

# Specyfikacja wymagań dla procesu biznesowego przewozu osób

## 1. Opis ogólny procesu biznesowego

### a. Ogólny opis procesu biznesowego i opis wydajności metryk generowanych przez ten proces, możliwe bieżące problemy analityczne.

Proces przewozu zaczyna się od zakupu biletu. Następuje to przez Internet, w kasie na dworcu lub u konduktora w pociągu. Pasażer podaje klasę, wybiera rodzaj ulgi (jeśli przysługuje), stację początkową, końcową i czas odjazdu, które zostają wpisane w system wraz miejscem zakupu biletu. Na ich podstawie znajdowane jest połączenie i jeśli jest w nim wolne miejsce, wyliczana jest cena przejazdu i generowany jest rachunek. Pasażer może wybrać sobie dowolne wolne miejsce w obrębie wybranej klasy. Klient płaci za pomocą gotówki, karty płatniczej lub płatności internetowej (np. Blik). Bilet daje mu możliwość podróży konkretnym kursem, między wybranymi stacjami. Po wejściu do pociągu zajmuje wybrane miejsce. Bilet może być sprawdzany podczas kursu. Bilet jest ważny tylko i wyłącznie do wybranej przez pasażera stacji końcowej. Proces przewozu kończy się, kiedy pasażer opuszcza pokład pociągu.

**Miesięczny przychód ze sprzedaży biletów na poszczególnych trasach jest większy o minimum 0.3% niż w miesiącu poprzedzającym.**

**Miesięczna liczba pasażerów na wszystkich trasach jest większa o minimum 1% niż w miesiącu poprzedzającym.**

### b. Przykładowe pytania

Porównaj liczbę sprzedanych biletów oraz średni dystans w zależności od miejsca zakupu (dworzec/Internet/konduktor).

Podaj średni miesięczny przychód ze sprzedanych biletów w zeszłym roku.

Porównaj przychód z tego miesiąca z przychodem z poprzedniego miesiąca.

Porównaj średni przychód w dniu pracy i dniu wolnym.

Podaj najczęściej uczęszczane trasy.

Porównaj liczbę pasażerów w pierwszej klasie w tym i w zeszłym miesiącu.

Porównaj liczbę sprzedanych biletów ulgowych w tym i w zeszłym miesiącu.

Podaj średnią wysokość ulgi na sprzedanych biletach w danym miesiącu.

Podaj liczbę sprzedanych biletów w poszczególnych miesiącach zeszłego roku.

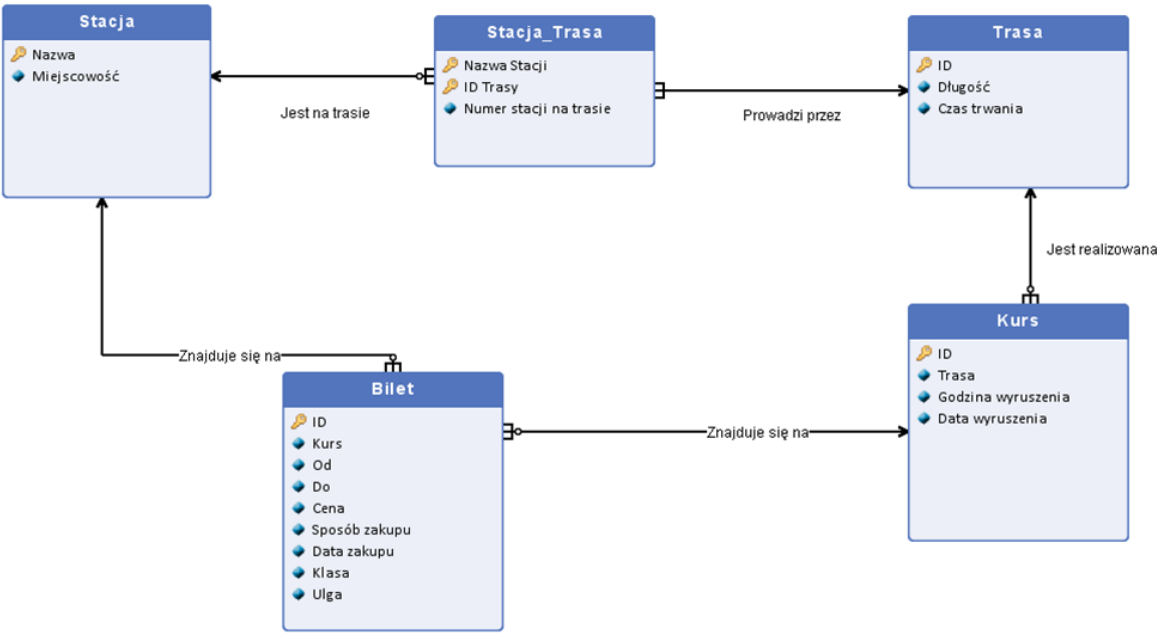
c. Dane

Wszystkie dane dotyczące sprzedaży pobierane są z systemu “TwojBilet”. System zawiera informacje o miejscu sprzedaży, stacji początkowej i końcowej wraz z ceną przejazdu, klasą, ulgą, datą i ID trasy.

Dodatkowo dane dotyczące dystansu między poszczególnymi stacjami zapisane są w pliku CSV.

2. Struktury źródeł danych

TwojBilet



Stacja			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
Nazwa	Tak	String (max 30 znaków)	Pełna nazwa stacji
Miejscowość	Nie	String (max 30 znaków)	Miejscowość w której znajduje się stacja.

Trasa			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID	Tak	Liczba naturalna	Unikalny identyfikator trasy
Długość	Nie	Liczba zmiennoprzecinkowa (km)	Długość całej trasy wyrażona w kilometrach
Czas trwania	Nie	Czas w formacie hh:mm	Czas trwania całej trasy

Stacja_Trasa			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
Nazwa Stacji	Tak	String (max 30 znaków)	Pełna nazwa stacji wchodzącej w skład trasy
ID Trasy	Tak	Liczba naturalna	ID trasy na której jest dana stacja
Numer stacji na trasie	Nie	Liczba naturalna	Liczba opisująca która na trasie jest dana stacja

Kurs			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID	Tak	Liczba naturalna	Unikalny identyfikator kursu
Trasa	Nie	Liczba naturalna z zakresu ID tabeli "Trasa"	Nr ID realizowanej trasy
Godzina wyruszenia	Nie	Czas w formacie hh:mm	Godzina, o której pociąg wyrusza z pierwszej stacji
Data wyruszenia	Nie	Data	Dzień, w którym kurs się odbywa

Bilet			
Nazwa	Klucz główny	Typ/Dziedzina	Opis
ID	Tak	Liczba naturalna	ID biletu
Kurs	Nie	Liczba naturalna	ID kursu, na który został zakupiony bilet
Od	Nie	Liczba naturalna	Nazwa stacji, na której wsiada pasażer
Do	Nie	Liczba naturalna	Nazwa stacji, na której wysiada pasażer
Cena	Nie	Liczba zmiennoprzecinkowa dodatnia	Cena biletu
Sposób zakupu	Nie	Kasa/Konduktor/Internet	Sposób zakupu biletu
Data zakupu	Nie	Data	Data zakupu biletu
Klasa	Nie	I / II	Wybrana klasa (pod kątem wygody podróży)
Ulga	Nie	0 - 100%	Ulga przysługująca pasażerowi

## Plik CSV

Plik *distances.csv* zawiera informacje o najkrótszym dystansie w kilometrach (po torach) między wszystkimi możliwymi parami stacji.

Struktura pliku:

Stacja1, Stacja2, Dystans(km)

XXX, YYY, 22.5

## 3. Scenariusze problemów analitycznych

**Dlaczego w zeszłym miesiącu zysk ze sprzedaży biletów wzrósł/zmalał:**

1. Porównaj łączny dystans przejechany przez osoby w I klasie w tym miesiącu w porównaniu z poprzednim.
2. Porównaj średnią długość trasy na bilecie w tym miesiącu w porównaniu z poprzednim.
3. Porównaj przychód ze sprzedaży biletów na poszczególnych trasach w tym miesiącu w porównaniu z poprzednim.

4. Porównaj liczbę sprzedanych biletów w dniach pracy w tym miesiącu w porównaniu z poprzednim.
5. Jaka była średnia długość trasy na bilecie, którego stacja początkowa to XXX w tym miesiącu w porównaniu z poprzednim?
6. Porównaj przychód ze sprzedaży biletów do/z danego miasta w okolicy dużego wydarzenia (np. Koncert, mecz) w porównaniu z dniem normalnym.

#### **Dlaczego liczba pasażerów w zeszłym miesiącu wzrosła/zmaląa:**

1. Porównaj liczbę sprzedanych biletów ulgowych w tym miesiącu w porównaniu z poprzednim.
2. Jaka była najczęściej wybierana stacja końcowa/początkowa w tym i w zeszłym miesiącu?
3. Które trasy były najpopularniejsze w godzinach szczytu w tym i w zeszłym miesiącu.
4. Które trasy były najpopularniejsze w okolicy świąt w tym i w zeszłym miesiącu.
5. Porównaj liczbę sprzedanych biletów I klasy w tym miesiącu w porównaniu z poprzednim.
6. Porównaj średnie zadowolenie pasażerów z przejazdu w tym i w zeszłym miesiącu.

## **4. Dane potrzebne do problemów analitycznych**

#### **Dlaczego w zeszłym miesiącu zysk ze sprzedaży biletów wzrósł/zmalął:**

1. Porównaj łączny dystans przejechany przez osoby w I klasie w tym miesiącu w porównaniu z poprzednim.

Dystans - *TwójBilet*, tabela *bilet*, atrybuty *Od* i *Do*, następnie dystans odczytany z pliku *distances.csv*

I klasa - *TwójBilet*, tabela *Bilet*, atrybut *Klasa*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

2. Porównaj średnią długość trasy na bilecie w tym miesiącu w porównaniu z poprzednim.

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

Długość trasy - *TwójBilet*, tabela *bilet*, atrybuty *Od* i *Do*, następnie dystans odczytany z pliku *distances.csv*

3. Porównaj przychód ze sprzedaży biletów na poszczególnych trasach w tym miesiącu w porównaniu z poprzednim.

Trasa - *TwójBilet*, tabela *Kurs*, atrybut *Trasa*

Przychód – suma cen biletów: *TwójBilet*, tabela *Bilet*, atrybut *Cena*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

4. Porównaj liczbę sprzedanych biletów w dniach pracy w tym miesiącu w porównaniu z poprzednim.

Liczba sprzedanych biletów - *TwójBilet*, tabela *Bilet*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

Dni pracy – odczytane z publicznie dostępnego kalendarza np. Google

5. Jaka była średnia długość trasy na bilecie, którego stacja początkowa to XXX w tym miesiącu w porównaniu z poprzednim?

Długość trasy - *TwójBilet*, tabela *bilet*, atrybuty *Od* i *Do*, następnie dystans odczytany z pliku *distances.csv*

Stacja początkowa - *TwójBilet*, tabela *stacja*, atrybut *nazwa*

6. Porównaj przychód ze sprzedaży biletów do/z danego miasta w okolicy dużego wydarzenia (np. Koncert, mecz) w porównaniu z dniem normalnym.

Przychód – suma cen biletów: *TwójBilet*, tabela *Bilet*, atrybut *Cena*

Miasto - *TwójBilet*, tabela *stacja*, atrybut *miescowość*

Okolice dużego wydarzenia – nie ma takiej informacji w źródłach danych. Można ją otrzymać za pomocą stron agregujących takie wydarzenia (np. [Trojmiasto.pl](http://Trojmiasto.pl))

#### **Dlaczego liczba pasażerów w zeszłym miesiącu wzrosła/zmaląa:**

1. Porównaj liczbę sprzedanych biletów ulgowych w tym miesiącu w porównaniu z poprzednim.

Liczba biletów ulgowych - *TwójBilet*, tabela *Bilet*, atrybut *Ulg*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

2. Jaka była najczęściej wybierana stacja końcowa/początkowa w tym i w zeszłym miesiącu?

Stacja początkowa/końcowa - *TwójBilet*, tabela *bilet*, atrybuty *Od* i *Do*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

3. Które trasy były najpopularniejsze w godzinach szczytu w tym i w zeszłym miesiącu.

Trasa - *TwójBilet*, tabela *Kurs*, atrybut *Trasa*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

Godziny szczytu - *TwójBilet*, tabela *Kurs*, atrybut *Godzina wyruszenia*

4. Które trasy były najpopularniejsze w okolicy świąt w tym i w zeszłym miesiącu.

Trasa - *TwójBilet*, tabela *Kurs*, atrybut *Trasa*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

Okolice świąt – odczytane z publicznie dostępnego kalendarza np. Google

5. Porównaj liczbę sprzedanych biletów I klasy w tym miesiącu w porównaniu z poprzednim.

Liczba biletów I klasy - *TwójBilet*, tabela *Bilet*, atrybut *Klasa*

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

6. Porównaj średnie zadowolenie pasażerów z przejazdu w tym i w zeszłym miesiącu.

Średnie zadowolenie klientów - nie ma takiej informacji w źródłach danych. Można ją uzyskać przeprowadzając ankietę wśród pasażerów.

Miesiąc - *TwójBilet*, tabela *Bilet*, atrybut *Data zakupu*

Aby pozyskać informację o zadowoleniu klientów należałoby przeprowadzić ankietę. Ankieta potrzebuje tylko jednego pytania - "W skali od 1-10, jak zadowolony/a jesteś ze swojego przejazdu?". Kod QR prowadzący do ankiety mógłby być nadrukowany na bilecie. Dodatkowo, do osób kupujących bilet przez Internet, mógłby zostać wysłany mail z prośbą o wypełnienie ankiety. Aby zachęcić pasażerów do wzięcia udziału w ankiecie, można zaoferować kupon zniżkowy na posiłki w pociągu w zamian za wypełnienie.