

# Analiza potrošnje kućanstava

## Motivacija i opis problema

Anketa o potrošnji kućanstava prikuplja godišnje informacije o prirodi i namjeni izdataka kućanstava, kao i različite karakteristike vezane uz njihove životne uvjete. Anketu provodi španjolski Institut za statistiku (*Instituto Nacional de Estadística* ili INE) u svrhu usklađivanja životnog standarda u državi.

## Opis skupa podataka i smjernice za izradu projekta

Skup podataka sadrži 24000 anketiranih kućanstava u periodu od 2009. do 2019. godine. Anketna pitanja podijeljena su u četiri kategorije: geografske informacije o kućanstvu, opće informacije o članovima kućanstva, informacije o prihodima te informacije o rashodima kućanstva. Svaka kategorija određena je svojim značajkama, primjerice u opće informacije o članovima kućanstva uključeni su broj članova, dobna distribucija, sektor zaposlenja članova i slične značajke.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Postoji li zavisnost između broja članova kućanstva i energenta koji se upotrebljava za dobivanje tople vode?
- Impliciraju li visoka mjesečna primanja posjedovanje vlastite nekretnine?
- Naručuju li kućanstva s većim brojem članova više hrane od onih s manje ukućana?
- Mogu li dostupne značajke predvidjeti ukupni mjesečni prihod kućanstva?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [andro.mercep@fer.hr](mailto:andro.mercep@fer.hr).

# Analiza tržišta nekretnina

## Motivacija i opis problema

Kupci nekretnina često žele zadovoljiti različite potrebe (npr. broj soba zbog djece ili potreba za većom garažom zbog dva auta), no takve “luksuze” ne žele preplatiti. Također, često su cijene nekretnina zbog različitih razloga napuhane, dok je bankama u interesu objektivno procijeniti vrijednost nekretnine za potrebe kreditiranja klijenata. Zbog toga se javlja potreba za prikupljanjem podataka o prodanim nekretninama.

## Opis skupa podataka i smjernice za izradu projekta

Prikupljeni podatci predstavljaju informacije o prodanim nekretninama u gradu Ames (SAD). Svaka je nekretnina opisana sa 79 značajki, a ukupni broj prikupljenih podataka je 1460. Nekretnine su opisane značajkama poput kvadrature, blizine javnog prijevoza, brojem spavaćih soba, posjedovanja podruma, orijentacije i slično.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Određuje li oblik zemljišne čestice broj katova kuće?
- Ovisi li veličina podruma o kvartu u gradu?
- Uvjetuje li broj spavaćih soba cijenu kvadrata nekretnine?
- Mogu li dostupne značajke predvidjeti cijenu nekretnine?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [tomislav.kovacevic@fer.hr](mailto:tomislav.kovacevic@fer.hr).

# Analiza UFC borbi

## Motivacija i opis problema

Analiza sportskih događaja (utakmica, mečeva, borbi...) kao i performansi pojedinih sportaša zadire u pore svakog dijela sportske industrije (menadžment, trening, sponzorstvo, marketing i sportske prognoze). Iz tog razloga se već duži niz godina prikupljaju podatci Ultimate Fighting Championship (UFC) borbi.

## Opis skupa podataka i smjernice za izradu projekta

Prikupljeni podatci predstavljaju informacije o UFC borbama (i borcima koji su u njima sudjelovali) u razdoblju od 1993. - 2021. godine. Borbama su pridodane značajke poput trajanja u rundama, sudca borbe te datuma i lokacije održavanja, dok su borcima pridružene značajke kao što su visina, težine, dužina ruke ili stav borca. Dodatno je poznat i pobjednik svake borbe (i način pobjede).

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Možemo li očekivati završetak borbe nokautom ovisno o razlici u dužini ruku između boraca?
- Razlikuje li se trajanje mečeva (u sekundama) između pojedinih kategorija?
- Traju li (u rundama) borbe za titulu duže od ostalih borbi u natjecanju?
- Mogu li dostupne značajke predvidjeti pobjednika?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [tomislav.kovacevic@fer.hr](mailto:tomislav.kovacevic@fer.hr).

# Antropometrija

## Motivacija i opis problema

Antropometrijske mjere opisuju morfološke osobine ljudskog tijela i njegovih pojedinih segmenata. Ovdje spadaju neke od vrlo učestalih mjera (kao što su visina, težina, širina struka ili indeks tjelesne mase (BMI)) koje redovno pratimo tokom života, ali i neke o kojima možda nikada nismo niti razmišljali (kao što su opseg gležnja, širina uha ili visina od poda do ručnog zgloba). Primjene antropometrijskih mjera mogu se naći u mnoštvu područja kao što su medicina, dizajn odjeće, ergonomija, paleoantropologija, forenzika i druga.

## Opis skupa podataka i smjernice za izradu projekta

Podatci se sastoje od odabranih antropometrijskih mjera 133 ispitanika prikupljenih u dvije faze. Prva faza se sastoji od ispitanika iz Sjedinjenih Američkih Država, dok se druga faza sastoji od ispitanika iz Španjolske. Svaka mjera je procijenjena više puta: od strane stručnjaka, ili od strane (polu)-automatskih 3D skenera i pripadnih softverskih rješenja. Zajedno sa podacima spomenutih mjera, svaki ispitanik ima više odgovarajućih 3D snimaka, dobivenih s različitim 3D skenerima koji se analiziraju.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Mjere li određeni 3D skeneri različito pojedinu mjeru?
- Imaju li određene antropometrijske mjere statistički značajne razlike između onih prikupljenih u Sjedinjenim Američkim Državama, te onih prikupljenih u Španjolskoj?
- Možemo li sa nekim danim mjerama predvidjeti neku drugu antropometrijsku mjeru?
- Na koji način biste izmjerili neku vama zanimljivu mjeru sa popratnih 3D podataka? Možete li takvu mjeru dobiti iz već danih podataka?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu david.bojanic@fer.hr.

# Goodreads – analiza popularnosti knjiga

## Motivacija i opis problema

Goodreads je društvena mreža za obožavatelje čitanja s preko 90 milijuna korisnika. Goodreads baza podataka sastoji se od iscrpne količine knjiga, autora i komentara korisnika. Između ostalog, korisnici mogu pretraživati veliki katalog knjiga te ih ocjenjivati. Pomoću ocjena i stvorene mreže profila korisnici lako dolaze do novih, njima zanimljivih publikacija.

## Opis skupa podataka i smjernice za izradu projekta

Dani skup sadrži podatke koji odgovaraju knjigama na popisu Goodreads Best Books Ever (najveći popis na portalu). Za svaku knjigu mogu se iščitati razne informacije, među ostalima: autor, prosječna ocjena, datum izdavanja i žanr.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Postoje li razlike u ocjenama knjiga s obzirom na žanr?
- Jesu li knjige s manje stranica jeftinije?
- Možete li odrediti popularnost knjige (po vašoj definiciji, npr. broj glasača, prosječna ocjena...) na temelju dostupnih varijabli?
- Postoje li razlike u popularnosti knjiga s obzirom na njihovu starost?
- Možete li na temelju dostupnih varijabli odrediti je li knjiga bila nagrađivana?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [tessa.bauman@fer.hr](mailto:tessa.bauman@fer.hr).

# Milijarderi

## Motivacija i opis problema

Trenutno na svijetu ima manje od 1% milijardera. Ljudsko bogatstvo se sa sve većim razmjerom razdvaja na ekstremno bogate i ekstremno siromašne ljude. Kako bi dobili veći uvid u bogatstvo onih najbogatijih, te stvorili šansu za povećanje tih 1%, prikupljeni su podatci onih najbogatijih sa Forbesove liste od 1994. do 2014.

## Opis skupa podataka i smjernice za izradu projekta

Svaki milijarder u danim podacima sadrži 21 atribut koji ga opisuje. Neki od njih su: broj godina, spol, državljanstvo, porijeklo bogatstva, struka, vrijednost imovine, itd.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Ima li neki kontinent statistički značajno više milijardera?
- Jesu li milijarderi koji su naslijedili bogatstvo statistički značajno bogatiji od onih koji nisu?
- Možete li iz danih varijabli predvidjeti njihovo bogatstvo?
- Kada biste birali karijeru isključivo prema kriteriju da se obogatite, koju biste industriju izabrali?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu david.bojanic@fer.hr.

# Mobilnost turista i poslovanje hrvatskog turističkog sektora

## Motivacija i opis problema

Kroz ovaj projektni zadatak analizira se veza mobilnosti turista kroz Hrvatsku i uspješnosti hrvatskog turističkog sektora, predstavljenog s nekoliko najznačajnijih domaćih poduzeća u turističkom sektoru. Mobilnost turista najjednostavnije se može izraziti temeljem prometa na graničnim prijelazima, odnosno brojem ulazaka i izlazaka u određenom periodu, a uspješnost poslovanja turističkih tvrtki kroz poslovne pokazatelje (fundamente) kao što su prihodi od prodaje i neto dobit. Budući da je poslovni model turističkih tvrtki zasnovan prije svega na broju gostiju, očekuje se da će s većim brojem ulazaka stranih državljana u Hrvatsku rasti i njihovi poslovni prihodi i dobit. Osim toga, kroz ovaj projekt ispitat će se i povezanost mobilnosti turista s cijenom dionica turističkih kompanija na Zagrebačkoj burzi, budući da one odražavaju konsenzus investitora o vrijednosti tih tvrtki.

## Opis skupa podataka i smjernice za izradu projekta

Podatci uključuju dnevne ulaske i izlaske gostiju u Hrvatskoj, podijeljene po vrsti graničnog prijelaza (cestovni, željeznički, riječni, pomorski, zračni), za period od 2013. do 2022. Također, za isti period su uključeni i poslovni pokazatelji prihoda od prodaje te neto dobiti za četiri velika turistička poduzeća u Hrvatskoj: Arena Hospitality Group d.d., Imperial Riviera d.d., Maistra d.d. te Valamar Riviera d.d. – ovi podatci dani su kvartalno, budući da se objavljuju u kvartalnim financijskim izvještajima. Uz to su dostupni i podatci o dnevnim vrijednostima indeksa CROBEXturist Zagrebačke burze čije su sastavnice dionice poduzeća iz turističkog sektora.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Analizirajte podatke o dnevnom prometu preko graničnih prijelaza. Postoje li razlike među vrstama graničnih prijelaza? Postoje li razlike među različitim danima u tjednu ili mjesecima?
- Analizirajte fundamentalne podatke turističkih tvrtki – postoje li sezonalnosti i u njima?
- Proučite vezu prometa po različitim graničnim prijelazima s fundamentima turističkih poduzeća. Mogu li se fundamenti objasniti mobilnošću turista preko hrvatskih granica i u kojoj mjeri? Koja je vrsta prometa najznačajnija za hrvatska turistička poduzeća?
- Istražite u kojoj mjeri se dnevna kretanja cijene indeksa CROBEXturist mogu objasniti kretanjem turista preko hrvatskih granica.

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [andro.mercep@fer.hr](mailto:andro.mercep@fer.hr).

# Statistika nogometaša engleske Premier lige

## Motivacija i opis problema

Svjesni smo da je statistika postala dio naše svakodnevice i da je kvalitetna statistička analiza podataka potrebna gdje god se okrenemo. Jedno područje u kojem se statistika već jako dugo primjenjuje je sport. S vremenom su sportske statističke analize jako napredovale i danas imamo ogromne količine podataka za razne sportove koje je potrebno dobro analizirati. Kvalitetna analiza tih podataka postala je iznimno važan faktor u razvoju sportova i pojedinaca i više si nitko ne može priuštiti da ne radi analizu prikupljenih podataka. Većina klubova u današnje vrijeme posjeduje tim za analizu izvedbi svojih igrača, a to je pogotovo slučaj u engleskoj Premier ligi.

## Opis skupa podataka i smjernice za izradu projekta

U datoteci koja će se koristiti u sklopu ovog projekta se nalaze razne statistike za sve igrače Premier lige za sezonu 2021-2022. Svaki redak datoteke odnosi se na nekog igrača, a u stupcima možemo naći razne statistike vezane za igrače poput dobi igrača, broja golova, broja odigranih minuta, broja dobivenih kartona i slično.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Postoji li razlika u broju odigranih minuta mladih igrača (do 25 godina) među premierligaškim ekipama?
- Dobivaju li u prosjeku više žutih kartona napadači ili igrači veznog reda?
- Možete li na temelju zadanih parametara odrediti uspješnost pojedinog igrača?
- Doprinosi li sveukupnom uspjehu svoga tima više "domaći" igrači (tj. igrači engleske nacionalnosti) ili strani igrači?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [krunoslav.jurcic@fer.hr](mailto:krunoslav.jurcic@fer.hr).



# Zločin i kazna – analiza kriminala

## Motivacija i opis problema

Recidivizam je jedan od temeljnih pojmova u kaznenom pravosuđu koji se odnosi na kriminalno ponašanje nakon što je osoba već kažnjena zbog prethodnog. Recidivizmom se smatra ponavljanje bilo kakvog kaznenog djela koje rezultira ponovnim uhićenjem ili ponovnom osudom tijekom razdoblja od tri godine nakon puštanja osobe na slobodu. Zbog opće sigurnosti, a i razvoja novih strategija za reintegraciju (bivših) osuđenika, bitno je proučiti i razumjeti koji čimbenici su učestaliji kod ponavljača zločina.

## Opis skupa podataka i smjernice za izradu projekta

Podatci su sakupljeni u saveznoj zemlji Georgia, SAD, od 2013. do 2015. godine, na zahtjev američkog Nacionalnog instituta za pravosuđe. Za svakog osuđenika dostupni su sociodemografski podatci (spol, dob, obrazovanje...), podatci vezani uz kriminalno ponašanje (tip kriminalnog djela, rezultati testova na narkotike, duljina zatvorske kazne...) te podatci o ponašanju nakon odrađene kazne, uključujući podatak o tome je li došlo do recidivizma ili ne.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Postoji li razlika u duljini zatvorske kazne s obzirom na razinu obrazovanja?
- Jesu li mlađe dobne skupine podložnije ponavljanju kriminala od starijih?
- Postoji li veza između korištene vrste narkotika i vrste zločina?
- Možete li na temelju dostupnih varijabli odrediti hoće li doći do ponavljanja kriminalnog ponašanja?
- Jesu li mlađi osuđenici skloniji drogama?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [tessa.bauman@fer.hr](mailto:tessa.bauman@fer.hr).

# Životni vijek ljudi

## Motivacija i opis problema

Kvaliteta života se neprestano poboljšava u modernom svijetu, a statistika je definitivno doprinijela toj činjenici. Uz pomoć statističkih podataka danas je jednostavnije analizirati razne parametre koji opisuju kvalitetu života i usporediti vrijednost navedenih parametara za zemlje cijelog svijeta. Očekivani životni vijek ljudi je jedan od bitnijih parametara za procjenu kvalitete života na određenom području.

## Opis skupa podataka i smjernice za izradu projekta

WHO (World Health Organization) datoteka koja će se koristiti u sklopu ovog projekta sadrži kombinaciju informacija skupljenih od strane GHO (Global Health Observatory) organizacije i informacija skupljenih od strane UNESCO (United Nations Educational Scientific and Culture Organization) organizacije. Skup sadrži razne parametre skupljene za 183 države u periodu od 2000-2016. godine. Među opisanim parametrima uključeni su očekivani životni vijek pri rođenju, očekivani životni vijek u 60. godini života, konzumacija alkohola po stanovniku, indeks tjelesne mase i drugo.

Pri izradi projekta vodite se ovim istraživačkim pitanjima:

- Postoji li razlika u konzumaciji alkohola tijekom godina među svjetskim regijama?
- Razlikuje li se količina uloženog novca u zdravstvenu skrb između zemalja jugoistočne Azije i zemalja zapadnog Pacifika?
- Može li se na temelju zadanih parametara objasniti očekivani životni vijek ljudi u Europi u 2015. godini?
- Ima li razlike u trendovima zaraženosti različitim bolestima među svjetskim regijama?

Također razmislite o vlastitim istraživačkim pitanjima i hipotezama koje želite analizirati koristeći dostupne podatke.

**Napomena:** Ako imate dodatnih pitanja vezanih uz ovaj konkretan projekt, javite se na e-mail adresu [krunoslav.jurcic@fer.hr](mailto:krunoslav.jurcic@fer.hr).