

Használtautó adatok elemzése

Balázs Frigyes, Nagy Balázs

Elméleti keret

Használt autók értéke → Jövőbeni hasznosságuk jelenértéke
(Kooreman, P., & Haan, M. A., 2006)

Mennyi ideig
használható

Milyen jó

Változók kiválasztása

DE információs asszimetria,
nem racionális viselkedés miatt messze nem tökéletes piac

Fontosnak ítélt változók

Kor/Évjárat

Teljesítmény

Kivitel

Megtett Km-k száma

Állapot

Üzemanyag

Kiszűrt változók

Tető típusa

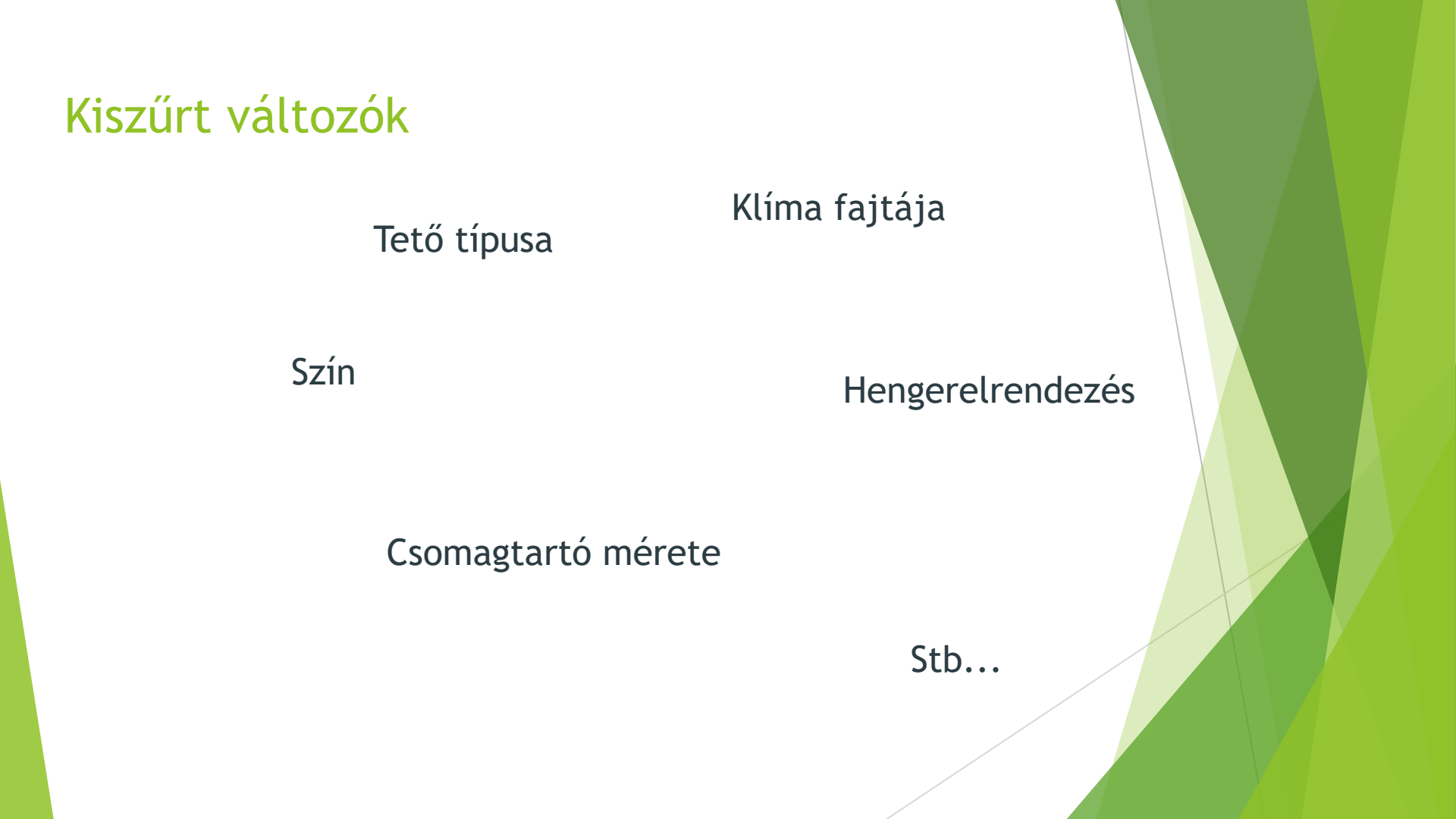
Klíma fajtája

Szín

Hengerelrendezés

Csomagtartó mérete

Stb...

The background of the slide features abstract, overlapping green geometric shapes, primarily triangles and polygons, in various shades of green, creating a modern and dynamic visual effect.

Adatok mennyisége

- 80253 megfigyelés
- Outlierek
- Hosszú futási idő
- Gyakran csak az adatok töredékét (1000-10000) használtuk
- Változókból is csak a legfontosabbakat használtuk

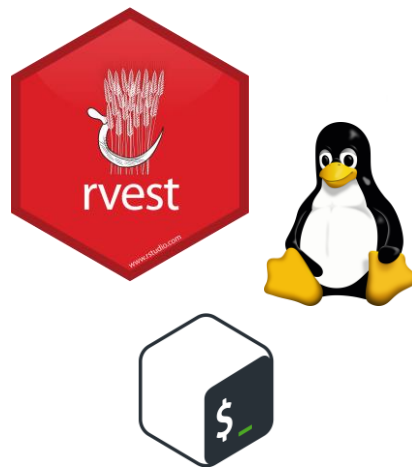
Adatok forrása

Hasznáلتautó●hu

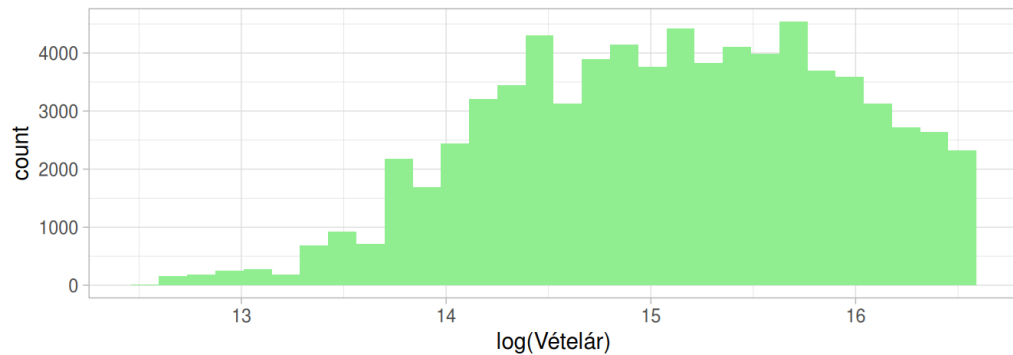
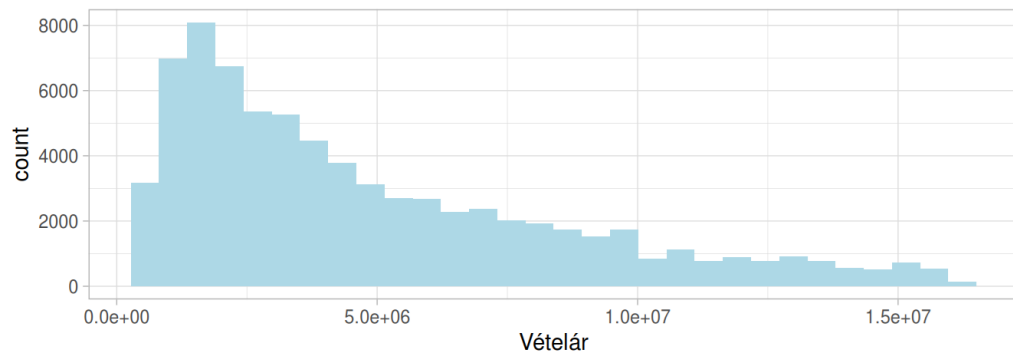
- hasznaltautok.hu
- Október eleji megfigyelések
- Bash, wget egy vpn és Rvest használatával

Nehézségek

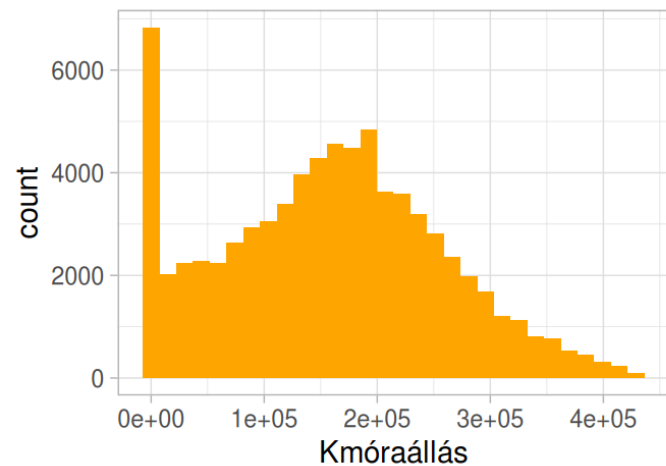
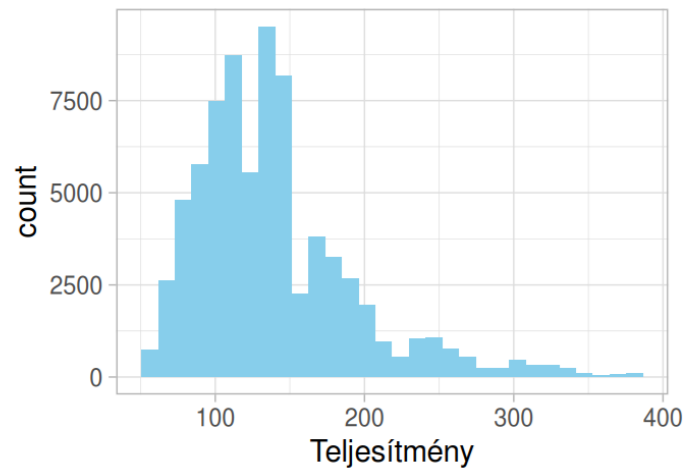
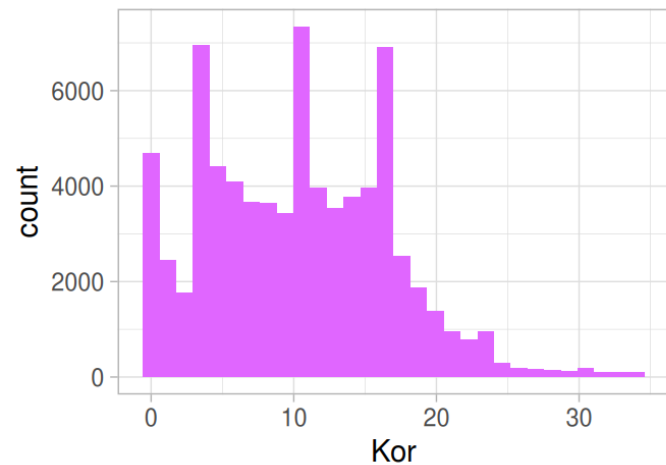
- Rate limit, lassú
- Ip címek letiltása
- HTML adathibák



Elemzett adatok leírása



	Vételár	Teljesítmény	Kmóraállás	Kor
Vételár	1.00	0.34	-0.49	-0.64
Teljesítmény	0.34	1.00	0.02	-0.10
Kmóraállás	-0.49	0.02	1.00	0.51
Kor	-0.64	-0.10	0.51	1.00



Állapot		Kivitel	Üzemanyag	
Kitűnő	:21819	Ferdehátú	:22244	Benzin :38124
Megkímélt	:18107	Kombi	:14798	Dízel :31384
Motorhibás	: 0	Városi terepjáró (crossover)	:13684	Egyéb : 161
Normál	:24713	Sedan	: 8391	Elektromos: 2119
Sérülésmentes	: 3563	Egyterű	: 8372	Hibrid : 2807
Újszerű	: 6393	Coupe	: 2374	
		(Other)	: 4732	

Lineáris Regresszió

Referencia modell

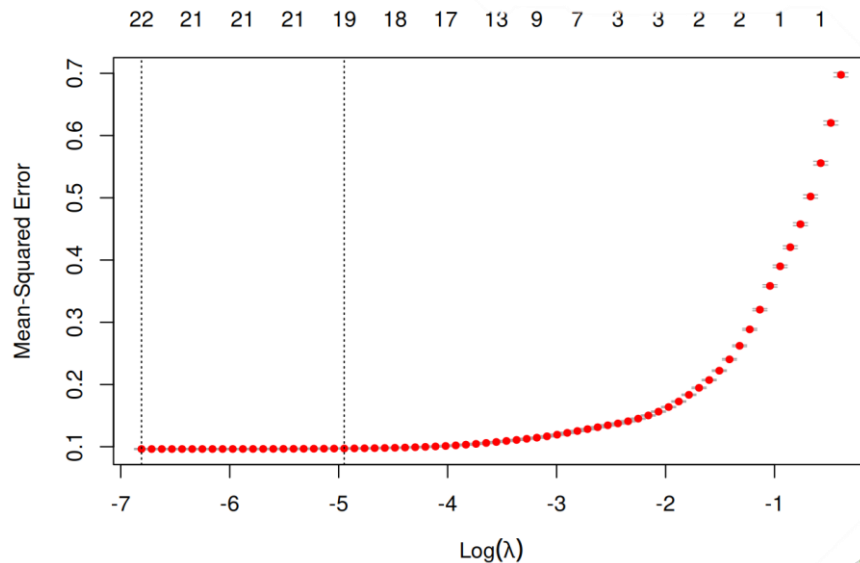
$R^2 = 0,863$

Szignifikáns változók:

- Kor (negatív),
- Kilométeróra állása (negatív),
- Állapot,
- Üzemanyag,
- Teljesítmény,
- Kivitel egyes kategóriái

LASSO regresszió

- R-négyzet: 0,861
- Hasonló a lineárishez
- A kategorikus változók pár kategóriájára nem becsült Bétát
- A kiválasztott lambda: 0,0071



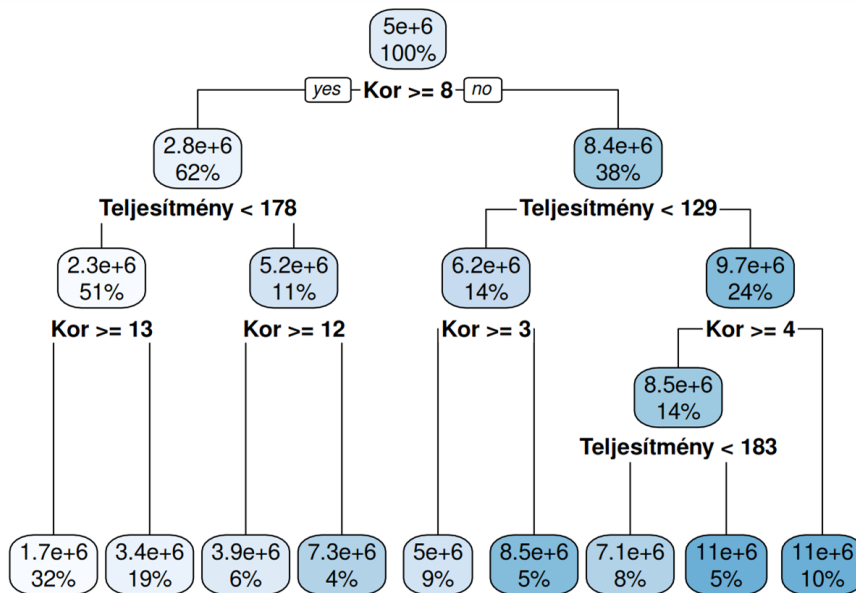
MARS regresszió

- R-négyzet: 0,896
- Teljesítmény, Kilóméteróra állása fontos változók, sok interakcióba bekerültek
- Meredekségváltozások:
 - Kor: 5, 10, 21 évnél
 - Megtett kilométerek: 40000 km-nél (39395 km)
 - Teljesítmény: 73 és 192 lóerőnél



Döntési fa

- R négyzet értéke: 0,77 (teszt adaton)
- Keresztvalidált hiperparaméter optimalizálással

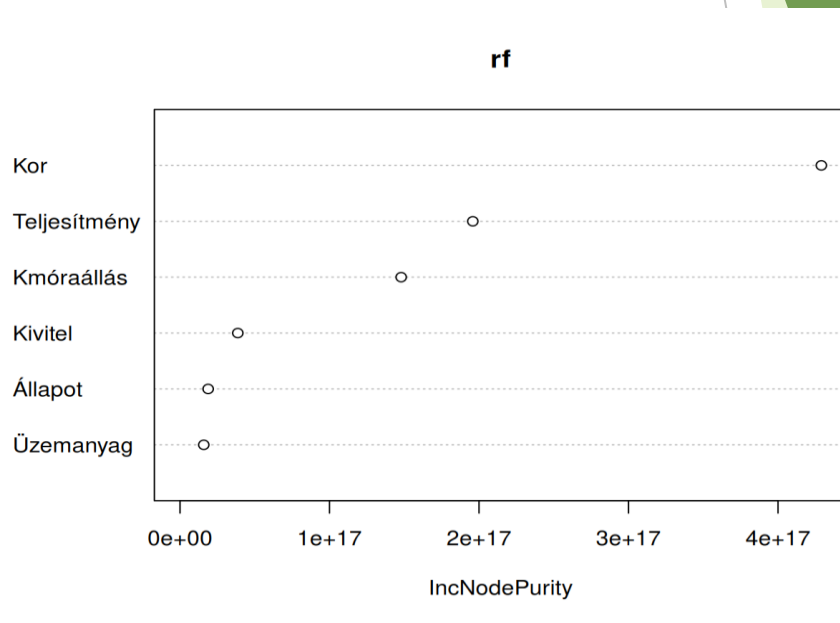


Véletlen erdő

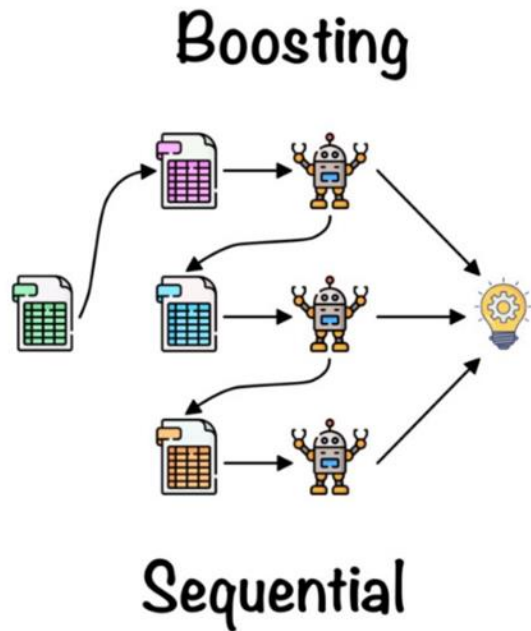
- R négyzet értéke: 0.9178 (teszt adaton)
- Eddigi legjobb modell
- Ugyanazok a fontos változók
- Hosszú futás (~20-30 perc)



- nem optimalizáltuk
- 100 fás erdő



Boosting



- R négyzet a teszt adatokon: 0.9162
- Nagyjából olyan mint a véletlenerdő
- Mivel adaboost klasszifikációs, ezért másikat használtunk
- XGBoost a boosting modellünk (100 iterációval)

Főkomponens elemzés

Kaiser-kritérium alapján 3 főkomponens felhasználása indokolt

Ez a teljes variancia 61,24%-át őrzi meg

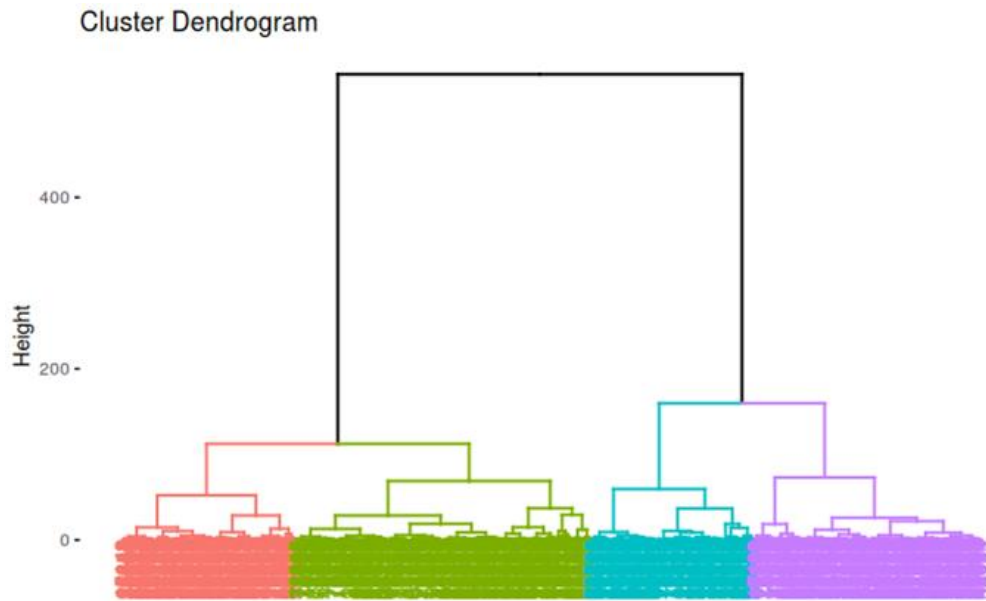
A három főkomponens:

1. mennyire jó az autó
2. új de nem kiemelkedő minőségű
3. autó méret (nem busz)

	1. főkomponens	2. főkomponens	3. főkomponens
Vételár	0,411	0,42	
Km óraállás		-0,591	
Szállítható személyek száma			-0,715
Teljesítmény	0,526	-0,118	0,224
Csomagtartó méret	0,176		-0,62
Tömeg	0,508	-0,167	
Hengerűrtartalom	0,465	-0,327	0,172
Kor	-0,199	-0,569	

Hierarchikus klaszterezés

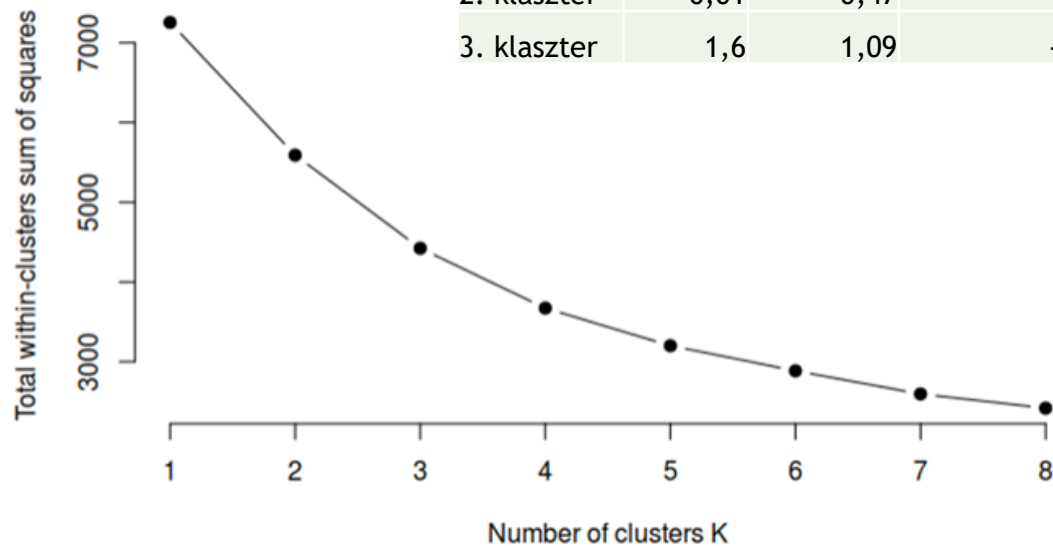
- Nehezen értelmezhető az ábra
- Csak numerikus változókat használ, így fontos változókat kihagy



K-középpontos klaszterezés

3 középpont mellett döntöttünk

	Vételár	Évjárat	Km óraállás	Teljesítmény
1. klaszter	-0,68	-0,82	0,62	-0,13
2. klaszter	0,01	0,47	-0,24	-0,06
3. klaszter	1,6	1,09	-1,01	0,51



K-prototípus klaszterezés

Használ kategorikus és numerikus változókat is

	Vételár	Évjárat	Km óraállás	Teljesítmény	Állapot	Kivitel	Üzemanyag
1. klaszter	0,01	0,54	-0,41	-0,18	Kitűnő	Ferdehátú	Benzin
2. klaszter	1,63	0,87	-0,7	0,65	Kitűnő	Városi terepjáró	Dízel
3. klaszter	-0,67	-0,81	0,58	-0,12	Normál	Ferdehátú	Benzin

Hasznáلتautó 🤔 hu

Ha ezzel az üzenettel találkozol, akkor a jelenleg használt IP-címről (80.99.193.105) korlátoztuk a weboldalunk elérhetőségét, mert korábban robotokra utaló (automatikus gépi) hozzáférést érzékeltünk. Ha erről az IP címről szeretnéd az oldal böngészését folytatni, kérjük küldd el ezt az IP-címet az ugyfelszolgalat@hasznaltauto.hu e-mail címre, ahol munkatársaink segítséget nyújtanak ehhez! Az adatkezelésre vonatkozóan bővebb információ az [Adatvédelmi tájékoztatóban](#) található.

Köszönjük a figyelmet!

