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.

Analiza podataka:

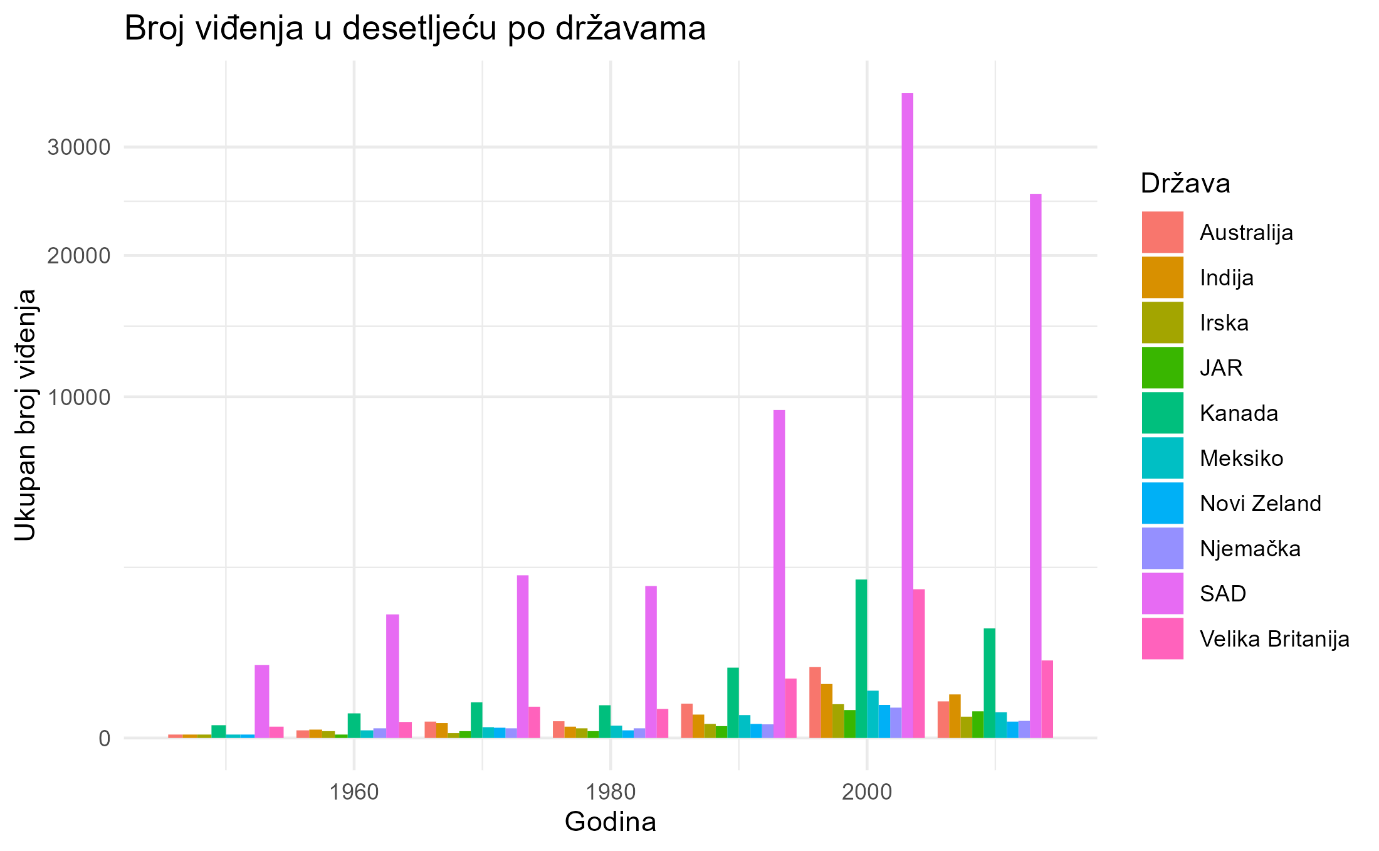
Prve pretpostavke koje smo željeli provjerit bile su da će broj prijavljenih UFO viđenja naglo rasti nakon što su dronovi postali lako dvostupačni i da je država s najviše prijavljenih viđenja SAD. Dronovi se u lošim uvjetima lako mogu zamijeniti za nepoznatu letjelicu. Da bi smo to provjerili stvorili smo graf(slika 1) koji prikazuje broj UFO viđenja svake godine za nekoliko različitih država. Na to me grafu očekujemo nagli rast oko 2000-te godine. A graph of a graph of a stock market

Description automatically generated with medium confidence

Slika Broj viđenja u godini po državama

Prikazano na graf(slika 1) možemo vidjeti da nije došlo do naglog rasta oko 2000-te godine nego je taj rast krenuo već u devedesetim godinama to nam je opovrgnulo prvu pretpostavku. Naša druga pretpostavka je točna SAD ima najviše viđenja.

Željeli smo provjerit kolika je rizika u broju viđenja u prošlih nekoliko desetljeća. Možemo vidjeti na sljedećem grafu(slika 2) kako od šezdesetih do osamdesetih godina broj viđanja ostaje malo promijenjen dok se u devedesetima naglo udvostručio, te nakon toga u dvije tisućitima utrostručio.



Slika Broj viđenja u desetljeću

Pretpostavili smo da broj viđanja 1969. kada je čovjek prvi put sletio na mjesec. Jer su ljudi u to vrijem, pogotovo u SAD-u, bili jako zainteresirani za svemirsku tehnologiju. To bi moglo povećati vjerojatnost da se neki obični događaj interpretira kao nešto nepoznato iz svemira. Na grafu (slika 3) smo prikazali razdoblje od 1960-1980 koko bi provjerili ovu pretpostavku.

A graph with lines and lines

Description automatically generated

Slika Broj viđanja od 1960-1980

Začuđujuće broj viđenja u SAD-u pada dok u ostalim država dolazi do malih promjena. Lokani minimum je oko 1972 ali možemo vidjeti da je broj slučajeva počeo padati već prije 1969 što je suprotno od naše pretopstavke