

Ad. 3:

- Dominują samochody o niskiej i średniej mocy i małej i średniej wielkości silnika (horsepower między 100, a 230 i engine_size między 1,92, a 4,5),
- Większość stanowią samochody tanie (price między 11, a 30 i średnia ok. 27),
- W zbiorze danych najczęściej znajduje się samochodów marki Dodge i Honda,
- Znacznie więcej jest samochodów typu 0 niż 1,
- Bardzo rzadko sprzedaż modelu była wysoka (sales głównie między 14, a 71 i średnia na poziomie ok. 53),
- resale również zwykle osiągało niskie wartości (głównie między 10, a 19 i średnia na poziomie ok. 18),
- Na wykresie zmiennej model każdy pojawia się raz poza Neon - ten ma wartość dwa,
- Większość stanowią małe i średnie auta (width w większości między 66, a 73 i length między 175, a 194),
- W przypadku zmiennych zscore również największą liczebność mają niskie i średnie wyniki

Ad. 5:

Wykorzystane będą:

- sieć Kohonena,
- metoda K-średnich,
- dwustopniowa analiza skupień

Ad. 6:

Mamy do dyspozycji trzy modele:

- W dwustopniowej analizie skupień powstały trzy grupy. Grupa 1 i 2 mają dość podobne rozmiary (44 i 45), a grupa 3 nieco mniejsza (rozmiar równy 28). Proporcja rozmiarów największej grupy do najmniejszej to 1,61. Jakość grupy została oceniona na dobrą.
- W K-średnich otrzymujemy 5 grup. Grupa 1 i 5 są największe (rozmiary równe 61 i 45). Pozostałe grupy mają rozmiary 30, 11 i 10. Proporcja rozmiarów największej grupy do najmniejszej to 6,1. Jakość grupy została oceniona na poprawną.
- W sieci Kohonena otrzymujemy 8 grup o rozmiarach 20, 11, 22, 26, 22, 14, 21 i 21. Proporcja rozmiarów największej grupy do najmniejszej to 2,36. Jakość grupy została oceniona na poprawną.

Ad. 7c:

- Dwustopniowa analiza skupień: Najważniejszym predyktorem jest ztype (ważność równa 1). Na drugim miejscu jest type z poziomem ważności zaledwie 0,32. Najmniej ważnym insales (ważność równa 0,04).
- K-średnich: Najważniejsze predyktory to zscore: fuel capacity i fuel capacity (ważność równa 1). Nieco mniej ważne są horsepower i zscore: horsepower (ważność równa 0,85). Najmniej ważny jest log-transformed sales (ważność równa 0,18). Widać tu bardziej równomierny rozkład ważności predyktorów niż w modelu wyżej.
- Sieci Kohonena: Najważniejsze predyktory to zscore: fuel capacity i fuel capacity (ważność równa 1). Na kolejnych miejscach są zscore: curb weight i curb weight (ważność równa 0,94), horsepower i zscore: horsepower

(ważność równa 0,9). Najmniej ważny jest ponownie log-transformed sales (ważność równa 0,14). Tu również jest bardziej równomierny rozkład ważności predyktorów niż w przypadku metody z pierwszego punktu.

Ad. 7d:

- Dwustopniowa analiza skupień: W grupie 2 pojawiają się wyniki między 14, a 25. Najczęściej pojawiają się wartości ze środka przedziału wartości fuel capacity. W grupie 1 pojawiają się wyniki między 0, a 20. Najczęściej pojawiają się wyniki z początku przedziału wartości fuel capacity. W grupie 3 pojawiają się wyniki między 14, a 34. Najczęściej pojawiają się wyniki z końca przedziału wartości fuel capacity.
- K-średnich: W grupie 1 wyniki z przedziału 0 do 22. Najczęściej pojawiają się wyniki z początku przedziału wartości fuel capacity. W grupie 5 wyniki między 10, a 23. Obejmują w większości wyniki ze środka przedziału wartości fuel capacity. W grupie 4 wyniki między 13, a 26. Najczęściej pojawiają się wyniki około 20, jednak rzadziej niż w przypadku grupy 5. W grupie 2 wyniki z przedziału 23, a 34. Najczęściej występują wyniki z końca przedziału fuel capacity. W grupie 3 wyniki z przedziału 15, a 25. Obejmuje głównie wyniki ze środka przedziału wartości fuel capacity, lecz częstość jest mniejsza niż w przypadku grupy 4 i 5.
- Sieć Kohonena: Grupa X=0, Y=2 - wyniki między 0, a 16. Najczęściej z początku wartości fuel capacity. Grupa X=0, Y=1 - wyniki między 11, a 20. Najczęściej pojawiają się wyniki około 16. Grupa X=1, Y=0 - wyniki między 15, a 21. Najczęściej około 18. Grupa X=2, Y=2 - wyniki między 14, a 23. Najczęściej około 20. Grupa X=3, Y=0 - wyniki między 10, a 25. Najczęstsze wyniki w okolicach 18. Grupa X=3, Y=2 - wyniki między 19, a 34. Najczęściej z końca przedziału wartości fuel capacity. Grupa X=2, Y=0 - wyniki między 15, a 23. Najczęściej około 18. Grupa X=0, Y=0 - wyniki między 14, a 20. Najczęściej około 16.

Ad. 7e:

Dla każdej zmiennej poza mpg i zmpg wartości w grupie 3 są wyższe niż w grupie 1. Dla zmiennych horsepower i zhorsepower, price, zprice łączna mediana jest zbliżona do mediany grupy 3. Dla zmiennej insales mediana łączna jest zbliżona do mediany dla grupy 1. Dla grupy 3 dla zmiennych zwheelbas, wheelbas mediana jest zbliżona do łącznego trzeciego kwartyłu. Dla grupy 1 dla zmiennych zresale, zlength, zwidth, width, price, zprice, length mediana jest zbliżona do łącznego pierwszego kwartyłu.

Ad. 8:

Pojawia się wiele pól \$null\$. Dzieje się tak, gdy w którymś z pól dla samochodu pojawiła się wartość \$null\$. W grupie 1 pojawiają się samochody małe, w grupie 2 średnia, a w grupie 3 największe.