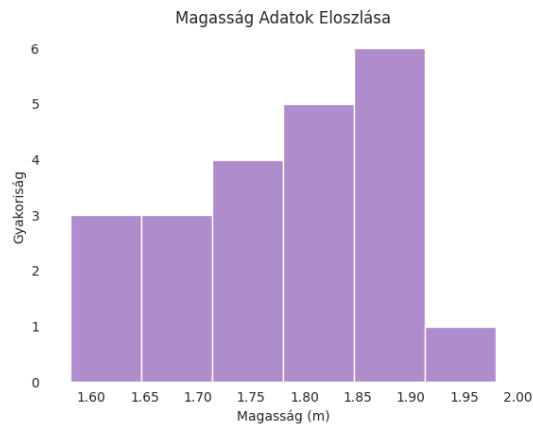


BMI Adatok Elemzése

- Rafain Emőke, Data Science –

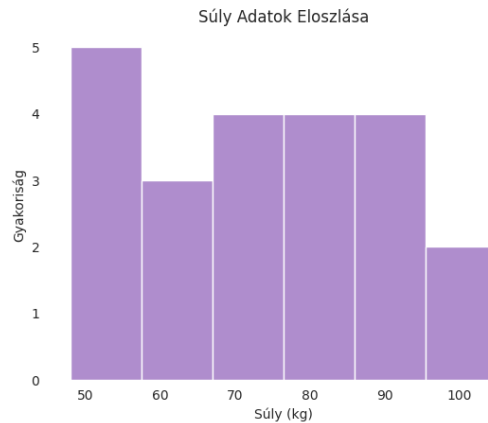
Színmagyarázat: **Saját adatok**; **Kaggle adatok**

1. Magasság Adatok Eloszlása (Saját Adatok):



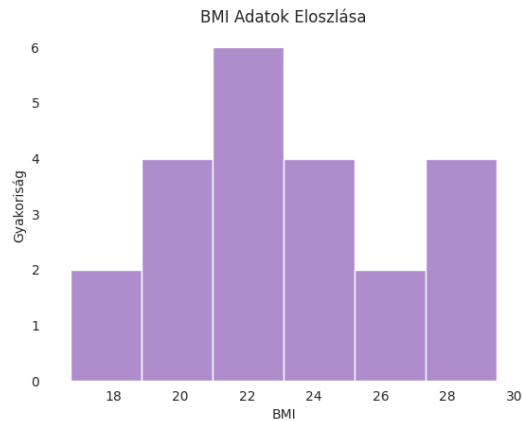
A magasságok eloszlása azt mutatja, hogy a legtöbb megfigyelt ember magassága 1.70-1.90 m közötti. A szélsőséges magasságok (nagyon alacsony vagy nagyon magas értékek) ritkábbak.

2. Súly Adatok Eloszlása (Saját Adatok):



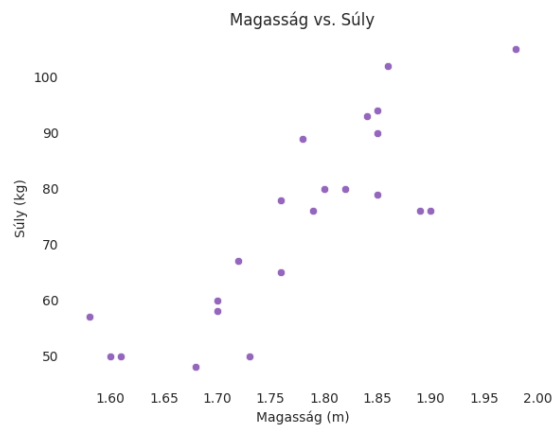
A súly adatok eloszlása azt mutatja, hogy a megfigyelt embereknél a leggyakoribb értékek 50 kg körüliek, azonban majdnem ugyanolyan számban megtalálhatóak az 60 és 90 körüli súlyok is. Kevesebb az olyan megfigyelés ahol a súly 100 kg vagy több.

3. BMI Adatok Eloszlása (Saját Adatok):



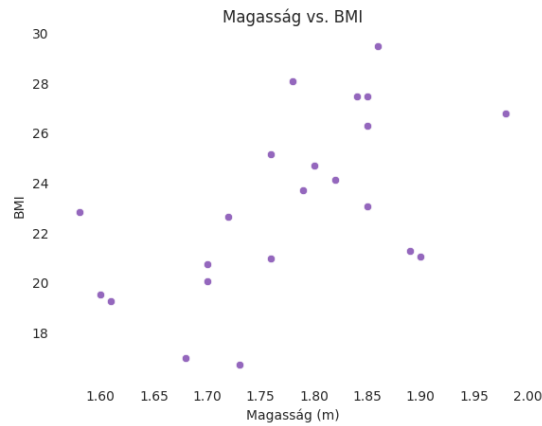
A megfigyelt adatok közül a legnagyobb gyakoriság a 22-es BMI értéknél látható, a vizsgált emberek többsége ebbe a kategóriába esik (az egészséges BMI értékek 18.5-24.9 közé esnek, <https://www.betterhealth.vic.gov.au/tools/body-mass-index-calculator-for-adult>). A BMI eloszlás viszonylag normál eloszlást mutat. A szélsőségesen alacsony vagy magas BMI értékek kevésbé gyakoriak, a 28-as érték feletti értékeket kivéve, amelyek szintén gyakoriak.

4. Magasság vs. Súly (Saját Adatok):



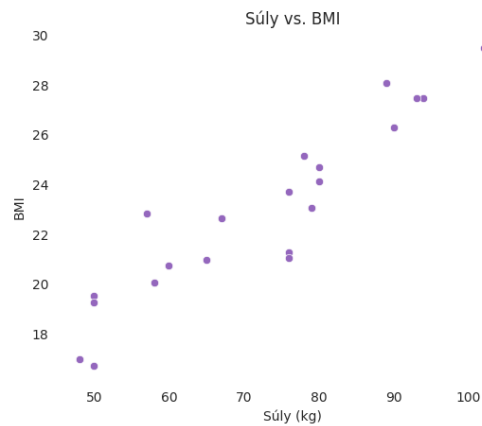
A magasság és súly összefüggését vizsgáló diagram egy erős pozitív korrelációt mutat: minél magasabb valaki, annál nagyobb a súlya.

5. Magasság vs. BMI (Saját Adatok):



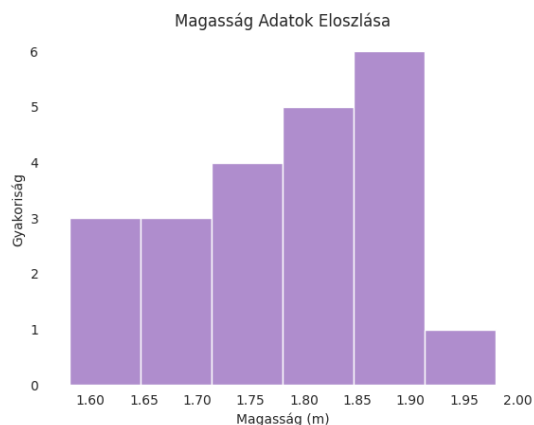
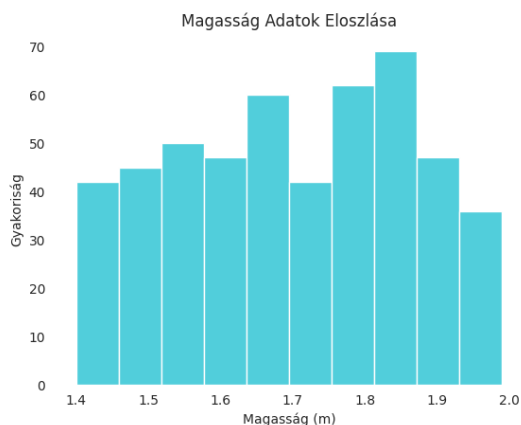
Az összefüggést vizsgáló ábra alapján úgy tűnik, hogy a magasság adatok és a BMI adatok értékei között egy gyenge, nem annyira egyértelmű, pozitív korreláció van (nem annyira egyértelmű mint a magasság és súly esetében). Azt láthatjuk, hogy a magasabb embereknek gyakran magasabb BMI értékei vannak.

6. Súly vs. BMI (Saját Adatok):



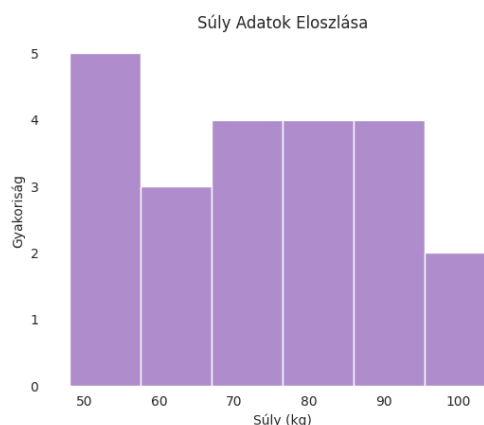
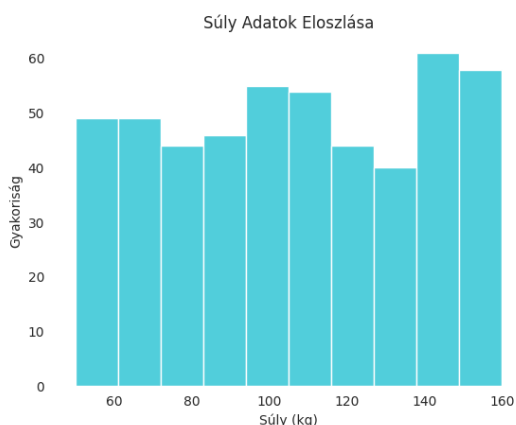
Ezen az ábrán is megfigyelhető az erős pozitív korreláció, ami logikus, hiszen a BMI képlete a súlyt és a magasságot használja fel. A magasabb testtömeghez magasabb BMI tartozik.

7. Magasság Adatok Eloszlása (Kaggle vs. Saját Adatok):



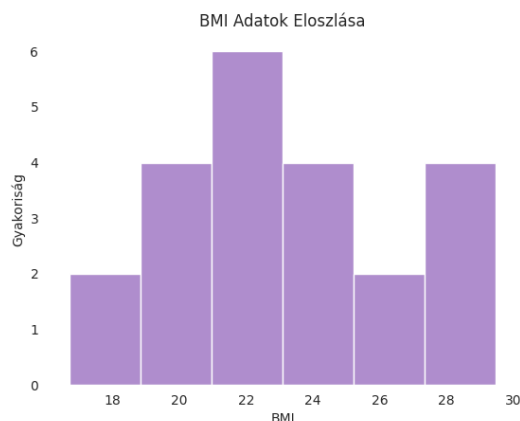
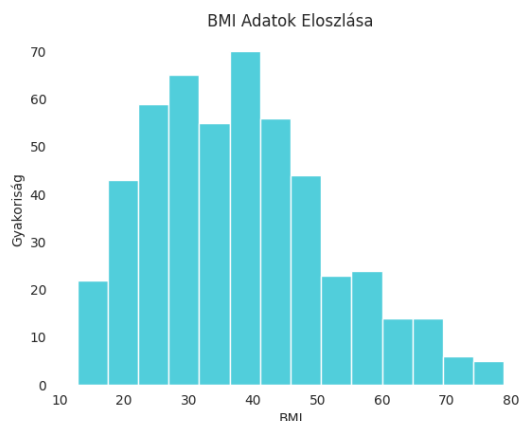
Ami a magasságot illeti, a Kaggle adatok esetében a legtöbb adat 1.65-1.85 m közötti. A saját adatoknál a legtöbb érték 1.70-1.90 m közé esik, itt viszont a szélsőségebb magasságok ritkábbak. Mindkét esetben a leggyakoribb értékek hasonló intervallumokat vesznek fel, azonban a Kaggle adatoknál változatosabbak az adatok.

8. Súly Adatok Eloszlása (Kaggle vs. Saját Adatok):



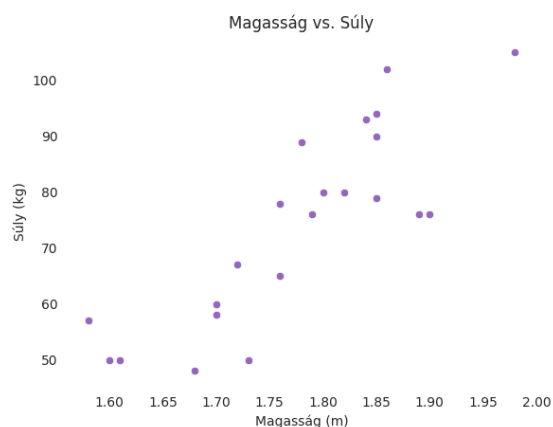
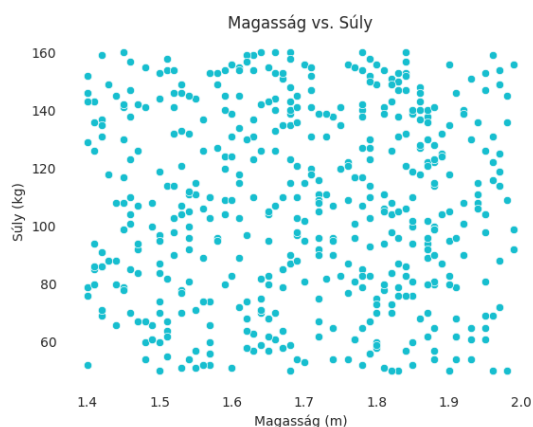
A Kaggle adatsor súlyeloszlását megfigyelve a legtöbb adat 140-160 kg közötti egyénekre vonatkozik, azonban viszonylag egyenletes az eloszlás. Hasonló a saját adatok eloszlása is, mely esetben azonban az 50 kg körüli értékekből van több, az elosztás itt is viszonylag egyenletes. A Kaggle adatok szerint a vizsgált populáció súlyosabb, míg a saját adatok populációja.

9. BMI Adatok Eloszlása (Kaggle vs. Saját Adatok):



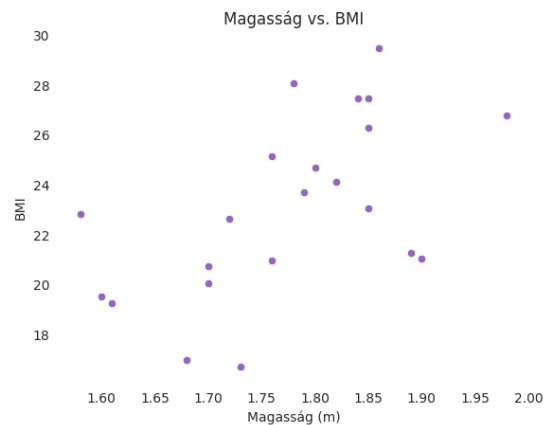
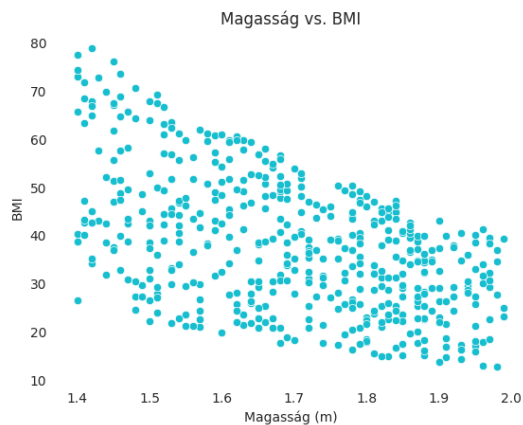
A Kaggle adatokat megfigyelve a BMI adatok eloszlása viszonylag normál eloszlást mutat, ahogy az megfigyelhető a saját adatok esetében is. Azonban a két adatsort figyelve a Kaggle adatok BMI értékei nagyban eltérnek a saját adatok értékeitől, ugyanis az előbbieket esetében több a súlyosabb egyed. A leggyakoribb értékek az egészséges 24.9 BMI feletti értékeket is gyakran felveszik, gyakoriak a 30-nál magasabb értékek, melyek a túlsúlyt jelzik. A saját adatoknál a 22-es BMI volt a leggyakoribb, mely egy egészséges érték, ebben az esetben több is az egészséges súlyú egyed.

10. Magasság vs. Súly (Kaggle vs. Saját Adatok):



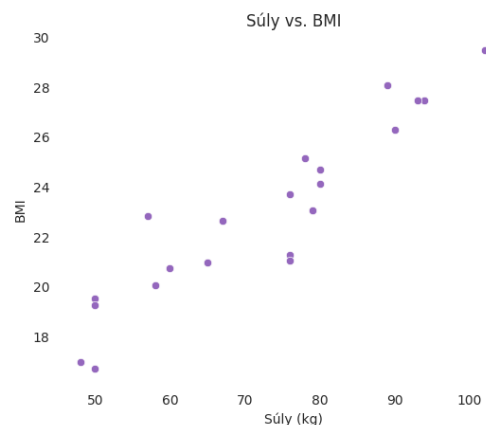
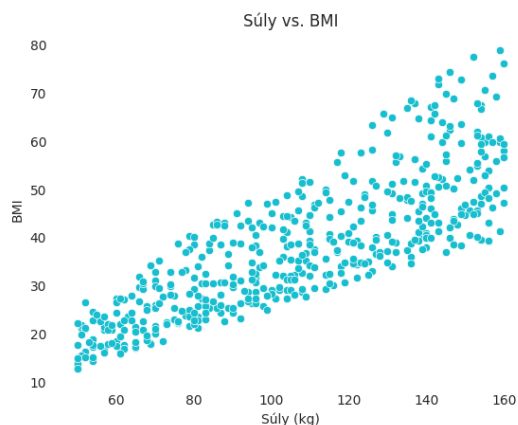
A Kaggle adatok magasság és súly összehasonlítását szemlélve nincs sem pozitív, sem negatív korreláció, az adatok szétszórtabbak. A saját adatoknál viszont egy viszonylag erős pozitív korrelációt láttunk: minél magasabb valaki, annál nagyobb a súlya. A Kaggle adatok a vizsgált személyek sokféleségére utalnak.

11. Magasság vs. BMI (Kaggle vs. Saját Adatok):



A Kaggle adatok esetében a magasság és BMI összehasonlítási ábrája egy viszonylag erős, negatív korrelációt mutat: a magasság növekedésével a BMI értékek csökkennek. A diagram szerint a magasabb emberekhez általában alacsonyabb BMI tartozik. Ezzel ellentétben a saját adatok ábrája egy gyenge pozitív korrelációt mutat, az egyre magasabb megfigyelt egyedek BMI értékei egyre magasabbak. Ez a megfigyelés nagyon is érdekes, mivel a Kaggle adatok által mutatott trend azt feltételezi, hogy az alacsonyabb emberek súlyosabbak, mint a magasabb emberek. Ennek a feltételezésnek pedig ellent mondanak a saját adatok.

12. Súly vs. BMI (Kaggle vs. Saját Adatok):



A Kaggle adatok súly és BMI ábrája ezúttal alátámasztja a saját adatok által megfigyelteket: a súly és a BMI között erős pozitív korreláció van, a súly növekedésével a BMI értéke is növekszik, egy egyértelmű lineáris kapcsolat figyelhető meg.

Általános megfigyelések, következtetések

A Kaggle adatsor inkább reprezentál egy általánosabb, változatosabb populációt, míg a saját adatok egy szűkebb, egészségesebb mintát fednek le. Mind a saját, mind a Kaggle adatok alapján a BMI eloszlás normálisnak mondható, bár a Kaggle adatsorban több túlsúlyos egyén található. A súly és BMI között mindkét esetben erős pozitív korreláció figyelhető meg: ahogy nő a súly, úgy nő a BMI is.