



Analiza podataka korišćenjem panelne regresije, radi istraživanja rasta broja turista po regionima Srbije.



Maksim Privalov SV81/2023,
Aleksa Ćetković SV77/2022



#Sklapanje podataka

Imali smo 6 datoteka sa:
~9.000 redova i 86 kolona
ukupno,

Morali smo izbaciti *kolone koje nisu validne*, analizirati
ručno *svaku tabelu* i sklopiti sve
u jednu datoteku.





Dobili smo

create-table.py

градски region	2013	718943	539937	52.8	83.7	<nul>
ran Vojvodine	2013	347138	151857	44.7	73.5	<nul>
ran Šumadije i Zapadne Srbije	2013	791545	149838	38.6	63.1	<nul>
ran Jужne i Istočne Srbije	2013	334809	83136	36.9	60.7	<nul>
REPUBLIKA SRBIJA	2014	2192268	1028732	55.1	74	1039599.999999
градски region	2014	753742	597765	65.2	81.3	421200.0
ran Vojvodine	2014	371690	168582	59.7	70.8	254900.0
ran Šumadije i Zapadne Srbije	2014	737507	165869	50.5	70.1	208700.0
ran Jужne i Istočne Srbije	2014	329529	96510	44.8	66.6	163800.0
REPUBLIKA SRBIJA	2015	2437165	1132221	56	75.2	1087700.0
градски region	2015	807607	650362	65.8	83.3	421300.0
ran Vojvodine	2015	413332	186041	59.1	71.5	253200.0
ran Šumadije i Zapadne Srbije	2015	854448	189950	51.8	70.5	224100.0
ran Jужne i Istočne Srbije	2015	361778	108868	46.5	68.5	189000.0
REPUBLIKA SRBIJA	2016	2753591	1281426	57.8	80.8	1114000.0
градски region	2016	913150	737303	68.5	84	420600.0
ran Vojvodine	2016	446492	230353	61.1	80.8	268300.0
ran Šumadije i Zapadne Srbije	2016	996565	218684	53.9	78.6	231000.0
ran Jужne i Istočne Srbije	2016	397384	122626	47	76.3	194100.0
REPUBLIKA SRBIJA	2017	3085866	1497173	61.9	80.4	1122780.0
градски region	2017	1035205	863162	72.7	85.3	434300.0
ran Vojvodine	2017	496625	227600	61.1	81.7	265700.0
ran Šumadije i Zapadne Srbije	2017	1086264	251190	61.5	76.2	225800.0
ran Jужne i Istočne Srbije	2017	467772	155221	54.6	71.8	187800.0
REPUBLIKA SRBIJA	2018	3630522	1710514	72.5	82.6	1149800.0
градски region	2018	1160582	971942	82.1	89	455300.0
ran Vojvodine	2018	538472	259956	70.3	83.2	270000.0
ran Šumadije i Zapadne Srbije	2018	1184051	293864	71.5	75.5	229400.0
ran Jужne i Istočne Srbije	2018	547417	184752	66	74.5	195100.0

1 datoteka sa svim bitnim kolonama

Zbog čudnosti podataka i velikih razlika između svake tabele, spajanje podataka je izvršeno samo korišćenjem *Python*-a i bibliotekom *csv*.



#”Ekstrapolacija”

Kao rezultat grupisanja podataka, ispostavilo se da za mnoge kriterijume (kolone) nemamo podatke pre 2014. godine.

Pošto je bilo teško izračunati ovoliki broj kriterijuma na osnovu samo malog broja vrednosti, odlučili smo popuniti podatke po formuli: *kriterijum* = *kriterijum u narednoj godini* – (*kriterijum za dve godine* – *kriterijum u narednoj godini*).

1	2	3	4	5
C1 Region	РЕПУБЛИКА СРБИЈА	Географски регион	Регион Војводине	Регион Шумадије и Западне Ср
C2 Year	2010	2010	2010	2010
C3 TotalTourists	2000597	618454	281342	747748
C4 ForeignTourists	682081	395488	102457	104443
C5 HousesWithInternet%	27.6	41.6	27.8	21.3
C6 CompaniesWithWebsites%	67.5	88.6	61.2	58.9
C7 AmountOfPeopleWithHighEducation	847200.0	395300.0	261700.0	142100.0
C8 PeopleWithHighEducationWithoutJob	87500.0	7500.0	39800.0	35000.0
C9 AmountOfPeopleWithSecondaryEducation	3226100.0	754100.0	849500.0	898600.0
C10 PeopleWithSecondaryEducationWithoutJob	562400.0	65200.0	187600.0	144500.0
C11 AmountOfPeopleWithoutEducation	2183100.0	293100.00000000023	572000.0	733900.0
C12 PeopleWithoutEducationWithoutJob	181000.0	14400.0	70000.0	45700.0
C13 AmountOfFemales	3036300.0	683100.0	810000.0	854400.0
C14 AmountOfMalesWithoutJob	450600.0	45200.0	174400.0	103700.0
C15 AmountOfFemales	3239400.0	757000.0	872700.0	906100.0
C16 AmountOfFemalesWithoutJob	382800.0	27200.0	123300.0	122000.0
C17 AmountOfPeopleInRegion	6256400.0	1422100.0	1682800.0	1771600.0
C18 AmountOfPeopleInRegionWithoutJob	833400.0	75000.0	297700.0	225700.0
C19 AmountOf_15-24_PeopleInRegion	902000.0	195000.0	249400.0	242300.0
C20 AmountOf_15-24_PeopleInRegionWithoutJ.	150800.0	26000.0	48400.0	27600.0
C21 AmountOf_15-64_PeopleInRegion	5105000.0	1181600.0	1392800.0	1439100.0
C22 AmountOf_15-64_PeopleInRegionWithoutJ.	833400.0	75000.0	296900.0	225200.0
C23 AmountOf_25-34_PeopleInRegion	1018500.0	206800.0	281000.0	206800.0
C24 AmountOf_25-34_PeopleInRegionWithoutJ.	245200.0	18000.0	97000.0	53900.0
C25 AmountOf_35-44_PeopleInRegion	952900.0	218100.0	253800.0	278300.0
C26 AmountOf_35-44_PeopleInRegionWithoutJ.	179400.0	13500.0	53900.0	72100.0
C27 AmountOf_45-54_PeopleInRegion	1040700.0	223000.0	294500.0	304500.0
C28 AmountOf_45-54_PeopleInRegionWithoutJ.	161700.0	10400.0	61000.0	53100.0
C29 AmountOf_55-64_PeopleInRegion	1106900.0	275900.0	315500.0	334400.0
C30 AmountOf_55-64_PeopleInRegionWithoutJ.	76300.0	5900.0	36200.0	18400.0
C31 AmountOf_65+ PeopleInRegion	1150400.0	240900.0	298000.0	333800.0

(Transposed csv)

Ovaj pristup u proseku odgovorio je opštoj tendenciji.



#Otkriti koji region ima najbolji rast broja turista

Procitali smo nas csv
fajl u DataFrame,
grupisani po *Region*'u, i
plotirali smo podatke sa
kolonama "Year" i
"TotalTourists"



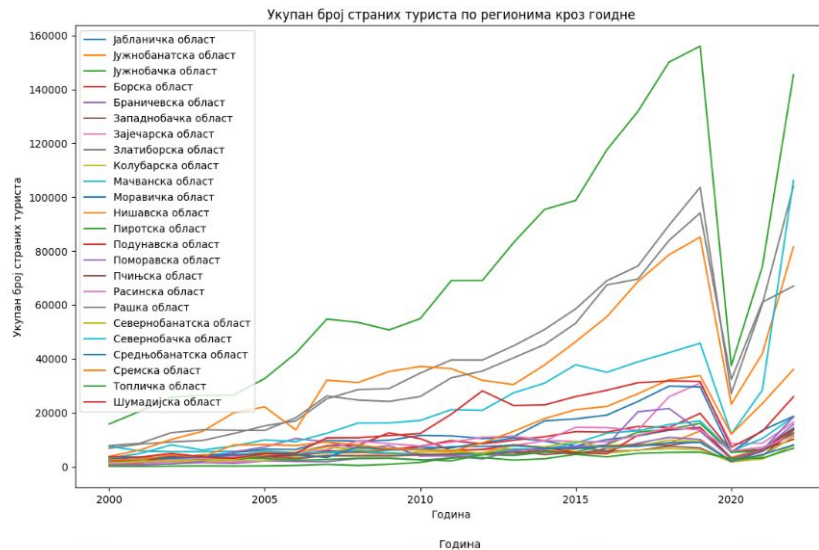
Iz grafika se vidi da
Beogradski region ima
najveći rast, ali i svi ostali
regioni takođe beleže porast
u broju turista.

S obzirom da ima ukupno
4 regiona, odlučili smo da
sprovedemo istraživanje i
po oblastima Srbije (naredni
slajd).



#Otkriti koja oblast ima najbolji rast broja turista

Mi smo već radili sa drugom bazom podataka i morali smo da uradimo dosta filtriranja pre nego što smo mogli grupisati podatke po oblastima i prikazati odnos broja turista prema godini.



*bez Beogradske oblasti

Naj "progresivnija" oblast je *Beogradska* (ako je prikažemo, sve linije na grafiku će biti na dnu).

Najprogresivnije od preostalih su *Raška oblast* i *Žlatiborska oblast*.

Najbolji rast je kod *Žuznobackoj oblasti*

P.s. *Žuznobačka oblast* je najprogresivnija kada je reč o stranim turistima.



Nama je bitna kolona *TotalTourists* – jer će pokazati kako će se povećanje određenih kriterijuma odraziti na broj turista.





#Otkriti koji faktori utiču na rast broja turista i pronaći najznačajnije faktore

Filtracijom smo dobili kriterijume koji najviše utiču na rast broja turista.

Regresija nad panelnim podacima

- Procenat domaćinstava koje posjeduju internet konekciju: “HousesWithInternet%”
- Procenat preduzeca koja imaju web sajt: “CompaniesWithWebSites%”
- Broj visokoobrazovanih stanovnika: “AmountOfPeopleWithHighEducation”
- Broj stanovnika starijih od 65 godina: “AmountOf_65+_PeopleInRegion”





		HousesWithInternet%	CompaniesWithWebSites%	AmountOfPeopleWithHighEducation	AmountOf_65+_PeopleInRegion
Region	Year				
Београдски регион	2018	82.1	89.0	455300.0	321000.0
	2019	89.0	84.0	472500.0	329400.0
	2020	93.9	89.5	470700.0	337600.0
Регион Војводине	2018	70.3	83.2	270000.0	354600.0
	2019	81.7	91.0	264100.0	361800.0
	2020	74.7	86.3	266100.0	369500.0
Регион Шумадије и Западне Србије	2018	71.5	75.5	229400.0	392600.0
	2019	75.0	78.9	235900.0	402300.0
	2020	77.5	75.5	243100.0	411600.0
Регион Јужне и Источне Србије	2018	66.0	74.5	195100.0	321800.0
	2019	71.1	75.4	190100.0	326300.0
	2020	77.3	80.5	194700.0	330300.0



#Pooled regression

Prvo smo isitali pretpostavku nezavisnosti reziduala i provjerili da li su reziduali normalno distribuirani.

Model nije ispunio ove pretpostavke pa smo odlučili da bi Fixed Effects ili Random Effects model možda bio prikladniji za naše podatke.



Fixed effects and Random effects model

Koristeci biblioteku 'linearmodels' fitovali smo podatke modelom sa fiksnim i nasumicnim efektima.

Hausman test:

- Pretpostavlja da ne postoji veza izmedju nezavisnih promjenljivih i pojedinačnih efekata (kovarijansa je nula)
- Ako je hipoteza ispunjena onda se preferira RE, u suprotnom FE



Hvala na pažnji!