Black Friday dataset

Metody analýzy dat III

Jan Vargovský

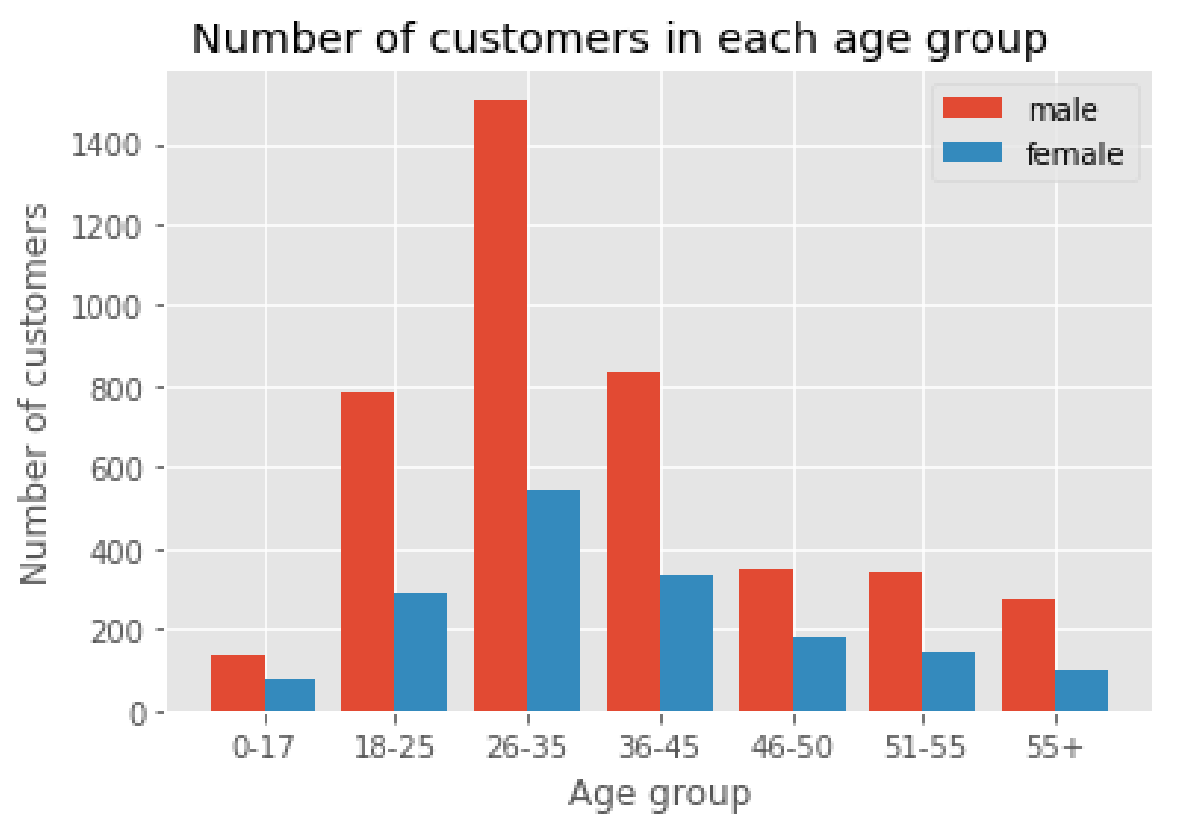
17. prosince 2018

Katedra informatiky, FEI, VŠB-TU Ostrava

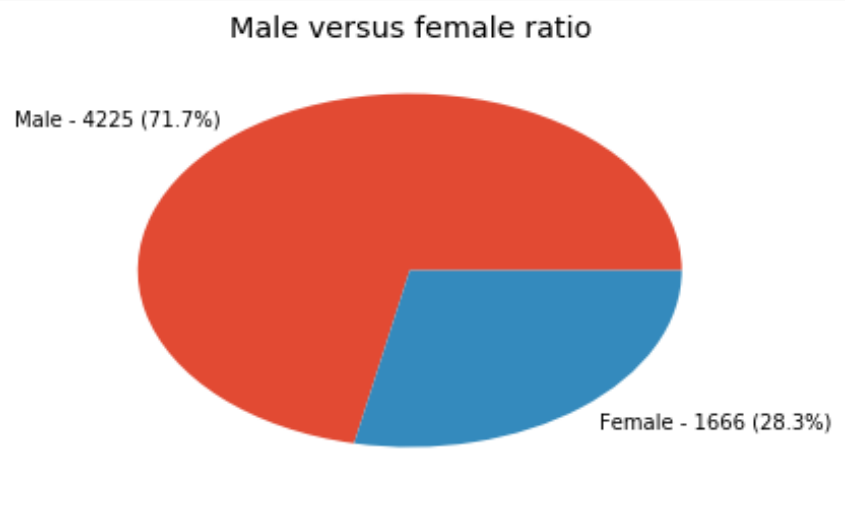
Základní info

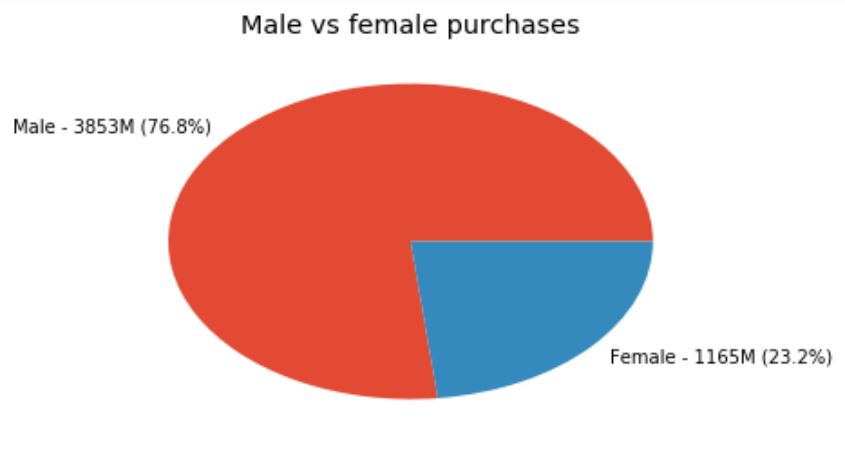
- Obsahuje 537577 nákupů od 5891 uživatelů.
- Každý zákazník má uvedeno:
  - věkovou kategorii (7),
  - pohlaví (2),
  - kde žije (3),
  - jak dlouho v daném místě žije (5),
  - jestli je sám nebo s někým (2),
  - a jaké má povolání (21).
- Transakce pak obsahuje:
  - ID výrobku (3623),
  - 1 až 3 kategorie (18),
  - a cenu ($185 - 23961$, $\mu = 9333$, $\sigma = 4981$).
- Dataset neobsahuje žádné chybějící data.

Distribuce zákazníků podle věkových kategorií

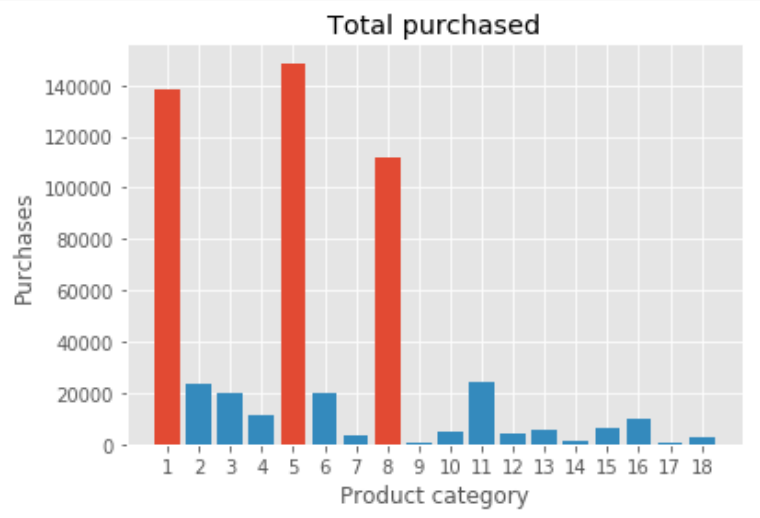


Mužů je více a utratí více jak ženy

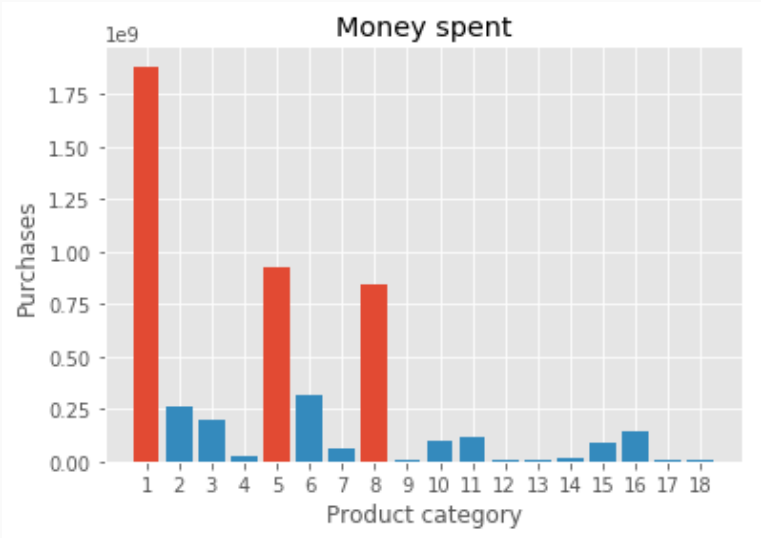




Nejvíce se nakupují výrobky z 1., 5. a 8. kategorie

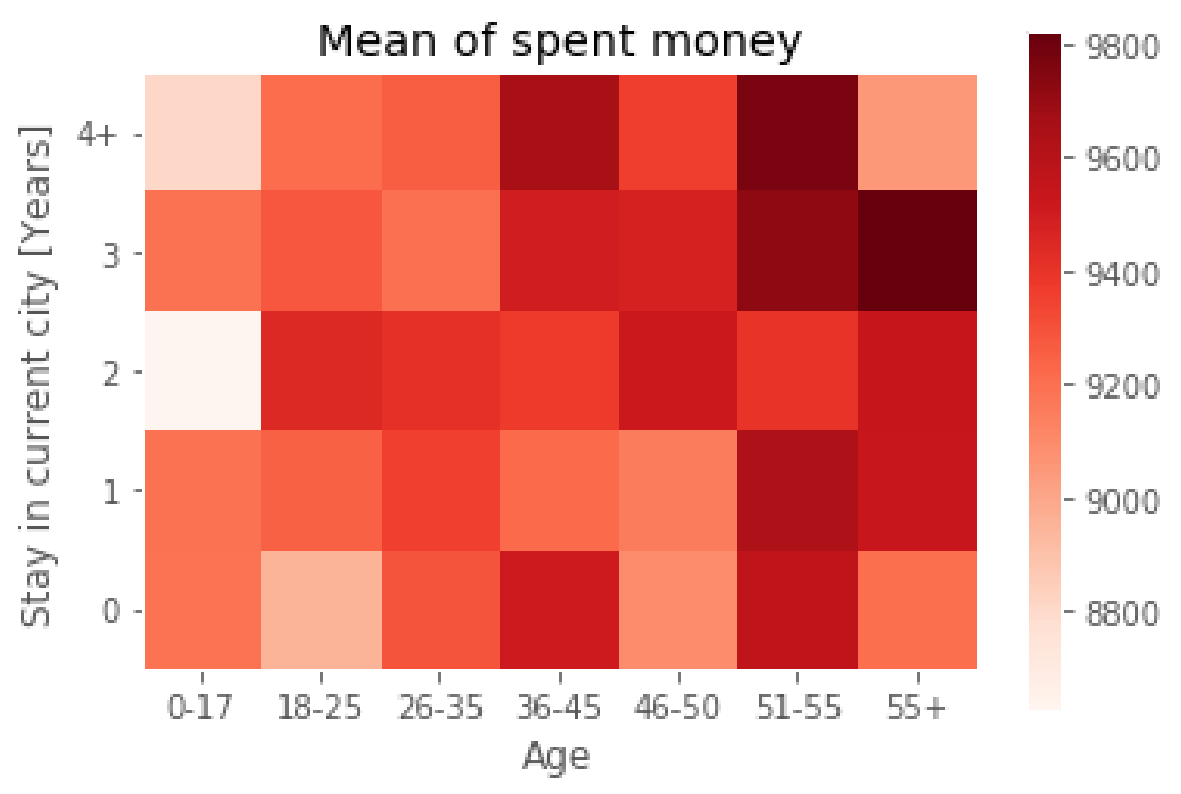


(a) Počet

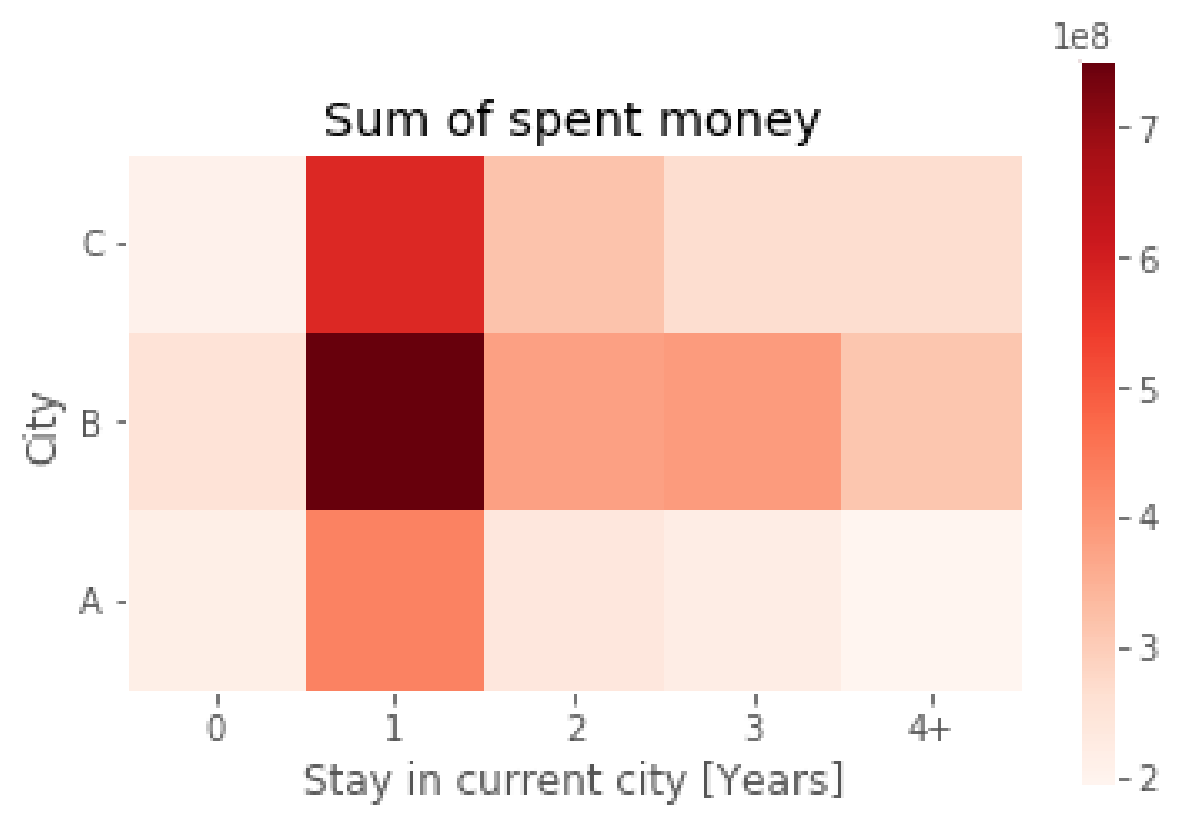


(b) Obrat

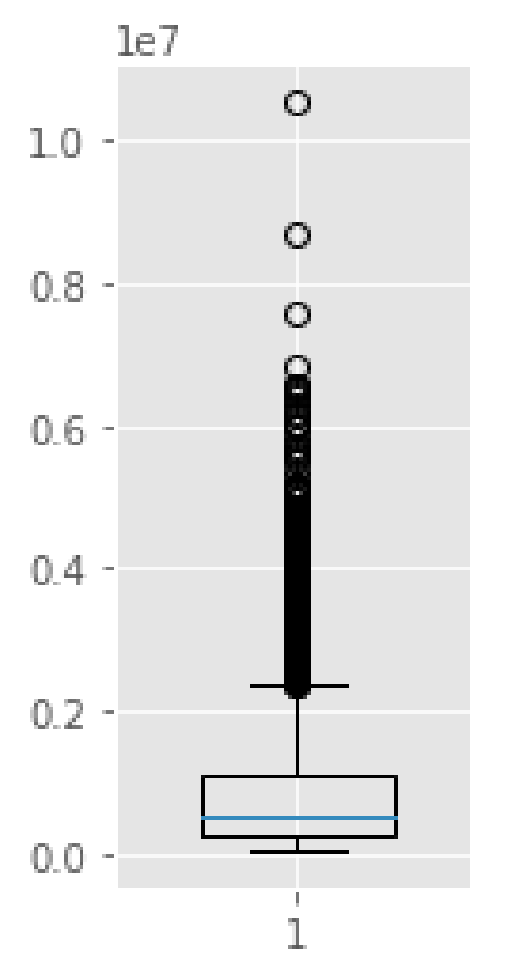
Čím starší jste, tím (pravděpodobně) více utratíte

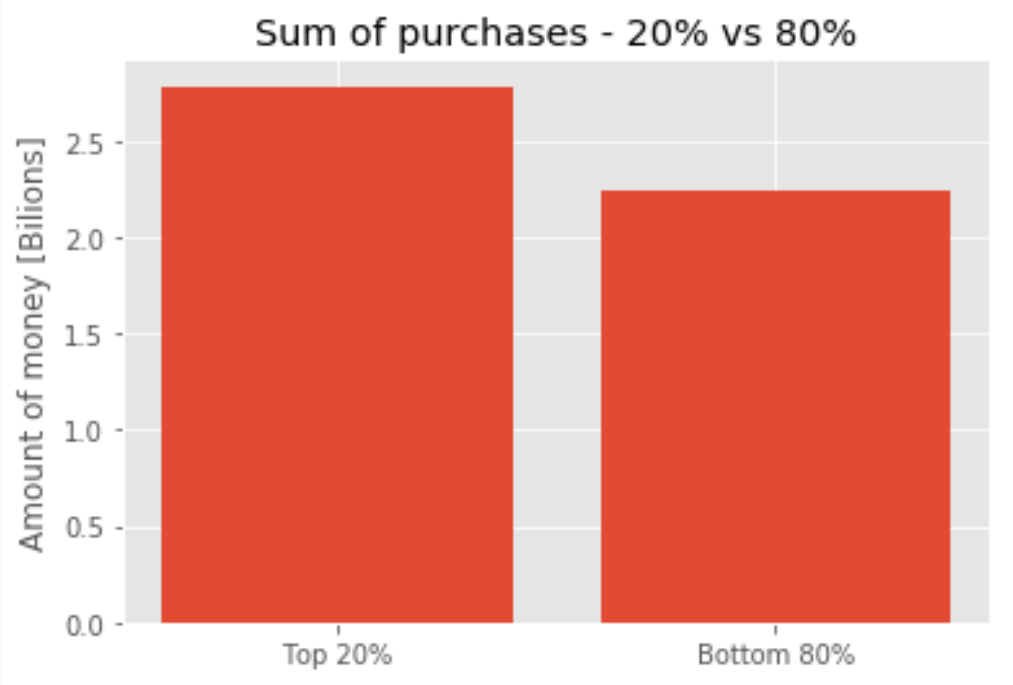


Nejvíce se nakupuje ve městě B a v 2. roce co tam žijete



Pravidlo 80/20 pro celkovou sumu nákupu





Odhad kolik zákazník utratí pomocí random forest

Dataset je dostupný zde:

`https://www.kaggle.com/mehdidag/black-friday`

Děkuji za pozornost