OSNOVE STATISTIČKOG PROGRAMIRANJA

Obrada skupa podataka: UFO Sightings

Projektni zadatak

2023/2024

Autori:

Luka Blagaić

Valentina Ivanić

Dominik Ožvald

Nikola Vlahović

Učitavanje i prilagodba podataka:

Kod učitavanja podataka koristili smo csv datoteku copleate.csv u kojoj su podaci odvojeni zarezom pa se njezino učitavanje ostvaruje funkcijom read\_csv(). Nakon učitavanja i upoznavanja s dokumentacijom vezanom uz zadani skup podatak primetali smo da se u stupcu comments nalaze skupine znakova poput &#44 koji predstavljaju točke, zareze i druge znakove koji bi mogli smetati kod spremanja u obliku csv datoteke. Skupine takvih znakova zamijenili smo razmakom, pri tome smo se koristili regularnim izrazima. Naziv država su bili predstavljeni kraticama us za Sjedinjene Američke Države, gb za Ujedinjeno Kraljevstvo, de za Njemačku … zamijenili smo kratice punim nazivima država na hrvatskom jeziku. Podacima koji su imali poznati grad u kojem je došlo do viđenja, ali nisu imali navedenu državu dodali smo državu. Sve navedene promjene spremili smo u posebnu csv datoteku koju smo koristili za daljinu analizu podataka.