**SVEUČILIŠTE U RIJECI  
FAKULTET INFORMATIKE I DIGITALNIH TEHNOLOGIJA**

**Preddiplomski studij informatike**

**Seminarski rad – Završni ispit iz kolegija  
Programiranje za podatkovnu znanost**

**na temu**

**Prodaja automobila**

Autor: Ahmičić Dino  
Mentori: Maja Matetić

Dejan Ljubobratović

U Rijeci, Kolovoz 2023.

**Sadržaj**

[1. Uvod 3](#_Toc143595128)

[2. Prvi zadatak – Učitavanje podataka 4](#_Toc143595129)

[3. Drugi zadatak – Uređivanje podataka 5](#_Toc143595130)

[4. Treći zadatak – Reprezentacija datuma 6](#_Toc143595131)

[5. Četvrti zadatak – Statistika 7](#_Toc143595132)

[6. Peti zadatak – Prilagodba podatkovnih okvira 8](#_Toc143595133)

[7. Šesti zadatak – Vizualizacija podataka 9](#_Toc143595134)

# Uvod

U ovom završnom ispitu iz kolegija programiranje za podatkovnu znanost, kao seminarski rad izradio sam R skriptu za prikaz statističkih podataka prodaje automobile. CSV tablica koju sam koristio, može se pronaći na web stranici kaggle.com na linku: <https://www.kaggle.com/datasets/gagandeep16/car-sales>

Praktični dio ovoga rada bio je podijeljen u 6 zadataka u kojim smo trebali izraditi, te objasniti postupak učitavanja podataka iz CSV datoteke unutar R studija, morali smo prikazati postupak uređivanja podataka korištenjem funkcija *pivot\_longer, pivot\_wider, separate, unite*, te *drop\_NA* i *replace\_NA*, prikazati datume i vremenske oznake, prikazati statističke mjere (primjerice mjere centralne tendencije, itd. ), prikazati funckije filtriranja podataka i odabire podskupova stupaca, te iskazati neke podatke kao grafičke prikaze.

# Prvi zadatak – Učitavanje podataka

Za početak rada, moramo prvo instalirati, te učitati sve biblioteke koje će nam biti potrebne za daljnji rad. Te biblioteke u ovom slučaju jesu ggplot2, plyr, readr, tidyr i dplyr.

A screen shot of a computer code

Description automatically generated

Sljedeće još samo moramo učitati našu CSV datoteku. To ćemo izvesti funkcijom read\_csv, te ćemo ju odmah spremiti u varijablu automobili.



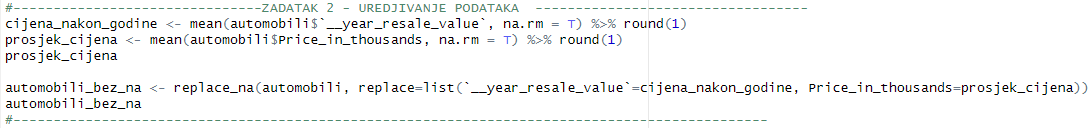
A group of text on a white background

Description automatically generated

# Drugi zadatak – Uređivanje podataka

U zadatku gdje moramo urediti podatke napraviti ćemo zamjenu praznih ćelija unutar stupaca „\_\_year\_resale\_value“ i „Price\_in\_thousands“ tako što ćemo u novi dataset „cijena\_nakon\_godine“ spremiti prosjek cijena automobila (maknemo prazne ćelije i zaokružimo rezultat na jednu decimalu), a u novi dataset „prosjek\_cijena“ spremimo prosjek cijena novih automobila (također maknemo prazne ćelije i rezultat zaokružimo na jednu decimalu).

Na kraju, sa funkcijom „replace\_na“ unutar stupca „\_\_year\_resale\_value“ prazne ćelije ispunimo vrijednostima „cijena\_nakon\_godine“ dataset-a, a unutar stupca „Price\_in\_thousands“ prazne ćelije ispunimo vrijednostima „prosjek\_cijena“ dataset-a.



A screen shot of a computer

Description automatically generated

# Treći zadatak – Reprezentacija datuma

Kao reprezentaciju datuma, izraditi ćemo prikaz dataset-a, u kojem su spremljeni marke i modeli automobila, te datumi koji predstavljaju dane ulaska automobila u pogon.

Unutar novog dataset-a pod imenom „datumi\_zadnjih\_modela“, funckijom „paste“ spremiti ćemo imena brandova, imena modela automobila, te funckijom „as.date“ spremiti datume u formatu „%m-%d-%Y“ (mjesec – dan – godina).



Sljedeće, iskazati ćemo vremensku razliku između najstarijeg i najnovijeg datuma ulaska automobila u pogon, iskazanu u tjednima. To radimo funckijom „difftime“:  
  


A close-up of a computer code

Description automatically generated

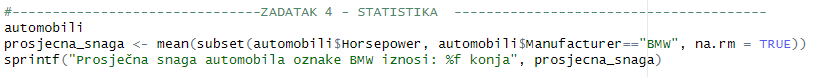
A close up of a text

Description automatically generated

# Četvrti zadatak – Statistika

Kao prikaz nekih statističkih obilježna našeg glavnog dataset-a automobili, koristiti ćemo funkcije mean, median, i mode.

Kao prvi dio, iskazati ćemo prosječnu konjsku snagu automobila marke BMW, te ćemo sve prazne ćelije zanemariti:





Sljedeće, median funkcijom prikazati ćemo prodaju automobila u jednoj godini:





Na kraju, funkcijom mode iskazati ćemo prosječnu zapremninu motora automobila kao boxplot:  
D  


A diagram of a graph

Description automatically generated

# Peti zadatak – Prilagodba podatkovnih okvira

U ovom zadatku, unutar dataset-a „automobili\_bez\_na“ izmjeniti ćemo stupac „Sales\_in\_thousands“ u stupac imena „Sales\_in\_10\_thousands“ koji će reprezentirati prethodni stupac sa izmjenjenim vrijednostima (iz prodaje iskazane u tisućama, u prodaja iskazana u deset tisuća).



A screenshot of a calculator

Description automatically generated

Sljedeće ćemo funkcijom „filter“ filtrirati sve automobile kojima je ukupna prodaja jedinica bila preko 50000, kojima je cijena ispod 35 000, te konjska snaga preko 150. Rezultate ćemo prikazati silazno rangirano prema količini prodanih jedinica.



A white background with many small letters

Description automatically generated with medium confidence

# Šesti zadatak – Vizualizacija podataka

Kao grafičku vizualizaciju podataka napraviti ćemo dva grafa. Prvi graf biti će točkasti graf koji reprezentira cijenu i prodaju jedinica automobila u tisućama:



A graph with black dots

Description automatically generated

Drugi graf biti će nam histogram koji reprezentira količinu auta u određenim cijenovnim rangovima:



A graph of a building

Description automatically generated