varchar(30)	Podział na przedziałowe/bezprzedziałowe
	KursID	int	FK (Kurs)
CzasPrzejazdu	KursID	int	PK – FK (Kurs)
	Data	DATETIME	PK - Dzień
	PrzystanekID	int	PK – FK (Przystanek)

	GodzinaPrzyjazdu	DATETIME	Godzina przyjazdu na przystanek
	GodzinaOdjazdu	DATETIME	Godzina odjazdu z przystanku

Specyfikacja pociągów w posiadaniu firmy CSV

Dane o pociągach:

Kolumna A - Numer identyfikacyjny pociągu (liczbowy, bez miejsc po przecinku) - Unikalny identyfikator pociągu.

Kolumna B - Model pociągu (tekst) - Model pociągu.

Kolumna C – Rok produkcji (liczbowy) - Rok wyprodukowania pociągu

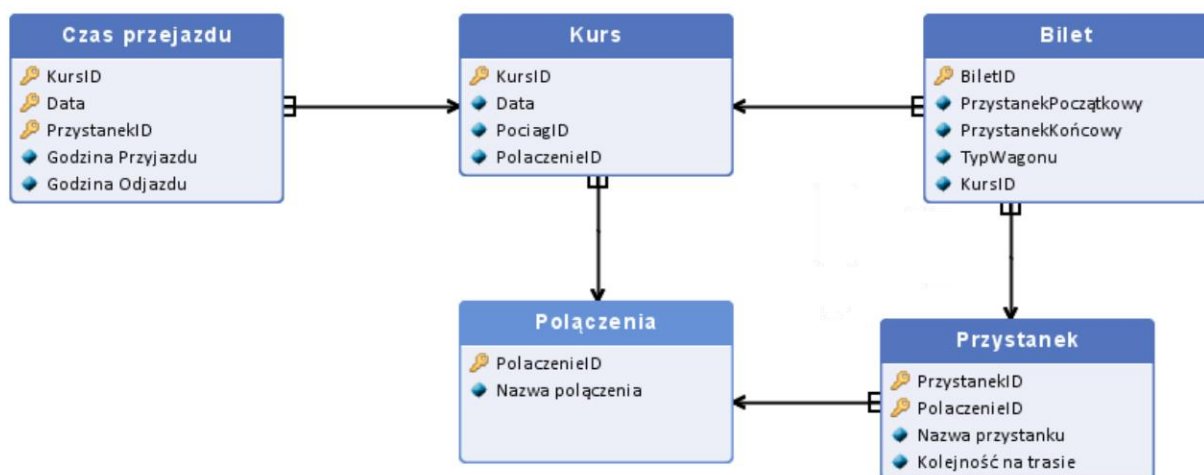
Kolumna D - Miejsca pasażerskie (liczbowy) - Liczba miejsc pasażerskich.

Kolumna E - Wagony bezprzedziałowe (liczbowy) - Liczba miejsc w wagonach bezprzedziałowych.

Kolumna F - Wagony przedziałowe (liczbowy) - Liczba miejsc w wagonach przedziałowych.

Kolumna G - Prędkość pociągu (liczbowy) - Maksymalna prędkość osiągalna przez pociąg

Kolumna H – Klimatyzacja (tak/nie) - Sprawna klimatyzacja w pociągu



3. Scenariusze problemów analitycznych

Problem: Jakie są czynniki wpływające na obłożenie pociągów niższe niż 60%?

1. Porównaj różnice w obłożeniu pociągów w różnych porach dnia.
2. Porównaj obłożenie pociągów w różnych dniach tygodnia.
3. Podaj połączenia, które mają najczęściej obłożenie poniżej 60%.
4. Wypisz, ile osób kończy trasę na jakim przystanku na danym kursie.
5. Podaj różnice w obłożeniu pociągów w różnych miesiącach roku na danych połączeniach.
6. Porównaj obłożenie pociągu na danym kursie z obłożeniem pociągu konkurencyjnego przewoźnika wyruszającym o podobnej godzinie.
7. Podaj średnie obłożenie pociągów o różnej ocenie czystości wagonów.

Problem: Jak specyfikacja techniczna pociągów wpływa na obłożenie?

1. Podaj średnie obłożenie wagonów przedziałowych/bezprzedziałowych danego pociągu na trasach jakie wykonał w ciągu miesiąca.
2. Porównaj obłożenie wszystkich modeli pociągów wykonujących dane połączenie o podobnej godzinie.
3. Porównaj obłożenie pociągów jeżdżących na tej samej trasie, ale z inną maksymalną prędkością.
4. Porównaj obłożenie pociągów jeżdżących na tej samej trasie, wyprodukowanych w innych latach.
5. Porównaj średnie obłożenie pociągów kursujących w okresie letnim, wyposażonych w klimatyzację, z pociągami bez klimatyzacji.

4. Dane potrzebne do problemów analitycznych

Sposób wyliczania średniego obłożenia w pociągu:

Obłożenie na kursie składającym się z n- przystanków będziemy liczyć w następujący sposób:

- Filtruujemy bilety po kolumnie KursID – tabela bilety
- Przystanki należące do danego kursu – tabela przystanki połączona z kursem poprzez FK PołączenieID
- Kolejność przystanków na danym kursie – tabela przystanek kolumna kolejność na trasie
- Ilość biletów na danym odcinku trasy będzie liczona sumarycznie na podstawie parametrów PrzystanekPoczątkowy PrzystanekKońcowy z tabeli bilety
Jeśli to PrzystanekPoczątkowy - liczba biletów (ilość pasażerów) +
Jeśli to PrzystanekKońcowy - liczba biletów (ilość pasażerów) -
- Godziny przyjazdu/odjazdu z danego przystanku dla danego kursu – tabela czas przejazdu kolumny godzina przyjazdu/odjazdu
- Ilość miejsc w pociągu z pliku CSV o specyfikacji (Pociąg łączymy z kursem po ID Pociągu)
- Średnie obłożenie kursu = średnia z każdego odcinka kursu (liczba biletów / liczbę miejsc w pociągu)

Utworzymy taką tabelę:

PołączenieID	KursID	Data	PrzystanekID	Kolejność na trasie	Godzina przyjazdu	Godzina odjazdu	Liczba biletów	PociągID	Liczba miejsc
	1	1234	09.10.2024	1001	1	06:30:00	06:32:00 0 + 80 - 0	333	200
	1	1234	09.10.2024	1002	2	06:45:00	06:47:00 80 + 10 - 2	333	200
	1	1234	09.10.2024	1003	3	06:55:00	06:57:00 88 + 5 - 20	333	200
	2	1235	09.10.2024	1001	1	14:50:00	14:52:00 0 + 65 - 0	444	120
	2	1235	09.10.2024	1003	2	14:59:00	15:09:00 65 + 11 - 3	444	120

1. Porównaj różnice w obłożeniu pociągów w różnych porach dnia.

- Obłożenie - wyliczone jak w przykładzie powyżej
- Pory dnia – tabela Czas Przejazdu kolumna godzina przyjazdu/odjazdu

2. Porównaj obłożenie pociągów w różnych dniach tygodnia.

- Obłożenie - wyliczone jak w przykładzie powyżej

- Dni tygodnia – tabela Kurs kolumna data
3. Podaj połączenia, które mają najczęściej obłożenie poniżej 60%.
 - Obłożenie - wyliczone jak w przykładzie powyżej
 - Połączenie - tabela Połączenie kolumna PołączenieID
 - Słowna nazwa połączenia - tabela Połączenie kolumna nazwa połączenia
 4. Wypisz, ile osób kończy trasę na jakim przystanku na danym kursie.
 - Jaki to kurs – tabela bilety kolumna KursID
 - Ile osób kończy trasę – tabela bilety kolumna PrzystanekKoncowy
 5. Podaj różnice w obłożeniu pociągów w różnych miesiącach roku na danych połączeniach.
 - Obłożenie - wyliczone jak w przykładzie powyżej
 - Połączenie - tabela Połączenie kolumna PołączenieID
 - Daty wykonanych kursów – tabela kurs kolumna data
 6. Porównaj obłożenie pociągu na danym kursie z obłożeniem pociągu konkurencyjnego przewoźnika wyruszającym o podobnej godzinie.
 - Obłożenie - wyliczone jak w przykładzie powyżej
 - Obłożenie konkurencji – brak danych – aby je uzyskać moglibyśmy skorzystać z usług firm analizujących rynek transportowy, które oferują raporty o liczbie pasażerów i stanie taboru u różnych przewoźników.
 7. Podaj średnie obłożenie pociągów o różnej ocenie czystości wagonów.
 - Ocena czystości - brak danych, moglibyśmy je otrzymać dając pasażerom możliwość oceny czystości po zakończeniu podróży na przykład poprzez ankietę pozwalającą na ocenę czystości pociągu w skali 1-10.
 8. Podaj średnie obłożenie wagonów przedziałowych/bezprzedziałowych danego pociągu na trasach jakie wykonał w ciągu miesiąca.
 - Obłożenie - wyliczone jak w przykładzie powyżej z tym, że liczymy osobno przedziałowe bilety i bezprzedziałowe
 - Przedziałowy/Bezprzedziałowy – tabela bilet kolumna typ wagonu
 9. Porównaj obłożenie wszystkich modeli pociągów wykonujących dane połączenie o podobnej godzinie.
 - Obłożenie - wyliczone jak w przykładzie powyżej

- Model pociągu - dane z pliku CSV o specyfikacji kolumna model
10. Porównaj obłożenie pociągów jeżdżących na tej samej trasie, ale z inną maksymalną prędkością.
- Obłożenie - wyliczone jak w przykładzie powyżej
 - Maksymalna prędkość pociągu - dane z pliku CSV o specyfikacji kolumna maksymalna prędkość
11. Porównaj obłożenie pociągów jeżdżących na tej samej trasie, wyprodukowanych w innych latach.
- Obłożenie - wyliczone jak w przykładzie powyżej
 - Trasa – tabela połączenia
 - Rok produkcji - dane z pliku CSV o specyfikacji kolumna rok produkcji
12. Porównaj średnie obłożenie pociągów kursujących w okresie letnim, wyposażonych w klimatyzację, z pociągami bez klimatyzacji.
- Obłożenie - jak w przykładzie powyżej
 - Okres letni – kolumna data z tabeli kurs
 - Dane o klimatyzacji - dane z pliku CSV o specyfikacji kolumna klimatyzacja