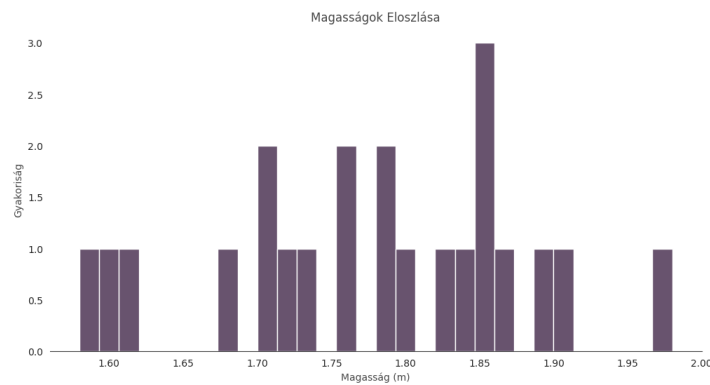


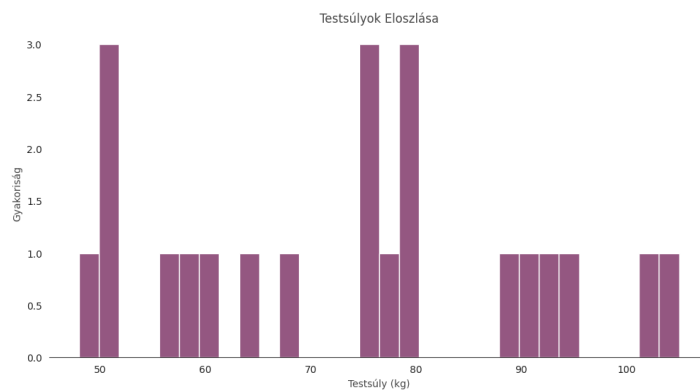
BMI fit – folytatás

Készítette: Póra Boglárka

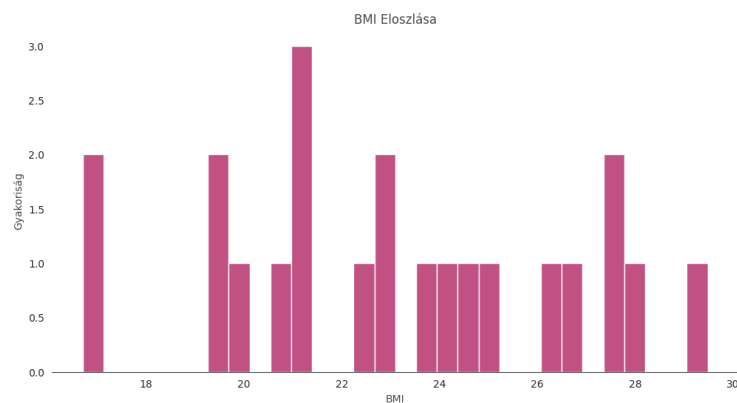
Megfigyelések a **saját adatokról**:



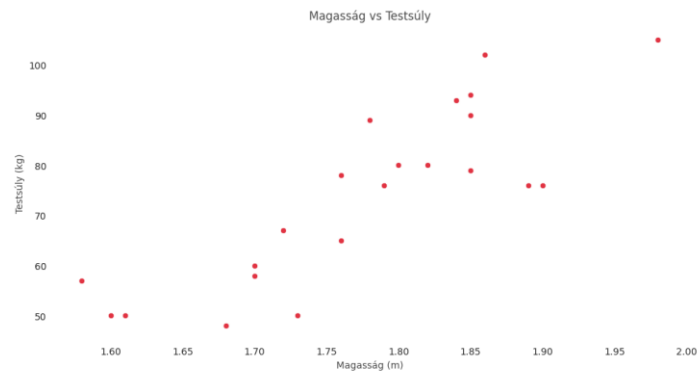
Az adatok szerint a legtöbb megfigyelés 1.70 és 1.85 méter közötti érték. A magasságok eloszlása nem föltetlen szimmetrikus, és viszonylag nagy szóródás figyelhető meg.



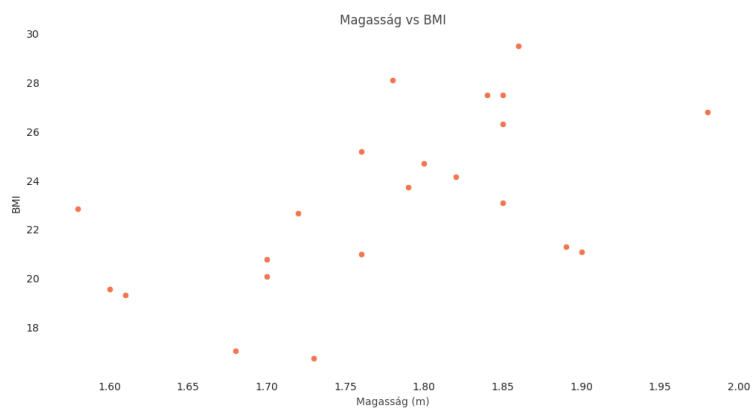
A testsúlyok eloszlásánál is sok féle érték látható. Jól látszik, hogy a 50 kg körüli és a 75-80 kg közötti tartományban van a legtöbb adatpont, de létrejön néhány kisebb csoport is.



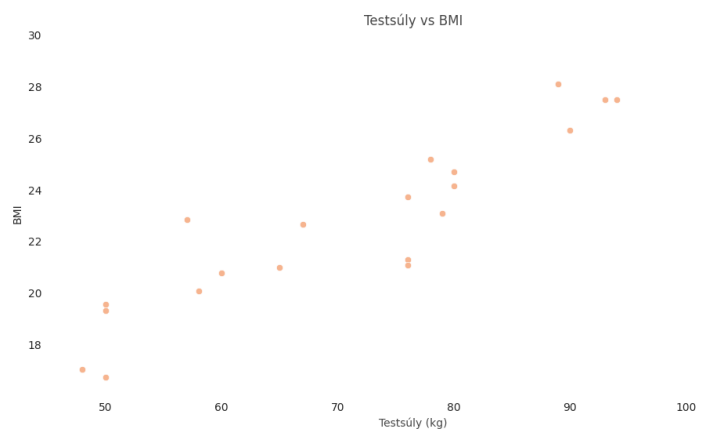
Az BMI eloszlás nagyjából hajaz a normális eloszlásra. Az adatok többsége a 20-24 közötti tartományban helyezkedik el, ami a normális testsúlykategória a magassághoz képest.



A fenti scatterplot szerint pozitív korreláció van a magasság és a testsúly között.



A magasság és BMI közötti scatter diagram kevésbé mutat szoros kapcsolatot. Bár van egy növekvő tendencia, a magasságnak nincs teljesen egyértelmű hatása a BMI-re, mivel a testsúlynak van jelentősebb szerepe.

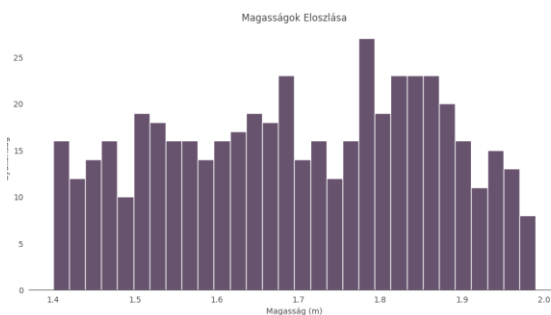
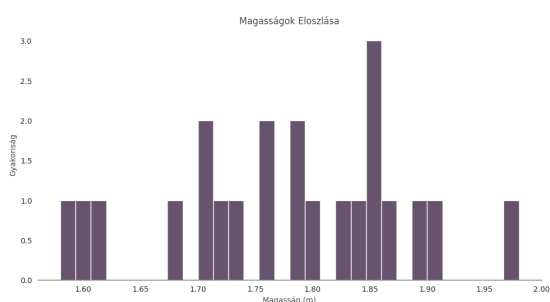


A testsúly és BMI közötti scatterplot egyértelműen erős pozitív kapcsolatot mutat. A BMI egyenesen arányosan emelkedik, ahogy növekszik a súly értéke

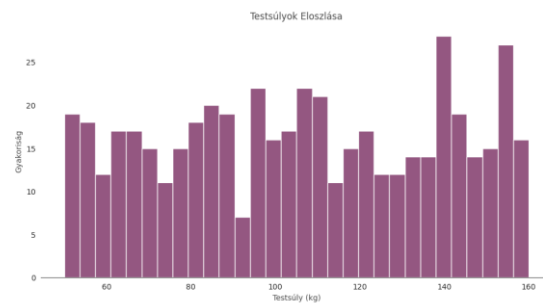
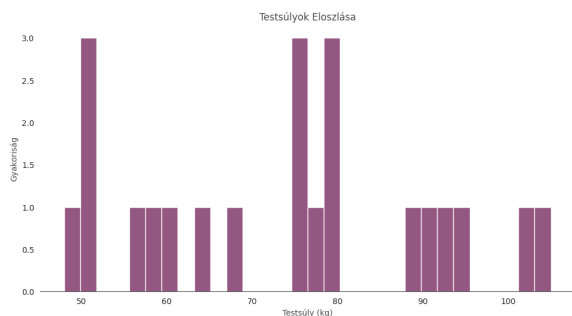
Következtetések:

- A magasság és testsúly átlagos értékeket mutat közepes szélességű tartományban, azonban egyes csoportok esetében kiemelkedhetnek az értékek.
- A testtömegindex adatok többnyire egészséges tartományban vannak, bár néhány adatpont túlsúlyt mutat.

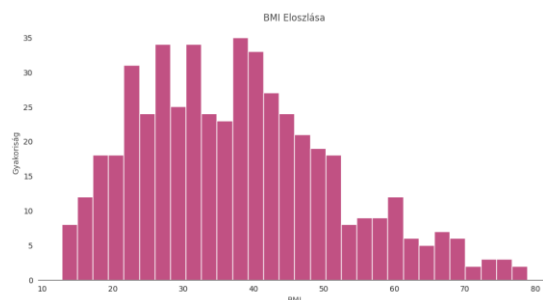
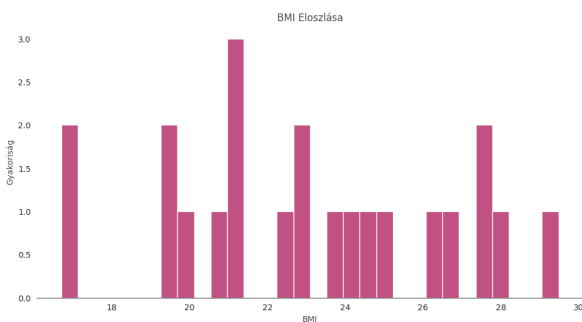
Összehasonlítva a sajátot a **Kaggle adatokkal**, az alábbi következtetéseket vonhatjuk le (bal oldalon saját adatok vizualizációk, jobb oldalon Kaggle adatokból készült plotok):



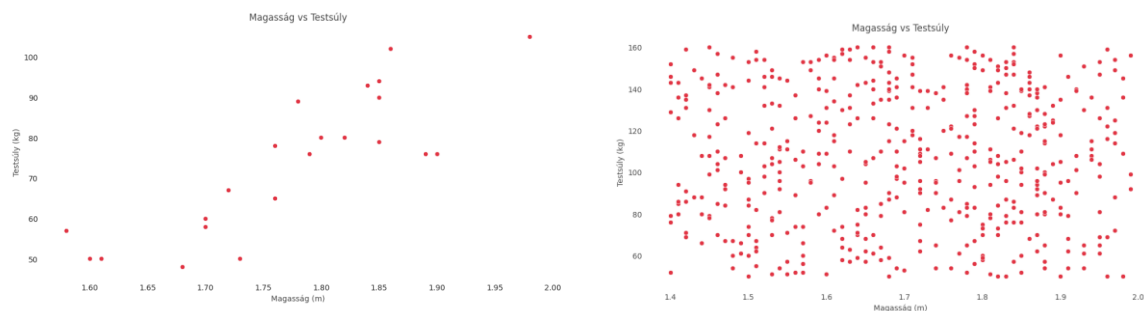
A Kaggle adatok alapján több magasságérték jelenik meg, és az eloszlás egyértelműbben követ egy normál görbét. A különbség valószínűleg abból adódik, hogy a Kaggle adatsor sokkal több adatpontot tartalmaz.



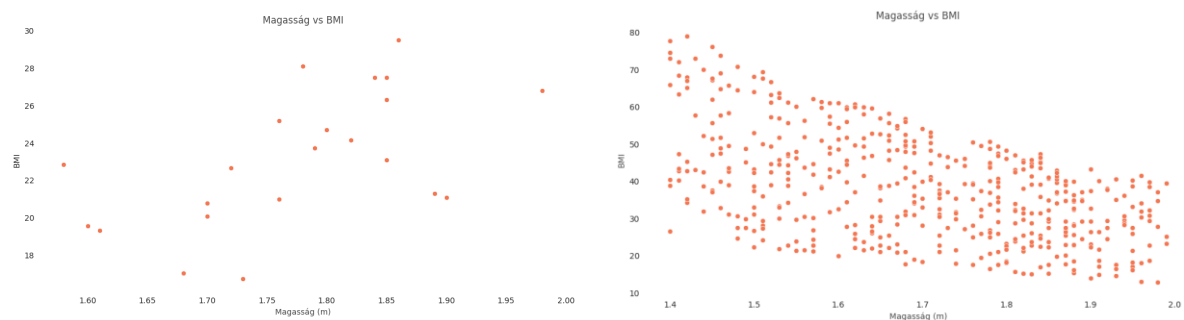
A Kaggle adatoknál a testsúly szélesebb tartományban helyezkedik el. Az eloszlás sokkal egyenletesebb, és nagyobb súlyok is szerepelnek az adatbázisban. Nagyjából hasonló tendenciát mutat, mint a saját adatok.



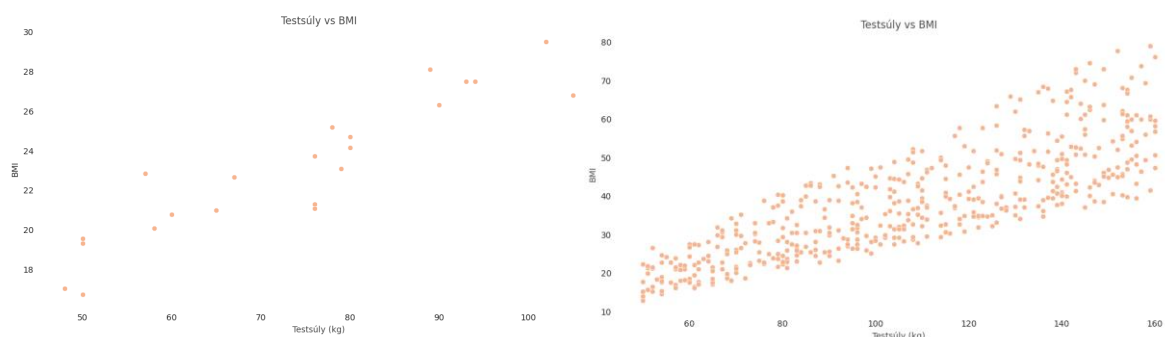
A Kaggle adatok szélesebb skálán mozognak, ahol az elhízás és extrém elhízás kategóriák is megjelennek. Az adatok alapján több súlyosabb eset szerepel, mint a saját adatoknál, viszont az eloszlás hasonló.



A Kaggle adatoknál az összefüggés a magasság és súly között kevésbé egyértelmű. A testsúly nem mutat olyan erős kapcsolatot a magassággal. Nem olvasható ki a pozitív korreláció, mint a saját adatoknál.



A Kaggle adatoknál egyértelmű negatív korreláció figyelhető meg: alacsonyabb embereknél nagyobb a BMI, míg magasabb embereknél kisebb. Habár a saját adatok esetén nincs erős korreláció, semmiképp sem negatív összefüggés jellemzi őket.



A Kaggle adatok is hasonló korrelációt mutatnak mint a saját adatok, de mivel a testsúly szélesebb skálán mozog, így a BMI is sokkal magasabb értékeket ér el.

Következtetések:

- A Kaggle adatok sokkal több megfigyelést tartalmaznak és kiugróbb eseteket is láthatunk, míg a saját adatok nagyjából közepes testsúly- és magasságértékeket mutatnak be.
- Nagyrészt reprezentatívnak mondható a saját adatok adatbázis, követi a Kaggle adatokban rejlő tendenciákat, kivéve a magasság és a BMI közötti kapcsolat esetén.