**‘**

**Hurtownie danych – Projekt – Etap 02**

PWr. Data: 25-26.04.2022

|  |  |  |
| --- | --- | --- |
| Student | ---------------------------------------------------------- | Ocena |
| Indeks | 256305 |  |
| Imię | Grzegorz |
| Nazwisko | Dzikowski |

Zestaw składa się z 2 zadań. Pamiętaj o podaniu nr. indeksu oraz imienia i nazwiska.

# Wstępna specyfikacja wybranego tematu projektu

# Tytuł projektu

Analiza wypadków samochodowych w Wielkiej Brytanii w latach 2005 -2015

# Charakterystyka dziedziny problemowej

# Opis obszaru analizy (wybrany fragment dziedziny, przeznaczony do szczegółowej analizy i opracowania hurtowni danych)

Wypadki samochodowe corocznie powodują wiele śmierci oraz kalectw ludzi. Pomimo kampanii społecznych, kontroli drogowych oraz coraz nowszych aut, wypadki dalej występują na drogach.

Faktem, jakim będziemy się zajmować podczas analizy, to wypadek drogowy. Wymiarami są wiek pojazdu, wiek kierowcy, warunki pogodowych, warunki drogowe, jakość drogi, ograniczenie prędkości, liczby poszkodowanych oraz poważność wypadku. Wymiarami są data, czas, lokalizacja, departament policji, kierowca, przyczyna, skutek, przeprowadzona akcja. Miarą faktu jest liczba zatrzymań.

# Problemy

P01 – rosnąca liczba wypadków samochodowych

P02 – wzrost liczby ofiar śmiertelnych

P03 – niszczenie infrastruktury przez wypadki samochodowe

P04 – nieefektywność regulacji na ograniczenie liczby wypadków

# Cel przedsięwzięcia

# Oczekiwania i potrzeby w zakresie wsparcia podejmowania decyzji (pytania badawcze)

Analiza udostępni analizę faktów dotyczących wypadków i odpowiedzi na, między innymi, następujące pytania:

1. Jakie jest przekrój wiekowy oraz płciowy ofiar oraz kierowców?
2. Czy starsze auta są bezpieczniejsze?
3. Czy starsi wiekowo kierowcy jeżdżą bezpieczniej?
4. Czy limit prędkości ma wpływ na bezpieczeństwo na drogach?
5. Czy warunki na drodze mają wpływ na bezpieczeństwo?
6. Czy typ drogi ma wpływ na liczbę wypadków?

Właściwa analiza powinna odpowiedzieć na powyższe pytania

# Zakres analizy – badane aspekty

Analiza odbędzie się na wielu płaszczyznach. Będzie można dzięki temu podjąć działania ograniczające liczbę wypadków na wielu poziomach, tj. miejsce zdarzenia, warunku pogodowe, profil kierowcy czy typ pojazdu.

# Potencjalni użytkownicy projektowanej hurtowni danych

Baza analityczna będzie wspierać ministerstwo transportu w decyzjach dotyczących bezpieczeństwa ruch drogowego, oraz potencjalnych kupujących używane auta jako wyznacznik, czy kupowanie starego auta ma uzasadnienie pod kątem bezpieczeństwa

# Dane źródłowe

# Źródła danych

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Lp.** | **Plik, bazy danych** | **Typ** | **Liczba rek.** | **Rozmiar**  **[MB]** | **Opis** |
| 1. | Accidents | .csv | ~ 1 780 000 | 238 | Wszystkie wypadki drogowe w latach 2005-2015 w UK |
| 2. | Casualities | .csv | ~2 400 000 | 105 | Ofiary w wypadkach drogowych |
| 3 | Vehicles | .csv | ~3 200 000 | 201 | Pojazdy uczestniczące w wypadkach |
| 4. | Road-Safety-Open-Dataset-Data-Guide | .xlsx | 1580 | 0.55 | Objaśnienie danych w tabelach wyżej |

# Lokalizacja, dostępność danych źródłowych

Dane pochodzą z <https://www.kaggle.com/datasets/silicon99/dft-accident-data?resource=download>, które dla odmiany są zebrane z <https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data>. Tam też znajduje się słownik pojęć i jego interpretacja

# Słownik danych – interpretacja

Tabela Accidents

|  |  |  |  |
| --- | --- | --- | --- |
| **Lp.** | **Kolumna** | **Znaczenie** | **Uwagi** |
| 1. | [1st\_Road\_Class] | Typ drogi głównej, na której był wypadek.   1. Motorway 2. A(M) 3. A 4. B 5. C 6. Nieklasyfikowany |  |
| 2. | [1st\_Road\_Number] | Numer drogi, od 1 do 9999, -1 to nieznana, 0 – C lub nieklasyfikowana (bez numeru oficjalnego) |  |
| 3. | [2nd\_Road\_Class] | Droga, która znajduje się na skrzyżowaniu 20 metrów od wypadku i głównej drogi  0 – brak skrzyżowania  1 – Motorway  2 – A(M)  3 – A  4 – B  5 – C  6 – Nieklasyfikowane |  |
| 4. | [2nd\_Road\_Number] | Numer drogi na skrzyżowaniu 20 metrów od wypadku. Wartośc od 1 do 9999m, -1 to Nieznany, 0 to brak numeru |  |
| 5. | [Local\_Authority\_(District)] | Dystrykt lokalnych władz |  |
| 6. | [Local\_Authority\_(Highway)] | Pod jaką juryzdykcją jest autostrada |  |
| 7. | [Pedestrian\_Crossing-Human\_Control] | Czy jakiś człowiek kontrolował przejście dla pieszych do 50 metrów od wypadku?  0 – brak przejścia  1 – Szkolny patrol  2 – Inna autoryzowana osoba  -1 – Brak danych  9 – Nieznane |  |
| 8. | [Pedestrian\_Crossing-Physical\_Facilities] | Charakterystyka fizyczna przejścia dla pieszych na wypadku lub 50 metrów od niego  0 – Brak przejścia  1 – Zebra  4 – Przejście dla pieszych ze światłami bez skrzyżowania  5 – Przejście dla pieszych ze światłami na skrzyżowaniu |  |
| 9. | Accident\_Index | Unikalny identyfikator wypadku, klucz sztuczny. | Używany do łączenia z Casuality i Vehicle |
| 10. | Accident\_Severity | Poważność wypadku: 1 – śmiertelny, 2 – poważny, 3 - lekki |  |
| 11. | Carriageway\_Hazards | Dodatkowe zagrożenia na drodze  1 – Ładunek pojazdu  2 – Inne obiekty na drodze  3 – Poprzedni wypadek  4 – Pies na drodze  5 – Inne zwierzę na drodze  6 – Pieszy na drodze – nie ranny  7 – dowolne zwierzę na drodze |  |
| 12. | Date | Data wystąpienia incydentu w formacie DD/MM/YYYY |  |
| 13. | Day\_of\_Week | Dzień tygodnia wypadku, liczba od 1 (Niedziela) do 7 (Sobota) |  |
| 14. | Did\_Police\_Officer\_Attend\_Scene\_of\_Accident | Czy był policjant na miejscu wypadku?  1 – Tak  2 – Nie  3 – Nie, wypadek był raportowany przez samodzielny formularz  -1 – Brak danych |  |
| 15. | Junction\_Control | Rodzaj sterowania na skrzyżowaniu 20 metrów od wypadku  -1 – brak danych  0 – brak skrzyżowania  1 – autoryzowana osoba  2 – Automatyczne światłą  3 – Znak STOP  4 – Brak kontroli  9 – nieznane |  |
| 16. | Junction\_Detail | Typ skrzyżowania w pobliżu 20 metrów od wypadku  0 – brak skrzyżowania  1 – Rondo  2 - Małe rondo  3 – Skrzyżowanie typu T lub przesunięte skrzyżowanie  5 - Zjazd  6 – Przejazd kolejowy  7 - Więcej niż 4 odnogi (nie rondo)  8 – Prywatna droga  9 – Inny typ  99 – nieznany  -1 – brak danych |  |
| 17. | Latitude | Szerokość geograficzna, wartość z zakresu od -90 do 90 |  |
| 18. | Light\_Conditions | Warunki oświetlenia na drodze  1 – dzień  4 - Ciemność – światła uliczne włączone  5 – Ciemność – światła uliczne wyłączone  6 – Ciemność – bez lamp ulicznych  7 – Ciemność – brak danych  -1 brak danych lub nieznane |  |
| 19. | Location\_Easting\_OSGR | Lokalizacja wschód zachód na oficjalnej siatce lokalizacji w UK | gridreferencefinder.com |
| 20. | Location\_Northing\_OSGR | Lokalizacja północ południe na oficjalnej siatce lokalizacji w UK | gridreferencefinder.com |
| 21. | Longitude | Długość geograficzna, poprawne dane od -180 do 180 |  |
| 22. | LSOA\_of\_Accident\_Location | Tylko Anglia i Walia – Lower Layer Super Output Areas (LSOA) to geograficzna hierarchia stworzona w celu ulepszenia raportowania statystyk lokalnych w Anglii I Walii | [Źródło](https://www.datadictionary.nhs.uk/nhs_business_definitions/lower_layer_super_output_area.html) |
| 23. | Number\_of\_Casualties | Liczba ofiar wypadku |  |
| 24. | Number\_of\_Vehicles | Liczba pojazdów uczestniczących w wypadku |  |
| 25. | Police\_Force | Oddział policji zajmujący się tym wypadkiem, lista w Tabela 3 Pole Police Force |  |
| 26 | Road\_Surface\_Conditions | Warunki na drodze  1 – sucho  2 – mokro  3 – śnieg  4 – Lód lub przymrozek  5 – Powódź powyżej 3 cm  6 – Olej  7 – Błoto  9 – Nieznane |  |
| 27. | Road\_Type | Typ drogi  1. Rondo, 2. Jednokierunkowa, 3. Droga dwujezdniowa, 6. Droga jednojezdniowa, 7. Zjazd, 9. Nieznana, 12. Jednokierunkowy zjazd, -1 – brak danych |  |
| 28 | Special\_Conditions\_at\_Site | Specjalne warunki na miejscu wypadku  0 – brak  1 – Nie działające światłą  2 – Część świateł nie działająca  3 – Uszkodzone znaki drogowe  4 – Prace drogowe  5 – Uszkodzona droga  6 – Olej  7 – Błoto  -1 – Brak danych  9 – nieznane |  |
| 29. | Speed\_limit | Ograniczenie prędkości na drodze. Poprawne dane to 20, 30, 40, 50, 60, 70, -1 (brak danych), 99 (nieznane) |  |
| 30. | Time | Godzina wypadku, z dokładnością do minut, format HH:MM |  |
| 31. | Urban\_or\_Rural\_Area | Czy teren miejscy czy wiejski?  1 – Miejski  2 – Wiejski  3 – Nieokreślony  -1 – Brak danych |  |
| 32. | Weather\_Conditions | Warunki pogodowe na drodze  1 – dobra pogoda, brak wiatru  2 – deszcz, brak wiatru  3 – śnieg, brak wiatru  4 – dobra, silne wiatry  5 – deszcz, silne wiatry  6 – śnieg, silne wiatry  7 – mgła lub mżawka  8 – inne  9 – nieznane  -1 – Brak danych |  |

Tabela Casulaities

|  |  |  |  |
| --- | --- | --- | --- |
| Lp. | Kolumna | Znaczenie | Uwagi |
| 1. | Accident\_Index | Unikalny identyfikator wypadku, klucz sztuczny. | Używany do łączenia z Accident i Vehicle |
| 2. | Age\_Band\_of\_Casualty | Przedział wiekowy ofiary, pokazany w tabeli Tabela 4 Przedział wiekowy |  |
| 3. | Age\_of\_Casualty | Wiek ofiary, -1 jak brak danych |  |
| 4. | Bus\_or\_Coach\_Passenger | Czy to osoba będąca w autobusie?  0 – Nie w autobusie  1 – wchodząca do autobusu  2 – wychodząca z autobusu  3 – pasażer stojący  4 – pasażer siedzący  9 – nieznany  -1 – brak danych |  |
| 5. | Car\_Passenger | Typ pasażera w aucie  0 – Nie w aucie  1 – pasażer z przodu auta  2 – pasażer z tyłu auta  9 – nieznany  -1 – Brak danych |  |
| 6. | Casualty\_Class | Typ ofiary, 1 – kierowca, 2 – pasażer, 3 – pieszy |  |
| 7. | Casualty\_Home\_Area\_Type | Pochodzenie ofiary, 1 – miejskie, 2 – małe miasto, 3 – wioska, -1 – brak danych |  |
| 8. | Casualty\_Reference | Unikalne ID ofiary w ramach wypadku, klucz obcy |  |
| 9. | Casualty\_Severity | Poważność obrażeń ofiary. 1 – śmiertelnie, 2 – poważnie, 3 – lekkie |  |
| 10. | Casualty\_Type | Rodzaj ofiary, opisany w tabeli Tabela 5 Klasa ofiary/pojazdu |  |
| 11. | Pedestrian\_Location | Lokalizacja pieszego w momencie wypadku, wartości w tabeli Tabela 6 Lokalizacja pieszego |  |
| 12. | Pedestrian\_Movement | Jak poruszał się pieszy? Wartości w tabeli Tabela 7 Sposób poruszania się pieszego |  |
| 13. | Pedestrian\_Road\_Maintenance\_Worker | Czy pieszy był pracownikiem budowy? 0 – Nie, 1 – Tak, 2 – nie wiadomo, 3 – prawdopodobnie, - 1 – brak danych |  |
| 14. | Sex\_of\_Casualty | Płeć kierowcy, Mężczyzna, kobieta, nieznany lub brak danych |  |
| 15. | Vehicle\_Reference | Unikalne ID każdego pojazdu w ramach wypadku. Klucz sztuczny | Umożliwia połączenie ofiary z pojazdem |

Tabela Vehicles

|  |  |  |  |
| --- | --- | --- | --- |
| **Lp.** | **Kolumna** | **Znaczenie** | **Uwagi** |
| 1 | [1st\_Point\_of\_Impact] | Miejsce pierwszego uderzenia, wartości w tabeli Tabela 8 Miejsce pierwszego kontaktu |  |
| 2 | [Engine\_Capacity\_(CC)] | Pojemność silnika w CC |  |
| 3 | [Vehicle\_Location-Restricted\_Lane] | Typ pasa awaryjnego, na jakim znajduje się pojazd, |  |
| 4 | [Was\_Vehicle\_Left\_Hand\_Drive?] | Czy pojazd miał kierownicę po lewej stronie? Tak , nie, brak danych |  |
| 5 | Accident\_Index | Unikalny identyfikator wypadku, klucz sztuczny. | Używany do łączenia z Accident i Vehicle |
| 6 | Age\_Band\_of\_Driver | Przedział wiekowy kierowcy, pokazany w tabeli Tabela 4 Przedział wiekowy |  |
| 7 | Age\_of\_Driver | Wiek kierowcy, -1 jak brak danych |  |
| 8 | Age\_of\_Vehicle | Wiek auta, -1 jak brak danych |  |
| 9 | Driver\_Home\_Area\_Type | Pochodzenie kierowcy, 1 – miejskie, 2 – małe miasto, 3 – wioska, -1 – brak danych |  |
| 10 | Driver\_IMD\_Decile | Wskaźnik IMD kierowcy, wskazujący na poziom miejsca, z którego pochodzi kierowca | [Źródło](https://en.wikipedia.org/wiki/Multiple_deprivation_index) |
| 11 | Hit\_Object\_in\_Carriageway | W jaki obiekt uderzył pojazd na drodze? |  |
| 12 | Hit\_Object\_off\_Carriageway | W jaki obiekt uderzył pojazd poza drogą |  |
| 13 | Journey\_Purpose\_of\_Driver | Cel podróży |  |
| 14 | Junction\_Location | Pozycja na skrzyżowaniu |  |
| 15 | Propulsion\_Code | Rodzaj napędu, wartości pod Tabela 10 Typ paliwa |  |
| 16 | Sex\_of\_Driver | Płeć kierowcy, Mężczyzna, kobieta, nieznany lub brak danych |  |
| 17 | Skidding\_and\_Overturning | Czy auto wpadło w poślizg lub wywróciło się?   |  |  | | --- | --- | | 0 | None | | 1 | Poślizg | | 2 | Wpadło w poślizg I wywróciło się | | 3 | Skręciło w niekontrolowany sposób | | 4 | Skręciło w niekontrolowany sposób i się wywróciło | | 5 | Przewróciło się | | 9 | Nieznane | | -1 | Brak danych | |  |
| 18 | Towing\_and\_Articulation | Czy posiadał Przyczepę?   |  |  | | --- | --- | | -1 | Data missing or out of range | | 0 | Brak przyczepy | | 1 | Pojazd przegubowy | | 2 | Powójna lub wielokrotna przyczepa | | 3 | Karawana | | 4 | Pojedyncza przyczepa | | 5 | Inna przyczepa | | 9 | Nieznany | |  |
| 19 | Vehicle\_Leaving\_Carriageway | Czy pojazd opuścił jezdnię i w jaki sposób?   |  |  | | --- | --- | | 0 | Nie opuścił jezdnii | | 1 | Krawędź jezdni | | 2 | Odbił się od krawędzi | | 3 | Prosto na skrzyżowaniu | | 4 | Od strony kierowcy w centrum | | 5 | Od strony kierowcy w centrum + odbity | | 6 | Od strony kierowcy w centrum, przekroczył środek jezdni | | 7 | Od strony kierowcy | | 8 | Od strony kierowcy, odbity | | 9 | Nieznane | | -1 | Brak danych | |  |
| 20 | Vehicle\_Manoeuvre | Co to był za typ manewru? |  |
| 21 | Vehicle\_Reference | Unikalny numer pojazdu w ramach wypadku, pozwala na powiązanie ofiary z pojazdem |  |
| 22 | Vehicle\_Type | Typ pojazdu, wartości w tabeli Tabela 5 Klasa ofiary/pojazdu |  |

# Ocena jakościowa danych

Analiza liczby poszkodowanych oraz poważność wypadków samochodowych w Wielkiej Brytanii w latach 2005 -2015 pod kątem wieku pojazdu, wieku kierowcy, warunków pogodowych, jakości drogi oraz ograniczenia prędkości

**Dane o wysokiej jakości**

**Dane o niskiej jakości**

**Dane nieistotne w analizie**

**Tabela Accidents**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lp.** | **Kolumna** | **Typ** | **Zakres** | **Ocena jakości** |
| 1. | [1st\_Road\_Class] | Numeryczny, Całkowity | 1 do 6 | 0% null |
| 2. | [1st\_Road\_Number] | Numeryczny, Całkowity | -1 do 9999 | 0% null |
| 3. | [2nd\_Road\_Class] | Numeryczny, Całkowity | -1 do 6 | 0% null, -1 ma 41% wartości. Wartość przydatna do analizy jakości dróg, ale niestety niskiej jakości |
| 4. | [2nd\_Road\_Number] | Numeryczny, Całkowity | -1 do 9999 | 0% null, |
| 5. | [Local\_Authority\_(District)] | Numeryczny, Całkowity | 1 do 941 | 0% null |
| 6. | [Local\_Authority\_(Highway)] | Tekstowy | 9 znaków | 0% null |
| 7. | [Pedestrian\_Crossing-Human\_Control] | Numeryczny, Całkowity | -1 do 2 | 0% null |
| 8. | [Pedestrian\_Crossing-Physical\_Facilities] | Numeryczny, Całkowity | -1 do 8 | 0% null |
| 9. | Accident\_Index | Tekstowy | 13 znaków | 0% null, 100% Key Strength |
| 10. | Accident\_Severity | Numeryczny, Całkowity | 1 do 3 | 0% null |
| 11. | Carriageway\_Hazards | Numeryczny, Całkowity | -1 do 7 | 0% null |
| 12. | Date | Data | 01.01.2005 do 31.12.2015 | 0% null |
| 13. | Day\_of\_Week | Numeryczny, Całkowity | 1 do 7 | 0% null |
| 14. | Did\_Police\_Officer\_Attend\_Scene\_of\_Accident | Numeryczny, Całkowity | -1 do 3 | 0% null |
| 15. | Junction\_Control | Numeryczny, Całkowity | -1 do 4 | 0% null, -1 ma 36% wartości |
| 16. | Junction\_Detail | Numeryczny, Całkowity | -1 do 9 | 0% null |
| 17. | Latitude | Numeryczny, Zmiennoprzecinkowy | 49.912941 do 60.757544 | < 1% null |
| 18. | Light\_Conditions | Numeryczny, Całkowity | 1 do 7 | 0% null |
| 19. | Location\_Easting\_OSGR | Numeryczny, Całkowity | 64950 do 655540 | < 1% null |
| 20. | Location\_Northing\_OSGR | Numeryczny, Całkowity | 10290 do 128800 | < 1% null |
| 21. | Longitude | Numeryczny, Całkowity | -7.516225 do 1.76201 | < 1% null |
| 22. | LSOA\_of\_Accident\_Location | Tekstowy | 9 znaków | 7% null |
| 23. | Number\_of\_Casualties | Numeryczny, Całkowity | 1 do 93 | 0% null |
| 24. | Number\_of\_Vehicles | Numeryczny, Całkowity | 1 do 67 | 0% null |
| 25. | Police\_Force | Numeryczny, Całkowity | 1 do 98 | 0% null |
| 26 | Road\_Surface\_Conditions | Numeryczny, Całkowity | -1 do 5 | 0% null |
| 27. | Road\_Type | Numeryczny, Całkowity | 1 do 9 | 0% null |
| 28 | Special\_Conditions\_at\_Site | Numeryczny, Całkowity | -1 do 7 | 0% null |
| 29. | Speed\_limit | Numeryczny, Całkowity | 0 do 70 | 0% null |
| 30. | Time | Godzina | Od 00:01:00 do 23:59:00 (1 minutowa ziarnistość) | < 1% null |
| 31. | Urban\_or\_Rural\_Area | Numeryczny, Całkowity | 1 do 3 | 0% null |
| 32. | Weather\_Conditions | Numeryczny, Całkowity | -1 do 9 | 0% null |

**Tabela Casualities**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lp.** | **Kolumna** | **Typ** | **Zakres** | **Ocena jakości** |
| 1. | Accident\_Index | Tekstowy | Tekst o długości 13 znaków, ale 49 wpisów ma długość 1 | 0.002% null |
| 2. | Age\_Band\_of\_Casualty | Numeryczny, Całkowity | -1 do 11 | 0.002% null, 2% ma wartość -1 |
| 3. | Age\_of\_Casualty | Numeryczny, Całkowity | -1 do 104 | 0.002% null, 2% ma wartość -1 |
| 4. | Bus\_or\_Coach\_Passenger | Numeryczny, Całkowity | -1 do 4 | 0.002% null |
| 5. | Car\_Passenger | Numeryczny, Całkowity | -1 do 2 | 0.002% null |
| 6. | Casualty\_Class | Numeryczny, Całkowity | 1 do 3 | 0.002% null |
| 7. | Casualty\_Home\_Area\_Type | Numeryczny, Całkowity | -1 do 3 | 0.002% null, 14% wartości ma -1 |
| 8. | Casualty\_Reference | Numeryczny, Całkowity | 1 do 852 | 0.002% null, 59% wartości ma -1 |
| 9. | Casualty\_Severity | Numeryczny, Całkowity | 1 do 3 | 0.002% null |
| 10. | Casualty\_Type | Numeryczny, Całkowity | 0 do 98 | 0.002% null |
| 11. | Pedestrian\_Location | Numeryczny, Całkowity | -1 do 10 | 0.002% null |
| 12. | Pedestrian\_Movement | Numeryczny, Całkowity | -1 do 9 | 0.002% null |
| 13. | Pedestrian\_Road\_Maintenance\_Worker | Numeryczny, Całkowity | -1 do 2 | 0.002% null |
| 14. | Sex\_of\_Casualty | Numeryczny, Całkowity | -1 do 2 | 0.002% null |
| 15. | Vehicle\_Reference | Numeryczny, Całkowity | 1 do 91 | 0.002% null |

**Tabela Vehicles**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lp.** | **Kolumna** | **Typ** | **Zakres** | **Ocena jakości** |
| 1. | [1st\_Point\_of\_Impact] | Numeryczny, Całkowity | -1 do 4 | 0.0019% null |
| 2. | [Engine\_Capacity\_(CC)] | Numeryczny, Całkowity | -1 do 99999 | 0.0019% null |
| 3. | [Vehicle\_Location-Restricted\_Lane] | Numeryczny, Całkowity | -1 do 9 | 0.0019% null |
| 4. | [Was\_Vehicle\_Left\_Hand\_Drive?] | Prawda/Fałsz, Nieznany (-1), Null | -1 , 1, 2 | 0.0019% null, -1 ma 0.5% wartości |
| 5. | Accident\_Index | Tekstowy | 2 do 13 znaków, 63 wartości -1, klucz sztuczny | 0 % null |
| 6. | Age\_Band\_of\_Driver | Numeryczny, Całkowity | -1 do 11 | 0.0019% null, 11% ma wartość -1 |
| 7. | Age\_of\_Driver | Numeryczny, Całkowity | -1 do 100 | 0.0019% null |
| 8. | Age\_of\_Vehicle | Numeryczny, Całkowity | -1 do 111 | 0.0019% null, 30% ma wartość -1 |
| 9. | Driver\_Home\_Area\_Type | Numeryczny, Całkowity | -1 do 3 | 0.0019% null, 20% ma wartość -1 |
| 10. | Driver\_IMD\_Decile | Numeryczny, Całkowity | -1 do 10 | 0.0019% null, 33% ma wartość -1 |
| 11. | Hit\_Object\_in\_Carriageway | Numeryczny, Całkowity | -1 do 12 | 0.0019% null, |
| 12. | Hit\_Object\_off\_Carriageway | Numeryczny, Całkowity | -1 do 11 | 0.0019% null |
| 13. | Journey\_Purpose\_of\_Driver | Numeryczny, Całkowity | -1 do 15 | 0.0019% null, 1% ma wartość -1 |
| 14. | Junction\_Location | Numeryczny, Całkowity | -1 do 8 | 0.0019% null |
| 15. | Propulsion\_Code | Numeryczny, Całkowity | -1 do 12 | 0.0019% null, 26% ma wartość -1 |
| 16. | Sex\_of\_Driver | Numeryczny, Całkowity | -1 do 3 | 0.0019% null |
| 17. | Skidding\_and\_Overturning | Numeryczny, Całkowity | -1 do 5 | 0.0019% null |
| 18. | Towing\_and\_Articulation | Numeryczny, Całkowity | -1 do 5 | 0.0019% null |
| 19. | Vehicle\_Leaving\_Carriageway | Numeryczny, Całkowity | -1 do 8 | 0.0019% null |
| 20. | Vehicle\_Manoeuvre | Numeryczny, Całkowity | -1 do 9 | 0.0019% null |
| 21. | Vehicle\_Reference | Numeryczny, Całkowity | 1 do 91 | 0.0019% null |
| 22. | Vehicle\_Type | Numeryczny, Całkowity | -1 do 98 | 0.0019% null |

# Analityczne modele wielowymiarowe

# Fakty podlegające analizie oraz ich miary

Tabela 1. Fakty oraz ich miary opracowywanych modeli analitycznych

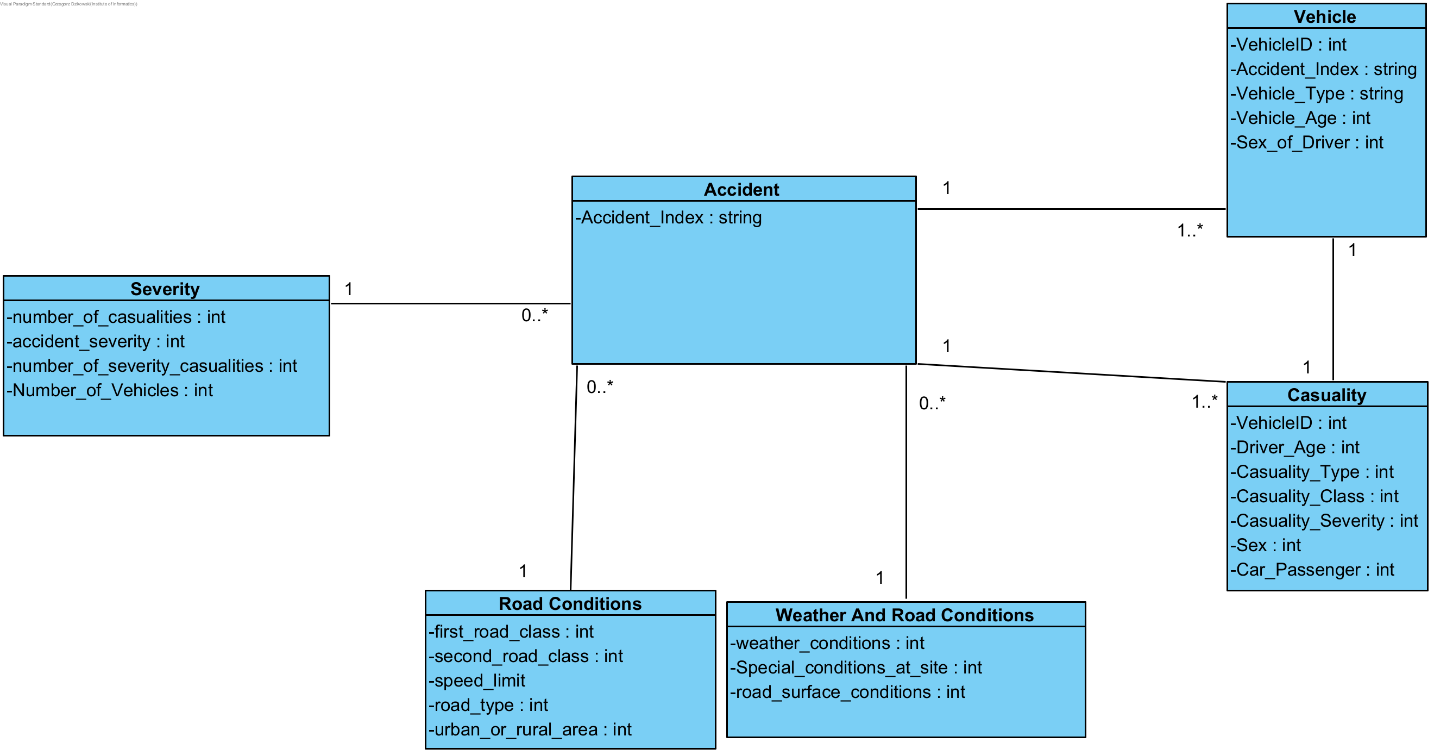
|  |  |  |
| --- | --- | --- |
| **Lp.** | **Fakt** | **Miary** |
| **1.** | **Wypadek drogowy** | **Wieku pojazdu, wiek kierowcy, warunki pogodowych, warunki drogowe, jakość drogi, ograniczenie prędkości, liczby poszkodowanych, poważność** |

# Kontekst analizy faktów

Tabela 2. Zidentyfikowane wymiary wraz z ich własnościami (charakterystykami) opracowywanych modeli analitycznych

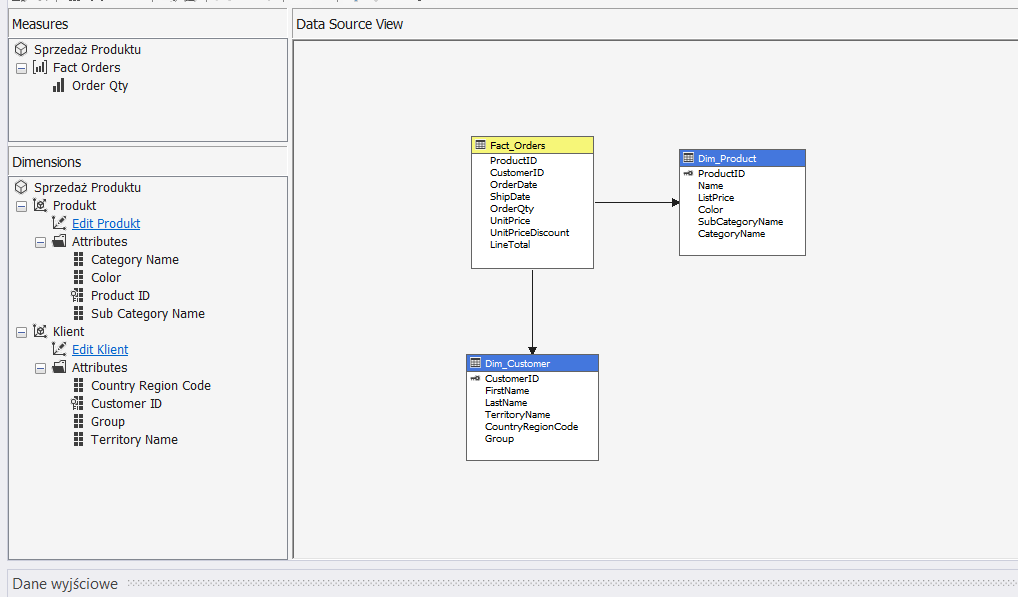
|  |  |  |
| --- | --- | --- |
| Lp. | Kontekst analizy - wymiary | Własności |
| 1. | Kierowca | Zbiór wartości dotyczących kierowcy. Zawiera informacje o wieku kierowcy, jego pochodzeniu itp. |
| 2. | **Pojazd** | Zawiera informacje o pojeździe, między innymi moc silnika,= lub wiek pojazdu |
| 3. | **Warunki pogodowych i drogowe** | Wymiar określa pogodą w momencie wypadku, oraz warunki na drodze, między innymi czy było ciemno |
| 4. | **Jakość drogi** | Typ drogi. Połączenie kilku wartości, np. czy droga jednokierunkowa, czy autostrada, czy polna, oraz ograniczenia drogowe na |
| 5. | **Poważność** | Każdy wypadek ma listę poszkodowanych, ich liczbę, oraz poważność każdego przypadku. Ten wymiar określa, jak poważny był wypadek |

# Modele wielowymiarowe (UML)

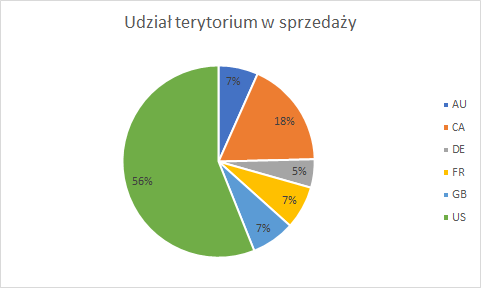


# Implementacja analitycznego modelu danych na podstawie bazy danych ,która została stworzona w sprawozdaniu 3.

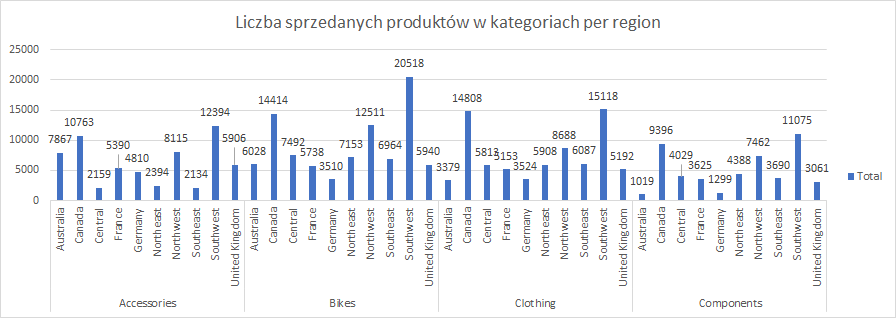
1. Analityczny model danych (kopia kostki)



1. Przykłady analizy danych wraz z dwoma, trzema wykresami (Excel)



Powyższy wykres pokazuje udział poszczególnych krajów w sprzedaży. Widać na nim, że głównym klientem firmy jest US, gdzie sprzedawane jest 56% produktów. Europa pokrywa sumarycznie tylko 26% udziału sprzedaży



Ten wykres pokazuje, ile produktów danej kategorii zostały sprzedane w poszczególnych regionach. Widać na nim, że najchętniej kupuje Southwest w każdej kategorii, oprócz ubrania, w której dominują kanadyjczycy

Obraz zawierający stół

Opis wygenerowany automatycznie

Wykres pokazuje, ile produktów danego koloru zostało sprzedanych w poszczególnych, NA nim widać, że USA, Kanada i Australia lubi kolory czerwone, natomiast Francja, Niemcy i Austria lubi bardziej kolor niebieski

**Uwaga:** - każdy diagram musi być uzupełniony ciekawymi wnioskami wynikającym z prezentowanych danych

# Wnioski z realizacji zadań

Analiza danych i ocena jakości to ciężkie zadanie. Nawet mając dostęp do konkretnych opisów najbardziej czasochłonna była ocena jakości i słownik. Dodatkowo, znalezienie pytań badawczych i poznanie tematu jest ciężkie dla osoby nieinteresującej się tematem

Analiza wielowymiarowa i kostka bardzo ułatwia analizę i pobieranie danych z bazy do analizy wielowymiarowej.

*Uwaga:*

* Sprawozdanie bez wniosków końcowych nie będzie sprawdzane i tym samym ocena jest negatywna!

Załączniki

|  |  |
| --- | --- |
| ID | Police\_Force |
| 1 | Metropolitan Police |
| 3 | Cumbria |
| 4 | Lancashire |
| 5 | Merseyside |
| 6 | Greater Manchester |
| 7 | Cheshire |
| 10 | Northumbria |
| 11 | Durham |
| 12 | North Yorkshire |
| 13 | West Yorkshire |
| 14 | South Yorkshire |
| 16 | Humberside |
| 17 | Cleveland |
| 20 | West Midlands |
| 21 | Staffordshire |
| 22 | West Mercia |
| 23 | Warwickshire |
| 30 | Derbyshire |
| 31 | Nottinghamshire |
| 32 | Lincolnshire |
| 33 | Leicestershire |
| 34 | Northamptonshire |
| 35 | Cambridgeshire |
| 36 | Norfolk |
| 37 | Suffolk |
| 40 | Bedfordshire |
| 41 | Hertfordshire |
| 42 | Essex |
| 43 | Thames Valley |
| 44 | Hampshire |
| 45 | Surrey |
| 46 | Kent |
| 47 | Sussex |
| 48 | City of London |
| 50 | Devon and Cornwall |
| 52 | Avon and Somerset |
| 53 | Gloucestershire |
| 54 | Wiltshire |
| 55 | Dorset |
| 60 | North Wales |
| 61 | Gwent |
| 62 | South Wales |
| 63 | Dyfed-Powys |
| 91 | Northern |
| 92 | Grampian |
| 93 | Tayside |
| 94 | Fife |
| 95 | Lothian and Borders |
| 96 | Central |
| 97 | Strathclyde |
| 98 | Dumfries and Galloway |
| 99 | Police Scotland |

Tabela 3 Pole Police Force

|  |  |
| --- | --- |
| ID | Przedział wiekowy |
| 1 | 0 - 5 |
| 2 | 6 - 10 |
| 3 | 11 - 15 |
| 4 | 16 - 20 |
| 5 | 21 - 25 |
| 6 | 26 - 35 |
| 7 | 36 - 45 |
| 8 | 46 - 55 |
| 9 | 56 - 65 |
| 10 | 66 - 75 |
| 11 | +75 |
| -1 | Brak danych |

Tabela 4 Przedział wiekowy

|  |  |
| --- | --- |
| ID | Typ pojazdu/ofiary |
| 0 | Pieszy |
| 1 | Rowerzysta |
| 2 | Motor poniżej 50cc |
| 3 | Motor poniżej 125cc |
| 4 | Motor poniżej 500 cc |
| 5 | Motor powyżej 500 cc |
| 8 | Taksówka lub wypożyczone auto |
| 9 | Auto |
| 10 | Minibus (8-16 osób) |
| 11 | Bus (+16 osób) |
| 16 | Koń |
| 17 | Pojazd rolniczy |
| 18 | Trwamwaj |
| 19 | Van poniżej 3.5t |
| 20 | Ciężarówka poniżej 7.5 tony |
| 21 | Ciężarówka powyżej 7.5 tony |
| 22 | Skuter mobilny |
| 23 | Elektryczny rower |
| 90 | Inne |
| 97 | Motocykl o nieznanej mocy |
| 98 | Ciężarówka o nieznanej wadze |
| 99 | Nieznany pojazd |

Tabela 5 Klasa ofiary/pojazdu

|  |  |
| --- | --- |
| ID | Lokalizacja pieszego |
| 0 | Not a Pedestrian |
| 1 | Crossing on pedestrian crossing facility |
| 2 | Crossing in zig-zag approach lines |
| 3 | Crossing in zig-zag exit lines |
| 4 | Crossing elsewhere within 50m. of pedestrian crossing |
| 5 | In carriageway, crossing elsewhere |
| 6 | On footway or verge |
| 7 | On refuge, central island or central reservation |
| 8 | In centre of carriageway - not on refuge, island or central reservation |
| 9 | In carriageway, not crossing |
| 10 | Unknown or other |
| -1 | Data missing or out of range |

Tabela 6 Lokalizacja pieszego

|  |  |
| --- | --- |
| 0 | Not a Pedestrian |
| 1 | Crossing from driver's nearside |
| 2 | Crossing from nearside - masked by parked or stationary vehicle |
| 3 | Crossing from driver's offside |
| 4 | Crossing from offside - masked by parked or stationary vehicle |
| 5 | In carriageway, stationary - not crossing (standing or playing) |
| 6 | In carriageway, stationary - not crossing (standing or playing) - masked by parked or stationary vehicle |
| 7 | Walking along in carriageway, facing traffic |
| 8 | Walking along in carriageway, back to traffic |
| 9 | Unknown or other |
| -1 | Data missing or out of range |

Tabela 7 Sposób poruszania się pieszego

|  |  |
| --- | --- |
| ID | Miejsce pierwszego kontaktu |
| 0 | Did not impact |
| 1 | Front |
| 2 | Back |
| 3 | Offside |
| 4 | Nearside |
| 9 | unknown (self reported) |

Tabela 8 Miejsce pierwszego kontaktu

|  |  |
| --- | --- |
| 0 | On main c'way - not in restricted lane |
| 1 | Tram/Light rail track |
| 2 | Bus lane |
| 3 | Busway (including guided busway) |
| 4 | Cycle lane (on main carriageway) |
| 5 | Cycleway or shared use footway (not part of main carriageway) |
| 6 | On lay-by or hard shoulder |
| 7 | Entering lay-by or hard shoulder |
| 8 | Leaving lay-by or hard shoulder |
| 9 | Footway (pavement) |
| 10 | Not on carriageway |
| 99 | unknown (self reported) |
| -1 | Data missing or out of range |

Tabela 9 Typ pasa awaryjnego, na którym znajdowało się auto

|  |  |
| --- | --- |
| -1 | Data missing or out of range |
| 1 | Petrol |
| 2 | Heavy oil |
| 3 | Electric |
| 4 | Steam |
| 5 | Gas |
| 6 | Petrol/Gas (LPG) |
| 7 | Gas/Bi-fuel |
| 8 | Hybrid electric |
| 9 | Gas Diesel |
| 10 | New fuel technology |
| 11 | Fuel cells |
| 12 | Electric diesel |
| -1 | Undefined |

Tabela 10 Typ paliwa

|  |  |
| --- | --- |
| -1 | Data missing or out of range |
| 1 | Journey as part of work |
| 2 | Commuting to/from work |
| 3 | Taking pupil to/from school |
| 4 | Pupil riding to/from school |
| 5 | Other |
| 6 | Not known |
| 15 | Other/Not known |
| -1 | Data missing or out of range |

Tabela 11 Cel podróży

|  |  |
| --- | --- |
| 1 | Reversing |
| 2 | Parked |
| 3 | Waiting to go - held up |
| 4 | Slowing or stopping |
| 5 | Moving off |
| 6 | U-turn |
| 7 | Turning left |
| 8 | Waiting to turn left |
| 9 | Turning right |
| 10 | Waiting to turn right |
| 11 | Changing lane to left |
| 12 | Changing lane to right |
| 13 | Overtaking moving vehicle - offside |
| 14 | Overtaking static vehicle - offside |
| 15 | Overtaking - nearside |
| 16 | Going ahead left-hand bend |
| 17 | Going ahead right-hand bend |
| 18 | Going ahead other |
| 99 | unknown (self reported) |
| -1 | Data missing or out of range |

Tabela 12 Typ manewru