

Analiza čimbenika rizika pretilosti primjenom statističkih metoda

Cukrov, Petra

Undergraduate thesis / Završni rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Rijeka / Sveučilište u Rijeci**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:195:125276>

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

Download date / Datum preuzimanja: **2025-02-22**



Repository / Repozitorij:

[Repository of the University of Rijeka, Faculty of Informatics and Digital Technologies - INFORI Repository](#)





Sveučilište u Rijeci

**Fakultet informatike
i digitalnih tehnologija**

Sveučilišni prijediplomski studij Informatika

Petra Cukrov

Analiza čimbenika rizika pretilosti primjenom statističkih metoda

Završni rad

Mentor: Doc. dr. sc. Lucia Načinović Prskalo

Rijeka, rujan 2024.



Sveučilište u Rijeci

Fakultet informatike
i digitalnih tehnologija

UNIRI



Rijeka, 9. travnja 2024.

Zadatak za završni rad

Pristupnik/ica: Petra Cukrov

Naziv završnog rada: Analiza čimbenika rizika pretilosti primjenom statističkih metoda

Naziv završnog rada na engleskom jeziku: Analysis of obesity risk factors using statistical methods

Sadržaj zadatka: Zadatak završnog rada je istražiti čimbenike koji utječu na pretilost te analiza i vizualizacija rezultata primjenom statističkih metoda. Glavni cilj analize je interpretirati raširenost pretilosti, istražiti razlike među državama i regijama te utvrditi veze između specifičnih prehrambenih i osobnih navika i razvoja pretilosti. Za analizu i vizualizaciju podataka koristit će se suvremeni alati koji će u radu također biti opisani.

Mentorica
Doc. dr. sc. Lucia Načinović Prskalo

Voditelj za završne radove
Izv. prof. dr. sc. Miran Pobar

Zadatak preuzet: 9.4.2024.

(potpis pristupnika/ice)

Sažetak

Pretilost je već dugu godinu jedan od ključnih zdravstvenih problema na globalnoj razini. Zbog svoje kompleksne prirode zahtjeva pregled iz različitih perspektiva, pa su tako dosada provedena mnoga istraživanja s ciljem boljeg razumijevanja njenih uzroka i posljedica.

Ovaj rad nadovezuje se na dosad provedena istraživanja i koristeći podatke o 30 Europskih država, preuzete s Eurostata, ispituje poveznice između različitih prehrambenih i osobnih navika i razvoja pretilosti.

Opisan je proces prikupljanja i pripreme podataka za analizu u kojoj su korištene metode deskriptivne i inferencijalne statistike, analiza korelacije i regresijska analiza.

Statističkim metodama pronađena je umjerena veza između konzumacije šećerom zaslađenih pića (engl. *Sugar sweetened beverages, SSB*), stope urbanizacije i pretilosti. Također su pronađene veze između različitih prehrambenih navika i fizičke aktivnosti koje upućuju na to da su zdrave navike međusobno povezane, kao i nezdrave.

Analizom su otkrivene regionalne razlike u životnim stilovima koje mogu pozitivno ili negativno utjecati na pretilost, a za bolje razumijevanje i dobivanje potpunije slike pretilosti u Europi predloženo je proširivanje analize na dodatne faktore.

Rezultati su vizualizirani pomoću Pythonove biblioteke Seaborn i Tableau-a.

Ključne riječi: analiza; vizualizacija; pretilost; korelacija; Europa; BMI

SADRŽAJ

1.Uvod.....	1
2.Pregled literature	2
3.Skup podataka	3
4.Obrada skupa podataka	5
5. Metode	13
6.Analiza podataka.....	14
6.1. Pregled stanja po državama	14
6.2. Korelacija Indeksa tjelesne mase (<i>BMI</i>) i faktora pretilosti po državama	20
6.3. Veze između različitih faktora pretilosti.....	25
6.3.1. Različite prehrambene navike.....	26
6.3.2. Prehrambene navike i fizička aktivnost	27
6.3.3. Različiti oblici fizičke aktivnosti	30
6.3.4. Alkohol i cigarete.....	30
6.3.5. Urbanizacija	31
6.4. Usporedba faktora pretilosti i načina života u Regijama Europe	32
6.4.1. Prehrambene navike.....	34
6.4.2. Fizička aktivnost	37
6.4.3. Alkohol i cigarete.....	39
6.4.4. Urbanizacija	41
7.Zaključak.....	42
8.Popis literature	43
9. Popis slika	45
10. Popis tablica	46

1. Uvod

U ovom radu provedena je analiza podataka prikupljenih s Eurostata, fokusirajući se na faktore koji utječu na pretilost u državama Europe. Skup podataka obuhvaća demografske podatke te informacije o osobnim i prehrambenim navikama kao što su konzumacija hrane i pića, konzumacija alkohola, pušenje te razina i učestalost bavljenja fizičkim aktivnostima.

Glavni cilj analize je interpretirati raširenost pretilosti u Europi, istražiti razlike među državama i regijama te utvrditi veze između specifičnih prehrambenih i osobnih navika i razvoja pretilosti.

Analiza podataka obuhvaća deskriptivnu statistiku za opis osnovnih karakteristika podataka te inferencijalnu statistiku, analizu korelacije i regresijsku analizu za istraživanje veza između različitih faktora.

Rad je razvijen u okruženju Jupyter Notebook koristeći programski jezik Python, integrirajući ključne biblioteke kao što su Pandas za manipulaciju podacima, Numpy za numeričke izračune i statistics i SciPy.stats za statističku analizu.

Rezultati analize su grafički prikazani kako bi se olakšalo njihovo razumijevanje, a vizualizacije su izrađene korištenjem Tableau-a i biblioteka programskog jezika Python: Seaborn i StatsModels.

2. Pregled literature

Pretilost je prema Svjetskoj zdravstvenoj organizaciji (engl. *World Health Organisation*, WHO) definirana kao "prekomjerno nakupljanje masnog tkiva koje može utjecati na zdravlje i dijagnosticira se pri indeksu tjelesne mase (BMI) većem od 30 kg/m²" (Blüher, 2019).

Predstavlja veliki zdravstveni problem na globalnoj razini jer znatno povećava rizik od raznih bolesti, poput dijabetesa tipa 2 i srčanih problema, što dovodi do smanjenja kvalitete života i očekivanog trajanja života (Blüher, 2019).

S ciljem boljeg razumijevanja složenih komponenti pretilosti i posljedica koje nosi sa sobom, istraženi su različiti biološki, okolišni i ekonomski čimbenici koji pridonose njenom razvoju. Provedena su brojna istraživanja usredotočena na utjecaj pretilosti na kvalitetu života, ekonomske aspekte povezane s troškovima liječenja, te specifične rizične faktore kod različitih dobnih skupina, pretežito djece i adolescenata.

Kao reference i inspiracija za ovu analizu poslužili su članci koji se bave pretilošću na svjetskoj i europskoj razini. U ovom odlomku izdvojeni su neki od njih.

Studija provedena na razini Europe, koja je pratila razvoj pretilosti od 1990. do 2016. godine prema spolu, pokazala je da je pretilost u Europskim državama sve češća pojava, te da su razlike u pretilosti među državama jako male. Predviđanja su tada sugerirala da bi postotak pretilosti veći od 30% mogao biti dostignut do 2020. godine među muškarcima u Bugarskoj, Češkoj, Irskoj, Španjolskoj i Ujedinjenom Kraljevstvu, dok bi u Mađarskoj i na Malti mogao doseći 33,3%. Kod žena bi se sličan postotak mogao postići u Hrvatskoj, Češkoj, Grčkoj, Irskoj, Litvi i na Malti, te do 33,1% u Ujedinjenom Kraljevstvu (Krzyszczoszek, Laudańska-Krzemińska, & Bronikowski, 2019). Koliko se točnom pokazala ta pretpostavka bit će ispitano dalje u radu.

Dokazano je postojanje povezanosti između prehrambenih navika i fizičke aktivnosti, te je zaključeno da u slučaju pretilosti kod djece i adolescenata značajnu ulogu igra način života roditelja, koji posljedično utječe na životne navike djece (Moschonis & Trakman, 2023).

Kod ispitivanja utjecaja konzumacije alkohola na razvoj pretilosti zaključeno je da iako s povećanjem količine alkohola koji se konzumira ne dolazi i do povećanja BMI-a, alkohol ipak može biti štetniji za osobe s prekomjernom težinom jer se njihovo tijelo teže oporavlja od posljedica konzumacije velike količine alkohola (Åberg & Färkkilä, 2020).

Istraživanje korelacije između pušenja cigareta i pretilosti pokazalo je da, iako je pušenje povezano s raznim bolestima i poremećajima, ono nema značajan utjecaj na pretilost. Štoviše, povećanjem broja popušanih cigareta dolazi do smanjenja BMI-a i krvnog tlaka (Gümüşa, Kayhan, & Kayhana, 2013).

Unatoč velikom broju provedenih istraživanja i preglednih članaka (engl. *review articles*), nova istraživanja i ponovno sagledavanje postojećih rezultata uvijek su potrebni zbog stalnih promjena u načinu života i novih faktora koje treba uključiti u analizu.

3. Skup podataka

Podaci korišteni za analizu preuzeti su s Eurostata (Eurostat, 2019) , a prikupljeni su tijekom istraživanja provedenog 2019. godine. Iako su dostupni noviji podaci za neke varijable poput broja stanovnika pojedinih zemalja, većina faktora obuhvaćenih ovim radom ima podatke iz 2019. godine, zbog čega je ta godina uzeta kao referentna.

Skup podataka sastoji se od normiranih podataka koji opisuju broj stanovnika pojedine države, postotak stanovništva s određenim indeksom tjelesne mase (eng. *Body Mass Index, BMI*) i faktore koji bi mogli utjecati na pretilost. Faktore pretilosti možemo podijeliti na prehrambene navike i način života. Svaki faktor podijeljen je u kategorije.

Indeks tjelesne mase (engl. *Body Mass Index, BMI*)

- Pothranjeni (engl. *Underweight*)
- Normalne težine (engl. *Normal*)
- Pretili (engl. *Obese*)

Prehrambene navike:

- Dnevna konzumacija voća i povrća (engl. *Daily consumption of fruit and vegetables*), s kategorijama:
 - 0 porcija tjedno (engl. *0 portions*)
 - 1 do 4 porcije tjedno (engl. *From 1 to 4 portions*)
 - 5 ili više porcija tjedno (engl. *5 or more portions*)
- Učestalost konzumacije pića zaslađenih šećerom (engl. *Frequency of drinking sugar-sweetened soft drinks (Sugar sweetened beverages, SSB)*), s kategorijama:
 - Barem jednom dnevno (engl. *At least once a day*)
 - 1 do 3 puta tjedno (engl. *From 1 to 3 times a week*)
 - 4 do 6 puta tjedno (engl. *From 4 to 6 times a week*)
 - Nikad ili povremeno (engl. *Never or occasionally*)

Način života:

- Bavljenje fizičkim aktivnostima koje nisu vezane za posao (engl. *Performing (non-work-related) physical activities*), s kategorijama:
 - Pješaćenje kao način kretanja (engl. *Walking to get to and from place*)
 - Bicikliranje kao način kretanja (engl. *Cycling to get to and from place*)

- Aerobni sportovi (engl. *Aerobic sports*)
- Jačanje mišića (engl. *Muscle-strengthening*)
- Učestalost konzumacije alkohola (engl. *Frequency of alcohol consumption*), s kategorijama:
 - Svaki dan (engl. *Every day*)
 - Svaki tjedan (engl. *Every week*)
 - Svaki mjesec (engl. *Every month*)
 - Manje od jednom mjesečno (engl. *Less than once a month*)
 - Nikad ili ne u posljednjih 12 mjeseci (engl. *Never or not in the last 12 months*)
- Aktivni pušači cigareta (engl. *Daily smokers of cigarettes*), s kategorijama:
 - Ukupni broj pušača u državi (engl. *Total*)
 - Manje od 20 cigareta dnevno (engl. *Less than 20 cigarettes pre day*)
 - 20 ili više cigareta dnevno (engl. *20 or more cigarettes per day*)
- Vrijeme provedeno bavljenjem fizičkim aktivnostima koje doprinose zdravlju nevezanima za posao (engl. *Time spent on health-enhancing (non-work-related) aerobic physical activity*), s kategorijama:
 - 0 minuta tjedno (engl. *Zero minutes*)
 - 1 do 149 minuta tjedno (engl. *From 1 to 149 minutes*)
 - više od 150 minuta tjedno (engl. *150 minutes or over*)
 - više od 300 minuta tjedno (engl. *300 minutes or over*)
- Stopa urbanizacije (engl. *Distribution of population by degree of urbanisation*) s kategorijama:
 - Stanovništvo koje živi u velikim gradovima (engl. *Cities*)
 - Stanovništvo koje živi u predgrađima i manjim gradovima (engl. *Towns and suburbs*)
 - Stanovništvo koje živi u ruralnim predjelima (engl. *Rural areas*)

4. Obrada skupa podataka

Podaci preuzeti sa Eurostata su u .xlsx formatu, pri čemu svaka radna knjiga sadrži tri radna lista. Prvi radni list (Slika 1) je sažetak koji uključuje naslov, izvore podataka, vrijeme zadnjeg ažuriranja i promjene strukture podataka, mjernu jedinicu, godinu iz koje su podaci, spol i dob ispitanika (ako je specificirano), te vremensku frekvenciju podataka. Drugi radni list (Slika 2) sadrži opis strukture podataka, uključujući nazive stupaca i redaka. Treći radni list (Slika 3) sadrži kratki opis podataka i tablicu.

Contents	Time frequency	Unit of measure	Income quantile	Sex	Age class	Time
Sheet 1	Annual	Percentage	Total	Total	Total	2019

Slika 1: Prvi radni list .xlsx datoteke

	A	B	C
1	Structure		
2	Dimension		Position
3	Label		Label
4	Time frequency		Annual
5	Unit of measure		Percentage
6	Body Mass Index		Underweight
7	Body Mass Index		Normal
8	Body Mass Index		Obese
9	Income quantile		Total
10	Sex		Total
11	Age class		Total
12	Geopolitical entity (reporting)		European Union - 27 countries (from 2020)
13	Geopolitical entity (reporting)		European Union - 28 countries (2013-2020)
14	Geopolitical entity (reporting)		Belgium
15	Geopolitical entity (reporting)		Bulgaria
16	Geopolitical entity (reporting)		Czechia
17	Geopolitical entity (reporting)		Denmark
18	Geopolitical entity (reporting)		Germany
19	Geopolitical entity (reporting)		Estonia
20	Geopolitical entity (reporting)		Ireland
21	Geopolitical entity (reporting)		Greece
22	Geopolitical entity (reporting)		Spain
23	Geopolitical entity (reporting)		France
24	Geopolitical entity (reporting)		Croatia
25	Geopolitical entity (reporting)		Italy
26	Geopolitical entity (reporting)		Cyprus
27	Geopolitical entity (reporting)		Latvia
28	Geopolitical entity (reporting)		Lithuania
29	Geopolitical entity (reporting)		Luxembourg
30	Geopolitical entity (reporting)		Hungary
31	Geopolitical entity (reporting)		Malta
32	Geopolitical entity (reporting)		Netherlands
33	Geopolitical entity (reporting)		Austria
34	Geopolitical entity (reporting)		Poland
35	Geopolitical entity (reporting)		Portugal
36	Geopolitical entity (reporting)		Romania
37	Geopolitical entity (reporting)		Slovenia

Slika 2: Drugi radni list .xlsx datoteke

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Data extracted on 06/05/2024 00:32:24	from [ESTAT]											
2	Dataset:	Body mass index (BMI) by sex, age and income quintile [hlth_ehis_bm1\$defaultview]											
3	Last updated:	24/10/2022 23:00											
4													
5	Time frequency		Annual										
6	Unit of measure		Percentage										
7	Income quantile		Total										
8	Sex		Total										
9	Age class		Total										
10	Time		2019										
11													
12		BMI (Labels)	Underweight	Normal	Obese								
13	GEO (Labels)												
17	Bulgaria		2.4	44.1	13.2								
18	Czechia		1.6	40.0	19.3								
19	Denmark		2.3	48.8	15.9								
20	Germany		2.7	45.2	18.5								
21	Estonia		2.5	42.4	21.1								
22	Ireland		:	u	:	u	:	u	:	u			
23	Greece		1.4	42.3	16.2								
24	Spain		2.4	45.3	15.4								
25	France		4.3	50.3	14.4								
26	Croatia		1.7	34.5	22.6								
27	Italy		3.9	51.4	11.4								
28	Cyprus		3.9	47.7	14.6								
29	Latvia		2.6	40.7	22.3								
30	Lithuania		2.3	42.7	18.3								
31	Luxembourg		3.7	49.2	16.1								
32	Hungary		2.7	38.9	23.9								
33	Malta		1.9	34.2	28.1								
34	Netherlands		2.9	48.9	14.1								
35	Austria		2.6	46.3	16.7								
36	Poland		2.7	40.6	18.5								
37	Portugal		2.0	43.4	17.2								
38	Romania		1.0	42.5	10.5								
39	Slovenia		1.6	41.8	19.4								
40	Slovakia		2.2	40.0	19.3								
41	Finland		1.7	40.6	20.3								
42	Sweden		2.4	48.0	14.7								
43	Iceland		1.4	38.5	21.7								
44	Norway		2.0	48.4	13.8								
45	United Kingdom		:	:	:								
46	Serbia		2.3	45.2	16.8								
47	Türkiye		3.8	40.1	21.1								
48													
49	Special value												
50	:		not available										
51	Available flags:												
52	u		low reliability										

Slika 3: Treći radni list .xlsx datoteke

Za učitavanje u Jupyter Notebook korišten je Pandas DataFrame i funkcija **pandas.read_excel()**. Pri učitavanju tablica iz Excel datoteka, učitani su samo treći radni list, preskačući retke koji sadrže opis tablice, dok su naslovi stupaca zadržani.

Nakon učitavanja podataka u Pandas DataFrame nazivi stupaca promijenjeni su u smislene nazive direktno na podatkovnom okviru, naredbom `inplace=True`.

Na isti način učitane su i Excel datoteke za ostale faktore.

```
df_population_total=pd.read_excel(r"C:\Users\petra\Desktop\faks\treca\ZAVRSNI\Eurostat_data\Original datasets\population_2019_total.xlsx", sheet_name='Sheet 1', header=1, skiprows=10)
df_population_total.rename(columns={"GEO (Labels)": "GEO labels", "Unnamed: 1": "Population_total", "Unnamed: 2": "Flag_t"},inplace=True)
```

Podatkovni okviri za pojedine varijable povezivani su u jedan podatkovni okvir naredbom **merge**, unutarnjim spajanjem podatkovnih okvira na varijabli 'GEO labels'.

```
population_merge.drop([0,1,2,3,4,10,34,35,36,37,60,61,62,63,64,65,66,67], inplace=True)
population_merge.reset_index(drop=True, inplace=True)
```

population_merge							
	GEO labels	Population_total	Flag_t	Population_male	Flag_m	Population_female	Flag_f
0	European Union - 27 countries (from 2020)	446446444	bep	218073691	bep	228372753	bep
1	European Union - 28 countries (2013-2020)	513093556	bep	250974727	bep	262118829	bep
2	European Union - 27 countries (2007-2013)	509017310	bep	249002155	bep	260015155	bep
3	Euro area – 20 countries (from 2023)	:	NaN	:	NaN	:	NaN
4	Euro area - 19 countries (2015-2022)	341524067	bep	166873066	bep	174651001	bep
5	Belgium	11455519	NaN	5644826	NaN	5810693	NaN
6	Bulgaria	7000039	NaN	3395701	NaN	3604338	NaN
7	Czechia	10649800	NaN	5244194	NaN	5405606	NaN
8	Denmark	5806081	NaN	2889073	NaN	2917008	NaN

Slika 4: Podatkovni okvir nakon učitavanja i preimenovanja stupaca

Nakon učitavanja Excel datoteke obrisani su podatci koji sadrže podatke u različitim vremenskim intervalima, kao i podatci za države koje su suvišne ili za koje nisu dostupni ključni podaci (Irska, Ujedinjeno Kraljevstvo, Metropolitan France, Germany Including former GDR). Nakon brisanja nepotrebnih redaka resetiran je indeks tablice naredbom `reset_index()`.

Na isti način uvezeni su podaci za sve faktore koji će se ispitivati u analizi.

Slika 5 prikazuje podatkovni okvir s podacima o broju stanovnika u državama.

population_merge							
	GEO labels	Population_total	Flag_t	Population_male	Flag_m	Population_female	Flag_f
0	Belgium	11455519	NaN	5644826	NaN	5810693	NaN
1	Bulgaria	7000039	NaN	3395701	NaN	3604338	NaN
2	Czechia	10649800	NaN	5244194	NaN	5405606	NaN
3	Denmark	5806081	NaN	2889073	NaN	2917008	NaN
4	Germany	83019213	NaN	40966691	NaN	42052522	NaN
5	Estonia	1324820	NaN	625635	NaN	699185	NaN
6	Ireland	4904240	e	2428881	e	2475359	e
7	Greece	10724599	NaN	5208293	NaN	5516306	NaN
8	Spain	46937060	NaN	23009259	NaN	23927801	NaN
9	France	67177636	p	32475891	p	34701745	p
10	Metropolitan France	:	NaN	:	NaN	:	NaN

Slika 5: Podatkovni okvir nakon uklanjanja suvišnih redaka

U skup podataka dodan je stupac “**Region**”, kako bi bilo moguće grupirati države po regijama u kojima se nalaze. Gledano iz socioekonomske, geopolitičke ili geografske perspektive, Europa se može podijeliti na regije na više načina. U ovom radu, podjela je izvedena prema podacima iz članka “Europske regije”(IZZI: Europske regije, n.d.), a države su razvrstane po regijama na sljedeći način:

- Zapadna Europa
 - Belgija, Francuska, Luksemburg, Monako, Nizozemska
- Sjeverna Europa
 - Estonija, Finska, Island, Danska, Latvija, Litva, Norveška, Švedska
- Središnja Europa
 - Austrija, Češka, Hrvatska, Lihtenštajn, Mađarska, Njemačka, Poljska, Slovačka, Slovenija, Švicarska
- Južna Europa
 - Albanija, Grčka, Italija, Malta, Portugal, Španjolska
- Jugoistočna Europa
 - Bosna i Hercegovina, Bugarska, Cipar, Crna Gora, Kosovo, Makedonija, Rumunjska, Srbija, Turska (Europski dio).

Države su najprije spremljene u liste, a zatim mapirane u odgovarajuće regije `mapping.update()` funkcijom.

```
western = ['Ireland', 'France', 'Netherlands', 'Belgium', 'Luxembourg', 'Monaco']
northern = ['Iceland', 'Denmark', 'Norway', 'Sweden', 'Finland', 'Lithuania', 'Latvia', 'Estonia']
central = ['Germany', 'Austria', 'Switzerland', 'Liechtenstein', 'Slovenia', 'Hungary', 'Czechia', 'Slovakia', 'Poland', 'Croatia']
southern = ['Portugal', 'Spain', 'Italy', 'Greece', 'Albania', 'Malta']
southeastern = ['Romania', 'Bulgaria', 'Macedonia', 'Bosnia and Herzegovina', 'Serbia', 'Montenegro', 'Kosovo', 'Türkiye', 'Cyprus']

region_mapping = {country: 'Western' for country in western}
region_mapping.update({country: 'Northern' for country in northern})
region_mapping.update({country: 'Central' for country in central})
region_mapping.update({country: 'Southern' for country in southern})
region_mapping.update({country: 'Southeastern' for country in southeastern})
dataset['Region'] = dataset['GEO labels'].map(region_mapping)
dataset.insert(1, 'Region', dataset.pop('Region'))
```

	GEO labels	Region	Population_total	Flag_t	Population_male	Flag_m	Population_female	Flag_f	Underweight_total
0	Belgium	Western	11455519.0	/	5644826.0	/	5810693.0	/	3.2
1	Bulgaria	Southeastern	7000039.0	/	3395701.0	/	3604338.0	/	2.4
2	Czechia	Central	10649800.0	/	5244194.0	/	5405606.0	/	1.6
3	Denmark	Northern	5806081.0	/	2889073.0	/	2917008.0	/	2.3
4	Germany	Central	83019213.0	/	40966691.0	/	42052522.0	/	2.7
5	Estonia	Northern	1324820.0	/	625635.0	/	699185.0	/	2.5
6	Greece	Southern	10724599.0	/	5208293.0	/	5516306.0	/	1.4
7	Spain	Southern	46937060.0	/	23009259.0	/	23927801.0	/	2.4
8	France	Western	67177636.0	p	32475891.0	p	34701745.0	p	4.3
9	Croatia	Central	4076246.0	/	1972572.0	/	2103674.0	/	1.7
10	Italy	Southern	59816673.0	b	29131195.0	b	30685478.0	b	3.9

Slika 6: Skup podataka nakon dodavanja stupca "Region"

Kako za sve države nisu prikupljeni svi podaci, na primjer, podaci o fizičkoj aktivnosti ili konzumaciji alkohola, nedostajući podaci zamijenjeni su medijanom pojedine varijable u regiji u kojoj se država nalazi, što je prikazano na Slici 7.

```
#Latvia - 'Muscle-strengthening'
latvia_ms_median = df_northern['Muscle-strengthening'].median()
print(latvia_ms_median)
```

46.55

```
#Nizozemska - "Sugar sweetened drinks"
niz_1_median = df_western['At least once a day'].median()
print(niz_1_median)
niz_1_3_median = df_western['From 1 to 3 times a week'].median()
print(niz_1_3_median)
niz_4_6_median = df_western['From 4 to 6 times a week'].median()
print(niz_4_6_median)
niz_never_median = df_western['Never or occasionally'].median()
print(niz_never_median)
```

9.6
15.9
5.3
69.9

```
#Turska - "Level of urbansation"
turk_city_median = df_southeastern['Living in cities'].median()
turk_town_median = df_southeastern['Living in towns_sub'].median()
turk_rural_median = df_southeastern['Living in rural areas'].median()
```

Slika 7: Računanje medijana regije za nedostajuće vrijednosti

Podaci koji nedostaju bili su zapisani ili kao „NaN“, ili kao „.“, svi su postavljeni na „NaN“ naredbom

```
dataset.replace(":", np.nan, inplace=True).
```

Prijašnje NaN vrijednosti zamijenjene su izračunatim medijanima naredbom

```
dataset['Muscle-strengthening'].replace({np.nan: latvia_ms_median}, inplace=True).
```

U ovom procesu iz skupa podataka je uklonjena Finska zbog nedostatka podataka za nekoliko faktora.

Tablice za neke varijable uz stupac vrijednosti sadrže i stupac rezerviran za zastavice tj. oznake koje pobliže opisuju broj ili nedostatak informacije. Takvi stupci su u podatkovnom okviru nazvani "**Flag_ime/kratica_atributa**", a objašnjenja zastavica su u nastavku:

- **:** not available (nedostupno)
- **u** low reliability (niska pouzdanost)
- **bep** break in time series, estimated, provisional (prekid u vremenskom nizu, procijenjeno, privremeno)
- **b** break in time series (prekid u vremenskom nizu)
- **e** estimated (procijenjeno)
- **p** provisional (privremeno)

Nakon čišćenja podataka, u tablici su podaci za 30 europskih država, organizirani u kategorije. Na slici 8 je prikazan dio podataka.

dataset

	GEO labels	Region	Population_total	Flag_t	Population_male	Flag_m	Population_female	Flag_f	Underweight_total	Normal_total	Obese_total	Under
0	Belgium	Western	11455519.0	/	5644826.0	/	5810693.0	/	3.2	48.0	15.9	
1	Bulgaria	Southeastern	7000039.0	/	3395701.0	/	3604338.0	/	2.4	44.1	13.2	
2	Czechia	Central	10649800.0	/	5244194.0	/	5405606.0	/	1.6	40.0	19.3	
3	Denmark	Northern	5806081.0	/	2889073.0	/	2917008.0	/	2.3	48.8	15.9	
4	Germany	Central	83019213.0	/	40966691.0	/	42052522.0	/	2.7	45.2	18.5	
5	Estonia	Northern	1324820.0	/	625635.0	/	699185.0	/	2.5	42.4	21.1	
6	Greece	Southern	10724599.0	/	5208293.0	/	5516306.0	/	1.4	42.3	16.2	
7	Spain	Southern	46937060.0	/	23009259.0	/	23927801.0	/	2.4	45.3	15.4	
8	France	Western	67177636.0	p	32475891.0	p	34701745.0	p	4.3	50.3	14.4	
9	Croatia	Central	4076246.0	/	1972572.0	/	2103674.0	/	1.7	34.5	22.6	
10	Italy	Southern	59816673.0	b	29131195.0	b	30685478.0	b	3.9	51.4	11.4	
11	Cyprus	Southeastern	875899.0	/	427831.0	/	448068.0	/	3.9	47.7	14.6	
12	Latvia	Northern	1919968.0	/	884909.0	/	1035059.0	/	2.6	40.7	22.3	
13	Lithuania	Northern	2794184.0	/	1295591.0	/	1498593.0	/	2.3	42.7	18.3	
14	Luxembourg	Western	613894.0	/	308720.0	/	305174.0	/	3.7	49.2	16.1	
15	Hungary	Central	9772756.0	/	4675821.0	/	5096935.0	/	2.7	38.9	23.9	
16	Malta	Southern	493559.0	/	251836.0	/	241723.0	/	1.9	34.2	28.1	

Slika 8: Finalni skup podataka za analizu

Za provođenje analize posebno za muškarce i žene, na isti način su dodani i očišćeni novi skupovi podataka s podacima za pojedini spol.

5. Metode

U ovom radu za analizu podataka korištene su metode deskriptivne i inferencijalne statistike i provedene su analiza korelacije i regresijska analiza.

Deskriptivna statistika je grana statistike koja se koristi za opisivanje i sažimanje skupa podataka (Assistance, 2006). Metode deskriptivne statistike korištene u ovom radu su **aritmetička sredina** (engl. *mean*), srednja vrijednost skupa dobivena zbrajanjem svih uzoraka i dijeljenjem na broj uzoraka, **medijan** (engl. *median*), koji opisuje srednju vrijednost skupa zanemarujući ekstremne vrijednosti (engl. *outliers*) i **kvartilni raspon** za opis raspršenosti vrijednosti (Walpole, Myers, Myers, & Ye, 2011).

Najčešće korištena metoda analize u radu je **analiza korelacije**. Korelacija je statistička mjera koja opisuje stupanj povezanosti između dvije varijable. Ispituje se pomoću procjenitelja, a u ovom radu kao procjenitelj je korišten **Pearsonov koeficijent korelacije**. On poprima vrijednosti između -1 i 1, pri čemu -1 ukazuje na savršenu negativnu korelaciju, a 1 na savršenu pozitivnu korelaciju (Assistance, 2006). Uz Pearsonov koeficijent korelacije izračunate su i p-vrijednosti kako bi se ispitalo jesu li pronađene korelacije statistički značajne. Prilikom ispitivanja značajnosti postavljaju se nulta i alternativna hipoteza. Nulta hipoteza tvrdi da ne postoji statistički značajna povezanost između dviju varijabli, dok alternativna hipoteza tvrdi da postoji. Ovisno o dobivenoj p-vrijednosti prihvaća se ili odbacuje nulta hipoteza. Ako je p-vrijednost manja od razine značajnosti, nulta hipoteza se odbacuje i prihvaća se alternativna, tj. varijable su zavisne i rezultat je statistički značajan. Ako je p-vrijednost veća ili jednaka razini značajnosti prihvaća se nulta hipoteza (Benšić & Šuvak, 2013). Razina značajnosti u ovom radu je 0.05, tako da se svi rezultati s p-vrijednostima manjim od toga smatraju značajnima (Benšić & Šuvak, 2013).

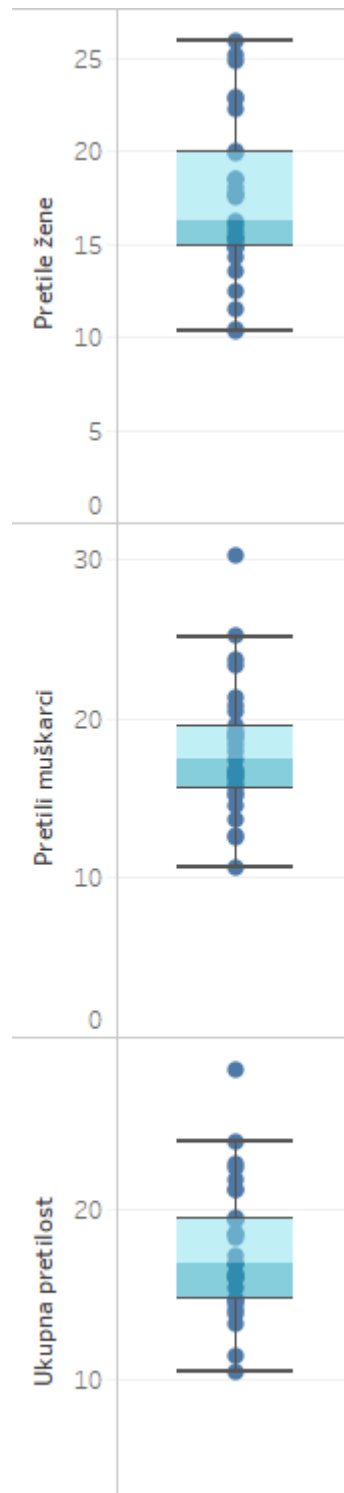
Osim analize korelacije, provedena je i **jednostavna linearna regresija** koja ima za cilj ustanoviti prirodu ovisnosti slučajne (zavisne) varijabli o nezavisnoj varijabli (Benšić & Šuvak, 2013). Dok korelacija pruža informacije o postojanju linearne veze između dvije varijable, regresija pomaže razumjeti kako promjene jedne varijable možemo objasniti putem druge varijable (Walpole, Myers, Myers, & Ye, 2011).

Za usporedbu srednjih vrijednosti između dviju nezavisnih skupina korišten je **neupareni T-test**, a za opis stanja po regijama korištena je **komparativna analiza**.

6. Analiza podataka

6.1. Pregled stanja po državama

Koristeći funkcije iz Python biblioteke `statistics` (`statistics` — Mathematical statistics functions, n.d.)



Slika 9: Box-plot

, ispitan je medijan pretilosti za ukupno stanovništvo te za mušku i žensku populaciju u europskim državama. Rezultati su prikazani *box-plot* dijagramom.

Tablica 1: Box-plot vrijednosti

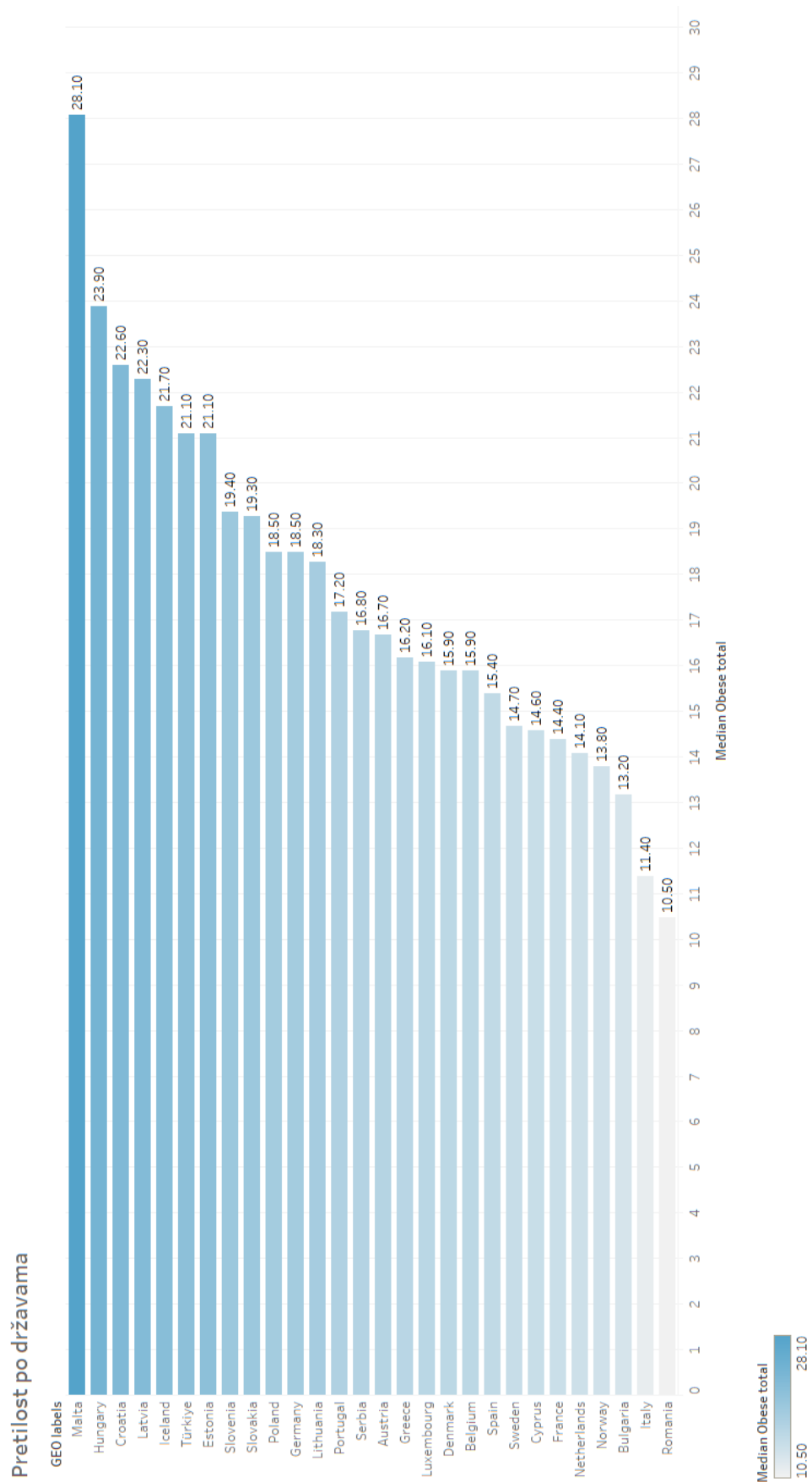
	Q1 (Prvi kvartil)	Medijan	Q3 (Treći kvartil)	IQR (Kvartilni raspon)
Ukupna populacija	14.7	16.8	19.4	4.7
Muškarci	15.6	16.2	19.5	3.9
Žene	14.8	17.4	19.9	5.1

Iz dijagrama na Slici 9 i podataka iz Tablice 1 vidi se da medijan ukupnog pretilog stanovništva po državi iznosi 16.8, što bi značilo da pola država u Europi ima stopu pretilosti manju od ove, a pola veću, odnosno da je u pola europskih država postotak stanovnika koji imaju BMI veći od 30 veći od 16.8 posto. Srednja vrijednost stope pretilosti nešto je veća kod žena nego kod muškaraca, a i kvartilni raspon (*IQR*) kod žena je veći, što ukazuje na veću raznolikost u razinama pretilosti među ženama u državama Europe, u usporedbi s vrijednostima kod muškaraca.

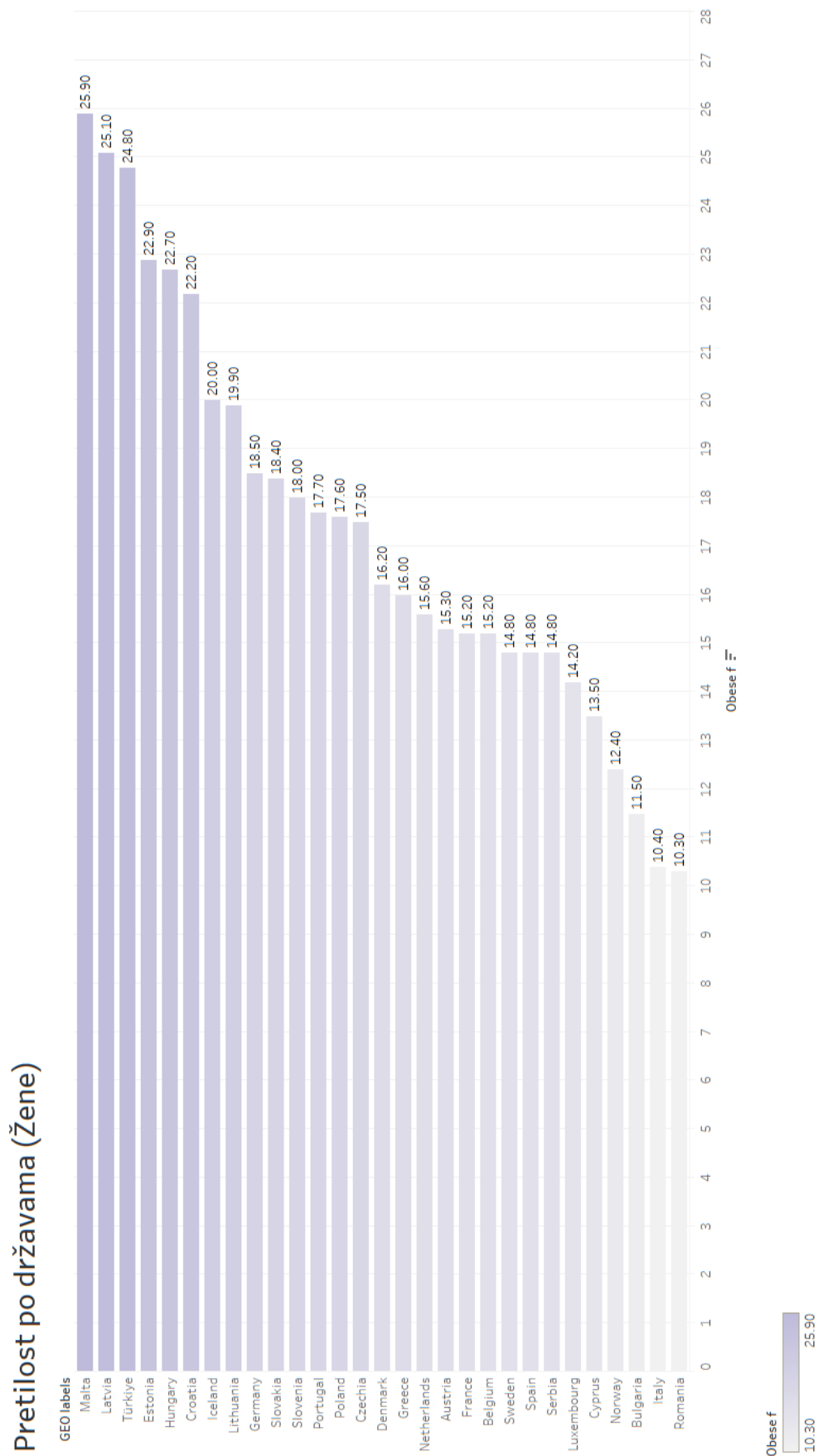
Primjenom neuparenog T-testa iz Pythonove biblioteke SciPy ispitana je statistička značajnost razlike u postotku pretilosti između muškaraca i žena.

```
t_stat, p_value = stats.ttest_ind(dataset['Obese_m'], dataset['Obese_f'])
```

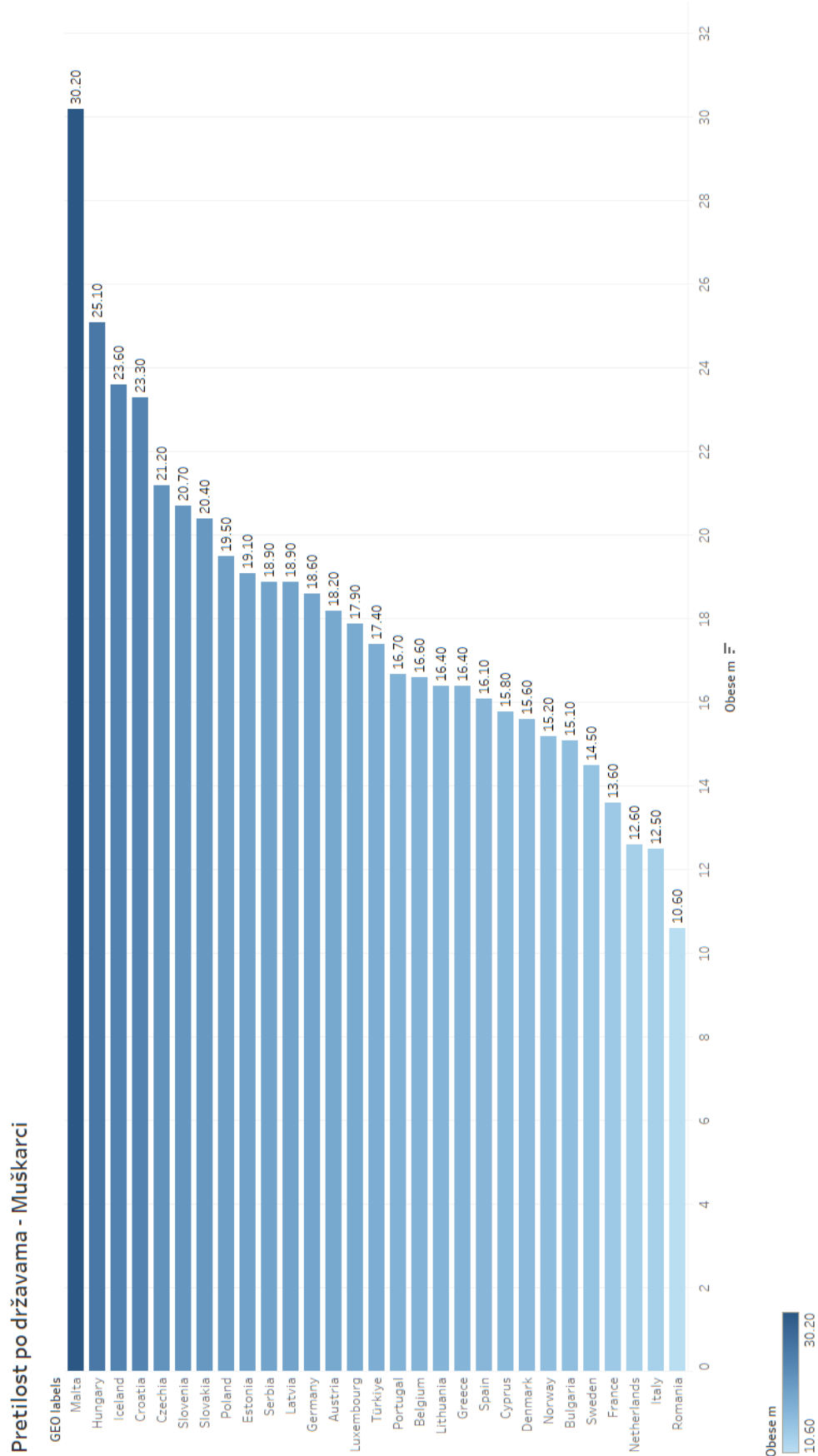
Dobiveni rezultati su 0.6053 za t-statistiku i 0.54 za p-vrijednost iz čega se može zaključiti da razlika između postotka pretilosti muškog i ženskog stanovništva u Europi nije statistički značajna, već je vjerojatno rezultat slučajnosti.



Slika 10: Pretilost po državama -Total



Slika 11: Pretilost po državama -Žene



Slika 12: Pretilost po državama -Muškarci

Iz dijagrama na Slikama 10, 11 i 12 iščitavamo da su europske države s najvećim postotkom ukupnog pretilog stanovništva Malta, Mađarska, Hrvatska, Latvija i Island. U broju pretilih muškaraca prednjače Malta, Mađarska, Island, Hrvatska i Češka, a pretilih žena je najviše na Malti, Latviji, Turskoj, Estoniji i Mađarskoj, tim redom.

Osvrnemo li se na istraživanje provedeno od 1990 do 2016. godine spomenuto u pregledu literature (Krzysztozek, Laudańska-Krzemińska, & Bronikowski, 2019), vidimo da se predviđanja nisu ostvarila, ali trendovi pretilosti ipak prate njihov smjer. Postotak pretilosti među muškarcima je stvarno dosegao 30% na Malti, a u Mađarskoj mu se približava sa 25,2%. Iako postotak pretilosti među ženama u ni jednoj državi nije dosegao predviđenih 30%, Hrvatska i Malta, za koje se to predviđalo, su u vrhu po postotku pretilog ženskog stanovništva.

6.2. Korelacija Indeksa tjelesne mase (*BMI*) i faktora pretilosti po državama

Kako bi se utvrdilo koji faktori najviše utječu na postotak pretilosti u državi, provedena je analiza korelacije između kategorija svakog faktora i BMI kategorije *Obese* za ukupni broj stanovnika, *Obese_total*. Pearsonov koeficijent korelacije, pripadajuća p-vrijednost i regresija računati su pomoću funkcija iz Pythonove biblioteke SciPy.stats.

```
corr, p_value = stats.pearsonr(dataset['Obese_total'], dataset['Walking to get to and from place'])
```

Nakon analize korelacije istaknulo se nekoliko umjerenih statistički značajnih korelacija čije su vrijednosti zapisane u Tablici 2. Za te faktore provedena je i regresijska analiza čiji su rezultati prikazani u Tablici 3. Slike 13, 14 i 15 prikazuju rezultate ovih analiza na regresijskim dijagramima.

```
X = dataset['Living in cities']
y = dataset['Obese_total']
X = sm.add_constant(X)
model = sm.OLS(y, X).fit()
print(model.summary())
```

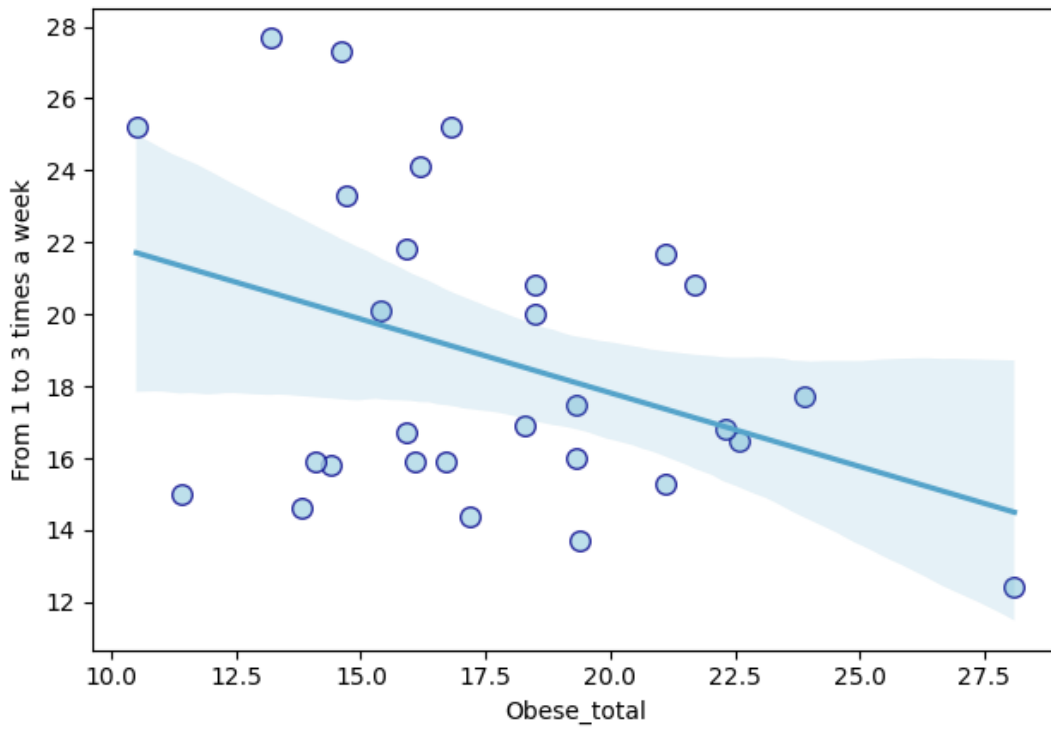
Tablica 2: Korelacija faktora pretilosti i BMI za ukupno stanovništvo

Varijabla 1	Varijabla 2	Pearsonov koeficijent korelacije	P-vrijednost
From 1 to 3 times a week (Konzumacija šećerom zaslađenih pića 1 do 3 puta tjedno)	Obese total	-0.37	0.04
Living in towns and subs	Obese total	-0.43	0.02
Living in cities	Obese total	0.38	0.04

Tablica 3: : Regresija faktora pretilosti i BMI za ukupno stanovništvo

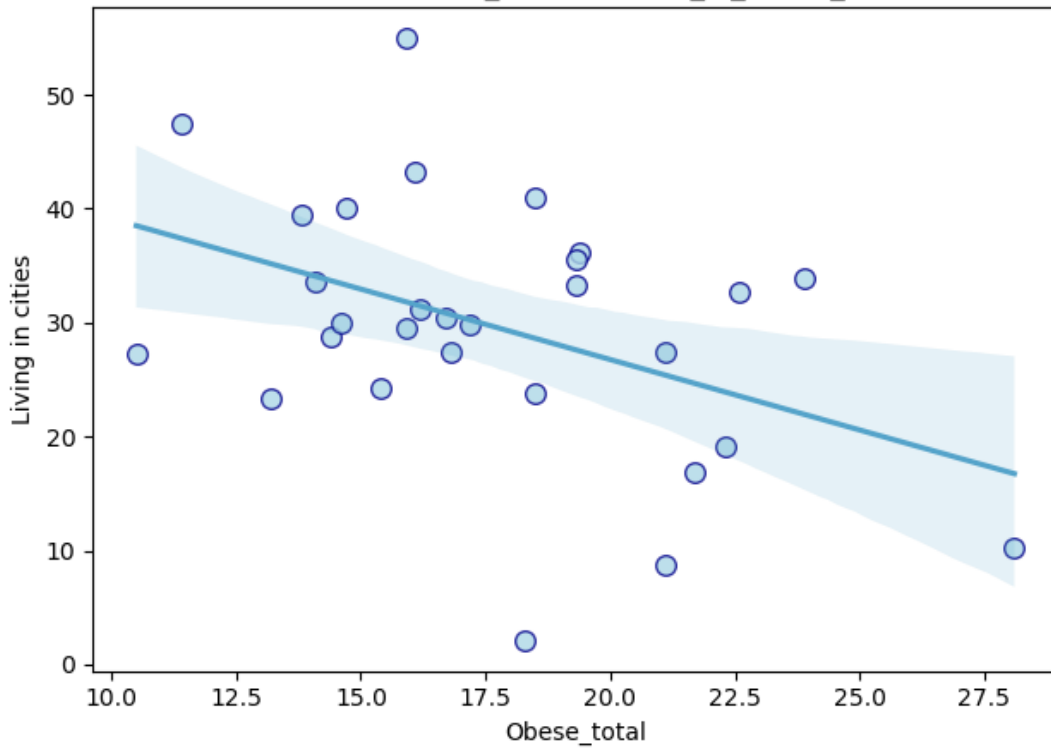
Nezavisna varijabla	Zavisna varijabla	Koeficijent regresije (beta)	Standardna pogreška (SE)	P- vrijednost	R²
From 1 to 3 times a week (Konzumacija šećerom zaslađenih pića 1 do 3 puta tjedno)	Obese total	-0.349	0.164	0.043	0.143
Living in towns and subs	Obese total	-0.1478	0.060	0.021	0.183
Living in cities	Obese total	0.1023	0.048	0.042	0.144

Korelacija Obese_total & From 1 to 3 times a week

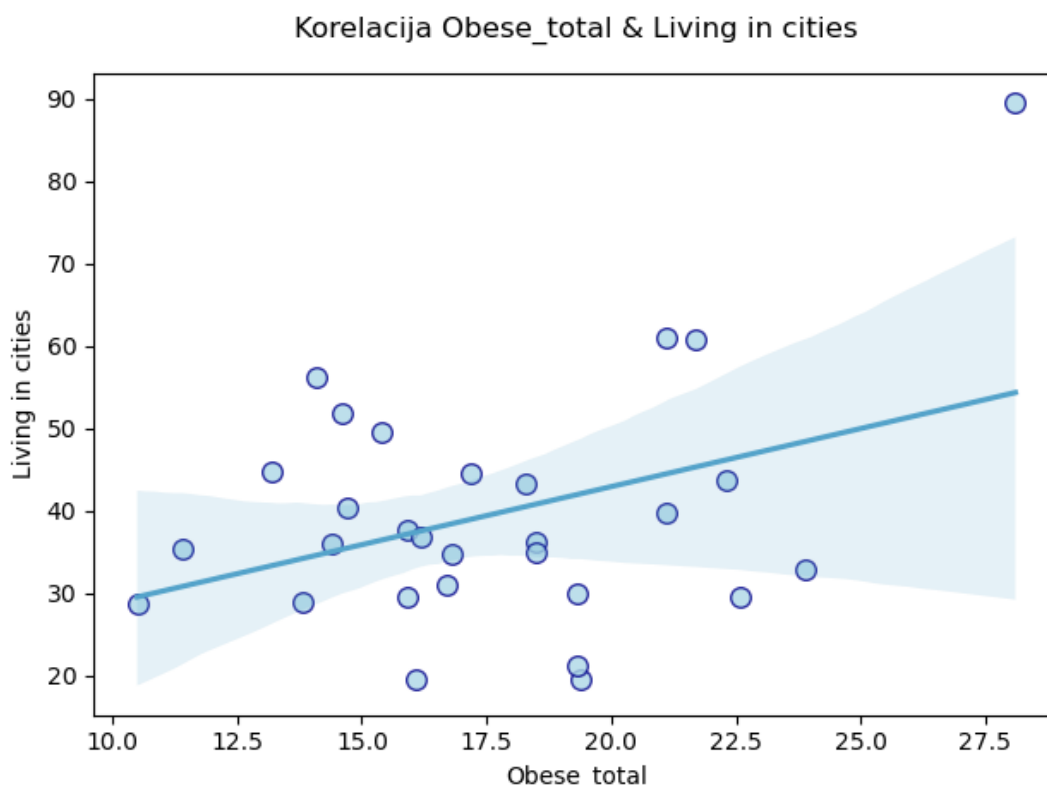


Slika 13: : Korelacija konzumacije SSB-a 1 do 3 puta tjedno i pretilosti

Korelacija Obese_total & Living_in_towns_sub



Slika 14: : Korelacija života u manjim gradovima i predgrađima i pretilosti



Slika 15: Korelacija života u velikim gradovima i pretilosti

Konzumiranje šećerom zaslađenih pića (*Sugar sweetened beverages, SSB*) 1 do 3 puta tjedno (*From 1 to 3 times a week*) pokazalo je umjerenu negativnu korelaciju i negativan koeficijent regresije s porastom pretilosti (*Obese_total*), što znači da je u državama gdje veći postotak ljudi konzumira 1 do 3 *SSB*-a tjedno stopa pretilosti manja. Iako su zaslađena pića na lošem glasu, ovi rezultati sugeriraju da njihovo umjereno konzumiranje ne utječe nužno na povećanje tjelesne mase.

Dosadašnja istraživanja nisu rezultirala jedinstvenim zaključkom o tome utječe li i u kolikoj mjeri konzumacija *SSB*-a na porast tjelesne mase. Nekoliko istraživanja pokazalo je da njihova konzumacija može utjecati na rad centra za nagradu u mozgu i poremetiti pravilno ponašanje hormona koji reguliraju apetit (hormona gladi i hormona sitosti) te time dovesti do prekomjerne konzumacije hrane. Ipak, zbog profila ispitanika na kojima su istraživanja provedena, kratkog trajanja istraživanja i nedostatka izokalorijske (*engl. isocaloric*¹) prehrane za vrijeme istraživanja, teško je utvrditi je li do povećanja težine došlo zbog konzumiranja *SSB*-a ili drugih faktora u prehrani ispitanika (Sigala & Stanhope, 2021).

¹ prehrana koja se sastoji od jednake količine kalorija iz masti, proteina i ugljikohidrata

Život u malim gradovima i predgrađima (*Living_in_towns_sub*) također pokazuje umjerenu negativnu korelaciju i negativan korelacijski koeficijent s porastom pretilosti, dok je porast postotka stanovništva koje živi u gradu (*Living_in_cities*) povezan s porastom pretilosti.

Okruženje u kojem čovjek živi utječe na njegove odluke, način života i prehrambene navike. Podijelimo li okruženja na ona koja promiču pretilost (*engl. obesogenic*) i ona koja potiču zdrav način života (*eng. leptogenic*) (Swinburn, Egger, & Raza, 1999), veliki gradovi češće su primjer ovih prvih.

Karakteristike velikih gradova često se preklapaju s karakteristikama *obesogenic*² okruženja, a neke od njih su olakšan pristup jeftinoj visokokaloričnoj, ali nutritivno siromašnoj hrani, infrastruktura i urbano planiranje koje otežava aktivno kretanje (hodanje, vožnja bicikla) i bavljenje fizičkim aktivnostima, nedostatak rekreativnih sadržaja i zelenih površina (al, 2023). Stanovnici gradova također su izloženi promociji brze hrane, kojoj često pribjegavaju zbog ubrzanog tempa života. Iako u novijim istraživanjima nije pronađena čvrsta poveznica između samih okruženja i pretilosti, postoje pokazatelji da neke od njihovih karakteristika mogu biti povezane s njom (Minh Lam, Vaartjes, Grobbee, Karssenber, & Lakerveld, 2021).

Iako je i u malim gradovima, rjeđe predgrađima, moguće pronaći neke karakteristike *obesogenic* okruženja, one su prisutne u manjim razmjerima. Količina dostupnih zelenih površina, dostupnost trgovina sa zdravom hranom i bavljenje fizičkim aktivnostima i dalje ovise o urbanom planu grada, ali ipak se pretpostavlja da su stanovnici malih gradova i predgrađa manje izloženi brznoj i nekvalitetnoj hrani, te da im sporiji životni ritam olakšava kultiviranje zdravih navika.

Regresijskom analizom dobivene su vrijednosti R^2 između 0.14 i 0.18, što znači da regresijski model objašnjava samo 14 do 18 posto varijabilnosti postotka pretilosti. Drugim riječima, razlike u stopi pretilosti po državama Europe možemo pripisati utjecaju ovih faktora u ograničenoj mjeri.

Postavlja se pitanje: postoje li međusobne veze između tri istaknuta faktora i drugih faktora koji bi, u teoriji, pridonijeli smanjenju ili porastu pretilosti u državi (npr. pozitivna korelacija između života u velikom gradu i konzumiranja 0 porcija voća i povrća dnevno ili negativna korelacija između života u velikom gradu i bavljenja HE aktivnostima (*engl. health-enhancing activities (HE activities), aktivnosti koje pridonose poboljšanju zdravlja*) više od 150 minuta tjedno)?

Postojanje veza je ispitano Pearsonovim koeficijentom korelacije, stvaranjem korelacijske matrice i regresijskom analizom.

Nisu pronađene statistički značajne korelacije koje bi dodatno opisale utjecaj tri istaknuta faktora na pretilost, pa se pretpostavlja da je za dobivanje cjelokupne slike pretilosti potrebno ispitati dodatne faktore koji nisu dio ove analize.

² Obesogenicity- „skup utjecaja koji okruženje i uvjeti života imaju na promicanje pretilosti kod pojedinca ili populacije“ (Swinburn, Egger, & Raza, 1999).

6.3. Veze između različitih faktora pretilosti

Analizom korelacije između različitih faktora pretilosti ispitano je koje se osobne i prehrambene navike vežu jedna uz drugu.

Korištenjem funkcije *corr()* ispitane su veze između:

- prehrambenih navika i fizičke aktivnosti
- prehrambenih navika i konzumiranja alkohola i cigareta
- fizičkih aktivnosti i konzumiranja alkohola i cigareta
- urbanizacije i prehrambenih navika
- urbanizacije i fizičke aktivnosti
- urbanizacije i konzumiranja alkohola i cigareta.

Rezultati *corr()* funkcije prikazani su u obliku korelacijske matrice, a jeda takva prikazana je na Slici 16.

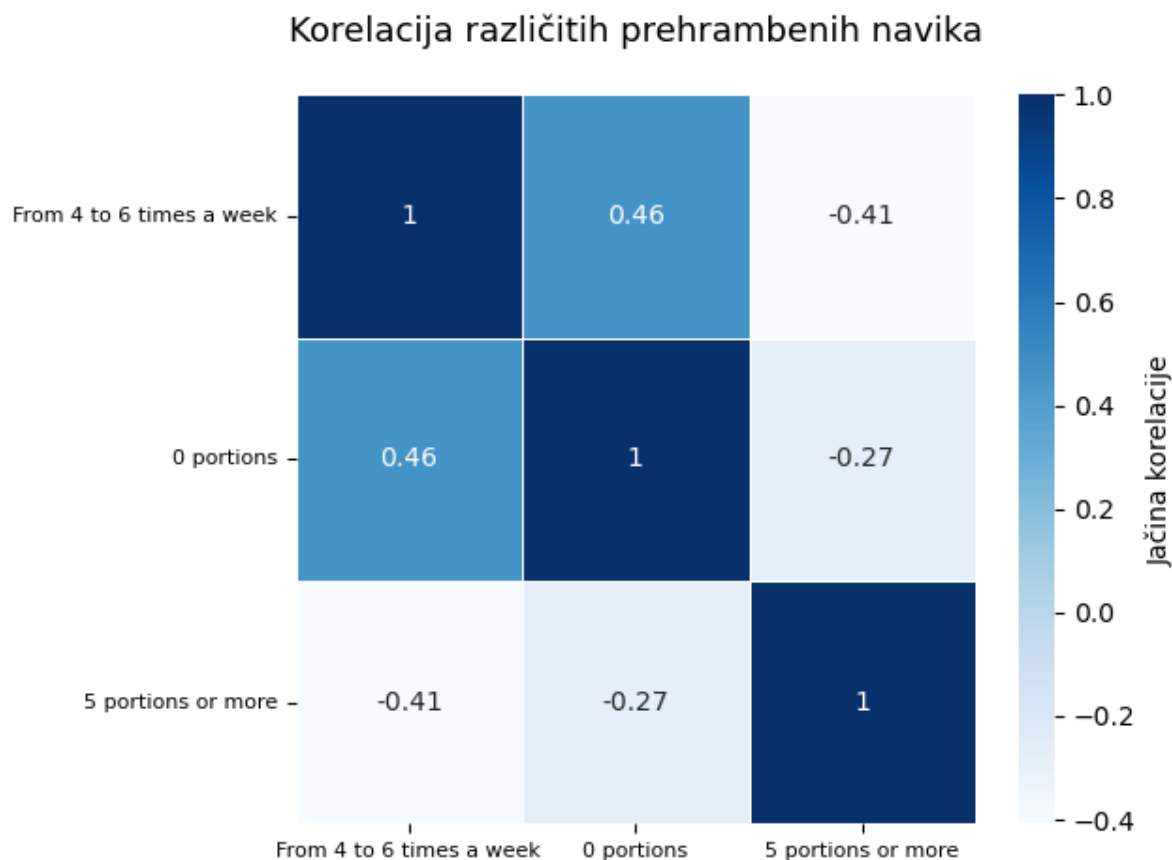
```
correlation_food_exercise=dataset[['0 portions', 'From 1 to 4 portions', '5 portions or more', 'At least once a day', 'From 1 to 3 times a week', 'From 4 to 6 times a week', 'Never or occasionally', 'Walking to get to and from place', 'Cycling to get to and from place', 'Aerobic sports', 'Muscle-strengthening', 'HE_0_MIN', 'HE_1_149', 'HE_150_MIN', 'HE_300_MIN']].corr()
```

	Total smokers	Less than 20 cigarettes per day	20 or more cigarettes per day	Every day	Every week	Every month	Less than once a month
Total smokers	1.000000	0.778253	0.882512	-0.093848	-0.564274	-0.244710	0.007943
Less than 20 cigarettes per day	0.778253	1.000000	0.391596	-0.007807	-0.302741	-0.119998	0.182919
20 or more cigarettes per day	0.882512	0.391596	1.000000	-0.131646	-0.599788	-0.266481	-0.125892
Every day	-0.093848	-0.007807	-0.131646	1.000000	0.367841	-0.479437	-0.562874
Every week	-0.564274	-0.302741	-0.599788	0.367841	1.000000	-0.062916	-0.433257
Every month	-0.244710	-0.119998	-0.266481	-0.479437	-0.062916	1.000000	0.688570
Less than once a month	0.007943	0.182919	-0.125892	-0.562874	-0.433257	0.688570	1.000000
Never or not in the last 12 months	0.591360	0.216738	0.703059	-0.131328	-0.675676	-0.605894	-0.270712
Living in cities	0.007649	0.068393	-0.035594	-0.021905	-0.088498	-0.037308	0.075242
Living_in_towns_sub	-0.246316	-0.278969	-0.155345	0.335084	0.530269	-0.303523	-0.531164
Living_in_rural_areas	0.237978	0.185950	0.207107	-0.308161	-0.418320	0.327571	0.414622

Slika 16: Korelacijska matrica

Uz matricu korelacija dodatno je provedeno ispitivanje Pearsonovog korelacijskog koeficijenta kako bi se ispitala statistička značajnost korelacija.

6.3.1. Različite prehrambene navike



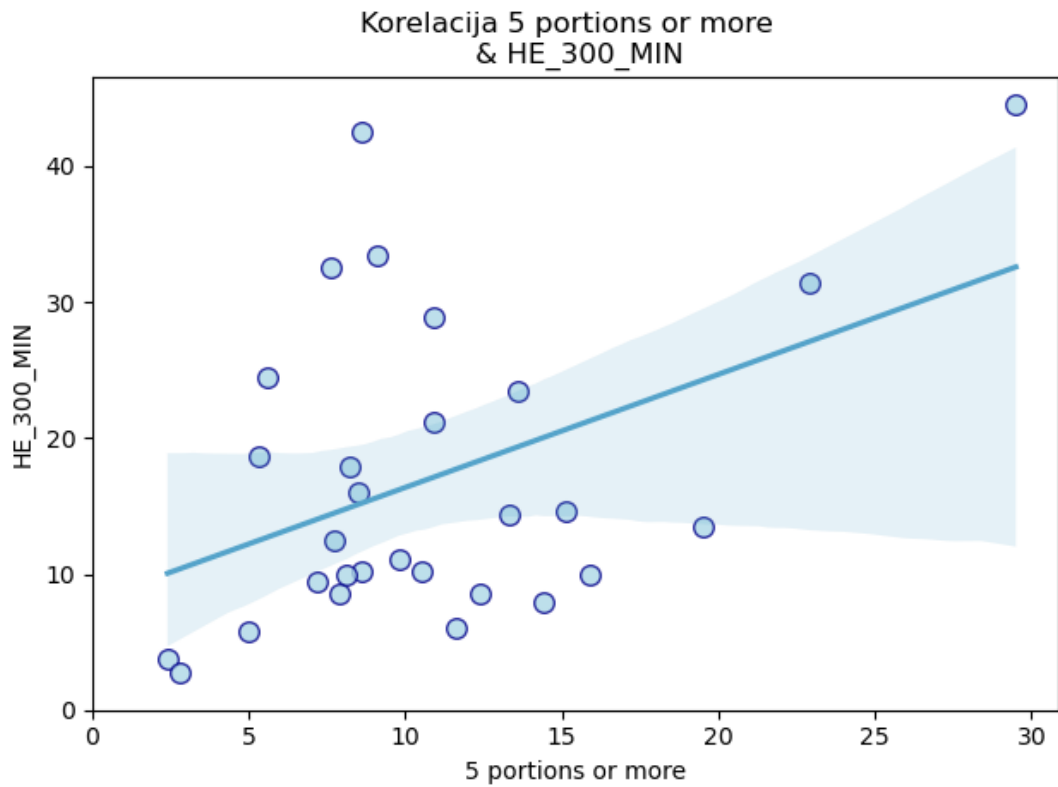
Slika 17: Korelacija različitih prehrambenih navika

Na Slici 17 prikazana je korelacija različitih prehrambenih navika, gdje se vidi da je povećan postotak onih koji konzumiraju 5 ili više porcija voća i povrća dnevno (*5 portions or more*) povezan sa manjim postotkom onih koji konzumiraju šećerom zaslađena pića 4 do 6 puta tjedno (*From 4 to 6 times a week*), dok sa porastom onih koji ne konzumiraju voće i povrće (*0 portions*) raste i broj onih koji konzumiraju SSB 4 do 6 puta tjedno, pa možemo zaključiti da su zdrave navike međusobno povezane, kao i nezdrave.

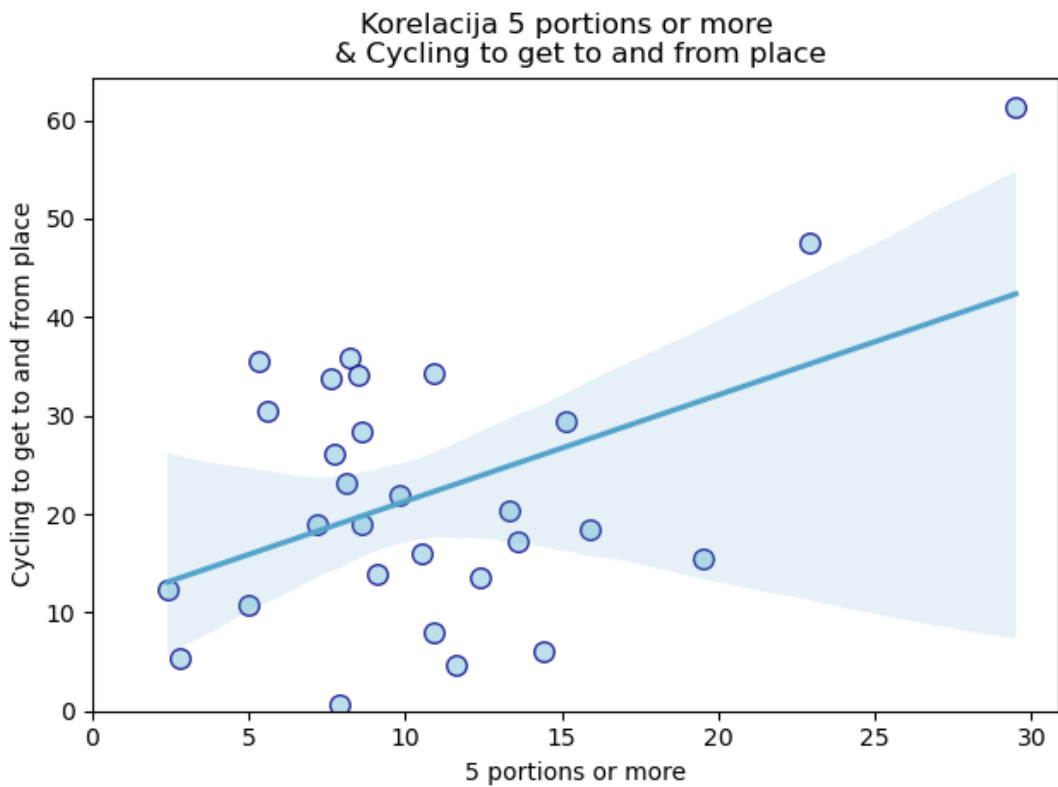
6.3.2. Prehrambene navike i fizička aktivnost

Tablica 4: Korelacija prehrambenih navika i fizičkih aktivnosti

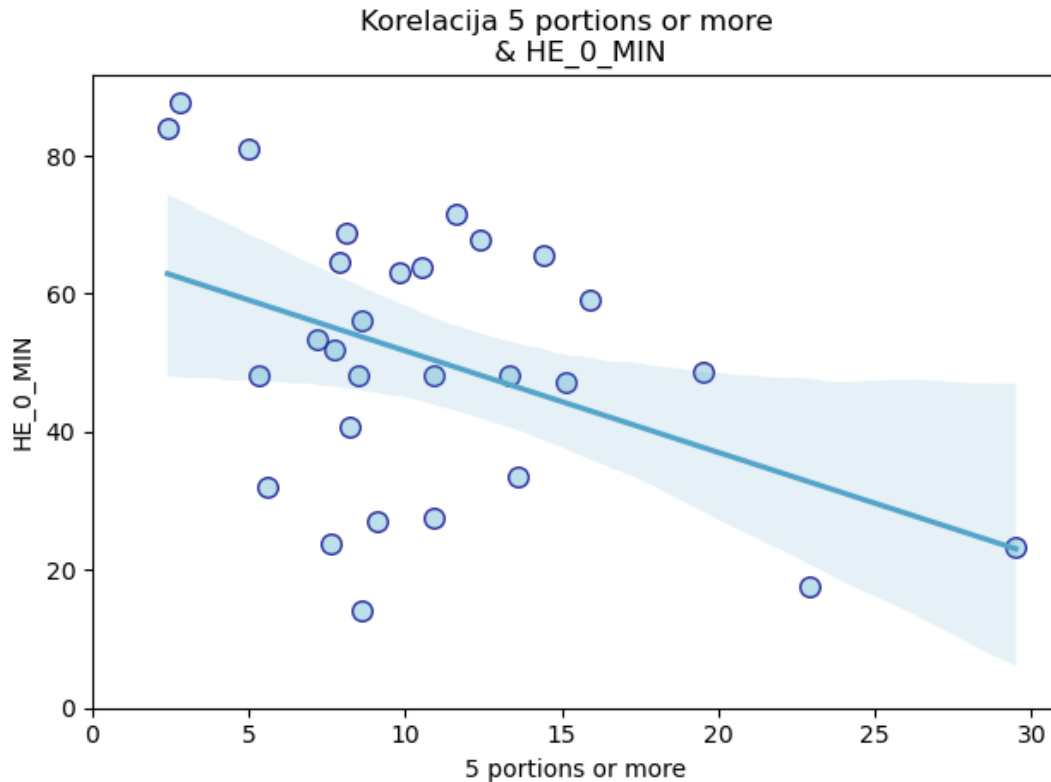
Varijabla 1	Varijabla 2	Koeficijent korelacije	P- vrijednost
5 ili više porcija voća i povrća tjedno (<i>5 portions or more</i>)	Bavljenje aktivnostima koje poboljšavaju zdravlje preko 300 minuta tjedno (<i>HE_300_MIN</i>)	0.43	0.02
5 ili više porcija voća i povrća tjedno (<i>5 portions or more</i>)	Bicikliranje kao način kretanja (<i>Cycling to get to and from place</i>)	0.46	0.01
5 ili više porcija voća i povrća tjedno (<i>5 portions or more</i>)	Bavljenje aktivnostima koje poboljšavaju zdravlje 0 minuta tjedno (<i>HE_0_MIN</i>)	-0.43	0.02



Slika 18: Korelacija konzumacije 5 ili više porcija voća i povrća i bavljenja aktivnostima koje poboljšavaju zdravlje više od 300 minuta tjedno



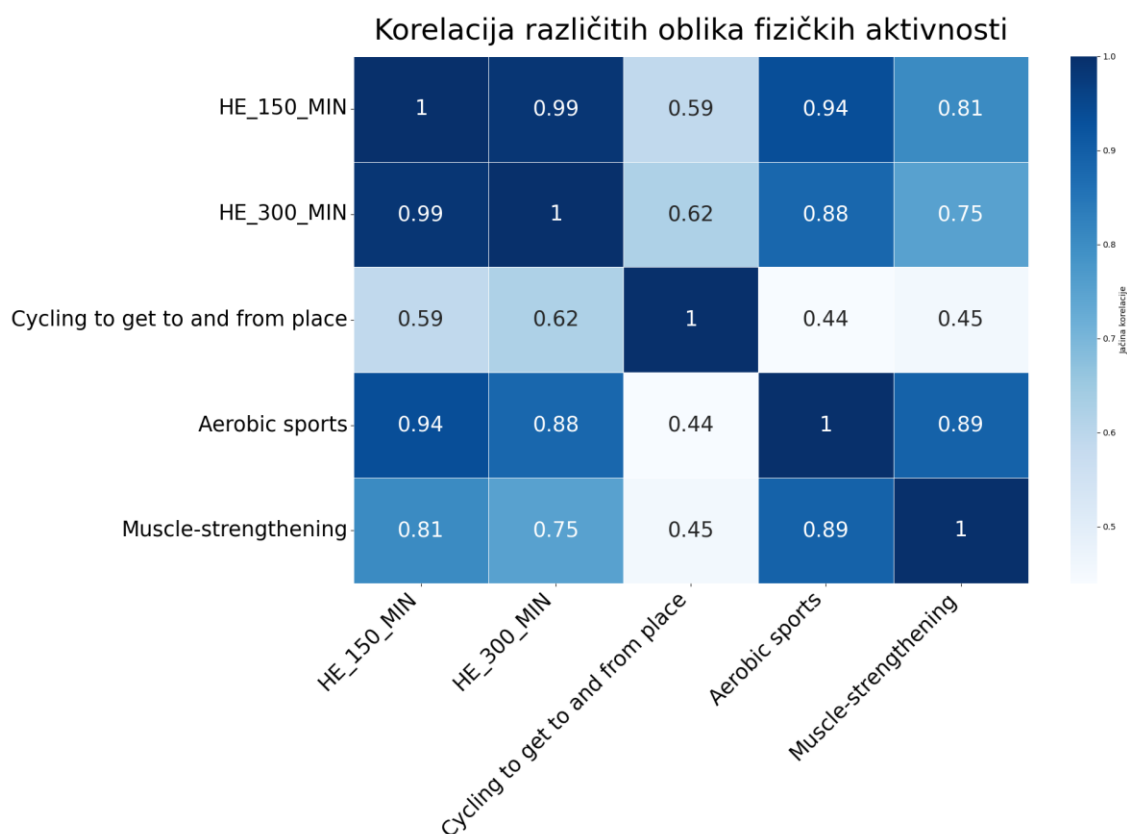
Slika 19: Korelacija konzumacije 5 ili više porcija voća i povrća i bicikliranja kao načina kretanja



Slika 20: Korelacija konzumacije 5 ili više porcija voća i povrća i bavljenja aktivnostima koje poboljšavaju zdravlje 0 minuta tjedno

Osim što se zdrave prehrambene navike vežu jedna na drugu, pokazuju i vezu sa povećanom fizičkom aktivnošću. Iz Tablice 4 i Slika 18, 19 i 20 vidi se da porastom broja ljudi koji konzumiraju 5 ili više porcija voća i povrća raste i broj onih koji se bave bicikliranjem i aktivnostima koje poboljšavaju zdravlje više od 300 minuta tjedno a smanjuje se broj onih koji se takvim aktivnostima uopće ne bave.

6.3.3. Različiti oblici fizičke aktivnosti



Slika 21: Korelacija različitih oblika fizičke aktivnosti

Najveće statistički značajne korelacije u skupu podataka pronađene su među faktorima koji se odnose na fizičku aktivnost. Na Slici 21 možemo vidjeti da je porast broja ljudi koji se bave *HE* aktivnostima preko preporučenih 150 minuta tjedno (*HE_150_MIN*) povezan s porastom onih koji se njima bave duplo, preko 300 minuta tjedno (*HE_300_MIN*).

Države koje imaju najveći postotak stanovnika koji se kreću preko 300 minuta tjedno su Nizozemska sa 44,5% i Norveška sa 42.5%. Odmah nakon njih su Island, Švedska, Danska i Njemačka.

Porastom vremena provedenog u *HE* aktivnostima raste i postotak ljudi koji koriste bicikl kao prijevozno sredstvo, bave se aerobnim sportovima (*Aerobic sports*) i jačanjem mišića (*Muscle-strengthening*).

Jačanje mišića i aerobni sportovi također pokazuju jaku vezu, što može sugerirati da su aktivnosti jednako popularne ili da su ljudi skloni baviti se paralelno s jednim i drugim aktivnostima.

6.3.4. Alkohol i cigarete

Iako ni u prijašnjim istraživanjima ni u ovom nije pronađena direktna poveznica između konzumiranja alkohola i cigareta i razvoja pretilosti, pronađene su neke zanimljive korelacije.

Tako je konzumacija alkohola na tjednoj bazi (*Every week*) u negativnoj korelaciji s pušenjem 20 ili više cigareta dnevno (-0.56), dok postotak ljudi koji ne konzumiraju alkohol ili ga nisu

konzumirali u posljednjih 12 mjeseci pokazuje jaku korelaciju s pušenjem 20 ili više cigareta dnevno (0.70).

6.3.5. Urbanizacija

Porast postotka stanovništva koje živi u velikim gradovima povezan je s padom postotka onih kojima je hodanje preferirani način kretanja (*Walking to get to and from place*), ali i s porastom postotka onih koji radije bicikliraju (*Cycling to get to and from place*).

Postotak ljudi koji žive u manjim gradovima (*Living_in_towns_sub*) pokazuje pozitivnu vezu s postotkom onih koji koriste bicikl kao način kretanja i konzumiraju alkohol na tjednoj bazi, a postotak stanovništva koje živi u ruralnim područjima (*Living_in_rural_areas*) proporcionalan je s postotkom stanovništva kojem je hodanje glavni način kretanja, onima koji konzumiraju alkohol rjeđe od jednom mjesečno (*Less than once a month*) i onima koji svakodnevno konzumiraju šećerom zaslađena pića (*At least once a day*).

6.4. Usporedba faktora pretilosti i načina života u Regijama Europe

Nakon podjele država po regijama, svaka je regija obuhvatila između 5 i 8 država. S obzirom na mali broj podataka za svaku regiju, provođenje korelacije i regresijske analize nije bilo preporučljivo zbog mogućnosti dobivanja rezultata koji nisu vjerodostojni. Zato su za analizu podataka korištene metode deskriptivne statistike uz komparativnu analizu dobivenih rezultata.

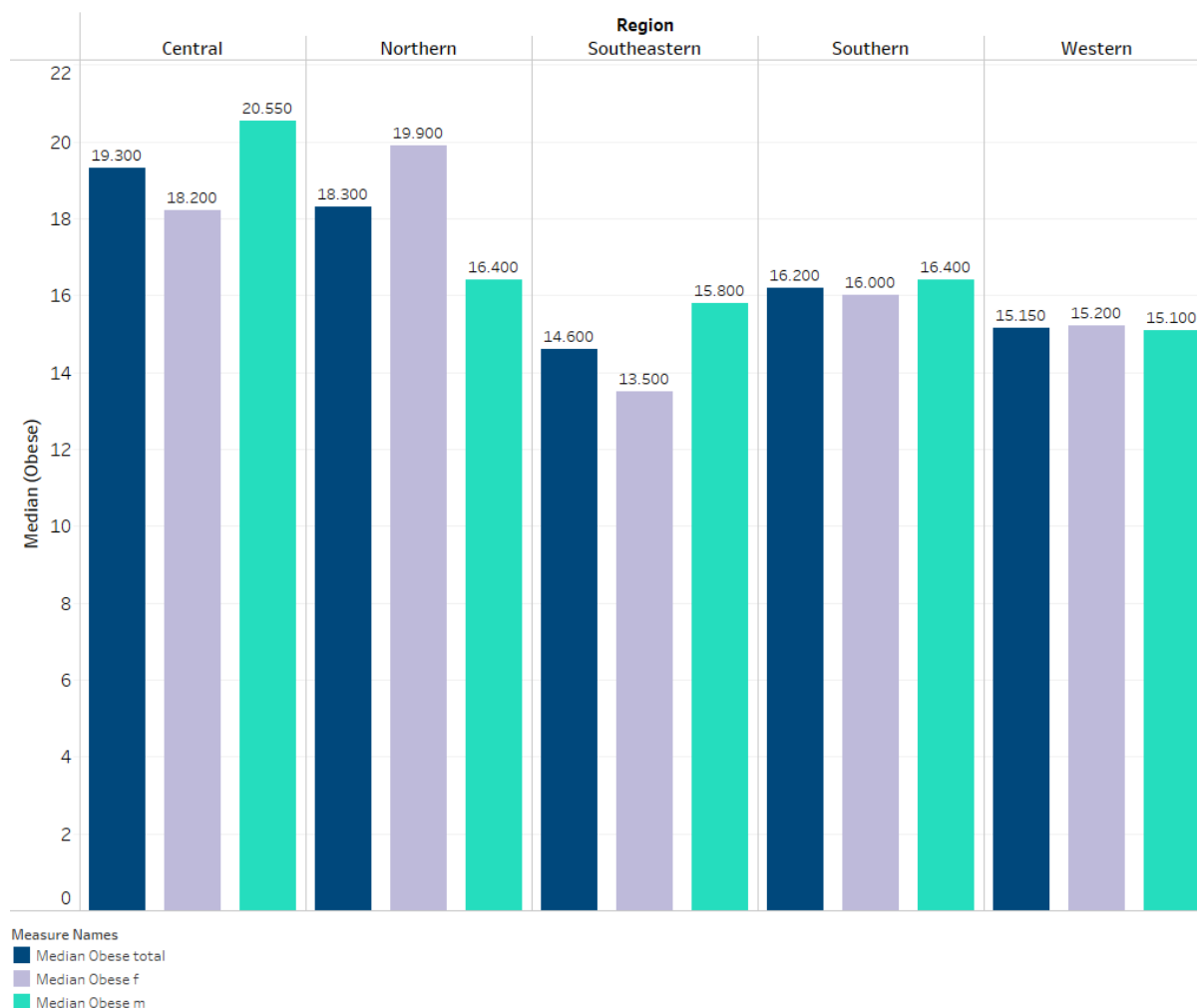
Podaci su grupirani po regijama i ispitan je medijan svakog faktora. Kao rezultat je dobiven podatkovni okvir na Slici 22.

```
region_medians = dataset.groupby('Region').median()  
region_medians
```

	Population_total	Population_male	Population_female	Underweight_total	Normal_total	Obese_total	Underweight_
Region							
Central	9315765.5	4516427.0	4799338.5	2.40	40.30	19.30	0.9
Northern	2794184.0	1295591.0	1498593.0	2.30	42.70	18.30	1.4
Southeastern	7000039.0	3395701.0	3604338.0	2.40	44.10	14.60	1.7
Southern	10724599.0	5208293.0	5516306.0	2.00	43.40	16.20	1.3
Western	14368841.0	7112956.0	7255885.0	3.45	49.05	15.15	2.0

Slika 22: Medijan pojedinih faktora po regijama

Pretilost po regijama

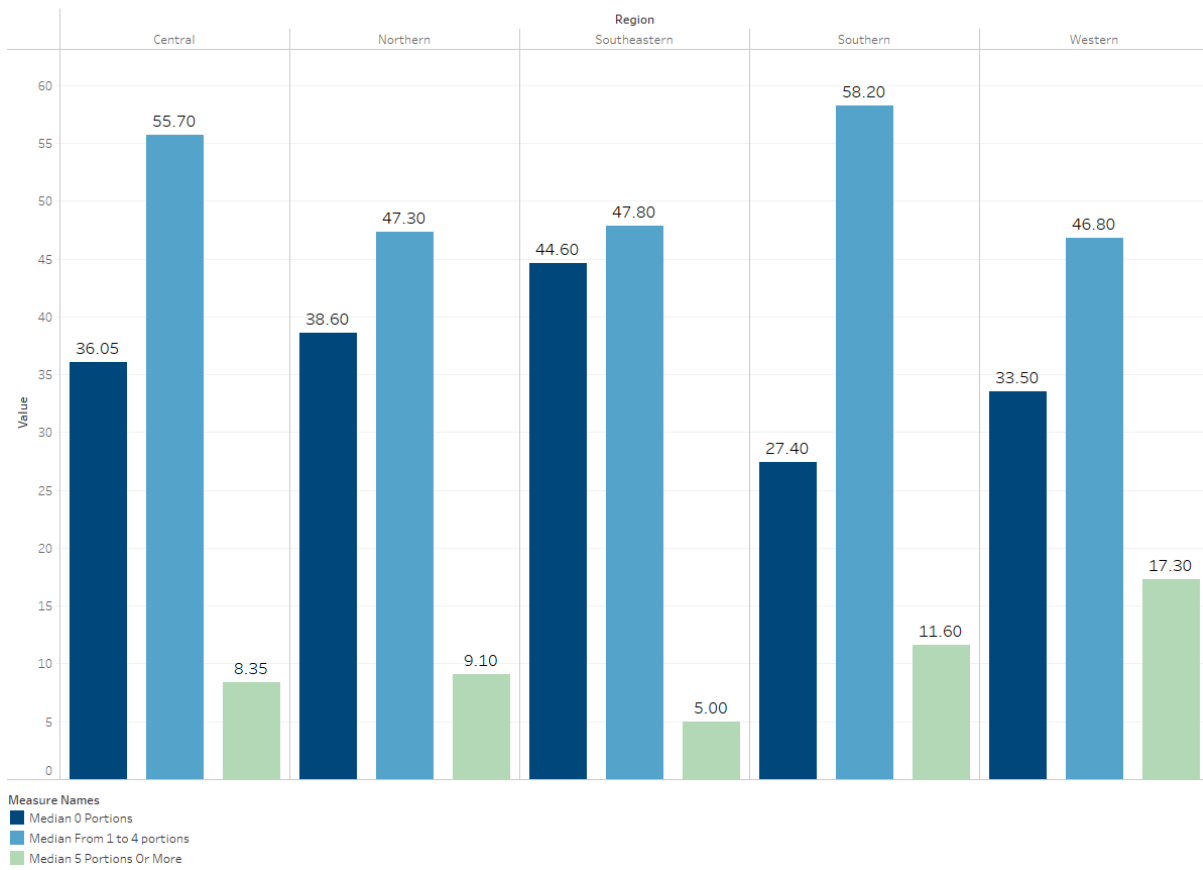


Slika 23: Pretilost po regijama

Na Slici 23 prikazan je postotak pretilosti ukupnog stanovništva (*Median Obese total*), žena (*Median Obese f*) i muškaraca (*Median Obese m*) za sve regije. Najveći postotak ukupnog pretilog stanovništva ima Srednja Europa, a najmanji Jugoistočna. Što se tiče razlike među spolovima, broj pretilih muškaraca i žena je gotovo izjednačen u Južnoj i Zapadnoj Europi, u Srednjoj i Jugoistočnoj je više pretilih muškaraca, dok u Sjevernoj u pretilosti prednjače žene.

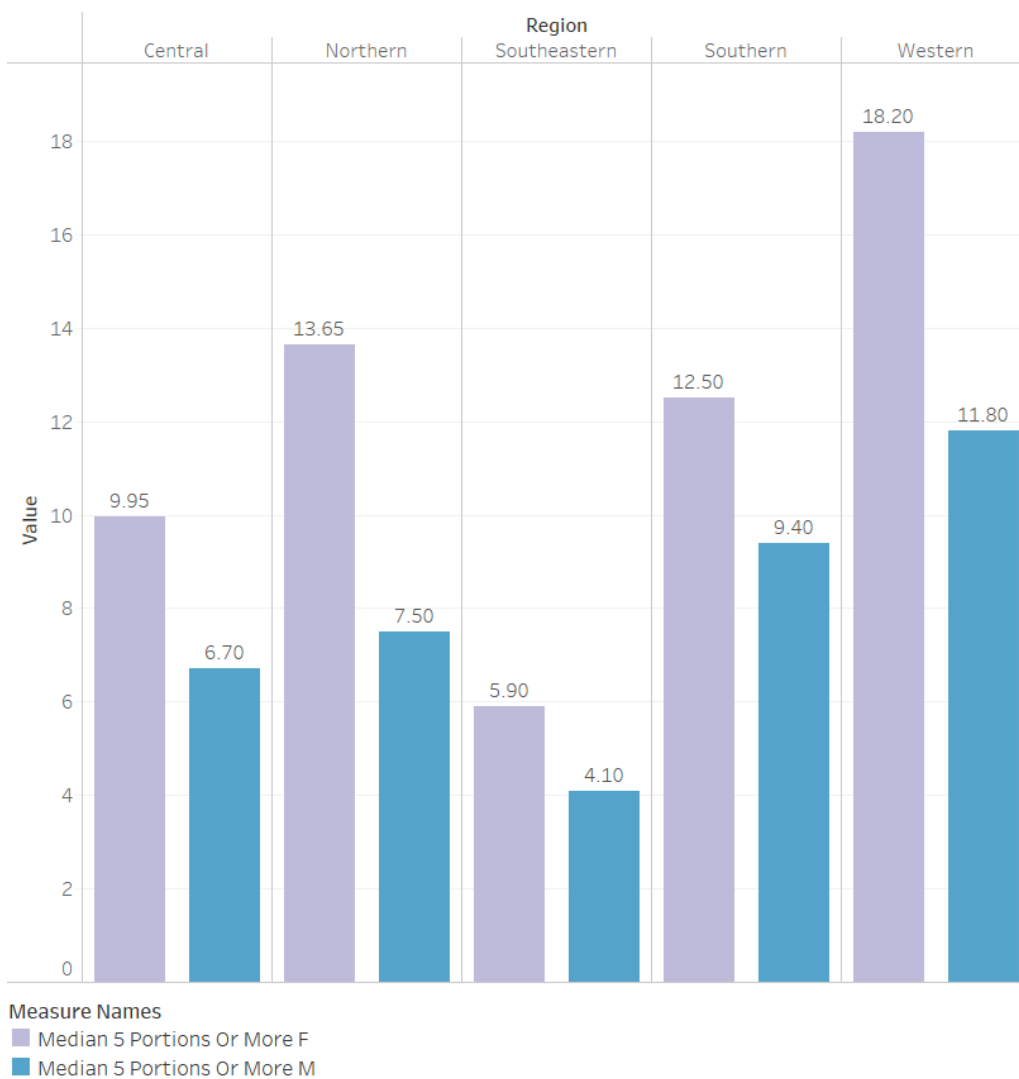
6.4.1. Prehrambene navike

Konzumacija voća i povrća po regijama



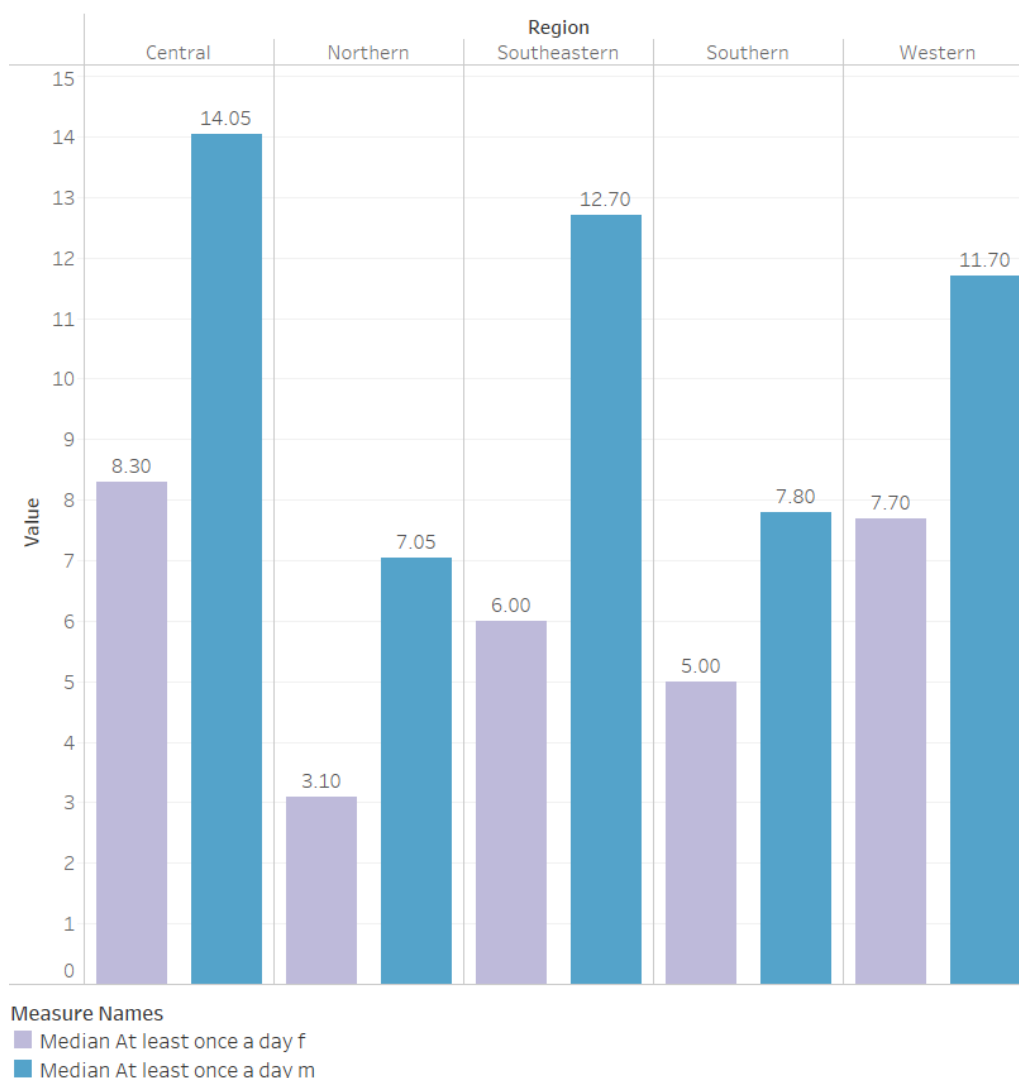
Slika 24: Konzumacija voća i povrća po regijama

Konzumacija 5 ili više porcija voća i povrća dnevno



Slika 25: Konzumacija 5 ili više porcija voća i povrća po regijama i po spolu

Konzumacija SSB-a barem jednom dnevno

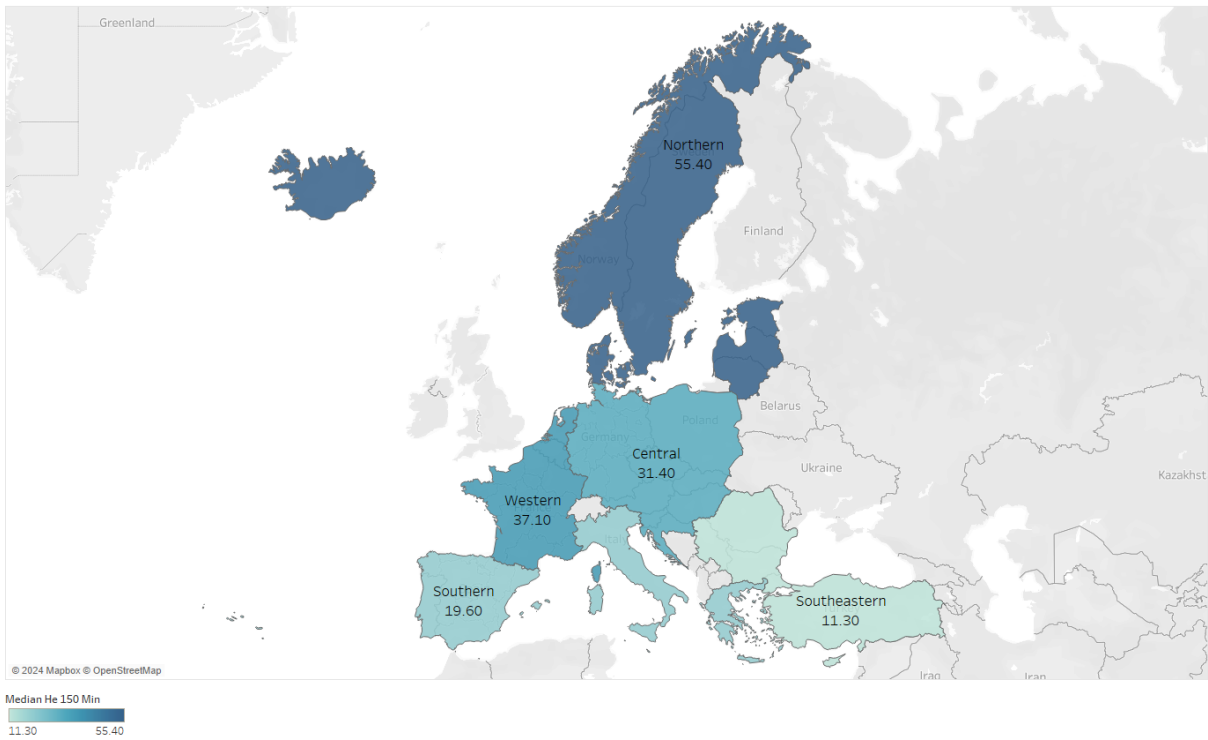


Slika 26: Konzumacija šećerom zaslađenih pića po regijama i po spolu

Pogledamo li sliku 24 možemo vidjeti da se u svim regijama najčešće konzumiraju 1 do 4 porcije voća i povrća dnevno (*From 1 to 4 portions*). Usporedimo li prehrambene navike po spolu i pogledamo dijagrame na Slikama 25 i 26, vidimo da je u svim regijama postotak žena koje konzumiraju preporučenu dozu voća i povrća (*5 portions or more*) veći od postotka muškaraca, dok je kod konzumiranje šećerom zaslađenih pića (*At least once a day*) situacija obrnuta.

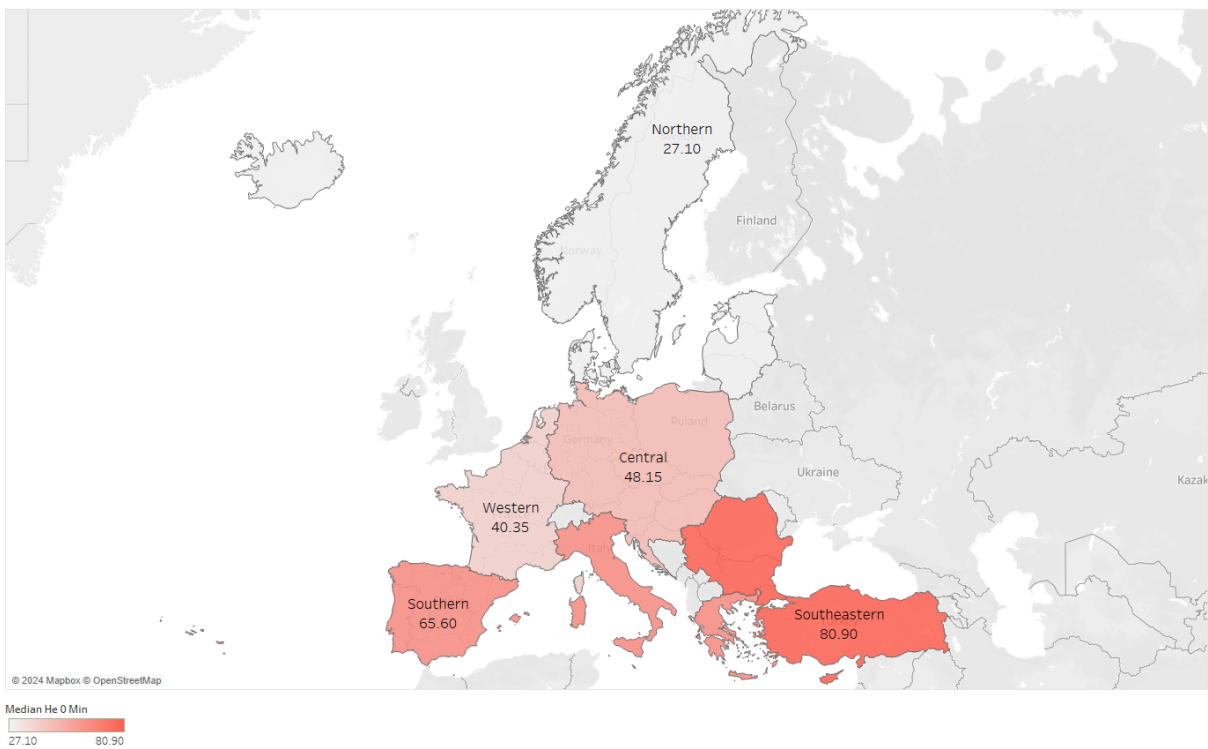
6.4.2. Fizička aktivnost

Bavljenje HE aktivnostima preko 150 minuta tjedno



Slika 27: Bavljenje HE aktivnostima preko 150 minuta tjedno, po regijama

Bavljenje HE aktivnostima 0 minuta tjedno



Slika 28: Bavljenje HE aktivnostima 0 minuta tjedno, po regijama

U svim regijama uočen je veliki postotak ljudi koji se od jednog mjesta do drugog kreću pješice (72,2% do 92,4% za žene i 73,8% do 93,8% za muškarce), ali se količina vremena provedena u bavljenju HE aktivnostima i vrsta aktivnosti kojom se bave razlikuju od regije do regije.

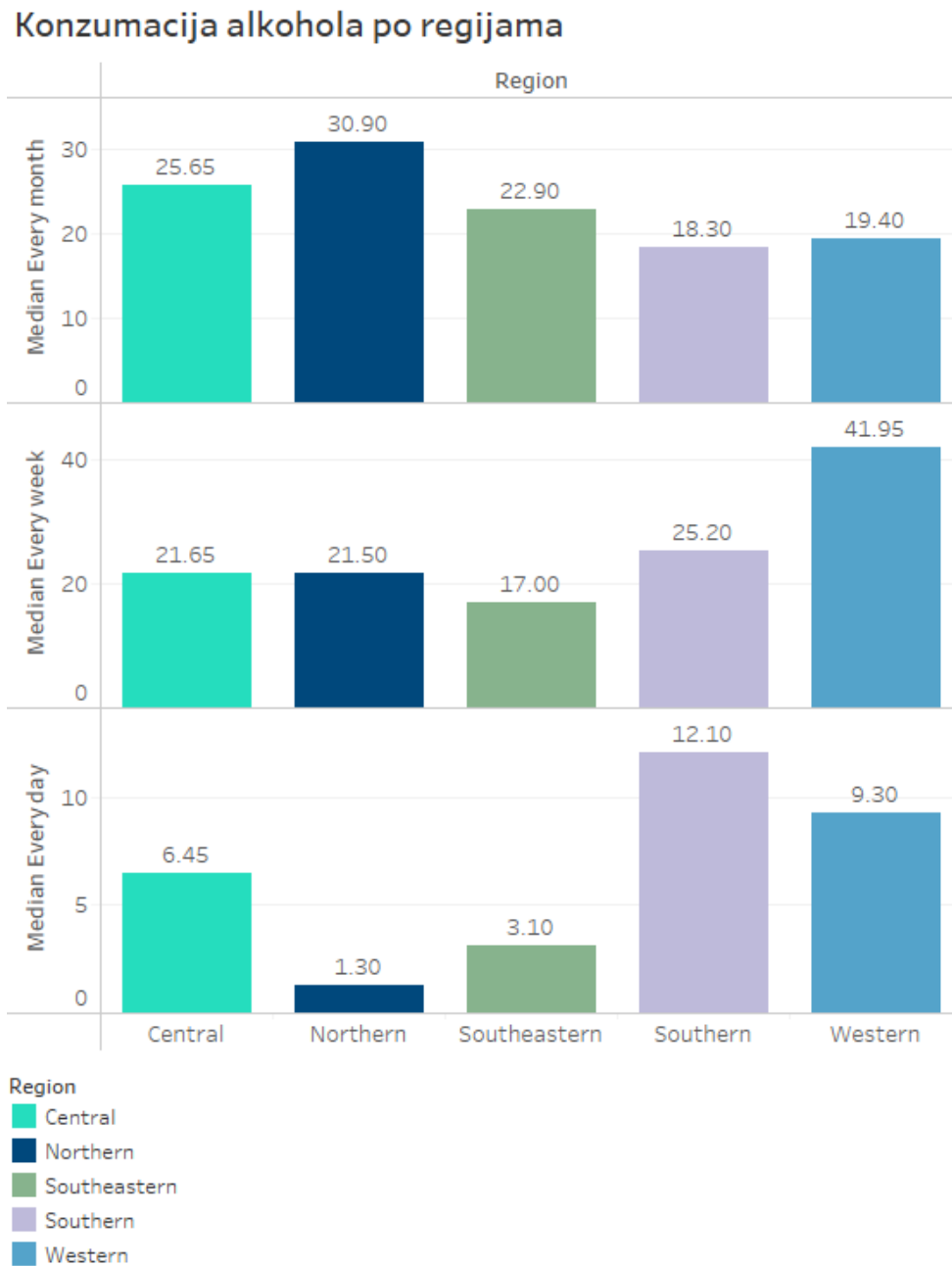
Na slici 27 vidi se da je Sjeverna Europa je najaktivnija, s više od 50% stanovništva koje se HE aktivnostima bavi barem 150 minuta (*HE_150_MIN*). Dodatno, 30% stanovništva bavi se njima barem 300 minuta tjedno. Također se ističe i u postotcima ljudi koji se bave aerobnim sportovima (71.20%) i jačanjem mišića (46.55%).

U Srednjoj Europi, u kojoj je ustanovljena najveća stopa pretilosti, skoro pola ukupnog stanovništva ne posvećuje niti malo vremena HE aktivnostima (*HE_0_MIN*), a 31,4% se njima bavi 150 ili više minuta tjedno. Oni koji se bave HE aktivnostima radije odabiru aerobne sportove nego jačanje mišića ili vožnju biciklom kao način kretanja.

Najveći postotak ljudi koji se ne bave HE aktivnostima (*HE_0_MIN*) je u Jugoistočnoj Europi gdje on iznosi čak 80,9%, što možemo vidjeti na Slici 28. Ovaj podatak zvuči zabrinjavajuće, i pomalo zbunjujuće uzmemo li u obzir da jugoistočna Europa ima najmanju stopu pretilosti.

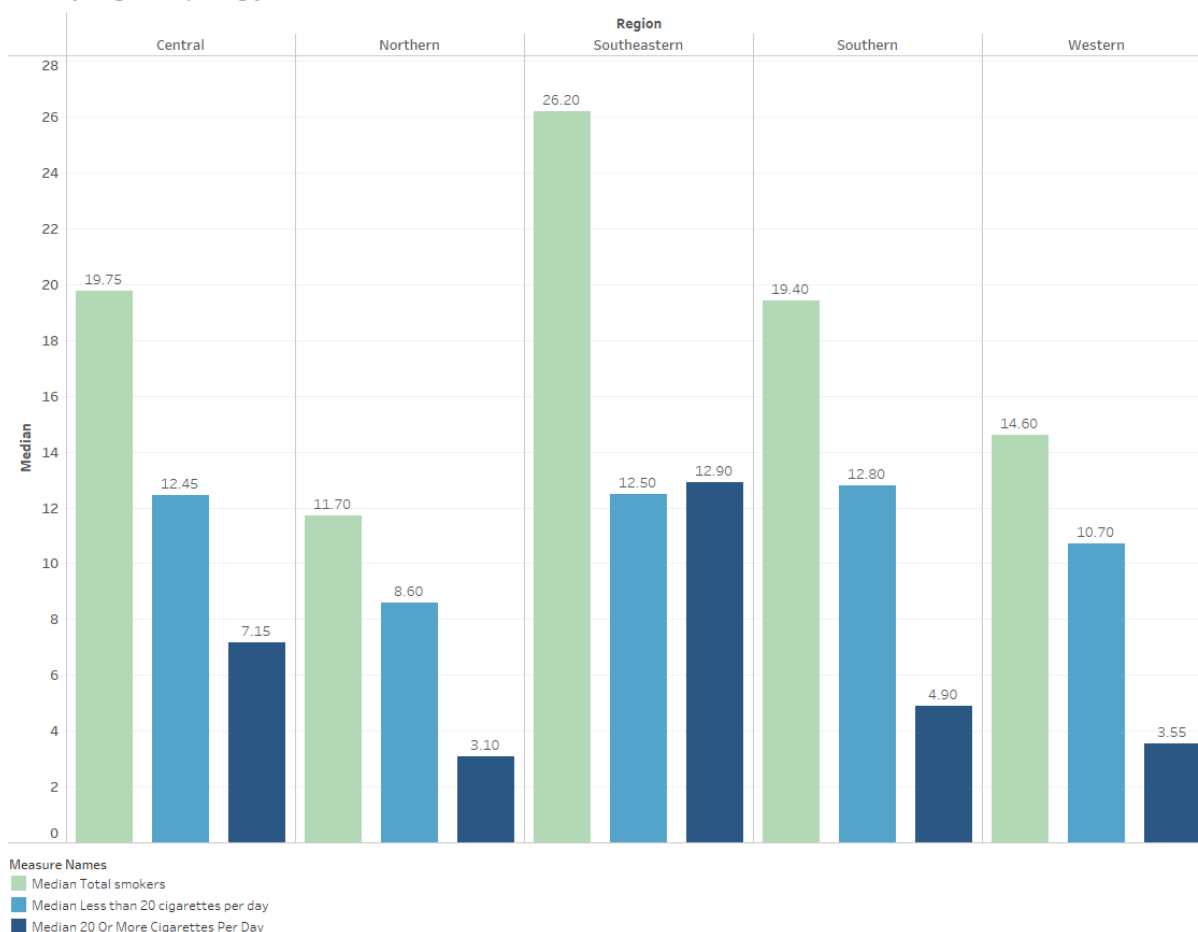
Muškarci su uglavnom fizički aktivniji od žena, osim u Sjevernoj Europi gdje je (iako ne znatno) postotak žena koje se bave HE aktivnostima barem 150 minuta tjedno veći od onoga za muškarce. Ipak kada se vrijeme posvećeno HE aktivnostima poveća na 300 minuta, muškarci su ponovo u prednosti.

6.4.3. Alkohol i cigarete



Slika 29: Konzumacija alkohola po regijama

Pušenje cigareta po regijama



Slika 30: Pušenje cigareta po regijama

Na Slici 29 vidi se da učestalost konzumacije alkohola varira po regijama. U Sjevernoj Europi najveći broj ljudi konzumira alkohol na mjesečnoj bazi (*Every month*), u Srednjoj na tjednoj (*Every week*), dok je u Južnoj Europi je najveći broj onih koji ga konzumiraju svakodnevno (*Every day*). Kako bi bolje razumjeli ove rezultate, potrebno je sjetiti se da ove navike ovise o kulturološkim razlikama i društvenim normama. Također ovise i o vrsti preferiranih pića za svaku regiju. Ako je vjerovati članku u The Guardianu (Guardian, n.d.) u Sjevernoj Europi najčešće se piju žestoka pića, u Srednjoj pivo, a u Južnoj vino. Uzmemo li u obzir preferirana pića i učestalost konzumacije alkohola možemo je kategorizirati kao konzumaciju žestokih pića u posebnim prigodama u Sjevernoj Europi, konzumaciju piva u društvu (vikendom ili nakon posla) u Srednjoj Europi i svakodnevnu konzumaciju vina uz obrok u Južnoj Europi, iako iz trenutno dostupnih izvora nije moguće potvrditi tu pretpostavku.

Iz grafa na Slici 30 vidi se da najveći ukupni postotak pušača cigareta (*Total smokers*) ima Jugoistočna Europa, a najmanji Sjeverna. Osim što je postotak pušača u Jugoistočnoj Europi znatno veći od onog u ostalim regijama, također je i jedini kojeg većinski čine pušači koji konzumiraju više od 20 cigareta tj. jedne kutije dnevno (*20 or More Cigarettes Per Day*). U ostalim regijama većinu ukupnog broja pušača čine oni koji puše manje od kutije cigareta dnevno (*Less than 20 Cigarettes Per Day*), što možemo, uz nedostatak znanstvene potvrde,

opisati kao pušenje u društvu ili uz piće. Bitno je naglasiti da manji postotak pušača ne garantira zdravije životne navike jer nam nedostaju podaci o alternativnim načinima konzumacije nikotina.

6.4.4. Urbanizacija

Što se tiče urbanizacije, najviše stanovnika u ruralnim krajevima (*Living_in_rural_areas*) ima Srednja Europa, u malim gradovima i predgrađima (*Living_in_towns_sub*) zapadna Europa, dok u Sjevernoj, Jugoistočnoj i Istočnoj najveći broj ljudi živi u velikim gradovima (*Living_in_cities*).

7. Zaključak

Analiza podataka o europskim državama pružila je uvid u odnos između prehrambenih i životnih navika i pretilosti. Metodama statističke analize ispitane su veze između različitih faktora i pretilosti stanovništva u pojedinim državama i regijama u svrhu boljeg razumijevanja ovog kompleksnog problema.

Rezultati analize pokazuju da najveću negativnu korelaciju s pretilosti imaju konzumacija 1 do 3 šećerom zaslađena pića tjedno i život u manjim gradovima i predgrađima, dok najveću pozitivnu ima život u velikim gradovima. Isti rezultati vrijede za oba spola. Nisu pronađene veze između ovih faktora i drugih koji bi objasnili utjecaj na pretilost, ali je daljnjim ispitivanjem veza između pojedinih faktora uočeno da se, između ostalog, zdrave prehrambene navike pojavljuju zajedno te da zdrave prehrambene navike pokazuju i pozitivnu vezu sa povećanim bavljenjem fizičkim aktivnostima.

Najveću stopu pretilosti ima Srednja Europa, a najmanju Jugoistočna. Prehrambene navike slične su u svim dijelovima Europe te se u svim regijama žene hrane zdravije nego muškarci. Navike vezane za fizičku aktivnost i konzumaciju alkohola i cigareta variraju ovisno o regiji. Dobiveni su i neki naizgled proturječni rezultati, primjerice to da Jugoistočna Europa kao regija sa najmanjom stopom pretilosti ima najviši postotak ljudi koji se u slobodno vrijeme ne bave nikakvom fizičkom aktivnošću. Ovakvi rezultati ilustriraju složenost faktora koji utječu na pretilost i ukazuju na potrebu za uključivanjem dodatnih faktora u analizu. Neki od njih mogli bi biti prihodi, zaposlenost i radni uvjeti, dob ispitanika, razina obrazovanja i kulturološke značajke.

Ovaj rad nudi okvirni pregled načina života u Europi i utjecaja različitih faktora na pretilost i naglašava potrebu za dodatnim istraživanjima kako bi se stekla cjelovita slika problema pretilosti u Europi.

8. Popis literature

- [1] M. Blüher, »Obesity: global epidemiology and pathogenesis,« *Nature Reviews Endocrinology*, pp. 288-298, 2019.
- [2] J. Krzysztozek, I. Laudańska-Krzemińska i M. Bronikowski, »Assessment of epidemiological obesity among adults in EU countries. « *Annals of Agricultural and Environmental Medicine*, p. 341–349, 2019.
- [3] G. Moschonis i G. L. Trakman, »Overweight and Obesity: The Interplay of Eating Habits and Physical Activity,« *Nutrients*, 2023.
- [4] F. Åberg i M. Färkkilä, »Drinking and Obesity: Alcoholic Liver Disease/Nonalcoholic Fatty Liver Disease Interactions ,« *Seminars in Liver Disease*, pp. 154-162, 2020.
- [5] A. Gümüşa, S. Kayhan i S. Kayhana, »The relationship between cigarette smoking and obesity,« *Journal of Experimental & Clinical Medicine*, 2013.
- [6] Eurostat, Population on 1 January by age group and sex, https://ec.europa.eu/eurostat/databrowser/view/demo_pjangroup/default/table?lang=en , 2019. [Pristupljeno 16 Rujan 2024].
- [7] Eurostat, Distribution of population by degree of urbanisation, dwelling type and income group - EU-SILC survey, https://ec.europa.eu/eurostat/databrowser/view/ilc_lvho01/default/table?lang=en&category=degurb.degurb_livcon , 2019. [Pristupljeno 16 Rujan 2024].
- [8] Eurostat, Body mass index (BMI) by sex, age and income quintile, https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_bm1i/default/table?lang=en&category=hlth.hlth_det.hlth_bmi , 2019. [Pristupljeno 16 Rujan 2024].
- [9] Eurostat, Daily consumption of fruit and vegetables by sex, age and income quintile, https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_fv3i/default/table?lang=en&category=hlth.hlth_det.hlth_cfv, 2019. [Pristupljeno 16 Rujan 2024].
- [10] Eurostat, Frequency of drinking sugar-sweetened soft drinks by sex, age and income quintile, https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_fv7i/default/table?lang=en&category=hlth.hlth_det.hlth_cfv, 2019. [Pristupljeno 16 Rujan 2024].
- [11] Eurostat: Frequency of alcohol consumption by sex, age and income quintile, https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_alli/default/table?lang=en&category=hlth.hlth_det.hlth_alc, 2019. [Pristupljeno 16 Rujan 2024].
- [12] Eurostat, Daily smokers of cigarettes by sex, age and income quintile, https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_sk3i/default/table?lang=en&category=hlth.hlth_det.hlth_smok, 2019. [Pristupljeno 16 Rujan 2024].
- [13] Eurostat, Performing (non-work-related) physical activities by sex, age and income quintile,

- https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_pe3i/default/table?lang=en&category=hlth.hlth_det.hlth_pha, 2019. [Pristupljeno 16 Rujan 2024].
- [14] Eurostat, Time spent on health-enhancing (non-work-related) aerobic physical activity by sex, age and income quintile, https://ec.europa.eu/eurostat/databrowser/view/hlth_ehis_pe2i/default/table?lang=en&category=hlth.hlth_det.hlth_pha, 2019. [Pristupljeno 16 Rujan 2024].
- [15] IZZI: Europske regije, <https://hr.izzi.digital/DOS/401324/410020.html> [Pristupljeno 1 Lipanj 2024].
- [16] M. & S. H. S. T. Assistance, Introduction to Data Analysis Handbook, AED/TAC, Washington, 2006.
- [17] R. E. Walpole, R. H. Myers, S. L. Myers i K. Ye, Probability & Statistics for Engineers & Scientists, Prentice Hall, 2011.
- [18] M. Benšić i N. Šuvak, Primijenjena statistika. Sveučilište J.J. Strossmayera, Odjel za matematiku, Osijek, 2013.
- [19] »statistics — Mathematical statistics functions,« [Mrežno]. Available: <https://docs.python.org/3/library/statistics.html>. [Pristupljeno 16 Rujan 2024].
- [20] D. M. Sigala i K. L. Stanhope, »An Exploration of the Role of Sugar-Sweetened Beverage in Promoting Obesity and Health Disparities,« *Current Obesity Reports*, pp. 39-52, 2021.
- [21] B. Swinburn, G. Egger i F. Raza, »Dissecting obesogenic environments: the development and application of a framework for identifying and prioritizing environmental interventions for obesity,« *Preventive Medicine*, pp. 563-570, 1999.
- [22] S. V. e. al, »Upstream Determinants of Overweight and Obesity in Europe,« *Current Obesity Reports*, pp. 417-428, Kolovoz 2023.
- [23] T. Minh Lam, I. Vaartjes, D. E. Grobbee, D. Karssenberga i J. Lakerveld, »Associations between the built environment and obesity: an umbrella review,« *International Journal of Health Geographics volume*, 2021.
- [24] T. Guardian, »The Guardian,«, <https://www.theguardian.com/society/article/2024/jun/25/wine-beer-or-spirits-europeans-cant-kick-their-traditional-drinking-habits>, 2024. [Pristupljeno 25 8 2024].

9. Popis slika

Slika 1: Prvi radni list .xlsx datoteke	5
Slika 2: Drugi radni list .xlsx datoteke.....	6
Slika 3: Treći radni list .xlsx datoteke	7
Slika 4: Podatkovni okvir nakon učitavanja i preimenovanja stupaca	8
Slika 5: Podatkovni okvir nakon uklanjanja suvišnih redaka	8
Slika 6: Skup podataka nakon dodavanja stupca "Region"	10
Slika 7: Računanje medijana regije za nedostajuće vrijednosti	10
Slika 8: Finalni skup podataka za analizu.....	12
Slika 9: Box-plot.....	14
Slika 10: Pretilost po državama -Total.....	16
Slika 11: Pretilost po državama -Žene	17
Slika 12: Pretilost po državama -Muškarci	18
Slika 13: : Korelacija konzumacije SSB-a 1 do 3 puta tjedno i pretilosti.....	22
Slika 14: : Korelacija života u manjim gradovima i predgrađima i pretilosti.....	22
Slika 15: Korelacija života u velikim gradovima i pretilosti	23
Slika 16: Korelacijska matrica	25
Slika 17: Korelacija različitih prehrambenih navika	26
Slika 18: Korelacija konzumacije 5 ili više porcija voća i povrća i bavljenja aktivnostima koje poboljšavaju zdravlje više od 300 minuta tjedno.....	28
Slika 19: Korelacija konzumacije 5 ili više porcija voća i povrća i bicikliranja kao načina kretanja.....	28
Slika 20: Korelacija konzumacije 5 ili više porcija voća i povrća i bavljenja aktivnostima koje poboljšavaju zdravlje 0 minuta tjedno	29
Slika 21: Korelacija različitih oblika fizičke aktivnosti.....	30
Slika 22: Medijan pojedinih faktora po regijama	32
Slika 23: Pretilost po regijama.....	33
Slika 24: Konzumacija voća i povrća po regijama	34
Slika 25: Konzumacija 5 ili više porcija voća i povrća po regijama i po spolu.....	35
Slika 26: Konzumacija šećerom zaslađenih pića po regijama i po spolu	36
Slika 27: Bavljenje HE aktivnostima preko 150 minuta tjedno, po regijama.....	37
Slika 28: Bavljenje HE aktivnostima 0 minuta tjedno, po regijama.....	37
Slika 29: Konzumacija alkohola po regijama	39
Slika 30: Pušenje cigareta po regijama	40

10. Popis tablica

Tablica 1: Box-plot vrijednosti	15
Tablica 2: Korelacija faktora pretilosti i BMI za ukupno stanovništvo.....	20
Tablica 3: : Regresija faktora pretilosti i BMI za ukupno stanovništvo	21
Tablica 4: Korelacija prehrambenih navika i fizičkih aktivnosti.....	27