**Hurtownie danych – Projekt – Etap 01**

PWr. WIZ, Data: 11-12.04.2022

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

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

# Uzasadnienie wyboru tematu projektu

1. **Propozycja tematu projektu**
   1. **Tytuł projektu**

Analiza zużycia mocy w poszczególnych regionach Kalifornii w kontekście średniej rocznej pogody z podziałem na źródło mocy oraz typ odbiornika

* 1. **Problemy**

Predykcja poboru prądu w poszczególnych godzinach w celu odpowiedniegodostosowania generacji prądu

Identyfikacja miejsc, w których infrastruktura potencjalnie może nie dać rady dostarczać zamodelowanego poboru prądu

Estymacja kosztów i zysków przez konkretnych dostawców

* 1. **Cel przedsięwzięcia**
* Zbadanie czynników: pora roku, rodzaj odbiorcy, pora dnia na zużycie prądu
* Ocena efektywności typów budynków pod kątem zużycia energii
* Określenie przyczyn zmian zużycia energii
  1. **Zakres analizy – badane aspekty**
* Porównanie regionów w ramach stanu Kalifornia
* Przygotowanie „odcisku palca” miasta w celu porównawczej analizy
* Charakterystyka zużycia energii np. źródła energii a lokalizacji na mapie stanu
  1. **Źródła danych**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Lp.** | **Plik, bazy danych** | **Typ** | **Liczba rek.** | **Rozmiar**  **[MB]** | **Opis** |
| 1. | California | .csv | ~ 10 000 000 | 1500 | Baza zawierający godzinowy pobór prądu przez różne sektory ze względu na godzinę w Kalifornii |

* 1. **Ocena jakości danych źródłowych**

Analiza zużycia mocy w poszczególnych regionach Kalifornii w kontekście średniej rocznej pogody z podziałem na źródło mocy oraz typ odbiornika

**Dane o wysokiej jakości**

**Dane o niskiej jakości**

**Dane nieistotne w analizie**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Lp.** | **Kolumna** | **Typ** | **Zakres** | **Ocena jakości** |
| 1. | [Date/Time] | Tekstowy, Data w formacie MM/DD HH:MM:SS | Cały rok godzinowo, 8760 unikalnych wartości | 0% null, 0 powtarzających się wartości |
| 2. | [Water Heater:WaterSystems:Gas (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0.02 - ~647 | 43% null |
| 3. | [Fans:Electricity (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0-~370 | 0% null |
| 4. | [Heating:Gas (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0-3460 | 0% null |
| 5. | [Electricity:Facility (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy |  | 0% null |
| 6. | [Heating:Electricity (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0-246 | 0% null |
| 7. | [InteriorEquipment:Electricity (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | ~1 - ~449 | 0% null |
| 8. | [Gas:Facility (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0-~3464 | 0% null |
| 9. | [InteriorLights:Electricity (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | ~0.27 - ~448 | 0% null |
| 10. | [InteriorEquipment:Gas (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0 - ~91 | 18% null |
| 11. | [Cooling:Electricity (kW)(Hourly)] | Numeryczny, Zmiennoprzecinkowy | 0-~996 | 0% null |
| 12. | Location | Tekstowy | 73 lokacje, 14061 każda | 0% null |
| 13. | Type | Tekstowy | 15 typów lokalizacji | 0% null |

* 1. **Fakty – co podlega ocenie**

|  |  |  |
| --- | --- | --- |
| **Lp.** | **Fakt** | **Miary** |
|  | **Model poboru energii w godzinie** | **Data, Lokalizacja, Typ Budynku, Typ Energii** |

* 1. **Kontekst analizy faktów**

|  |  |  |
| --- | --- | --- |
| Lp. | Kontekst analizy - wymiary | Własności |
| 1. | Data | Przedział 01/01 00:00 do 31/12 24:00 – ziarnistość 1h co daje 8070 wpisów na jedną lokalizację na jeden budynek |
| 3. | Miasto | 73 różnych miast z Kalifornii |
| 4. | Typ budynku | 16 różnych budynków o różnym charakterze cieplnym czy wielkościowym. Niejawnie typ budynku zależy od strefy klimatycznej |
| 5. | Typ energii | Oznacza źródło energii dla danej mocy (kW) |

1. **Propozycja tematu projektu**
   1. **Tytuł projektu**

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

* 1. **Problemy**

P01 – rosnąca liczba wypadków samochodowych

P02 – wzrost liczby ofiar śmiertelnych

P03 – niszczenie infrastruktury przez wypadki samochodowe

P04 – nieefektywność regulacji na ograniczenie liczby wypadków

* 1. **Cel przedsięwzięcia**
* Wykrycie przyczyn wypadków drogowych
* Określenie profilu kierowców powodujących wypadki
* Zbadanie wpływu wieku pojazdu na wypadki
* Zbadanie wpływu pogody na wypadki
  1. **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.

* 1. **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 |

* 1. **Ocena jakości danych źródłowych**

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 | 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 |

* 1. **Fakty – co podlega ocenie**

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

* 1. **Kontekst analizy faktów**

|  |  |  |
| --- | --- | --- |
| Lp. | Kontekst analizy - wymiary | Własności |
| 1. | Wiek pojazdu | -1 do 111 lat, ziarnistość 1 rok, definiuje, jak stary pojazd jest |
| 2. | **Wiek kierowcy** | -1 do 100 lat, ziarnistość 1 rok, definiuje, ile lat miał kierowca w momencie wypadku |
| 3. | **Warunki pogodowych** | Wartość od 1 do 9, określający stan pogody. Wartości podane w tabeli Tabela 1 Warunki Pogodowe |
| 4. | **Jakość drogi** | Typ drogi. Połączenie kilku wartości, np. czy droga jednokierunkowa, czy autostrada, czy polna |
| 5. | Ograniczenie prędkości | Ograniczenie prędkości na drodze, poprawne wartości to 20, 30, 40, 50, 60 lub 70 mil na godzinę |
| 6. | **Liczby poszkodowanych** | Ile osób ucierpiało w wyniku wypadku? Składa się z liczy powiązań między tabelą casualities i Accidents po kluczu głównym |
| 7. | **Poważność** | Każda ofiara ma przypisaną poważność, w skali od 1 (not-injured) do 3 (fatal) |
| 8. | **Warunki drogowe** | Warunki drogowe, jakie panowały na drodze. Wpisy w tabeli Tabela 2 Warunki drogowe |

# Podsumowane i wnioski

W pierwszym momencie wydawało mi się, że temat o Poborze Prądu będzie ciekawszy i będzie więcej danych. Ale po dokładnej analizie okazało się, że dane nie są tym, czym mi się wydawały. Myślałem, że będą bardziej szczegółowe, jakieś realne czy powiązane, a okazało się, że jest to tylko model. Według mnie analizowanie tylko modelu pod kątem hurtowni danych nie ma sensu, bo są to dane już zagregowane, przetworzone i jeszcze dodatkowo matematycznie przeliczone.

Natomiast drugi temat o wypadkach jest bardziej „surowy” i według mnie bardziej nadaje się na analizę pod kątem hurtowni danych. Dodatkowo, temat wydaje mi się dużo bardziej przydatniejszy i konkretny w tym zakresie, ponieważ opiera się na danych rzeczywistych i pozwala wyciągnąć konkretne dane.

# Decyzja

Wybrany temat: Analiza wypadków samochodowych w Wielkiej Brytanii w latach 2005 -2015 pod kątem wieku pojazdu, wieku kierowcy, warunków pogodowych, jakości drogi oraz ograniczenia prędkości

Załączniki

|  |  |
| --- | --- |
| 1 | Fine no high winds |
| 2 | Raining no high winds |
| 3 | Snowing no high winds |
| 4 | Fine + high winds |
| 5 | Raining + high winds |
| 6 | Snowing + high winds |
| 7 | Fog or mist |
| 8 | Other |
| 9 | Unknown |
| -1 | Data missing or out of range |

Tabela 1 Warunki Pogodowe

|  |  |
| --- | --- |
| 1 | Dry |
| 2 | Wet or damp |
| 3 | Snow |
| 4 | Frost or ice |
| 5 | Flood over 3cm. deep |
| 6 | Oil or diesel |
| 7 | Mud |
| -1 | Data missing or out of range |
| 9 | unknown (self reported) |

Tabela 2 Warunki drogowe