

Određivanje mikroklimatskih područja analizom integriranih meteoroloških i geografskih podataka

Samardžić, Nikola

Master's thesis / Diplomski rad

2022

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka, Faculty of Engineering / Sveučilište u Rijeci, Tehnički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:190:455501>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-07-20**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Engineering](#)



SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski sveučilišni studij računarstva

Diplomski rad

**Određivanje mikroklimatskih područja
analizom integriranih meteoroloških i
geografskih podataka**

Rijeka, rujan 2022.

Nikola Samardžić
0069078531

SVEUČILIŠTE U RIJECI
TEHNIČKI FAKULTET
Diplomski sveučilišni studij računarstva

Diplomski rad

**Određivanje mikroklimatskih područja
analizom integriranih meteoroloških i
geografskih podataka**

Mentori:

red. prof. dr. sc. Siniša Družeta

izv. prof. dr. sc. Stefan Ivić

Rijeka, rujan 2022.

Nikola Samardžić
0069078531

Rijeka, 14. ožujka 2022.

Zavod: **Zavod za mehaniku fluida i računarsko inženjerstvo**
Predmet: **Analiza sustava i podataka**
Grana: **2.09.03 obradba informacija**

ZADATAK ZA DIPLOMSKI RAD

Pristupnik: **Nikola Samardžić (0069078531)**
Studij: **Diplomski sveučilišni studij računarstva**
Modul: **Računalni sustavi**

Zadatak: **Određivanje mikroklimatskih područja analizom integriranih meteoroloških i geografskih podataka / Determination of microclimatic areas by analysis of integrated meteorological and geographical data**

Opis zadatka:

Opisati dostupne izvore i parametre meteoroloških i geografskih podataka. Pripremiti i provesti automatsko programsko preuzimanje podataka za referentne lokalitete odabrane geografske regije te provesti osnovnu statističku obradu prikupljenih podataka. Pomoću adekvatnih numeričko-analičkih tehnika istražiti prostorno-vremenske korelacije u podacima. Na temelju rezultata analize odrediti područja koja pokazuju mikroklimatsko ponašanje. Vizualizirati rezultate analize. U završnom dijelu rada istaknuti najvažnije zaključke, komentare i preporuke za daljnji razvoj korištene metodologije.

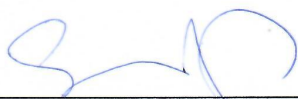
Polje znanstvenog područja: Računarstvo
Grana znanstvenog polja: Obradba informacija

Rad mora biti napisan prema Uputama za pisanje diplomskih / završnih radova koje su objavljene na mrežnim stranicama studija.



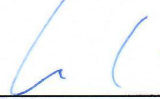
Zadatak uručen pristupniku: 21. ožujka 2022.

Mentor:



Prof. dr. sc. Siniša Družeta

Predsjednik povjerenstva za
diplomski ispit:



Prof. dr. sc. Kristijan Lenac



Izv. prof. dr. sc. Stefan Ivić (komentor)

Izjava o samostalnoj izradi rada

Izjavljujem da sam samostalno izradio ovaj rad.

Rijeka, rujan 2022.

Nikola Samardžić

Zahvala

Zahvaljujem se mentorima red. prof. dr. sc. Siniši Družeti i izv. prof. dr. sc. Stefanu Iviću na vodstvu, savjetima i stečenom znanju tijekom izrade diplomskog rada.

Hvala mojim prijateljima na podršci kroz sve godine prije, za vrijeme i poslije studija.

Najveće zahvale mojoj obitelji na bezuvjetnoj i beskonačnoj ljubavi, podršci i motivaciji.

Sadržaj

Popis slika	ix
Popis tablica	xii
Popis isječaka programskog koda	xiii
1 Uvod	1
2 Meteorološki podaci	3
2.1 Preuzimanje podataka	3
2.2 Opis podataka	4
2.3 Priprema podataka	8
2.4 Validacija podataka	9
3 Metodologija	11
3.1 Tehnologije i alati	13
3.1.1 <i>NumPy</i>	13
3.1.2 <i>Pandas</i>	14
3.1.3 <i>Plotly</i>	14
3.1.4 <i>NetworkX</i>	15
3.1.5 <i>Jupyter Notebook</i>	15

Sadržaj

3.1.6	<i>VS Code</i>	15
4	Korelacija	17
4.1	Unakrsna korelacija	17
4.2	Pearson korelacija	19
4.3	Računanje Pearson korelacije	19
4.4	Particioniranje grafa	21
5	SVD dekompozicija	23
5.1	Rang matrice	24
5.2	Lijevi singularni vektor - U	24
5.3	Singularne vrijednosti - S	24
5.4	Desni singularni vektor - V	25
6	Rezultati analize	26
6.1	Temperatura zraka	28
6.1.1	Cijeli skup podataka	28
6.1.2	Skraćeni skup podataka - ljetni mjeseci	33
6.1.3	Skraćeni skup podataka - zimski mjeseci	37
6.2	Tlak zraka	41
6.2.1	Cijeli skup podataka	41
6.2.2	Skraćeni skup podataka	44
6.3	Vlažnost zraka	45
6.3.1	Cijeli skup podataka	45
6.3.2	Skraćeni skup podataka - ljetni mjeseci	49
6.3.3	Skraćeni skup podataka - zimski mjeseci	51
7	Zaključak	55

Sadržaj

Bibliografija

57

Sažetak

60

Popis slika

2.1	Prikaz promatranih lokaliteta na geografskoj karti	7
2.2	Primjer preuzetih podataka za Bakar	8
3.1	Predviđanje rasta analitičara podataka	12
4.1	Grafički prikaz rezultata unakrsne korelacije u pet različitih promatranih točaka	18
6.1	Preciznost rekonstrukcije početne matrice	27
6.2	Korelacijska matrica za temperaturu zraka na čitavom skupu podataka	29
6.3	Prikaz četiri particije za temperaturu zraka na čitavom čitavom skupu podataka	30
6.4	Prikaz šest particija za temperaturu zraka na čitavom skupu podataka	31
6.5	Vrijednosti lijevog singularnog vektora s fokusom na koncept $k = 0$ za temperaturu zraka na čitavom skupu podataka	32
6.6	Vrijednosti desnog singularnog vektora s fokusom na koncept $k = 0$ za temperaturu zraka na čitavom skupu podataka	32
6.7	Prikaz vrijednosti desnog singularnog vektora koncept $k = 0$ na geografskoj karti za temperaturu zraka na čitavom skupu podataka . .	33
6.8	Korelacijska matrica za temperaturu zraka u ljetnim mjesecima . . .	34
6.9	Prikaz sedam particija za temperaturu zraka u ljetnim mjesecima . .	35

Popis slika

6.10	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za temperaturu zraka u ljetnim mjesecima	36
6.11	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 3$ na geografskoj karti za temperaturu zraka u ljetnim mjesecima	37
6.12	Korelacijska matrica za temperaturu zraka u zimskim mjesecima	38
6.13	Prikaz šest particija za temperaturu zraka u zimskim mjesecima	39
6.14	Vrijednosti lijevog singularnog vektora koncepta $k = 0$ za temperaturu zraka u zimskim mjesecima	40
6.15	Korelacijska matrica za tlak zraka na čitavom skupu podataka	42
6.16	Vrijednosti desnog singularnog vektora za tlak zraka na čitavom skupu podataka	43
6.17	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za tlak zraka na čitavom skupu podataka	43
6.18	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za tlak zraka u ljetnim mjesecima	45
6.19	Korelacijska matrica za vlažnost zraka na čitavom skupu podataka	46
6.20	Prikaz četiri particije za vlažnost zraka na čitavom čitavom skupu podataka	47
6.21	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za vlažnost zraka na čitavom čitavom skupu podataka	48
6.22	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za vlažnost zraka na čitavom čitavom skupu podataka	48
6.23	Prikaz četiri particije za vlažnost zraka u ljetnim mjesecima	49
6.24	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za vlažnost zraka u ljetnim mjesecima	50
6.25	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za vlažnost zraka u ljetnim mjesecima	50
6.26	Korelacijska matrica za vlažnost zraka u zimskim mjesecima	52

Popis slika

6.27	Prikaz četiri particije za vlažnost zraka u zimskim mjesecima	53
6.28	Prikaz pet particija za vlažnost zraka u zimskim mjesecima	53
6.29	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za vlažnost zraka u zimskim mjesecima	54
6.30	Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za vlažnost zraka u zimskim mjesecima	54

Popis tablica

2.1	Popis promatranih lokaliteta u analizi	4
2.2	Razlika u prosječnoj temperaturi zraka	9
2.3	Razlika u prosječnom tlaku zraka	10
6.1	Najniže izmjerene temperature zraka u zimskim mjesecima	40

Popis isječaka programskog koda

4.1	Računanje korelacija za svaki par lokaliteta	20
4.2	Funkcija za stvaranje grafa	21

Poglavlje 1

Uvod

Meteorologija je znanost koja se bavi proučavanjem atmosfere s posebnim naglaskom na vremenske procese i prognoziranje vremena. S obzirom na velik značaj vremenskih prilika na ljude i ljudske aktivnosti, meteorološka mjerenja i predviđanja uvelike utječu na svakodnevne aktivnosti. Također, ona svoju primjenu pronalazi u suradnji s mnogim granama znanosti kao što su, primjerice, aeronautika, arhitektura i medicina. Razvitak meteorologije kao znanosti danas je otvorila mogućnost njenog iskorištavanja u svakodnevnom životu za potrebe čovjeka. Za precizan opis određenog vremenskog stanja, uvjeta ili pojava, koristi se stručno nazivlje i prikladne mjerne jedinice. Primjeri meteoroloških pojava su oblaci i oborine, temperatura zraka te ciklona i anticiklona, dok mjerne jedinice opisuju pojave, a kao primjer možemo navesti tlak zraka te brzinu vjetra [1].

Iako su u većini slučajeva široj javnosti zanimljivi isključivo podaci koji govore o budućnosti i predviđanju vremenskih prilika koje se još nisu dogodile, jednako tako je važno imati širu sliku te biti upoznat s meteorološkim uvjetima u duljem ili kraćem prošlom vremenskom periodu. Takav period promatra činjenično stanje za razliku od još nerealiziranih predviđanja. Preduvjet za predviđanje vremenskih prilika u budućnosti zahtjeva poznavanje podneblja koje se istražuje te znanje o prilikama iz prošlosti. Upravo su podaci iz prošlosti u fokusu ovog rada te se njihovom analizom pokušavaju pronaći zanimljive informacije i zaključci. Osim meteoroloških podataka, bitno je u analizi iskoristiti i geografske podatke - geografsku dužinu i širinu mjesta. Prikaz odnosa među mjestima na geografskoj karti donosi atraktivan

Poglavlje 1. Uvod

prikaz povezanosti među lokalitetima. Ovom prostorno-vremenskom analizom želimo saznati kada i gdje uočavamo određene meteorološke prilike, ali također, pokušava se odgovoriti i na pitanje zašto dolazi do istih. Pronalaskom ponavljajćih uzoraka i ponašanja želimo dokazati da čak i naizgled nasumićne pojave imaju pozadinski razlog i moguće ih je znanstveno opisati.

Glavni cilj ove analize je pokušaj pronalaska mikroklimatskih područja koja se svojim meteorološkim obilježjima na određeni način karakteristićno razlikuju u odnosu na geografski bliska mjesta. Osim mikroklimatskih područja, rezultatima analize dobiti ćemo jasniji uvid meteoroloških prilika na području Primorko-goranske i Istarske županije. Prikazani rezultati predstavljaju odabrane analize iz kojih je moguće izvući zanimljive informacije.

Kroz rad biti će detaljnije opisani korišćeni podaci te njihovo preuzimanje i validacija, tehnologije i alati korišćeni za dobivanje rezultata, opis korišćenih numerićko-analitićkih tehnika te sami rezultati. Rezultati su podijeljeni prema promatranom meteorološkom obilježju i korišćenom skupu podacima.

Poglavlje 2

Meteorološki podaci

Podatak je činjenica koja može biti predočena u formaliziranom obliku, kao što su broj, riječ ili slika. Takva činjenica opisuje i kvalificira stanje određenog procesa u realnom svijetu. Možemo ga shvatiti kao apstraktnu strukturu sastavljenu od značenja (naziv i opis značenja određenog svojstva), vrijednosti (mjera i iznos) te vremena (trenutak u kojem je podatak prikupljen ili izmjeren) [2]. Dodatnom analizom, takvu apstraktnu strukturu pretvaramo u jasne informacije koje definiraju određenu situaciju. Konkretno, podaci će biti uspoređivani među lokalitetima u vremenu, no isto tako i u prostoru. Takvu analizu još nazivamo i prostorno-vremenska analiza.

2.1 Preuzimanje podataka

Prvi korak analize zahtijeva pronalazak kvalitetnih meteoroloških podataka. Kvaliteta podataka podrazumijeva dovoljno dug vremenski period, dovoljan broj mjerenja kroz dan, širok izbor dostupnih lokaliteta te više dostupnih meteoroloških obilježja kao što su, primjerice, tlak i vlažnost zraka. Uzmemo li u obzir sve navedene značajke vezane uz podatke, lako je zaključiti da je takav skup podataka izrazito teško za pronaći u besplatnoj verziji.

World Weather Online je internet stranica sa sjedištem u Londonu koja se bavi predviđanjem vremena. Kako njihov algoritam za predviđanje vremena koristi upravo podatke iz njihove baze, tako su odlučili iste podijeliti s ostatkom svijeta. Njihova

Poglavlje 2. Meteorološki podaci

glavna prednost je izrazito gust skup podataka, ali isto tako i širok raspon mogućih lokaliteta. Primjerice, osim većih mjesta, čiji su meteorološki podaci lako dostupni, *World Weather Online* također sadrži podatke za manja mjesta, pa čak i okolice velikih gradova. Podatke je moguće preuzeti koristeći njihov *Application Programming Interface (API)*. Za pristup, potrebno je prijaviti se na njihovu dvotjednu besplatnu inačicu, gdje korisnik dobiva određeni broj zahtjeva koji može uputiti po danu. Za skup podataka korišten u analizi, bilo je potrebno nekoliko dana kako bi se preuzeli željeni podaci. Sama kvaliteta i vjerodostojnost podataka detaljnije je opisana u potpoglavlju "Validacija podataka".

2.2 Opis podataka

Za uspješnu analizu od presudne važnosti bilo je pronaći i prikupiti podatke koji su raspoređeni kroz dug period, no isto tako imati dovoljan broj lokaliteta koji se nalaze na uskom području. Uzevši navedeno u obzir, odlučeno je da će promatrani lokaliteti biti oni iz Primorsko-goranske te Istarske županije. Sveukupno u ovoj analizi uspoređivati će se meteorološka obilježja 72 različita lokaliteta, pri čemu je njih 44 u Primorsko-goranskoj, a preostalih 28 u Istarskoj županiji. Popis lokaliteta uz odgovarajuću geografsku širinu i dužinu možemo vidjeti u Tablici 2.1.

Tablica 2.1 Popis promatranih lokaliteta u analizi

Mjesto	Geografska širina (N)	Geografska dužina (E)
Klana	45.4444467	14.3743376
Bakar	45.3071461	14.5342673
Kanfanar	45.1228399	13.8384985
Drenova	45.3525337	14.4297354
Tribalj	45.2231439	14.6783868
Radetići	45.1655761	13.807981
Prezid	45.6369972	14.5782552
Kraljevica	45.2736336	14.5702723
Matulji	45.3637986	14.3241057

Poglavlje 2. Meteorološki podaci

Klenovica	45.1027535	14.8457182
Motovun	45.3367292	13.8282067
Hreljin	45.2840364	14.5989301
Krk	45.0901954	14.5496681
Pula	44.8702281	13.8455311
Baška	44.9709329	14.7508013
Njivice	45.1588133	14.5395314
Skrad	45.4332541	14.9059561
Rovinj	45.0807411	13.6417282
Radovani	45.2504633	13.744803
Mrkopalj	45.3154497	14.8557128
Funtana	45.1745816	13.6052853
Mali Lošinj	44.531663	14.4694205
Radoši	45.2210988	13.6578424
Boljun	45.3019621	14.1213576
Prempen	45.15222	14.86333
Vodnjan	44.9589555	13.8513371
Opatija	45.332782	14.3029538
Punat	45.0195884	14.631247
Rabac	45.0804894	14.1530529
Tar	45.2999319	13.6262747
Kastav	45.3731673	14.3487225
Lupoglav	45.3519805	14.1088008
Fužine	45.3046471	14.7135873
Bale	45.041053	13.786641
Kampor	44.7768332	14.7279432
Poreč	45.2273829	13.5956941
Cres	44.8890769	14.383259
Buje	45.4083025	13.6588349
Jadranovo	45.2266115	14.6157613
Fažana	44.9273327	13.8050769

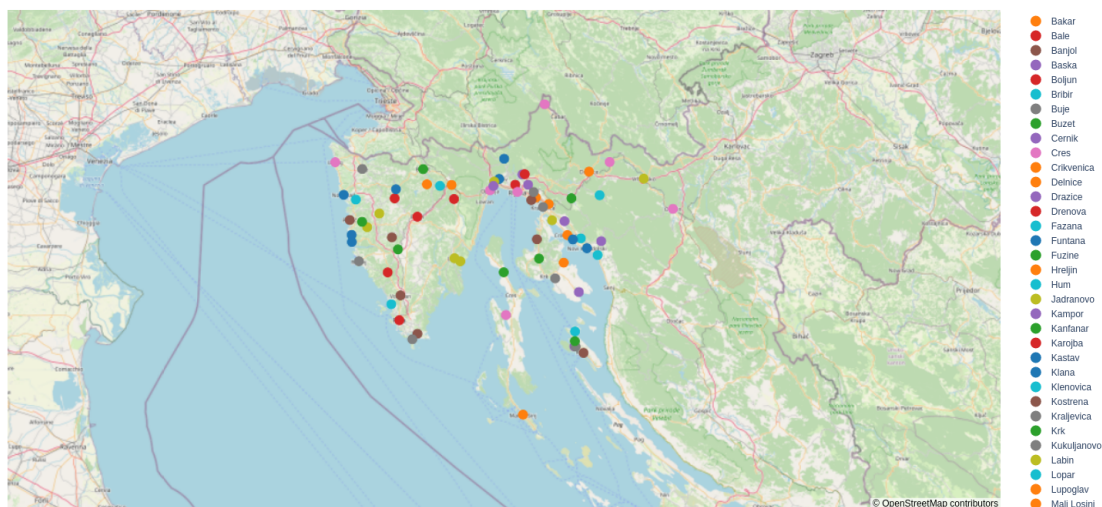
Poglavlje 2. Meteorološki podaci

Selce	45.1582241	14.7206658
Delnice	45.3991761	14.8020393
Supetarska Draga	44.7954875	14.7304429
Pazin	45.2391995	13.9363313
Cernik	45.3522868	14.4946827
Medulin	44.821786	13.9367142
Ogulin	45.2668211	15.2250286
Kukuljanovo	45.3264757	14.5227691
Radmani	45.2018779	13.6833586
Lopar	44.8299167	14.7312832
Vrsar	45.1490782	13.6063472
Crikvenica	45.1737049	14.6927906
Buzet	45.4078767	13.9655905
Podhum	45.3900903	14.4780274
Dražice	45.3889312	14.4659426
Banjol	44.7533091	14.77446
Umag	45.4323587	13.5224098
Rab	44.7756866	14.7347122
Vrbovsko	45.3740037	15.0781121
Bribir	45.1618724	14.7608879
Novi Vinodolski	45.1270044	14.7914067
Vrbnik	45.0757279	14.6744188
Predosljica	45.04139	14.37194
Karolja	45.3043468	13.8220017
Volosko	45.3483902	14.3200644
Kostrena	45.2981626	14.510908
Rijeka	45.3271752	14.4412309
Premantura	44.8021166	13.9111814
Labin	45.0913954	14.1236014
Hum	45.348526	14.0503632
Novigrad	45.3168509	13.5649454

Poglavlje 2. Meteorološki podaci

Račice	45.3541699	13.9844129
--------	------------	------------

Mjesta iz tablice prikazana na geografskoj karti mogu se vidjeti na Slici 2.1.



Slika 2.1 Prikaz promatranih lokaliteta na geografskoj karti

Svi lokaliteti imaju identičan vremenski period promatranja: početni datum promatranja je 31.10.2016., a posljednji 31.10.2021. Također, kako je jedan od atributa koji koristimo pri instanciranju objekta frekvencija, odlučeno je da će se prikupljati podaci u vremenskom razmaku od tri sata. Jednostavnim računom možemo dobiti da je sveukupno prikupljeno 1066968 redaka, a svaki od njih ima 36 stupaca. Istraživanjem podataka utvrđeno je da ne postoje prazne vrijednosti što znači da ne moramo brinuti o potencijalnim prekidima na grafovima. Primjer prikupljenih podataka nalazi se na Slici 2.2.

Poglavlje 2. Meteorološki podaci

date_time	sunrise	sunset	moonrise	moonset	tempC	moon_phase
2016-10-31 00:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	5	New Moon
2016-10-31 03:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	3	New Moon
2016-10-31 06:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	2	New Moon
2016-10-31 09:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	5	New Moon
2016-10-31 12:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	9	New Moon
2016-10-31 15:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	12	New Moon
2016-10-31 18:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	8	New Moon
2016-10-31 21:00:00	06:40 AM	04:50 PM	07:02 AM	05:38 PM	6	New Moon
2016-11-01 00:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	4	Waxing Crescent
2016-11-01 03:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	7	Waxing Crescent
2016-11-01 06:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	7	Waxing Crescent
2016-11-01 09:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	10	Waxing Crescent
2016-11-01 12:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	13	Waxing Crescent
2016-11-01 15:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	12	Waxing Crescent
2016-11-01 18:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	10	Waxing Crescent
2016-11-01 21:00:00	06:41 AM	04:49 PM	08:00 AM	06:11 PM	10	Waxing Crescent
2016-11-02 00:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	7	Waxing Crescent
2016-11-02 03:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	8	Waxing Crescent
2016-11-02 06:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	10	Waxing Crescent
2016-11-02 09:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	11	Waxing Crescent
2016-11-02 12:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	12	Waxing Crescent
2016-11-02 15:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	14	Waxing Crescent
2016-11-02 18:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	14	Waxing Crescent
2016-11-02 21:00:00	06:43 AM	04:47 PM	08:56 AM	06:48 PM	14	Waxing Crescent
2016-11-03 00:00:00	06:44 AM	04:46 PM	09:50 AM	07:30 PM	10	Waxing Crescent
2016-11-03 03:00:00	06:44 AM	04:46 PM	09:50 AM	07:30 PM	5	Waxing Crescent
2016-11-03 06:00:00	06:44 AM	04:46 PM	09:50 AM	07:30 PM	1	Waxing Crescent
2016-11-03 09:00:00	06:44 AM	04:46 PM	09:50 AM	07:30 PM	1	Waxing Crescent

Slika 2.2 Primjer preuzetih podataka za Bakar

2.3 Priprema podataka

Rezultat preuzimanja podataka bila je velika količina podataka - što gusto vremenski, što različitih meteoroloških pojava i mjernih jedinica. Kako bi ti podaci bili pregledniji, prvotno su se spremali u zasebne datoteke. Spremanje se odvijalo po principu da se stvori *Comma-separated values (CSV)* datoteka za svaki od promatranih lokaliteta. S obzirom da je glavna značajka analize i da sam cilj iste zahtjeva usporedbu podataka, bilo je smisleno izraditi novu datoteku u kojoj će se spojiti

Poglavlje 2. Meteorološki podaci

sadržaj svih datoteka. Primjerice, takva datoteka sadržavati će vrijednosti za svaki lokalitet u određenom vremenu, a usporedba među lokalitetima zahtijevati će čitanje isključivo novonastale, a ne svake od zasebnih datoteka. Uz nekoliko jednostavnih tekstualnih prepravki za određene stupce, potrebno je bilo dodati novi stupac koji označava koji red pripada kojem lokalitetu jer, nakon preuzimanje, svaka datoteka je jedinstveni lokalitet. Problem takve datoteke je njezina veličina koja iznosi oko 279 MB. Kako bi ona bila dostupna za lakšu distribuciju, kompresirana je te je kao takva dostupna za jednostavno učitavanje. Kompresijom, rezultirajuća datoteka je veličine oko 12 MB, što zasigurno olakšava distribuciju te zauzima znatno manje prostora na disku.

2.4 Validacija podataka

Državni hidrometeorološki zavod (DHMZ) je državna upravna organizacija i znanstvena organizacija koja se bavi praćenjem hidroloških i meteoroloških procesa. Osim praćenjem, bavi se i prikupljanjem, obrađivanjem i objavljivanjem podataka. Ti podaci koriste se za poslove istraživanja atmosfere i vodnih resursa te različite primjene meteorologije i hidrologije u područjima klimatologije, pomorske meteorologije, agro-meteorologije, zrakoplovne meteorologije, prostornog planiranja i projektiranja i ostalih primjena. Također, DHMZ upravlja mrežom meteoroloških i hidroloških postaja te nacionalnim bazama podataka [3]. Kako bi se ustvrdila vjerodostojnost podataka, uspoređeni su preuzeti podaci sa službenim podacima DHMZa. Prva uočena razlika je da su preuzeti podaci iz skupa racionalnih brojeva, dok su podaci DHMZa iz skupa racionalnih brojeva. Ova činjenica pokazuje da su službeni podaci precizniji. S obzirom na dostupne službene podatke, provedena je usporedba podataka za tri mjesta: Mali Lošinj, Pazin i Rijeku. Uspoređivane su prosječne vrijednosti temperature i tlaka zraka. Rezultati se nalaze u Tablici 2.2 i Tablici 2.3.

Tablica 2.2 Razlika u prosječnoj temperaturi zraka

Mjesto	Temperatura DHMZ [$^{\circ}C$]	Temperatura World Weather Online [$^{\circ}C$]
Mali Lošinj	16.69	16.95
Pazin	12.51	15.43

Poglavlje 2. Meteorološki podaci

Rijeka	15.46	11.78
--------	-------	-------

Tablica 2.3 Razlika u prosječnom tlaku zraka

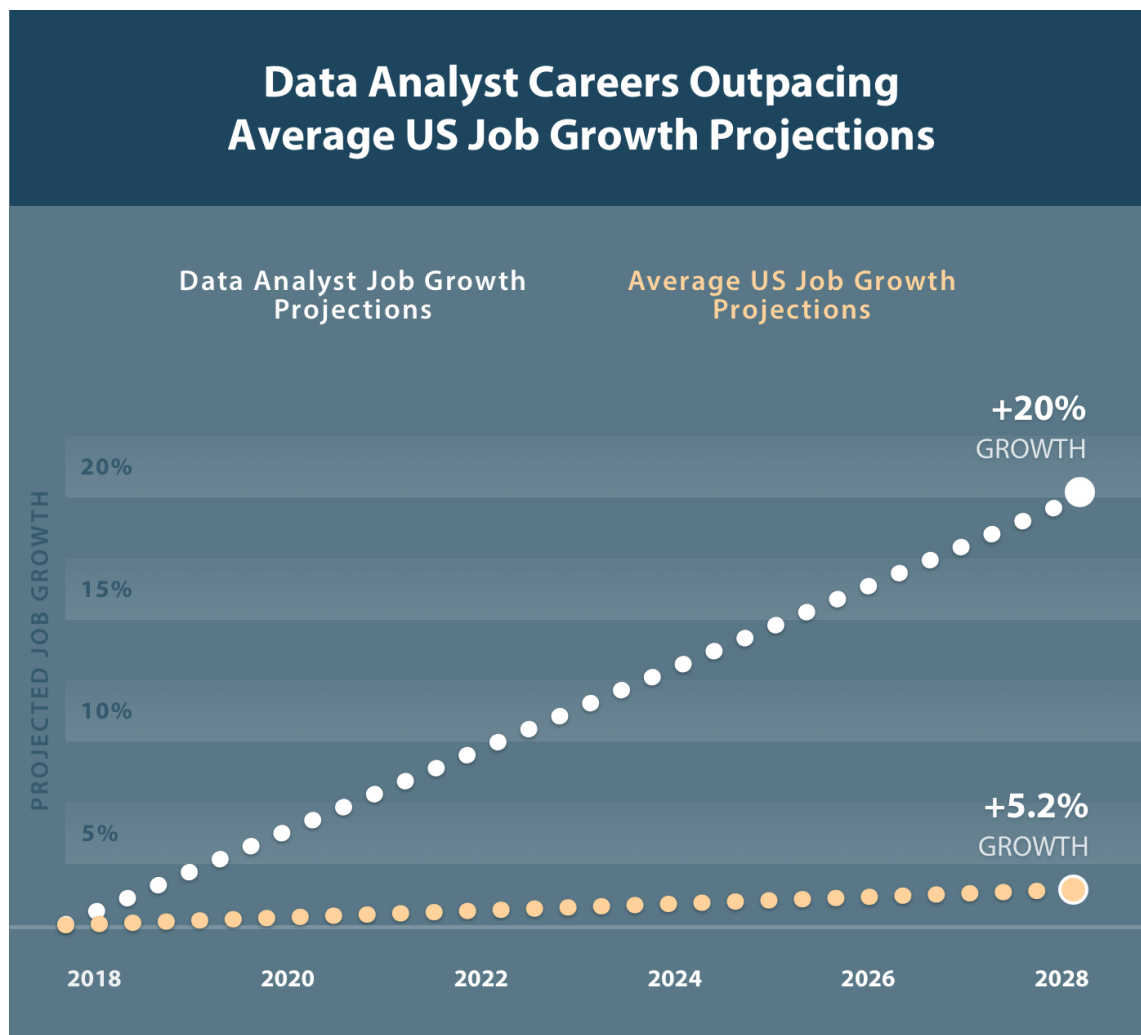
Mjesto	Tlak zraka DHMZ [hPa]	Tlak zraka World Weather Online [hPa]
Mali Lošinj	1009.71	1016.3
Pazin	981.39	1016.66
Rijeka	1001.64	1016.95

Rezultati ove jednostavne analize nam govore da za promatrane lokalitete postoji prosječno odstupanje koje može ići i do 20%. Samim time, dobiveni rezultati trebaju se uzeti s dozom opreza i imati na umu da je moguće da ne prikazuju stvarno meteorološko stanje na području Primorsko-goranske i Istarske županije.

Poglavlje 3

Metodologija

Analiza podataka je proces ispitivanja skupa podataka s ciljem pronalaska trendova i izvlačenja određenih zaključaka iz informacija koje se nalaze u podacima [4]. Neki od glavnih zadataka kod analize podataka su njihova priprema, statistička analiza te odgovarajuća vizualizacija. Danas, proporcionalno s rastom količine podataka, raste i potreba za njihovom analizom što čini posao analitičara podataka jednim od najtraženijih. Prikaz predviđanja rasta poslova analitičara podataka možemo vidjeti na Slici 3.1.



Source: Bureau of Labor



Slika 3.1 Predviđanje rasta analitičara podataka [5]

Samu analizu podataka možemo podijeliti na četiri različita tipa: opisna, dijagnostička, prediktivna i preskriptivna analiza. Opisna analiza koristi povijesne podatke iz jednog ili više izvora kako bi opisala ili trenutno ili određeno prošlo stanje kroz pronalazak trendova i uzoraka. Nastavno na opisnu, dijagnostička analiza koristi dobivene podatke i rezultate te pronalazi razloge i čimbenika za određene događaje

i situacije. S druge strane, prediktivna i preskriptivna analiza su više usmjerene budućnosti te uz upotrebu strojnog učenja, predviđanja, testiranja te ostalih tehnika pokušavaju predvidjeti budućnost te predložiti korake koje bi vodili ka željenom cilju [6]. U ovom radu, analizu možemo okarakterizirati kao opisnu, no s obzirom da je cilj pronaći uzorke i čimbenike među dobivenim rezultatima, jednako tako može biti i dijagnostička analiza.

Ovaj rad moguće je podijeliti u tri cjeline: preuzimanje podataka, korelacijska analiza i Singular value decomposition (SVD) analiza, pri čemu su svaka od analiza odvojene i neovisne jedna o drugoj, dok je preuzimanje podataka preduvjet za početak.

3.1 Tehnologije i alati

U izradi programskog koda, koji je omogućio analizu podataka i detekciju značajnih mikroklima iz meteroloških i geografskih podataka, koristio se programski jezik *Python* u verziji 3.9. Prema glasovima programera, upravo je on najkorišteniji jezik kada je u pitanju rad s podacima [7]. U svojoj srži, *Python* je objektno-orijentiran programski jezik koji se ističe svojom jednostavnom sintaksom čineći kod lako čitljivim te jednostavnijim za održavanje i dodavanje novih dijelova. Mogućnost korištenja raznih knjižnica omogućuje da *Python* svoju primjenu nalazi u gotovo svim granama programiranja. Upravo su knjižnice ono što omogućuje povezivanje s bazama podataka ili s datotekama koje sadržavaju podatke te rad s njima. Tako su se, primjerice, u ovoj analizi koristile knjižnice koje su olakšale i optimizirale pribavljanje podataka, statističke izračune i vizualizaciju te svojim klasama i metodama omogućile da kod bude čitljiv te jednostavan za potencijalne promjene ili dodatke. U nastavku, opisane su najkorištenije knjižnice koje su bile neophodne za analizu.

3.1.1 *NumPy*

NumPy je *Python* knjižnica koja podržava rad na velikim količinama podataka koji se nalaze u poljima ili matricama. Također, uz nju dolazi i velik broj matematičkih operacija koje se mogu izvršavati na danim poljima i matricama. Za

Poglavlje 3. Metodologija

razliku od liste iz standardne *Python* knjižnice, obvezno je da svaki element bude istog tipa podatka, što pridonosi brzini izvršavanja određenih operacija, no i čitavog koda. Konkretno, u ovoj analizi, glavni zadatak ove knjižnice je, uz pripremu podataka, izvršavanje Pearson korelacije i SVDA - dvije statističke analize kojima tražimo posebnosti iz meteoroloških podataka te pokušavamo pronaći mikroklima.

3.1.2 *Pandas*

Pandas je knjižnica napisana za programski jezik *Python*, a koristi se najčešće za učitavanje, obradu i spremanje podataka. Nudi razne strukture podataka i metode kojima je znatno olakšan rad te izvođenje matematičkih operacija, bilo na malom ili većem skupu podataka [8]. Među najpopularnijim klasama unutar knjižnice je *DataFrame* - struktura podataka koja nalikuje na tablicu relacijskih baza podataka. Upravo je ona temeljna struktura koja je korištena za učitavanje, obradu i spremanje podataka. Metode koje koristimo, vrlo lako se mogu povezati sa Structured Query Language (SQL) programskim jezikom koji služi upravo za operacije na tablici te manipuliranjem podataka unutar iste.

3.1.3 *Plotly*

Plotly je kompanija koja se bavi razvojem alata koji se koriste za analizu podataka te njihovu vizualizaciju. Njihovi alati za grafiku, analizu i vizualizaciju su lako dostupni na internetu, no isto tako i u obliku knjižnice. Takva knjižnica, kao i ostale, može lako biti iskorištena u kodu [9]. Kako je *Python* knjižnica razvijena kao nastavak na *Javascript* verziju knjižnice, dobiveni grafovi, karte te ostali vizuali su interaktivni te je moguće klikati po njima, skraćivati promatrani period, zumirati, odabrati što se želi prikazati te druge slične funkcionalnosti [10]. Upravo te funkcionalnosti, čine dobivene rezultate vrlo atraktivnima te u dosta slučajeva, zahvaljujući raznim mogućnostima ove knjižnice, nije potrebno prilagođavati kod ili dodavati nove grafove.

3.1.4 *NetworkX*

U matematici i računarstvu, teorija grafova je grana koja se bavi proučavanjem matematičkih struktura (grafova) korištenih radi predstavljanja odnosa između dva elementa određene korelacije. Prema gruboj definiciji, graf je skup vrhova (čvorova) povezanih linijama koje nazivamo bridovi. Prema tome, graf je moguće definirati kao odnos između tih skupova: svaki brid spaja dva vrha. Grafove se prikazuju crtanjem točaka za svaki vrh i povlačenjem linija između dvaju vrhova, ako ih povezuje brid [11]. *NetworkX* je *Python* knjižnica koja služi za izradu, manipulaciju i proučavanje strukture, dinamike i funkcija kompleksnih grafova. Omogućava nam strukturiranje podataka u obliku željenog grafa te analizu i rad na istom [12].

3.1.5 *Jupyter Notebook*

Jupyter Notebook je server-klijent aplikacija koja dozvoljava uređivanje i pokretanje bilježnica u internet pregledniku ili u nekoj drugoj aplikaciji. Bilježnice su dokumenti koji sadrže programski kod koji se izvršava, rezultate izvršavanja koda i komentare te opise koji pojašnjavaju sam kod ili rezultate [13]. Nadalje, rezultirajuću bilježnicu moguće je jednostavno podijeliti s drugim osobama, preuzeti te spremi lokalno i pretvoriti u druge formate, kao što su *Portable Document Format (PDF)* ili *HyperText Markup Language (HTML)*. Bilježnica je podijeljena u ćelije - dijelove programskog koda koje je moguće izvršiti zasebno, bez potrebe da se pokrenu ostale. Iako se ćelije mogu pokretati zasebno, svaka ima "pamćenje" i može koristiti knjižnice, varijable, funkcije i objekte iz prethodnih ćelija. Jezgra bilježnice je "motor" koji je zadužen za pokretanje koda koji se nalazi u ćelijama. Osim što *Jupyter Notebook* podržava izvršavanje *Python* koda, moguće je stvaranje bilježnica i u drugim programskim jezicima, kao što je *R*.

3.1.6 *VS Code*

Visual Studio Code, poznat i kao *VS Code*, je alat za uređivanje koda kojeg je razvio *Microsoft*. Prvi put se pojavio 2015. godine, a nedugo nakon izvorni kod je podijeljen na *GitHub* te je dostupan svima. Glavne značajke ovog popularnog alata

Poglavlje 3. Metodologija

su podrška za korištenje *debuggera*, isticanje sintakse, prijedlozi za kod, integracija s *Gitom* i mnoge druge [14]. Dodavanjem proširenja, moguće je koristiti ga i kao alat u kojem možemo uređivati i pokretati bilježnice. Zbog svoje široke primjere, razvijena je zajednica što olakšava rješavanje problema, no isto tako prati trendove te razvija proširenja i dodatke za gotovo sve programske jezike. Također, jedan od glavnih razloga zašto je korišten u ovom radu, je mogućnost kreiranja *Python* bilježnica ekstenzije `.ipynb` te nije potrebno koristiti dodatan alat, već su i pomoćne skripte (npr. preuzimanje podataka) i same analize odrađene kroz *VS Code*.

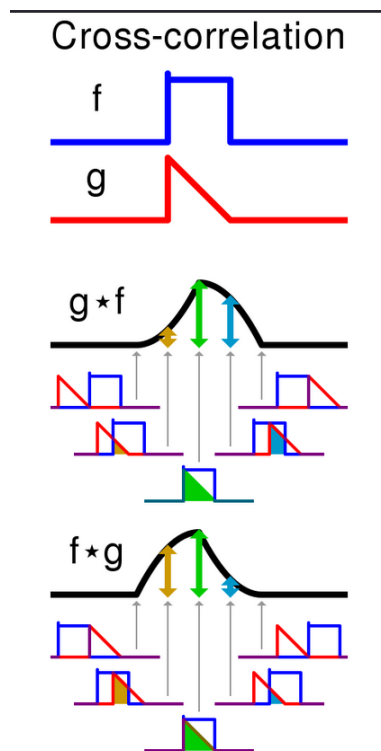
Poglavlje 4

Korelacija

Prema definiciji, u statistici, korelacija ili ovisnost je statistička povezanost, bilo slučajna ili namjerna, između dvije nasumične varijable ili dva vektora. Povezanost dvaju varijabli znači da porast jedne znači i proporcionalan porast druge, ili obrnuto proporcionalan pad kod negativne korelacije. Razlikujemo dva tipa varijabli: neovisnu i ovisnu varijablu [15]. Kao meteorološki primjer, možemo uzeti vektor koji predstavlja datume i drugi vektor koji sadrži vrijednosti prosječne temperature u nekom mjestu. Vektor s datumima je konstantan i vrijednosti prosječne temperature ne utječu na njega te je zbog toga neovisna varijabla. S druge strane, znamo da datum i koje je godišnje doba utječu na to koja je prosječna temperatura te je zbog toga vektor temperature zavisna varijabla. Također, moguća je situacija kada su i jedna i druga varijabla zavisne varijable gdje su vrijednosti oba vektora ovisni o vrijednostima onog drugog.

4.1 Unakrsna korelacija

Unakrsna korelacija (eng. *cross correlation*) je mjera sličnosti između dva signala u funkciji pomicanja jedne u odnosu na drugu. Grafički prikaz unakrsne korelacije može se vidjeti na Slici 4.1.



Slika 4.1 Grafički prikaz rezultata unakrsne korelacije u pet različitim promatranih točaka [15]

Često se koristi u prepoznavanju uzoraka, neuropsihologiji i kriptanalizi. U vjerojatnosti i statistici, ona označava podudarnost između dva vektora u određenom vremenskom periodu. Moguće ju je računati i na diskretnim, ali i na kontinuiranim signalima [16]. Kod kontinuiranih signala, unakrsnu korelaciju može se opisati formulom:

$$(f * g)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{f(t)}g(t + \tau)dt \quad (4.1)$$

Što je isto kao:

$$(f * g)(\tau) \triangleq \int_{-\infty}^{\infty} \overline{f(t - \tau)}g(t)dt \quad (4.2)$$

pri čemu $\overline{f(t)}$ označava konjugirano kompleksni broj funkcije $f(t)$, a τ predstavlja

Poglavlje 4. Korelacija

kašnjenje. Kao što je i u slučaju ove analize, signali f i g mogu biti kontinuirani i imati poznat period T . U takvom slučaju, moguće je beskonačne granice integrala zamijeniti granicama perioda te dobivamo sljedeću formulu [16]:

$$(f * g)(\tau) \triangleq \int_{t_0}^{t_0+T} \overline{f(t)}g(t + \tau)dt \quad (4.3)$$

4.2 Pearson korelacija

Pearson korelacija je podvrsta unakrsne korelacije za koju vrijedi da ne vršimo pomak na bilo kojem od signala. Drugim riječima, uspoređuju se vrijednosti koje se nalaze na istoj poziciji u vektoru ili, u našem slučaju, vrijednosti koje su izmjerene u istom trenutku. Prema definiciji, ona se dobiva dijeljenjem koeficijenta kovarijance između dva signala (vektora) s produktom njihovih standardnih devijacija [17].

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (4.4)$$

Kako nam je ideja uspoređivati vrijednosti izmjerene u identičnom trenutku, dovoljno je koristiti samo Pearson korelaciju jer pomak u vremenu bilo kojeg signala nije potreban. Kod Pearson korelacije, rezultat može biti veći ili jednak od -1.0 i manji ili jednak 1.0. Ta vrijednost opisuje "jačinu" korelacije između dva promatrana signala, u našem slučaju dva lokaliteta za odabranu meteorološku pojavu ili mjernu jedinicu. Korelacija u vrijednosti 1.0 predstavlja savršeno podudaranje između dva signala, a korelacija od -1.0 predstavlja negativno koreliranje među signalima.

4.3 Računanje Pearson korelacije

Prvi preduvjet izračunavanja Pearson korelacije je da imamo dva signala. U našem slučaju, signali bi predstavljali sve vrijednosti lokaliteta za odabranu meteorološku pojavu ili mjernu jedinicu. Iako je moguće računati korelaciju korištenjem formule, knjižnica *NumPy* u sebi ima gotovu funkciju *corrcoef*, koja kao argumente, u

Poglavlje 4. Korelacija

našem slučaju, prima dva signala. Rezultat takvog poziva funkcije je matrica dimenzija 2 x 2, pri čemu su vrijednosti na sporednoj dijagonali predstavljaju vrijednost Pearson korelacije između dva signala [18].

Računanje Pearson korelacije odvija se u parovima: za svaki lokalitet računa se korelacija s ostalim mjestima. Kako imamo ukupno 72 mjesta koja promatramo, želimo dobiti vrijednost korelacije svakog mjesta s preostalim 71. Ako uzmemo u obzir da vrijedi sljedeća jednakost, tada znamo da je izračun za dva lokaliteta potrebno računati samo jednom:

$$\text{correlation}(f, g) = \text{correlation}(g, f) \quad (4.5)$$

```
{
# function to calculate correlation matrix values
def create_correlation_matrix(data, towns, field):
    if len(field) == 1:
        field = field[0]

    towns_cnt = len(towns)
    # init zero matrix with m=n=count of cities
    # set values to -13, just to be sure it is an imposible correlation value
    ret_matrix = np.zeros((towns_cnt, towns_cnt)) - 13

    # iterate through every city combination and calculate the correlation
    for i, town1 in enumerate(towns):
        town1_values = np.array(data.loc[data['city'] == town1][field])

        # correlation 1 on diagonal
        ret_matrix[i, i] = 1.0

        # having in mind that ret_matrix[i, j] == ret_matrix[j, i]
        for j, town2 in enumerate(towns[i+1:], i+1):
            town2_values = np.array(data.loc[data['city'] == town2][field])

            ret_matrix[i, j] = np.corrcoef(town1_values, town2_values)[0, 1]
            ret_matrix[j, i] = ret_matrix[i, j]

    return ret_matrix
}
```

Ispis 4.1 Računanje korelacija za svaki par lokaliteta

Iz koda u Ispisu 4.1, možemo vidjeti da se rezultati koje dobijemo spremaju u matricu. Takvu kvadratnu matricu još zovemo i korelacijska matrica. Značajke takve matrice su da je broj redaka i broj stupaca jednak broju promatranih lokaliteta, da su sve vrijednosti na dijagonali jednake 1.0 jer svaki lokalitet u potpunosti korelira sa samim sobom i da je matrica dijagonalno simetrična zbog jednakost koju vidimo

na formuli 4.5.

4.4 Particioniranje grafa

Rezultirajuća korelacijska matrica daje nam vrijednosti koje opisuju postotak korelacije između svakog para lokaliteta. Osim same matrice, zanimljivo bi bilo prikazati rezultate na geografskoj karti. Ideja iza particioniranja grafa je dobiti više manjih particija te pokušati vidjeti odgovaraju li dobivene particije geografskom položaju mjesta ili će postojati neka mjesta koja se nalaze u particiji s mjestima s kojima, u teoriji, ne bi trebala biti.

Prvi korak je kreiranje grafa u kojem je svako mjesto povezano sa svakim, osim sa sobom. Kako je opisano ranije, graf je skup vrhova povezanih bridovima. U ovom slučaju imamo 72 vrha - po jedan za svaki lokalitet. S obzirom su svi međusobno povezani, jednostavnim računom dobijemo da broj bridova iznosi 2556. Kako je korelacija parametar koji opisuje povezanost između dva lokaliteta, bridu je pridružena težina u odgovarajućoj vrijednosti. Kreiranje grafa izvedeno je uz korištenje knjižnice *NetworkX* (*nx*) i možemo ga vidjeti u Ispisu 4.2.

```
{
# function to create graph from correlation matrix
def create_graph(corr_matrix, towns_index, data_geo):
    G = nx.Graph()
    for i in towns_index:
        for j in towns_index[i+1:]:
            G.add_edge(i, j, weight=corr_matrix[i, j])

    return G
}
```

Ispis 4.2 Funkcija za stvaranje grafa

Tako dobiveni graf potrebno je podijeliti u više manjih. Istraživanjem se došlo do algoritma *Louvain Community Detection* u sklopu *NetworkX* knjižnice koji svojim idejama i funkcionalnostima zadovoljava potrebe analize.

Poglavlje 4. Korelacija

Louvain Community Detection je algoritam čiji je zadatak, uz korištenje heurističke metode bazirane na optimizaciji modularnosti, pronaći particije inicijalnog grafa. Algoritam se izvršava u dva koraka. U prvom koraku, svaki vrh je dodijeljen u vlastitu particiju te se za svaki ostali pokušava pronaći maksimalna pozitivna modularnost na način da se pokuša pridružiti svakom od susjeda. U slučaju neuspjeha, vrh ostaje u originalnoj particiji. Sama modularnost vrha računa se na sljedeći način:

$$\Delta Q = \frac{k_{i,in}}{2m} - \gamma \frac{\sum_{tot} \cdot k_i}{2m^2} \quad (4.6)$$

pri čemu je m veličina grafa, $k_{i,in}$ zbroj težina bridova koji idu iz i u druge vrhove unutar particije, k_i predstavlja zbroj težina bridova povezanih s vrhom i , \sum_{tot} je zbroj svih bridova unutar particije, a γ predstavlja rezoluciju - parametar koji određuje hoće li algoritam favorizirati manje ili veće particije. Prva faza završava u trenutku kada nije moguće više poboljšati modularnost. Drugi korak sastoji se od stvaranja nove mreže čiji su vrhovi sada particije pronađene u prvom koraku. To se postiže na način da su težine bridova između vrhova sada zbroj težine bridova između povezanih vrhova [19].

Sama funkcija *louvain_communities* kao argument prima graf koji se dobije korištenjem funkcije s Ispisa 4.2 te rezoluciju. Početno zadana vrijednost rezolucije je 1.0, a njenim povećavanjem algoritam favorizira manje particije (veći broj particija), odnosno smanjivanjem veće particije (manji broj particija). Postepenim povećavanjem vrijednosti rezolucije kroz petlju možemo pokušati vratiti željeni broj particija. Također, zbog sličnosti u promatranim podacima, treba paziti da je broj željenih particija realan te kroz iteracije pronaći optimalan maksimalan broj particija koje se mogu pronaći iz inicijalno stvorene mreže.

Poglavlje 5

SVD dekompozicija

Dekompozicija singularnih vrijednosti (eng. *Singular value decomposition*) ili skraćeno SVD je metoda faktorizacije složenih realnih ili kompleksnih matrica. Početno zadana matrica rastavlja se na nekoliko manjih od koje svaka otkriva zanimljive i korisne osobine inicijalne [20]. Dimenzija je $m \times n$ pri čemu reci predstavljaju datum to jest vrijeme mjerenja, s stupci promatrane lokalitete. Same vrijednosti unutar matrice su vrijednosti nekog od atributa kojeg koristimo u mjerenjima. Tako ćemo, primjerice, u početnoj matrici imati vrijednosti temperature gdje će vrijednost ćelije predstavljati iznos temperature zraka određenog lokaliteta u danom trenutku. Takva početna matrica rastavlja se na umnožak triju matrica:

$$M = USV^T \tag{5.1}$$

pri čemu se matrica U naziva lijevi singularni vektor, matrica S sadrži singularne vrijednosti, a V^T predstavlja vrijednosti desnog singularnog vektora [21]. U sklopu knjižnice *NumPy*, u *Pythonu* možemo dobiti tražene matrice. Funkcija *svd* kao argument prima početnu matricu koju želimo rastaviti. Također, dobro je postaviti argument *full_matrices* kao *True* kako bi rezultirajući lijevi i desni singularni vektor mogli biti kvadratne matrice [22].

5.1 Rang matrice

Rang matrice u SVDu je metoda kojom aproksimiramo vrijednosti unutar dobivenih matrica. Aproksimacija se vrši na način da odaberemo vrijednost k za koju vrijedi:

$$k \leq \min(m, n) \quad (5.2)$$

Pri čemu su vrijednosti m i n broj redaka, odnosno broj stupaca početne matrice na kojoj vršimo dekompoziciju singularnih vrijednosti. U slučaju da se odluči koristiti rang matrice, novonastale tri matrice moguće je aproksimirati korištenjem prvih k singularnih vektora. Na taj način, u matricama ostavljamo samo one koncepte za koje smatramo da su vrijedni istraživanja. Također, dimenzije matrice mogu se značajno promijeniti čime smanjujemo njihovo korištenje memorije te optimiziramo daljnju analizu. Kao dodatan primjer korištenja ranga matrice, možemo uzeti digitalne slike koje, nakon pronalaženja optimalne vrijednosti k , ne gube previše na kvaliteti, ali zbog kompresije zauzimaju manje prostora te su jednostavnije za distribuciju.

5.2 Lijevi singularni vektor - U

Matrica U je kvadratna matrica dimenzija $m \times r$ gdje je m broj redaka originalne matrice, a r rang matrice. Kako se na x osi nalaze vrijednosti koje predstavljaju datum mjerenja podataka, tako možemo zaključiti da ova matrica govori o odnosu datuma i koncepta. Sami rezultati prikazani na grafu nisu intuitivni te u početku nismo sigurni što predstavljaju. Detaljnijom analizom i poznavanjem podataka i teme koju analiziramo, moguće je otkriti značenje određenih konceptata. Shvaćanje značenja koncepta daje nam jasniji uvid u rezultate te pojašnjava iste.

5.3 Singularne vrijednosti - S

Matrica singularnih vrijednosti S je matrica dimenzija $r \times r$ gdje je r rang matrice. Glavna značajka matrice je da se vrijednosti koje nisu nula nalaze isključivo na

Poglavlje 5. SVD dekompozicija

dijagonali matrice čineći je dijagonalnom matricom. Iz toga razloga, matrica S se može shvatiti kao vektor. Vrijednosti koje se nalaze u vektoru govore o važnosti koncepta. Također, u takvom vektoru vrijedi da je svaka vrijednost veća od sljedeće:

$$s_0 > s_1 > s_2 > \dots > s_n > 0 \quad (5.3)$$

Same dobivene vrijednosti ove matrice (vektora) govore nam o "jačini" koncepta u odnosu na ostale.

5.4 Desni singularni vektor - V

Matrica V^T je kvadratna matrica dimenzija $r \times n$ gdje je r rang matrice, a n broj stupaca originalne matrice. Za razliku od matrice U , ova matrica govori o odnosi između gradova i koncepta. Sam graf je čitljiviji i jednostavniji nego kod odnosa datuma i koncepta te je intuitivan i bez dodatne analize može ponuditi određene zaključke. Također, kao i kod korelacije, ove rezultate moguće je prikazati na geografskoj karti što je još jedan dodatan i zanimljiv način prikaza odnosa gradova i koncepta.

Poglavlje 6

Rezultati analize

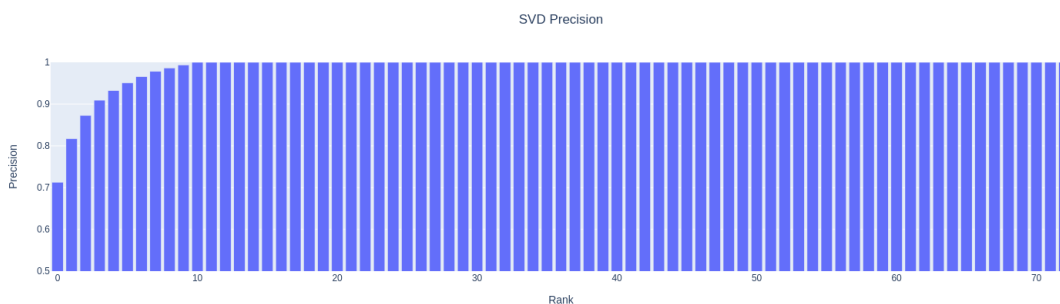
Kako je i rečeno ranije, glavni cilj analize je pokušati pronaći mikroklimatska područja koja se ističu od geografski bliskih mjesta. Također, želimo provjeriti općenita meteorološka obilježja promatranog područja te pokušati opisati razlog za uočene pojave. Dvije korištene numeričko-analitičke tehnike su Pearson korelacija i SVD. Obje tehnike implementirane su u *Python* knjižnici *NumPy* te se u samom kodu koriste kao metode.

Rezultati korelacije koje ćemo proučavati su korelacijska matrica te particioniranje grafa. Kako imamo 72 promatrana lokaliteta, svaka rezultirajuća matrica imati će isto toliko redaka i stupaca. Vrijednosti na dijagonali su uvijek 1.0, a ostale vrijednosti govore o odnosu među dva lokaliteta. Kod particioniranja grafa, istraživanje je pokazalo da je optimalno pokušati pronaći najmanje četiri i ne više od sedam particija. Manje od četiri particije su uvijek povezane kako bismo i očekivali s obzirom na geografska obilježja, a osam ili više particija se pokazalo kao pretjerano te se algoritam forsira pronaći nove particije što često rezultira stvaranjem njihovog velikog broja. Primjerice, nakon što algoritam pronađe sedam particija, u sljedećoj iteraciji ih pronađe više od 20. Tako velik broj particija ne dijele lokalitete prema dobivenim rezultatima, već zbog petlje nastoji pronaći nove particije čak i kada one ne postoje.

S druge strane, SVD koristimo kako bismo analizirali dva odnosa: datuma i koncepta te mjesta i koncepta. Kako je vektor singularnih vrijednosti sortiran od najvećeg elementa prema najmanjem, tako uvijek možemo znati da su i koncepti po-

Poglavlje 6. Rezultati analize

redani po važnosti koncepta. U svakoj od analiza, koristio se rang matrice k čija je vrijednost iznosila pet. U prijevodu, u analizi će uvijek biti promatrano pet najvažnijih koncepata te će se zaključci donositi na osnovu njih. Preciznost rekonstrukcije s tako zadanim k uvijek iznosi preko 95%, što pokazuje da je postavljena vrijednosti vjerodostojna i dovoljno precizna da takvi budu i rezultati. Primjer preciznosti rekonstrukcije početne matrice za sve vrijednosti k može se vidjeti na Slici 6.1.



Slika 6.1 Preciznost rekonstrukcije početne matrice

U rezultatima ove analize promatrati će se nekoliko različitih vrijednosti. Točnije, u fokusu će biti temperatura zraka (tempC) izražena u Celzijevim stupnjevima, tlak zraka (pressure) izražen u hektopaskalima te vlažnost zraka (humidity) prikazana u postocima. Svaka od navedenih vrijednosti uspoređivati će se na svim promatranim lokalitetima. Također, osim prostorne analize na čitavom prikupljenom periodu, promatrati će se i isključivo ljetni ili zimski mjeseci. Kod ljetnih mjeseci, korišteni su podaci prikupljeni u lipnju, srpnju i kolovozu dok su za zimske mjesece podaci prikupljeni u prosincu, siječnju i veljači. Prikazivati će se rezultati isključivo ukoliko opisuju neke dodatne i zanimljive vrijednosti. Primjerice, ukoliko su za vlažnost zraka rezultati slični i ljeti i zimi, opisati će se samo jedna situacija. Drugim riječima, prikazani rezultati su odabrane analize iz kojih je moguće izvući zanimljive informacije i saznanja o meteorološkim obilježjima promatranog područja.

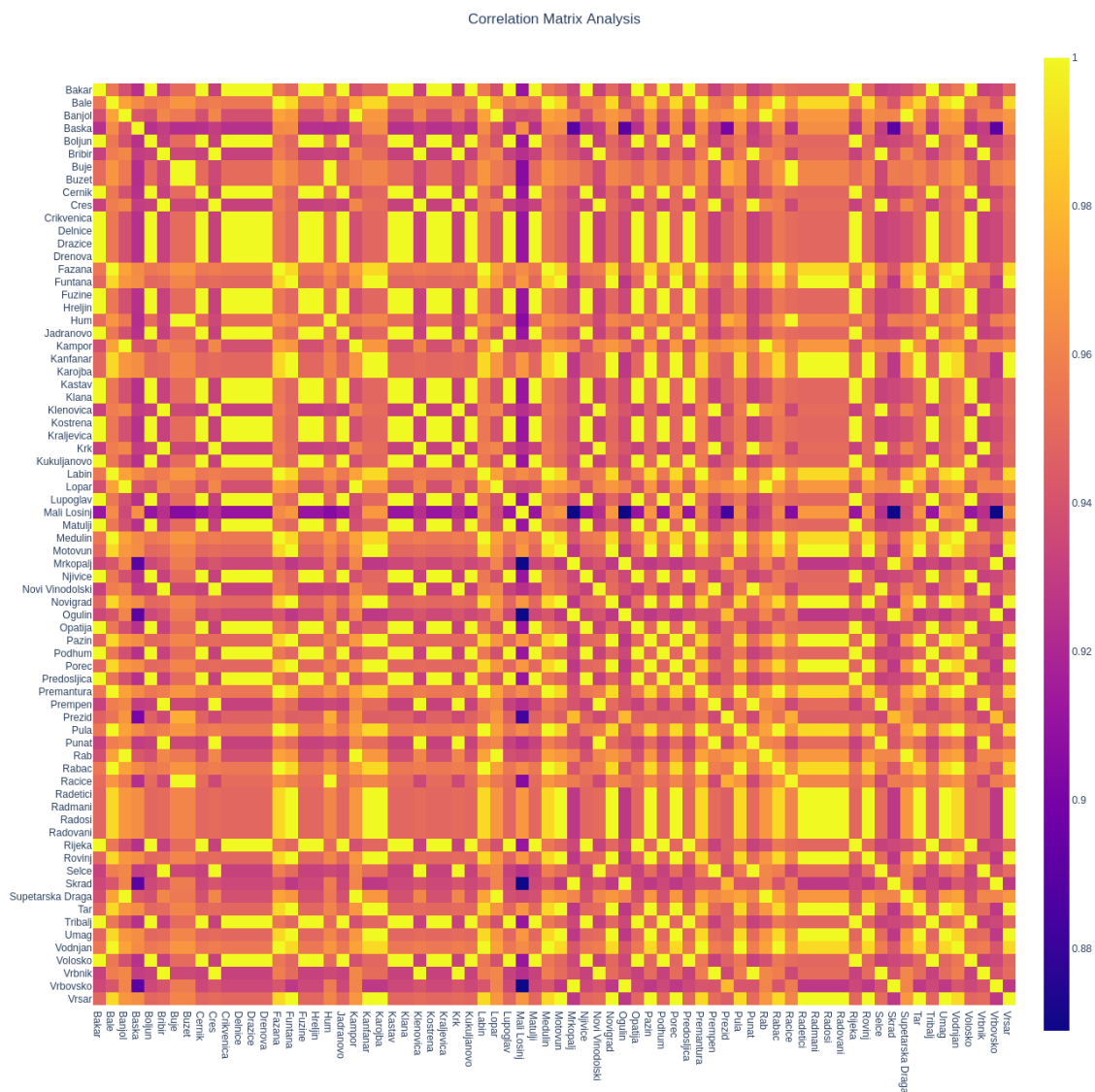
6.1 Temperatura zraka

Temperatura zraka je mjera koja nam govori koliko je zrak topao ili hladan. Jedna je od najpopularnijih i najpraćenijih meteoroloških pojava, a informacije vezane uz nju dostupne su i praćene na dnevnoj bazi. Može se izražavati u više mjernih jedinica, no u ovoj analizi korištene su vrijednosti u Celzijevim stupnjevima.

6.1.1 Cijeli skup podataka

Analiza na čitavom skupu podataka potvrdila je određena očekivanja, no isto tako istaknula određene lokalitete. Zbog velikog broja podataka i geografske blizine među lokalitetima, vrijednosti u korelacijskoj matrici se visoke. Vrijednosti koje se nalaze u dobivenoj matrici su veće od 0.86 i manje ili jednake 1.0. Iako su vrijednosti korelacije između lokaliteta 1.0, zbog velikog broja podataka možemo zaključiti da je ta vrijednost zaokružena te da ipak postoje razlike u uspoređivanim izmjerenim vrijednostima. Dobivena korelacijska matrica za cijeli skup podataka nalazi se na Slici 6.2.

Poglavlje 6. Rezultati analize

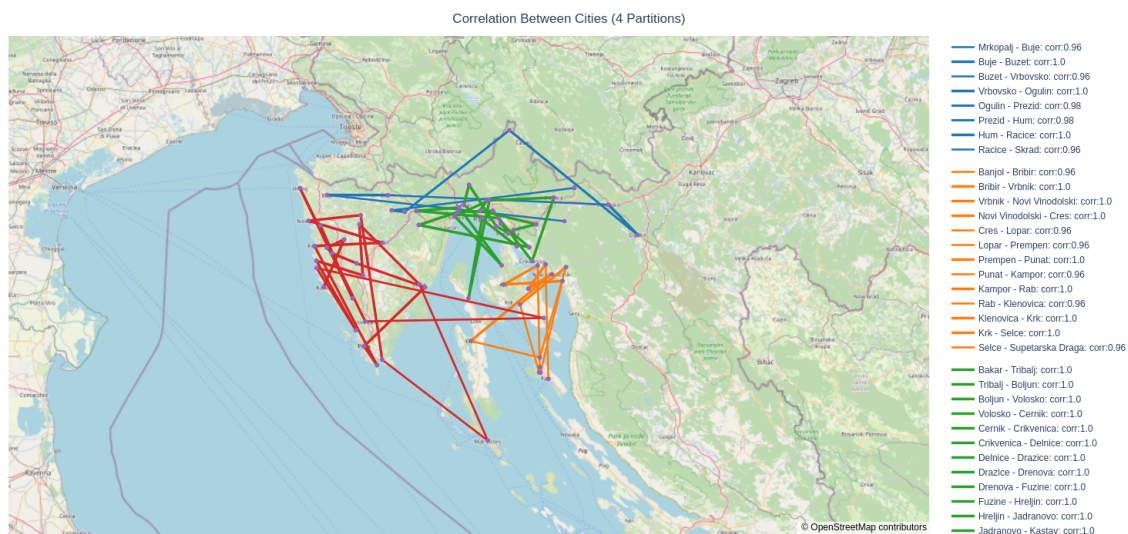


Slika 6.2 Korelacijska matrica za temperaturu zraka na čitavom skupu podataka

Već na korelacijskoj matrici možemo vidjeti da je većina polja "svjetlije" boje te da lokaliteti imaju izraženu korelaciju. Međutim, možemo uočiti dva lokaliteta koja se ističu: Baška i Mali Lošinj. Vrijednosti korelacije za ta dva lokaliteta su često bliže donjoj granici vrijednosti nego što su gornjoj. Sudeći po rezultatima s matrice, Baška i Mali Lošinj su zasigurno lokaliteti na koje treba obratiti pažnju u prostornoj analizi. Upravo takav prikaz promatranih lokaliteta na geografskoj karti možemo

Poglavlje 6. Rezultati analize

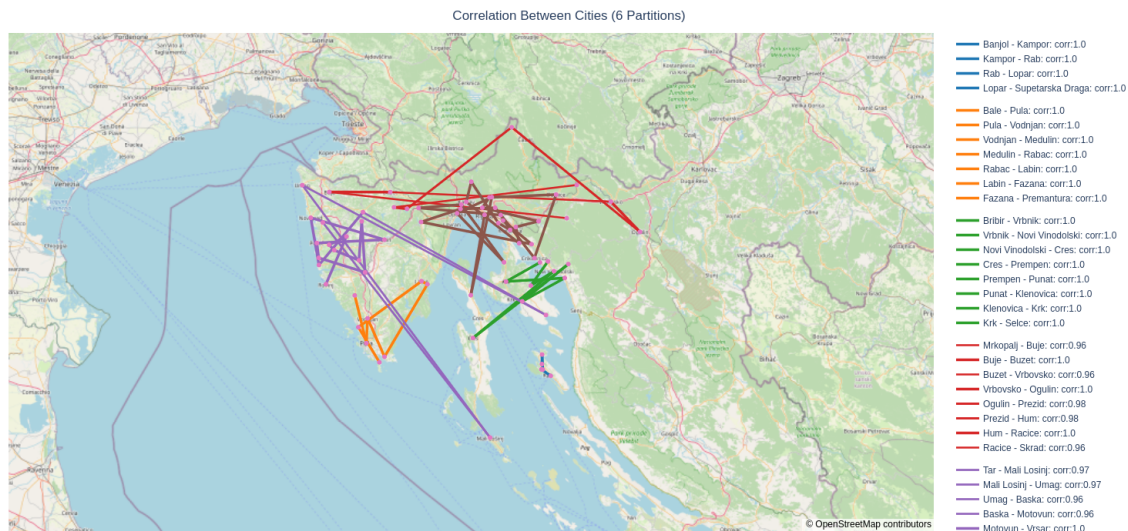
vidjeti na Slici 6.3. Uočavamo da su lokaliteti podijeljeni u četiri particije, što je ujedno i najmanji broj particija koji promatramo.



Slika 6.3 Prikaz četiri particije za temperaturu zraka na čitavom čitavom skupu podataka

Uočavamo da se već u grafu s četiri particije Baška i Mali Lošinj potvrđuju kao lokaliteti koji su povezani s udaljenijim mjestima, a ne s geografski bližim. U ovom slučaju, algoritam nije pronašao najveći mogući broj particija (sedam), već je prvotno zadan graf moguće rastaviti u najviše šest particija. Takav prikaz sa šest particija nalazi se na Slici 6.4. Možemo uočiti nastanak dvaju novih particija: područje otoka Raba te područje južne Istre. Također, Baška i Mali Lošinj su povezani sa Zapadnom Istrom, iako postoji nekoliko particija koje bi geografski bile prikladnije.

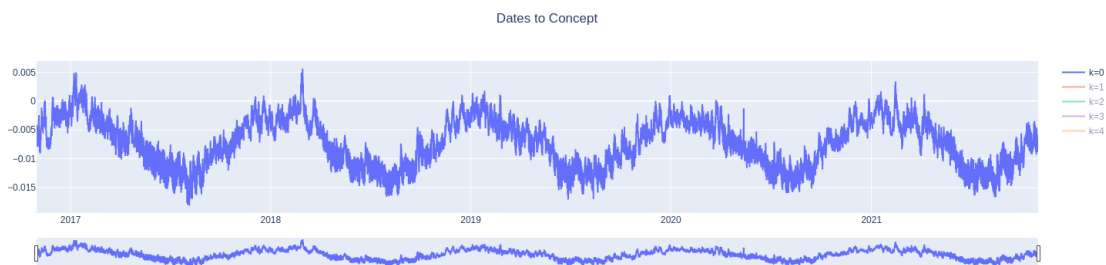
Poglavlje 6. Rezultati analize



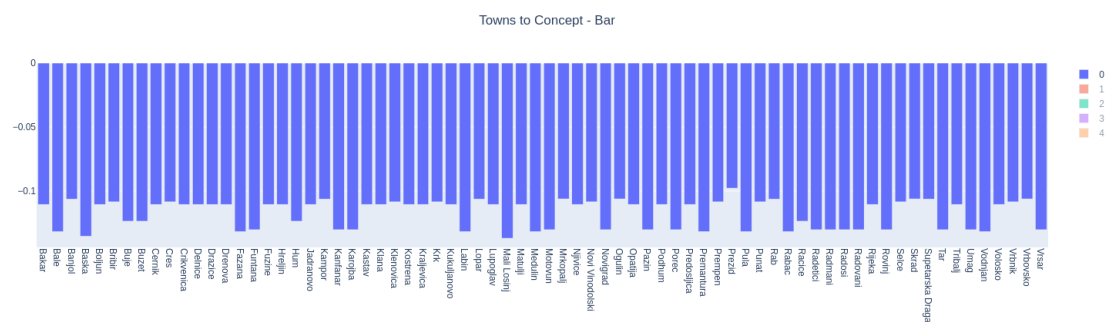
Slika 6.4 Prikaz šest particija za temperaturu zraka na čitavom skupu podataka

U slučaju kada promatramo temperaturu zraka na cijelom skupu podataka, uz rang matrice $k = 5$, prosječno odstupanje pri rekonstrukciji za sve lokalitete ne prelazi vrijednost od 1.2 Celzijeva stupnja. Većinom, ta vrijednost je ispod jednog Celzijevog stupnja te zbog toga možemo zaključiti da je zadani rang matrice vjerodostojan i dobra opcija za analizu. Među singularnim vrijednostima ističe se prvi koncept čija relativna vrijednost je 75%. Zbog tolike vrijednosti, koja je znatno veća nego sve ostale zajedno, jednostavno zaključujemo da upravo taj koncept ima najveći utjecaj na temperaturu. Izgled ovog koncepta je intuitivan i možemo uočiti da prati trend godišnjeg doba. Kako su vrijednosti desnog singularnog vektora negativne, graf lijevog singularnog vektora ima minimume u ljetnim mjesecima, a maksimume u zimskim. Vrijednosti lijevog i desnog singularnog vektora mogu se vidjeti na Slici 6.5, odnosno na Slici 6.6.

Poglavlje 6. Rezultati analize



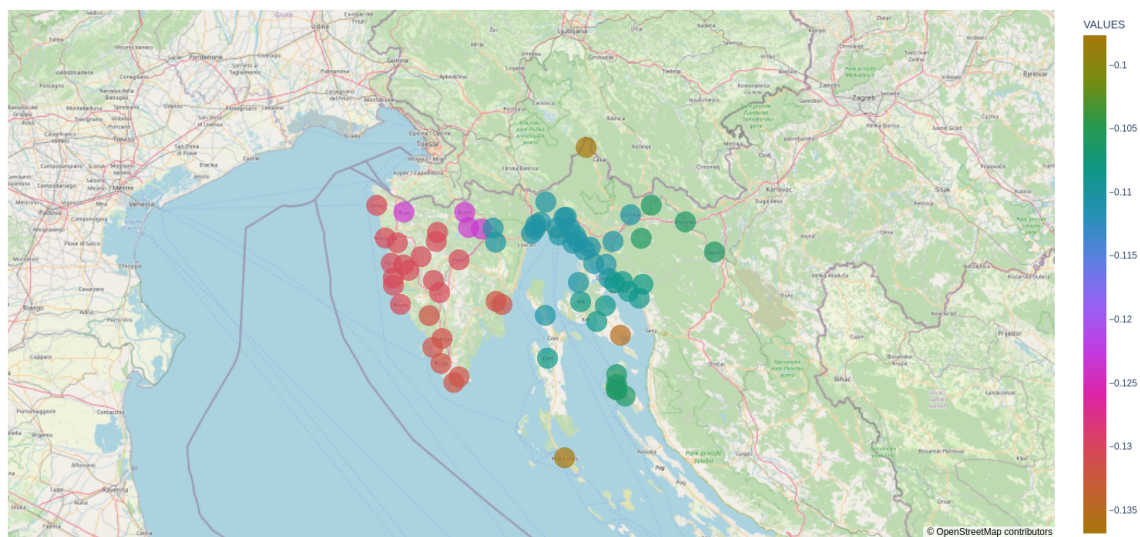
Slika 6.5 Vrijednosti lijevog singularnog vektora s fokusom na koncept $k = 0$ za temperaturu zraka na čitavom skupu podataka



Slika 6.6 Vrijednosti desnog singularnog vektora s fokusom na koncept $k = 0$ za temperaturu zraka na čitavom skupu podataka

Kako se vrijednosti za ovaj koncept nalaze unutar uskog raspona, moglo bi se pogrešno zaključiti da i svi lokaliteti imaju slične temperature, no prikazom vrijednosti na geografskoj karti dobivamo jasniju sliku. Takav prikaz možemo vidjeti na Slici 6.7.

Poglavlje 6. Rezultati analize



Slika 6.7 Prikaz vrijednosti desnog singularnog vektora koncept $k = 0$ na geografskoj karti za temperaturu zraka na čitavom skupu podataka

Rezultati prikazuje postojanje triju mikroklima kod promatranja temperature zraka na čitavom skupu podataka. Uz već istaknutu Bašku i Mali Lošinj, SVD analiza istaknula je i Prezid. Za sva tri lokaliteta sa sigurnošću možemo reći da njihova povezanost, to jest "boja" na karti ne odgovara bližim lokalitetima. Od preostala četiri koncepta, Baška i Mali Lošinj uvijek su istaknutiji od geografski bliskih lokaliteta, dok je Prezid istaknut i u konceptu $k = 2$.

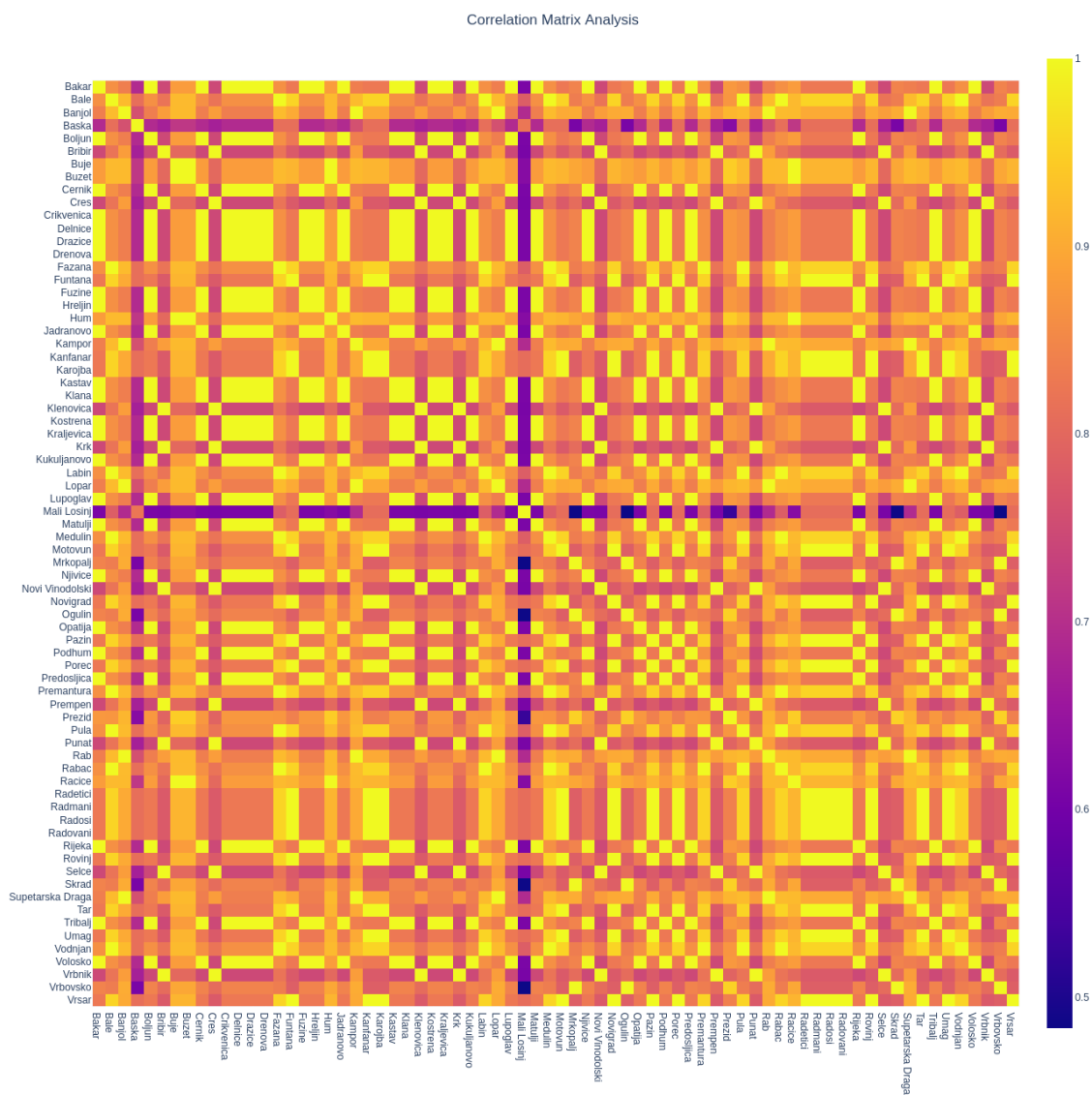
6.1.2 Skraćeni skup podataka - ljetni mjeseci

Analiza na skraćenom skupu podataka ima nekoliko razlika u odnosu na cijeli skup podataka. Prva i glavna karakteristika je da je broj vrijednosti znatno manji. U našem slučaju, kako uzimamo četvrtinu dostupnih mjeseci, pretpostavka je da je i vrijednosti oko četiri puta manje. Zbog velike količine podataka, čak i četvrtina podataka rezultira s velikim brojem mjerenja te je analizu moguće provesti i na takvom skupu podataka te dobiti određene zaključke.

Za razliku od cijelog skupa podataka, u korelacijskoj matrici rezultirajuće vrijednosti su veće ili jednake 0.48, a manje ili jednake 1.0. S obzirom da znamo da je

Poglavlje 6. Rezultati analize

na cijelom skupu podataka donja granica bila približno 0.86, možemo uočiti da je u ljetnim mjesecima najniža korelacija gotovo dvostruko manja nego na cijelom skupu podataka. Prikaz korelacijske matrice nalazi se na Slici 6.8.



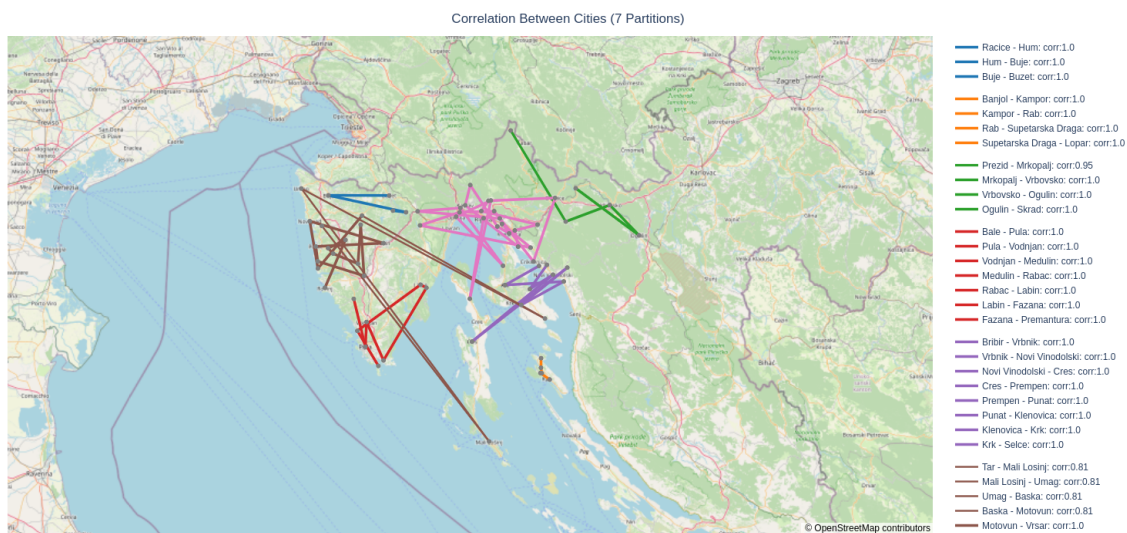
Slika 6.8 Korelacijska matrica za temperaturu zraka u ljetnim mjesecima

Uočavamo da su korelacijski faktori to jest "boje" slične čitavom skupu podataka, no kako je donja granica značajno pomaknuta isti izgled matrice ne znači i

Poglavlje 6. Rezultati analize

istu povezanost. Ponovno možemo uočiti značajno isticanje Baške i Malog Lošinja. Gledajući njihove vrijednosti u odgovarajućim recima, ističu se raznolikim bojama uz ipak znatno više tamnijih boja koje predstavljaju niže vrijednosti korelacije.

Samo particioniranje grafova na osnovu korelacije rezultiralo je istim prikazom koji smo vidjeli kod čitavog skupa podataka na Slici 6.3. Ipak, ovoga puta algoritam uspješno izdvaja i sedmu particiju koju nije bilo moguće pronaći na cijelom skupu podataka te se izdvaja Sjeverna Istra kao zasebna particija. Prikaz svih sedam particija nalazi se na Slici 6.9. Kao i u prijašnjoj analizi temperature zraka na cijelom skupu podataka, Baška i Mali Lošinj nalaze se u particiji sa Zapadnom Istrom te se već u korelacijskoj analizi ističu kao dobri kandidati za lokalitete s mikroklimom.

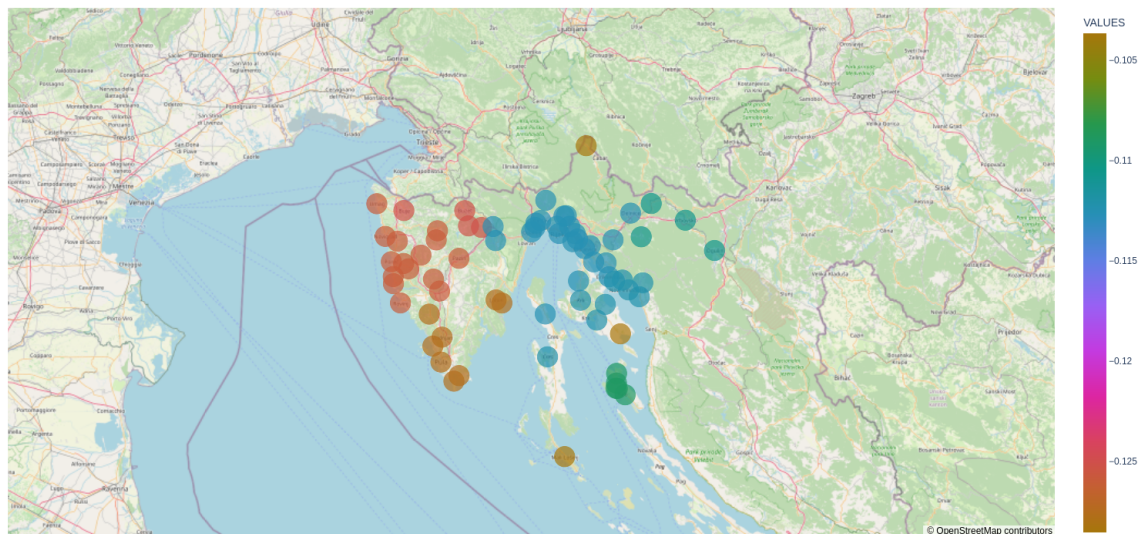


Slika 6.9 Prikaz sedam particija za temperaturu zraka u ljetnim mjesecima

U SVD analizi, rezultati su slični onima u analizi na cijelom skupu podataka. Baška i Mali Lošinj nastavljaju biti istaknuti lokaliteti te im se ponovno pridružuje Prezid kao mikroklima. Ipak, u ljetnim mjesecima, vrijednosti za tri lokaliteta koja imaju mikroklimu su sličnija ostalim gradovima nego što je to slučaj kod cijelog skupa podataka. Tako, primjerice, kod prvog koncepta, čija je singularna vrijednost gotovo dvostruko veća od vrijednosti za drugi koncept, vrijednosti Baške, Malog Lošinja i Prezida su slične onima Južne Istre. Prikaz desnog singularnog vektora za prvi

Poglavlje 6. Rezultati analize

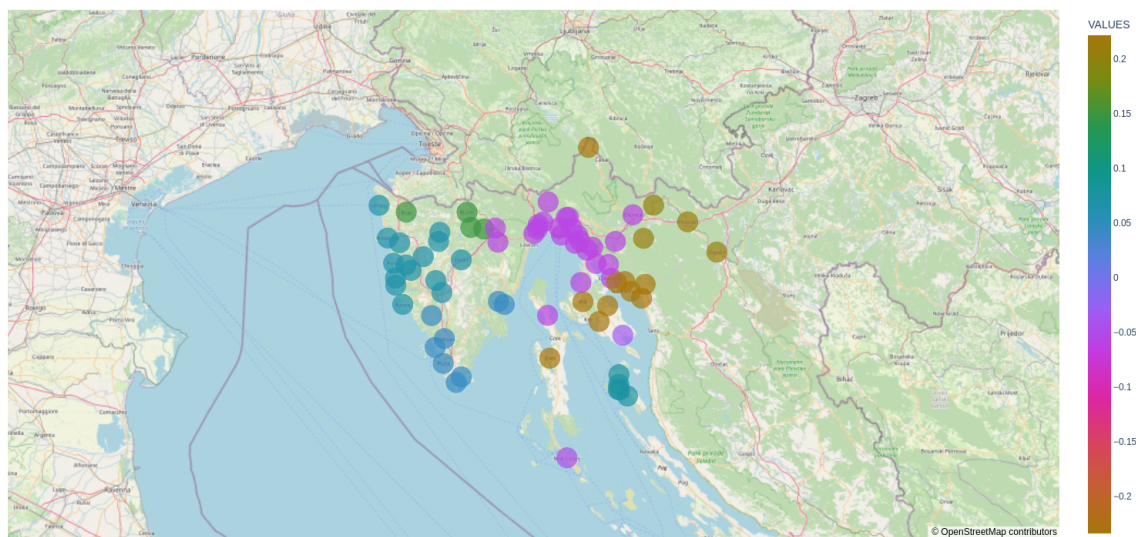
koncepta na geografskoj karti nalazi se na Slici 6.10.



Slika 6.10 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za temperaturu zraka u ljetnim mjesecima

Zanimljiva činjenica je da postoje i dva koncepta u kojima Baška i Mali Lošinj nisu istaknuti kao mikroklima, a čak za četvrti koncept vrijedi da ni Prezid nije mikroklima. Baška i Mali Lošinj imaju vrijednosti slične Kvarneru, dok vrijednost Prezida odgovara Gorskoj Hrvatskoj. Nažalost, iz lijevog singularnog vektora nije moguće razaznati što četvrti koncept $k = 3$ predstavlja. Ipak, jasno pokazuje da postoje vremenski uvjeti i situacije u kojima za ljetne mjesece nije moguće istaknuti niti jedan lokalitet kao mikroklimu. Desni singularni vektor za taj koncept prikazan je na Slici 6.11

Poglavlje 6. Rezultati analize

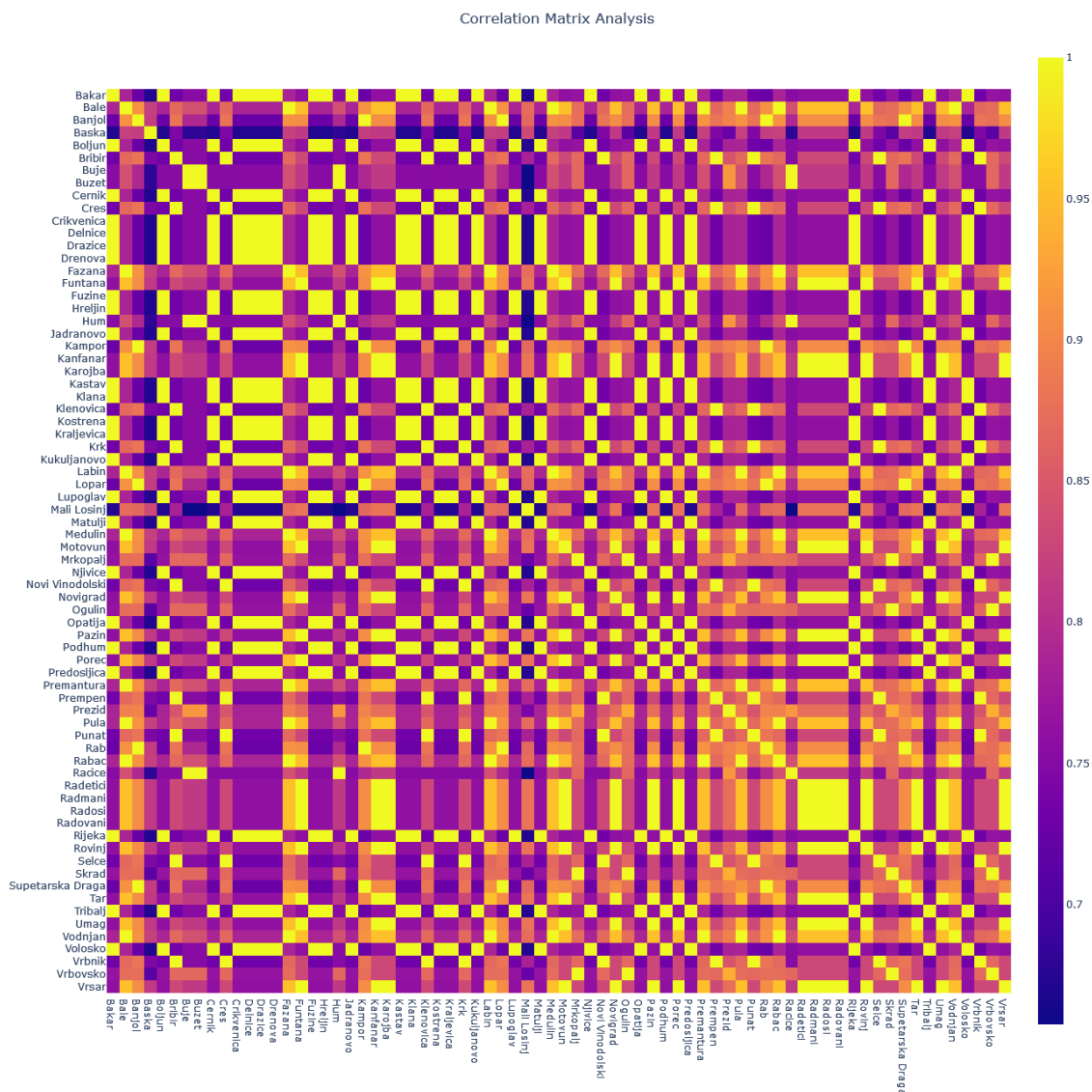


Slika 6.11 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 3$ na geografskoj karti za temperaturu zraka u ljetnim mjesecima

6.1.3 Skraćeni skup podataka - zimski mjeseci

Kako je područje koje promatramo Primorsko-goranska i Istarska županija, znamo da Jadransko more ima jak utjecaj na meteorološke prilike područja. Isto tako, gorski dio županije i Sjeverna Istra su područja koja su zimi zasigurno hladnija od obalnih područja te samim time postoji mogućnost da će razlike u vrijednosti korelacije među određenim lokalitetima biti veće. Upravo nam korelacijska matrica na slici 6.12 potvrđuje ovu hipotezu.

Poglavlje 6. Rezultati analize



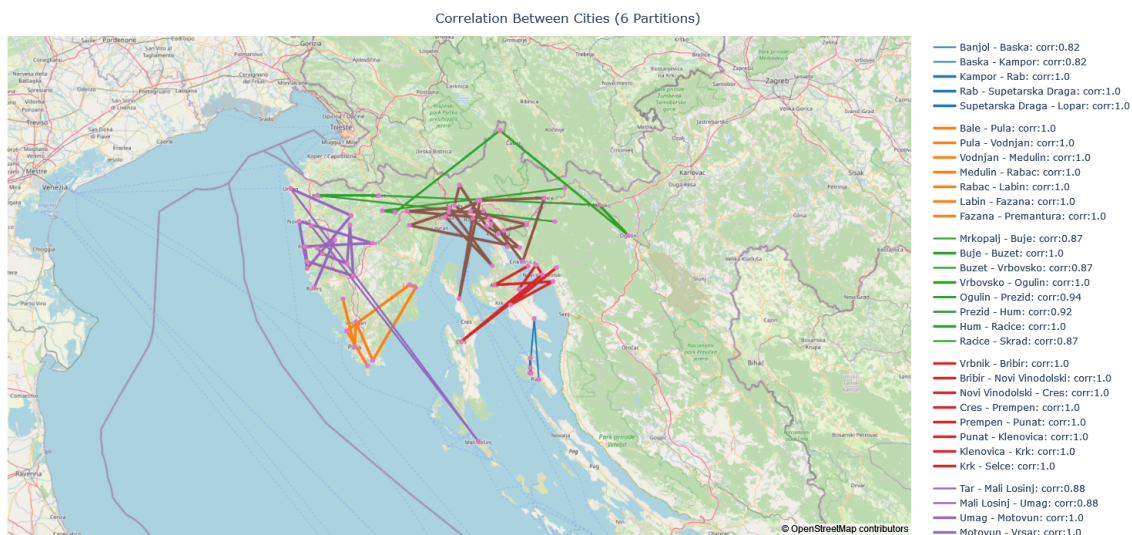
Slika 6.12 Korelacijska matrica za temperaturu zraka u zimskim mjesecima

Iako donja granica nije toliko niska kao u ljetnim mjesecima, u ovoj matrici uočavamo dominaciju tamnijih boja. Samim time, zaključujemo da je povezanost među određenim promatranim lokalitetima osjetno slabija nego ljeti, a pogotovo na skupu podataka gdje uzimamo svih dvanaest mjeseci. Iz same matrice nemamo lokalitet koji se jasno ističe od ostalih u mjeri kao što su to bili Baška i Mali Lošinj u prethodnim slučajevima. Svaki od promatranih lokaliteta s određenim mjestima ima snažnu

Poglavlje 6. Rezultati analize

korelaciju, dok s određenim mjestima ima slabiju.

Algoritam za particioniranje grafa pronašao je najviše šest particija. Po prvi puta, Baška se nalazi u particiji s geografski bliskim mjestima otoka Raba te, uzimajući Sliku 6.13 u obzir, nije kandidat za mikroklimu. S druge strane, Mali Lošinj je, kao i u prethodim promatranjima temperature zraka, povezan sa Zapadnom Istrom te je zbog toga ponovno dobar kandidat za mikroklimu.



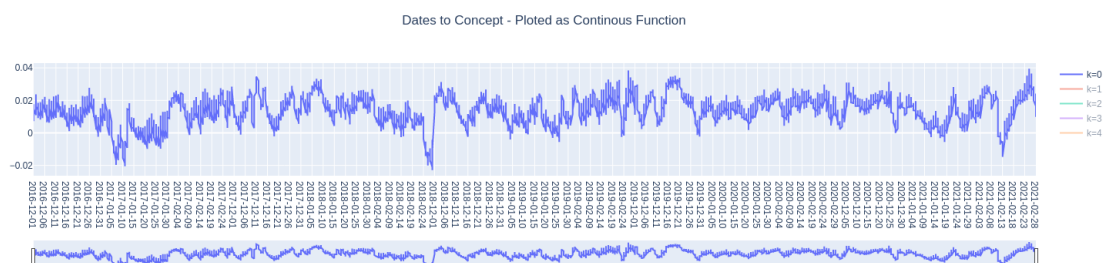
Slika 6.13 Prikaz šest particija za temperaturu zraka u zimskim mjesecima

Suprotno rezultatima korelacije i particioniranja grafa, SVD ponovno uspješno izdvaja iste lokalitete kao one s mikroklimom. Baška i Mali Lošinj su u svim promatranim konceptima istaknuti, dok se Prezid ističe u prva tri koncepta čija je i vrijednosti singularnog vektora najveća. U posljednja dva promatrana koncepta, Prezid se ne ističe od geografski bliskih mjesta. Zanimljivo je da u konceptu $k = 0$ lijevi singularni vektor koji opisuje odnos datuma i koncepta prati minimalne vrijednosti izmjerenih temperatura. Taj koncept nalazi se na Slici 6.14. Uzmemo li u obzir vrijednosti u Tablici 6.1, jasno je da minimumi grafa prate trend s najnižim izmjerenim temperaturama zraka.

Poglavlje 6. Rezultati analize

Tablica 6.1 Najniže izmjerene temperature zraka u zimskim mjesecima

Godina	Mjesec	Temperatura [$^{\circ}\text{C}$]
2016	12	-8
2017	1	-17
2017	2	-8
2017	12	-8
2018	1	-4
2018	2	-16
2018	12	-7
2019	1	-8
2019	2	-7
2019	12	-8
2020	1	-6
2020	2	-5
2020	12	-10
2021	1	-10
2021	2	-12



Slika 6.14 Vrijednosti lijevog singularnog vektora koncepta $k = 0$ za temperaturu zraka u zimskim mjesecima

Nakon analize rezultata temperature zraka na cijelom skupu podataka, u ljetnim i u zimskim mjesecima možemo zaključiti da se izdvajaju tri lokaliteta kao mikroklima.

Vrijednosti za Bašku, Mali Lošinj i Prezid se u većini koncepata ističu, a osobito u onima s većom singularnom vrijednošću. Baška i Mali Lošinj, kada se vrši particioniranje grafova, se nađu u particiji sa Zapadnom Istrom, dok Prezid svoju mikroklimu jasnije istakne tek u SVD analizi.

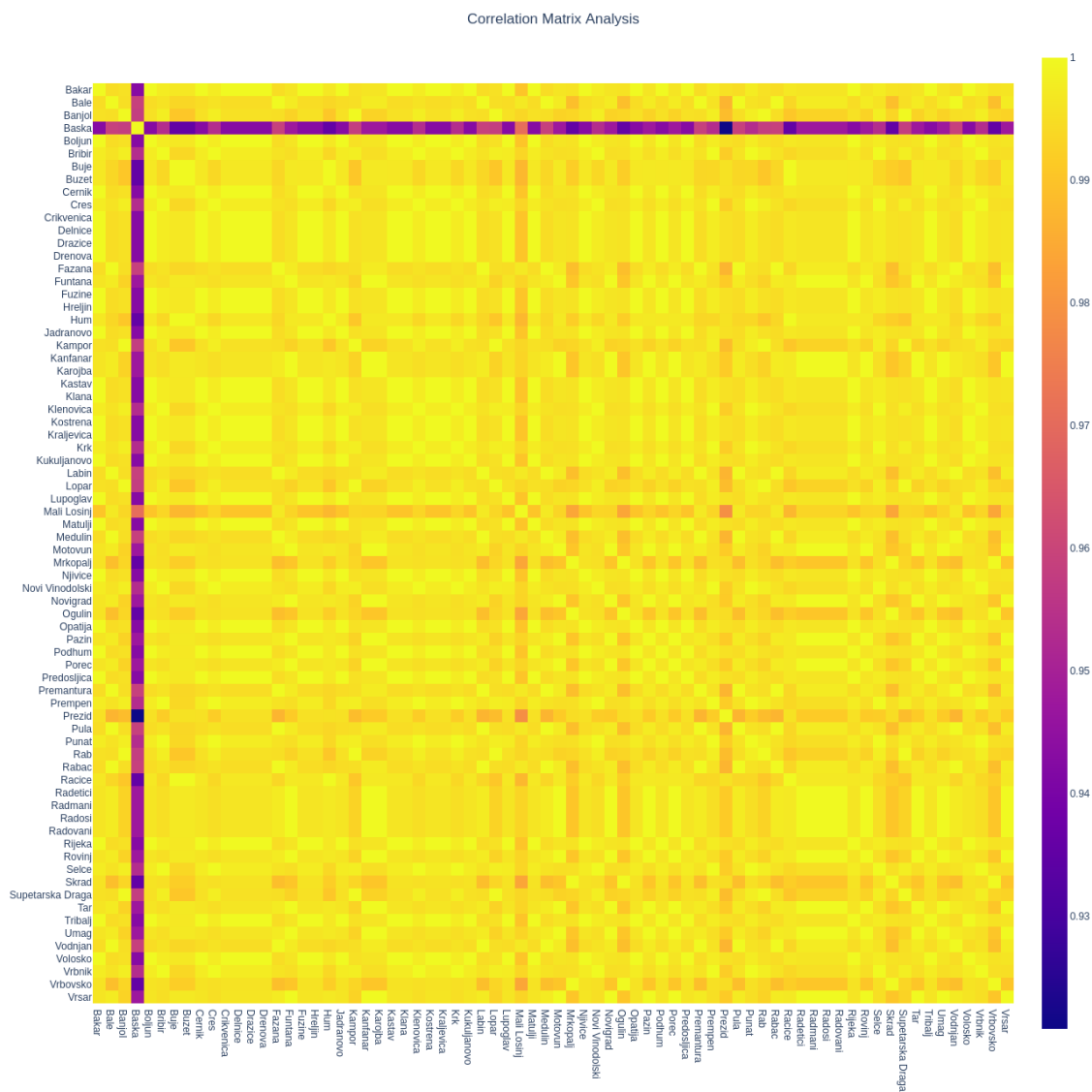
6.2 Tlak zraka

Tlak zraka ili atmosferski tlak je težina kojom zrak pritišće na površinu zemlje [23]. Osim u hektopaskalima, mjeri se i u barima. Vrijednost od jednog bara jednaka je vrijednosti od 1000 hektopaskala. U meteorologiji, češće se koriste vrijednosti iskazane u hektopaskalima, pa će tako i promatrane vrijednosti u ovoj analizi biti iskazane u toj mjernoj jedinici.

6.2.1 Cijeli skup podataka

Analiza tlaka zraka na čitavom skupu podataka zasigurno donosi jedne od najzanimljivijih rezultata analize. Prema rezultatima korelacijske matrice sa Slike 6.15, sva mjesta imaju izrazitu povezanost. Međutim, Baška se ističe kao lokalitet koji u svakom slučaju ima tamnije boje i najslabije korelira s ostalim mjestima. Treba imati na umu da su rezultati svi veći od 0.92 što ukazuje na to da i tamne boje u ovom slučaju označavaju jaku korelaciju. Sljedeći korak korelacijske analize trebao je biti odvajanje particija prema grafu kojem su bridovi vrijednosti jačine korelacije. Zbog izrazite povezanosti, algoritam nije pronašao niti minimalan broj particija što ponovno pokazuje izrazitu sličnost tlaka zraka na promatranom području.

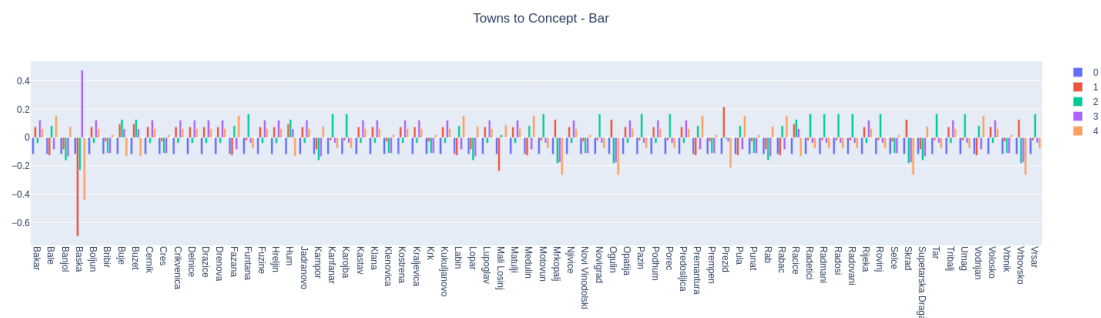
Poglavlje 6. Rezultati analize



Slika 6.15 Korelacijska matrica za tlak zraka na čitavom skupu podataka

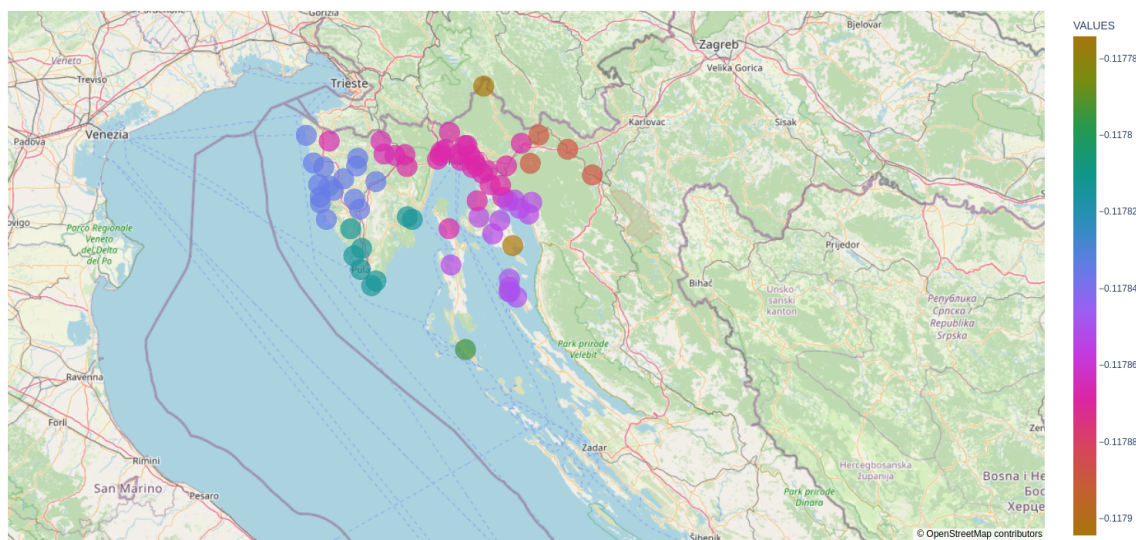
SVD analiza, točnije desni singularni vektor sa Slike 6.16 koji predstavlja odnos mjesta i koncepta, potvrđuje korelacijsku analizu i ističe Bašku kao lokalitet koji se izdvaja od ostalih mjesta. Međutim, ukoliko detaljnije pogledamo koncept $k = 0$, možemo uočiti da su vrijednosti gotovo identične. Vrijednost tog koncepta (1043165) je znatno veća nego vrijednosti svih ostalih zajedno ukazuje na njegovu važnost te uzak raspon vrijednosti naglašava povezanost među mjestima.

Poglavlje 6. Rezultati analize



Slika 6.16 Vrijednosti desnog singularnog vektora za tlak zraka na čitavom skupu podataka

Iako su razlike teže uočive na korelacijskoj matrici, u svakom konceptu ponavljaju se istaknuti lokaliteti s mikroklimom kao i kod temperature. Baška, Mali Lošinj i Prezid su lokaliteti koji se tlakom zraka izdvajaju od geografskog područja na kojem se nalaze. Tako koncept $k = 0$ prikazan na Slici 6.17 uspješno prikazuje minimalne razlike te ističe navedene lokalitete. Ostali koncepti također ističu iste lokalitete.

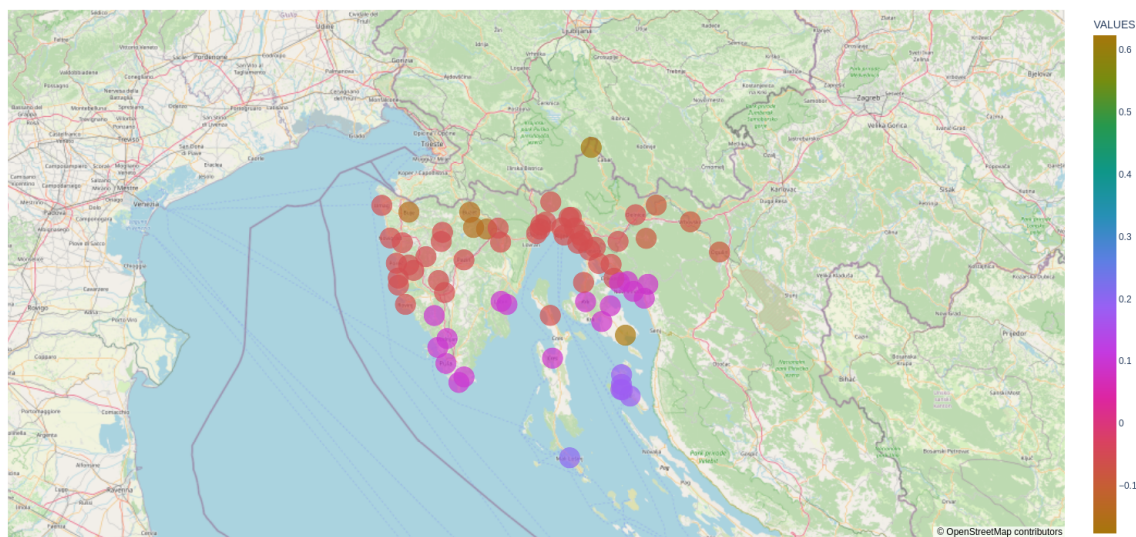


Slika 6.17 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za tlak zraka na čitavom skupu podataka

6.2.2 Skraćeni skup podataka

Za razliku od temperature zraka, gdje su rezultati u ljetnim i zimskim mjesecima izraženiji, u slučaju tlaka zraka dobiveni rezultati su vrlo slični onima na cijelom skupu podataka. Korelacijska analiza u ljetnim i zimskim mjesecima ponovno znatno ističe Bašku kao mjesto koje lošije korelira sa ostalim promatranim mjestima. U ljetnim mjesecima, donja granica se spušta s 0.94 na 0.86, no grafički prikaz korelacijske matrice gotovo da je nepromijenjen. Također, ponovno imamo slučaj u kojem se ne uspijeva odvojiti graf u particije. Kao i kod cijelog skupa podataka, lokaliteti snažno koreliraju i minimalan broj particija je ponovno viši od onog kojeg algoritam uspijeva podijeliti. S druge strane, SVD analiza pronalazi koncepte koji ne ističu neke od tri istaknuta lokaliteta. U ljetnim mjesecima, za koncept $k = 2$ uočavamo da Prezid ima povezanost sa Sjevernom Istrom, a Mali Lošinj s Rabom. Takav koncept na geografskoj karti nalazi se na Slici 6.18. Koncept sličnog izgleda pronalazimo i u zimskim mjesecima, samo je ondje pod konceptom $k = 3$. Važno je naglasiti da oba koncepta nemaju veliku singularnu vrijednost, no dokaz su da postoje meteorološki uvjeti koji izjednačavaju tlak mjesta koja imaju mikroklimu s bliskim mjestima. Nažalost, iz lijevog singularnog vektora nije moguće razaznati što predstavljaju ti koncepti.

Poglavlje 6. Rezultati analize



Slika 6.18 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za tlak zraka u ljetnim mjesecima

6.3 Vlažnost zraka

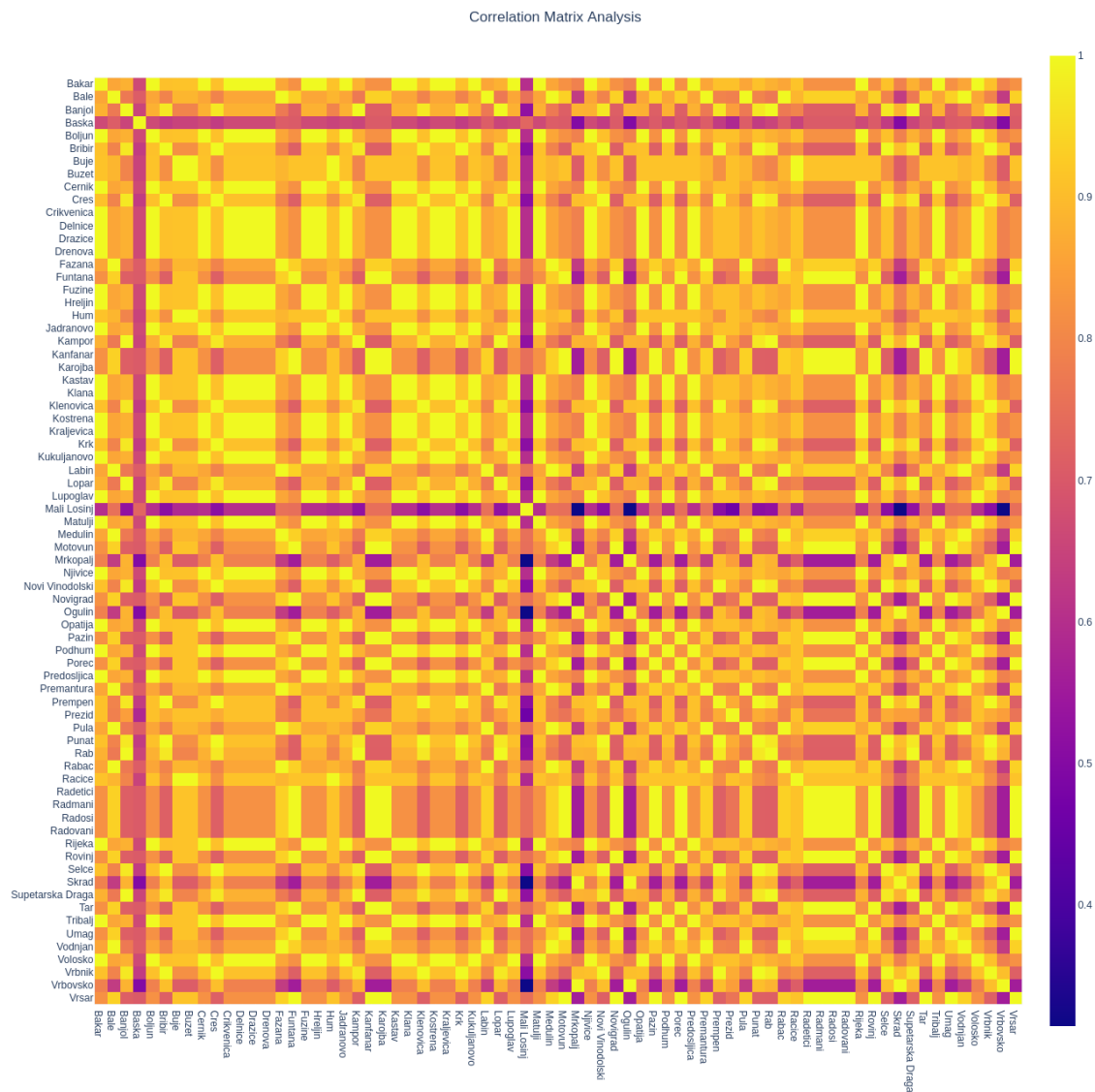
Vlažnost zraka je količina vodene pare koja je sadržana u zraku ili u atmosferi. Iako postoji više načina izražavanja, u ovoj analizi govorimo o relativnoj vlažnosti zraka. Izražava se omjerom parcijalnog tlaka prisutne vode u zraku i parcijalnog tlaka zasićene vode pri određenoj temperaturi i tlaku zraka. Kako govorimo o omjeru, rezultat se izražava u postocima. Relativna vlažnost iznosi 0% kada je zrak potpuno suh, a 100% kada je zrak potpuno zasićen vodenom parom [24].

6.3.1 Cijeli skup podataka

Korelacijska analiza vlažnosti zraka na čitavom skupu podataka donosi nam zanimljive rezultate te ponovno naglašava određene mikrokline. Već korelacijska matrica sa Slike 6.19 ističe Bašku i Mali Lošinj kao mjesta koja su kandidati za one s mikroklimom. Također, primjećujemo da su dobivene vrijednosti korelacije veće od 0.3, što je zasigurno najniža donja granica u analizama na svim dostupnim poda-

Poglavlje 6. Rezultati analize

cima. Prema tim rezultatima, možemo uočiti da bez obzira na geografsku sličnost lokaliteta imamo parove koji imaju srednje jaku korelaciju, pa čak i slabu.

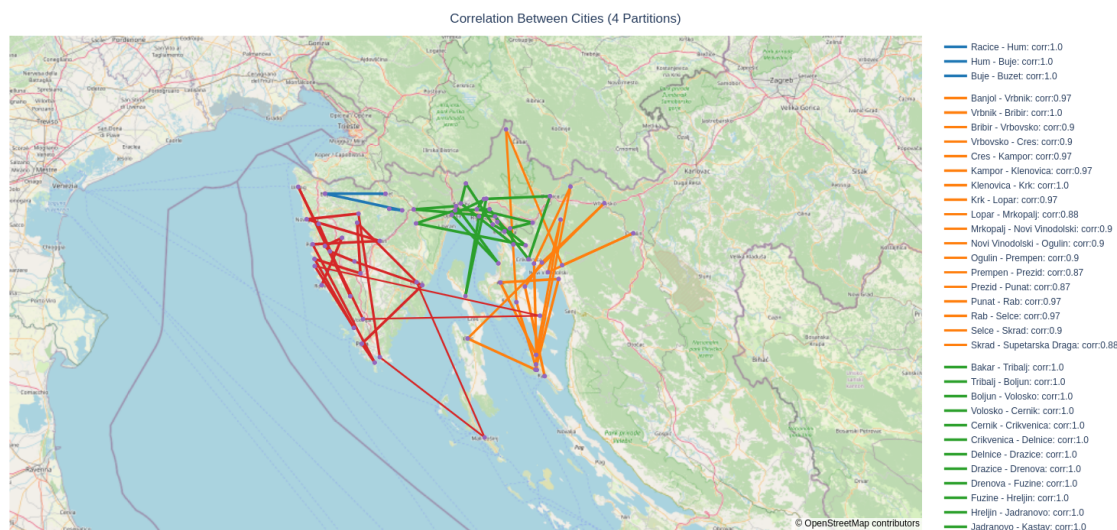


Slika 6.19 Korelacijska matrica za vlažnost zraka na čitavom skupu podataka

Za razliku od tlaka zraka, uspješno dobivamo minimalan broj particija grafa prikazan na Slici 6.20. Kao i u prijašnjim analizama, Baška i Mali Lošinj povezani su s Istrom. Ta povezanost čini ih kandidatima za mikroklimu. U analizi na cijelom

Poglavlje 6. Rezultati analize

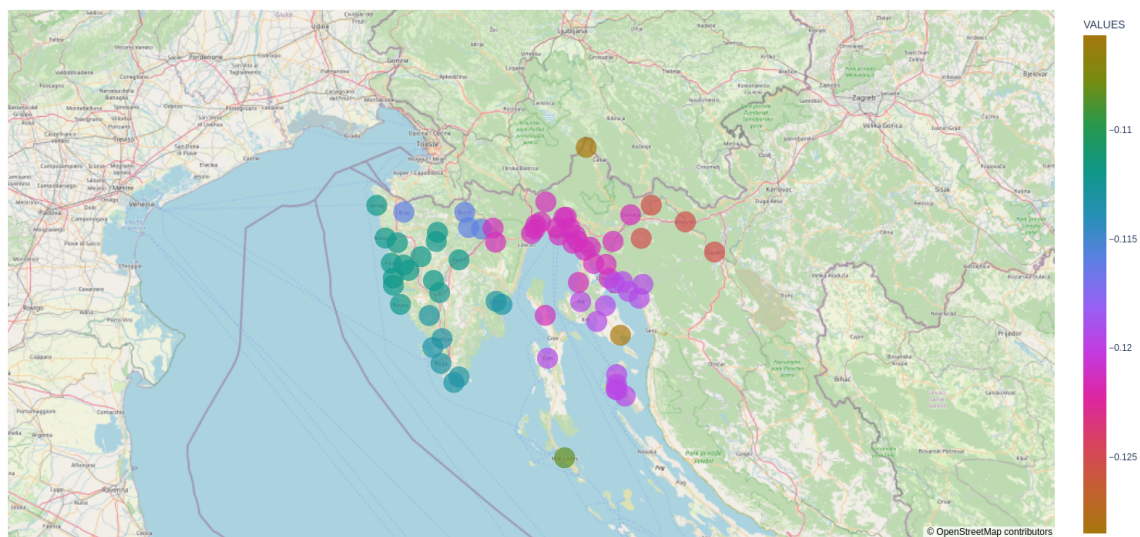
skupu podataka nije bilo moguće izdvojiti više od četiri particije, no analiza tlaka zraka na skraćenom skupu podataka će biti uspješnija u tome.



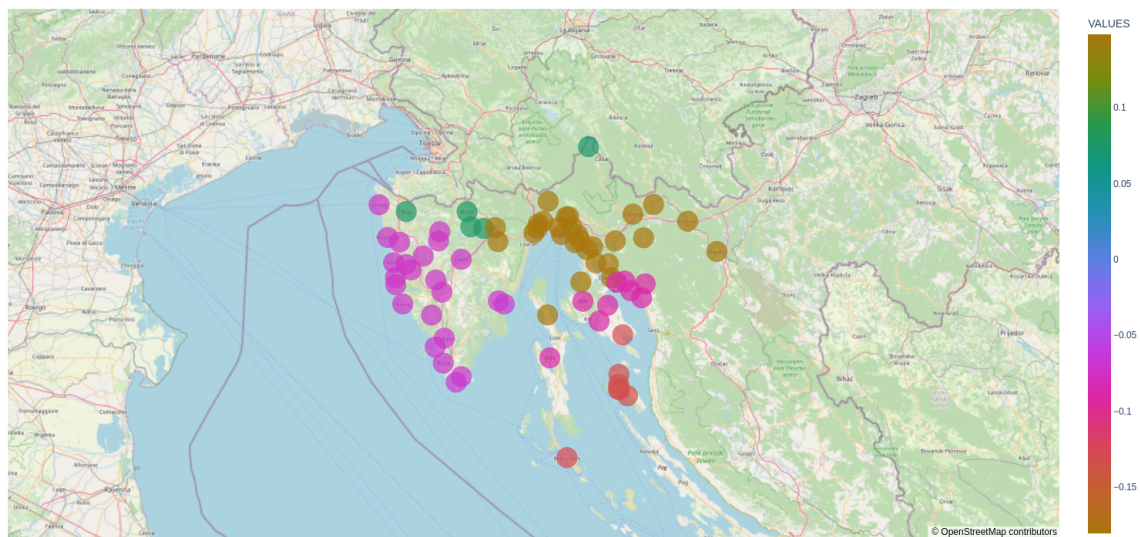
Slika 6.20 Prikaz četiri particije za vlažnost zraka na čitavom čitavom skupu podataka

SVD analiza potvrđuje očekivanja. Kao i kod temperature zraka, uz očekivanu Bašku i Mali Lošinj kao mikroklimu ističe i Prezid. Odnos datuma i koncepta ne donosi nam jasne grafove te nije moguće bez detaljne analize shvatiti njihovo značenje. Iako koncept $k = 0$, koji se nalazi na Slici 6.21, ima najznačajniju vrijednost singularnog vektora, zanimljiv je i koncept $k = 2$ prikazan na Slici 6.22. Njegova vrijednost od 2387 jasno nam govori da je važan i da postoje meteorološke prilike u kojima ne postoji mikroklima. Na njemu uočavamo da su Mali Lošinj i Baška povezani s Rabom, ali i međusobno. Također, Prezid je povezan sa Sjevernom Istrom. Iako bi geografski više odgovarala Gorska Hrvatska, geografska obilježja Sjeverne Istre su zasigurno sličnija obilježjima Prezida, za razliku od Baške s kojom je povezan u konceptu $k = 0$.

Poglavlje 6. Rezultati analize



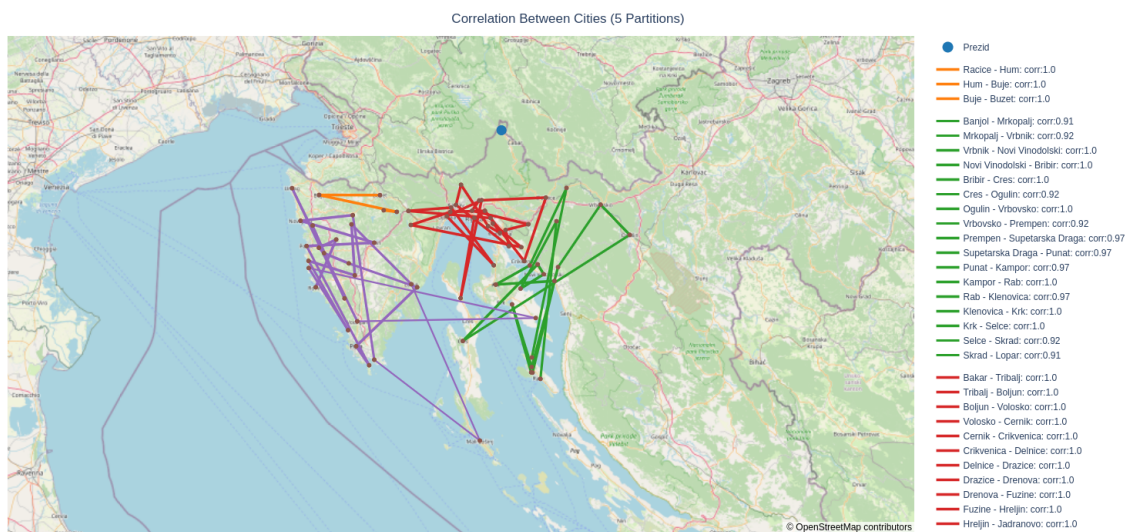
Slika 6.21 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za vlažnost zraka na čitavom čitavom skupu podataka



Slika 6.22 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za vlažnost zraka na čitavom čitavom skupu podataka

6.3.2 Skraćeni skup podataka - ljetni mjeseci

Sama korelacijska matrica u ljetnim mjesecima je izrazito slična onoj na cijelom skupu podataka. Baška i Mali Lošinj su istaknuti, a ostala mjesta većinom snažno koreliraju. Uzevši samo nju u obzir, pretpostavka bi bila da su rezultati i zaključci korelacijske analize isti kao i za vlažnost zraka na čitavom skupu podataka. Međutim, particioniranje grafova uspjeva pronaći particiju više. Iako prikaz četiriju particija na geografskoj karti izgleda identično, po prvi put imamo lokalitet koji se nalazi u zasebnoj particiji. Na Slici 6.23 možemo uočiti da je peta particija upravo Prezid - lokalitet za koji znamo da gotovo uvijek ima mikroklimu. Uz jasno istaknuti Prezid, Baška i Mali Lošinj svojom povezanošću ponovno se nameću kao kandidati koji bi se u SVD analizi mogli istaknuti.

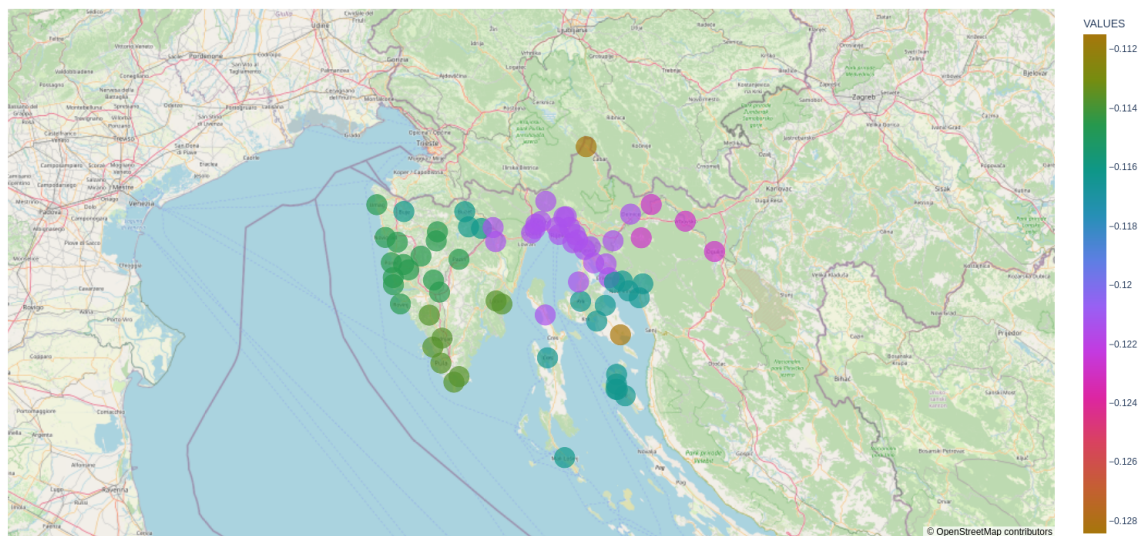


Slika 6.23 Prikaz četiri particije za vlažnost zraka u ljetnim mjesecima

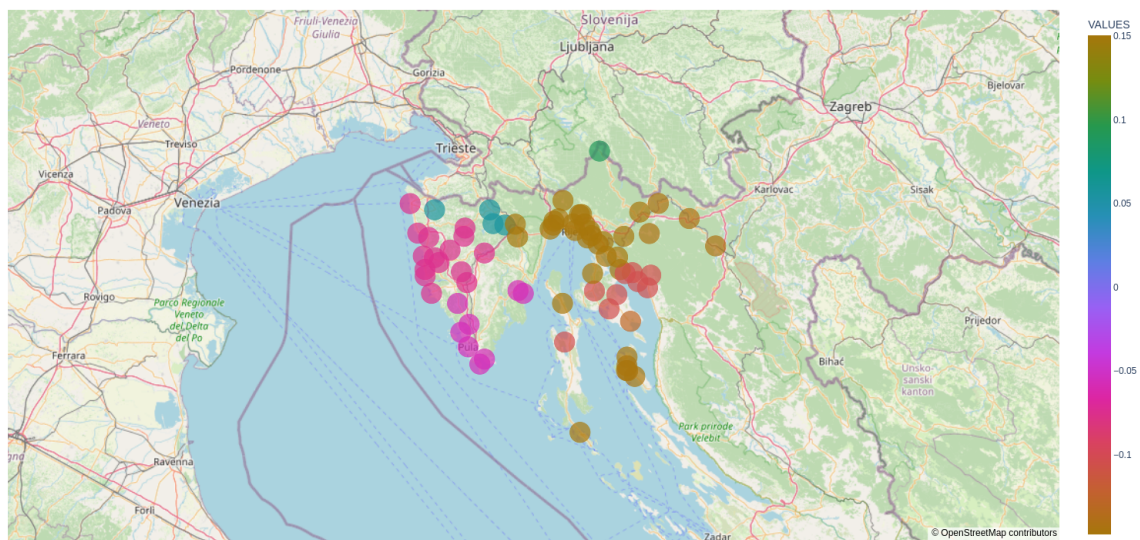
Kod SVD analize, najvažniji koncept $k = 0$ povezuje Mali Lošinj s geografski bliskim mjestima, što ne odgovara rezultatima korelacijske matrice i pokazuje da u najbitnijem konceptu za vlažnost zraka u ljetnim mjesecima nema mikroklimu. U tom konceptu, kojeg možemo vidjeti na Slici 6.24, Baška i Prezid su istaknuti. Na Slici 6.25 možemo vidjeti koncept $k = 2$ čija singularna vrijednosti iznosi 1215, prikazuje meteorološka obilježja u kojima je Mali Lošinj povezan s obližnjim mjestima,

Poglavlje 6. Rezultati analize

a vrijednosti Baške i Prezida su veoma blizu vrijednostima područja Krka, odnosno Sjeverne Istre. U preostala tri promatrana koncepta, sva tri lokaliteta su izražene mikroklimе.



Slika 6.24 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za vlažnost zraka u ljetnim mjesecima

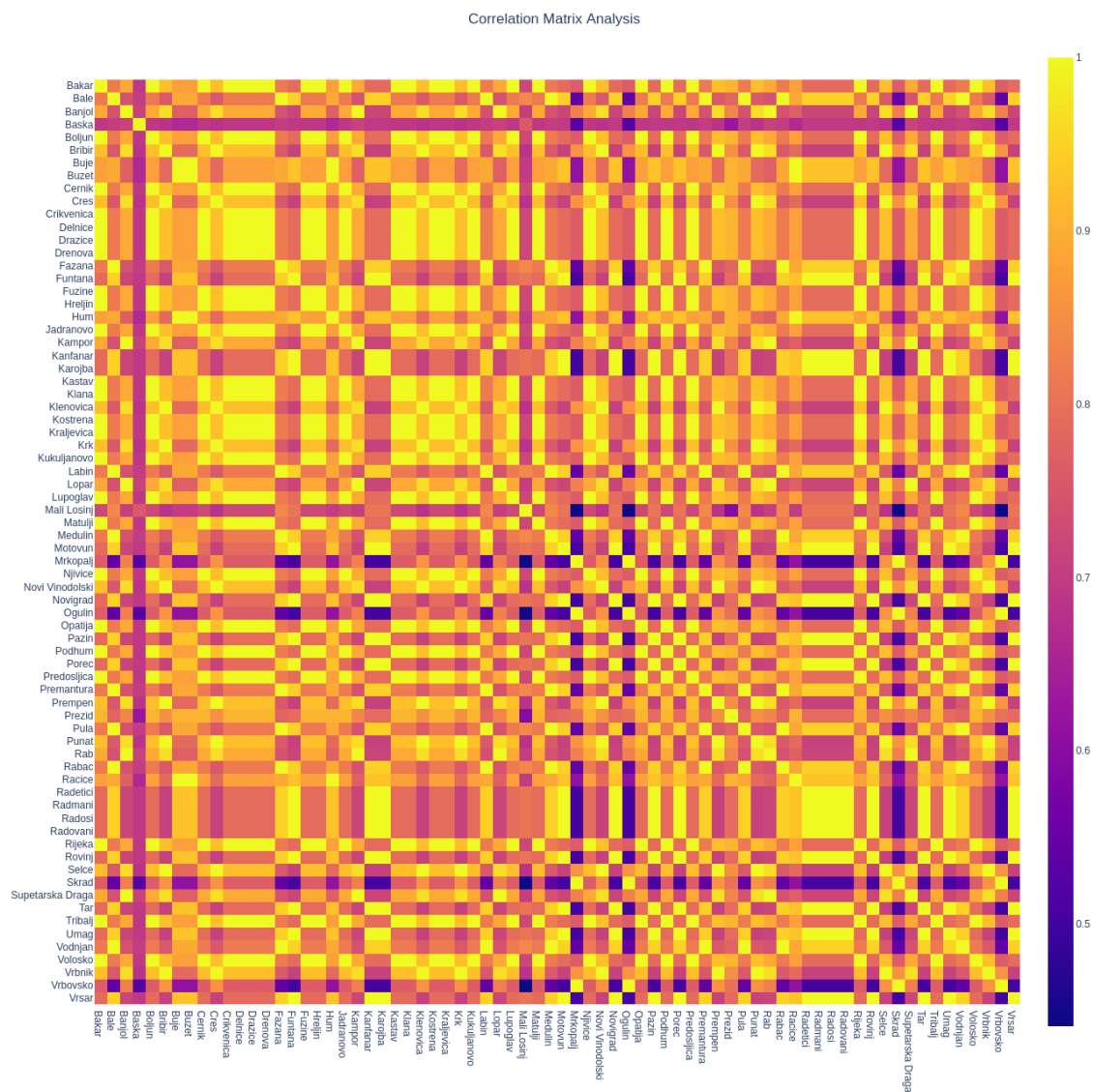


Slika 6.25 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za vlažnost zraka u ljetnim mjesecima

6.3.3 Skraćeni skup podataka - zimski mjeseci

Za razliku od cijelog skupa podataka i ljetnih mjeseci, kod zimskih mjeseci korelacijska matrica za Bašku i Mali Lošinj nema toliko "tamnih" polja. Drugim riječima, povezanost ta dva mjesta s ostalim lokalitetima nije toliko izražena jer vrijednosti nisu na samoj minimalnoj vrijednosti u korelacijskoj matrici. Međutim, možemo primijetiti da isto tako ne postoje mjesta s kojima navedeni lokaliteti imaju izrazito jaku korelaciju. Uzevši u obzir korelacijsku matricu na Slici 6.26, ponovno moramo uzeti u obzir da će i Baška i Mali Lošinj biti dobri kandidati za mikroklimu.

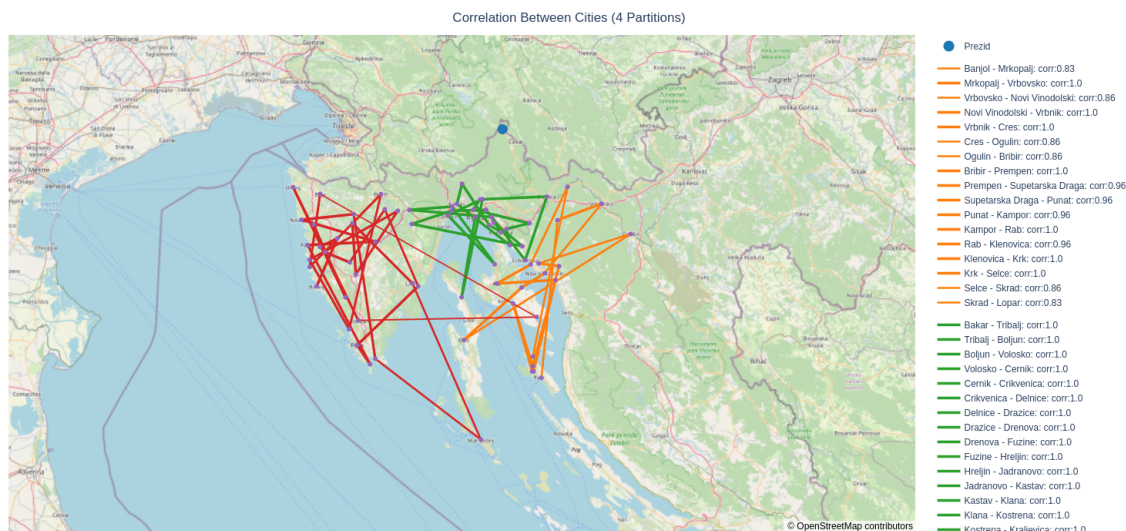
Poglavlje 6. Rezultati analize



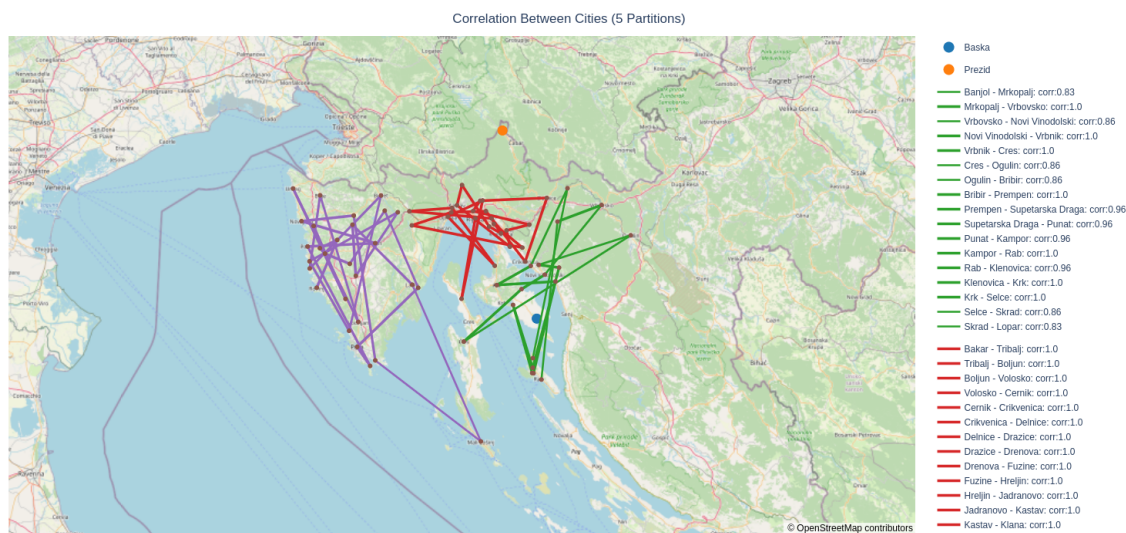
Slika 6.26 Korelacijska matrica za vlažnost zraka u zimskim mjesecima

Za razliku od cijelog skupa podataka gdje je graf uspješno podijeljen u samo četiri particije, u zimskim mjesecima ponovno dobivamo i petu. Međutim, već u prikazu četiri particije na Slici 6.27 uočavamo da je Prezid istaknut kao mikroklima, a Sjeverna Istra je povezana s ostatkom Istre. Peta particija sa Slike 6.28 ističe i Bašku i već u particioniranju grafa imamo uspješno izdvojena dva poznata lokaliteta.

Poglavlje 6. Rezultati analize



Slika 6.27 Prikaz četiri particije za vlažnost zraka u zimskim mjesecima

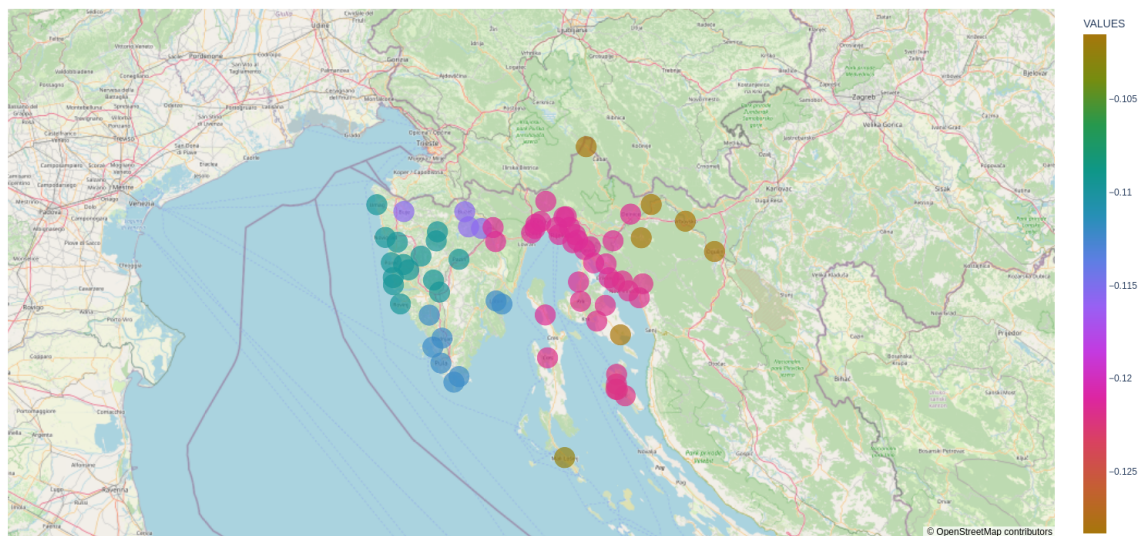


Slika 6.28 Prikaz pet particija za vlažnost zraka u zimskim mjesecima

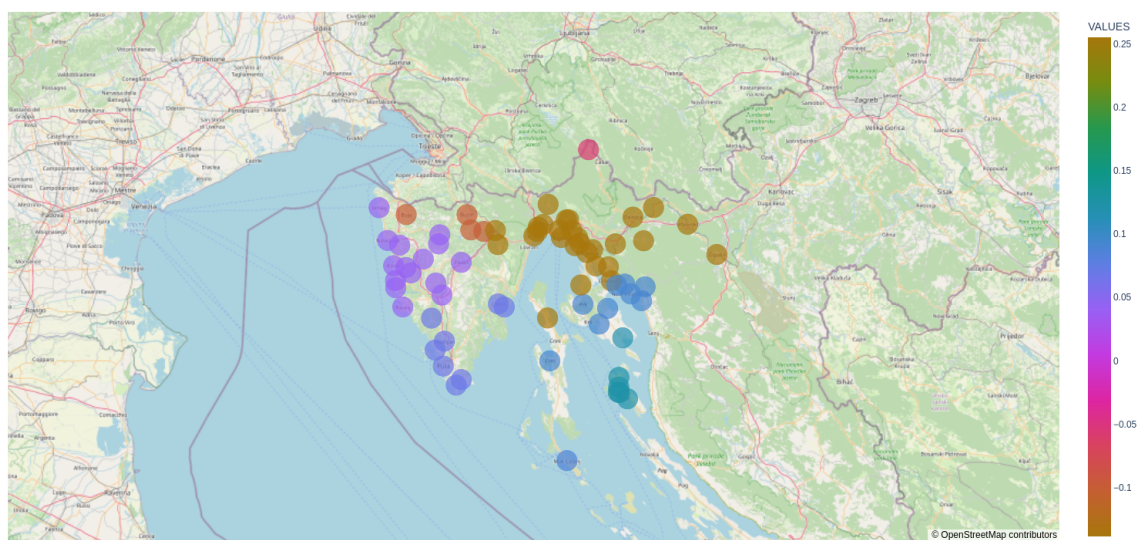
U SVD analizi zanimljivo je da nakon jasno istaknute mikroklimе Prezida, koja se izvđaja već u prikazu četiri particije, glavni koncept $k = 0$ ne uspijeva potvrditi isto. Na Slici 6.29 možemo uočiti da je Prezid povezan s Gorskom Hrvatskom. Baška i Mali Lošinj su ponovno istaknuti te potvrđeni kao mikrokorelacije. Iako ovaj koncept

Poglavlje 6. Rezultati analize

ne ističe Prezid, svi ostali ga jasno ističu kao mikroklimu. Kao i kod cijelog skupa podataka i zimskih mjeseci, koncept $k = 2$ pokazuje se kao zanimljiv. Na Slici 6.30 vidimo da Mali Lošinj i Baška nisu mikroklimi, dok Prezid to je.



Slika 6.29 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 0$ na geografskoj karti za vlažnost zraka u zimskim mjesecima



Slika 6.30 Prikaz vrijednosti desnog singularnog vektora koncepta $k = 2$ na geografskoj karti za vlažnost zraka u zimskim mjesecima

Poglavlje 7

Zaključak

Kroz ovaj rad opisana je analiza meteoroloških podataka na području Primorsko-goranske i Istarske županije u razdoblju od pet godina počevši od 31.10.2016. do 31.10.2021. Glavini cilj analize je pokušaj pronalaska mikroklimе - lokaliteta koji bi se svojim meteorološkim obilježjima isticao od lokaliteta sličnih geografskih obilježja. Osim pronalaska mikroklimе, ovom prostorno-vremenskom analizom želimo saznati kada i gdje uočavamo određene meteorološke prilike. Iste pokušavamo objasniti te znanstveno opisati.

Podaci su preuzeti sa stranice *World Weather Online* gdje je bilo moguće postaviti frekvenciju koja je iznosila tri sata. Od ukupno 72 promatrana mjesta, 44 se nalaze u Primorsko-goranskoj, a 28 u Istarskoj županiji. Iako podaci imaju mnogo različitih meteoroloških obilježja, kroz analizu promatrane su tri vrijednosti: temperatura, tlak i vlažnost zraka. Osim prostorne analize koja uključuje 72 lokaliteta, za svako obilježje izdvojeni su ljetni i zimski mjeseci koji često donose dodatne zaključke u odnosu na cijeli skup podataka.

Kroz analizu korištena su dva dvije numeričko-analitičke tehnike na osnovu kojih se pokušavaju dobiti zaključci: korelacija i SVD. Kod korelacije, korištena je Pearsonova korelacija koja uspoređuje vrijednosti istih pozicija dvaju vektora. Rezultati korelacijske analize su korelacijska matrica koja prikazuje povezanost među lokalitetima te particioniranje grafa u kojem je težina brida jednaka korelaciji između dva povezana mjesta. Korelacijska analiza uspješno je izdvojila određene mikroklimе koje

Poglavlje 7. Zaključak

bi naknadno bile potvrđene i nadopunjene od strane SVDa. Za dobivanje rezultata korišten je rang matrice $k = 5$ čime se aproksimira rekonstrukcija originalne matrice. Kako je preciznost aproksimirane matrice uvijek veća od 0.95 možemo zaključiti kako je uzeti rang vjerodostojan. Rezultati SVD analize bila su tri singularna vektora. Lijevi singularni vektor predstavlja odnos datuma i koncepta, desni singularni vektor predstavlja odnos mjesta i koncepta dok singularne vrijednosti ukazuju na jačinu promatranih koncepata. Kako bismo imali atraktivan prikaz koji dodatno naglašava mikroklimu određenog mjesta, desni sigurlarni vektor prikazivan je na geografskoj karti.

Analiza jasno ističe tri lokaliteta koja se ističu od geografskog područja u kojem se nalaze. Baška, Prezid i Mali Lošinj se u analizama temperature, tlaka i vlažnosti zraka pokazuju kao mjesta čija meteorološka obilježja ne odgovaraju geografskim čineći ih mikroklimama. Ipak, gotovo u svakoj od promatranih analiza, SVD pronalazi koncepte u kojima neki od navedenih lokaliteta nije istaknut. Postoje i koncepti koji dokazuju da za određene meteorološke uvjete nije moguće izdvojiti mikroklimu jer su sva promatrana mjesta povezana sukladno geografskom položaju. Osim pojedinih lokaliteta, analiza je pokazala da je moguće izdvojiti i područja koja imaju karakteristične klimatske uvjete. Područja Zapadne Istre, Južne Istre, Riječkog zaljeva i Istočnog Kvarnera su četiri cjeline koje kontinuirano imaju snažnu korelaciju.

Bibliografija

- [1] Wikipedia, “Meteorology,” <https://en.wikipedia.org/wiki/Meteorology>, 2022, s interneta; pristupljeno 10.06.2022.
- [2] —, “Data,” <https://en.wikipedia.org/wiki/Data>, 2022, s interneta; pristupljeno 18.06.2022.
- [3] DHMZ, “Dhmz - o nama,” https://meteo.hr/o_nama.php, 2022, s interneta; pristupljeno 08.09.2022.
- [4] TechTarget, “Data analytics,” <https://www.techtarget.com/searchdatamanagement/definition/data-analytics>, 2022, s interneta; pristupljeno 04.08.2022.
- [5] F. University, “How to become a data analyst,” <https://www.franklin.edu/blog/how-become-a-data-analyst>, 2022, s interneta; pristupljeno 04.08.2022.
- [6] CIO, “What is data analytics,” <https://www.cio.com/article/191313/what-is-data-analytics-analyzing-and-managing-data-for-decisions.html>, 2022, s interneta; pristupljeno 04.08.2022.
- [7] KDnuggets, “Why python is one of the most preferred languages for data science?” <https://www.kdnuggets.com/2020/01/python-preferred-languages-data-science.html>, 2022, s interneta; pristupljeno 07.08.2022.
- [8] Wikipedia, “Pandas,” [https://en.wikipedia.org/wiki/Pandas_\(software\)](https://en.wikipedia.org/wiki/Pandas_(software)), 2022, s interneta; pristupljeno 07.08.2022.
- [9] —, “Plotly,” <https://en.wikipedia.org/wiki/Plotly>, 2022, s interneta; pristupljeno 07.08.2022.
- [10] GitHub, “Plotly,” <https://github.com/plotly/plotly.py>, 2022, s interneta; pristupljeno 07.08.2022.

Bibliografija

- [11] Wikipedia, “Teorija grafova,” https://hr.wikipedia.org/wiki/Teorija_grafova, 2022, s interneta; pristupljeno 07.08.2022.
- [12] Networkx, “Software for complex networks,” <https://networkx.org/documentation/stable/index.html>, 2022, s interneta; pristupljeno 07.08.2022.
- [13] J. Notebook, “What is the jupyter notebook?” https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html, 2022, s interneta; pristupljeno 07.08.2022.
- [14] Wikipedia, “Visual studio code,” https://en.wikipedia.org/wiki/Visual_Studio_Code, 2022, s interneta; pristupljeno 07.08.2022.
- [15] —, “Korelacija,” <https://hr.wikipedia.org/wiki/Korelacija>, 2022, s interneta; pristupljeno 08.08.2022.
- [16] —, “Cross-correlation,” <https://en.wikipedia.org/wiki/Cross-correlation>, 2022, s interneta; pristupljeno 08.08.2022.
- [17] —, “Pearson correlation,” https://en.wikipedia.org/wiki/Pearson_correlation_coefficient, 2022, s interneta; pristupljeno 14.08.2022.
- [18] NumPy, “Numpy corrcoef,” <https://numpy.org/doc/stable/reference/generated/numpy.corrcoef.html>, 2022, s interneta; pristupljeno 14.08.2022.
- [19] NetworkX, “Louvain communities,” https://networkx.org/documentation/latest/reference/algorithms/generated/networkx.algorithms.community.louvain.louvain_communities.html#networkx.algorithms.community.louvain.louvain_communities, 2022, s interneta; pristupljeno 10.08.2022.
- [20] S. Direct, “Singular value decomposition,” <https://www.sciencedirect.com/topics/engineering/singular-value-decomposition>, 2022, s interneta; pristupljeno 10.08.2022.
- [21] MIT, “Singular value decomposition,” https://web.mit.edu/be.400/www/SVD/Singular_Value_Decomposition.htm, 2022, s interneta; pristupljeno 10.08.2022.
- [22] NumPy, “numpy.linalg.svd,” <https://numpy.org/doc/stable/reference/generated/numpy.linalg.svd.html>, 2022, s interneta; pristupljeno 10.08.2022.
- [23] Wikipedia, “Atmospheric pressure,” https://en.wikipedia.org/wiki/Atmospheric_pressure, 2022, s interneta; pristupljeno 09.09.2022.
- [24] —, “Vlažnost zraka,” https://hr.wikipedia.org/wiki/Vla%C5%BEnost_zraka, 2022, s interneta; pristupljeno 10.09.2022.

Pojmovnik

API Application Programming Interface. 4

CSV Comma-separated values. 8

DHMZ Državni hidrometeorološki zavod. 9

HTML HyperText Markup Language. 15

PDF Portable Document Format. 15

SQL Structured Query Language. 14

SVD Singular value decomposition. 13, 14, 23, 24, 26, 33, 35, 39, 41, 42, 44, 47, 49, 53, 55, 56, 60

Sažetak

Glavni fokus ovog rada jest analiza meteoroloških prilika i pokušaj pronalaska mikroklima na području Primorsko-goranske i Istarske županije. Prostorno-vremenskom analizom pokušali smo ponaći sličnosti i razlike između 72 promatrana mjesta. Baza analize jesu preuzeti podaci s frekvencijom od tri sata u periodu od 31.10.2016. do 31.10.2021. Korelacijska i SVD analiza detaljnije su pojasnile meteorološke prilike promatranog područja. Baška, Mali Lošinj i Prezid su tri lokaliteta koja se jasno ističu od onih sličnih geografskih obilježja. Promatranjem temperature, tlaka i vlažnosti zraka na čitavom i na skraćenom skupu podataka dokazano je da navedena tri mjesta imaju mikroklima. Ipak, rezultate trebamo uzeti s dozom opreza jer postoje određena odstupanja preuzetih podataka od onih službenih.

Ključne riječi — korelacija, mikroklima, prostorno-vremenska analiza, SVD

Abstract

The main focus of this thesis is the analysis of meteorological conditions and an attempt to find the microclimate in the area of Primorsko-goranska and Istria counties. Through spatio-temporal analysis, we tried to find similarities and differences between the 72 observed towns. The foundation of the analysis are downloaded data with a frequency of three hours in the period from 31.10.2016. until 31.10.2021. Correlation and SVD analysis explained meteorological conditions of observed area in more detail. Baška, Mali Lošinj and Prezid are three localities that clearly stand out from those with similar geographical features. By observing the temperature, pressure and humidity of the air on the full and on the partial data set, it was proven that the three mentioned towns have microclimates. However, we should take the results with a dose of caution because there are certain deviations in downloaded data from the official ones.

Keywords — correlation, microclimate, spatio-temporal analysis, SVD