

Основне академске студије
Информациони инжењеринг

Основи рачунарске интелигенције

Обрада и анализа података у језику Python

(материјали за вежбе)

Библиотека *NumPy*

функција **`numpy.around(...)`**

заокруживање вредности

функција **`numpy.linspace(...)`**

формирање низа равномерно распоређених вредности у датом интервалу

функција **`numpy.repeat(...)`**

умножавање елемената низа

Одабране функције

Библиотека *pandas*

функција **`pandas.read_csv(...)`**

учитавање података из датотеке у формату CSV

функција **`pandas.cut(...)`**

распоређивање вредности по интервалима

функција **`pandas.Series.astype(...)`**

трансформација серије у задати тип података

функција **`pandas.Series.value_counts(...)`**

пребрајање вредности по различитим елементима серије

функција **`pandas.Series.replace(...)`**

замена вредности серије

функција **`pandas.Series.isin(...)`**

провера заступљености елемената серије међу датим вредностима

особина **`pandas.Series.dt.month`**

издвајање месеца из датумског садржаја серије

Библиотека *pandas*

функција **`pandas.DataFrame.rename(...)`**

промена ознака оса за скуп података

функција **`pandas.Index.to_series(...)`**

трансформација индекса у серију

Библиотека *Matplotlib*

функција **`matplotlib.pyplot.plot(...)`**

исцртавање графикана с линијама и ознакама у две димензије

функција **`matplotlib.pyplot.step(...)`**

исцртавање степеничастог графикана

функција **`matplotlib.pyplot.subplot(...)`**

манипулација простором за цртање с координатним системом

функција **`matplotlib.pyplot.xlim(...)`**

манипулација распоном *X* осе

функција **`matplotlib.pyplot.ylim(...)`**

манипулација распоном *Y* осе

класа **`matplotlib.patches.Patch`**

дводимензионални елемент с бојама

Библиотека *Matplotlib*

функција **`matplotlib.cm.get_cmap(...)`**

очитавање мапе боја

особина **`matplotlib.colors.ListedColormap.colors`**

издвајање боја из мапе боја

Подаци

скуп метеоролошких података из 2020. за Беверли Хилс (САД)

постоји 266 појава над 17 обележја

дан, датум, температура (највиша, просечна и најнижа), тачка росе (највиша, просечна и најнижа), влажност (највиша, просечна и најнижа), брзина ветра (највећа, просечна и најмања), притисак (највиши и најнижи) и акумулација падавина

свака појава обухвата календарске и метеоролошке податке за један дан

извор скупа података

скуп података *Weather*

матична Интернет страна

<https://www.beverlyhills.org/opendata/datasets/weather/>

у оквиру портала отворених података Беверли Хилса (САД)

подаци доступни у формату *CSV* засебно по годинама од 2017. до 2020. године

може бити искоришћен скуп података за 2020. годину

скуп података ажуриран јануара 2021. године

податке пружио Град Беверли Хилс (*City of Beverly Hills*) а одржава Одељење за информационе технологије (*Information Technology Department*)

лиценца и могућности коришћења скупа података

Open Data Commons Public Domain Dedication and License (PDDL)

<http://www.opendefinition.org/licenses/odc-pddl>

Задатак 1.

Учитати скуп података из одговарајуће *CSV* датотеке!

Приликом учитавања, водити рачуна о кодном распореду и парсирању датума!

Уклонити из учитаног скупа података све редове у којима су недостајуће вредности присутне!

Проверити садржај учитаног скупа података и типове колона!

Задатак 2.

Унутар учитаног скупа података задржати само обележја о датуму, просечној температури, просечној влажности и просечној брзини ветра!

Преименовати колоне скупа података тако да уместо изворних буду коришћени преведени називи!

Проверити садржај скупа података и типове колона!

Задатак 3.

У скуп података додати обележје о дану у месецу и обележје о месецу, а вредности тих обележја извести из одговарајућих вредности обележја о датуму!

За вредности обележја о месецу користити пуне називе месеци!

Проверити садржај скупа података и типове колона!

Задатак 4.

У скуп података додати нова бројчана обележја о просечној температури израженој у степенима Целзијуса, просечној влажности израженој у процентима и просечној брзини ветра израженој у метрима у секунди!

Вредности нових обележја о просечној температури и просечној брзини ветра заокружити на две децимале!

Из скупа података уклонити претходна знаковна обележја о просечној температури, просечној влажности и просечној брзини ветра!

Проверити садржај скупа података и типове колона!

Задатак 5.

За податке о месецу, просечној температури, просечној влажности и просечној брзини ветра спровести груписање на основу месеца а потом и агрегирање бројчаних вредности помоћу функције за аритметичку средину!

Приказати груписане и агрегиране податке хронолошки по месецима!

Утврдити на нивоу груписаних и агрегираних података која је највећа вредност у вези с просечном температуром!

Приказати груписане и агрегиране податке за просечно најтоплији месец!

Задатак 6.

Утврдити која је највиша просечна температура унутар скупа података и приказати је!

Нацртати графикон расејања тако да X оса одговара просечној влажности а Y оса просечној температури!

Подесити и X и Y осу на графикону тако да прикладно покривају распоне приказаних вредности!

Подесити наслов графикона и називе оса!

Задатак 7.

Формирати подскуп података који се односи само на месеце октобар, новембар и децембар!

Проверити садржај подскупа података прескачући обележје о датуму!

Над подскупом података израчунати и приказати дескриптивне статистике за просечну температуру!

На основу подскупа података одредити колико дана се по својој просечној температури уклапа у сваки од следећих температурних опсега: $[0, 10)$ °C, $[10, 20)$ °C и $[20, 30)$ °C!

...

Задатак 7.

...

Нацртати тортни графикон који одговара расподели данâ по датим температурним опсезима!

Подесити ознаке поред исечака на графикону тако да буду приказани одговарајући температурни опсези!

Подесити ознаке на самим исечцима тако да буду приказани одговарајући удели у процентима!

Исечке одмакнути једне од других!

Подесити боје исечака користећи неку од уграђених мапа боја!

Подесити наслов графикона!

Задатак 8.

Утврдити која је највећа просечна брзина ветра унутар скупа података и приказати је!

Нацртати два хистограма за просечну брзину ветра, при чему се један хистограм односи на податке за јул а други на податке за децембар!

За оба хистограма користити идентичне границе интервала: оквирна ширина појединачног интервала је 0,25 а први интервал креће од 0!

Хистограме треба приказати тако да буду један до другог!

Подесити за оба хистограма Y осу тако да прикладно покрива распон одговарајућих вредности!

Подесити наслов графикона и називе оса за оба хистограма!

Задатак 9.

Проверити податке који се односе на мај!

Одредити медијану за просечну влажност на нивоу свих података и приказати је!

Нацртати стубичасти графикон који приказује просечну влажност по појединачним данима током маја!

Стубове на графикону обојити тако да је могуће разликовати дане по томе да ли је одговарајућа просечна влажност била до медијане или изнад ње!

Подесити на графикону ознаке за дане испод стубова и подесити Y осу тако да прикладно покрива распон одговарајућих вредности!

Подесити наслов графикона и називе оса!

Поставити одговарајућу легенду!

Задатак 10.

Нацртати линијски графикон где X оса одговара дану у месецу а Y оса просечној температури тако да подаци за месеце јануар и октобар буду приказани засебним линијама!

За линије на графикону користити различите боје, различите врсте ознаке и различите врсте линије!

Подесити прозирност линија на графикону!

Подесити Y осу тако да почетна вредност буде 0!

Подесити наслов графикона и називе оса!

Поставити одговарајућу легенду!

...

Задатак 10.

...

Нацртати степеничасти графикон где X оса одговара дану у месецу а Y оса просечној температури тако да подаци за месеце јануар и октобар буду приказани засебним степеничастим линијама!

За степеничасте линије на графикону користити различите боје и различите врсте линије!

Подесити прозирност степеничастих линија на графикону!

Подесити Y осу тако да почетна вредност буде 0!

Подесити наслов графикона и називе оса!

Поставити одговарајућу легенду!

Основна литература

NumPy. NumPy. Internet: <https://numpy.org/>

NumPy. NumPy documentation. Internet:
<https://numpy.org/doc/stable/index.html>

pandas. Python data analysis library. Internet:
<https://pandas.pydata.org/>

pandas. pandas documentation. Internet:
<https://pandas.pydata.org/docs/>

Matplotlib. Matplotlib: Visualization with Python. Internet:
<https://matplotlib.org/>

Matplotlib. Users guide. Internet:
<https://matplotlib.org/stable/users/index>