Robert Kozakiewicz Gdańsk, 26.04.2015

Dawid Piasecki

Eksploracja danych

Zrozumienie warunków biznesowych i identyfikacja źródeł danych

**1. Cele biznesowe organizacji.**

1.1. Krótka charakterystyka organizacji

Nasza organizacja jest firmą ubezpieczeniową, zajmującą się ubezpieczeniami majątkowymi, a dokładnie komunikacyjnymi. Organizacja ta oferuje ubezpieczenia mieszkańcom stanu Nowy Jork.

1.2. Cele biznesowe organizacji

* zwiększenie bazy klientów "bezkolizyjnych" i bezpiecznych
* zwiększenie stawek klientom niebezpiecznym
* znalezienie grupy wiekowej o największej rentowności ubezpieczania
* zwiększenie zysków znajdując najmniej kolizyjne obszary Nowego Jorku

1.3. Kryteria sukcesu organizacji

* dziesięcioprocentowy spadek kosztów wypłacanych odszkodowań
* wzrost liczby klientów o pięć procent

**2. Cele biznesowe eksploracji danych.**

2.1.

a) Jak rozkłada się liczba przyznanych mandatów w stosunku do wieku kierowców?

b) Jak rozkłada się liczba przyznanych mandatów w stosunku do płci kierowców?

c) Jakie rodzaje mandatów dostają osoby w danej grupie wiekowej?

d) Jak pora roku wpływa na liczbę wypisywanych mandatów?

e) W jakich dzielnicach stanu Nowy Jork przyznawana jest największa liczba mandatów?

2.2.

a) Wyznaczenie trzech grup wiekowych (o maksymalnej rozpiętości pięciu lat) o największej liczbie przyznanych mandatów.

b) Wyznaczenie stosunku procentowego uzyskanych mandatów przez przedstawicieli obu płci.

c) Wyznaczenie trzech rodzajów najczęściej wypisywanych mandatów dla grup wiekowych o rozpiętości 10 lat (do 20, 20 - 30, 30 - 40 itd.).

d) Wyznaczenie stosunku procentowego wypisywanych mandatów w poszczególne pory roku.

e) Wyznaczenie trzech dzielnic, w których wypisano najwięcej mandatów.

**3. Identyfikacja źródeł danych.**

Będziemy korzystać z danych w formacie csv pobranych ze strony rządowej Stanów Zjednoczonych (http://catalog.data.gov/dataset/traffic-tickets-issued-beginning-2008). Dane te zawierają informacje o mandatach przyznanych kierowcom w stanie Nowy Jork w 2013 roku: rodzaj wykroczenia, kod wykroczenia, rok, miesiąc oraz czas wykroczenia, wiek oraz płeć kierowcy, stan, w którym wydano prawo jazdy kierowcy, oddział policji oraz sąd zajmujący się sprawą wykroczenia.

**4. Analiza źródeł danych pod kątem ich kompletności względem celów zdefiniowanych celów eksploracji danych.**

a) Potrzebne dane: wiek karanych kierowców - dostępny w zidentyfikowanym źródle.

b) Potrzebne dane: płeć karanych kierowców - dostępna w zidentyfikowanym źródle.

c) Potrzebne dane: rodzaj wykroczenia - dostępny w zidentyfikowanym źródle.

d) Miesiąc, w którym popełniono wykroczenie - dostępny w zidentyfikowanym źródle.

e) Miejsce popełnienia wykroczenia - niedostępne, jednak możliwe do wywnioskowania na podstawie oddziału policji wypisującego mandat.