

## Hurtownie danych – Projekt – Etap 01

PWr. WIZ, Data: 11-12.04.2022

Student	-----	Ocena
Indeks	<u>256305</u>	
Imię	<u>Grzegorz</u>	
Nazwisko	<u>Dzikowski</u>	

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

### Uzasadnienie wyboru tematu projektu

#### 1. Propozycja tematu projektu

##### 1.1. Tytuł projektu

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

##### 1.2. Problemy

Predykcja poboru prądu w poszczególnych godzinach w celu odpowiedniego dostosowania generacji prądu

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

Estymacja kosztów i zysków przez konkretnych dostawców

##### 1.3. 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

##### 1.4. 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

## 1.5. 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

## 1.6. 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

### 1.7. Fakty – co podlega ocenie

Lp.	Fakt	Miary
1.	Model poboru energii w godzinie	Data, Lokalizacja, Typ Budynku, Typ Energii

### 1.8. 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)

## **2. Propozycja tematu projektu**

### **2.1. 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

### **2.2. Problemy**

P01 – rosnąca liczba wypadków samochodowych

P02 – wzrost liczby ofiar śmiertelnych

P03 – niszczenie infrastruktury przez wypadki samochodowe

P04 – nieefektywność regulacji na ograniczenie liczby wypadków

### **2.3. Cel przedsięwzięcia**

- Wykrycie przyczyn wypadków drogowych
- Określenie profilu kierowców powodujących wypadki
- Zbadanie wpływu wieku pojazdu na wypadki
- Zbadanie wpływu pogody na wypadki

### **2.4. 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, warunków pogodowych, profil kierowcy czy typ pojazdu.

## 2.5. 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.	Casualties	.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

## 2.6. 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

## 2.7. Fakty – co podlega ocenie

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

## 2.8. 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.	<b>Liczby poszkodowanych</b>	Ile osób ucierpiało w wyniku wypadku? Składa się z liczy powiązań między tabelą casualties i Accidents po kluczu głównym
7.	<b>Poważność</b>	Każda ofiara ma przypisaną poważność, w skali od 1 (not-injured) do 3 (fatal)
8.	<b>Warunki drogowe</b>	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*