

Uzdevums Datu analītiķa vakancei

Termiņš - 2 nedēļas

Mājas uzdevums sastāv no trīs dažādām daļām. Protams, tiek sagaidīts, ka pretendents aplūkos visus uzdevumus, bet, ja nu tomēr kāds no tiem šķiet pārāk neinteresants, vai kompetences dēļ pagaidām vēl nav izpildāms, droši var sūtīt arī kādu no darbiem neizpildītu, vai tikai daļēji izpildītu.

Visus rezultātus nepieciešams apkopot vienkopus un sagatavot rezultātu aprakstu. Izskatot pretendenta pieteikumu, tiks vērtēta gan rezultātu, gan apraksta kvalitāte!

Aprakstu un programmas kodus 2 nedēļu laikā atsūtīt uz e-pastu dace.zandersone@lvgmc.lv

Neskaidrību vai jautājumu gadījumā rakstīt uz e-pastu dace.zandersone@lvgmc.lv

I.

Uzdevumam nepieciešamie dati ir failā:

https://drive.google.com/file/d/18bxURbIoI9zXxZvum_pWsMt4BcJo2NYp/view?%20usp=sharing

Datu failā ir apkopoti meteoroloģiskā parametra minūšu novērojumu dati par vienu gadu trīs blakus esošajās stacijās (x_1 , x_2 un x_3)

	Datums.un.laiks	x_1	x_2	x_3
1	01.01 00:00	2.9	NA	2.7
2	01.01 00:01	2.2	NA	2.7
3	01.01 00:02	2.0	NA	2.5
4	01.01 00:03	2.7	NA	2.4
5	01.01 00:04	2.6	NA	2.0
...

1. Vizualizēt datus pa stacijām, izveidot kastu grafikus (*boxplot*).
2. Pārbaudīt datu kvalitāti un identificēt potenciālos izlēcējus (*outlier*) vai kļūdainas vērtības, savstarpēji salīdzinot datus starp stacijām. Izdarīt secinājumus. Ja nepieciešams, koriģēt datus.
3. Katrā stacijā aprēķināt mēneša vidējo vērtību, ņemot vērā nosacījumu, ka vidējā vērtība netiek rēķināta, ja iztrūkst vairāk par 20% vērtību mēneša laikā.
4. Modelēt datus ar sadalījumu – piemeklēt sadalījuma formas (*shape*) un mēroga (*scale*) parametrus. Piemeklēt tādu sadalījumu, kas Jūsaprāt labāk atspoguļotu datu uzvedību.

II.

Uzdevumam nepieciešamie dati ir sekojošos failos:

<https://wis.wmo.int/operational-info/VolumeC1/VolC1.txt>

https://wis.wmo.int/operational-info/GTS_routeing/ESWI/ESWIroca.txt

Fails *VolC1.txt* apraksta visu reģistrēto starptautisko telegrammu sarakstu, ko savstarpēji pārsūta meteoroloģiskie dienesti.

Datu saturs ir aprakstīts faila pirmajā rindā.

Fails ***ESWIroca.txt*** apraksta konkrētu RTH (*Regional Telecommunication Hub*) telegrammu apmaiņas shēmu starp meteo dienestiem. Tabula apraksta, kādu telegrammu, kurš un kuram meteoroloģiskajam dienestam ir pārsūtījis. Katrs meteoroloģiskais dienests var pārsūtīt savu telegrammu vairākiem datu saņēmējiem.

Telegrammu sūtītājs	Telegrammu saņēmēji
"CSEQ01 SEQU"	"RUMS", "ESWI", "XXXX"
"CSEQ01 SEQU"	"EDZW", "ESWI", "XXXX"
"CSES01 MSLP"	"EDZW", "ESWI", "XXXX"
"CSFI01 EFKL"	"EFKL", "EDZW", "EKMI", "ENMI", "ESWI", "RJTD", "RUMS", "XXXX"
"CSFI01 EFKL"	"EDZW", "EDZW", "EKMI", "ENMI", "ESWI", "RJTD", "RUMS", "XXXX"
"CSET01 ECVI"	"ECVI", "ENZI", "EVMT", "EAMT", "ECUT", "RJTD", "RUMS", "XXXX"

VoIC1. TTAaii

VoIC1. CCCC

VoIC1 CCCC

1. Veikt abu failu datu ielādi tiešā veidā ar programmēšanas valodas palīdzību. Pievienot kodu un tā aprakstu, kādā veidā tika ielādēta nepieciešamā informācija.
2. WMO datu analīze un vizualizācija. Failā **ESWIroca.txt** atrast, kādas telegrammas saņem LVĢMC. LVĢMC abreviātūra ir UMRR (CCCC). Apkopot sekojošo informāciju:
 - a. no kurām valstīm un cik daudz telegrammu ir saņemtas;
 - b. veikt telpisku datu vizualizāciju.
3. Failam VolC1.txt katru rindu saglabāt kā savu .txt failu, izveidot non-sql failu struktūru. Uzdevuma atskaitē pievienot cīļņu struktūru ar list.files vai alternatīvām komandām. Failu nosaukumus parādīt zem sekojošiem CCCC: UMRR, ESWI, EEMH.
 - a. Cīļņu struktūrai jāatbilst paraugam:
%Region%/ %RTH%/ %Country%/ %CCCC%
 - b. Failu nosaukumiem jāatbilst paraugam:
%Country% %TTAAii% %CCCC% TimeGroup%.txt

III.

Uzdevumam nepieciešamie dati ir failā:

<https://drive.google.com/file/d/1BmxOUfiEt6XpCIImKEO0xGphBIImKtV9U5/view>

Datu failā ir apkopoti prognožu dati Baltijas teritorijai no vairākiem meteoroloģiskajiem parametriem. Fails ir saglabāts .nc formātā (netcdf).

1. Veikt parametra ssr (summārā saules radiācija) vizualizāciju Latvijas apgabalam pirmajai, piecpadsmītajai un 22. stundai.
2. No faila izvilkt radiācijas datus (ssr) sekojošām pilsētām: Liepāja, Alūksne un Dagda. Veidot laikrindu, attēlojot uz grafika stundu vidējās radiācijas vērtības, konvertējot mērvienības uz W/m^2