

Hurtownie danych – Projekt HD

PWr. Wydział Informatyki i Telekomunikacji Data: **15.06.2023**

Student	-----	Ocena
Indeks	<u>260330</u>	
Imię	<u>Kamil</u>	
Nazwisko	<u>Krawiec</u>	

Co zostało zmienione	Kiedy	Przez kogo	Uwagi
Pkt 4	13.06.23	Kamil Krawiec	Dodano atrybuty do tabeli 6. Zmieniono rys 1. Dodano atrybuty do modelu ww.

1. Tytuł projektu

Wypadki rowerowe w Wielkiej Brytani na przestrzeni lat 1979-2018

2. Charakterystyka dziedziny problemowej

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

Hurtownia danych dotycząca wypadków rowerowych będzie się składać z 4 wymiarów i jednego faktu, które umożliwi szczegółową analizę bezpieczeństwa ruchu rowerowego i czynników na to wpływających.

Fakt wypadek będzie analizowany pod kątem warunków zewnętrznych takich jak pogoda i stan nawierzchni, kolejnym ciekawym kontekstem będzie droga na której doszło do zdarzenia, jej stan, typ jak i ograniczenie prędkości na niej występujące. Analiza w kontekście czasu pozwoli określić cykliczności oraz okresy większego niebezpieczeństwa na drogach. Do tego dochodzi obraz poszkodowanych, czyli w jakim byli wieku oraz jakie obrażenia doznali.

Pozwoli to na zwrócenie uwagi w obszary które wymagają poprawy lub baczniejszej obserwacji co może skutkować większym bezpieczeństwem dla rowerzystów jak i kierujących pojazdem.

2.2 Problemy

Warunki atmosferyczne: Pogoda ma istotny wpływ na bezpieczeństwo ruchu rowerowego, a analiza danych może pomóc ustalić, jakie warunki atmosferyczne (np. deszcz, śnieg, mgła) są najbardziej ryzykowne dla rowerzystów.

Stan nawierzchni: Analiza danych może pomóc w określeniu, jakie czynniki związane z nawierzchnią (np. nierówności, błoto, śliska powierzchnia) przyczyniają się do wypadków rowerowych.

Charakterystyka dróg: Badanie danych może umożliwić identyfikację konkretnych typów dróg, na których często dochodzi do wypadków rowerowych. Możemy również zbadać, jak ograniczenia prędkości na drogach wpływają na bezpieczeństwo rowerzystów.

Analiza czasu: Analiza danych w kontekście czasu może pomóc w identyfikacji okresów, w których występuje większe ryzyko wypadków rowerowych.

Profil poszkodowanych: Badanie danych pozwoli na zrozumienie charakterystyki poszkodowanych w wypadkach rowerowych, takich jak ich wiek i rodzaj odniesionych obrażeń. To może pomóc w lepszym ukierunkowaniu działań na rzecz poprawy bezpieczeństwa rowerzystów.

2.3 Cel przedsięwzięcia

2.3.1 Oczekiwania i potrzeby w zakresie wsparcia podejmowania decyzji

Pytania badawcze

1. Jak wygląda ogólna sytuacja z wypadkami rowerowymi w Wielkiej Brytanii?
2. Czy występują określone cykle lub okresy, w których dochodzi do większej liczby wypadków rowerowych?
3. Czy można zidentyfikować czynniki sezonowe lub godzinowe, które mają wpływ na bezpieczeństwo rowerzystów?
4. Kiedy jest najbezpieczniejszy czas na jazdę rowerem?
5. Czy można zauważyć pewne grupy wiekowe, które są bardziej narażone na ryzyko wypadków?

6. Jakie osoby najczęściej odnoszą obrażenia i kto jest najbardziej narażony jeśli chodzi o wypadki rowerowe?
7. Jaki jest najważniejszy czynnik wpływający na powagę wypadku?

2.3.2 Zakres analizy

Badane aspekty

Badanymi aspektami będą wypadki rowerowe z udziałem aut w kontekście: warunków pogodowych, nawierzchniowych oraz w perspektywie czasowej.

Analiza będzie zawierała w sobie:

1. - Częstość wypadków rowerowych
2. - Czynniki ryzyka
3. - Powaga wypadków
4. - Wzorce czasowe
5. - Korelacje między zmiennymi

2.3.3 Potencjalni użytkownicy

Społeczność rowerowa: Organizacje i grupy reprezentujące społeczność rowerową mogą wykorzystać dane z hurtowni do identyfikacji problematycznych obszarów, podejmowania działań lobbingowych i prowadzenia kampanii na rzecz poprawy bezpieczeństwa rowerzystów.

Projektanci infrastruktury: Osoby odpowiedzialne za projektowanie i modernizację infrastruktury drogowej, w tym ścieżek rowerowych i dróg, mogą wykorzystać dane z hurtowni do lepszego zrozumienia potrzeb rowerzystów, identyfikowania słabych punktów i dostosowywania projektów w celu zwiększenia bezpieczeństwa i komfortu dla rowerzystów.

Policja: Policja może wykorzystać dane z hurtowni do analizy wzorców wypadków rowerowych, identyfikacji obszarów o podwyższonym ryzyku, zaplanowania patroli i skierowania zasobów w celu zapewnienia większego bezpieczeństwa na drogach.

3. Dane źródłowe

3.1. Źródła danych

Źródłem danych są dwa pliki .csv które są połączone ze sobą atrybutem Accident_index.

Lp.	Plik	Typ	Liczba rek.	Rozmiar[MB]	Opis
1.	Accidents	.csv	Ok .830 tyś.	~75 mb	Plik zawierający w sobie wszystkie szczegóły dotyczące wypadku
2.	Bikers	.csv	Ok .830 tyś.	~30mb	Plik zawierający w sobie wszystkie szczegóły dotyczące kierowców rowerów.

Tabela 1. Zbiory danych źródłowych

W celu ułatwienia analizy będę posługiwał się połączonym plikiem:

Lp.	Plik	Typ	Liczba rek.	Rozmiar[MB]	Opis
1.	Accidents_info	.csv	Ok .827 tyś.	~100 mb	Plik zawierający w sobie wszystkie szczegóły dotyczące wypadku jak i kierowców rowerów.

Tabela 2. Informacje na temat danych źródłowych

3.2. Lokalizacja, dostępność danych źródłowych

Dane są dostępne pod linkiem: [link](#)

Na stronie Kaggle w formacie .csv

3.3. Słownik danych – interpretacja

Tabela 3. Słownik atrybutów

Plik: Accidents_info.csv			
Lp.	Kolumna	Typ	ZNACZENIE
1.	Accident_Index	String	Numer identyfikujący wypadek podawany dla każdego incydentu, łączy dwa pliki w jeden.
2.	Age_Grp	String	Zakres wieku poszkodowanych
3.	Date	Date	Data zdarzenia
4.	Day	String	Dzień tygodnia zdarzenia
5.	Gender	String	Płeć poszkodowanych, Female- oznacza, że poszkodowanymi były kobiety, Male- oznacza, że poszkodowanymi były mężczyźni, Other - oznacza, że poszkodowanymi były osoby różnych płci
6.	Light_conditions	String	Warunki świetlne podczas wypadku
7.	Number_of_Casualties	Numeryczny	Liczba osób poszkodowanych ogółem w wyniku wypadku
8.	Number_of_Vehicles	Numeryczny	Liczba zamieszanych pojazdów
9.	Road_conditions	String	Warunki nawierzchni podczas zdarzenia, czy była mokra, oblodzona.
10.	Road_type	String	Typ drogi na której doszło do zdarzenia
11.	Severity	String	Powaga wypadku/ stan obrażeń poszkodowanych Fatal- śmiertelne Serious – poważne obrażenia Slight – lekkie obrażenia
12.	Speed_Limit	Numeryczny	Limit prędkości na miejscu zdarzenia
13.	Time	String	Godzina zdarzenia
14.	Weather_conditions	String	Warunki pogodowe podczas zdarzenia

3.4. Ocena jakościowa danych

Wynik analizy jakościowej przeprowadzonej za pomocą programu Tableau oraz profilu danych SSIS został przedstawiony w tab. 3.

Tabela 4. Ocena jakościowa danych

Plik: Accidents_info.csv				
Lp.	Atrybut	Typ danych	Zakres wartości	Ocena jakości/znaczenie
1.	Accident_Index	String	-	Numer identyfikujący wypadek 0% null
2.	Age_Grp	String	11 to 15/...	Zakres wieku poszkodowanych 0% null
3.	Date	Date	dd-mm-rrrr	Data zdarzenia 0% null
4.	Day	String	7 dni tygodnia	Dzień tygodnia zdarzenia 0% null
5.	Gender	String	Male/Female/...	Płeć poszkodowanych 0% null 3% Other
6.	Light_conditions	String	Daylight/ Darkness lights lit/...	Warunki świetlne podczas wypadku 0% null
7.	Number_of_Casualties	Numeryczny	Liczby naturalne i 0	Liczba osób poszkodowanych ogółem w wyniku wypadku 0% null
8.	Number_of_Vehicles	Numeryczny	Liczby naturalne i 0	Liczba zamieszanych pojazdów 0% null 1%< złe dane
9.	Road_conditions	String	Dry/Wet/..	Warunki nawierzchni podczas zdarzenia ~7% unknown
10.	Road_type	String	Single carriageway/...	Typ drogi na której doszło do zdarzenia 0 % null 4% unknown
11.	Severity	String	Slight/Serious/...	Powaga wypadku 0% null
12.	Speed_Limit	Numeryczny	X mph	Limit prędkości na miejscu zdarzenia 0% null 1% złe dane
13.	Time	String	Hh:mm	Godzina zdarzenia 0% null
14.	Weather_conditions	String	Clear/Rain/...	Warunki pogodowe podczas zdarzenia 0% Null 3% unknown

4. Analityczne modele wielowymiarowe

4.1. Fakty podlegające analizie oraz ich miary

Analizie będzie podlegał zbiór zarejestrowanych zdarzeń (tab. 4.)

Tabela 5. Fakty podlegające analizie

Lp.	Fakt	Miary	Uwagi
1.	Wypadek	Liczba zaangażowanych pojazdów, Liczba poszkodowanych, Liczba wypadków	

4.2. Kontekst analizy faktów

Ustalony kontekst analizy faktów został przedstawiony w tab. 5.

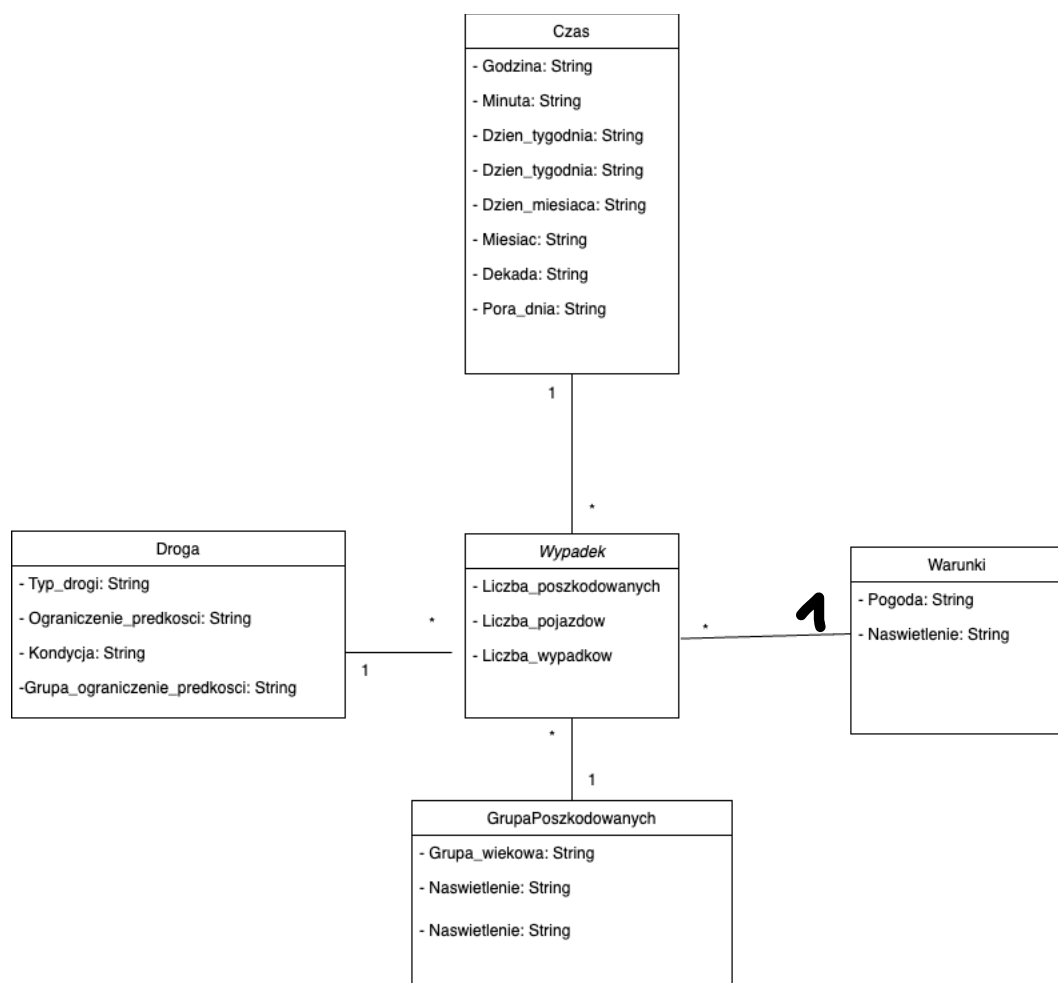
Tabela 6. Wymiary analizy faktów

Lp.	Kontekst analizy - wymiary	Własności
1.	Warunki	Naświetlenie drogi, pogoda
2.	Poszkodowany	Plec, grupa wiekowa, powaga obrażeń
3.	Droga	Typ drogi, ograniczenie prędkości, kondycja, grupa ograniczeń prędkości

4.	Czas	Dzien miesiaca, Miesiac, godzina, rok, dzien tygodnia, pora roku, dekada, pora dnia
-----------	-------------	--

4.3. Modele wielowymiarowe (UML)

Po przeanalizowaniu atrybutów źródła danych oraz ustalonego faktu i kontekstu analizy zaproponowano wielowymiarowy model konceptualny (rys. 1.). Składa się on z 1 faktu – wypadek , oraz 4 wymiarów. Model ten reprezentowany jest w postaci schematu na rys. 1.



Rysunek 1. Wielowymiarowy model analityczny przedstawiony na poziomie konceptualnym

5. Projekt procesu ETL

5.1. Schemat bazy danych HD (skrypt SQL)

```
-- Tworzenie tabeli Czas
CREATE TABLE DimCzas (
  Id_DimCzas INT IDENTITY(1,1) PRIMARY KEY ,
  Time VARCHAR(10),
  Day VARCHAR(10),
  Date DATE
);

-- Tworzenie tabeli Droga
CREATE TABLE DimDroga (
  Id_DimDroga INT IDENTITY(1,1) PRIMARY KEY,
  Road_conditions VARCHAR(50),
  Speed_Limit INT,
  Road_type VARCHAR(50)
);

-- Tworzenie tabeli Warunki
CREATE TABLE DimWarunki (
  Id_DimWarunki INT IDENTITY(1,1) PRIMARY KEY,
  Weather_conditions VARCHAR(50),
  Light_conditions VARCHAR(50)
);

-- Tworzenie tabeli GrupaPoszkodowanych
CREATE TABLE DimGrupaPoszkodowanych (
  Id_DimGrupaPoszkodowanych INT IDENTITY(1,1) PRIMARY KEY,
  Age_Grp VARCHAR(50),
  Gender VARCHAR(10),
  Severity VARCHAR(50)
);

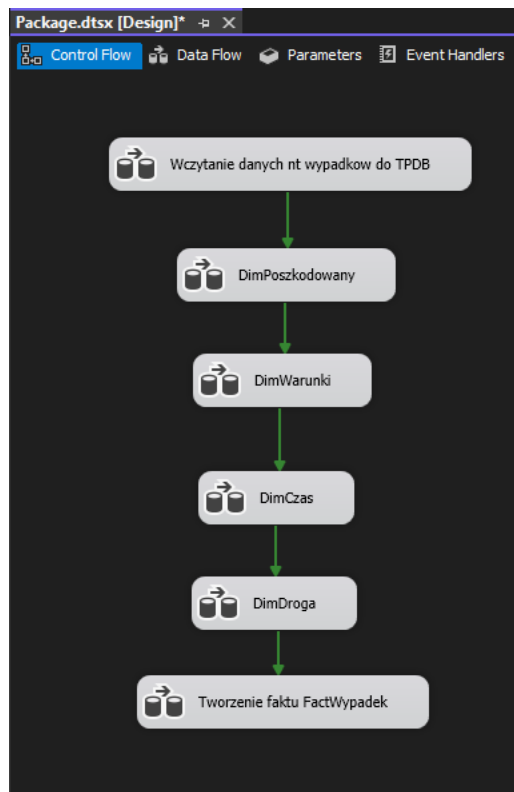
-- Tworzenie tabeli Wypadek
CREATE TABLE FactWypadek (
  Id_Fakt_Wypadek INT IDENTITY(1,1) PRIMARY KEY,
  Number_of_Casualties INT,
  Number_of_Vehicles INT,
  DimCzas_Id INT,
```

```

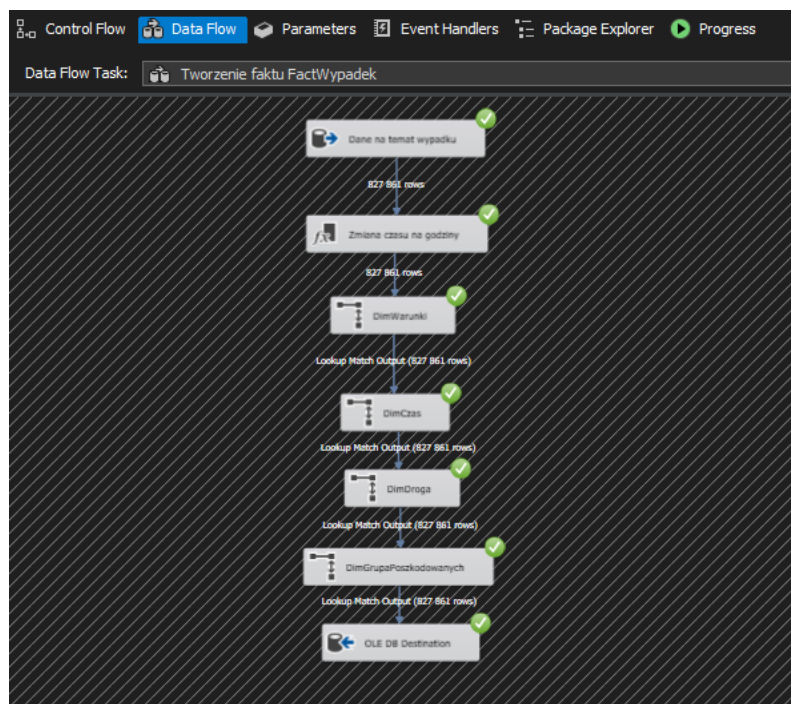
DimDroga_Id INT,
DimWarunki_Id INT,
DimGrupaPoszkodowanych_Id INT,
FOREIGN KEY (DimCzas_Id) REFERENCES DimCzas(Id_DimCzas),
FOREIGN KEY (DimDroga_Id) REFERENCES DimDroga(Id_DimDroga),
FOREIGN KEY (DimWarunki_Id) REFERENCES DimWarunki(Id_DimWarunki),
FOREIGN KEY (DimGrupaPoszkodowanych_Id) REFERENCES
Id_DimGrupaPoszkodowanych(Id_DimGrupaPoszkodowanych)
);

```

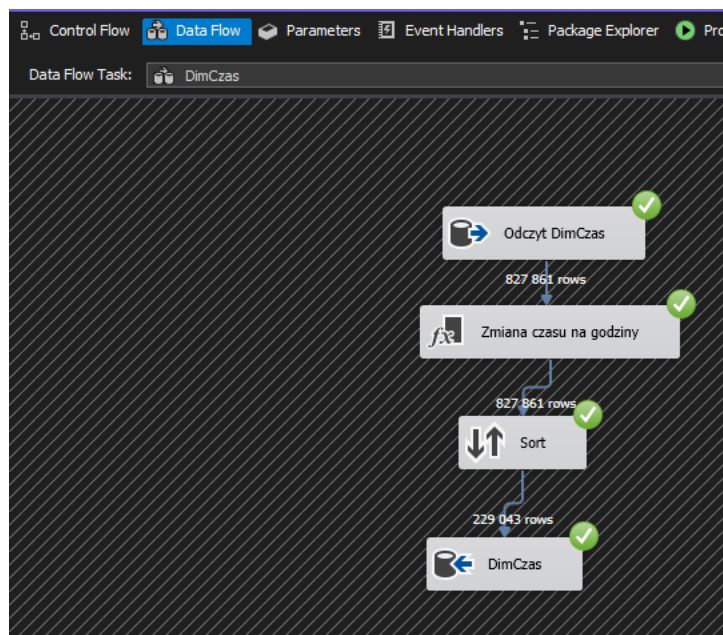
5.2 Specyfikacja procesów ETL (Control Flow + Data Flow)



Rysunek 2. Proces ETL Control Flow, dla całego projektu.



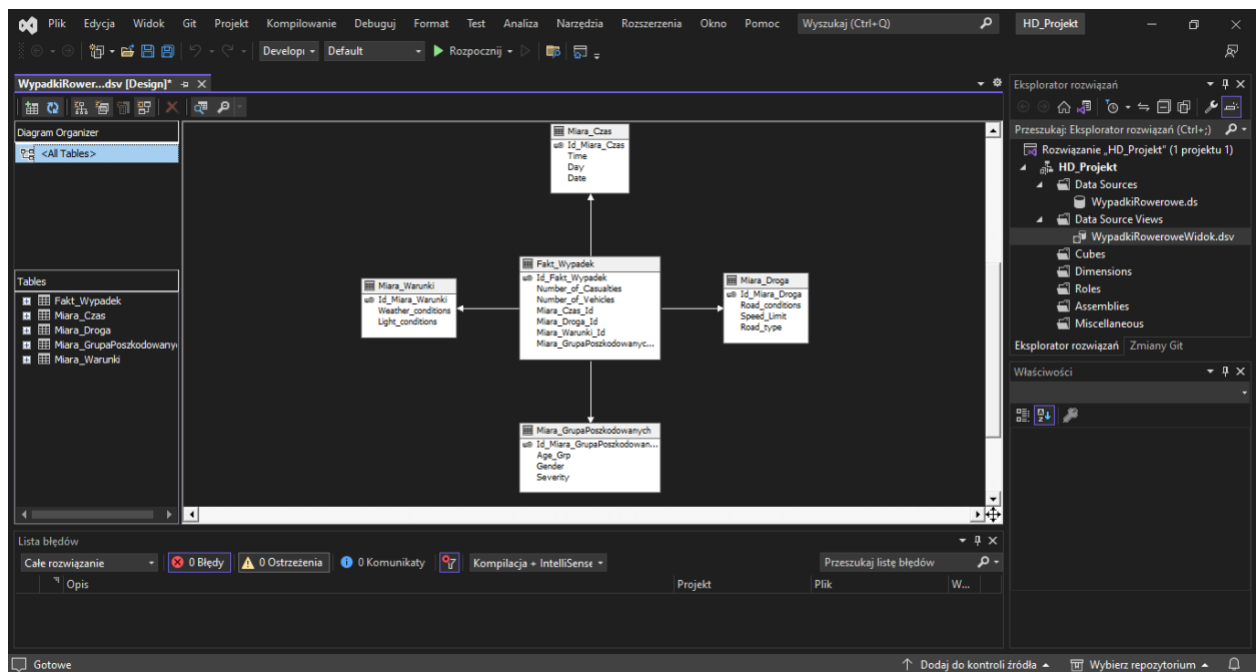
Rysunek 3. Proces ETL Data Flow, uzupełnienie tabeli FaktWypadek.



Rysunek 4. Przykładowy proces ETL Data Flow, uzupełnienie tabeli DimCzas.

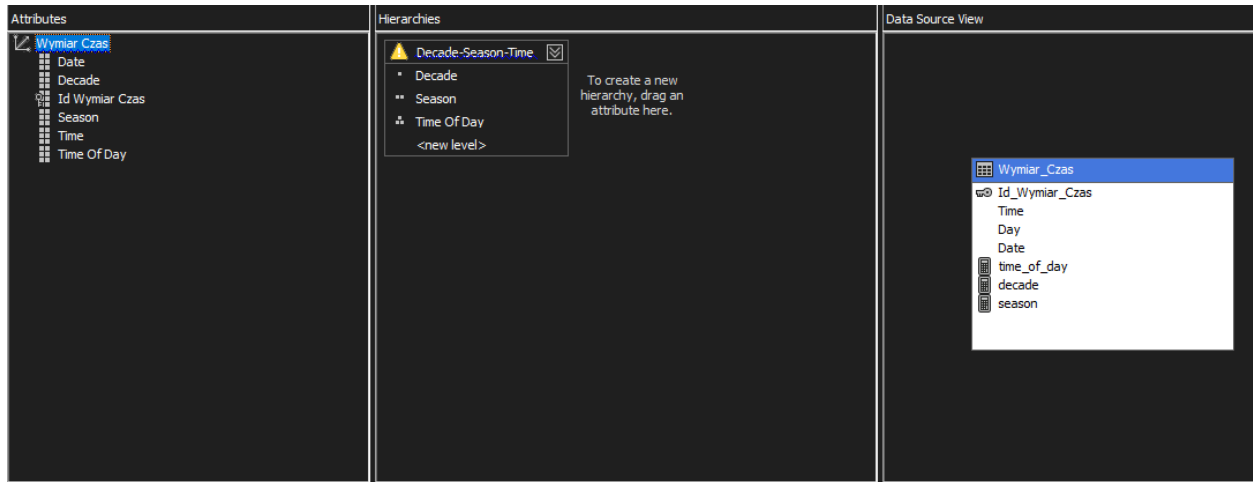
6. Implementacja modeli wielowymiarowych

6.1 Widok danych

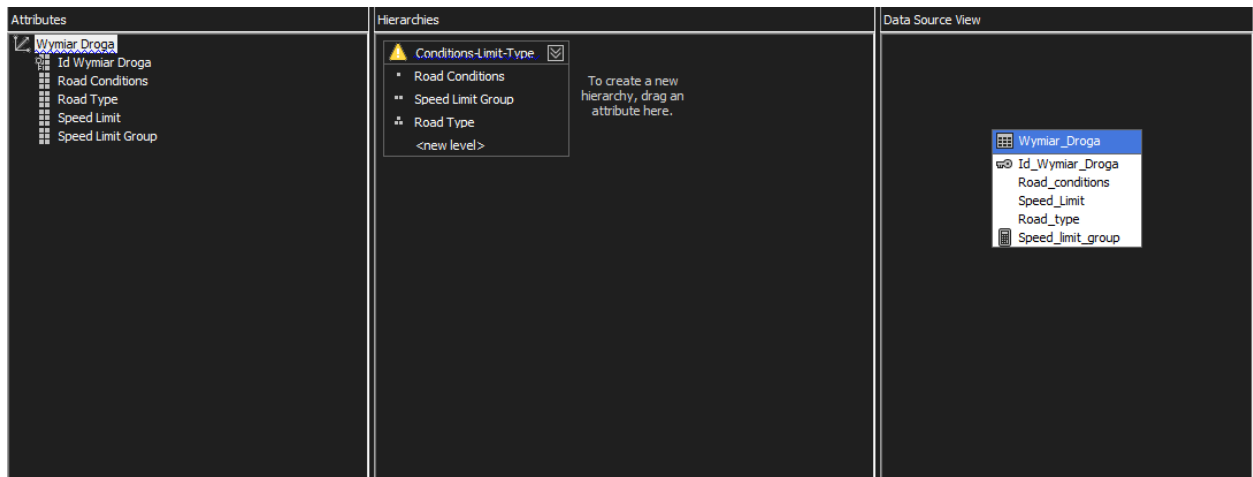


Rysunek 5. Widok danych w projekcie AS

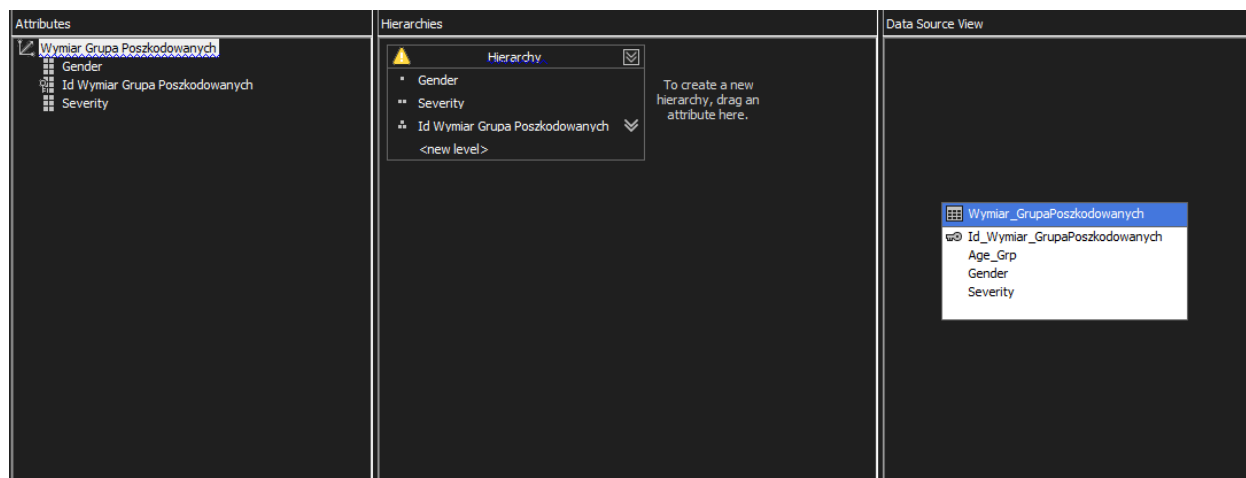
6.2 Wymiary



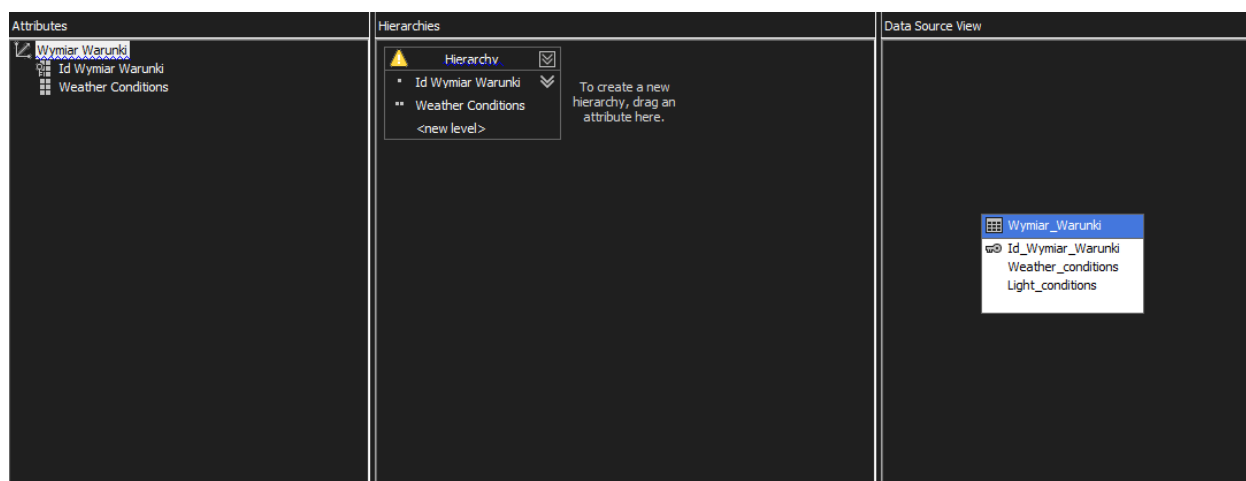
Rysunek 6. Wymiar Czas



Rysunek 7. Wymiar Droga

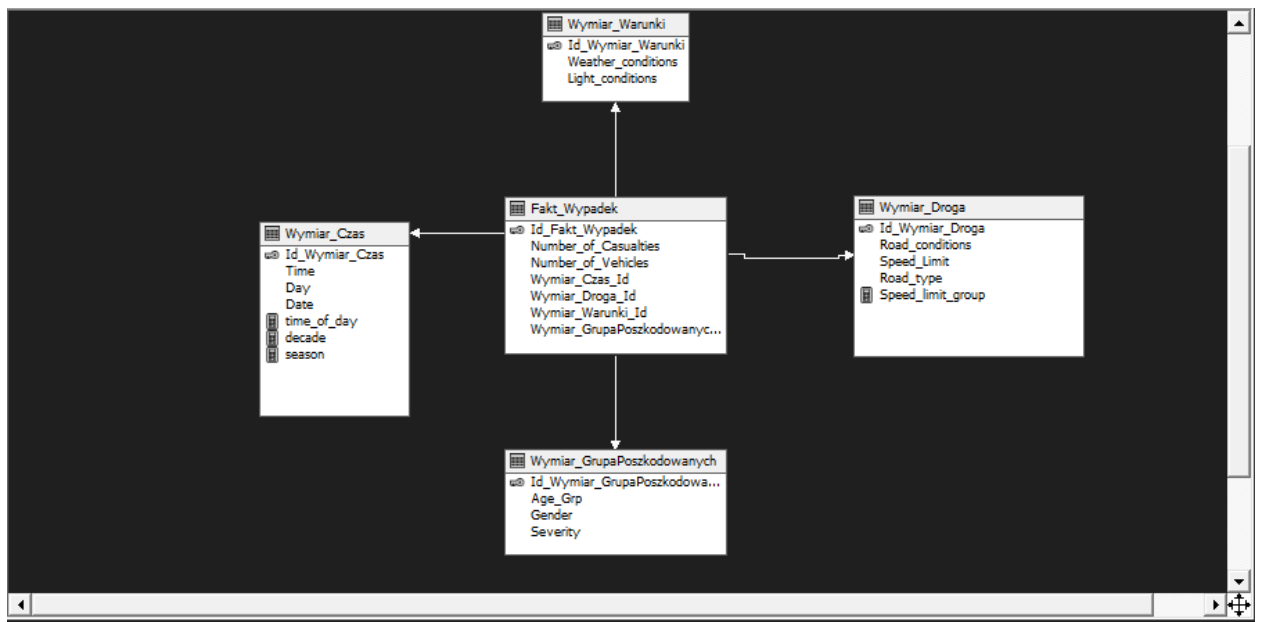


Rysunek 8. Wymiar Grupa Poszkodowanych



Rysunek 9. Wymiar Warunki

6.3 Modele wielowymiarowe – Kostki

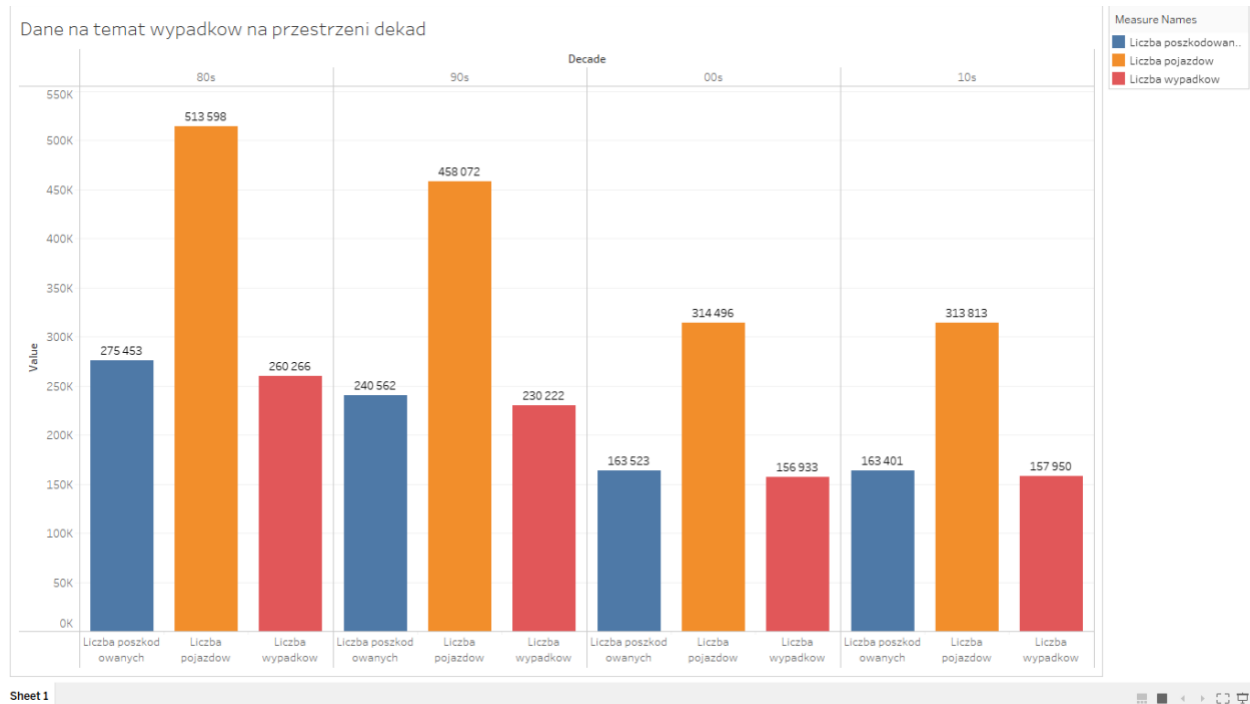


Rysunek 10. Model kostki w projekcie AS

7.1 Analiza danych

1.1. Realizacja procesów analitycznych

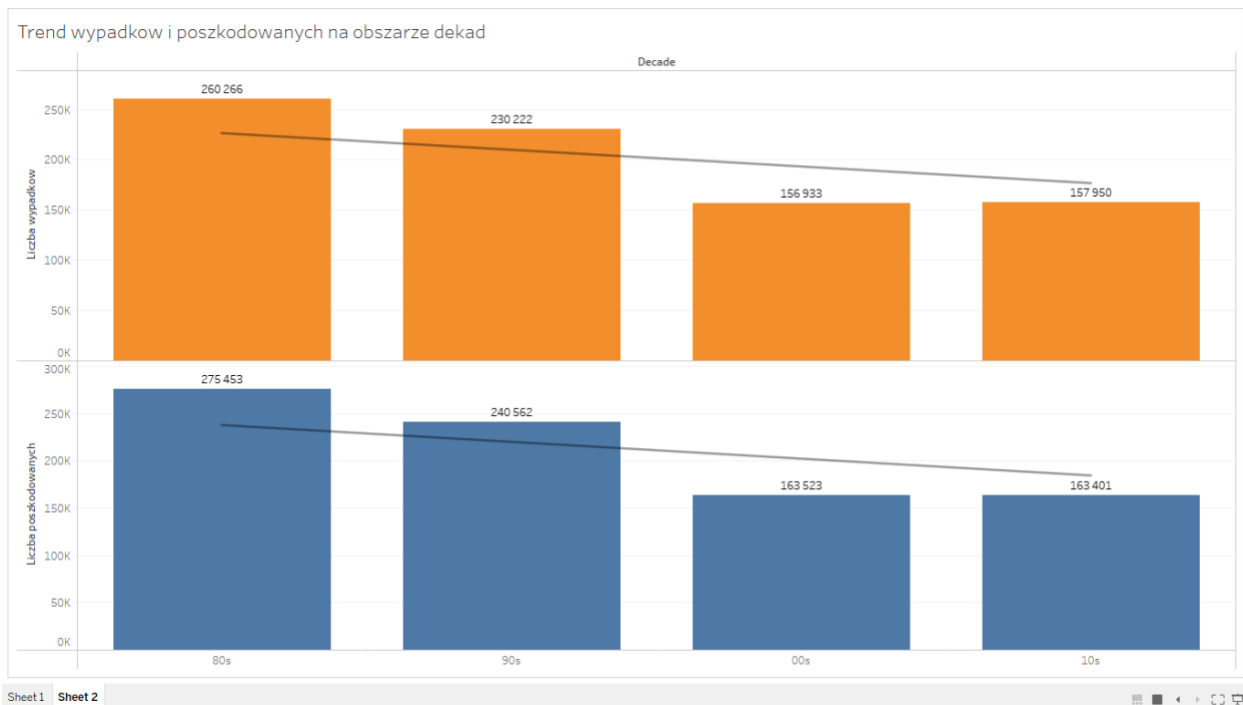
1. Jak wygląda ogólna sytuacja z wypadkami rowerowymi w Wielkiej Brytanii?



Rysunek 11. Wykres wszystkich miar na przestrzeni dekad.

Obserwacje:

Na rysunku 11 widać ogólną sytuację na przestrzeni dekad. Można zauważyć, że dane układają się w trendzie malejącym.



Rysunek 12. Wykres trendów na przestrzeni dekad.

Obserwacje:

Na rysunku 12 widać trend już trend malejący zaznaczony linią trendu. Jednak liczba wszystkich miar w latach 10' XXI wieku jest nieznacznie większa od lat 00'. Niepokojący jest fakt, że dane z lat 10' są tylko z okresu 8 lat, więc brakuje 2 lat do pełnego zakończenia, co może znaczyć, że trend zostanie zaburzony. Podobnie sytuacja ma się z poszkodowanymi w wypadkach jak i pojazdami.

Wnioski:

W XXI wieku liczba wszystkich miar znacząco zmalała, ponieważ prawie 2 krotnie w każdym przypadku, względem XX wieku, jednak druga dekada XXI wieku ma się gorzej w stosunku do pierwszej, co może znaczyć zahamowaniem trendu lub nawet jego odwróceniem.

2. Czy występują określone cykle lub okresy, w których dochodzi do większej liczby wypadków rowerowych?

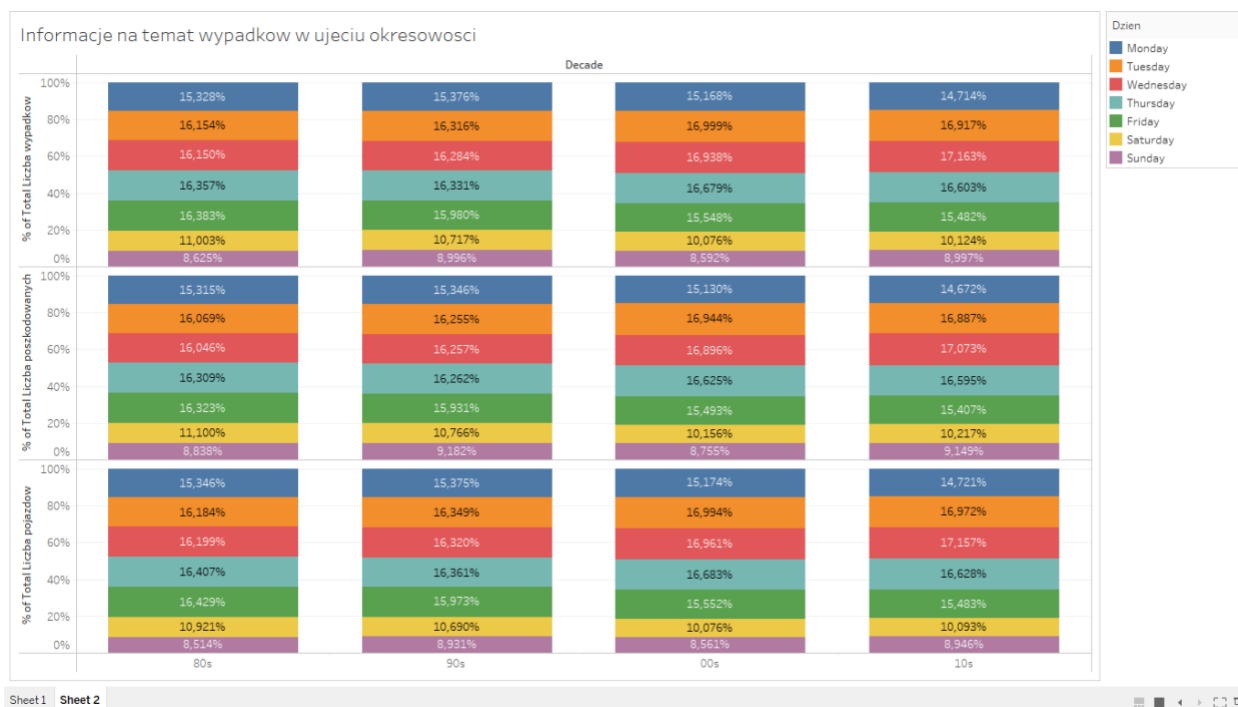
3. Czy można zidentyfikować czynniki sezonowe lub godzinowe, które mają wpływ na bezpieczeństwo rowerzystów?



Rysunek 13. Cykliczność wypadków rozpatrywana w okresie pór roku.

Obserwacje:

Z rysunku 9 można wyciągnąć ciekawą zależność, że najwięcej wypadków jest w lecie, a najmniej w zimie. W przypadku wszystkich miar i dekad zmienność co do procentowego udziału w całej dekadzie praktycznie nie ulega zmianie. Daje nam to ciekawą podstawę do dalszego zagłębienia się w cykliczności na mniejszych interwałach czasowych.



Rysunek 14. Cykliczność wypadków rozpatrywana w okresie dni tygodnia

Obserwacje:

Schodząc głębiej w podział ze względu na dni tygodnia, na rysunku 14 możemy zauważyć, że jeśli chodzi o cykliczność w wymiarze dni tygodnia, rozkład jest stosunkowo równy. Najmniej wypadków jest w weekendy, najwięcej w tygodniu z w miarę równym występowaniem. Rozkład w dekadach również nieznacznie się zmienia, odchylenia sięgają max 1.5%.



Rysunek 15. Cykliczność wypadków rozpatrywana w okresie pór dnia.

W przypadku pór dnia, również istnieje zauważalny trend. Nie jest on tak silny jak poprzednie ponieważ waha się znacznie bardziej bo odchylenie sięga nawet 8% względem dekad. Jednak niezmiennie jest to, że większość wypadków dzieje się w godzinach popołudniowych od 12 do 18. Jednak spora ilość jest również odnotowywana o poranku to znaczy 4-11. W nocy jest bardzo nieznaczna ilość wypadków co może być spowodowane niskim zainteresowaniem rowerzystów jazdą.

Wnioski:

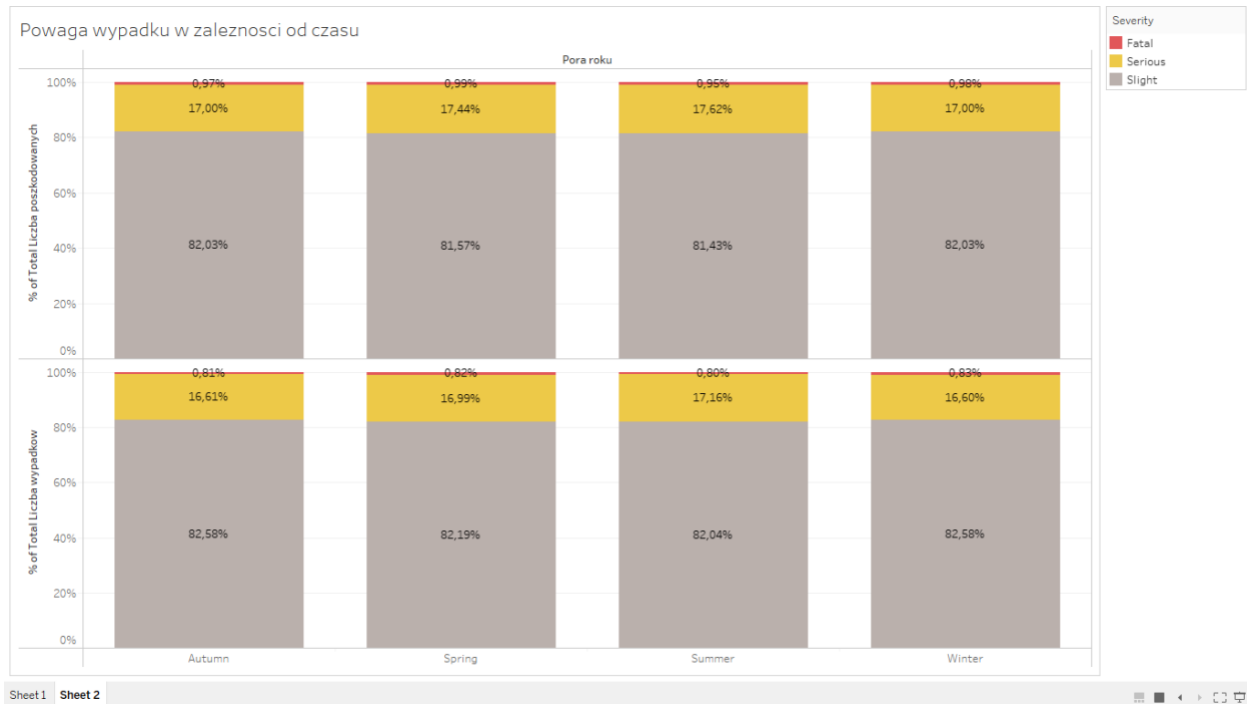
Cykliczność względem pór roku występuje i to w bardzo silnej relacji. Najwięcej wypadków jest w lecie, najmniej zaś w zimę.

Jeśli chodzi o dni tygodnia również występuje silna zależność cykliczna, mianowicie – najbezpieczniej jeździć w weekend.

W przypadku pór dnia istnieje już nie tak silna, jednak zauważalna zależność, większość wypadków ma miejsce w godzinach 12-18, najmniej w nocy.

4. Kiedy jest najbezpieczniejszy czas na jazdę rowerem?

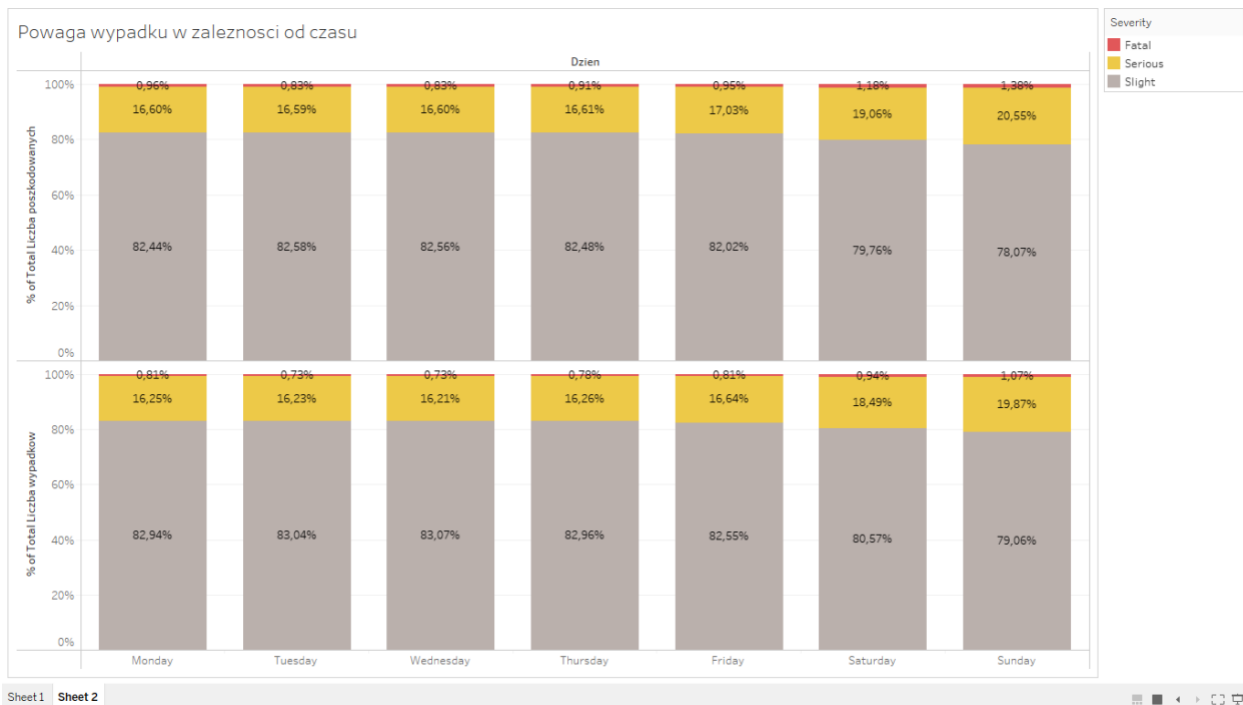
Czy zatem najbezpieczniej jeździć w zimie i najlepiej w weekend wieczorem?



Rysunek 16. Powaga wypadków rozpatrywana w okresie pór roku.

Obserwacje:

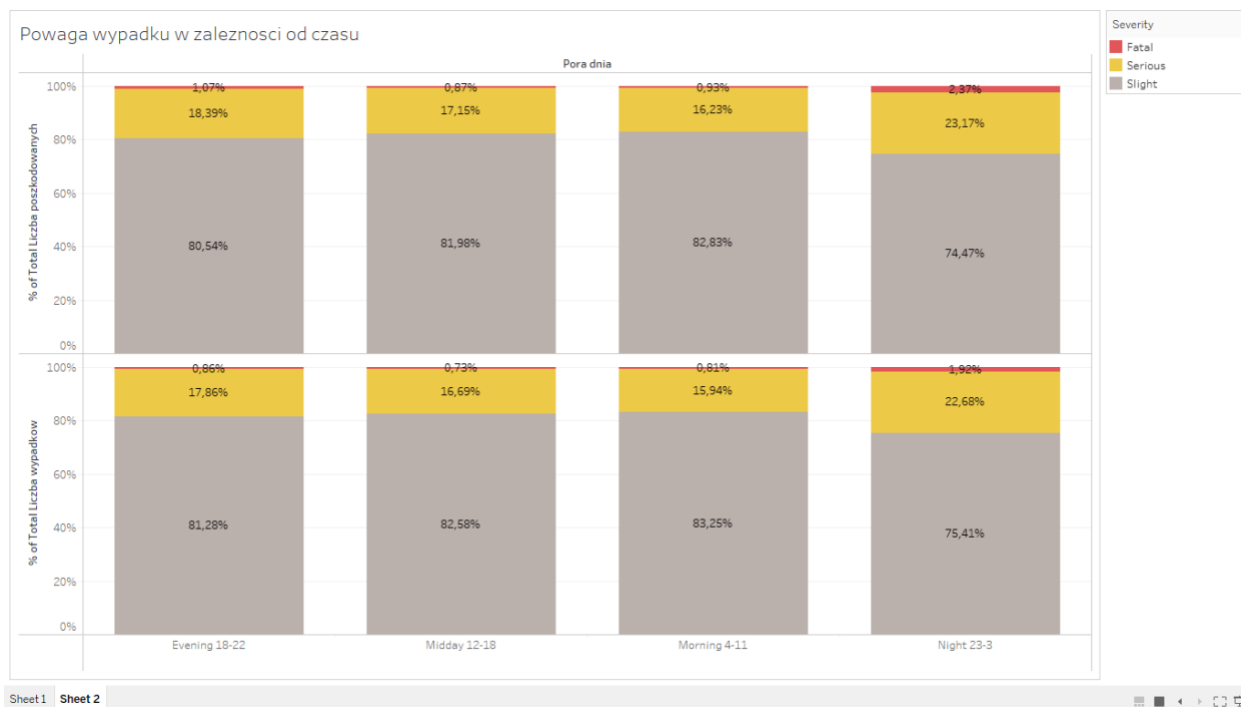
Pomimo wcześniejszych cykliczności dotyczących ilości wypadków i poszkodowanych nie braliśmy pod uwagę obrażeń odniesionych przez ofiary. Skupiając się na tym aspekcie, na rysunku 16 można zauważyć, że proporcje w każdej porze roku są takie same, około 1% osób poszkodowanych ponosi śmierć, a około 80% odnosi lekkie urazy.



Rysunek 17. Powaga wypadków rozpatrywana w okresie dni tygodnia.

Obserwacje:

Jeśli nasz podział będzie ze względu na dzień, tak jak na rysunku 17, zauważymy coś ciekawego. Pomimo, że w weekend jest najmniej wypadków są one najbardziej poważne! Podczas gdy w tygodniu procent śmiertelnych wypadków dochodzi do 1% w weekend przebija go znacząco. Procent osób odnoszących lekkie urazy spada poniżej 80%.



Rysunek 18. Powaga wypadków rozpatrywana w okresie pory dnia.

Ta sama ciekawa obserwacja dotyczy pór dnia. Na rysunku 15 powiedzieliśmy, że rzadkością jest, aby do wypadków dochodziło nocą. Teraz na rysunku 18 widzimy, że spośród wszystkich, to właśnie w nocy jest największy procent wypadków śmiertelnych i poszkodowanych odnoszących dotkliwsze urazy! Liczba wypadków kończących się śmiercią w nocy to prawie 2%, gdy w innych procent plasuje się na poziomie znacząco poniżej 1%.

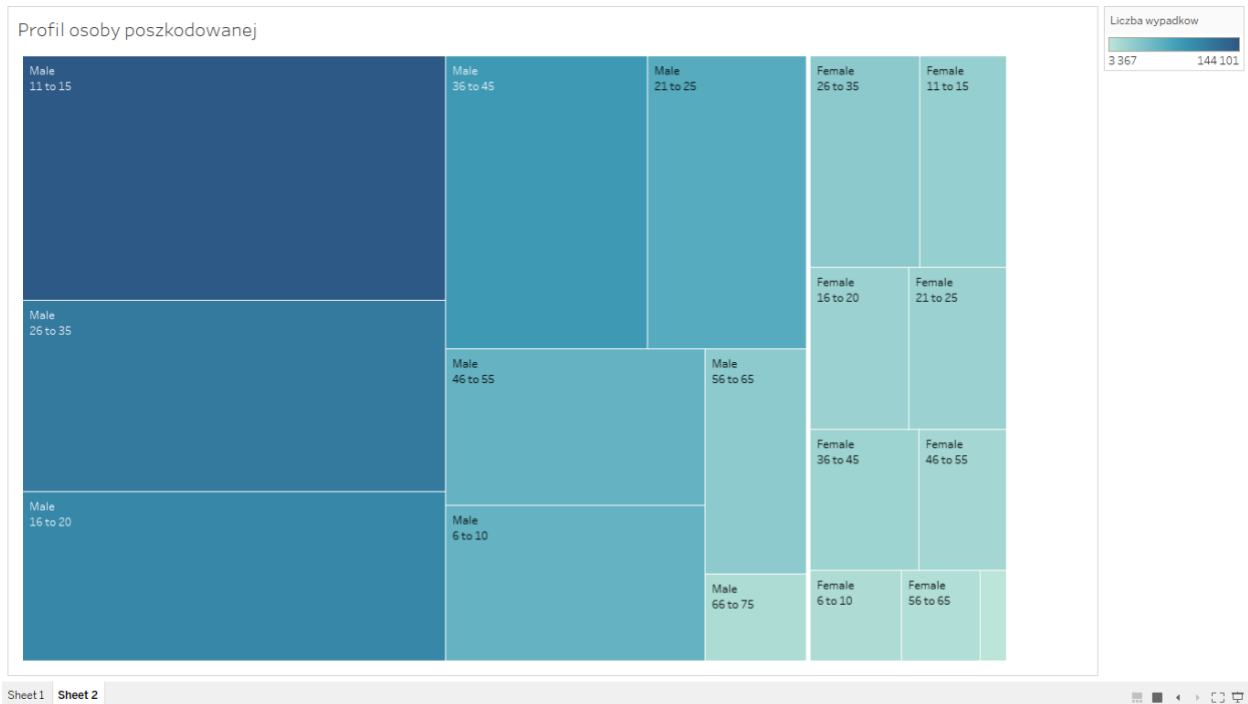
Wnioski:

Pomimo, że w weekendy jest mniej wypadków i osób poszkodowanych, są one poważniejsze i dotkliwsze.

To samo tyczy się pór dnia. Pomimo, że w nocy dochodzi rzadziej do wypadków i mniej osób jest poszkodowanych, są to wypadki bardziej drastyczne, a poszkodowani częściej tracą życie. Sama analiza ilości wypadków może nam dać błędny obraz bezpieczeństwa.

Pomimo większej liczby wypadków, najlepiej jeździć we wtorki najlepiej w godzinach popołudniowych, jeśli chodzi o porę roku, każda ma podobne statystyki.

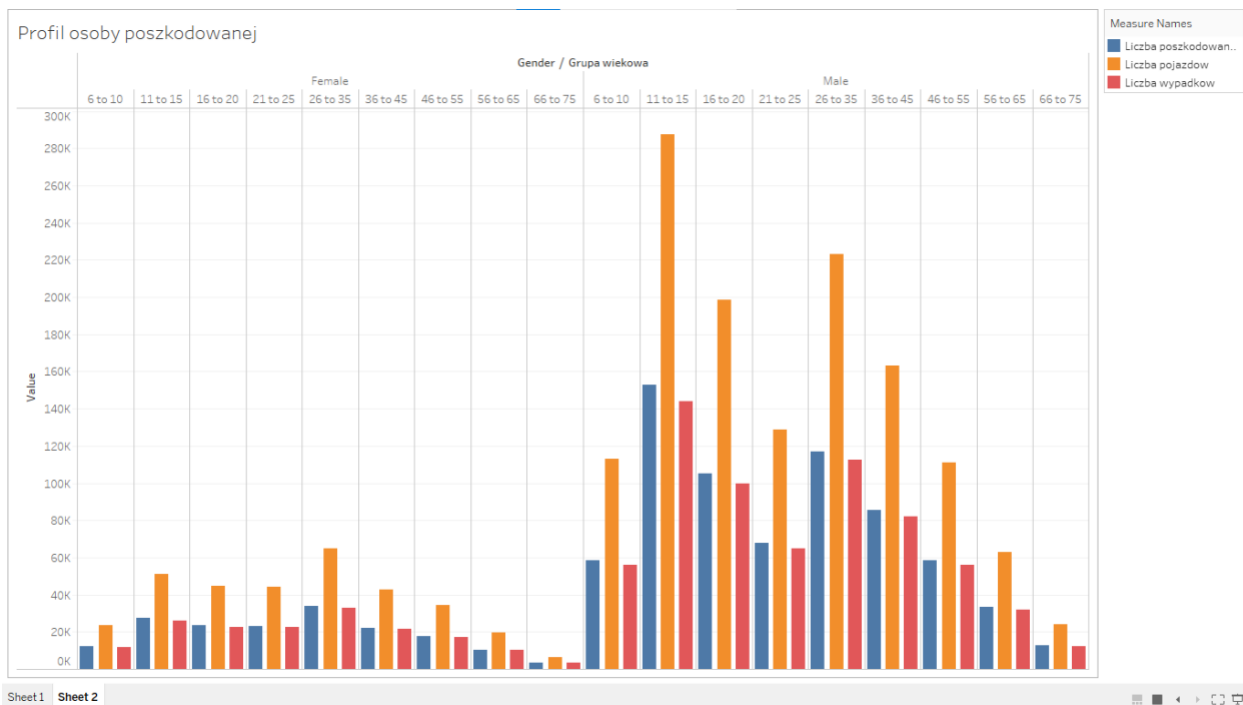
5. Czy można zauważyć pewne grupy wiekowe, które są bardziej narażone na ryzyko wypadków?



Rysunek 19. Profil osoby poszkodowanej

Obserwacje:

Rysunek 19 przedstawia profil osoby poszkodowanej w wypadkach rowerowych. Możemy zaobserwować z niego, że w większości, wypadkom ulegają mężczyźni w każdej grupie wiekowej jest to duża liczba w porównaniu do kobiet. Jeśli chodzi o kobiety, znacząco wyróżniają się dwie grupy 26-35 jak i 11-15. W dolnych widełkach występują osoby starsze. Lepiej zobrazować te dysproporcje pomoże nam rysunek 20.



Rysunek 20. Profil osoby poszkodowanej bardziej szczegółowo.

Obserwacje:

Patrząc bardziej szczegółowo na rysunek 20, możemy zauważyć znaczącą różnicę w proporcjach.

Jeśli chodzi o profil osoby poszkodowanej, zauważalnie wyróżniają się dwie grupy.

1. Chłopcy w wieku od 11 do 15 lat.
2. Mężczyźni w wieku 26-36.

W przypadku kobiet jest to mniejsza liczba, lecz ten sam przypadek, czyli:

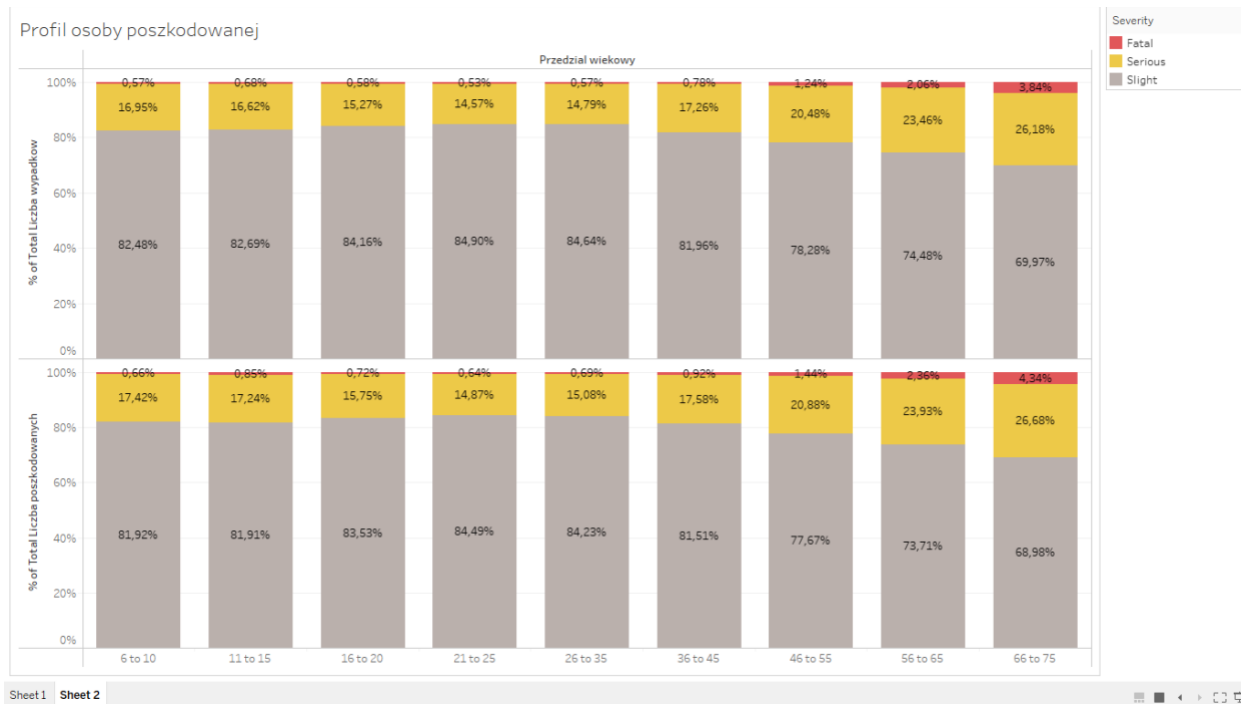
1. Dziewczynki w wieku 11 do 15
2. Kobiety w wieki 26-35

Wnioski:

Najbardziej narażeni są nastolatki w wieku od 11-15 lat. Grupa występuje licznie w obu podziałach. W wielu wypadkach bierze również udział grupa 26 do 35 latków w obu płci.

Najmniej liczną grupą są osoby starsze.

6. Jakie osoby najczęściej odnoszą obrażenia i kto jest najbardziej narażony jeśli chodzi o wypadki rowerowe?



Rysunek 21. Profil osoby poszkodowanej w kontekście powagi obrażeń.

Obserwacje:

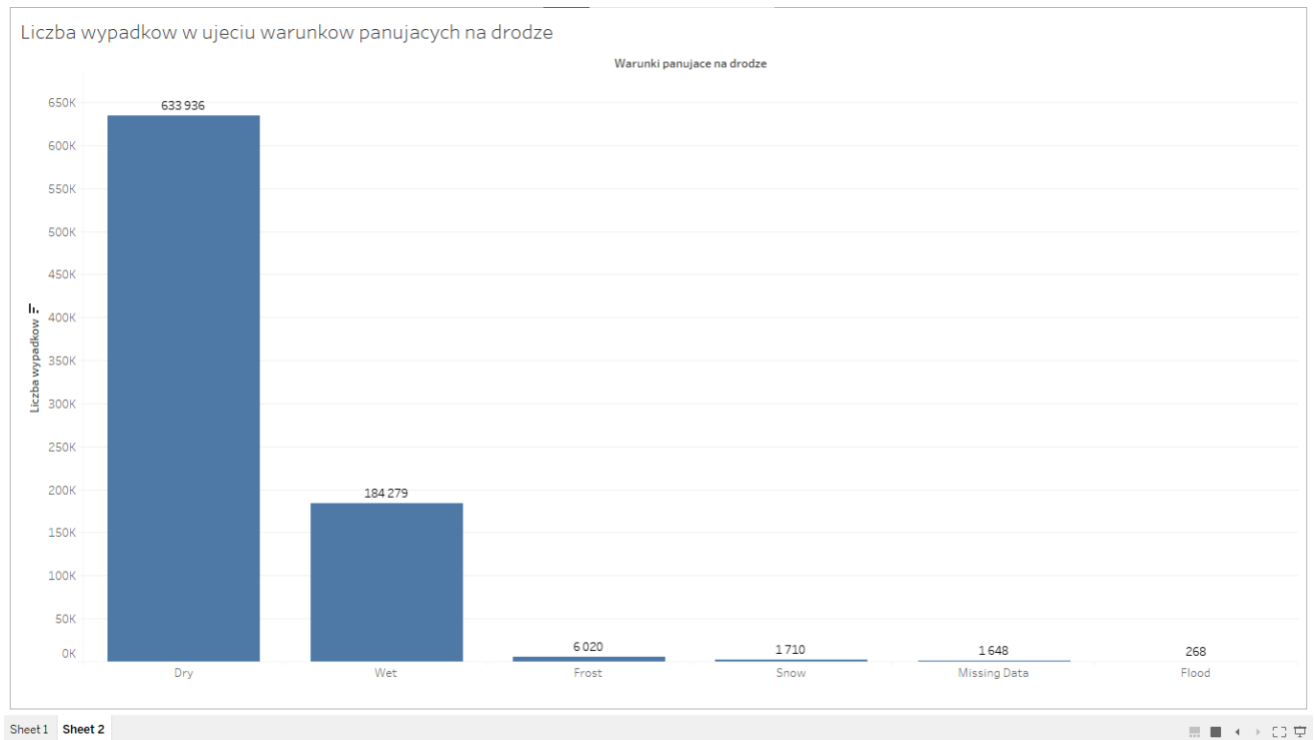
Tak jak wcześniej, okazuje się, że najmniej liczna grupa starszych osób pomimo małej ilości wypadków z ich udziałem, zaobserwowaną na rysunkach 20 i 19, są one poważniejsze i zdecydowanie częściej kończą się śmiercią.

Wnioski:

Powaga wzrasta wraz z wiekiem poszkodowanego. Ilość zgonów ludzi do lat 45 plasuje się na poziomie 1%, powyżej tej granicy, wypadki są bardziej śmiertelne oraz poszkodowani odnoszą większe obrażenia.

7. Jaki jest najważniejszy czynnik wpływający na powagę wypadku?

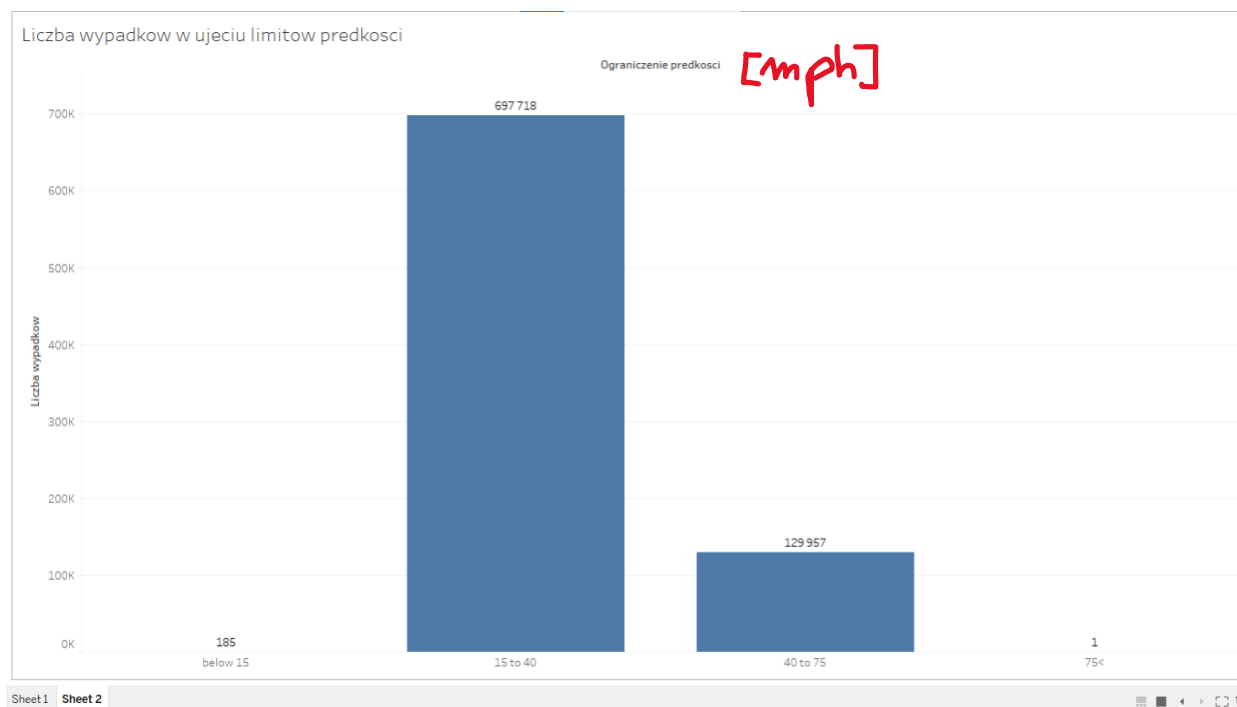
Zacznijmy naszą analizę od uwarunkowań drogowych:



Rysunek 22. Liczba wypadków w zależności od warunków drogowych.

Obserwacje:

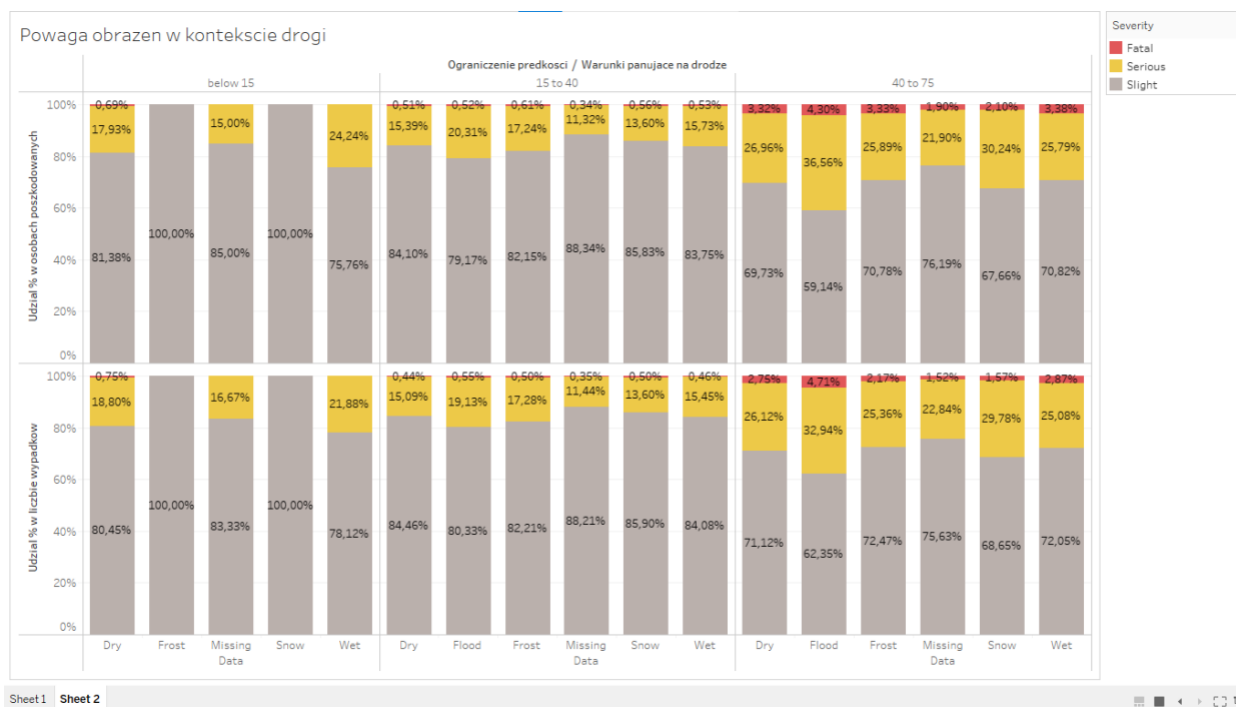
Zaczynając naszą analizę od ogólnego poglądu na sytuację, którą umożliwia nam rysunek 22, możemy zauważyć, że większość wypadków dzieje się na suchej powierzchni. Na drugim miejscu plasuje się mokra nawierzchnia. Jednak niewiele nam to mówi o powadze zdarzeń mających miejsce w takich warunkach.



Rysunek 23. Liczba wypadków w zależności od ograniczenia prędkości.

Obserwacje:

Kolejny ogólny pogląd na ograniczenia prędkości zaprezentowany na rysunku 23. Pomaga nam on zauważyć, że bardzo mało wypadków ma miejsce na drogach z ograniczeniem prędkości mniejszym niż 15. Zdecydowana większość występuje na drogach gdzie ograniczenie wynosi od 15 do 40. Czego jednak nie pokazuje rysunek 23 jest powaga obrażeń odniesionych w skutek wypadku.



Rysunek 24 Powaga wypadku w zależności od ograniczenia prędkości i warunków drogowych.

Obserwacje:

Na bardziej dogłębną analizę tego przypadku pozwoli nam rysunek 24.

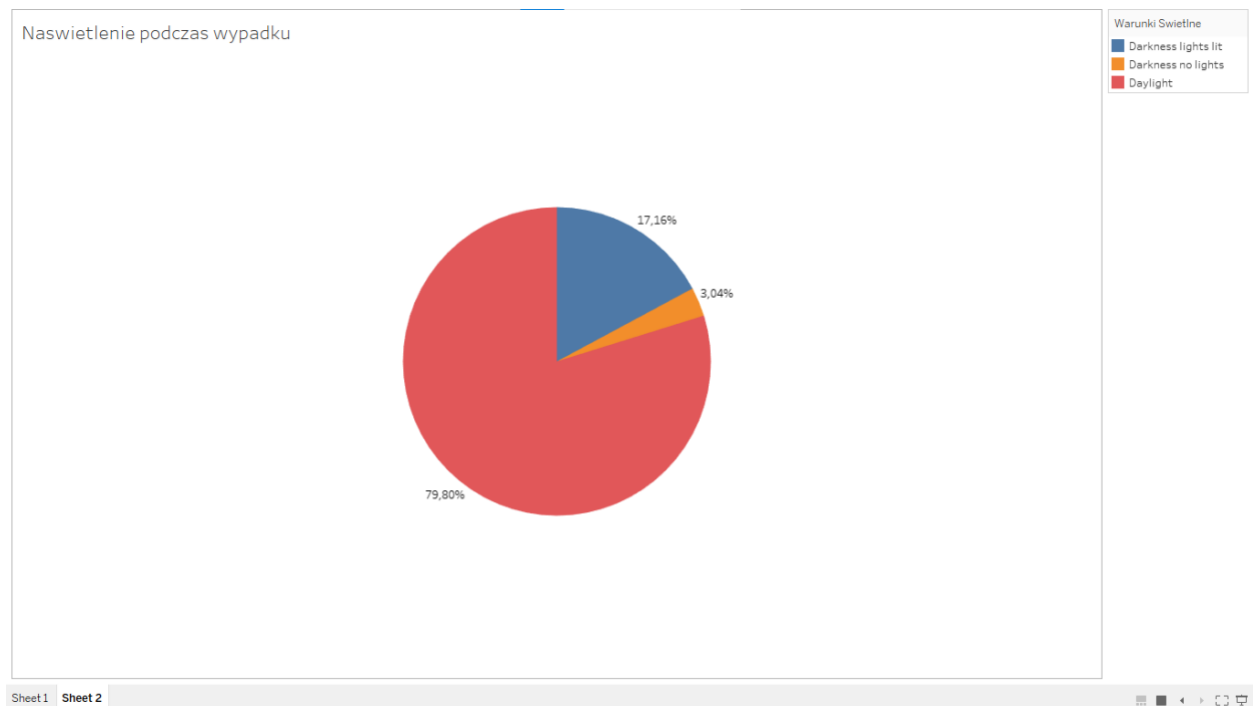
Dzięki niemu, możemy zauważyć, że pomimo mniejszej ilości wypadków na drogach z wyższym ograniczeniem prędkości, powaga tych zdarzeń jest zdecydowanie większa. Gdy poniżej 15 mph nie występują śmiertelne wypadki, a w przedziale 15-40 są to wartości niskie bo do 0.55%, w momencie przejścia do przedziału 40-75 mph, są to wartości powyżej 1.5% sięgające nawet pod 5%.

Wnioski:

Największy procent śmiertelnych oraz poważnych wypadków jest podczas występowania deszczu lub pozostałej po opadach wody na drogach. Szczególnie niebezpieczne są drogi z wysokim ograniczeniem prędkości tzn. 40-75 mph.

Skończywszy aspekt uwarunkowań drogowych kolejnym rozpatrywanym kontekstem będą uwarunkowania zewnętrzne.

Przejdźmy teraz z naszą analizą do uwarunkowań zewnętrznych, takich jak pogoda oraz naświetlenie:

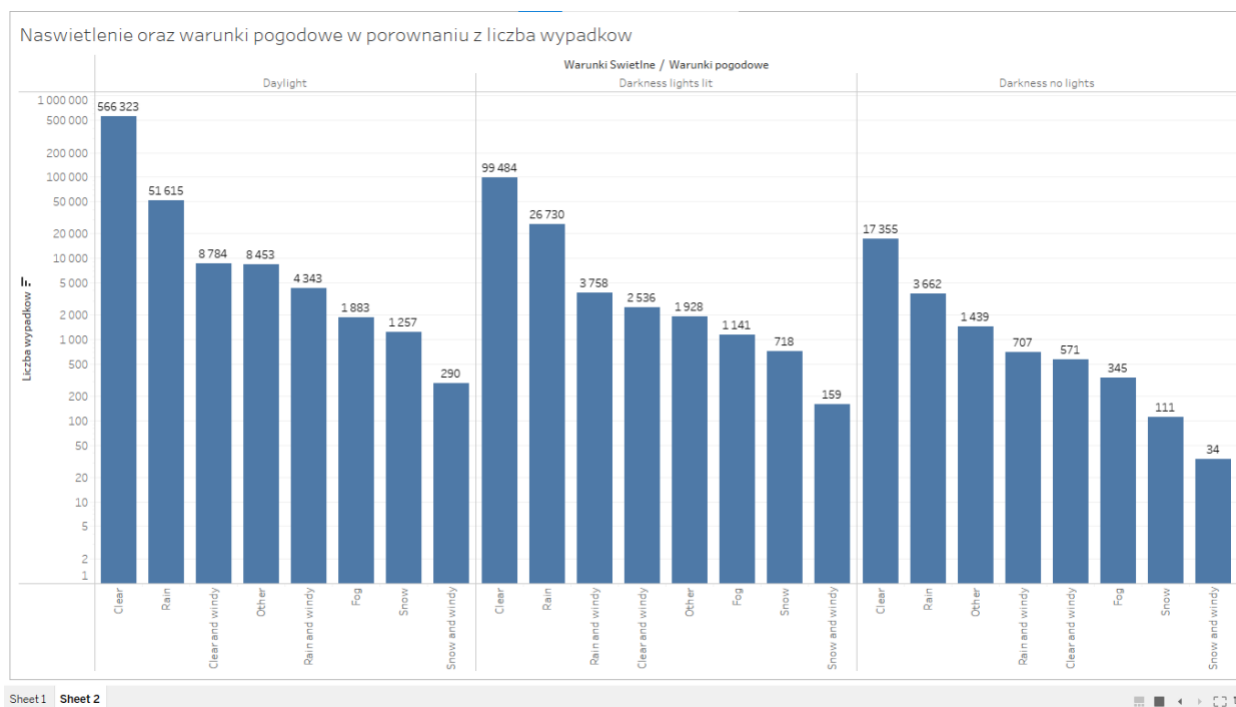


Rysunek 25. Ogólny pogląd na wpływ naświetlenia podczas wypadków.

Obserwacje:

Rysunek 25 pozwala nam zobaczyć bardzo ogólnie na wpływ naświetlenia przy wypadkach.

Jeśli chodzi o ogólny pogląd udziału poszczególnego naświetlenia w liczbach wypadku, to można jasno zauważyć, że zdecydowana większość miała miejsce podczas światła dziennego. Wypadki w ciemności zajmują ok. 20%. Niestety nie daje nam to wystarczających informacji na sprawdzenie czy jest to czynnik ważny przy wypadkach. Dlatego warto go wzbogacić o kolejne atrybuty.

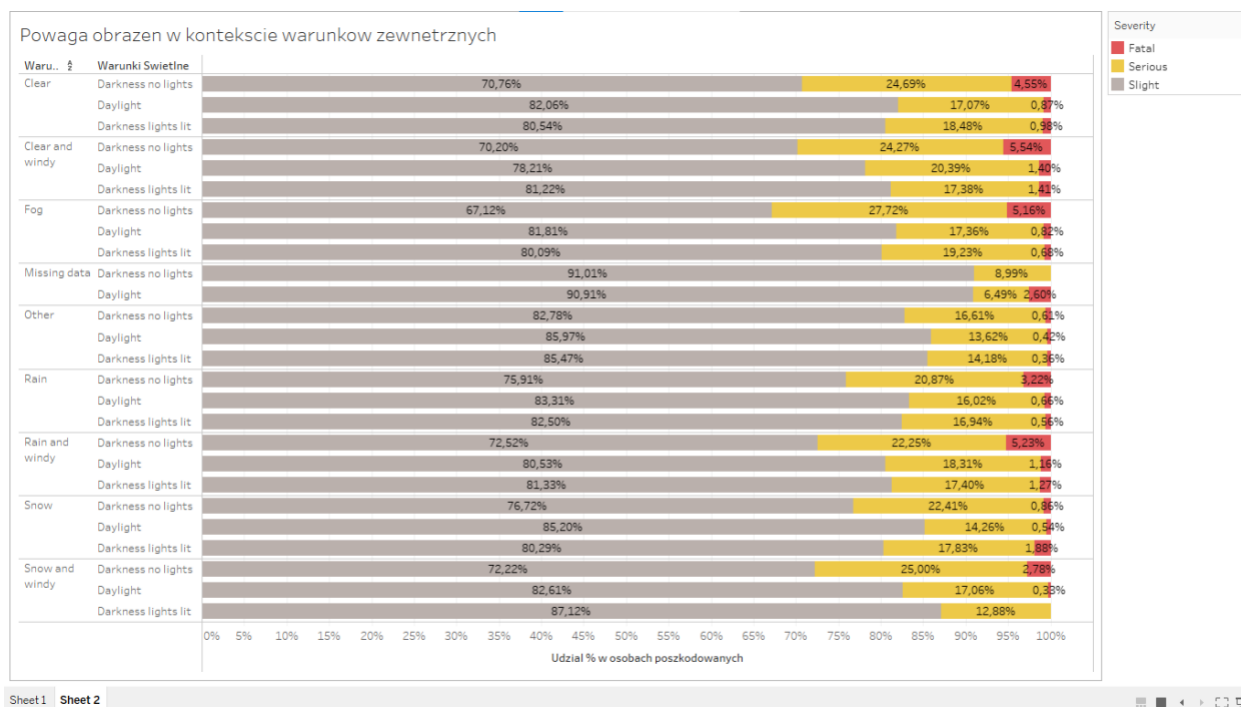


Rysunek 26. Wpływ światła oraz warunków pogodowych na liczbę wypadków.

Obserwacje:

Dorzucając do naświetlenia kolejny atrybut wymiaru, na rysunku 26 w postaci pogody, zauważymy, że zdecydowana większość miała miejsce gdy pogoda była przejrzysta/dobra. Wykres jest w skali logarytmicznej, więc należy zwrócić szczególną uwagę na liczby nad słupkami. Pierwsze dwa zdecydowane miejsca to dobra pogoda i deszcz, nieważne jakie naświetlenie.

Więc może to właśnie pogoda ma większy wpływ niż oświetlenie.



Rysunek 27. Wpływ światła oraz warunków pogodowych na powagę wypadków.

Obserwacje:

Łącząc nasze dwa wymiary z powagą wypadku dostajemy rysunek 27 na którym figurują ciekawe liczby, które kolejny raz potwierdzają, że mniej nie znaczy lepiej. Największe współczynniki śmiertelności i dużych obrażeń występuje w przypadku ciemności bez żadnego światła. Niezależnie od pogody procent jest w miarę stały i wynosi 5% +/- 0.5%, jedynym wyjątkiem jest pogoda śnieżna i wiatrzysta gdzie procent ten wynosi ok. 3%. Nie tylko śmiertelność jest większa w tym naświetleniu ale także poszkodowani odnoszą poważniejsze obrażenia.

Wnioski:

Jazda w miejscach nieoświetlonych lub po zmroku niesie ze sobą większe ryzyko poważnych obrażeń, a nawet śmierci. Jeśli chodzi o pogodę, najniebezpieczniej jest gdy pada deszcz lub jest mgła. Należy pamiętać, że podczas dobrej pogody jest zdecydowanie najwięcej przypadków śmiertelnych.

Wnioski końcowe do pytania 7:

Biorąc pod uwagę wszystkie konteksty analizy, czynniki zewnętrzne jak i czynniki drogowe. Z danych i obserwacji można wyciągnąć, że największy wpływ na wypadek (jego liczbę jak i powagę) ma naświetlenie jak i stan nawierzchni.

7.2 Podsumowanie - wnioski z analizy

1. Jak wygląda ogólna sytuacja z wypadkami rowerowymi w Wielkiej Brytanii?

Sytuacja na przestrzeni dekad ulega zmianie, trend wypadków oraz ofiar jest malejący, jednak trend ten został zachwiany w 10' latach XXI wieku. Może to wskazywać na odwrócenie trendu. Malejąca ilość wypadków może wynikać z większej świadomości ludzi jeśli chodzi o bezpieczeństwo na drodze. Przepisy również zmieniły się na korzyść dla rowerzystów, którzy mogą czuć się bezpieczniej na drodze za sprawą ograniczeń prędkości lub lepiej oświetlonych dróg.

2. Czy występują określone cykle lub okresy, w których dochodzi do większej liczby wypadków rowerowych?

Cykliczności w tym przypadku są bardzo widoczne. Większość wypadków dzieje się podczas lata, mało w zimie. Nie powinno nas to dziwić ponieważ rowerzyści wyjeżdżają na ulicę chętniej gdy jest ciepło.

Jeśli chodzi o dni tygodnia, to również widzimy tutaj równy rozkład pomiędzy dniami. Najwięcej najmniej wypadków dzieje się w weekend. Możliwe, że wtedy odchodzą do pracy które często przebiegają przez główne drogi.

Pory dnia również mogą się pochwalić dobrą cyklicznością, najwięcej wypadków jest w godzinach od 12-18 czyli powroty z pracy. Bardzo mało dzieje się w nocy od 23-3 prawie nikt nie jeździ.

3. Czy można zidentyfikować czynniki sezonowe lub godzinowe, które mają wpływ na bezpieczeństwo rowerzystów?

Na bezpieczeństwo rowerzystów może wpływać w dużym stopniu czas w którym sięgną po rower. Warunki pogodowe, świetlne jak i sezonowe mają bardzo duży wpływ na powagę wypadku jak i na jego szansę.

4. Kiedy jest najbezpieczniejszy czas na jazdę rowerem?

Tutaj nie ma jednej odpowiedzi, bo gdy powiemy- zima weekend najlepiej wieczorem, ponieważ liczba wypadków jest najmniejsza, nie będzie to do końca poprawne, ponieważ wypadki w tym czasie są zdecydowanie dotkliwsze od pozostałych. Ich szansa jest mniejsza, ale za to bardziej

odczuwalna. Gdy zaś powiemy Lato popołudniu najlepiej w tygodniu, również będzie to nie do końca poprawne, ponieważ jest w tym momencie najwięcej wypadków więc nasza szansa na uczestniczenie w jednym rośnie.

5. Czy można zauważyć pewne grupy wiekowe, które są bardziej narażone na ryzyko wypadków?

Tak i tutaj odpowiedź jest dość jasna. Najbardziej narażone są dzieci. W przypadku mężczyzn jak i kobiet, ta grupa stanowiła dużą część całości.

Bardzo dużą częścią jest grupa w wieku od 26 do 35 lat. Może to być spowodowane popularnym

6. Jakie osoby najczęściej odnoszą obrażenia i kto jest najbardziej narażony jeśli chodzi o wypadki rowerowe?

Zagłębiając się bardziej w temat, możemy jeszcze dojść do kolejnej grupy która jest narażona na wysokie ryzyko w wypadkach. Są to osoby starsze, których obrażenia zdecydowanie częściej od innych grup kończą się śmiercią.

7. Jaki jest najważniejszy czynnik wpływający na powagę wypadku?

Rozpatrując czynniki takie jak pogoda, stan nawierzchni oraz ograniczenia prędkości, możemy zauważyć, że najbardziej poważne wypadki w skutkach są podczas braku oświetlenia jak i przy słabym świetle, do tego przy wyższych ograniczeniach prędkości, wypadki są poważniejsze co może służyć za wskazówkę jak poruszać się bezpiecznie po drogach – wybierać drogi boczne, gruntowe. Bardzo dużo wypadków ma miejsce przy świetnej widoczności i suchej nawierzchni. Jednak to naświetlenie odgrywa największą rolę. Więc może to być wskazówka dla władz, aby lepiej oświetlić miasta/drogi, lub dla rowerzystów aby po zmroku starali się jeździć jak najmniej.

Wszystkie poruszone kwestie oraz pytania zostały poruszone i rozwinięte. Analiza była przyczynowo-skutkowa co pozwala na jej lepsze przyswojenie.

8. Wnioski końcowe z realizacji projektu

8.1 Problemy

Głównym problemem na początku było wyszukanie odpowiednich danych. Zdecydowanie od tego w dużej mierze zależy powodzenie i zabawa przy tworzeniu naszego projektu. Następnym problemem było rozłożenie sytuacji na fakty i miary, co było ciężkie przy tak małej ilości atrybutów, w szczególności miar było bardzo mało. Jednak analiza wydaje mi się ciekawa, poza tym, że brakuje w danych lokalizacji, bardziej krajowe spojrzenie na problem rowerzystów wydaje się równie ciekawym wyzwaniem.

Niestety oprogramowanie jest bardzo trudne do skonfigurowania, do tego nie działa na każdym sprzęcie oraz bardzo zwalnia komputer, jest to zdecydowany problem w tego typu wyzwaniach/projektach, czyli wymagania sprzętowe.

8.2 Pozyskana wiedza i doświadczenie

Pozyskana przeze mnie wiedza dotyczy obszarów posługiwania się oprogramowaniami do baz danych, tworzenia ich jak i pracy z nimi. Rozkład wycinka rzeczywistości na części również jest jedną z cennych i wyniesionych z projektu doświadczeń. Praca z narzędziem Tableau była bardzo udana i zdecydowanie przydatna. Praca z danymi i z hurtowniami to ciekawe doświadczenie.

9. Źródła informacji użyte w etapie analizy danych

Podczas analizy danych nie były użyte żadne inne źródła oprócz danych źródłowych:
dane: [link](#)

Uwaga:

- Niekompletny projekt nie będzie sprawdzany i tym samym ocena będzie negatywna!
- Kompletna dokumentacja musi być przesłana do sprawdzenia w formie pliku pdf nie później niż trzy dni przed terminem odbioru i prezentacji opracowanej hurtowni danych!
- Osoby, które chcą zaliczyć projekt i jednocześnie kurs HD w terminie zerowym muszą przekazać kompletną dokumentację projektową nie później niż 9.06.2023 do godz. 18:00.