

Hurtownie danych – Spr. 5.

PWr. WIZ, Data: 11-12.04.2022

Student	-----	Ocena
Indeks	<u>254489</u>	
Imię	<u>Jakub</u>	
Nazwisko	<u>Wrześniak</u>	

Zestaw składa się z 1 zadania. Pamiętaj o podaniu nr. indeksu oraz imienia i nazwiska.

Zad. 1. (Prezentacja wszystkich punktów tego zadania na zajęciach 11-12.04.2022)

Proces tworzenia hurtowni danych powinien być poprzedzony zrozumieniem „potrzeb biznesu” oraz rzeczywistości (dziedziny problemowej) reprezentowanej przez dostępne zasoby danych. Realizacja poniższego zadania ma uzmysłowić 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.**

Źródło badanych danych zostało zmienione, ponieważ baza danych zawierająca wypadki drogowe w Stanach Zjednoczonych, nie dostarczała wystarczających informacji o kierowcach czy ofiarach. Głównym obszarem zgromadzonych danych były warunki pogodowe oraz okoliczności zdarzenia.

Nowo wybrana baza danych – Wypadki w Wielkiej Brytanii, dostarcza bardziej szczegółowe informacje o sprawcach wypadków, typie pojazdów, ofiarach, co pozwoli na dokładniejszą, wieloobszarową analizę.

Zakres opracowania projektu HD – cz. 1.

1.1. Tytuł projektu

Analiza wypadków samochodowych w Wielkiej Brytanii w latach 2005 - 2015

1.2. Charakterystyka dziedziny problemowej

Pierwszy samochód powstał 136 lat temu w 1886 roku. Dziś, jest jednym z najbardziej popularnych środków transportu. Pomimo upływu wielu lat, wprowadzeniu praw regulujących zasady użytkowania, poruszania się samochodem oraz nakładaniu wysokich kar za ich nieprzestrzeganie, wciąż dochodzi do dużej liczby wypadków, w których giną ludzie.

Łączna liczba ofiar śmiertelnych wypadków samochodowych jest znacznie większa niż w przypadku katastrof lotniczych czy morskich i zwiększa się z każdym kolejnym rokiem.

Od kilkadziesiąt lat, dane o okolicznościach wypadków są gromadzone przez władze różnych krajów. Analiza tych danych pozwoli na zbadanie wpływu różnych czynników będących powodem wypadków.

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

Analiza będzie dotyczyć wypadków samochodowych w Wielkiej Brytanii w Latach 2005 – 2015. Dane gromadzone przez Departament Transportu rządu WB – „Road accidents and safety statistics” uwzględniają czas oraz okoliczności wypadku jak i warunki pogodowe, szczegóły pojazdu oraz dane o ofiarach. Identyfikacja przyczyn, okoliczności wypadków oraz miejsce ich najczęstszego występowania, pozwoli na podjęcie odpowiednich działań w celu zmniejszenia liczby ich wystąpień. Analiza rodzaju pojazdów czy ich wieku pozwoli na wprowadzanie dodatkowych obostrzeń co do wymagań pojazdu w celu wyeliminowania wypadków powodowanych przez stan pojazdu. Ponadto, dane o kierowcach takiej jak płeć czy grupa wiekowa, pozwolą na stworzenie profilu kierowcy, który jest najczęstszym sprawcą wypadków. Warto również zwrócić uwagę na dane ofiar oraz okoliczności w jakich zginęły. Informacje te pozwolą na zwiększenie bezpieczeństwa osób, które będą mogły zostać potencjalnymi ofiarami oraz wprowadzaniu odpowiednich zabezpieczeń, by w przypadku wypadku, poniesione obrażenia były jak najmniejsze.

1.4. Problemy

- P01 – rosnąca liczba wypadków samochodowych
- P02 – wzrost liczby ofiar śmiertelnych w wypadkach samochodowych

1.5. Cel przedsięwzięcia

1.5.1. Oczekiwania i potrzeby w zakresie wsparcia podejmowania decyzji

- Wykrycie najczęstszych przyczyn wypadków drogowych
- Zbadanie wpływu rodzaju pojazdu na ilość powodowanych wypadków
- Określenie profilu kierowców najczęściej powodujących wypadki
- Ograniczenie liczby poszkodowanych (oraz stopnia zadanych obrażeń)
- Identyfikacja miejsc, w których najczęściej dochodzi do wypadków

1.5.2. Zakres analizy – badane aspekty

Analiza pozwoli na zdobycie informacji dla wyżej określonych oczekiwań. Analiza odbędzie się na wielu płaszczyznach, dzięki czemu będzie można podjąć działania ograniczające liczbę wypadków na wielu poziomach tj. czas, w którym doszło do

wypadku, miejsce zdarzenia, panujące warunki pogodowe, doświadczenie jak i wiek kierowcy, typ pojazdu, okoliczności towarzyszące zdarzeniu oraz typ drogi. Powyższe aspekty mogą mieć wpływ na ilość wypadków, do których dochodzi na drogach.

1.6. Źródła danych

1.6.1. Lokalizacja, format, dostępność

Dane są dostępne pod na stronie [kaggle](#). Ich źródło można odnaleźć na stronie rządowej Wielkiej Brytanii - [www.gov.uk](#). Szczegółowe informacje na temat danych znajdują się pod adresem [Road Safety Open Dataset Data Guide](#). W nim też są definicje danych zawartych w plikach. Dane są zapisane w formacie csv, dla różnych lat. Informacje zostały podzielone na trzy różne pliki:

- Accident – informacje o wypadku
- Casualties – informacje o ofiarach wypadku
- Vehicles – informacje o pojeździe biorącym udział w wypadku

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

Zgromadzone dane posiadają informację o 1780653 wypadkach drogowych (ilość rekordów) między 01.01.2005- 31.12.2015.

Dane są poprawne, bardzo dobrej jakości. Brak problemu z brakującymi danymi bądź niepoprawnymi danymi.

Na stronie [kaggle](#) niestety nie wszystkie pliki z odniesieniami do danych są uwzględnione (np. weatherCondition), jednakże, nie ma problemu by je uzupełnić korzystając z [Road Safety Open Dataset Data Guide](#). Wystąpił problem przy ustawianiu referencji do tabel. W plikach z danymi na stronie [kaggle](#) występują dane o id -1, jednakże w danych z [Road Safety Open Dataset Data Guide](#), takie dane nie występują. Wymaga to interwencji zmienny danych klucza obcego np. z -1 na null, lub na dopisaniu do tabeli pozycji z id -1 o wartości np. „Unknown”.

Nie wszystkie dane zgromadzone w plikach nadają się do tworzenia hurtowni danych. Np. wartości takie jak Longitude czy Latitude zbyt szczegółowo określają położenie. Sprawia to, że wypadki, do których doszło w bliskiej odległości od siebie, będą uznane jako zdarzenia mająca miejsce w różnych lokalizacjach.

Wartości takie jak np, Carriageway_Hazards czy Special_Conditions_At_Site przyjmują w 99% jedną wartość. Oznacza to, że zmiana tych okoliczności nie wpływa na liczbę wypadków i można je wykluczyć już przy wstępnej analizie.

49 wierszy danych Casualties nie ma powiązania z żadnym wypadkiem. Dane te można usunąć.

Podobnie w przypadku Pojazdów. 63 dane nie są powiązane z żadnym wypadkiem.

Table z plików Casualties oraz Vehicles nie posiadają klucza identyfikującego wiersz (jedynie klucze obce do pliku Accdient), należy je utworzyć.

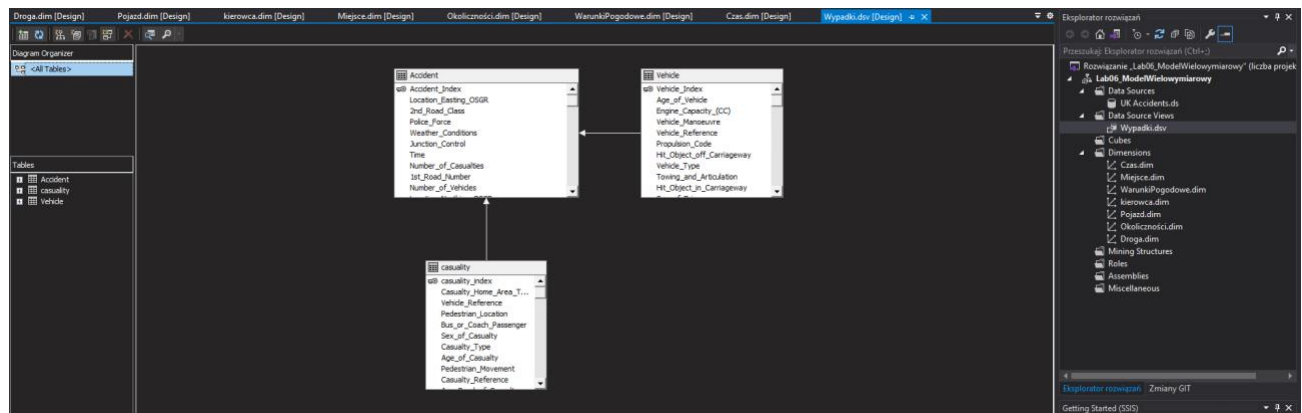
1.6.3. Fakty

Lp.	Fakt	Miary
1.	Wypadek	Liczba wypadków, liczba ofiar, wiek ofiar, rodzaj drogi, powaga wypadku, interwencja policji

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

Lp.	Kontekst analizy - wymiary	Własności
1.	Czas	Dzień, godzina, dzień tygodnia
2.	Miejsce	Lokalizacja, miasto, skrzyżowanie, teren miejski/wiejski
3.	Warunki pogodowe	Rodzaj panujących warunków
4.	Kierowca	Wiek, płeć, czas posiadania prawo jazdy
5.	Pojazd	Wiek, rodzaj pojazdu, rodzaj silnika
6.	Okoliczności	Wykonywany manewr, ograniczenie prędkości, ilość pasażerów
7.	Droga	Numer drogi, rodzaj

Prezentacja danych w modelu wielowymiarowym



Droga.dim [Design]Pojazd.dim [Design]kierowca.dim [Design] Miejsce.dim [Design]Okolicznosci.dim [Design]WarunkiPogodowe.dim [Design]Czas.dim

Dimension StructureAttribute RelationshipsTranslationsBrowser

Vehicle Index

→ Accident Index

→ Age Band Of Driver

→ Sex Of Driver

Attributes	Attribute Relationships
<div>Accident Index</div>	<div>Vehicle Index → Accident Index</div>
<div>Age Band Of Driver</div>	<div>Vehicle Index → Age Band Of Driver</div>
<div>Sex Of Driver</div>	<div>Vehicle Index → Sex Of Driver</div>
<div>Vehicle Index</div>	



Wrocław
University
of Science
and Technology

Road accidents and safety statistics

Politechnika Wrocławska
Hurtownie danych
Jakub Wrzesniak
254489
Wrocław, 12.04.2022r.

Źródło danych



GOV.UK

Dane

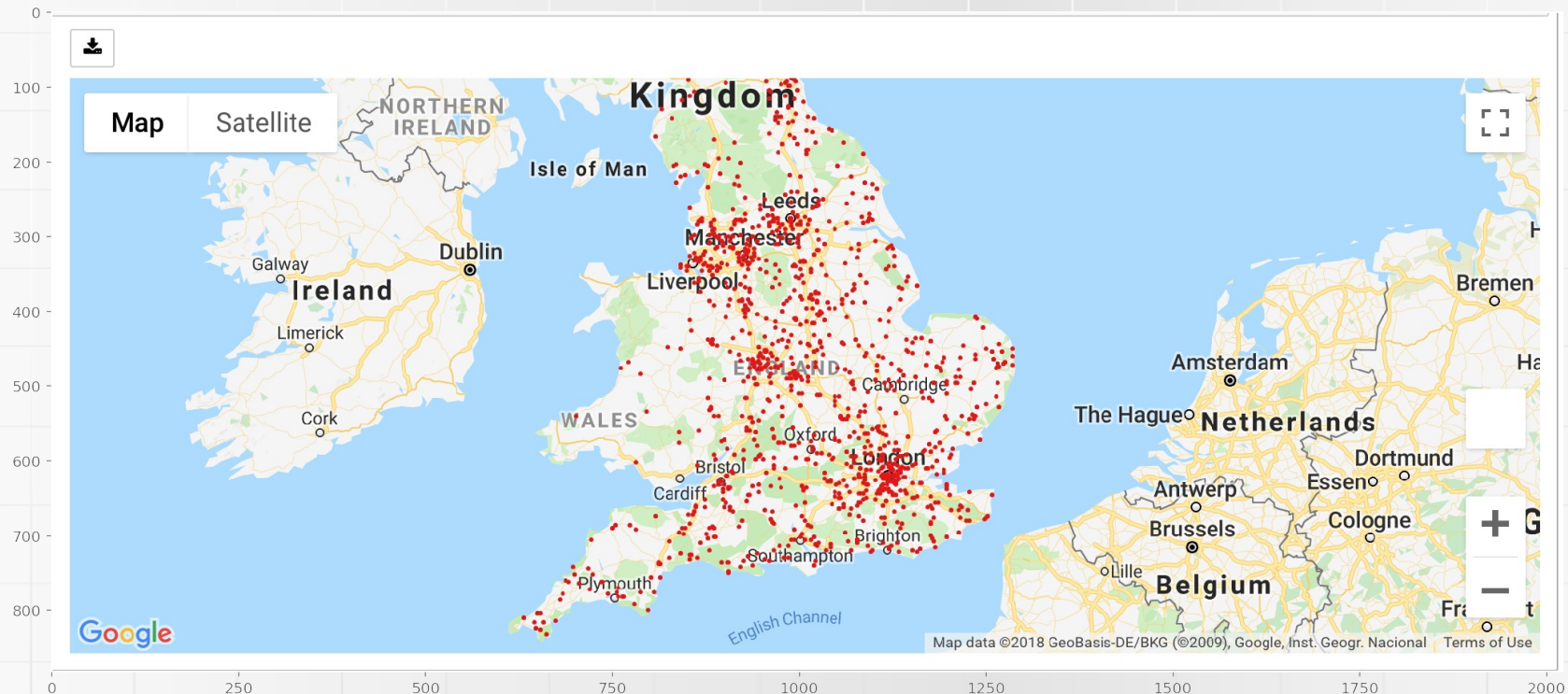
Dane wypadków w Wielkiej brytani, w latach 2005- 2015. Zwierają one ifnroamcje o:

- Czasie I miejscu zdarzenia
- Kierowcy pojazdu
- Rodzaju pojazdu
- Warunkach pogodowych
- Ofirach



Analiza

Zgromadzone dane pozwalają na analize wypadków i zbadanie czynników, które mogą się przyczyniać do ich występowania.



Wypadek

- Czas
- Miejsce
- Interwencja policji
- Powaga wypadku
- Liczba pojazdów
- Liczba ofiar
- Władze lokalne
- Numer drogi i jej rodzaj
- Ograniczenie prędkości
- Skrzyżowani
- Przejście dla pieszych
- Warunki pogodowe
- Stan drogi
- Teren miejski/wiejski

Pojazd

- Typ
- Wykonywany manewr
- Obecność na skrzyżowaniu
- Poślizk i przerwócenie pojazdu
- Wiek pojazdu
- Kod napędu
- Pojemność silnik
- Wiek oraz płeć kierowcy
- Cel podróży

Poszkodowany

- Rodzaj osoby poszkodowanej (kierowca, pasażer, ...)
- Płeć
- Wiek
- Poziom poniesionych obrażeń
- Obecność przejścia dla pieszych
- Rodzaj wypadku