**Electric Vehicle Population**

Analiza Statystyczna Zestawu Danych

Kacper Dusza, Oscar Borowiec, Radosław Cebula, Łukasz Babiś

Spis Treści:

1. Opis Zbioru Danych
2. Opis Cech
3. Analiza Cech
4. Analiza Współzależności
5. Opis Zbioru Danych

Zestaw danych przedstawia pojazdy elektryczne zasilane akumulatorami (BEV) i hybrydowe pojazdy elektryczne typu plug-in (PHEV), które są obecnie zarejestrowane w Departamencie Licencji Stanu Waszyngton (DOL). Jako że zestaw ten zawiera ponad 10000 rekordów, do celów projektu postanowiliśmy go okroić, stosując kryterium cechy „**Base MSRP”** (Podstawowa sugerowana cena detaliczna producenta). W konsekwecji do analizy trafiły tylko te dane w których ta cecha mieściła się w zakresie **89100 – 184400.** W ten sposób uzyskaliśmy **99 rekordów.**

Link:

https://www.kaggle.com/datasets/sahirmaharajj/electric-vehicle-population?resource=download

1. Opis Cech

Orginalna Baza „Electric Vehicle Population” zawiera 17 cech. Z nich wybraliśmy 2 cechy jakościowe i 3 cechy ilościowe:

**Cechy jakościowe:**

1. **Electric Vehicle Type** – Zmienna posiadające 2 wartości, określający typ pojazdu.
2. **Clean Alternative Fuel Vehicle (CAFV) Eligibility –** Zmienna posiadająca 3 wartosci, określająca kwalifikowalność do pojazdów zasilanych paliwem alternatywnym

**Cechy ilościowe:**

1. **Electric Range –** Zmienna określająca jak daleko pojazd może przejechać, korzystając wyłącznie z zasilania elektrycznego. **Skala ilorazowa.**
2. **Base MSRP –** Zmienna określająca Podstawową sugerowaną cenę detaliczną producenta. **Skala ilorazowa.**
3. **Model Year** – Zmienna określająca rok modelowy pojazdu, ustalony na podstawie rozkodowania numeru identyfikacyjnego pojazdu (VIN). **Skala interwałowa.**
4. Analiza Cech

**Cechy jakościowe**

Obraz zawierający tekst, zrzut ekranu, diagram, krąg

Opis wygenerowany automatycznie

**Cechy ilościowe**

Obraz zawierający tekst, zrzut ekranu, numer, Czcionka

Opis wygenerowany automatycznie

-**Model Year**

Porównując średnią, odchylenie standardowe, minimum i maksimum widzimy że wykres jest daleko do rozkładu normalnego. Cecha przyjmuje cechy z całego zakresu z podobną częstotliwością. Niska skośność oznacza duża symetryczność. Minusowa wartość kurtozy oznacza że ogony wykresu byłyby lżejsze niż w rozkładnie normalnym

**Electric Range**

Kwartyl I i mediana są takie same, czyli mamy sporo wartości równych 14. Bardzo duża różnica między średnią a medianą, plus duże odchylenie standardowe w stosunku do średniej, oznaczają że sporo jest wartości skrajnych. To samo oznacza spora, ujemna kurtoza. Skośność bliska 0 oznacza że obydwa ogony wykresy miałyby podobną długość

**Base MSRP**

Spora dodatnia skośność oznacza że wykres byłby odchylono w lewą stronę, a prawy ogon byłby wydłużony. Spora dodatnia kurtoza, czyli dużo jest wyników w pobliżu średniej.

1. Analiza Współzależności

**Cechy jakościowe**

**Test Chi Kwadrat:**

Obraz zawierający tekst, zrzut ekranu, Czcionka, numer

Opis wygenerowany automatycznie

**Współczynnik V-Cramera:**

Obraz zawierający tekst, pismo odręczne, papier, Produkty papierowe

Opis wygenerowany automatycznie

Współczynnik ma wartość 1, co oznacza silną zależność pomiędzy obydwoma zmiennymi jakościowymi. Wartość cechy Plug-in Hybrid Electric Vehicle (PHEV) pociąga za sobą wartości cechę Not eligible due to low battery range I Eligibility unknown as battery range has not been researched. Zależność ta jest łatwa to wytłumaczenia – samochody Hybrydowe (PHEV) mają w założeniu służyć do przejeżdżania krótkich odcinków, nie można ich przez to zakwalifikować jako samochody korzystające ściśle z źródeł paliwa określanych jako alternatywne.

**Cechy ilościowe**

Obraz zawierający zrzut ekranu, Wykres, linia

Opis wygenerowany automatycznie



Istnieje negatywna korelacja między zasięgiem a rocznikiem samochodu, co oznacza że mniejszy zasięg często oznacza nowsze auto. To prawdopodobnie oznacza że w ostatnich latach w tym przedziale cenowym popularniejsze od aut elektrycznych są hybrydy, które generalnie mają mniejszy zasięg.



Istnieje bardzo mała korelacja między ceną o rocznikiem.