

Uloga analize i vizualizacije podataka u prevenciji kardiovaskularnih bolesti u medicini

Alija, Helena

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:195:954878>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

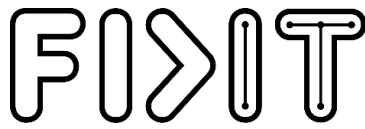
Download date / Datum preuzimanja: **2024-09-13**



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

Helena Alija

Uloga analize i vizualizacije podataka u prevenciji kardiovaskularnih bolesti u medicini

Završni rad

Mentor: doc.dr.sc. Lucia Načinović Prskalo

Rijeka, svibanj 2024.

Rijeka, 13.11.2023.

Zadatak za završni rad

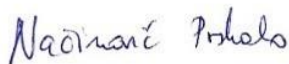
Pristupnik/ica: Helena Alija

Naziv završnog rada: Uloga analize i vizualizacije podataka u prevenciji kardiovaskularnih bolesti u medicini

Naziv završnog rada na engleskom jeziku: The role of data analysis and visualization in the prevention of cardiovascular diseases in medicine

Sadržaj zadatka: Zadatak završnog rada je istražiti ulogu analize i vizualizacije podataka u prevenciji kardiovaskularnih bolesti, s fokusom na upotrebu informatičkih alata u medicinskim istraživanjima. U radu će se na temelju različitih skupova medicinskih podataka izraditi vizualizacije u svrhu analize uzročnika i faktora rizika oboljenja od kardiovaskularnih bolesti, analizu trendova bolesti i slično. Opisat će se izrada svake vizualizacije te mogući zaključci na koje analiza i vizualizacija skupa podataka upućuju.

Mentorica
Doc. dr. sc. Lucia Načinović Prskalo



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



Zadatak preuzet: 13.11.2023.


(potpis pristupnika/ice)

Sažetak

Tema ovoga rada istražuje utjecaj i doprinos analize i vizualizacije podataka u prevenciji kardiovaskularnih bolesti, s fokusom na upotrebu informatičkih alata u medicinskim istraživanjima. Pregledava se veza između informatike i medicine, posebno u kontekstu kardiovaskularnih bolesti, ističući značaj vizualizacije podataka u analizi uzročnika i faktora rizika. Rad koristi razne metodologije i alate za vizualizaciju trendova bolesti, analize faktora rizika te odnos između dobi, tjelesne mase i krvnog tlaka. Identificira ključne obrasce iz grafikona, naglašavajući važnost zaključaka izvedenih iz podataka za prevenciju budućih patoloških stanja. Zaključno, rad potvrđuje doprinos u analize i vizualizacije podataka u medicini na način da primjećuje obrasce koji omogućuju rano otkrivanje rizičnih čimbenika za razvitak bolesti, preciznije dijagnostičke metode, optimizaciju terapijskih postupaka, poboljšanje javnozdravstvenih strategija i unapređenje istraživačkih inicijativa.

Ključne riječi

Analiza podataka; vizualizacija podataka; kardiovaskularne bolesti; medicinska informatika; Tableau; faktori rizika; statistička analiza; dashboardi; trendovi incidencije; heatmap analiza; scatter plot; strojno učenje; personalizirana medicina; prevencija

Sadržaj

1. Uvod u analizu i vizualizaciju podataka u medicini	1
2. Kardiovaskularne bolesti	2
2.1. Što su kardiovaskularne bolesti (KVB)?	2
2.2 Faktori rizika za KVB	2
3. Metodologija analize podataka	3
3.1. Upotreba informatičkih alata za analizu medicinskih podataka.....	3
3.2. Korištenje Tableau-a za vizualizaciju podataka.....	3
3.3. Proces vizualizacije podataka u istraživanju kardiovaskularnih bolesti.....	4
3.3.1. Prikupljanje i priprema podataka	4
3.3.2. Analiza podataka	5
3.3.3. Vizualizacija podataka	5
4. Odabir podataka za analizu i vizualizaciju	6
5. Grafički prikazi podataka o kardiovaskularnim bolestima	8
5.1. Trendovi u incidenciji kardiovaskularnih bolesti kroz vremenske serije.....	8
5.1.1. Incidencija KDV-a tijekom posljednjih deset godina u svijetu.....	8
5.1.2. Incidencija KDV-a rangirana po zemljama svijeta.....	10
5.1.3. Zemlje s najvišom incidencijom KDV-a.....	12
5.1.4. Razlike u incidenciji KDV-a između Europe i Azije.....	14
5.1.5. Smrtnost uzrokovana kardiovaskularnim bolestima.....	15
5.1.6. Raspodjela incidencije KDV-a po godinama	18
5.2. Indikatori kardiovaskularnih bolesti.....	20
5.2.1. Proporcija općeg zdravlja	20
5.2.2. Broj dana lošeg fizičkog zdravlja po općem zdravlju	22
5.2.3. Proporcija općeg zdravlja po spolu	24
5.2.4. 3D prikaz odnosa između BMI-a, broja sati sna i broja dana lošeg fizičkog zdravlja	26
5.2.5. Odnosi između različitih indikatora kardiovaskularnih bolesti	27
5.2.6. Korelacija između indikatora kardiovaskularnih bolesti	29
5.2.7. Prikaz korelacije sa više indikatora kardiovaskularnih bolesti	31
5.2.8. Povezanost bolesti srca s drugim bolestima i faktorima	32
5.2.9. Slučajevi bolesti kardiovaskularnog sustava po spolu i statusu pušenja	35
5.2.10. Trend slučajeva kardiovaskularnih bolesti po starosnim kategorijama	37
5.2.11. Broj slučajeva kardiovaskularnih bolesti po dobnim skupinama i spolu.....	39

5.2.12. Prisutne poteškoće kod kardiovaskularnih bolesnika	42
5.2.13. Povezanost cjevica protiv gripe sa srčanim udarom	43
5.3. Mogućnosti dijagnostike, prevencije i liječenja kardiovaskularnih bolesti – vizualizacija u Tableau	44
5.3.1. Udio uzroka smrti od kardiovaskularnih bolesti u ukupnom broju smrti u svijetu	44
5.3.2. Broj kardiologa po zemljama u svijetu – Tableau vizualizacija.....	45
5.3.3. Broj kardiovaskularnih kirurga na milijun ljudi po državama	47
5.3.4. Koronarna angiografija i analiza stentova na milijun ljudi	49
5.3.5. Broj koronarnih stentova na milijun ljudi.....	50
5.3.6. Ugrađeni pacemakeri na milijun stanovnika	52
7. Zaključak	54
8. Popis literature	55
9. Popis slika	57

1. Uvod u analizu i vizualizaciju podataka u medicini

Analiza podataka uključuje korištenje statističkih i matematičkih metoda za istraživanje, tumačenje i zaključivanje o podacima. Proces analize omogućuje korisnicima da identificiraju trendove, uzorke i anomalije unutar podataka, čime se olakšava razumijevanje kompleksnih skupova informacija i donošenje temeljitih odluka [1].

S druge strane, vizualizacija podataka transformira rezultate analize u grafičke prikaze poput grafikona, dijagrama, mapa topline i infografika. Vizualizacije pružaju intuitivan način za prezentaciju i razumijevanje složenih podataka, omogućujući korisnicima da brzo uoče ključne informacije, trendove i odnose bez potrebe za detaljnom analizom osnovnih podataka [2].

U kontekstu eksponencijalnog rasta razvoja tehnologije i dostupnosti podataka, ove metode postaju sve važnije za efikasno upravljanje i strategijsko planiranje, ali također imaju sve veću važnost u medicinskoj znanosti za prevenciju i suzbijanje primarnih faktora morbiditeta i mortaliteta na globalnoj razini [3]. Integracija naprednih informatičkih tehnologija u istraživačke protokole ne samo da omogućuje obradu obimnih podatkovnih setova, već i potencira efikasnu vizualizaciju, omogućujući tako dublje razumijevanje kompleksnih patoloških stanja poput kardiovaskularnih i neuroloških poremećaja. Ovaj pristup ne samo da unapređuje preventivne strategije, već i značajno doprinosi optimizaciji terapijskih intervencija.

Cilj ovog rada je pokazati kako informatičke discipline, koristeći napredne analitičke metode i vizualizacije, mogu značajno poboljšati medicinska istraživanja. Takav pristup omogućava bolje razumijevanje složenih skupova podataka i olakšava izdvajanje ključnih informacija važnih za razumijevanje uzroka i razvoja kardiovaskularnih bolesti. Rad se fokusira na prikaz potencijala različitih alata i metoda za analizu medicinskih podataka, potičući razvoj novih preventivnih i terapijskih pristupa.

2. Kardiovaskularne bolesti

2.1. Što su kardiovaskularne bolesti (KVB)?

Kardiovaskularne bolesti (KVB) predstavljaju grupu bolesti srca i krvnih žila, uključujući koronarnu bolest srca, cerebrovaskularne bolesti, reumatske bolesti srca, kao i druga stanja koja utječu na srčani i krvožilni sustav. Ove bolesti su vodeći uzrok smrti globalno, odgovorne za više od 17 milijuna smrti godišnje prema Svjetskoj zdravstvenoj organizaciji (WHO). Razumijevanje što su KVB, koji su faktori rizika i kako se mogu spriječiti ključno je za smanjenje njihove prevalencije.

2.2 Faktori rizika za KVB

Kardiovaskularne bolesti su kompleksna stanja na čiji razvoj utiče kombinacija genetskih, bioloških i okolišnih faktore te, načina života. Genetski faktori igraju značajnu ulogu jer obiteljska anamneza bolesti srca može povećati rizik [4]. Starost i spol također utječu; muškarci su generalno izloženiji većem riziku u mlađoj dobi, dok se rizik za žene povećava nakon menopauze.

Načini života poput pušenja, nezdrave prehrane, nedostatka tjelesne aktivnosti i prekomjernog konzumiranja alkohola značajno povećavaju rizik od kardiovaskularnih bolesti. Pušenje direktno oštećuje krvne žile i srce, dok nezdrava prehrana može dovesti do povišenog kolesterola, visokog krvnog tlaka (hipertenzije) i dijabetesa. Nedostatak tjelesne aktivnosti doprinosi prekomjernoj tjelesnoj težini i smanjenju kardiovaskularne kondicije. Prekomjerno konzumiranje alkohola može uzrokovati hipertenziju, srčane aritmije i druge probleme sa srcem.

Osim toga, okolišni faktori, kao što je zagađenje zraka, mogu dodatno povećati rizik od kardiovaskularnih bolesti. Zagađenje zraka, posebno čestice poput PM2.5, može uzrokovati upalu i oštećenje krvnih žila, što doprinosi razvoju ateroskleroze i drugih kardiovaskularnih stanja [5].

Razumijevanje i prepoznavanje ovih faktora rizika ključni su za implementaciju preventivnih mjera koje mogu smanjiti incidenciju i težinu kardiovaskularnih bolesti.

3. Metodologija analize podataka

3.1. Upotreba informatičkih alata za analizu medicinskih podataka

U modernoj medicini, upotreba informatičkih alata postala je neizostavan dio analize medicinskih podataka. Ovi alati omogućuju istraživačima i zdravstvenim radnicima da obrade velike količine podataka efikasno i precizno. Python je jedan od najpopularnijih jezika za analizu podataka, zahvaljujući svojim bogatim bibliotekama kao što su Pandas, NumPy, SciPy i Seaborn [6]. Ove biblioteke omogućuju naprednu obradu i analizu podataka, uključujući statističku analizu, manipulaciju podacima i matematičke operacije.

Pandas pruža jednostavne metode za rad s velikim skupovima podataka, omogućujući filtriranje, grupiranje i agregaciju podataka. NumPy je osnova za mnoge druge biblioteke u Pythonu i omogućava efikasne numeričke operacije. SciPy proširuje mogućnosti Pythona za napredne matematičke, znanstvene i inženjerske funkcije. Seaborn, s druge strane, specijaliziran je za statističku vizualizaciju i pruža alatke za lako kreiranje složenih grafika i dijagrama koji pomažu u razumijevanju podataka.

Seaborn se često koristi u kombinaciji s Pandasom za kreiranje vizualizacija koje ističu ključne statističke odnose u podacima. Na primjer, Seaborn može biti korišten za kreiranje boxplotova, violin plotova, heatmapa i drugih vizualizacija koje pomažu u identifikaciji obrazaca i distribucija unutar podataka. Ovi alati omogućuju istraživačima da dobiju dublje uvide u medicinske podatke i identificiraju važne trendove i korelacije.

Korištenje ovih alata omogućuje detaljnu analizu podataka iz različitih izvora, kao što su elektronički zdravstveni zapisi (EHR), genetski podaci i podaci s mobilnih zdravstvenih aplikacija. Integracijom i analizom ovih podataka, istraživači mogu otkriti obrasce i korelacije koje mogu biti ključne za rano otkrivanje i prevenciju kardiovaskularnih bolesti.

3.2. Korištenje Tableau-a za vizualizaciju podataka

Tableau predstavlja neprocjenjiv alat za analizu i vizualizaciju podataka u istraživanju kardiovaskularnih bolesti, omogućujući istraživačima i zdravstvenim radnicima da efikasno upravljaju i interpretiraju kompleksne medicinske podatke kako bi unaprijedili prevenciju, dijagnostiku i liječenje ovih bolesti [7].

Jedan je od najmoćnijih alata za vizualizaciju podataka koji se koristi u medicinskim istraživanjima, posebno u kontekstu analize i vizualizacije podataka vezanih za kardiovaskularne bolesti. Njegova sposobnost da stvori interaktivne i intuitivne vizualizacije čini ga idealnim za analizu složenih medicinskih podataka koji se odnose na kardiovaskularno zdravlje.

U istraživanju kardiovaskularnih bolesti, Tableau omogućava istraživačima da kreiraju dinamičke grafike, karte i druge vizualne prikaze koji olakšavaju dublje razumijevanje podataka. Primjerice, epidemiološki podaci mogu se vizualizirati kroz interaktivne mape koje prikazuju regionalne razlike u učestalosti kardiovaskularnih bolesti. Ovo omogućuje

identificiranje područja s većim rizikom, što je ključno za usmjeravanje preventivnih mjera i resursa.

Tableau također olakšava analizu trendova u incidenciji i prevalenciji kardiovaskularnih bolesti tokom vremena. Vizualizacije koje prikazuju promjene u broju slučajeva bolesti kroz godine mogu pomoći istraživačima da prepoznaju uzorke i promjene u epidemiologiji bolesti. Na primjer, analizom podataka o incidenciji infarkta miokarda ili moždanog udara, istraživači mogu uočiti sezonske varijacije ili trendove povezane s promjenama u društvenim i okolišnim faktorima.

Rezultati kliničkih studija također se mogu efikasno prikazati putem Tableau-a. Na primjer, vizualizacije koje prikazuju usporedne stope preživljavanja kod pacijenata koji su primili različite vrste terapija mogu pružiti važne uvide za optimizaciju liječenja.

Korištenje Tableau-a za vizualizaciju podataka u kardiovaskularnim istraživanjima ne samo da pomaže u identificiranju ključnih trendova i obrazaca, već također olakšava komunikaciju tih nalaza sa širom publikom, uključujući medicinske stručnjake i kreatore politika [7] Interaktivne vizualizacije omogućuju dublje istraživanje podataka, omogućujući korisnicima da filtriraju i prilagode prikaze prema specifičnim interesima i potrebama.

3.3. Proces vizualizacije podataka u istraživanju kardiovaskularnih bolesti

Proces vizualizacije podataka u istraživanju kardiovaskularnih bolesti započinje prikupljanjem i pripremom podataka. Ovaj inicijalni korak je ključan za osiguranje točnosti i konzistentnosti podataka, koji su neophodni za validne analize i interpretacije. Prikupljanje podataka može uključivati različite izvore, poput elektroničkih zdravstvenih zapisa (EHR), podataka iz kliničkih studija, epidemioloških baza podataka i rezultata laboratorijskih testova.

3.3.1. Prikupljanje i priprema podataka

Podaci prikupljeni iz različitih izvora često dolaze u nehomogenim formatima i mogu sadržavati pogreške ili nedostatke. Stoga je potrebno provesti postupak čišćenja podataka, koji uključuje uklanjanje ili ispravljanje netočnih zapisa, popunjavanje nedostajućih vrijednosti i standardizaciju varijabli kako bi se osigurala konzistentnost. Standardizacija podrazumijeva prilagođavanje podataka tako da svi zapisi budu u istim jedinicama i formatima, čime se olakšava njihova kasnija analiza.

3.3.2. Analiza podataka

Nakon što su podaci očišćeni i standardizirani, pristupa se analizi podataka s ciljem identifikacije ključnih pokazatelja i trendova. U ovom koraku se koriste različiti statistički i analitički alati kako bi se razumjeli obrasci i odnosi unutar podataka. Python, sa svojim bogatim ekosustavom biblioteka kao što su Pandas, NumPy i SciPy, omogućava naprednu obradu podataka i izvođenje kompleksnih statističkih analiza. Pandas olakšava manipulaciju i analizu podataka, dok NumPy omogućava efikasno izvođenje numeričkih operacija, a SciPy pruža dodatne funkcionalnosti za znanstvenu i inženjersku analizu.

3.3.3. Vizualizacija podataka

Nakon analize, podaci se transformiraju u vizualne prikaze koristeći alate poput Matplotliba i Seaborna, pri čemu se kreiraju raznovrsni grafovi i dijagrami. Time se istraživačima omogućava intuitivna vizualizacija podataka.

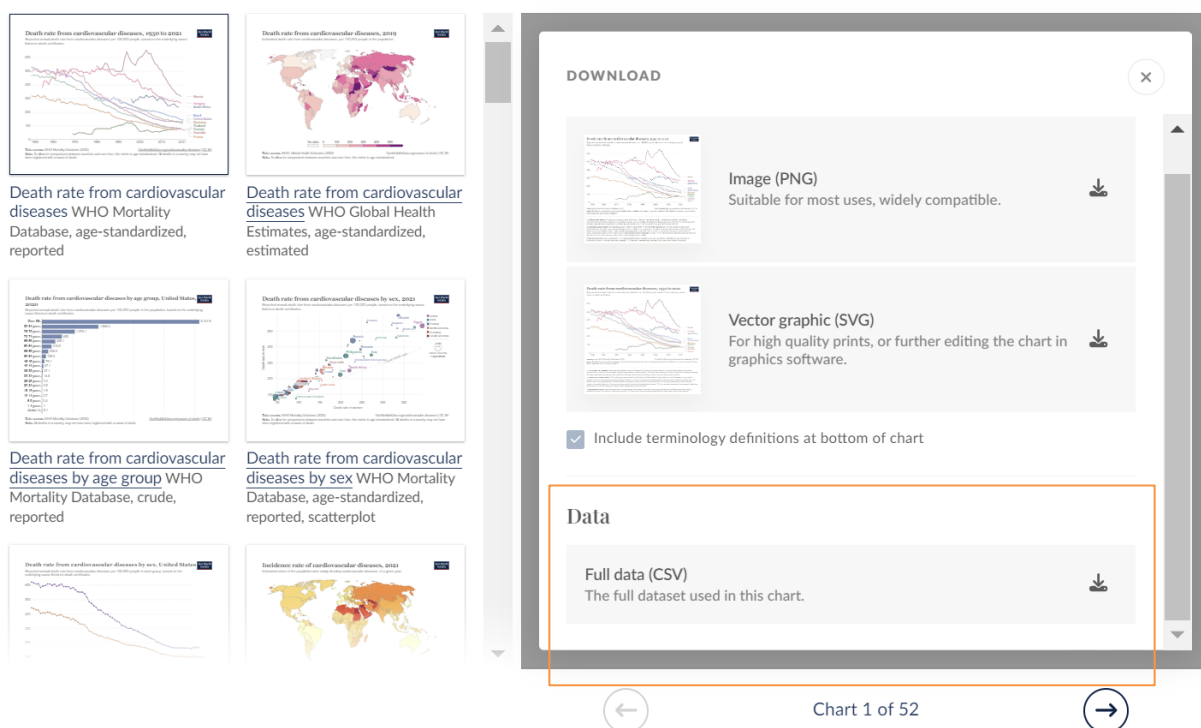
4. Odabir podataka za analizu i vizualizaciju

Za analizu i vizualizaciju podataka na ovu temu, u radu se koristi nekoliko relevantnih izvora s raznovrsnim podacima i uvidima u kardiovaskularne bolesti. Svaki izvor za skup podataka nudi specifične informacije koje su ključne za analizu i vizualizaciju rezultata.

Skupovi podataka preuzeti su sa sljedećih izvora:

- *'Our World in Data: Cardiovascular Diseases'* je sveobuhvatni izvor sa skupom podataka o prevalenciji, uzrocima i posljedicama kardiovaskularnih bolesti diljem svijeta. Podaci uključuju informacije o stopi mortaliteta, faktorima rizika i trendovima incidencije, a prikazani su u obliku interaktivnih grafova kao što je prikazano u nastavku (Slika 1). U uzradi završnog rada nisu korišteni interaktivni grafovi, već CSV datoteka koja se krije iza njih i koja u tabličnom obliku pohranjuje sve potrebite informacije za izradu analize i vizualizacije u Pythonu ili Tableau [8].

Interactive Charts on Cardiovascular Diseases [↗](#)

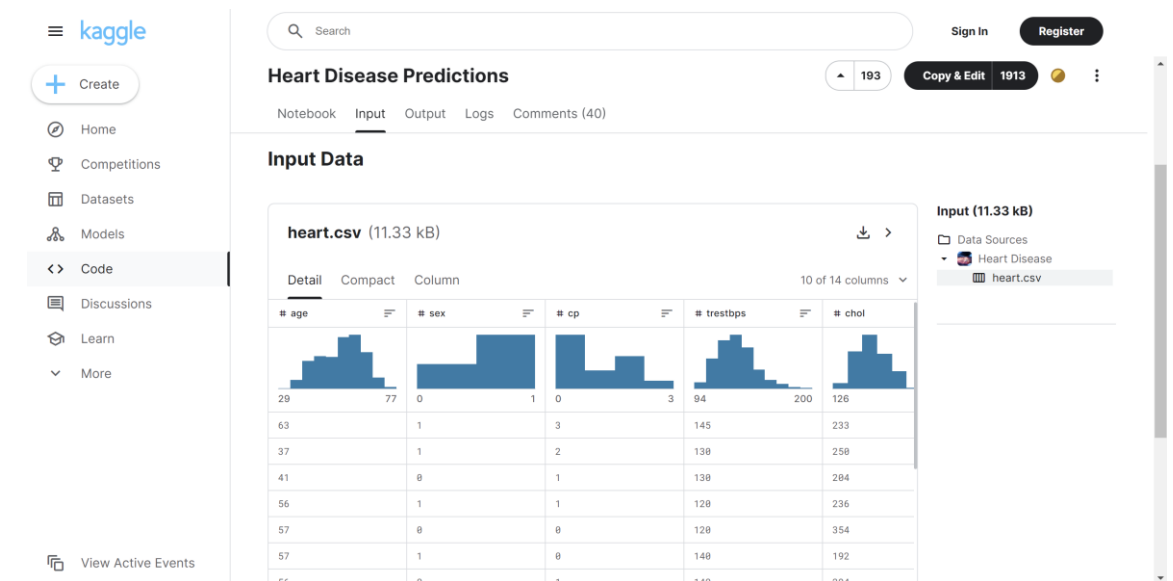


Slika 1 - Prikaz CSV datoteke za preuzimanje u sklopu interaktivnog grafa

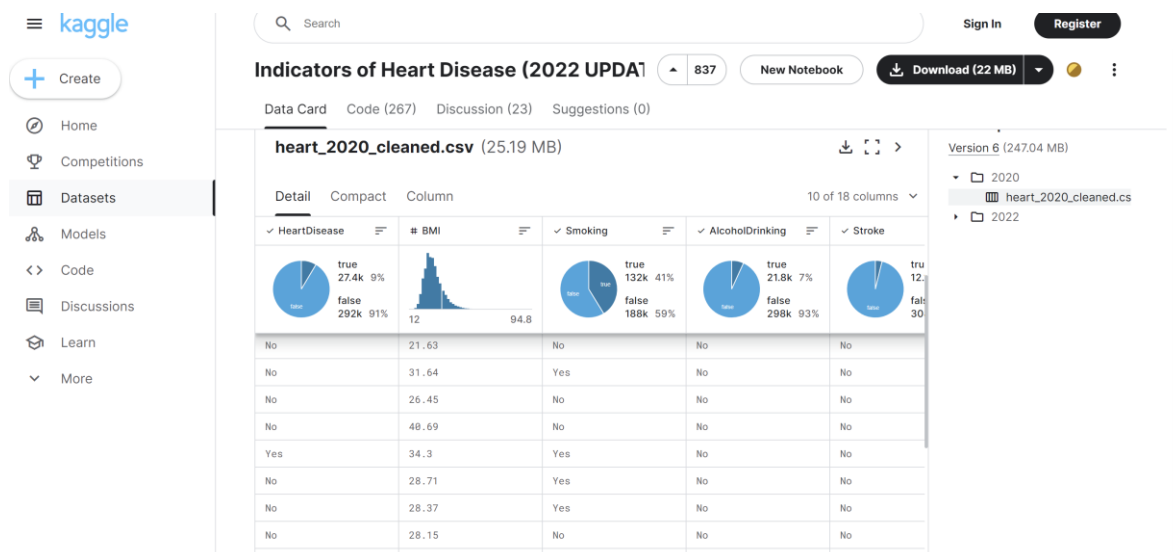
- Kaggle platforma: sa Kaggle platforme preuzeti su skupovi podataka sa naslovima Heart Disease Predictions (Slika 2) i Personal Key Indicators of Heart Disease (Slika 3) [9] [10]. Ovi izvori uključuju informacije o životnim navikama, medicinskoj povijesti i fizičkim karakteristikama ispitanika. Analiza ovih podataka pomaže u identifikaciji

ključnih faktora rizika i njihovom utjecaju na razvoj srčanih bolesti. Koristeći ove podatke, moguće je razviti personalizirane preporuke za prevenciju i liječenje, kao i ciljati specifične grupe s visokim rizikom u populaciji.

Kombiniranjem ovih izvora, moguće je dobiti sveobuhvatan uvid u različite aspekte kardiovaskularnih bolesti, od globalnih trendova do znanstvenih istraživanja. Ovi podaci omogućuju preciznu analizu i efikasnu vizualizaciju rezultata, što je ključno za donošenje informiranih odluka u aspektu zdravstvene politike i medicinske prakse.



Slika 2 - Heart Disease prediction skup podataka



Slika 3 - Indicators od Heart Disease skup podataka

5. Grafički prikazi podataka o kardiovaskularnim bolestima

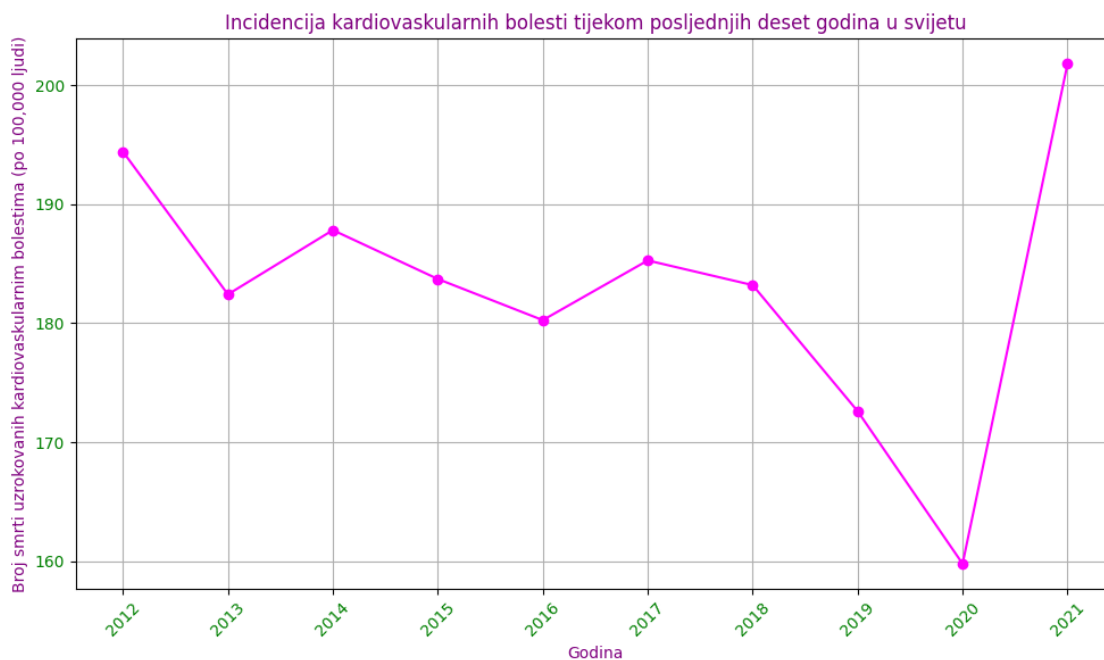
5.1. Trendovi u incidenciji kardiovaskularnih bolesti kroz vremenske serije

Praćenje incidencije kardiovaskularnih bolesti kroz vremenske serije pruža ključne uvide u efikasnost preventivnih mjera i tretmana, te omogućava planiranje budućih strategija u javnom zdravstvu. Analiza i vizualizacija podataka igraju ključnu ulogu u razumijevanju ovih trendova i njihovom prikazivanju na pristupačan način.

Analiza trendova incidencije kardiovaskularnih bolesti (KVD) tijekom posljednjeg desetljeća nudi važne uvide u dinamiku globalnog zdravlja. Podaci koji su obrađeni za analizu, preuzeti su s platforme Our World in Data te omogućuju detaljno razumijevanje promjena u incidenciji KVD-a kroz vrijeme te njihove varijacije među različitim zemljama i regijama [8].

5.1.1. Incidencija KVD-a tijekom posljednjih deset godina u svijetu

Prvi grafikon (Slika 4), linijski prikaz incidencije KVD-a na globalnoj razini od 2012. do 2021. godine, otkriva značajne obrasce na temelju skupa podataka o stopama smrtnosti [8]. Na osi Y prikazan je broj smrtnih slučajeva uzrokovanih kardiovaskularnim bolestima po 100.000 ljudi, dok je na osi X prikazano vremensko razdoblje (Slika 4). Općenito, podaci pokazuju stabilan trend s blagim smanjenjem od 2012. do 2020. godine. Međutim, 2021. godina bilježi nagli porast, što može biti izravno povezano s pandemijom COVID-19, koja je značajno opteretila zdravstvene sustave globalno i indirektno utjecala na povećanje broja smrtnih slučajeva od KVD-a.



Slika 4 - Linijski grafikon; Incidencija KVD-a tijekom posljednjih deset godina u svijetu (graf 1)

Za izradu grafikona korišten je programski jezik Python unutar alata Google Colab. Google Colab alat je koji je razvio Google Research i omogućuje pisanje te izvršavanje koda putem preglednika [11]. Python, sa svojim bibliotekama Pandas i Matplotlib, koje su opsežno korištene za obradu podataka i njihovu vizualizaciju, omogućuje jednostavan i pregledan prikaz te analizu podataka.

Biblioteka Pandas iznimno je korisna zbog svojih moćnih alata namijenjenih analizi i manipulaciji podacima strukturiranog formata, poput tablica i CSV datoteka. Za izradu ovog grafikona vizualizacije, Pandas je korišten za učitavanje podataka iz CSV datoteke naredbom `pd.read_csv(file_path)`, za filtriranje podataka na osnovu godina, te za grupiranje i računanje prosječnih vrijednosti po godini [12].

Matplotlib biblioteka korištena je za kreiranje različitih vrsta vizualizacija u Pythonu. U ovome kodu korištena je za izradu linijskog grafikona koji prikazuje trendove incidencije kardiovaskularnih bolesti unazad deset godina [13].

Kako bi se proces analize i vizualizacije započeo, najprije je napravljena Google Colab bilježnica koja služi za zapisivanje koda koji se zatim izvršava. Kako bi kod (Slika 5) mogao biti izvršen, najprije je bilo potrebno uvesti .csv datoteku s podacima unutar same bilježnice. Kroz nekoliko koraka, kod čita podatke unutar CSV datoteke, filtrira relevantne podatke unutar učitanoj skupa podataka te zatim prelazi na izračun prosječne incidencije na način da se podaci grupiraju po godinama i računa se prosječna vrijednost incidencije kardiovaskularnih bolesti za svaku godinu. Koristeći se Matplotlibom, kreira se linijski grafikon s odgovarajućim oznakama i bojama koje su definirane unutar koda. Konačno, graf se prikazuje naredbom `plt.show()`.

```
import pandas as pd
import matplotlib.pyplot as plt

# čitanje podataka iz csv datoteke
file_path = '/content/cardiovascular-disease-death-rate-who-mdb.csv'
data = pd.read_csv(file_path)

# Filtriranje podataka za posljednjih 10 godina dostupnih u data setu
last_ten_years = data['Year'].max() - 9
recent_data = data[data['Year'] >= last_ten_years]

# Identificiranje ispravnog stupca za podatke o incidenciji
incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages'

# Izračun prosječne incidencije po godini
average_incidence_per_year = recent_data.groupby('Year')[incidence_column].mean().reset_index()

# crtanje grafa
plt.figure(figsize=(10, 6))
plt.plot(average_incidence_per_year['Year'],
         average_incidence_per_year[incidence_column],
         marker='o', linestyle='-', color='fuchsia') # Grafikon u boji fuksije

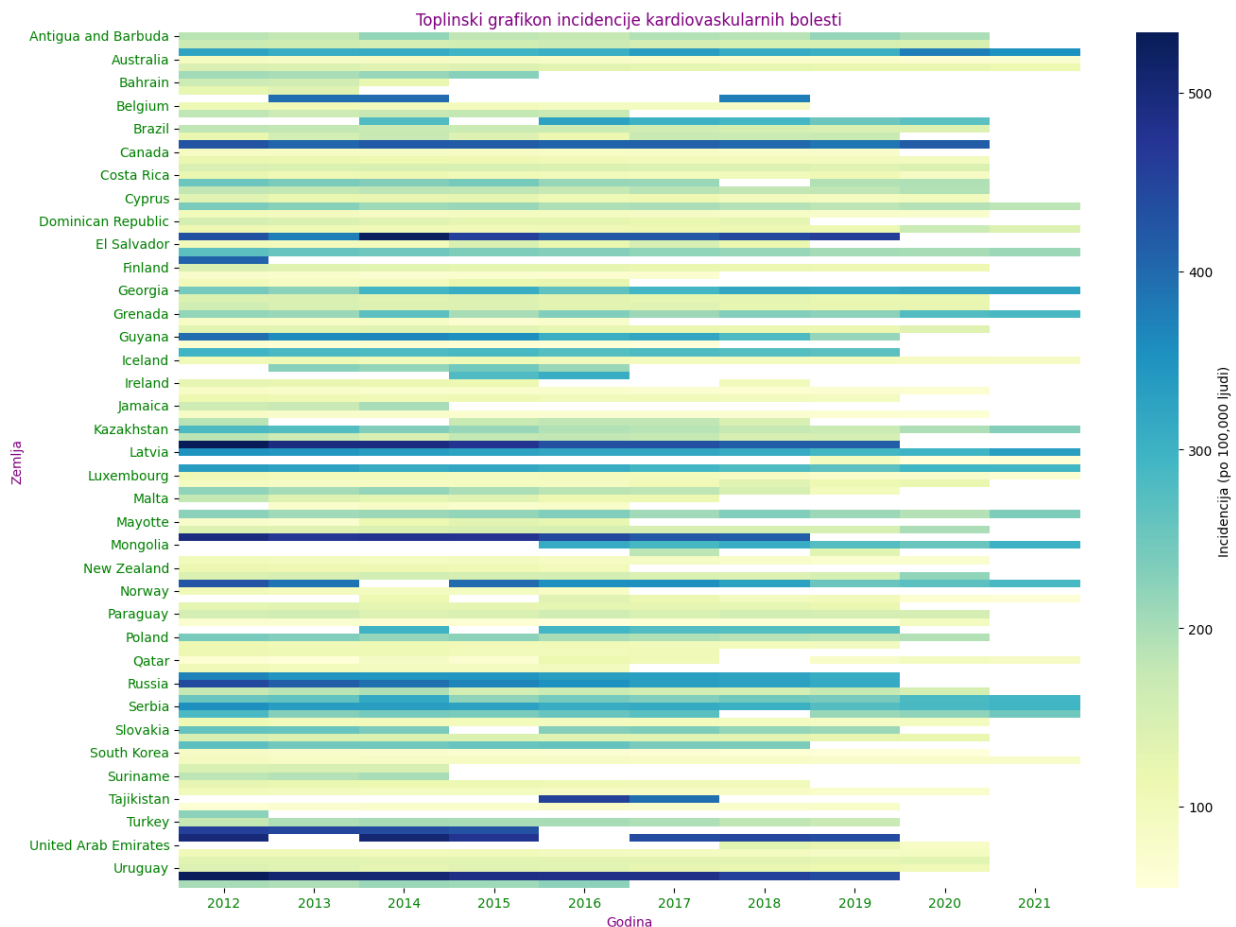
plt.title('Incidencija kardiovaskularnih bolesti tijekom posljednjih deset godina u svijetu', color='purple') # Tekst u ljubičastoj boji
plt.xlabel('Godina', color='purple') # Tekst u ljubičastoj boji
plt.ylabel('Broj smrti uzrokovanih kardiovaskularnim bolestima (po 100,000 ljudi)', color='purple') # Tekst u ljubičastoj boji
plt.grid(True)
plt.xticks(average_incidence_per_year['Year'], rotation=45, color='green') # Godine u zelenoj boji
plt.yticks(color='green') # Vrijednosti na y-osi u zelenoj boji
plt.tight_layout()

# prikaz grafikona
plt.show()
```

Slika 5 - Kod za graf 1; linijski grafikon

5.1.2. Incidencija KVD-a rangirana po zemljama svijeta

Toplinski grafikon (Slika 6) daje složeniji prikaz incidencije KVD-a isključivo po zemljama koje se nalaze u skupu podataka [8]. Osi X prikazuje godine, dok osi Y prikazuje različite zemlje. Intenzitet boje odražava razinu incidencije, gdje tamnije nijanse označavaju veći broj smrtnih slučajeva. Ovaj vizualni prikaz omogućuje brzu identifikaciju zemalja s najvišom i najnižom incidencijom kroz analizirano razdoblje. Primjerice, Latvija, Kazahstan i Mongolija pokazuju konzistentno visoke razine incidencije, što je jasno vidljivo kroz tamnije nijanse boja u svim godinama (Slika 6).



Slika 6 - Toplinski grafikon; Incidencija po nekim od zemalja svijeta (graf 2)

Prilikom izrade toplinskog grafikona (eng. Heat map) pomoću programskog jezika Python, ponovno su radi svoje jednostavnosti korištene biblioteke Pandas i Matplotlib. Osim toga, za izradu ovog grafikona, korištena je i biblioteka – Seaborn.

Seaborn je biblioteka za vizualizaciju podataka koja je temeljena na osnovi Matplotlib-a. Napravljena je tako da bude jednostavna za upotrebu, a da istovremeno pruža moćne mogućnosti prilikom stvaranja atraktivnih i informativnih grafova. Kao takav, Seaborn olakšava vizualizaciju statističkih odnosa između više podataka [14].

U kodu (Slika 7) su najprije uvezene spomenute biblioteke kako bi se moglo ispravno manipulirati podacima i vizualizirati ih. Naredbom `pd.read_csv` učitavamo podatke iz CSV datoteke u `DataFrame` objekat `df` koji zatim prikazuje prvih nekoliko redova podataka da bismo dobili uvid u samu strukturu dataset-a.

Stil vizualizacije od strane Seaborna postavljen je pomoću `sns.set(style="whitegrid")` te je time osigurana profesionalnost i preglednost grafova.

Da bismo kreirali scatter plot, korišten je `sns.scatterplot` gdje podatke prosljeđuje `DataFrame df`, dok `x` i `y` parametri označavaju kolone koje će biti prikazane na `x` i `y` osi. Parametrima `hue` i `style` omogućeno je dodatno označavanje podataka prema kategoriji, pri čemu su kategorije definirane različitim bojama radi veće preglednosti.

Naredbama `plt.title`, `plt.xlabel` i `plt.ylabel` dodan je naslov grafa i imenovane su osi istoga. Time je graf iznimno intuitivan i lak za razumijevanje.

Konačno, naredbom `plt.show()` graf se ispisuje i prikazuje unutar Google Collaba te ima izgled kao na Slici 3.

```
+ Code + Text

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# čitanje podataka iz CSV datoteke
file_path = '/content/cardiovascular-disease-death-rate-who-mdb.csv'
data = pd.read_csv(file_path)

# filtriranje podataka za posljednjih 10 godina unutar dataseta
last_ten_years = data['Year'].max() - 9
recent_data = data[data['Year'] >= last_ten_years]

# identificiranje ispravnog stupca o incidenciji
incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages'

# Heatmap (Toplinski grafikon)
pivot_table = recent_data.pivot_table(values=incidence_column, index='Entity', columns='Year')

plt.figure(figsize=(14, 10))
sns.heatmap(pivot_table, cmap='YlGnBu', cbar_kws={'label': 'Incidencija (po 100,000 ljudi)'})

plt.title('Toplinski grafikon incidencije kardiovaskularnih bolesti', color='purple')
plt.xlabel('Godina', color='purple')
plt.ylabel('Zemlja', color='purple')
plt.xticks(color='green')
plt.yticks(color='green', rotation=0)
plt.tight_layout()

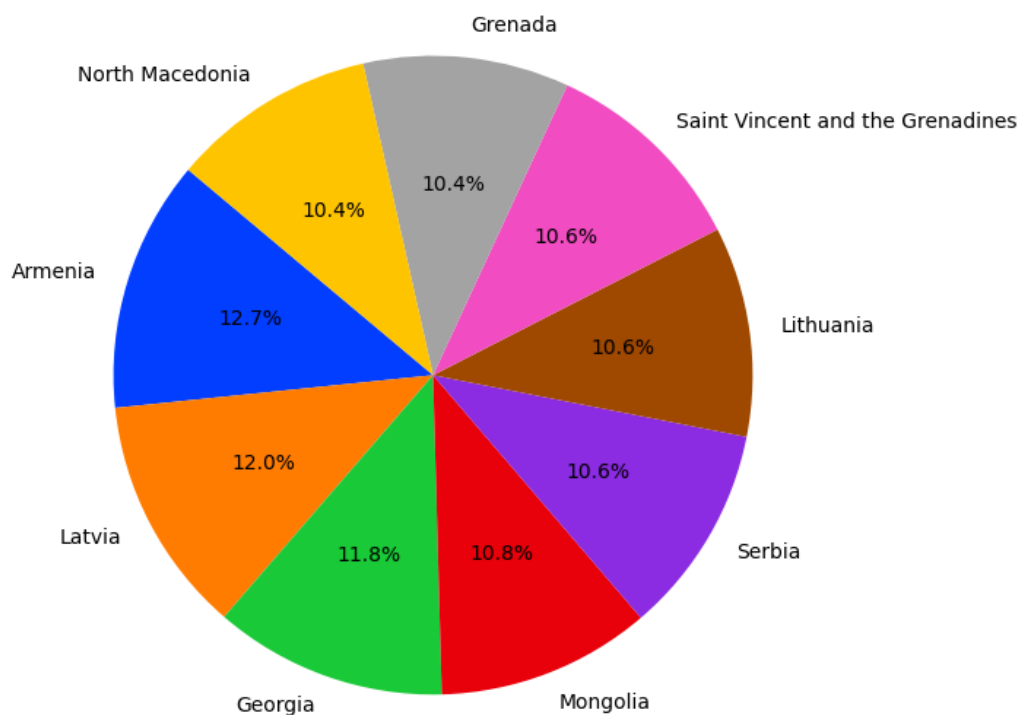
# prikaz grafikona
plt.show()
```

Slika 7 - Kod za graf 2; toplinski grafikon

5.1.3. Zemlje s najvišom incidencijom KVD-a

Pie chart pruža uvid u zemlje s najvišom incidencijom KVD-a u posljednjoj godini analize. Armenia (12.7%), Latvija (12.0%), i Gruzija (11.8%) prednjače s najvišim postotkom smrtnih slučajeva uzrokovanih kardiovaskularnim bolestima (Slika 8). Ovaj dijagram jasno ilustrira distribuciju incidencije među najpogođenijim zemljama, omogućujući bolje razumijevanje globalne raspodjele bolesti koristeći se podacima o stopama smrtnosti od kardiovaskularnih bolesti [8].

Udio zemalja s najvišom incidencijom kardiovaskularnih bolesti u posljednjoj godini



Slika 8 - Pie chart; zemlje s najvišom incidencijom (graf 3)

Pie chart ili tortni grafikon prikazan je također pomoću programskog jezika Python i njegovih biblioteka za vizualizaciju – Matplotlib, Seaborn i Pandas. Kao i u prethodnim grafikonima, ove biblioteke su ključne kako bi se podacima moglo upravljati da rezultat bude estetski ugodan, a analitički pogodan.

Kod (Slika 9) započinje uvozom spomenutih biblioteka, a zatim se podaci učitavaju iz CSV datoteke smještene na specificiranoj putanji. Slijedi filtriranje i priprema podataka iz CSV datoteke, što je postignuto identifikacijom maksimalne godine u skupu podataka i zatim odabirom zapisa iz posljednjeg desetljeća. Cijeli taj postupak definiran je sa `last_ten_years = data['Year'].max() - 9` i `recent_data = data[data['Year'] >=`

last_ten_years]. Ovdje data['Year'].max() pronalazi najnoviju godinu u skupu podataka, a oduzimanje devet osigurava da se uzima u obzir desetljeće podataka.

Pomoću incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages' python skripta identificira specifični stupac koji sadrži relevantne podatke potrebite za izradu vizualizacije, konkretno, podatke o incidenciji kardiovaskularnih bolesti.

latest_year = recent_data['Year'].max() i latest_data = recent_data[recent_data['Year'] == latest_year] korišteni su kako bi skup podataka sveli na najnoviju dostupnu godinu, a zatim slijedi izračun prosječne incidencije po zemlji za tu istu godinu.

Korakom average_incidence_per_country = latest_data.groupby... grupiraju se podaci po entitetima, u ovome slučaju zemljama te se izračunava prosječna stopa incidencije. Na taj način osigurano je da su podaci lakši za rangiranje i vizualizaciju.

Kako bi grafikon bio u formi pie charta, odnosno tortnog grafikona, korištena je funkcija plt.pie koja kreira tortni grafikon. autopct prikazuje vrijednosti u postocima, dok startangle postavlja početni kut grafikona. Naredbama poput plt.title određene su estetske karakteristike grafikona, dok je zaključnom naredbom plt.show grafikon prikazan.

Cijeli kod prikazan je na slici 9.

```
+ Code + Text
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# čitanje podataka iz CSV datoteke
file_path = '/content/cardiovascular-disease-death-rate-who-mdb.csv'
data = pd.read_csv(file_path)

# Filtriranje podataka za posljednjih deset godina dostupnih u datasetu
last_ten_years = data['Year'].max() - 9
recent_data = data[data['Year'] >= last_ten_years]

# Identificiranje ispravnog stupca za podatke o incidenciji
incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages'

# podacie za posljednju godinu
latest_year = recent_data['Year'].max()
latest_data = recent_data[recent_data['Year'] == latest_year]

# prosječni incidencij po zemlji za posljednju godinu
average_incidence_per_country = latest_data.groupby('Entity')[incidence_column].mean().reset_index()

# top 10 zemalja s najvišom incidencijom
top_countries = average_incidence_per_country.nlargest(10, incidence_column)

# crtanje - tortni grafikon
plt.figure(figsize=(10, 6))
colors = sns.color_palette('bright')
plt.pie(top_countries[incidence_column], labels=top_countries['Entity'], colors=colors,
        autopct='%1.1f%%', startangle=140)

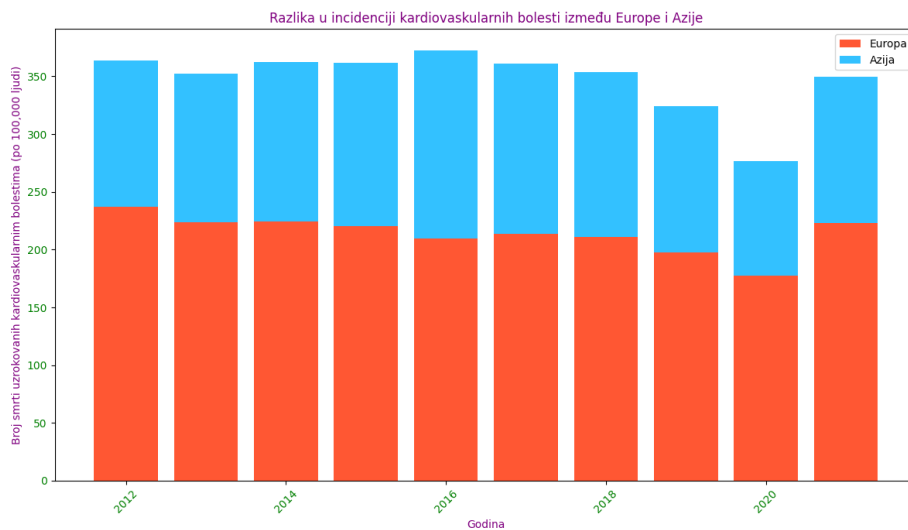
plt.title('Udio zemalja s najvišom incidencijom kardiovaskularnih bolesti u posljednjoj godini', color='purple')
plt.tight_layout()
```

Slika 9 - Kod za graf 3; pie chart

5.1.4. Razlike u incidenciji KVD-a između Europe i Azije

Složeni stupčasti grafikoni (Slika 10) uspoređuje incidenciju KVD-a između Europe i Azije na temelju podataka o stopama smrtnosti od kardiovaskularnih bolesti [8].

Na osi Y prikazan je broj smrtnih slučajeva po 100.000 ljudi, dok osi X prikazuje godine. Boje stupaca diferenciraju regije: Europa je prikazana narančastom, a Azija plavom bojom. Unatoč općem trendu blagog pada incidencije, u 2020. godini primjećuje se pad, dok 2021. godina bilježi porast. Ovi podaci sugeriraju kako su obje regije bile podjednako pogođene pandemijom, s time da je pandemija možda uzrokovala dodatne komplikacije kod pacijenata s kardiovaskularnim bolestima (Slika 10).



Slika 10 - Složeni stupčasti grafikoni - Razlike incidencije između Europe i Azije (graf 4)

Nakon što su primijenjene i uvezene prethodno opisane biblioteke programskog jezika Python, za vizualizaciju u formi složenog stupčastog grafikona potrebno je najprije učitati i filtrirati podatke kako bi se analizirala i prikazala razlika u incidenciji kardiovaskularnih bolesti između Europe i Azije. Kako bi podaci bili ispravno filtrirani, nakon što su učitani, grupiraju se u relevantne zapise pojedinačno za zemlje Azije i za zemlje Europe. Potom se izračunava prosječna godišnja incidencija kardiovaskularnih bolesti za svaki kontinent grupiranjem po godinama.

Konačno, koristi se stupčasti grafikoni kako bi se vizualizirala razlika u incidenciji kardiovaskularnih bolesti između Europe i Azije tijekom posljednjih deset godina. Na grafu se koriste različite boje za svaki kontinent (narančasta za Europu i plava za Aziju), što omogućuje jasno razlikovanje između dviju regija. Naslov grafikona i oznake osi su obojani u ljubičastu boju, dok su oznake na osima u zelenoj boji kako bi se dodatno poboljšala čitljivost i estetika grafikona (Slika 11).

```

+ Code + Text
import pandas as pd
import matplotlib.pyplot as plt

# putanje podataka
file_path = '/content/cardiovascular-disease-death-rate-who-mdb.csv'
data = pd.read_csv(file_path)

# zemlje koje pripadaju Europi i Aziji
european_countries = ['Albania', 'Andorra', 'Armenia', 'Austria', 'Azerbijan', 'Belarus', 'Belgium',
                      'Bosnia and Herzegovina', 'Bulgaria', 'Croatia', 'Cyprus', 'Czech Republic', 'Denmark',
                      'Estonia', 'Finland', 'France', 'Georgia', 'Germany', 'Greece', 'Hungary', 'Iceland',
                      'Ireland', 'Israel', 'Italy', 'Kazakhstan', 'Kosovo', 'Kyrgyzstan', 'Latvia', 'Lithuania',
                      'Luxembourg', 'Malta', 'Maldives', 'Moldova', 'Montenegro', 'Netherlands', 'North Macedonia',
                      'Norway', 'Poland', 'Portugal', 'Romania', 'Russia', 'San Marino', 'Serbia', 'Slovakia',
                      'Slovenia', 'Spain', 'Sweden', 'Switzerland', 'Tajikistan', 'Turkey', 'Turkmenistan',
                      'Ukraine', 'United Kingdom', 'Uzbekistan']

asian_countries = ['Afghanistan', 'Bahrain', 'Bangladesh', 'Bhutan', 'Brunei', 'Cambodia', 'China', 'East Timor',
                  'India', 'Indonesia', 'Iran', 'Iraq', 'Japan', 'Jordan', 'Kuwait', 'Laos', 'Lebanon', 'Malaysia',
                  'Maldives', 'Mongolia', 'Myanmar', 'Nepal', 'North Korea', 'Oman', 'Pakistan', 'Palestine',
                  'Philippines', 'Qatar', 'Saudi Arabia', 'Singapore', 'South Korea', 'Sri Lanka', 'Syria', 'Taiwan',
                  'Thailand', 'United Arab Emirates', 'Vietnam', 'Yemen']

# podaci za zadnjih deset godina dostupnih u datoteci
last_ten_years = data['Year'].max() - 9
recent_data = data[data['Year'] >= last_ten_years]

# stupac stupa za podatke o incidenciji
incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages'

# filtriranje podataka za 20 Europe i Azije
europe_data = recent_data[recent_data['Entity'].isin(european_countries)]
asia_data = recent_data[recent_data['Entity'].isin(asian_countries)]

# prosječne incidencije po godini za Europu i Aziju
europe_avg_incidence = europe_data.groupby('Year')[incidence_column].mean().reset_index()
asia_avg_incidence = asia_data.groupby('Year')[incidence_column].mean().reset_index()

# stapani grafikon
plt.figure(figsize=(12, 7))
plt.bar(europe_avg_incidence['Year'], europe_avg_incidence[incidence_column], color='#F5733', label='Europa') # Narandžasta boja
plt.bar(asia_avg_incidence['Year'], asia_avg_incidence[incidence_column], bottom=europe_avg_incidence[incidence_column], color='#33CFF', label='Azija') # Plava boja

plt.title('Razlika u incidenciji kardiovaskularnih bolesti između Europe i Azije', color='purple')
plt.xlabel('Godina', color='purple')
plt.ylabel('Broj smrti uzrokovanih kardiovaskularnim bolestima (po 100,000 ljudi)', color='purple')
plt.xticks(rotation=45, color='green')
plt.yticks(color='green')
plt.legend()
plt.tight_layout()

# prikaz grafikona
plt.show()

```

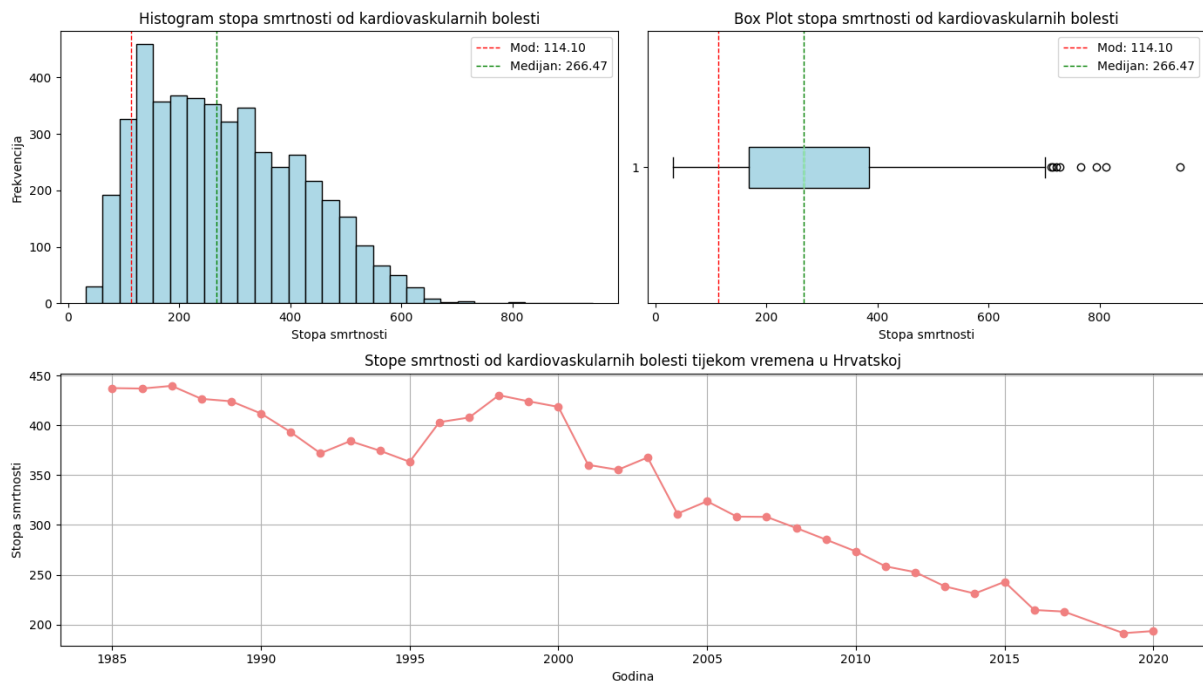
Slika 11 – Kod za graf 4 – složeni stupčasti grafikon

5.1.5. Smrtnost uzrokovana kardiovaskularnim bolestima

Histogram stopa smrtnosti od kardiovaskularnih bolesti prikazuje frekvenciju i distribuciju stopa smrtnosti od KVD-a (Slika 12), a na temelju podataka o stopama smrtnosti od kardiovaskularnih bolesti [8]. Os X predstavlja stope smrtnosti, dok os Y prikazuje frekvenciju. Mod (najčešća vrijednost) je 114.10 (označeno crvenom linijom), dok je medijan 266.47 (označeno zelenom linijom). Histogram pokazuje širok raspon stopa smrtnosti, s najvećom frekvencijom oko medijana (Slika 12).

Box Plot grafikon daje pregled raspodjele stopa smrtnosti s modom i medijanom jasno označenim. Kutija prikazuje interkvartilni raspon, dok linije (brkovi) prikazuju varijaciju izvan ovog raspona. To omogućava identifikaciju ekstremnih vrijednosti i uvid u varijabilnost podataka.

Stope smrtnosti od kardiovaskularnih bolesti tijekom vremena u Hrvatskoj jest linijski grafikon koji prikazuje trend smrtnosti od KVD-a u Hrvatskoj od 1985. do 2020. godine. Os X prikazuje godine, dok os Y prikazuje stope smrtnosti. Primjetan je opći trend smanjenja smrtnosti od sredine 1990-ih, s značajnim padom nakon 2005. godine. Ovaj pad može se pripisati poboljšanjima u zdravstvenoj zaštiti, prevenciji i svijesti o važnosti zdravih životnih stilova.



Slika 12 - Smrtnost od KVD-a (graf 5)

Sljedeći kod (Slika 13) analizira i vizualizira podatke o stopi smrtnosti od kardiovaskularnih bolesti koristeći se skupom podataka o stopama smrtnosti od kardiovaskularnih bolesti [8]. Nakon učitanih i pripremljenih podataka, prvi korak za izradu jest preimenovanje stupca radi lakšeg pristupa, a zatim slijedi izračun statističkih mjera moda i medijana. Mod je najčešća vrijednost u skupu podataka, dok je medijan srednja vrijednost koja dijeli podatke na dvije jednake polovice. Ove statističke mjere u kodu su definirane sa `mode = data['Death_Rate'].mode()[0]` i `median = data['Death_Rate'].median()`.

Isertavanje se u kodu izvršava pomoću `plt.figure(figsize=(14, 8))`. Zatim slijedi histogram koji koristi 30 binova, odnosno intervala, i prikazuje frekvenciju stopa smrtnosti. Crvenom linijom označen je mod, dok zelena linija označava medijan.

Box plotom je prikazana raspodjela stopa smrtnosti i identificirane su vrijednosti koje se značajno razlikuju od ostalih vrijednosti iz skupa podataka, tzv. outlieri.

Linijski graf prikazuje promjene u stopi smrtnosti kroz godine koristeći se podacima koji su specifični za Hrvatsku, a koje smo specificirali kao entitet (Slika 12).

Kako se grafovi ne bi preklapali već bili ravnomjerno raspoređeni, korištena je naredba `plt.tight_layout()` neposredno prije standardne naredbe `plt.show()`.

```
+ Code + Text
Add text cell
import pandas as pd
import matplotlib.pyplot as plt

# Učitaj dataset
file_path = '/content/cardiovascular-disease-death-rate-who-mdb.csv'
data = pd.read_csv(file_path)

# Preimenuj kolonu radi lakšeg pristupa
data.rename(columns={'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages': 'Death_Rate'}, inplace=True)

# Izračunaj mod i medijan
mode = data['Death_Rate'].mode()[0]
median = data['Death_Rate'].median()

# Postavljanje pastelnih boja
plt.style.use('seaborn-pastel')

# Plotanje
plt.figure(figsize=(14, 8))

# Histogram
plt.subplot(2, 2, 1)
plt.hist(data['Death_Rate'], bins=30, edgecolor='black', color='lightblue')
plt.axvline(mode, color='red', linestyle='dashed', linewidth=1, label=f'Mod: {mode:.2f}')
plt.axvline(median, color='green', linestyle='dashed', linewidth=1, label=f'Medijan: {median:.2f}')
plt.title('Histogram stopa smrtnosti od kardiovaskularnih bolesti')
plt.xlabel('Stopa smrtnosti')
plt.ylabel('Frekvencija')
plt.legend()

# Box Plot
plt.subplot(2, 2, 2)
plt.boxplot(data['Death_Rate'], vert=False, patch_artist=True, boxprops=dict(facecolor='lightblue'))
plt.axvline(mode, color='red', linestyle='dashed', linewidth=1, label=f'Mod: {mode:.2f}')
plt.axvline(median, color='green', linestyle='dashed', linewidth=1, label=f'Medijan: {median:.2f}')
plt.title('Box Plot stopa smrtnosti od kardiovaskularnih bolesti')
plt.xlabel('Stopa smrtnosti')
plt.legend()

# Linijski graf za Hrvatsku
croatia_data = data[data['Entity'] == 'Croatia']

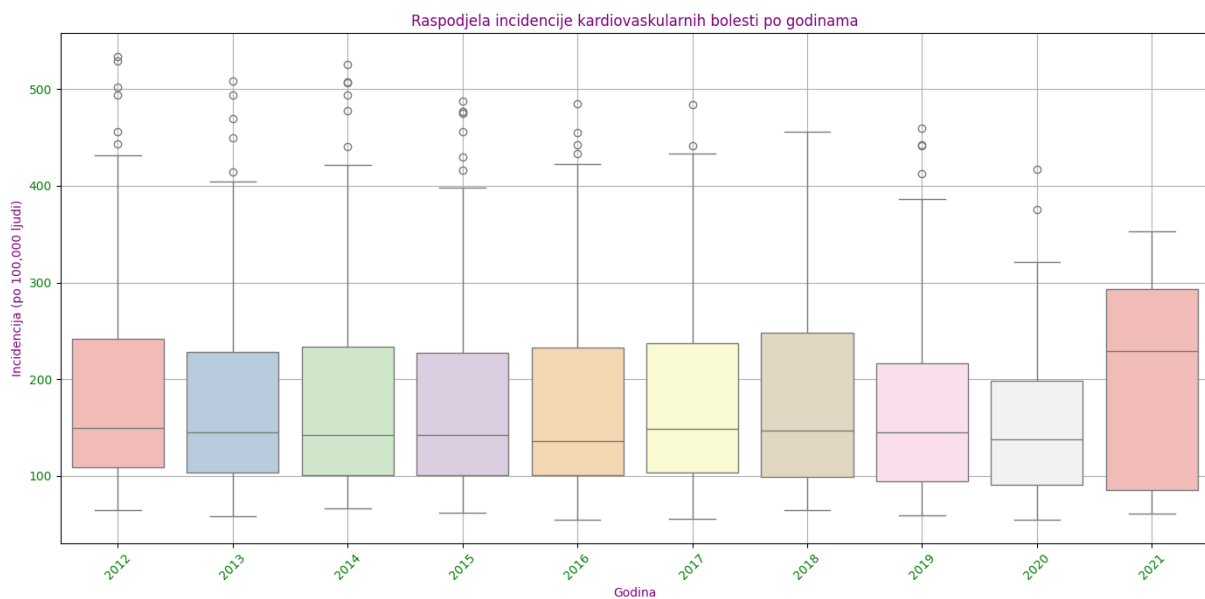
plt.subplot(2, 1, 2)
plt.plot(croatia_data['Year'], croatia_data['Death_Rate'], marker='o', color='lightcoral')
plt.title('Stope smrtnosti od kardiovaskularnih bolesti tijekom vremena u Hrvatskoj')
plt.xlabel('Godina')
plt.ylabel('Stopa smrtnosti')
plt.grid(True)

plt.tight_layout()
plt.show()
```

Slika 13 - Kod za graf 5; smrtnost uzrokovana KVD-om

5.1.6. Raspodjela incidencije KVD-a po godinama

Box plot prikazuje raspodjelu incidencije KVD-a po godinama (Slika 14). Na osi Y prikazana je incidencija po 100.000 ljudi, dok osi X prikazuje godine. Svaka kutija prikazuje raspodjelu podataka za određenu godinu, uključujući medijan, kvartile i ekstremne vrijednosti. Ovaj grafikon otkriva da, unatoč stabilnim medijanima, postoje značajne varijacije između godina, što ukazuje na fluktuacije u broju smrtnih slučajeva među različitim zemljama kroz analizirano razdoblje što sugerira da specifične godine imaju različite distribucije incidenata kardiovaskularnih bolesti među zemljama. Grafikon je izrađen na temelju podataka o stopama smrtnosti od kardiovaskularnih bolesti preuzetih sa stranice OurWorldInData [8].



Slika 14 Box plot - raspodjela incidencije po godinama (graf 6)

Izrada box plot-a, tzv. kutijastog grafikona, u početnom dijelu ne razlikuje se od prethodnih grafova. Uvoz potrebnih biblioteka i učitavanje podataka iz CSV datoteke standardni su postupci i za ovu vrstu vizualizacije.

Najprije su filtrirani podaci za posljednjih deset godina iz učitanoj skupa podataka. Izračunata je najnovija godina u skupu podataka pomoću `data['Year'].max()`, a zatim su podaci filtrirani od te godine pa unatrag devet godina, koristeći `last_ten_years = data['Year'].max() - 9` i `recent_data = data[data['Year'] >= last_ten_years]`.

Nadalje, bilo je potrebno odabrati ispravni stupac za incidenciju pomoću `incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages'`. Nakon što je stupac odabran, kreira se kutijasti grafikon. Naredbom `plt.figure(figsize=(14, 7))` postavljena je

veličina figure, dok se sa `sns.boxplot(x='Year', y=incidence_column, data=recent_data, palette='Pastel1')` kreira kutijasti grafikon. Argument `data=recent_data` specificira skup podataka koji se koristi, dok `palette='Pastel1'` postavlja paletu boja na pastelne nijanse.

Kako bi se poboljšala čitljivost grafikona, dodijeljeni su mu naslov i oznake. Funkcija `plt.title()` postavlja naslov grafikona, `plt.xlabel()` i `plt.ylabel()` postavljaju oznake za x i y os, `plt.xticks(rotation=45, color='green')` postavlja kut rotacije oznaka na x-osi na 45 stupnjeva i boju na zelenu, `plt.grid(True)` uključuje mrežu za bolju čitljivost grafikona, dok `plt.tight_layout()` osigurava da se elementi grafikona ne preklapaju i da su pravilno raspoređeni.

Na kraju, graf se standardno prikazuje pomoću `plt.show()` naredbe (Slika 15).

```
+ Code + Text

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

# čitanje podataka
file_path = '/content/cardiovascular-disease-death-rate-who-mdb.csv'
data = pd.read_csv(file_path)

# podaci za posljednjih deset godina dostupnih u datasetu
last_ten_years = data['Year'].max() - 9
recent_data = data[data['Year'] >= last_ten_years]

# ispravan stupac za podatke o incidenciji
incidence_column = 'Age-standardized deaths that are from cardiovascular diseases per 100,000 people, in both sexes aged all ages'

# kutijasti grafikon
plt.figure(figsize=(14, 7))
sns.boxplot(x='Year', y=incidence_column, data=recent_data, palette='Pastel1')

plt.title('Raspodjela incidencije kardiovaskularnih bolesti po godinama', color='purple')
plt.xlabel('Godina', color='purple')
plt.ylabel('Incidencija (po 100,000 ljudi)', color='purple')
plt.xticks(rotation=45, color='green')
plt.yticks(color='green')
plt.grid(True)
plt.tight_layout()

# prikaz grafikona
plt.show()
```

Slika 15 - Kod za graf 6; kutijasti grafikon

Analizom i vizualizacijom podataka iz spomenute .csv datoteke, uočava se nekoliko ključnih trendova u incidenciji kardiovaskularnih bolesti (KVD):

Rani Porast: U prvim godinama promatranja, primjećujemo porast incidencije KVD-a. Ovaj porast može biti povezan s faktorima kao što su nezdrav životni stil, povećana prevalencija pušenja, nezdrave prehrane i fizičke neaktivnosti.

Stabilizacija: Sredinom promatranog perioda, incidencija se stabilizirala. Stabilizacija može biti rezultat uspješnih kampanja za prevenciju KVD-a, bolje edukacije javnosti i unapređenja u dijagnostici i liječenju.

Pad: U posljednjih nekoliko godina bilježimo pad incidencije. Ovaj pad može biti posljedica daljnjeg poboljšanja preventivnih mjera, povećane svijesti o važnosti zdravih životnih stilova i napredaka u medicinskoj tehnologiji.

Nagli Rast: U 2021. godini primijećen je nagli rast incidencije kardiovaskularnih bolesti što je vjerojatno posljedica kombinacije zdravstvenih, društvenih i ekonomskih faktora uzrokovanih pandemijom COVID-19. Smanjena dostupnost redovne medicinske skrbi, povećani stres i promjene u životnim navikama igraju ključnu ulogu u ovom povećanju.

Zaključno, analiza ovih grafova pruža dubok uvid u globalne i regionalne trendove incidencije kardiovaskularnih bolesti. Stabilne, ali visoke razine incidencije globalno, te nagli porast u 2021. godini, naglašavaju potrebu za kontinuiranim praćenjem i implementacijom preventivnih mjera. Identifikacija najpogođenijih zemalja omogućuje ciljano usmjeravanje resursa kako bi se smanjila incidencija KVD-a i poboljšalo globalno javno zdravlje.

5.2. Indikatori kardiovaskularnih bolesti

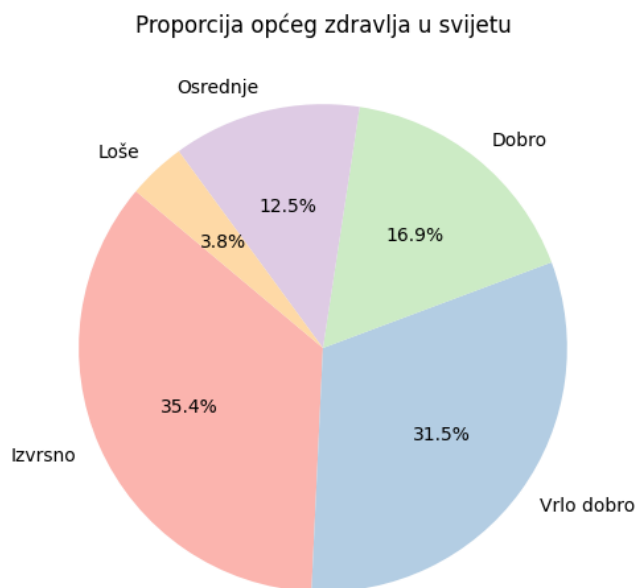
Podaci korišteni za analizu indikatora kardiovaskularnih bolesti skinuti su sa Kaggle stranice pod nazivom „Indicators of Heart Diseases“, a u CSV formatu [10]. Podaci sadrže različite zdravstvene pokazatelje i čimbenike rizika koji su usko povezani s kardiovaskularnim bolestima. U skupu podataka obuhvaćene su informacije poput općeg zdravstvenog stanja, općeg mentalnog stanja, broja dana lošeg fizičkog i mentalnog zdravlja, vremena od zadnjeg liječničkog pregleda, sudjelovanja u fizičkim aktivnostima, sati sna, indeksa tjelesne mase (BMI), potencijalnoj povijesti srčanih udara i zastoja, te konzumacije alkohola i pušenja.

Ovakvi podaci i njihovo praćenje, analiza i vizualizacija od izuzetne su važnosti za razumijevanje kako različiti čimbenici utječu na pojavu i razvoj kardiovaskularnih bolesti, ali i kako se njihovim praćenjem može potaknuti prevencija istih. Takav pothvat automatski pomaže u prepoznavanju trendova i anomalija koje bez analize inače mogu ostati neprimijećene. Na taj način, podaci i njihove analize igraju ključnu ulogu za unaprjeđenje javnog zdravstva i kvalitete života, ali mogu rezultirati i u smanjenju stope mortaliteta od KVD-a u svijetu.

5.2.1. Proporcija općeg zdravlja

Pie chart u nastavku vizualizira proporcije općeg zdravlja među ispitanicima diljem svijeta. Ispitanici su status svog općeg zdravlja ispunili prema pet sljedećih kategorija: izvrsno, vrlo dobro, dobro, osrednje i loše.

Analizom vizualiziranih podataka vidljivo je da je čak 35.4% ispitanika svoje zdravlje procijenilo kao izvrsno, 31.5% kao vrlo dobro, 16.9% kao dobro, 12.5% kao osrednje i tek 3.8% kao loše (Slika 16). Iz analize je jasno vidljivo da najveći postotak ispitanika ocjenjuje svoje zdravlje kao izvrsno i vrlo dobro, dok tek manji postotak svoje zdravlje ocjenjuje kao loše.



Slika 16 - Pie chart; proporcije općeg zdravlja (graf 7)

Kako bi se izradila prikazana vizualizacija, prvo su učitane potrebne biblioteke 'pandas' i 'matplotlib.pyplot' koje su i ranije bile potrebite za stvaranje grafova. Podatke učitavamo iz CSV datoteke s određene putanje. Slijedi računanje frekvencije svake kategorije općeg zdravlja koristeći se s 'value_counts' metodom na stupcu 'GeneralHelath'. Nazivi kategorija definirani su da pišu na hrvatskom jeziku, a za bolju estetiku odabrane su pastelne boje.

Konačno, kreiramo pie chart sa točno definiranim veličinama, nazivima kategorija, automatskim postocima i definiranim bojama. Grafu je dodijeljen naslov i graf se prikazuje. Kod opisanog postupka prikazan je na slici 17.

```
+ Code + Text

2s ▶ import pandas as pd
import matplotlib.pyplot as plt

# Učitavanje podataka
file_path = '/content/rizici_za_bolest.csv'
risk_factors_data = pd.read_csv(file_path)

# Kreiranje pie chart-a
general_health_counts = risk_factors_data['GeneralHealth'].value_counts()

# nazivi kategorija
labels_hr = ['Izvršno', 'Vrlo dobro', 'Dobro', 'Osrednje', 'Loše']

# Boje
colors = plt.cm.Pastel1(range(len(general_health_counts)))

plt.figure(figsize=(10, 6))
plt.pie(general_health_counts, labels=labels_hr, autopct='%1.1f%%', startangle=140, colors=colors)
plt.title('Proporcija općeg zdravlja u svijetu')
plt.show()
```

Slika 17 - Kod za graf 7; pie chart

5.2.2. Broj dana lošeg fizičkog zdravlja po općem zdravlju

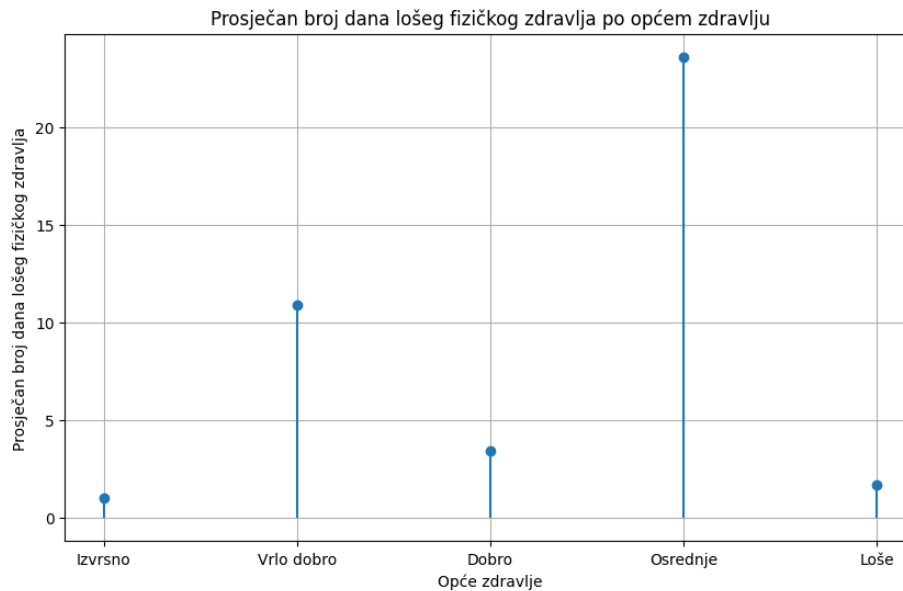
Broj dana lošeg fizičkog zdravlja vizualiziran je lizalica grafikonom (eng. Lollipop chart) koji predstavlja inovativan i vizualno privlačan način prikazivanja podataka, kombinirajući elemente stupčastog grafikona i scatter plot-a. Lollipop chart sastoji se od točaka koje predstavljaju vrhove lizalica i linija koje ih povezuju s osnovom grafikona. (Slika 18)

Prikazan lollipop chart u nastavku ilustrira prosječan broj dana lošeg fizičkog zdravlja po različitim kategorijama općeg zdravlja ispitanika (Slika 18). Podaci o općem zdravlju ispitanika preuzeti su sa Kaggle stranice, konkretno skup podataka podataka 'Key indicators of heart diseases' [10].

Rezultati vizualizacije pokazuju sljedeće:

- Ispitanici koji su svoje zdravlje percipirali izvrsnim imaju minimalan broj dana lošeg fizičkog zdravlja (<5 dana)
- Ispitanici koji su svoje zdravlje percipirali kao vrlo dobrim u prosjeku imaju oko desetak dana lošeg fizičkog zdravlja
- Ispitanici koji su svoje zdravlje percipirali kao dobrim imaju prosječno oko pet dana lošeg fizičkog zdravlja
- Ispitanici koji su svoje zdravlje percipirali kao osrednje imaju najveći prosječan broj dana lošeg fizičkog zdravlja, čak preko petnaestak dana
- Ispitanici koji su svoje zdravlje percipirali kao lošim imaju relativno manji broj dana lošeg fizičkog zdravlja u odnosu na kategoriju ispitanika koji su svoje zdravlje percipirali osrednjim

Ovom vizualizacijom uočljivo je kako se percepcija općeg zdravlja odražava na broj dana lošeg fizičkog zdravlja. Ovakvi prikazi pomažu prilikom identifikacije skupina koje su pod većim rizikom od lošeg fizičkog zdravlja što je osnova za ciljane zdravstvene intervencije i mjere preventive.



Slika 18 - Lollipop chart; dani lošeg fizičkog zdravlja (graf 8)

Za kreiranje vizualizacije grafikona lizalice, osim standardnih, korištene su i neke druge metode i postupci(Slika 19). Najprije se izračunava prosječni broj dana lošeg fizičkog zdravlja pomoću 'groupby' metode za grupiranje podataka. Groupby grupira podatke prema stupcu 'GeneralHealth', a zatim se računa prosječan broj dana lošeg fizičkog zdravlja za svaku grupu. Cijeli postupak izračuna definiran je unutar `avg_physical_health_days = risk_factors_data.groupby('GeneralHealth')['PhysicalHealthDays'].mean()`.

Lollipop chart se zaključno kreira pomoću 'stem' funkcije koja je dio matplotlib biblioteke. Prosječni broj dana lošeg fizičkog zdravlja prikazuje se u obliku točaka povezanih vertikalnim linijama s osnovom grafikona.

```

import pandas as pd
import matplotlib.pyplot as plt

# Učitavanje podataka
file_path = '/content/rizici_za_bolest.csv'
risk_factors_data = pd.read_csv(file_path)

# Izračunavanje prosječnog broja dana lošeg fizičkog zdravlja po općem zdravlju
avg_physical_health_days = risk_factors_data.groupby('GeneralHealth')['PhysicalHealthDays'].mean()

# Prijevod kategorija na hrvatski
labels_hr = ['Izvršno', 'Vrlo dobro', 'Dobro', 'Osrednje', 'Loše']
avg_physical_health_days.index = [labels_hr[i] for i in range(len(avg_physical_health_days))]

# Kreiranje lollipop chart-a
plt.figure(figsize=(10, 6))
plt.stem(avg_physical_health_days.index, avg_physical_health_days.values, basefmt=" ", use_line_collection=True)
plt.title('Prosječan broj dana lošeg fizičkog zdravlja po općem zdravlju')
plt.xlabel('Opće zdravlje')
plt.ylabel('Prosječan broj dana lošeg fizičkog zdravlja')
plt.grid(True)
plt.show()

```

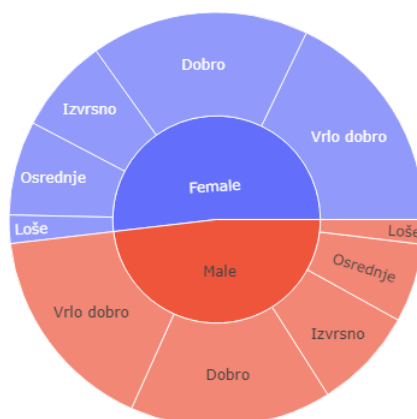
Slika 19 - Kod za graf 8; lollipop chart

5.2.3. Proporcija općeg zdravlja po spolu

Vizualizacija proporcije općeg zdravlja po spolu kreirana je pomoću Sunburst grafikona koji jasno prikazuje distribuciju općeg zdravlja između muškaraca i žena (Slika 20). Podaci o općem zdravlju ispitanika preuzeti su sa Kaggle stranice, konkretno skup podataka 'Key indicators of heart diseases' [10].

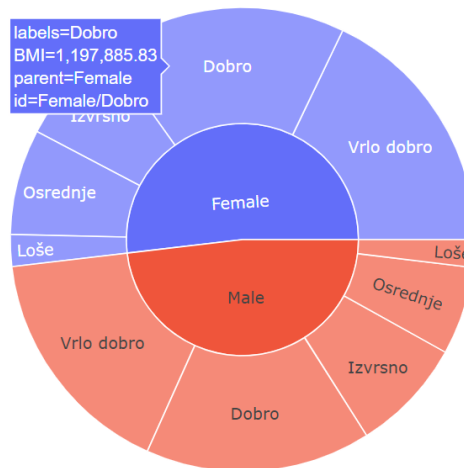
Uočava se da su i žene (Female) i muškarci (Male) skloniji ocijeniti svoje zdravlje kao "Dobro" ili "Vrlo dobro". Kategorije "Izvršno" i "Osrednje" su ravnomjernije raspoređene među oba spola, dok je kategorija "Loše" najmanje zastupljena u oba slučaja.

Proporcija općeg zdravlja po spolu



Slika 20 - Sunburst grafikon; opće zdravlje po spolovima (graf 9)

Kod za kreiranje Sunburst grafikona (Slika 22) koristi 'plotly.express' biblioteku za vizualizaciju. Plotly.express biblioteka je za vizualizaciju podataka koja omogućava jednostavno kreiranje interaktivnih grafova. U ovom primjeru, korištena je funkcija 'px.sunburst' za kreiranje grafikona koji omogućava intuitivno prikazivanje hijerarhijskih podataka [15].



Slika 21 - Interaktivni sunburst

U kodu su kategorije općeg zdravlja prevedene na hrvatski jezik kao i u nekim od prethodnih primjera obzirom da su u skupu podataka kategorije na engleskom jeziku. Cijeli postupak prijevoda smješten je unutar `risk_factors_data['GeneralHealth']`.

Za kreiranje grafikona, podatke je najprije bilo potrebno grupirati prema spolu (Sex), a zatim prema općemu zdravlju na temelju stupca GeneralHealth. Vrijednosti BMI-a korišteni su za proporcije.

```

import pandas as pd
import plotly.express as px

# Učitavanje podataka
file_path = '/content/rizici_za_bolest.csv'
risk_factors_data = pd.read_csv(file_path)

# Prijevod kategorija na hrvatski
translation = {
    'Excellent': 'Izvršno',
    'Very good': 'Vrlo dobro',
    'Good': 'Dobro',
    'Fair': 'Osrednje',
    'Poor': 'Loše'
}

risk_factors_data['GeneralHealth'] = risk_factors_data['GeneralHealth'].map(translation)

# Kreiranje sunburst chart-a
fig = px.sunburst(risk_factors_data, path=['Sex', 'GeneralHealth'], values='BMI',
                 title='Proporcija općeg zdravlja po spolu')

# Prikazivanje grafikona
fig.show()

```

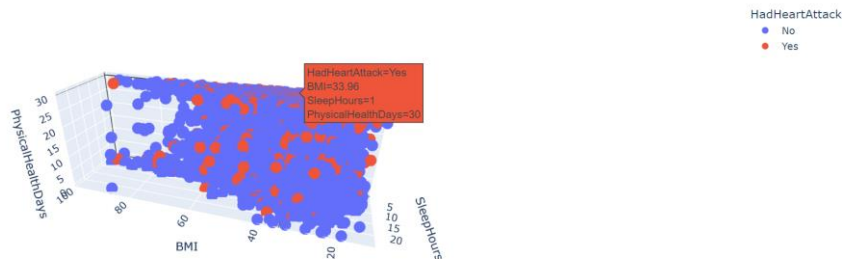
Slika 22 - Kod za graf 9; sunburst grafikon

5.2.4. 3D prikaz odnosa između BMI-a, broja sati sna i broja dana lošeg fizičkog zdravlja

Prikazani 3D scatter plot grafikon vizualizira odnos između indeksa tjelesne mase (BMI), broja sati sna i broja dana lošeg fizičkog zdravlja kod ispitanika (Slika 23). Podaci o indikatorima iz grafikona, preuzeti su sa Kaggle stranice, konkretno skup podataka podataka 'Key indicators of heart diseases' [10]. Na slici 24 prikazana je interaktivnost 3D scatter plot grafikona.

Točke su obojane prema povijesti srčanog udara. Crvenom bojom označene su točke za ispitanike koji posjeduju povijest srčanog udara, dok su plavom bojom označene točke za ispitanike bez povijesti srčanih udara.

Odnos između BMI-a, broja sati sna i broja dana lošeg fizičkog zdravlja



Slika 23; 3D Scatter plot; odnosi BMI, san, loše fizičko zdravlje (graf 10)

Odnos između BMI-a, broja sati sna i broja dana lošeg fizičkog zdravlja



Slika 24 - interaktivnost 3D Scatter plot-a

Kod (Slika 25) za kreiranje ovog 3D grafikona također se koristi plotly.express bibliotekom. Os X predstavlja BMI, os Y broj sati sna, dok os Z predstavlja broj dana lošeg fizičkog zdravlja. Interaktivnost grafikona omogućena je pomoću plotly.express biblioteke te je samim time dozvoljen dublji i bolji uvid u skup podataka, omogućavajući tako bolje razumijevanje indikatora za kardiovaskularne bolesti.

```
import plotly.express as px

# Kreiranje 3D scatter plot-a
fig = px.scatter_3d(risk_factors_data, x='BMI', y='SleepHours', z='PhysicalHealthDays',
                   color='HadHeartAttack', title='Odnos između BMI-a, broja sati sna i broja dana lošeg fizičkog zdravlja')
fig.show()
```

Slika 25 - Kod za graf 10; 3D Scatter plot

5.2.5. Odnosi između različitih indikatora kardiovaskularnih bolesti

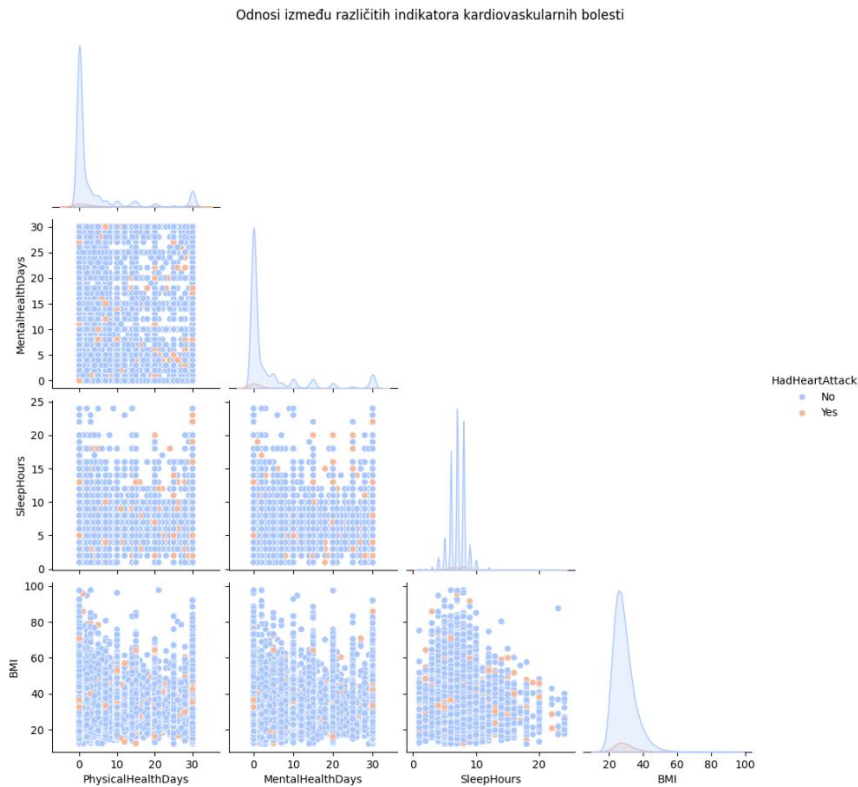
Odnosi između različitih indikatora kardiovaskularnih bolesti: broja dana lošeg fizičkog zdravlja (PhysicalHealthDays), broja dana lošeg mentalnog zdravlja (MentalHealthDays), broja sati sna (SleepHours) i indeksa tjelesne mase (BMI) prikazani su pomoću Pair Plot grafikona (Slika 26). Podaci za vizualizaciju i analizu, preuzeti su sa Kaggle stranice iz skupa podataka 'Key indicators of heart diseases' [10].

Boje točaka označavaju ispitanikovu povijest srčanog udara, pri čemu plava boja predstavlja ispitanike koji nemaju povijest srčanih udara, dok narančasta boja predstavlja ispitanike sa poviješću srčanih udara.

Iz grafikona je jasno vidljivo da je raspon BMI-a kod većine ispitanika između 20 i 40, sa tek malim brojem ekstremno visokih vrijednosti. Također, distribucija broja sati sna pokazuje da većina ispitanika spava između šest i deset sati dnevno.

Promatrajući odnose između indikatora, vidljivo je da ispitanici s višim brojem dana lošeg fizičkog zdravlja često imaju i povišen broj dana lošeg mentalnog zdravlja.

Indeks tjelesne mase ispitanika nema jasnu korelaciju s brojem sati sna, ali postojeća je disperzija, odnosno varira između ispitanika bez jasnog uzroka. Također, ispitanici s poviješću srčanog udara su ravnomjerno raspoređeni u svim kategorijama indikatora, dok su ispitanici bez povijesti srčanog udara dominantni u nižim vrijednostima indikatora.



Slika 26 - Pair plot grafikon; odnosi između indikatora bolesti (graf 11)

Za izradu grafikona, u kodu (Slika 27) je najprije potrebno napraviti izbor indikatora koje želimo koristiti. Isti je definiran unutar `indicators = ['PhysicalHealthDays', 'MentalHealthDays', 'SleepHours', 'BMI']`.

'`hue='HadHeartAttack'`' boja točke prema povijesti srčanih udara kod ispitanika, `palette='coolwarm'` koristi paletu boja za jasniji prikaz i razliku između ispitanika sa poviješću srčanog udara i ispitanika bez povijesti srčanih udara.

Naslov grafikona dodan je pomoću `pairplot.fig.suptitle('Odnosi između različitih indikatora kardiovaskularnih bolesti', y=1.02)`.

```

import seaborn as sns
import matplotlib.pyplot as plt

# Izbor indikatora kardiovaskularnih bolesti
indicators = ['PhysicalHealthDays', 'MentalHealthDays', 'SleepHours', 'BMI']

# Kreiranje pair plot-a s boljim bojama i hue za povijest srčanog udara
pairplot = sns.pairplot(risk_factors_data[indicators + ['HadHeartAttack']],
                        hue='HadHeartAttack',
                        palette='coolwarm',
                        corner=True)

# Dodavanje naslova
pairplot.fig.suptitle('Odnosi između različitih indikatora kardiovaskularnih bolesti', y=1.02)

# Prikazivanje grafikona
plt.show()

```

Slika 27 -Kod za graf 11; Pair plot

5.2.6. Korelacija između indikatora kardiovaskularnih bolesti

Prikazani toplinski grafikon (eng. Heat map) vizualizira korelaciju između različitih indikatora kardiovaskularnih bolesti danih unutar skupa podataka (Slika 28). Indikatori koji se promatraju su broj dana lošeg fizičkog zdravlja (DaniLošegFizičkogZdravlja), broj dana lošeg mentalnog zdravlja (DaniLošegMentalnogZdravlja), broj sati sna (SatiSna), indeks tjelesne mase (BMI), konzumacija alkohola (KonzumacijaAlkohola) i povijest srčanog udara (SrčaniUdar), a preuzeti su sa Kaggle stranice iz skupa podataka ' Key indicators of heart diseases' [10].

Analizom napravljene vizualizacije mogu se uočiti pozitivne, negativne i slabe korelacije.

Pozitivna korelacija od 0.31 vidljiva je između broja dana lošeg fizičkog zdravlja i broja dana lošeg mentalnog zdravlja. Također, pozitivna korelacija pojavljuje se i kod povijesti srčanog udara s brojem dana lošeg fizičkog zdravlja u iznosu od 0.13. Samim time, možemo zaključiti da ispitanici s poviješću srčanog udara imaju više dana lošeg fizičkog i mentalnog zdravlja.

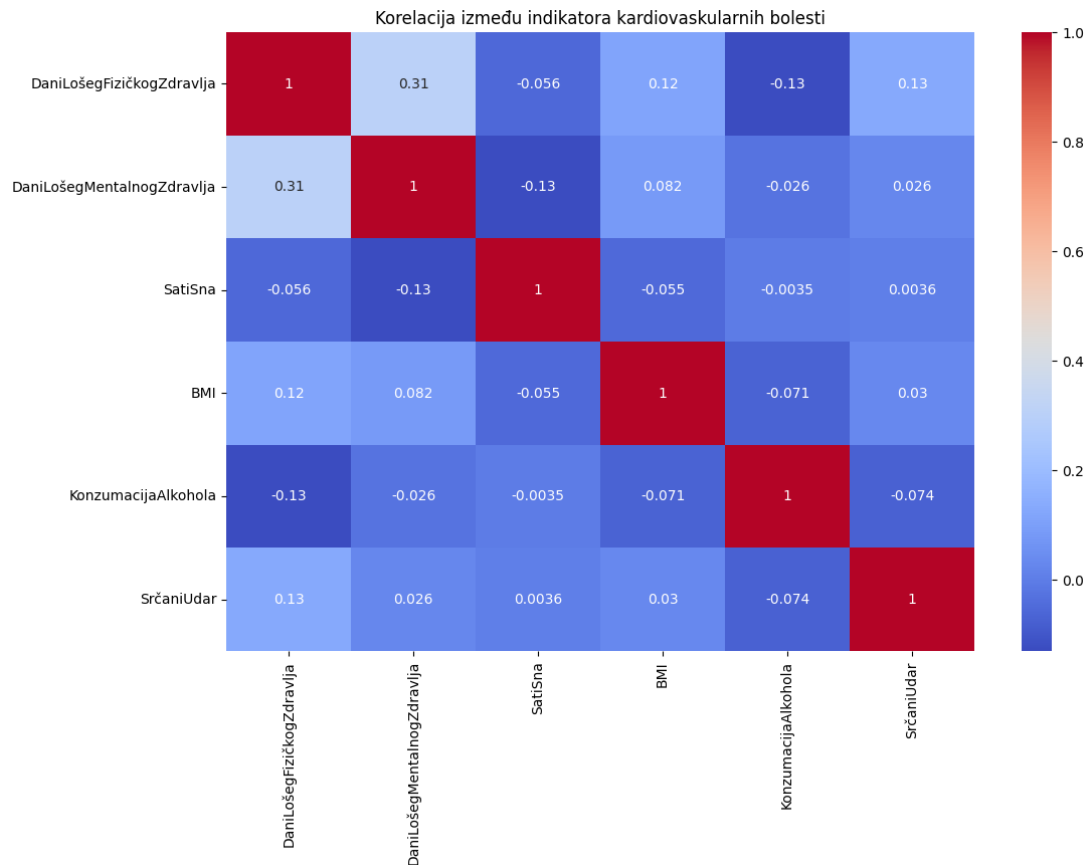
Negativna korelacija uočena je kod broja sati sna. Broj sati sna pokazuje blago negativnu korelaciju s brojem dana lošeg fizičkog zdravlja u iznosu od -0.056 i brojem dana lošeg mentalnog zdravlja u iznosu od -0.13. Može se zaključiti da se više sati sna povezuje sa manjim brojem dana lošeg fizičkog i mentalnog zdravlja.

Indeks tjelesne mase (BMI) ima vrlo slabu korelaciju s ostalim indikatorima. Iz toga možemo zaključiti da BMI nije striktno povezan s brojem dana lošeg fizičkog i mentalnog zdravlja, brojem sati sna i konzumacijom alkohola.

Konzumacija alkohola također pokazuje negativnu korelaciju s brojem dana lošeg fizičkog zdravlja u iznosu od -0.13. To ukazuje da ispitanici koji konzumiraju alkohol imaju tendenciju imati manji broj dana lošeg fizičkog zdravlja. Međutim, obzirom da je ta korelacija izuzetno

niska, jasno je da konzumacija alkohola nije značajan faktor za broj dana lošeg fizičkog zdravlja.

Ovom korelacijskom analizom prikazano je nekoliko ključnih povezanosti između indikatora kardiovaskularnih bolesti. Premda neki indikatori pokazuju blage ili gotovo nikakve korelacije, važno je istaknuti kako različiti aspekti zdravlja međusobno djeluju kako bi se razvile bolje strategije za prevenciju i liječenje kardiovaskularnih bolesti.



Slika 28 - Toplinski grafikon; korelacija indikatora (graf 12)

Da bi vizualizacija toplinskim grafikonom (graf 12) bila što bolja, najprije je bilo potrebno definirati indikatore koje promatramo, a zatim ih zamijeniti hrvatskim nazivima. Bilo je potrebno kopirati originalne podatke i napraviti konverziju tekstualnog stupca u numerički. 'map' metoda korištena je za konverziju tekstualnih vrijednosti 'yes' i 'no' u '1' i '0'.

Nakon što su podaci pripremljeni, izračunava se korelacija između danih indikatora i vizualizira se toplinski grafikon korelacijske matrice. Cijeli kod prikazan je na slici 29.

```
+ Code + Text

import seaborn as sns
import matplotlib.pyplot as plt

# Zamjena naziva indikatora na hrvatski
indicators_hr = {
    'PhysicalHealthDays': 'DaniLošegFizičkogZdravlja',
    'MentalHealthDays': 'DaniLošegMentalnogZdravlja',
    'SleepHours': 'SatiSna',
    'BMI': 'BMI',
    'AlcoholDrinkers': 'KonzumacijaAlkohola',
    'HadHeartAttack': 'SrčaniUdar'
}

# Izbor indikatora kardiovaskularnih bolesti, konvertirajući tekstualne kolone u numeričke
indicators = list(indicators_hr.keys())

# Kopiranje originalnih podataka i konverzija tekstualnih kolona u numeričke
data_for_corr = risk_factors_data.copy()
data_for_corr['AlcoholDrinkers'] = data_for_corr['AlcoholDrinkers'].map({'Yes': 1, 'No': 0})
data_for_corr['HadHeartAttack'] = data_for_corr['HadHeartAttack'].map({'Yes': 1, 'No': 0})

# Preimenovanje kolona na hrvatski
data_for_corr.rename(columns=indicators_hr, inplace=True)

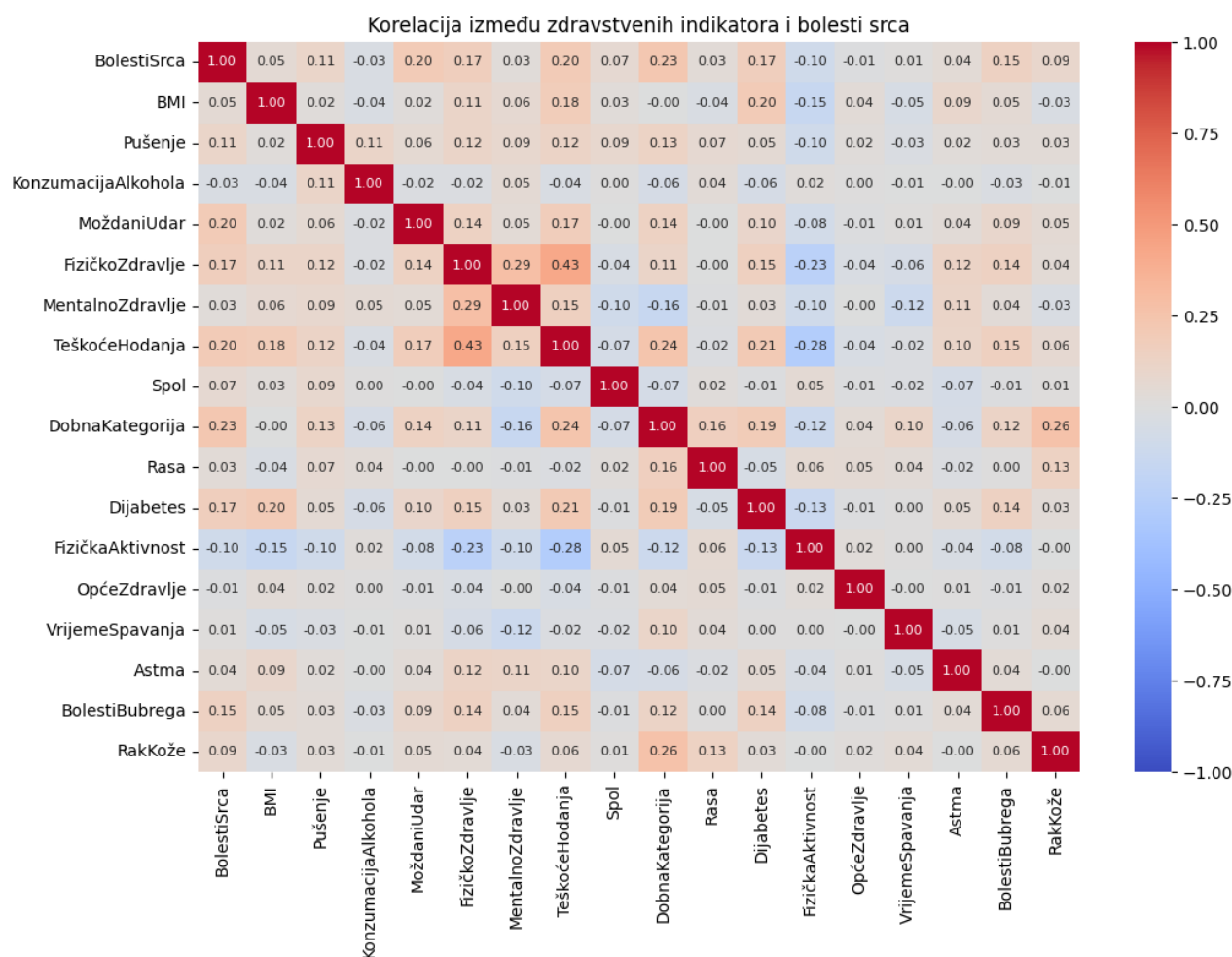
# Priprema podataka
corr_matrix = data_for_corr[list(indicators_hr.values())].corr()

# Kreiranje heatmap-a
plt.figure(figsize=(12, 8))
sns.heatmap(corr_matrix, annot=True, cmap='coolwarm', xticklabels=indicators_hr.values(), yticklabels=indicators_hr.values())
plt.title('Korelacija između indikatora kardiovaskularnih bolesti')
plt.show()
```

Slika 29 - Kod za graf 12; toplinski grafikon

5.2.7. Prikaz korelacije sa više indikatora kardiovaskularnih bolesti

U odnosu na prijašnji prikaz korelacije, za vizualizaciju u nastavku (Slika 30) korišten je 'key indicators of heart diseases' skup podataka koji nudi veći spektar indikatora za kardiovaskularne bolesti [10]. Vizualizacija je napravljena u obliku toplinskog grafikona, te se raspon korelacije kreće od -1 do 1 pri čemu 1 označava savršenu pozitivnu korelaciju, -1 označava savršenu negativnu korelaciju, a 0 označava da nema linearne korelacije između varijabli.



Slika 30 - Toplinski grafikon; korelacija s više indikatora (graf 13)

Ključni zaključci vizualizacije su sljedeći:

- Postojeća je umjerena pozitivna korelacija u iznosu od 0.17 između dijabetesa i kardiovaskularnih bolesti, što potvrđuje zaključak iz prethodnog grafikona da je dijabetes često prisutan kod osoba sa kardiovaskularnim bolestima
- Korelacija teškoće hodanja sa kardiovaskularnim bolestima je 0.20, što ukazuje na povezanost incidencije kardiovaskularnih bolesti s teškoćom hodanja pacijenata
- Govoreći o dobnim kategorijama, prisutna je korelacija u iznosu od 0.23 koja ukazuje da starije osobe imaju veću vjerojatnost za razvoj bolesti srca
- Bolest bubrega pokazuje korelaciju u iznosu od 0.15 i time iskazuje povezanost između bolesti bubrega i bolesti kardiovaskularnog sustava

5.2.8. Povezanost bolesti srca s drugim bolestima i faktorima

Povezanost bolesti srca s drugim bolestima i faktorima vizualizirana je pomoću Sankey interaktivnog grafikona koji je specifično fokusiran na prikazivanje relevantnih povezanosti,

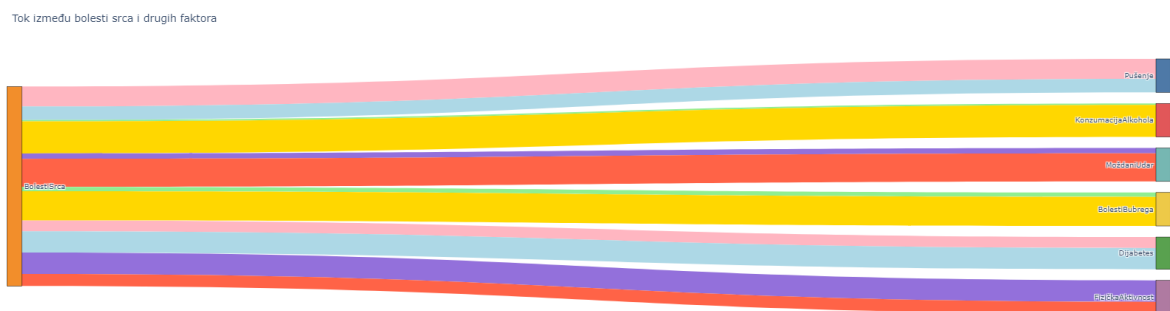
konkretno, kako su bolesti srca povezane s drugim relevantnim faktorima poput pušenja, konzumacije alkohola, moždanog udara, šećerne bolesti, bolesti bubrega i fizičke aktivnosti (Slika 31).

Sankey grafikon sastoji se od čvorova (eng. Nodes), traka (eng. Links) i boja. Čvorovi predstavljaju različite zdravstvene faktore, te su u ovome primjeru to pušenje, konzumacija alkohola, moždani udar, bolesti bubrega, dijabetes i fizička aktivnost. Trake predstavljaju veze između bolesti srca i drugih zdravstvenih faktora. Debljinom trake označen je broj slučajeva gdje su oba faktora prisutna zajedno. Različitim bojama dodatno je omogućeno razlikovanje faktora i veza između njih.

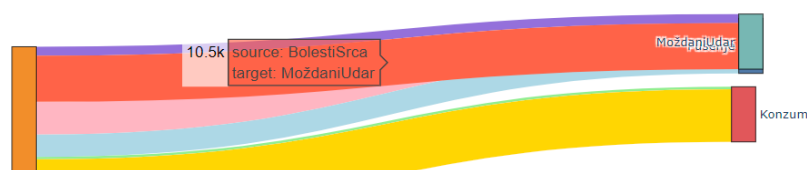
Interaktivnost grafikona (Slika 33) ostvarena je na način da korisnik ima mogućnost pomicanja traka, a upravo tako omogućeno je dublje istraživanje veza između različitih faktora. Pomicanje traka također omogućuje i bolju razlučivost između najviše povezanih faktora.

Prilikom izrade vizualizacije korišteni su podaci iz skupa podataka 'key indicators of heart diseases' [10].

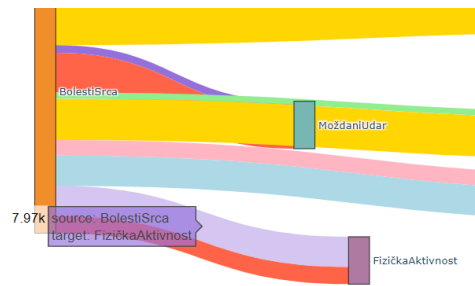
Vizualizacija jasno pokazuje kako postoji značajan broj slučajeva bolesti srca koji su povezani s pušenjem, samim time jasno je da postoji i snažna korelacija ovih dvaju faktora. Povezanost između bolesti srca i moždanog udara (Slika 32) također je vidljiva te ukazuje na učestalu prisutnost u paru. Osim moždanog udara, dijabetes je jednako prisutan kod većeg broja pacijenata sa već prisutnom kardiovaskularnom bolesti pa možemo zaključiti da se uz kardiovaskularnu bolest često veže i šećerna bolest, odnosno dijabetes. Povezanost između bolesti srca i bolesti bubrega se također može opaziti, a također postoji i vidljiv odnos između fizičke aktivnosti i prisutnosti bolesti srca na način da su fizički aktivnije osobe manje sklonije kardiovaskularnim bolestima, sve dok je fizička aktivnost umjerena i u normalnim granicama.



Slika 31 - Sankey grafikon; tok između bolesti srca i drugih bolesti i faktora (graf 14)



Slika 32 - povezanost bolesti srca s moždanim udarom; Sankey grafikon



Slika 33 - interaktivnost; pomicanje traka; Sankey grafikon

Za izradu grafikona najprije je bilo potrebno učitati podatke iz CSV datoteke 'podaci o ispitanicima.csv' preuzete sa Kaggle stranice [10]. Da bi nazivi stupca bili dosljedni dosadašnjim grafikonima, prilagođeni su da budu na hrvatskom jeziku što smo napravili unutar `data.columns` naredbe.

Sljedeće je bilo potrebno definirati kategorije (labels) koje će se prikazivati u dijagramu, ali i kreirati liste za izvore (source), ciljeve (target) i vrijednosti veza između njih. Cijeli taj tok definiran je unutar for petlje u kodu.

U kodu (Slika 34) su definirane i boje koje se koriste za čvorove i veze kako bi se osigurala što bolja čitljivost i preglednost grafikona.

Sam Sankey grafikon kreiran je korištenjem biblioteke 'plotly.graph_objects' koja omogućuje kreiranje složenih i interaktivnih grafikona [16]. Nudi širok raspon grafičkih objekata i grafikona koji su izuzetno prilagodljivi, što definitivno olakšava analiziranje i prezentiranje podataka iz skupa podataka.


```

import pandas as pd
import plotly.graph_objects as go

# Učitaj podatke
data_path = '/content/podaci o ispitanicima.csv'
data = pd.read_csv(data_path)

# Prilagodi nazive stupaca na hrvatski
data.columns = ['BolestiSrca', 'BMI', 'Pušenje', 'KonzumacijaAlkohola', 'MoždaniUdar',
                'FizičkoZdravlje', 'MentalnoZdravlje', 'TeškoćeHodanja', 'Spol', 'DobnaKategorija',
                'Rasa', 'Dijabetes', 'FizičkaAktivnost', 'OpćeZdravlje', 'VrijemeSpavanja',
                'Aстма', 'BolestiBubrega', 'RakKože']

# Pripremi podatke za Sankey dijagram fokusiran na bolesti srca
labels = ['BolestiSrca', 'Pušenje', 'KonzumacijaAlkohola', 'MoždaniUdar', 'Dijabetes', 'BolestiBubrega', 'FizičkaAktivnost']
source = []
target = []
value = []

# Kreiraj tokove između bolesti srca i drugih faktora
for i, label in enumerate(labels[1:]): # Počinje od 1 da bi preskočio 'BolestiSrca' kao izvor
    count_yes = len(data[(data['BolestiSrca'] == 'Yes') & (data[label] == 'Yes')])
    count_no = len(data[(data['BolestiSrca'] == 'Yes') & (data[label] == 'No')])
    if count_yes > 0:
        source.append(0) # BolestiSrca kao izvor
        target.append(i + 1) # Cilj je indeks trenutnog faktora + 1
        value.append(count_yes)
    if count_no > 0:
        source.append(0) # BolestiSrca kao izvor
        target.append(i + 1) # Cilj je indeks trenutnog faktora + 1
        value.append(count_no)

# boje linija
colors = ['#f28e2b', '#4e79a7', '#e15759', '#76b7b2', '#59a14f', '#edc948', '#af7aa1']

# Dodaj boje linkova za bolje razlikovanje
link_colors = ['#FFB6C1', '#ADD8E6', '#90EE90', '#FFD700', '#9370DB', '#FF6347']

# Kreiraj Sankey dijagram
fig = go.Figure(go.Sankey(
    node=dict(
        pad=15,
        thickness=20,
        line=dict(color="black", width=0.5),
        label=labels,
        color=colors
    ),
    link=dict(
        source=source,
        target=target,
        value=value,
        color=link_colors * ((len(value) // len(link_colors)) + 1)
    )
))

fig.update_layout(title_text="Tok između bolesti srca i drugih faktora", font_size=10)
fig.show()

```

Slika 34 - Kod za graf 14; Sankey grafikon

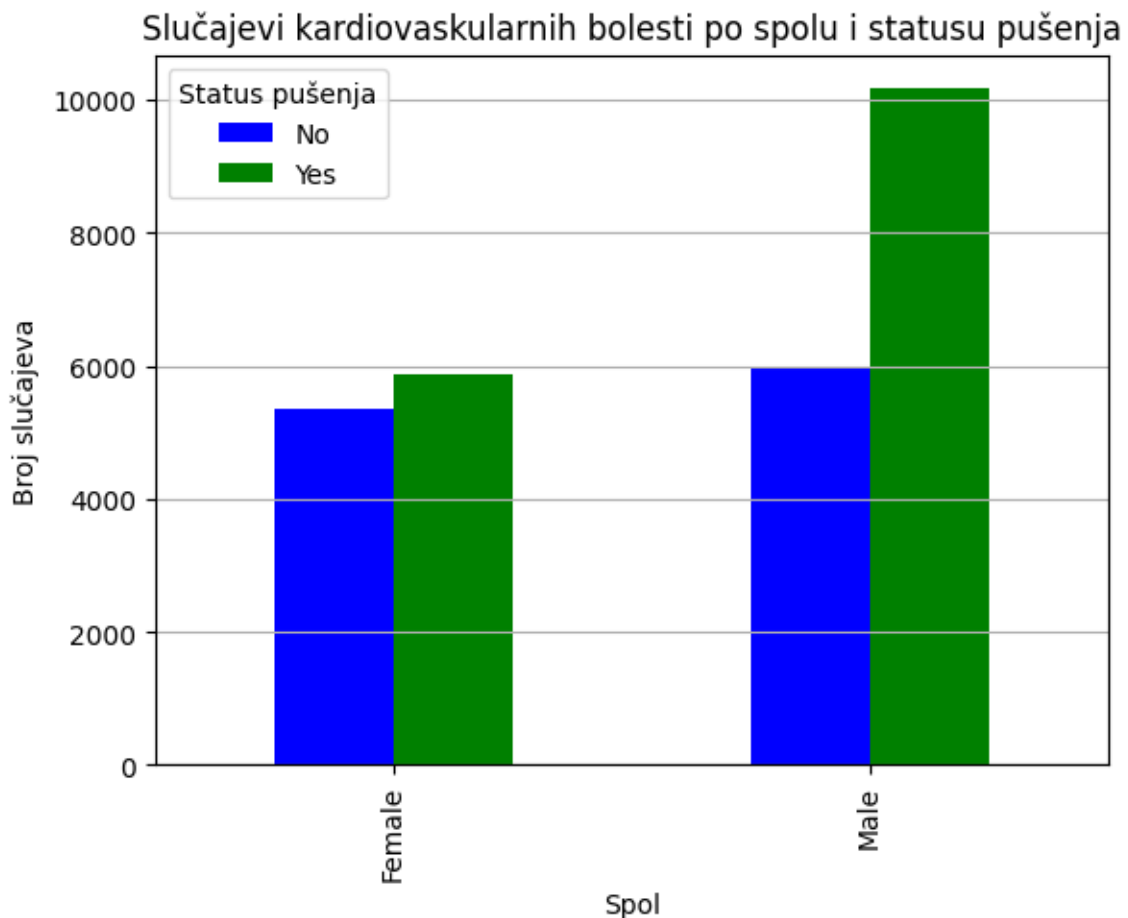
5.2.9. Slučajevi bolesti kardiovaskularnog sustava po spolu i statusu pušenja

Obzirom da je pušenje cigareta jedan od kardinalnih uzročnika kardiovaskularnih bolesti, za prevenciju istih izuzetno je važna statistika kardiovaskularnih bolesti uzrokovana pušenjem, a posebice rangirana po spolu kako bi se lakše moglo zaključiti jesu li muškarci ili žene rizičniji za razvitak bolesti.

Stupčasti grafikon u nastavku (Slika 35) prikazuje slučajeve kardiovaskularnih bolesti po spolu i statusu pušenja. Za vizualizaciju je korišten 'key indicators of heart diseases' skup podataka [10]. Ovom vizualizacijom omogućeno je da vidimo kako pušenje utječe na učestalost bolesti kod muškaraca, a kako kod žena.

Analiza napravljene vizualizacije pokazuje da je ukupan broj slučajeva kardiovaskularnih bolesti veći kod muškaraca nego kod žena. Također, muškarci pušači imaju značajno veći broj slučajeva bolesti u usporedbi s nepušačima. Ukupan broj slučajeva kardiovaskularnih bolesti kod žena generalno je manji nego kod muškaraca, ali ipak žene pušači imaju veći broj slučajeva bolesti u usporedbi s nepušačicama.

Rezultati vizualizacije sugeriraju da pušenje u velikoj mjeri povećava rizik od kardiovaskularnih bolesti kod oba spola, ali kod muškaraca je zabilježen generalno veći broj slučajeva bolesti nego kod žena na svjetskoj razini.



Slika 35 - Stupčasti grafikon; bolesti po spolu i statusu pušenja (graf 15)

Za izradu grafikona korišten je standardni postupak koda u Pythonu kao i kod prethodno napravljenih stupčastih grafikona. Najprije se učitavaju podaci iz CSV datoteke i pohranjuju se u DataFrame, zatim slijedi grupiranje po spolu i statusu pušenja pomoću `heart_disease_gender_smoking=data[data['HeartDisease']=='Yes'].group by(['Sex', 'Smoking']).size().unstack()`.

Grafikon se kreira pomoću Matplotlib biblioteke kao i kod prethodnih primjera. Kod za izradu grafikona prikazan je na slici 36.

```

import pandas as pd
import matplotlib.pyplot as plt

# Učitaj podatke
data_path = '/content/podaci_o_iscipanicima.csv'
data = pd.read_csv(data_path)

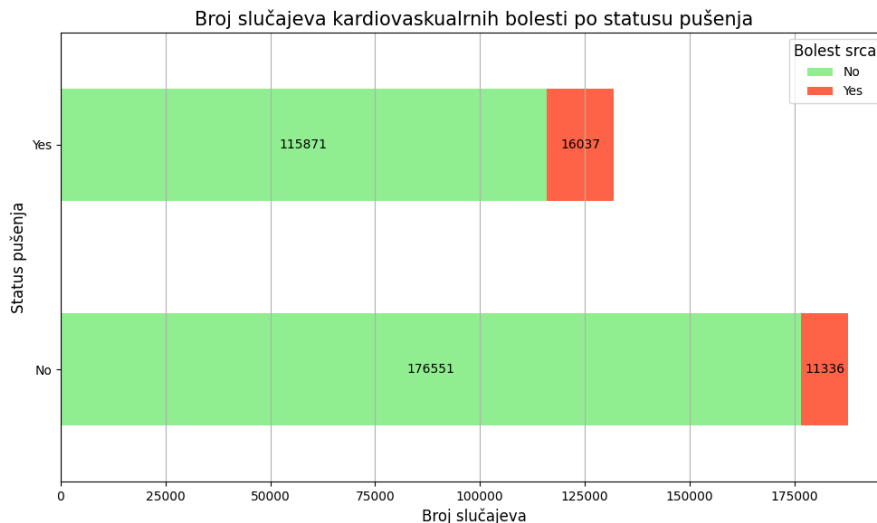
# Pretpostavimo da stupci za spol i status pušenja imaju nazive 'Sex' i 'Smoking'
# Grupiranje po spolu i statusu pušenja
heart_disease_gender_smoking = data[data['HeartDisease'] == 'Yes'].groupby(['Sex', 'Smoking']).size().unstack()

# Kreiraj grafikon
plt.figure(figsize=(10, 6))
colors = ['blue', 'green'] # Prilagođene boje za stupce
heart_disease_gender_smoking.plot(kind='bar', stacked=False, color=colors)
plt.title('Slučajevi kardiovaskularnih bolesti po spolu i statusu pušenja')
plt.xlabel('Spol')
plt.ylabel('Broj slučajeva')
plt.legend(title='Status pušenja')
plt.grid(axis='y')
plt.show()

```

Slika 36 - Kod za graf 15; stupčasti grafikon

Za prikaz štetnosti pušenja cigareta na razvitak oboljenja od kardiovaskularnih bolesti napravljen je i horizontalni *stacked* bar grafikon koji potvrđuje prikaz prethodnog stupičastog grafikona u svezi povezanosti pušenja s razvitkom kardiovaskularnih bolesti (Slika 37). Grafikon jasno daje rezultat da je broj slučajeva bez kardiovaskularnih bolesti mnogo veći među nepušačima u odnosu na pušače.



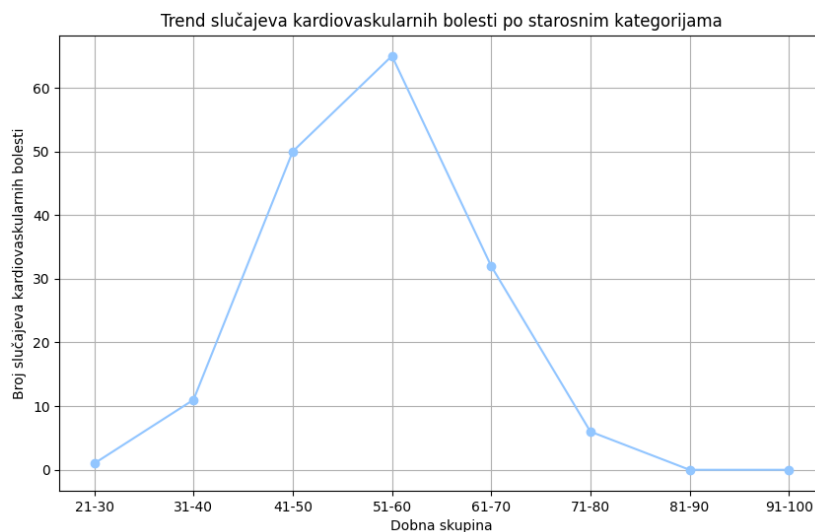
Slika 37 - Horizontalni stacked bar; pušenje cigareta (graf 16)

5.2.10. Trend slučajeva kardiovaskularnih bolesti po starosnim kategorijama

Linijski grafikon u nastavku prikazuje slučajeve kardiovaskularnih bolesti obzirom na starosne kategorije (Slika 38). Na X osi prikazane su dobne skupine, raspoređene u intervalima po 10 godina, a od 21 godine do 100 godina. Y os prikazuje broj slučajeva kardiovaskularnih bolesti

unutar svake dobne skupine. Svaka točka na liniji definira broj slučajeva kardiovaskularnih bolesti za određenu dobnu skupinu, a linija koja povezuje točke omogućuje precizan prikaz trenda bolesti. Podaci za ovu vizualizaciju preuzeti su pod naslovom 'Heart Disease Predictions' sa stranice Kaggle [9].

Iz vizualizacije je vidljivo da dobna skupina od 51-60 godina ima najveći broj slučajeva kardiovaskularnih bolesti. Dobne skupine od 41-50 i 61-70 također imaju povećan broj slučajeva, ali i dalje prednjači broj slučajeva za dobnu skupinu od 51 do 60 godina. Za dobne skupine ispod 40 godina i iznad 70 godina zabilježen je manji broj slučajeva kardiovaskularnih bolesti, dok dobna skupina od 81 godine pa na dalje ima zabilježen najmanji broj slučajeva.



Slika 38 - Linijski grafikon; trend bolesti po dobnim skupinama (graf 17)

Kao i ranije, linijski grafikon napravljen je standardnim postupkom koji započinje učitavanjem podataka. Zatim je potrebno definirati dobne skupine, što je u kodu (Slika 39) učinjeno pomoću 'pd.cut()' funkcije koja dobne podatke dijeli u unaprijed definirane intervale. Podaci se zatim grupiraju prema definiranim dobnim skupinama i slijedi zbrajanje slučajeva kardiovaskularnih bolesti za svaku skupinu, što je u kodu napravljeno pomoću `trend_data = data[data['target'] == 1].groupby('DobnaSkupina').size().reset_index(name='BrojSlučajeva')`.

Nakon toga, slijedi postupak kreacije grafikona pomoću funkcija iz matplotlib biblioteke, te se grafikon konačno prikazuje pomoću '`plt.show()`' funkcije.

```
#graf 16
import pandas as pd
import matplotlib.pyplot as plt

# Učitaj podatke
data_path = '/content/heart.csv'
data = pd.read_csv(data_path)

# Pretvorba dobne kategorije u odgovarajuće skupine
bins = [20, 30, 40, 50, 60, 70, 80, 90, 100]
labels = ['21-30', '31-40', '41-50', '51-60', '61-70', '71-80', '81-90', '91-100']
data['DobnaSkupina'] = pd.cut(data['age'], bins=bins, labels=labels, right=False)

# Brojanje slučajeva bolesti srca po dobnim skupinama
trend_data = data[data['target'] == 1].groupby('DobnaSkupina').size().reset_index(name='BrojSlučajeva')

# Kreiranje grafikona
plt.figure(figsize=(10, 6))
plt.plot(trend_data['DobnaSkupina'], trend_data['BrojSlučajeva'], marker='o')
plt.title('Trend slučajeva kardiovaskularnih bolesti po starosnim kategorijama')
plt.xlabel('Dobna skupina')
plt.ylabel('Broj slučajeva kardiovaskularnih bolesti')
plt.grid(True)
plt.show()
```

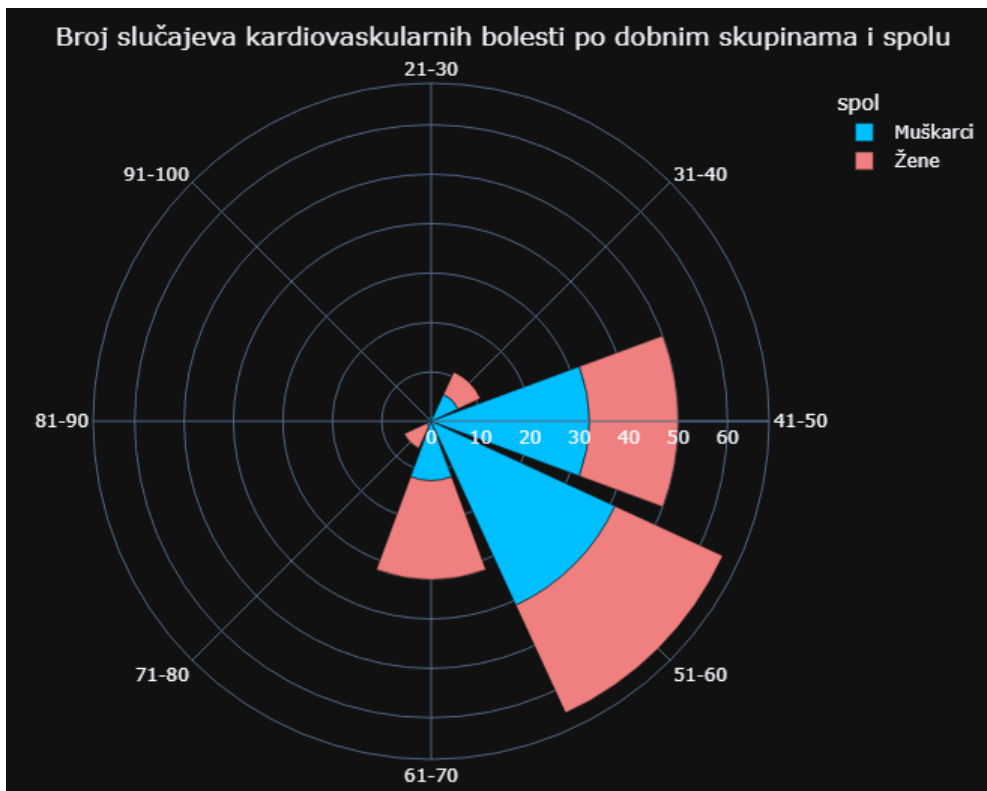
Slika 39 - Kod za graf 17; linijski grafikon

5.2.11. Broj slučajeva kardiovaskularnih bolesti po dobnim skupinama i spolu

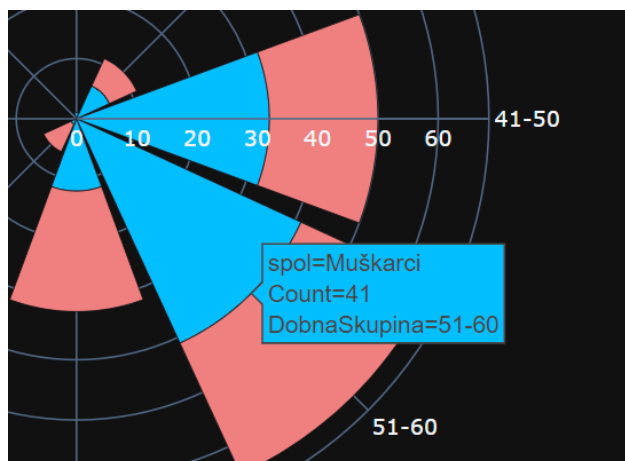
Vizualizacija broja slučajeva kardiovaskularnih bolesti po dobnim skupinama i spolu napravljena je sa bar polar grafikonom koji koristi polarne koordinate za prikaz broja slučajeva kardiovaskularnih bolesti po dobnim skupinama i spolu (Slik3 40, 41 i 42). Podaci za ovu vizualizaciju preuzeti su pod naslovom 'Heart Disease Predictions' sa stranice Kaggle [9]. Polarnim koordinatama vizualizacija omogućuje inovativni i privlačan prikaz podataka iz skupa podataka, pri čemu se različiti segmenti protežu iz središta grafa prema van.

Bar polar grafikon čine tri ključna elementa: theta (kut), R (radijus) i boje. Theta ili kut prikazuje dobne skupine u intervalima po deset godina, a od 21 do 100 godina. Radijus prikazuje broj slučajeva kardiovaskularnih bolesti unutar svake prisutne dobne skupine, dok su bojama predstavljeni muški i ženski spol.

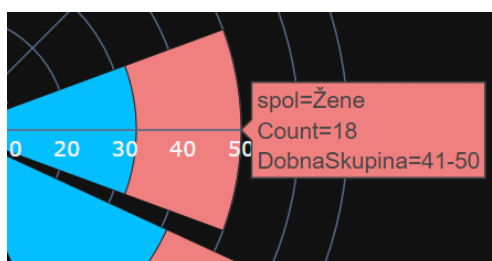
Rezultati analize vizualizacije pokazuju da dobna skupina između 51 i 60 godina ima najveći broj slučajeva kardiovaskularnih bolesti za oba spola. U dobnjoj skupini od 41 do 50 godina, zabilježen je veći broj kardiovaskularnih bolesti kod muškaraca nego kod žena. Za ostale dobne skupine postoje određene varijacije, ali broj slučajeva je manji za oba spola. Dobna skupina od 71 do 80 godina bilježi veći broj slučajeva kardiovaskularnih bolesti kod žena nego kod muškaraca.



Slika 40 - Bar polar grafikon; slučajevi po dobnjoj skupini i spolu (graf 18)



Slika 41 - Bar polar grafikon; interaktivnost; muškarci



Slika 42 - Bar polar grafikon; interaktivnost; žene

Prvi korak za izradu bar polar grafikona jest učitavanje podataka iz CSV datoteke i pohranjivanje istih u DataFrame. Najprije je potrebno kao i kod prethodnih primjera pretvoriti godine ispitanika u dobne skupine pomoću 'pd.cut()' funkcija koja omogućuje grupiranje po dobnim kategorijama. Numeričke vrijednosti u stupcu 'sex' zamjenjuju se tekstualnim oznakama na način da 0 postaje 'žene', a 1 'muškarci'. Podaci se zatim filtriraju i grupiraju po dobnim skupinama i spolu. Podaci su filtrirani na način da je za slučajeve kardiovaskularnih bolesti target==1. Slijedi zbrajanje slučajeva za svaku kombinaciju. Za bolje razlikovanje boja, u kodu su definirane boje između spolova pa tako 'deepskyblue' boja se koristi za muškarce, dok 'lightcoral' boja za žene.

Kreacija bar polar grafikona omogućena je plotly.express bibliotekom i definirana je unutar 'px.bar_polar' funkcije. Cjelokupni kod za izradu vizualizacije prikazan je na slici 43.

```
import pandas as pd
import plotly.express as px

# Učitaj podatke
data_path = '/content/heart.csv'
data = pd.read_csv(data_path)

# Kreiraj dobne skupine
bins = [20, 30, 40, 50, 60, 70, 80, 90, 100]
labels = ['21-30', '31-40', '41-50', '51-60', '61-70', '71-80', '81-90', '91-100']
data['DobnaSkupina'] = pd.cut(data['age'], bins=bins, labels=labels, right=False)

# Zamijeni numeričke oznake za spol tekstualnim oznakama
data['spol'] = data['sex'].replace({0: 'Žene', 1: 'Muškarci'})

# Filtriraj podatke za slučajeve bolesti srca
filtered_data = data[data['target'] == 1].copy()

# Grupiranje podataka po dobnim skupinama i spolu
grouped_data = filtered_data.groupby(['DobnaSkupina', 'spol']).size().reset_index(name='Count')

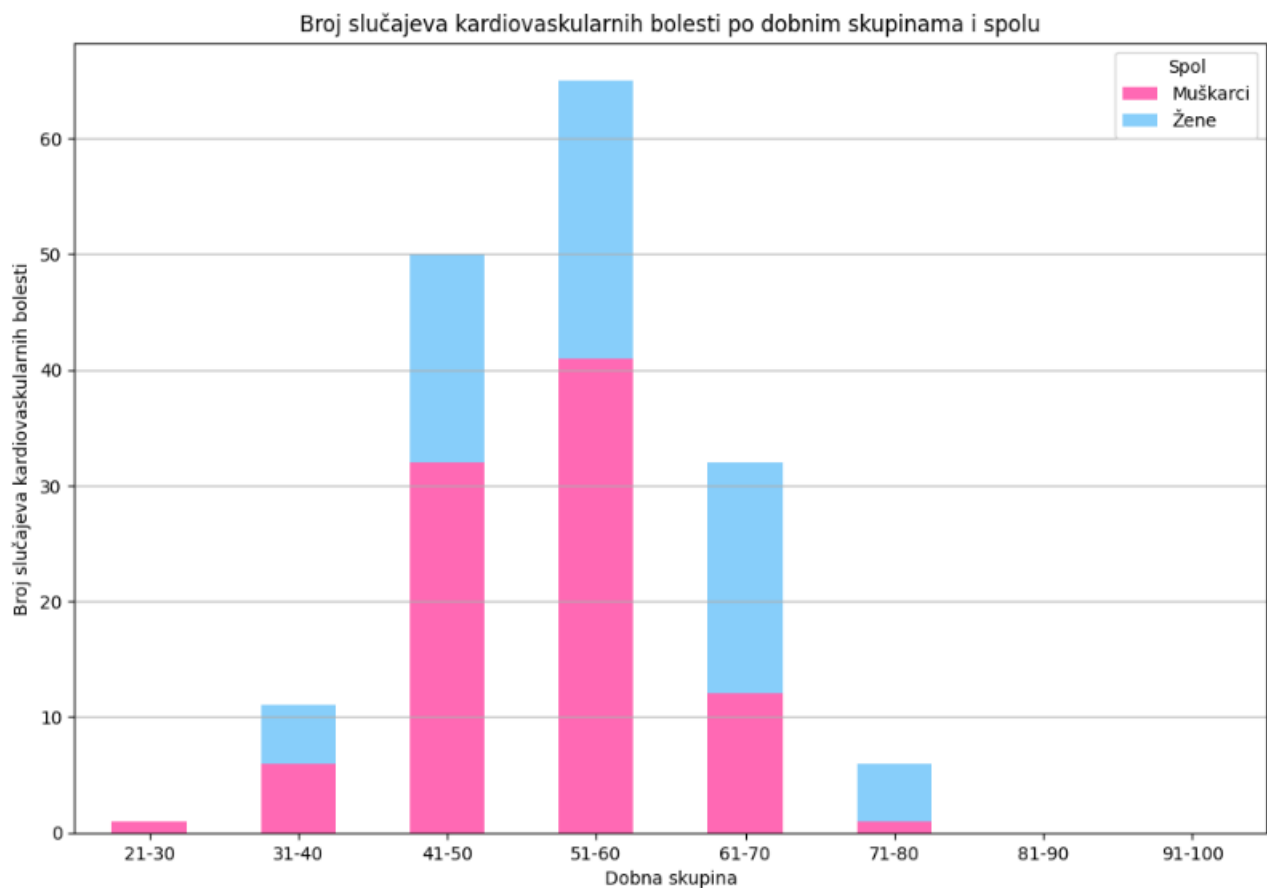
# Definiranje boja
color_discrete_map = {'Muškarci': 'deepskyblue', 'Žene': 'lightcoral'}

# Kreiraj bar polar grafikon
fig = px.bar_polar(grouped_data, r='Count', theta='DobnaSkupina', color='spol',
                  color_discrete_map=color_discrete_map,
                  title='Broj slučajeva kardiovaskularnih bolesti po dobnim skupinama i spolu',
                  template='plotly_dark')

fig.update_layout(margin=dict(t=50, l=25, r=25, b=25))
fig.show()
```

Slika 43 - Kod za graf 18; bar polar grafikon

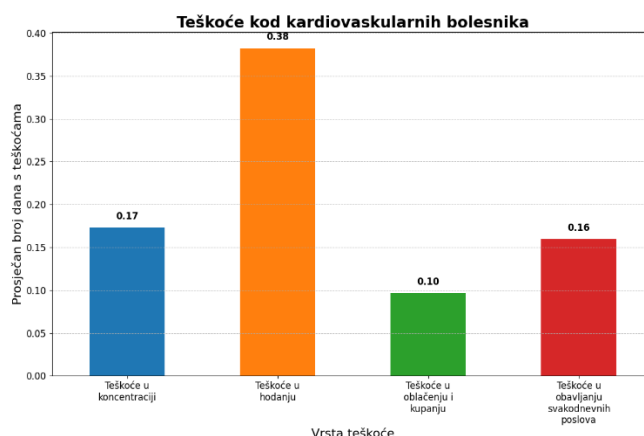
Prikaz slučajeva po dobnim skupinama i spolu napravljen je i pomoću stupčastog grafikona, pri čemu je rozom bojom označen ženski spol, a plavom bojom muški spol. Stupčasti grafikon (Slika 44) daje jednake rezultate analize s malo preglednijim i jasnijim izgledom.



Slika 44 - Stupčasti grafikon; slučajevi bolesti prema dobnj skupini i spolu (graf 19)

5.2.12. Prisutne poteškoće kod kardiovaskularnih bolesnika

Stupčasti grafikon u nastavku (Slika 45) prikazuje prosječne vrijednosti različitih vrsta teškoća kod pacijenata koji su proživjeli srčani udar. Podaci koji su korišteni za analizu preuzeti su sa stranice OurWorldInData, a konkretno su to podaci o općim karakteristikama kardiovaskularnih bolesnika [8]. Rezultati analize ukazuju da pacijenti proživljavaju 0.17 dana s poteškoćom nedostatka koncentracije. Teškoća koja je najzastupljenija jest teškoća u hodanju koja je prisutna prosječno 0.38 dana, što sugerira na značajne probleme mobilnosti kardiovaskularnih bolesnika. Najmanje poteškoća pacijenti su prijavili prilikom oblačenja i obavljanja higijene, a teškoće prilikom obavljanja svakodnevnih poslova prisutne su 0.16 dana. Vrijednosti su u grafu prikazane kao decimalni brojevi između 0 i 1, što odgovara postocima kada se pomnoži sa 100. Analiza se odnosi na period od godinu dana, pa tako 17% dana u godini dana pacijenti proživljavaju poteškoće sa koncentracijom, 38% dana poteškoće sa hodanjem, 10% dana poteškoće u oblačenju i kupanju, a 16% dana doživljavaju poteškoće u obavljanju svakodnevnih poslova.

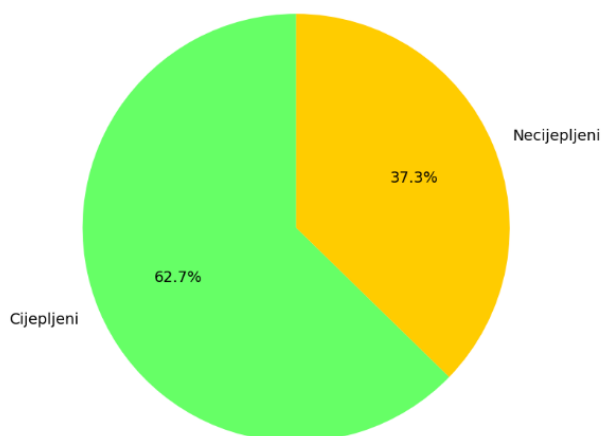


Slika 45 - Stupčasti grafikon; poteškoće s kojim se susreću srčani bolesnici (graf 22)

5.2.13. Povezanost cijepliva protiv gripe sa srčanim udarom

Pie chart u nastavku (Slika 46) prikazuje povezanost između cijepljenja protiv gripe i srčanog udara. Podaci koji su korišteni za analizu preuzeti su sa stranice OurWorldInData, a konkretno su to podaci o općim karakteristikama kardiovaskularnih bolesnika [8]. Iz analize možemo zaključiti da 62.7% pacijenata koji su cijepljeni protiv gripe su i preboljeli srčani udar, dok 37.3% pacijenata koji su imali srčani udar nisu bili cijepljeni protiv gripe.

Povezanost cijepljenja protiv gripe sa srčanim udarom



Slika 46 - Pie chart; cjevivo protiv gripe i srčani udar (graf 23)

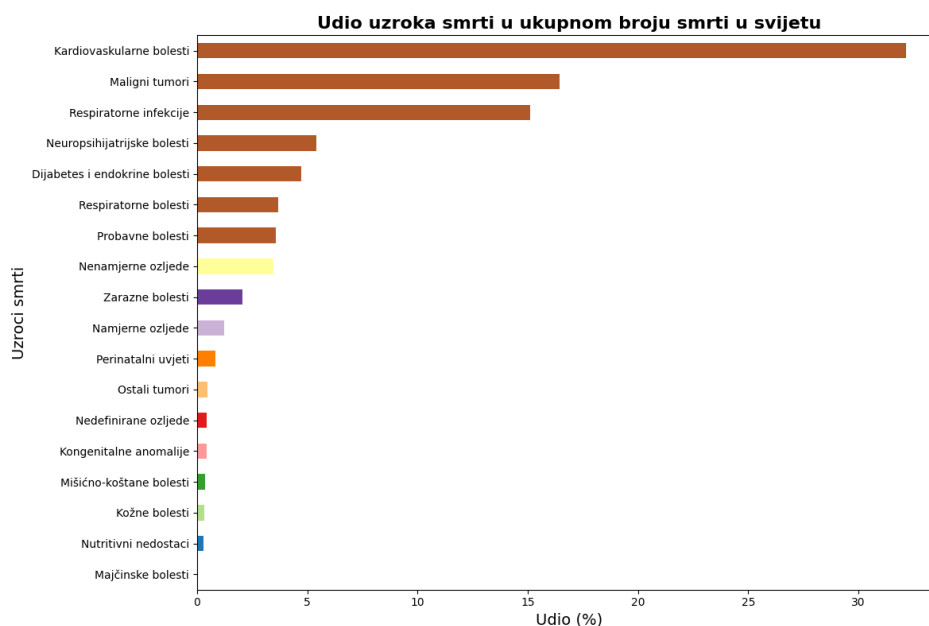
5.3. Mogućnosti dijagnostike, prevencije i liječenja kardiovaskularnih bolesti – vizualizacija u Tableau

Tableau je vodeći alat za vizualizaciju podataka koji sirove podatke prevodi u interaktivne i razumljive prikaze. Omogućuje analizu podataka, pomaže doći do često skrivenih zaključaka i rezultate prezentira na vrlo jasan i lako razumljiv način [17].

5.3.1. Udio uzroka smrti od kardiovaskularnih bolesti u ukupnom broju smrti u svijetu

Koliko su kardiovaskularne bolesti prisutne u svijetu i u kojoj mjeri, najbolje govori horizontalni stupčasti grafikoni u nastavku (Slika 47). Iz skupa podataka sa najčešćim uzročnicima smrti, kardiovaskularne bolesti zauzimaju najveći udio. Obzirom na rezultate, izuzetno je važno naglasiti važnost prevencije, rane dijagnoze i liječenja kardiovaskularnih bolesti kako bi se smanjila stopa smrtnosti.

Drugi najčešći uzročnik su maligni tumori i respiratorne infekcije, a veliki postotak smrtnosti uzrokuju i uzročnici poput neuropsihijatrijskih bolesti te šećernih, endokrinih, respiratornih i probavnih bolesti.



Slika 47 - Horizontalni stupčasti grafikoni; uzročnici smrti u svijetu

5.3.2. Broj kardiologa po zemljama u svijetu – Tableau vizualizacija

U ovome primjeru demonstrirati će se korištenje Tableau alata za kreiranje napredne vizualizacije podataka u formi interaktivne karte svijeta koja će prikazivati ukupan broj kardiologa na milijun stanovnika svake od zemalja dostupnih iz skupa podataka.

Kako bi se izradila vizualizacija u Tableau, najprije je bilo potrebno napraviti račun i prijaviti se sa njime. Nakon prijave, uslijedilo je preuzimanje Tableau Desktop inačice kako bi se mogla započeti izrada vizualizacije.

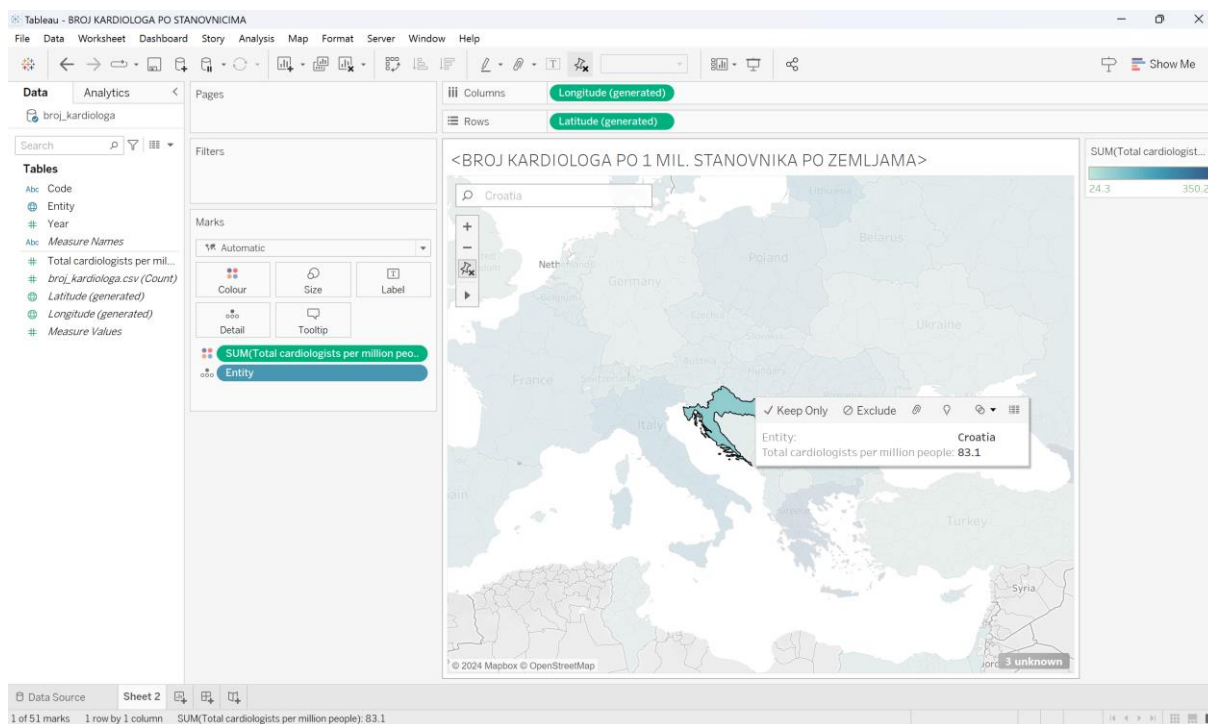
Neposredno nakon preuzimanja Tableau Desktop-a, potrebno je bilo ponovo se ulogirati svojim korisničkim računom te zatim otvoriti novu radnu knjigu za vizualizaciju.

Nakon otvaranja radne knjige, potrebno je povezati se sa CSV datotekom koja sadrži sirove podatke koje želimo vizualizirati, u ovome slučaju to je datoteka 'broj_kardiologa.csv'. Datoteka je preuzeta kao skup podataka sa stranice OurWorldInData, konkretno pod nazivom 'Cardiologists per million people, 2021' [8].

Nakon što su podaci učitani i nakon što je provjereno jesu li stupci ispravno prepoznati, stupac 'Entity' prepoznat je kao dimenzija koja sadrži nazive država koje analiziramo, a stupac 'Total cardiologists per milion people' kao mjera koja prikazuje broj kardiologa za svaku državu.

Stupac 'Entity' potrebno je povući na radni prostor i desnim klikom odabrati 'Geographic role' pa zatim 'Country/Region' kako bi alat Tableau prepoznao da se radi o državama. Da bi države bile obojane bojama prema broju kardiologa, stupac 'total cardiologist per milion people' premješten je na polje 'Color'. 'Show me' panel omogućuje odabir ikone mape kako bi se uspješno kreirala geografska mapa.

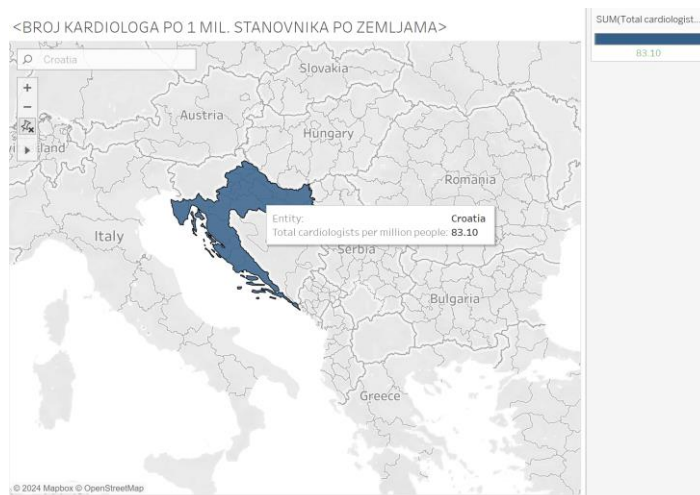
Kao rezultat opisanog postupka, pojavljuje se interaktivna karta koja prikazuje broj kardiologa na milijun stanovnika za svaku državu iz skupa podataka (Slike 48, 50). Karta omogućuje jasan i razumljiv prikaz podataka te bolje razumijevanje distribucije kardiologa unutar svijeta.



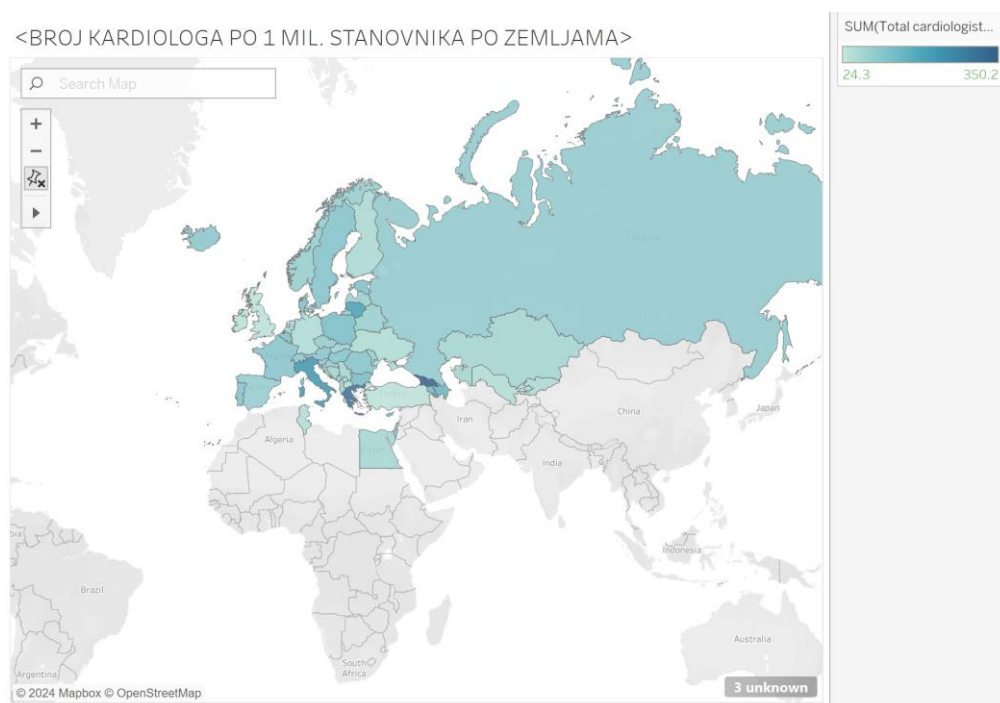
Slika 48 - Tableau vizualizacija; izgled sučelja, raspodjela stupaca i krajnji rezultat

Vizualizacije prikazane na Tableau karti jasno ukazuju na manjak broja kardiologa po milijun stanovnika u različitim zemljama. Na prvoj uvećanoj vizualizaciji (Slika 49) možemo vidjeti kako Hrvatska ima 83,10 kardiologa po milijun stanovnika. Međutim, promatrajući ostale zemlje, posebice zemlje istočne Europe i dijelova Azije, vidljivo je kako oni imaju daleko manji broj kardiologa po milijun stanovnika (npr. Kazahstan sa 61 kardiologom na milijun stanovnika).

Obzirom na vizualizaciju iz poglavlja 5.3.1. (Slika 51) koja nam prikazuje kako najviše smrtnosti u svijetu ima upravo od kardiovaskularnih bolesti, ovaj nedostatak stručnjaka može imati ozbiljne posljedice za prevenciju i liječenje kardiovaskularnih bolesti. Zemlje koje imaju niži broj kardiologa nemaju dovoljno specijalističkih kapaciteta za adekvatno liječenje kardiovaskularnih bolesnika što posljedično rezultira većom stopom smrtnosti za tu zemlju.



Slika 49 - Broj kardiologa/mil. stanovnika za Hrvatsku



Slika 50 – Prikaz broja kardiologa za neke od zemalja svijeta

5.3.3. Broj kardiovaskularnih kirurga na milijun ljudi po državama

Za kreaciju vizualizacije o broju kardiovaskularnih kirurga na milijun stanovnika po državama (Slika 52) koristeći alat Tableau, postupak je sljedeći. Prvo je potrebno pokrenuti Tableau Desktop alat na vašem računalu. Nakon što se alat pokrene, povežite se s podacima koje želite analizirati i vizualizirati. U ovom slučaju, radi se o datoteci cardiac-surgeons-per-million-people.csv koja je preuzeta sa stranice OurWorldInData [8]. Da biste se povezali s ovim skupom podataka, potrebno je kliknuti na 'File' > 'Open' i odabrati tu datoteku iz repozitorija u kojem se nalazi.

Nakon što su podaci učitani, pregled istih bit će vidljiv u kartici 'Data Source'. Unutar ove kartice, lista na dnu sučelja omogućuje da se otvori nova radna knjiga za vizualizaciju. Kada je radnja knjiga otvorena, podaci se dodaju u vizualizaciju na način da se polje koje predstavlja

države (Entity) povuče u područje sučelja 'Rows' (redovi), dok se polje koje predstavlja broj kardiokirurga po milijunu ljudi povuče u polje 'Columns' (stupci).

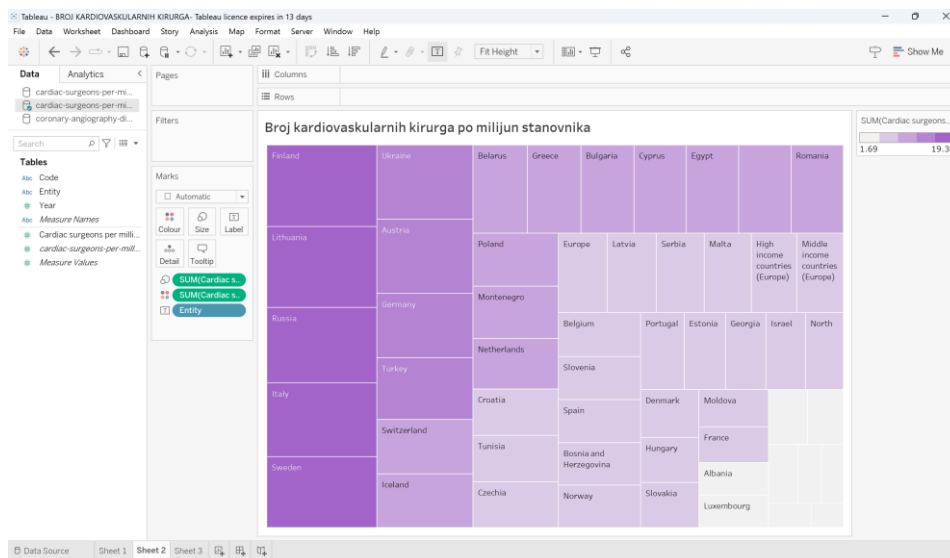
Klikom na karticu 'Show me', koja se nalazi u gornjem desnom kutu sučelja alata, prikazane su različite mogućnosti prikaza vizualizacije. Za ovaj primjer odabrana je vizualizacija 'tree map' koja koristi spektar različitih nijansi ljubičaste boje. Ovaj postupak omogućava vizualno pregledavanje podataka o broju kardiovaskularnih kirurga po državama, olakšavajući analizu i donošenje zaključaka.

Vizualizacija prikazuje broj kardiovaskularnih kirurga po milijun stanovnika pri čemu najtamnija nijansa ljubičaste boje prikazuje države s najvećim brojem kardiovaskularnih kirurga po milijun stanovnika, a najsvjetlija i bijela boja prikazuju države s najmanjim brojem kardiovaskularnih kirurga po milijun stanovnika.

Iz vizualizacije (Slika 52) je jasno vidljivo da postoji značajna varijabilnost u dostupnosti kardiovaskularnih kirurga među državama. Zemlje poput Finske i Latvije imaju najveći broj kirurga po milijun stanovnika, dok zemlje poput Albanije (5 kardiovaskularnih kirurga na milijun stanovnika) i Luksemburga (4 kardiovaskularna kirurga na milijun stanovnika) imaju znatno manji broj kardiovaskularnih kirurga.



Slika 51 - Broj kardiovaskularnih kirurga na milijun stanovnika u Hrvatskoj



Slika 52 - Broj kardiovaskularnih kirurga po milijun stanovnika - Tableau vizualizacija; Tree map

5.3.4. Koronarna angiografija i analiza stentova na milijun ljudi

Koronarna angiografija je medicinska procedura koja koristi rendgenske zrake za pregled krvnih žila srca. Ovim postupkom, liječnicima je omogućeno da identificiraju suženja krvnih žila ili blokade u koronarnim arterijama koje uzrokuju *anginu pectoris* (bol u psima) ili infarkt miokarda, odnosno srčani udar [18].

Skladno tome, koronarna angiografija jest medicinska procedura koja je od iznimne važnosti za prevenciju kardiovaskularnih bolesti jer njenim izvođenjem liječnici mogu sanirati problem suženja krvnih žila i blokada koronarnih arterija uvođenjem stentova, pa je samim time i pacijent spašen od srčanog udara.

Vizualizacija u nastavku (Slika 53) prikazuje broj izvođenja postupka koronarne angiografije i pregleda stentova istom na milijun stanovnika po državama.

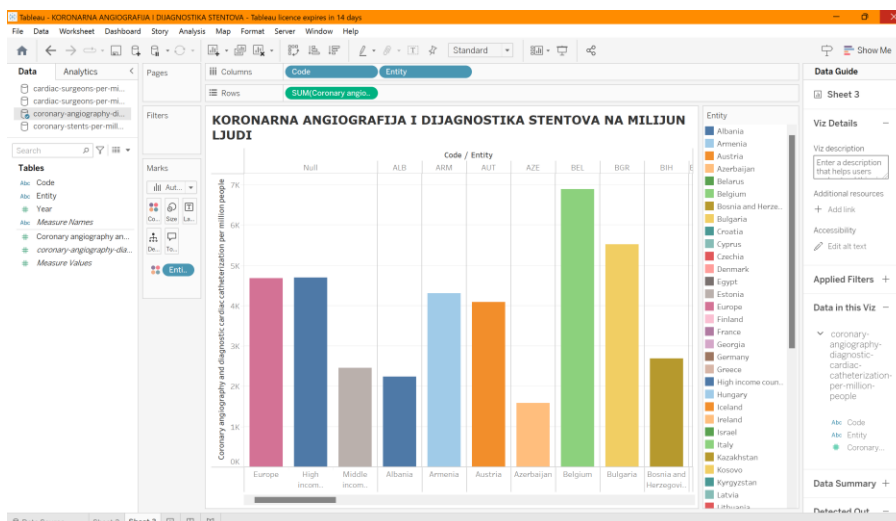
Vizualizacija je izrađena u Tableau alatu u formi stupčastog interaktivnog grafikona koji prikazuje broj postupaka koronarne angiografije na milijun stanovnika po državama. Povlačenjem miša po stupcima koji identificiraju države, vidljiv je broj koronarnih angiografija za svaku državu.

Izrada vizualizacije zahtijevala je jednak postupak kao i druga vizualizacija u ovome poglavlju, jedino su za vizualizaciju korišteni i učitani drugi podaci. Podaci za prikaz ove vizualizacije preuzeti su sa OurWorldInData stranice i nalaze se unutar *coronary-angiography-diagnostic-cardiac-catheterization-per-million-people.csv* datoteke [8].

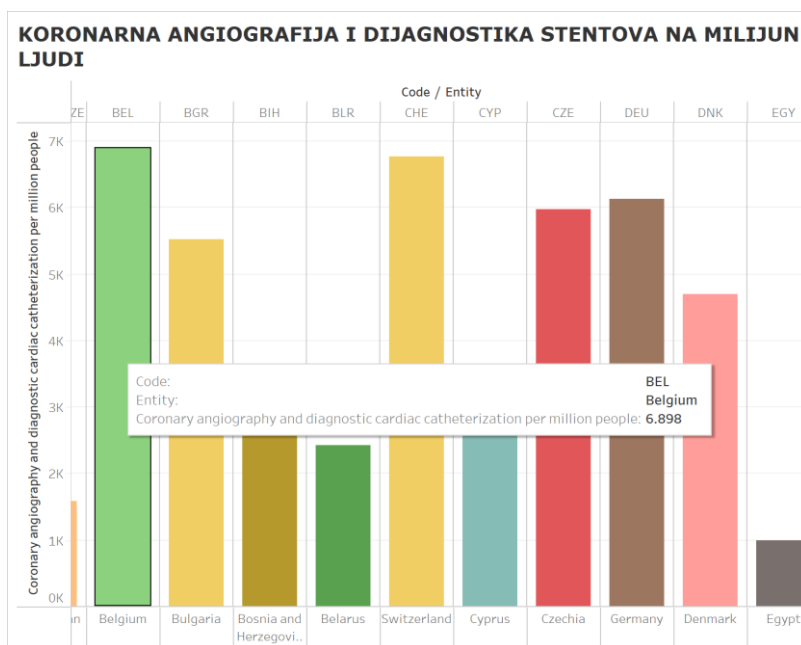
Također, u kartici 'Show me' je na kraju postupka odabran stupčasti grafikon kao željeni prikaz.

Iz vizualizacije možemo zaključiti da neke zemlje, poput Njemačke, Belgije (Slika 54) i Italije, imaju veći broj provedenih procedura na milijun ljudi. Druge zemlje, poput Hrvatske (3393) i Slovenije (5991) imaju manji broj provedenih procedura koronarne angiografije. Značajna je i razlika između Istočne i Zapadne Europe. Skandinavske zemlje poput Norveške (5770) i Švedske (1992) također pokazuju varijacije u broju procedura.

Visoka stopa provedenih procedura koronarne angiografije u nekim zemljama ukazuje u napredniju medicinsku infrastrukturu, a samim time i manji broj oboljelih od kardiovaskularnih bolesti.



Slika 53 - Koronarna angiografija i pregled stentova na milijun ljudi za neke od zemalja svijeta; Tableau vizualizacija; Stupčasti grafikon



Slika 54 - Koronarna angiografija i dijagnostika stentova na milijun ljudi u Belgiji

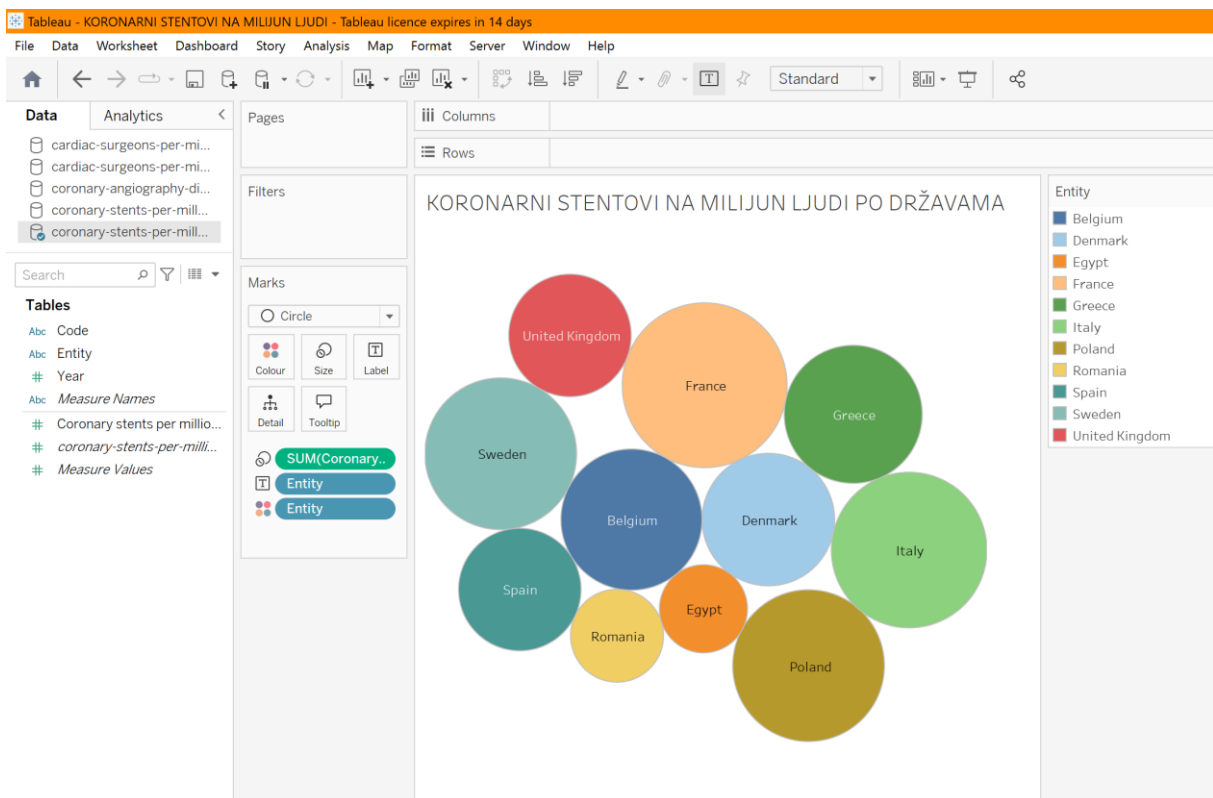
5.3.5. Broj koronarnih stentova na milijun ljudi

Vizualizacija u nastavku (Slika 55) napravljena je također u alatu Tableau i zove se 'packed bubbles' vizualizacija pri čemu veći krugovi predstavljaju veću vrijednost iz skupa podataka, a manji krugovi manju vrijednost.

U krugovima je različitim bojama vizualiziran broj ugrađenih koronarnih stentova na milijun stanovnika za neke od država svijeta koje se nalaze u skupu podataka iz datoteke *coronary-*

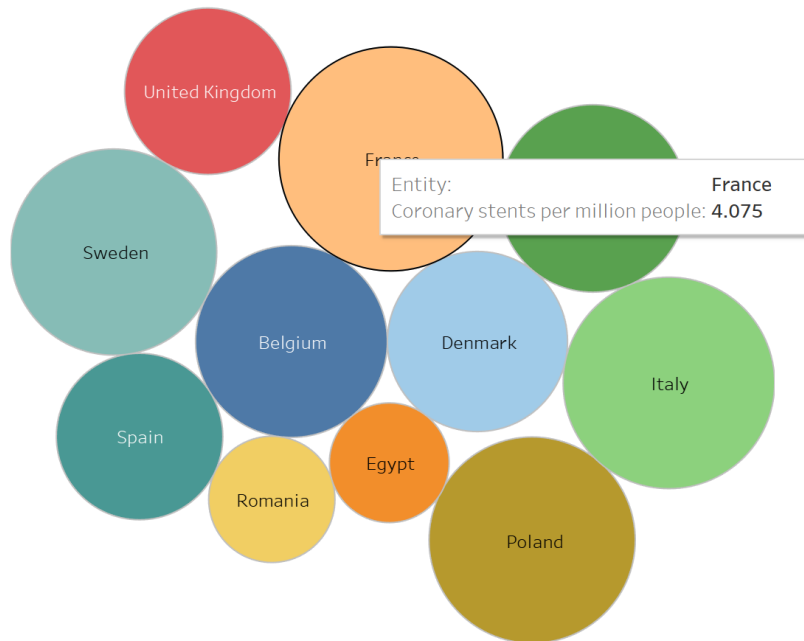
stents-per-million-people.csv [8]. Vizualizacija je napravljena istim postupkom kao prethodne dvije iz ovoga poglavlja, samo je u 'Show me' kartici odabran 'packed bubbles' grafikon za vizualizaciju.

Iz vizualizacije možemo zaključiti da je iz skupine država obuhvaćene podacima, Francuska (Slika 56) vodeća u ugradnji koronarnih stentova na milijun stanovnika (4075 stentova/mil. stanovnika) što ukazuje na velik broj kardiovaskularnih bolesnika, ali i veliki broj prevencija srčanih udara ugradnjom stentova. Egipat bilježi najmanji broj ugradnje stentova na milijun stanovnika (1165/mil. stanovnika).



Slika 55 - Koronarni stentovi na milijun stanovnika

KORONARNI STENTOVI NA MILIJUN LJUDI PO DRŽAVAMA



Slika 56 - Koronarni stentovi na milijun stanovnika u Francuskoj

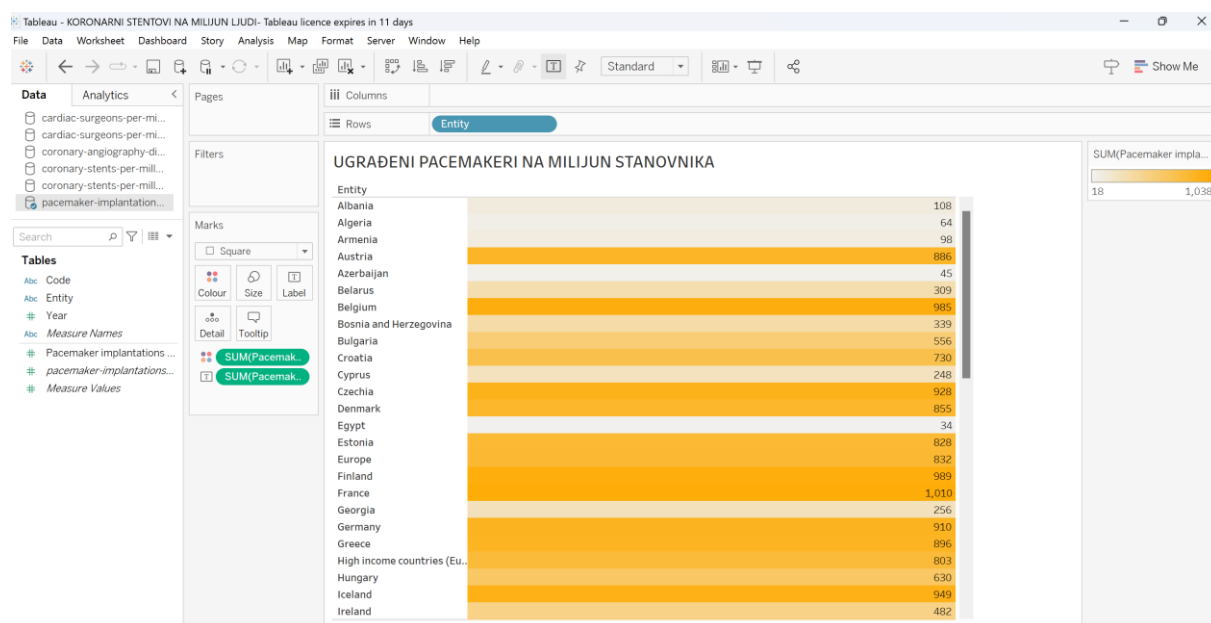
5.3.6. Ugrađeni pacemakeri na milijun stanovnika

Pacemaker ili kardioverter defibrilator je medicinski uređaj koji se koristi za reguliranje srčanog ritma kod osoba koje nemaju pravilan srčani ritam, koje su proživjele srčani udar ili srčani zastoj [19]. Uređaj funkcionira na način da šalje električne impulse koji pomažu srcu da održava ritam i brzinu otkucaja. Ugrađuje se kirurškim putem u predjelu prsnog koša, i povezan je sa srcem putem elektroda koje prenose impulse direktno u srčani mišić. U slučaju srčanog udara ili srčanog zastoja, uređaj se aktivira. Pacemaker je od izuzetne važnosti za prevenciju daljnjih srčanih udara ili zastoja kod već oboljelih osoba od kardiovaskularnih bolesti.

Vizualizacija (Slika 57) je izrađena u alatu Tableau i koristi podatke sa OurWorldInData stranice iz datotekepacemaker-implantations-per-million-people.csv [8]. Vizualizacija je napravljena standardnim postupkom izrade kao i kod prethodnih primjera, te je u formi horizontalnog grafikona koji spektrom boja izražava veličinu određenog podatka u skupu podataka. Tako je najtamnijom nijansom narančaste boje označen broj država iz skupa podataka s najvećim brojem ugrađenih pacemakera, dok je najsvjetlijom nijansom i bijelom bojom označen broj država s najmanjim brojem ugrađenih pacemakera.

Izradom vizualizacije, jasno je vidljivo da Švedska uvelike prednjači s brojem ugrađenih pacemakera (1038/mil. stanovnika), a slijede je Portugal, Belgija i Italija s nešto manjim brojem ugrađenih pacemakera [20].

Najmanje ugrađenih pacemakera po milijun stanovnika iz država koje su obuhvaćene podacima imaju Kosovo, Kirgistan, Kazahstan, Egipat i Uzbekistan. Taj rezultat možemo povezati i sa vizualizacijom o broju kardiologa (Slika 52) i kardiovaskularnih kirurga (Slika 56) na milijun stanovnika koji su potrebni da bi se ovakvi zahvati izvršili.



Slika 57 - Pacemakeri na milijun stanovnika

7. Zaključak

Na temelju istraživanja provedenog u ovome radu, jasno je da analiza i vizualizacija podataka igraju jednu od ključnih uloga za pomoć pri prevenciji kardiovaskularnih bolesti. Upotrebom informatičkih alata omogućena je detaljna obrada velikih količina podataka, što doprinosi boljem razumijevanju uzročnika i faktora rizika povezanih s ovim oboljenjem.

Različitim metodama vizualizacije podataka, poput linijskih grafikona, toplinskih grafikona, tortnih grafikona, dijagrama raspršivanja i ostalih vrsta grafikona, moguće je jasnije i razumljivije prikazati ključne trendove i indikatore za razvitak kardiovaskularnih bolesti. Analizom takvih podataka, informatički alati uvelike pomažu u donošenju odluka u medicinskoj praksi.

Važno je istaknuti i značaj alata poput alata Tableau, koji omogućava stvaranje interaktivnih vizualizacija, čime je olakšana interpretacija kompleksnih informacija o pacijentima, a u svrhu prevencije bolesti.

Kroz analizu trendova incidencije kardiovaskularnih bolesti u posljednjem desetljeću, identificirane su značajne varijacije među različitim zemljama i regijama, kao i utjecaj pandemije COVID-19 na povećanje broja smrtnih slučajeva. Vizualizacija podataka također je jasnije prikazala korelacije između različitih zdravstvenih indikatora, poput lošeg fizičkog i mentalnog zdravlja, broja sati sna, BMI-a i povijesti srčanog udara.

Zaključno, integracijom analitičkih metoda i vizualizacijskih tehnika u medicinu, omogućeno je dublje razumijevanje složenih stanja pacijenata, a samim time poboljšane su preventivne strategije i optimizirane kardiološke intervencije.

8. Popis literature

- [1] Peters MDJ, Godfrey C, McInerney P, Munn Z, Tricco AC, i Khalil H., „ Data visualisation in scoping reviews and evidence maps on health topics: a cross-sectional analysis. Syst Rev.“, 2021.
- [2] Xu Y, Min W, i Zhou Q, „Editorial: Biomedical Data Visualization: Methods and Applications. Front Med (Lausanne)“, 2021.
- [3] Noor U, Muhammad R, Ali SA, Sarwar R, i Abbas A, „Primjena analitike velikih podataka i organizacijska izvedba: posrednička uloga praksi upravljanja znanjem. J Big Data.“, 2020.
- [4] Cho SMJ, Koyama S, i Honigberg MC, „Genetic, sociodemographic, lifestyle, and clinical risk factors of recurrent coronary artery disease events: a population-based cohort study. Eur Heart J.“, 2023.
- [5] Newby DE, Mannucci PM, i Tell GS, „Expert position paper on air pollution and cardiovascular disease. Eur Heart J. “, 2024.
- [6] Ahmet Erdemir *i ostali*, „Credible practice of modeling and simulation in healthcare: ten rules from a multidisciplinary perspective. Journal of Translational Medicine“, 2020.
- [7] Chaithra N i Madhu B, „Analysis of Cardiovascular Diseases Data using Tableau Visualization Tool. Issues and Developments in Medicine and Medical Research.“, 2022.
- [8] Saloni Dattani, Veronika Samborska, Hannah Ritchie, i Max Roser, „Cardiovascular Diseases“, 2023.
- [9] Desalegngeb, „Heart Disease Predictions“, 2023.
- [10] Ala'a Abdu Saleh Alawdi, „Personal Key Indicators of Heart Disease“, 2022.
- [11] Isaac, „Google Collab ili Google Colaboratory: što je to“.
- [12] Inc. NumFOCUS, „pandas“, [Na internetu]. Dostupno na: <https://pandas.pydata.org/> (Pristupljeno 15.6.2024.)
- [13] „Matplotlib: Visualization with Python“, [Na internetu]. Dostupno na: <https://matplotlib.org/> (Pristupljeno 15.6.2024.)
- [14] Michael Waskom, „seaborn: statistical data visualization“, [Na internetu]. Dostupno na: <https://seaborn.pydata.org/> (Pristupljeno 15.6.2024.)
- [15] „Plotly Express in Python“, [Na internetu]. Dostupno na: <https://plotly.com/python/plotly-express/> (Pristupljeno 15.6.2024.)
- [16] „Graph Objects in Python“, [Na internetu]. Dostupno na: <https://plotly.com/python/graph-objects/> (Pristupljeno 15.6.2024.)
- [17] „What is Tableau?“, [Na internetu]. Dostupno na: <https://www.tableau.com/why-tableau/what-is-tableau> (Pristupljeno 20.06.2024.)

- [18] Anavekar NS, „Mayo Clinic coronary angiogram“, 2023.
- [19] Nikola Pavlović, „Pacemaker - elektrostimulacija srca“, 2013, [Na internetu]. Dostupno na: <https://www.plivazdravlje.hr/aktualno/clanak/24397/Pacemaker-elektrostimulacija-srca.html> (Pristupljeno 10.06.2024.)
- [20] „American Heart Association – Pacemaker“, 2022.

9. Popis slika

Slika 1 - Prikaz CSV datoteke za preuzimanje u sklopu interaktivnog grafa.....	6
Slika 2 - Heart Disease prediction skup podataka	7
Slika 3 - Indicators od Heart Disease skup podataka.....	7
Slika 4 - Linijski grafikon; Incidencija KVD-a tijekom posljednjih deset godina u svijetu (graf 1).....	8
Slika 5 - Kod za graf 1; linijski grafikon	9
Slika 6 - Toplinski grafikon; Incidencija po nekim od zemalja svijeta (graf 2)	10
Slika 7 - Kod za graf 2; toplinski grafikon.....	11
Slika 8 - Pie chart; zemlje s najvišom incidencijom (graf 3).....	12
Slika 9 - Kod za graf 3; pie chart	13
Slika 10 - Složeni stupčasti grafikon - Razlike incidencije između Europe i Azije (graf 4).....	14
Slika 11 - Kod za graf 4 – složeni stupčasti grafikon	15
Slika 12 - Smrtnost od KVD-a (graf 5)	16
Slika 13 - Kod za graf 5; smrtnost uzrokovana KVD-om	17
Slika 14 Box plot - raspodjela incidencije po godinama (graf 6)	18
Slika 15 - Kod za graf 6; kutijasti grafikon.....	19
Slika 16 - Pie chart; proporcije općeg zdravlja (graf 7).....	21
Slika 17 - Kod za graf 7; pie chart	22
Slika 18 - Lollipop chart; dani lošeg fizičkog zdravlja (graf 8).....	23
Slika 19 - Kod za graf 8; lollipop chart	24
Slika 20 - Sunburst grafikon; opće zdravlje po spolovima (graf 9)	24
Slika 21 - Interaktivni sunburst.....	25
Slika 22 - Kod za graf 9; sunburst grafikon	26
Slika 23; 3D Scatter plot; odnosi BMI, san, loše fizičko zdravlje (graf 10).....	26
Slika 24 - interaktivnost 3D Scatter plot-a.....	27
Slika 25 - Kod za graf 10; 3D Scatter plot	27
Slika 26 - Pair plot grafikon; odnosi između indikatora bolesti (graf 11)	28
Slika 27 -Kod za graf 11; Pair plot.....	29
Slika 28 - Toplinski grafikon; korelacija indikatora (graf 12)	30
Slika 29 - Kod za graf 12; toplinski grafikon.....	31
Slika 30 - Toplinski grafikon; korelacija s više indikatora (graf 13).....	32
Slika 31 - Sankey grafikon; tok između bolesti srca i drugih bolesti i faktora (graf 14).....	33
Slika 32 - povezanost bolesti srca s moždanim udarom; Sankey grafikon	33
Slika 33 - interaktivnost; pomicanje traka; Sankey grafikon	34
Slika 34 - Kod za graf 14; Sankey grafikon.....	35
Slika 35 - Stupčasti grafikon; bolesti po spolu i statusu pušenja (graf 15)	36
Slika 36 - Kod za graf 15; stupčasti grafikon	37
Slika 37 - Horizontalni stacked bar; pušenje cigareta (graf 16).....	37
Slika 38 - Linijski grafikon; trend bolesti po dobnim skupinama (graf 17)	38
Slika 39 - Kod za graf 17; linijski grafikon	39
Slika 40 - Bar polar grafikon; slučajevi po dobnoj skupini i spolu (graf 18).....	40
Slika 41 - Bar polar grafikon; interaktivnost; muškarci	40
Slika 42 - Bar polar grafikon; interaktivnost; žene	40
Slika 43 - Kod za graf 18; bar polar grafikon.....	41

Slika 44 - Stupčasti grafikon; slučajevi bolesti prema dobnoj skupini i spolu (graf 19).....	42
Slika 45 - Stpčasti grafikon; poteškoće s kojim se susreću srčani bolesnici (graf 22).....	43
Slika 46 - Pie chart; cjepivo protiv gripe i srčani udar (graf 23).....	43
Slika 47 - Horizontalni stupčasti grafikon; uzročnici smrti u svijetu.....	44
Slika 48 - Tableau vizualizacija; izgled sučelja, raspodjela stupaca i krajnji rezultat.....	46
Slika 49 - Broj kardiologa/mil. stanovnika za Hrvatsku	47
Slika 50 – Prikaz broja kardiologa za neke od zemalja svijeta.....	47
Slika 51 - Broj kardiovaskularnih kirurga na milijun stanovnika u Hrvatskoj	48
Slika 52 - Broj kardiovaskularnih kirurga po milijun stanovnika - Tableau vizualizacija; Tree map	48
Slika 53 - Koronarna angiografija i pregled stentova na milijun ljudi za neke od zemalja svijeta; Tableau vizualizacija; Stupčasti grafikon	50
Slika 54 - Koronarna angiografija i dijagnostika stentova na milijun ljudi u Belgiji.....	50
Slika 55 - Koronarni stentovi na milijun stanovnika	51
Slika 56 - Koronarni stentovi na milijun stanovnika u Francuskoj.....	52
Slika 57 - Pacemakeri na milijun stanovnika	53