

Zad. 3. (Prezentacja wszystkich punktów tego zadania na zajęciach 14-16.04.2021)

Proces tworzenia hurtowni danych powinien być poprzedzony zrozumieniem „potrzeb biznesu” oraz rzeczywistości (dziedziny problemowej) reprezentowanej przez dostępne zasoby danych. Realizacja poniższego zadania ma uzmysłwić występujące problemy w określonym (wybranym) wycinku rzeczywistości, a następnie umożliwić zidentyfikowanie (określenie) potrzeb, celu i możliwości analiz biznesowych, by wspierać procesy decyzyjne (podejmowanie właściwych decyzji biznesowych).

Projekt HD – propozycja tematu

Proszę przygotować zakres realizacji projektu zgodnie z poniższą specyfikacją oraz przedyskutować propozycję projektu z osobą prowadzącą zajęcia. Poczynione uzgodnienia zarejestrować w formie wniosków. Na zajęciach laboratoryjnych należy przedstawić na forum grupy swoją propozycję tematu projektu (uzasadniając celowość i jego główne elementy 1.1 – 1.6) wykorzystując PowerPoint.

1.1. Tytuł projektu

Przyczyny opóźnień lotów w latach 2009 - 2018

1.2. Charakterystyka dziedziny problemowej

Transport lotniczy jest jednym z najważniejszych rodzajów transportu, niezastąpionym w przypadku szybkiego przewożenia ludzi i towarów na duże odległości, a liczba lotów wykonanych w ciągu jednego dnia potrafi sięgać 200 tys. [1]. Każde opóźnienie wiąże się z problemami logistycznymi oraz finansowymi (wynikającymi choćby z odszkodowań).

...

1.3. Opis obszaru analizy wraz z uzasadnieniem (wybrany fragment dziedziny, przeznaczony do szczegółowej analizy i opracowania hurtowni danych)

Analiza będzie dotyczyła lotów pasażerskich na całym świecie uwzględniając przewoźników oraz lotnisk zarówno krajowych, jak i międzynarodowych. Identyfikacja przyczyn opóźnień oraz ich wpływ na funkcjonowanie firm lotniczych oraz liczbę przewożonych pasażerów pozwoli wykryć główne czynniki obniżające jakość usług lotniczych. Wielozakresowa analiza będzie podstawą wskazania zarówno miejsc (krajów, lotnisk), jak różnych połączeń lotniczych, które wiążą się z potencjalnymi problemami w trakcie podróżowania. Ponadto, ustalenie źle funkcjonujących lotnisk i przewoźników dostarczy użytecznej wiedzy zarówno dla ich właścicieli, jak i osób zainteresowanych skorzystaniem z usług transportowych. ...

1.4. Problemy

P01 - Rosnące koszty funkcjonowania firm lotniczych

P02 – Niezadowolenie pasażerów

...

Zakres opracowania projektu HD – cz. 1.

Opóźnienia lotów w latach 2009 - 2018

Analiza będzie dotyczyła opóźnień lotów w wybranych latach. Analiza może wykryć kłopotliwe połączenia, mniej i bardziej zawodnych przewoźników, kłopotliwe lotniska, czy też pory, w których problemy zdarzają się najczęściej. Ponadto, analizie można poddać przyczyny opóźnień. Takie informacje pozwolą określić gdzie i dlaczego powstają opóźnienia, co daje szansę przeciwdziałać im. Problemem jest fakt występowania opóźnień, celem zminimalizowanie ich. W tym celu można zbadać następujące rzeczy, aby zdobyć wiedzę umożliwiającą przeciwdziałanie opóźnieniom:

1.4. Cel przedsięwzięcia

1. Wykrycie czynników ryzyka, które zwiększają statystyczne ryzyko pojawienia się opóźnień
2. Ocena wpływu warunków pogodowych na opóźnienia lotów
3. Zbadanie czynników czasowych: pora roku, dzień tygodnia, pora dnia
4. Określenie przyczyn, które powodują najbardziej dotkliwe opóźnienia
5. Ustalenie wpływu opóźnień na zainteresowanie usługami lotniczymi: liczba pasażerów, liczba rezygnacji z lotu, liczba negatywnych opinii i reklamacji

1.5.1. Oczekiwania i potrzeby w zakresie wsparcia podejmowania decyzji

Również z perspektywy pasażera badanie może być przydatne – może pomóc wybrać najpewniejszego przewoźnika. Analiza powinna dostarczyć informacji na temat opóźnień, takich jak:

- Czy są pory, kiedy opóźnienia zdarzają się częściej?
- Gdzie i komu najczęściej zdarzają się opóźnienia?
- Jak duży jest udział małych, a jak dużych opóźnień?
- Czy czas trwania lotu (lub odległość między punktami), ma wpływ na wielkość opóźnienia
- Sprawdzenie, w jaki dzień tygodnia najczęściej zdarzają się opóźnienia
- Sprawdzenie, jak duże opóźnienia powodują poszczególne przyczyny
- Sprawdzenie, w którym etapie opóźnienia są generowane
- ...

1.5.2. Zakres analizy – badane aspekty

Analiza dostarczając odpowiedzi na te pytania, powinna dostarczyć informacji, które będą pomocne w zidentyfikowaniu przyczyn opóźnień i ograniczenia ich, czyli pozwoli rozwiązać problem. Analizie będą poddane wielkość opóźnień, przewoźnik i trasa lotu (jej punkty i długość, na podstawie czasu). Ponadto, sprawdzany będzie czas wylotu i jego data. Dodatkowo, wyniki analizy na podstawie powyższych czynników, można nałożyć na przyczyny opóźnień.

...

1.6. Źródła danych

1.6.1. Lokalizacja, format, dostępność

Dane są dostępne pod na stronie [kaggle](#). Autor zbioru wskazuje pierwotne źródło danych - <https://www.transtats.bts.gov>, jednak dokładny link jest błędny.

Dane są dostępne w postaci plików csv, jeden dla każdego roku.

1.6.2. Wstępna ocena (liczba rekordów, zakres czasowy danych - faktów)

Zakres czasowy jest duży. Dla każdego roku jest około 6mln lotów, więc danych jest wiele, być może nawet będzie można rozważyć ograniczenie analizy do wybranych lat. Dane są dość kompletne, braki wynoszą do 2%, nawet po odrzuceniu wszystkich danych z brakami, co nie powinno być konieczne, danych pozostanie dużo.

Lp.	Nazwa pliku	Format	Rok	Liczba rekordów	Komentarz
1.					
2.					
3.	...				
	...				

1.6.3. Fakty

1.6.4. Kontekst analizy faktów np. czas (ziarnistość), lokalizacja, warunki pogodowe, itd.