

## Način odabira zadatka za pojedinog studenta

Skupovi podataka:

<https://www.kaggle.com/datasets/steверusso/cincinnati-car-crash-data> (27)

<https://www.kaggle.com/datasets/arcticgiant/french-financial-news> (20)

<https://www.kaggle.com/datasets/piterfm/2022-ukraine-russian-war> (18)

<https://www.kaggle.com/datasets/equinxx/spotify-top-50-songs-in-2021> (18)

<https://www.kaggle.com/datasets/vivek468/superstore-dataset-final?select=Sample++Superstore.csv>

(21)

<https://www.kaggle.com/datasets/jaytilala/global-power-plant> (9)

Ukupno u danim skupovima podataka imamo 113 stupaca/varijabli koje opisuju podatkovne instance (podatkovna instanca je jedan redak Excel tablice). Završni projekt možete izraditi samostalno ili u paru. Ukoliko radite u paru, tada sortirajte prezimena i najprije uzmite ID osobe koja je prva gledano uzlaznim abecednim redom. Ukoliko oba studenta imaju isto prezime, tada za sortiranje koristite ime. Podijelite četveroznamenkasti broj (zadnje 4 znamenke tog studentskog ID broja - broj koji se prikazuje na Merlinu) sa 113 i ostatak dijeljenja uvećajte za 1 (primjerice,  $4235\%41+1=55$ ). Dobit ćete broj iz raspona [1,113] koji predstavlja redni broj varijable od ovih 113 gore navedenih za skupove podataka. Tako ste dobili varijablu koja će imati ulogu ključa. Zbog jednostavnosti odaberite po želji samo po jedan redak tablice sa određenom vrijednosti tog atributa (bez duplikata) da bi u stablu u svim čvorovima bile različite vrijednosti ključa (dovoljno je odabrati 20 do 30 takvih redaka). Ako odabrana varijabla nema bar 5 različitih vrijednosti u različitim recima, onda odaberite atribut koji je sljedeći na redu, itd., dok ne naiđete na atribut s tim svojstvom. Varijablu možete preskočiti ako nije odgovarajućeg tipa za implementaciju 2. dijela zadatka.

Drugi dio zadatka odnosi se na odabir algoritma ili strukture podataka koji/koju je potrebno implementirati. Uzmite ID drugog člana para (ili ponovno svoj vlastiti ukoliko zadatak izrađujete samostalno) i podijelite ga sa 7 te ostatak dijeljenja uvećajte za 1.

1. Quick sort
2. Merge sort
3. Ruksak sa svojstvom dijeljenja
4. Ruksak bez svojstva dijeljenja
5. Uređeno binarno stablo
6. Gomila
7. AVL stablo

Zadatak je za zadani skup podataka implementirati algoritam ili strukturu podataka koji ste dobili u drugom dijelu zadatka. Ukoliko je rezultat 2, tada trebate na odabranim podacima (čitaju se iz datoteke) pokrenuti implementaciju merge algoritma sortiranja na temelju ključa odabranog u 1. dijelu zadatka i ispisati sortirane podatke. Radi preglednosti izlaza formatirajte izlaz u stupce fiksne širine, a možete izbaciti i prikaz nepotrebnih stupaca ukoliko skup sadrži puno stupaca.

Seminarski rad (Word dokument) treba sadržavati opis izračuna pri odabiru atributa, prvih 20 redaka tablice s podacima na temelju kojih je testirana implementacija i prikaz izvođenja programa. Ispis pri izvođenju programa treba biti pregledan. Potrebno je predati i datoteku sa izvornim kodom programa, te datoteku sa podacima koje ste koristili pri oblikovanju i testiranju programa.

### **Prezentiranje projekta**

Student rezultate projekta prezentira na ispitnom roku. Nije potrebno izraditi prezentaciju, već je potrebno objasniti dijelove koda programa. Za pristup ispitnom roku student prijavljuje ispit. Podsjetimo se što piše u DINP-u: Završni ispit nosi udio od maksimalno 30 ocjenskih bodova, a smatra se položenim samo ako na njemu student postigne minimalno 50%-ni uspjeh. Studenti su dužni predati sve datoteke izvornog koda, kao i seminarski rad najkasnije tri dana prije datuma ispitnog roka. Na dan ispitnog roka studenti su dužni izložiti svoj rad i prikazati rješenje. Popis točnih termina bit će objavljen najkasnije 1 dan prije datuma ispitnog roka.