



Универзитет у Београду
Математички факултет

СТАТИСТИЧКИ СОФТВЕР 1 (3В)

9. фебруар 2026. године

- База података `iris` садржи флоралне мере (дужина и ширина листова) за три врсте цвета (`setosa`, `versicolor`, `virginica`).
 - [2 поен] Креирати нову променљиву `Sepal.Area` као производ `Sepal.Length` и `Sepal.Width`. Затим, користећи функцију `summarise()`, израчунати минималну вредност, максималну вредност и интерквartilни распон променљиве `Sepal.Area`, без обзира на врсту.
 - [2 поен] Користећи функције `group_by()` и `summarise()`, израчунати за сваку врсту цвета проценат опсервација за које важи `Petal.Length > 4`.
 - [3 поен] Креирати променљиву `petal_size` са вредностима:
 - `small` ако је `Petal.Length < 1.5`,
 - `medium` ако је `1.5 <= Petal.Length <= 5`,
 - `large` ако је `Petal.Length > 5`.Затим одредити број опсервација за сваку комбинацију (`Species`, `petal_size`) и резултат сортирати по `Species`, а затим опадајуће по броју опсервација.
 - [1 поен] Користећи пакет `ggplot2`, приказати `boxplot` за променљиву `Sepal.Area` по врстама цвета (`Species`). На графику додати наслов и означити осе.
- У пакету `nycflights13` налазе се табеле `flights` (подаци о летовима) и `airports` (подаци о аеродромима). Анализирају се летови који полазе из Њујорка током 2013. године.
 - [3 поена] Из табеле `flights` издвојити само летове за које су дефинисане вредности `dep_delay` и `arr_delay`. Затим груписати податке по комбинацији (`origin`, `dest`) и израчунати:
 - `n_flights` – укупан број летова за дату комбинацију,
 - `avg_arr_delay` – просечно долазно кашњење (у минутима),
 - `pct_late` – проценат летова са долазним кашњењем већим од 15 минута.Резултат треба да садржи по један ред за сваку комбинацију (`origin`, `dest`).
 - [2 поена] Из табеле `flights` издвојити летове за које је дефинисана вредност `air_time`. Затим, за сваки полазни аеродром `origin` израчунати просечно време лета (`avg_air_time`) и максимално време лета (`max_air_time`).
 - [2 поена] Из табеле `flights` издвојити летове за које су дефинисане вредности `dep_delay` и `arr_delay`. Затим, за сваки полазни аеродром `origin` израчунати проценат летова за које важи услов `dep_delay > arr_delay`.
 - [2 поена] Из табеле `flights` издвојити летове за које је дефинисана вредност `distance`. Добијену табелу спојити са табелом `airports` тако да се дода назив полазног аеродрома (`name`). Након спајања, за сваки полазни аеродром `origin` израчунати просечну дужину лета (`avg_distance`).
 - [2 поена] На основу табеле `flights` издвојити летове који полазе 9. фебруара 2013. године и за које су дефинисане вредности `dep_delay`, `arr_delay` и `distance`. Затим приказати `scatter plot` који илуструје однос између полазног кашњења (`dep_delay`) и долазног кашњења (`arr_delay`). На графику боју тачака одредити у зависности од полазног аеродрома (`origin`), величину тачака одредити у зависности од дужине лета (`distance`).
 - [1 поена] За летове из табеле из дела под (д) приказати `scatter plot` који илуструје однос између дужине лета (`distance`) и времена лета (`air_time`). На графику додати полиномијалну регресиону криву.