

# Vizualizacija zdravstvenih podataka: mogućnosti i izazovi

---

**Bačić, Ana**

**Undergraduate thesis / Završni rad**

**2023**

*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:195:489809>

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

*Download date / Datum preuzimanja:* **2024-11-23**



*Repository / Repozitorij:*

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



Sveučilište u Rijeci – Fakultet informatike i digitalnih tehnologija

Preddiplomski studij informatike

Ana Bačić

# Vizualizacija zdravstvenih podataka: mogućnosti i izazovi

Završni rad

Mentor: doc. dr. sc. Lucia Načinović Prskalo

Rijeka, rujan 2023.

Rijeka, 20.4.2023.

## Zadatak za završni rad

Pristupnik: Ana Bačić

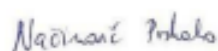
Naziv završnog rada: Vizualizacija zdravstvenih podataka: mogućnosti i izazovi

Naziv završnog rada na engleskom jeziku: Visualization of Health Data: Opportunities and Challenges

Sadržaj zadatka: Zadatak završnog rada je istražiti postojeće radove na temu vizualizacije zdravstvenih podataka te prednosti i nedostatke korištenja metoda vizualizacije u analizi zdravstvenih podataka. U radu će se također primijeniti metode vizualizacije na odabrani skup podataka iz zdravstva.

Mentor

Doc. dr. sc. Lucia Načinović Prskalo



Voditelj za završne radove

Doc. dr. sc. Miran Pobar



Zadatak preuzet: 20.4.2023.



(potpis pristupnika)

## Sadržaj

Sažetak .....	3
1. Uvod .....	4
2. Pregled literature .....	5
2.1. Dosadašnja istraživanja o vizualizaciji podataka i njenom utjecaju na donošenje odluka u zdravstvenim organizacijama .....	5
3. Metode i tehnike vizualizacije podataka u zdravstvenim organizacijama .....	7
3.1. Uvjeti za vizualizaciju.....	12
4. Prednosti i izazovi korištenja vizualizacije podataka u zdravstvu.....	13
4.1. Dobre strane vizualizacije podataka u zdravstvu .....	13
4.2. Izazovi vizualizacije podataka u zdravstvu.....	14
5. Analiza podataka, korištene metode i skupovi podataka.....	15
5.1. Kružni dijagram.....	18
5.2. Stupčasti dijagram .....	19
5.3. Geografska mapa .....	20
5.4. Dijagram raspršenosti .....	23
6. Diskusija .....	25
6.1. Glavni zaključci vlastite vizualizacije .....	25
6.2. Prijedlozi za buduća istraživanja .....	25
7. Zaključak .....	27
8. Popis literature .....	28
9. Popis slika.....	31
10. Prilozi .....	33

## Sažetak

Tema ovog završnog rada je vizualizacija podataka u zdravstvu. Napravljena je analiza dosadašnjih istraživanja o vizualizaciji podataka i njenom utjecaju na donošenje odluka u zdravstvenim organizacijama. Uz mogućnosti koje ona pruža, navedeni su i izazovi koji se susreću. Nadalje, naveden je pregled dostupnih metoda i tehnika koje se koriste za vizualizaciju te dosadašnjih primjena vizualizacije podataka u zdravstvu. Na praktičnim primjerima prikazano je kako se vizualizacija može koristiti za donošenje zaključaka. Za kraj, navedeni su prijedlozi za maksimalno iskorištavanje potencijala vizualizacije u budućim istraživanjima.

**Ključne riječi:** vizualizacija podataka, podaci velikog obujma, zdravstveni sektor, COVID, grafovi, izazovi vizualizacije, Kaggle, Google Colab, Matplotlib, Seaborn, Pandas, GeoPandas, kružni dijagram, stupčasti dijagram, dijagram raspršenosti, geografska mapa

## 1. Uvod

U današnjem digitalnom svijetu, količina informacija sve više raste zajedno sa širenjem upotrebe interneta. Upravo zbog toga, velika količina podataka se konstantno generira te je poznata pod nazivom „Big Data“. Ti su podaci vrlo zahtjevni za obradu korištenjem konvencionalnih sustava za upravljanje bazom podataka. Oni su skup podataka velikog volumena, velike brzine i raznolikosti te zahtijevaju novi stil obrade kako bi se olakšalo donošenje odluka i istraživanje znanja i optimizacija tehnika. Upravo zbog toga, jedna od važnijih tehnologija koja se razvija u ovom stoljeću je analiza tih podataka. Zbog velikog uspjeha analize podataka velikog obujma, zdravstveni sektor se počeo sve više usmjeravati prema usvajanju ovih pristupa (Tenali & Babu, 2023).

Vizualizacija podataka u zdravstvu je predstavljanje složenih medicinskih statističkih podataka u grafičkom formatu za njihovo lakše shvaćanje i donošenje bolje informiranih odluka. Korisnicima se omogućuje jednostavno dobivanje kvalitativnih uvida, promjena, obrazaca i trendova dok pokušavaju doći do vrijednih zaključaka (Khrstich, 2023).

Zdravstveni sektor oduvijek je generirao veliku količinu podataka zbog povećanih potreba vođenja evidencije o skrbi pacijenata. Pošto je velik dio tih podataka u polustrukturiranom ili nestrukturiranom obliku, raznoliki su i dinamično se mijenjaju, izvlačenje vrijednih uvida korištenjem tradicionalnih metoda je izazovno. Ljudski kapacitet za obradu ovih podataka je ograničen te time dolazi do potrebne učinkovite podrške pri donošenju odluka. Digitalizacija kliničkih pregleda i medicinske dokumentacije u zdravstvenim sustavima postala je široko rasprostranjena i prihvaćena norma od početka razvoja računalnih sustava (Berros, Mendili, Filaly, & Idrissi, 2023).

U ovom radu objasniti će se značajke zdravstvenih podataka, njihove posebnosti i dostupni alati za njihovo korištenje. Također, osvrnut će se na tehničke i organizacijske izazove o kojima se raspravlja.

U sljedećem poglavlju dan je pregled literature dosadašnjih istraživanja. U poglavlju 3 navedene su i opisane metode za vizualizaciju te su za svaku navedeni primjeri. Prednosti i izazovi korištenja vizualizacije u zdravstvu opisani su u poglavlju 4. U 5. poglavlju je izrađeno nekoliko vizualizacija zdravstvenih podataka za odabrani skup podataka. Diskusija o vizualizaciji podataka u zdravstvu dana je u poglavlju 6, dok je zaključak naveden u 7. poglavlju.

## 2. Pregled literature

Analitika podataka velikog obujma može pretvoriti velike količine kontinuiranih podataka u korisne uvide analizom i povezivanjem informacija iz više izvora. Sposobnost pružanja ove vrste uvida posebno je bitna u hitnim medicinskim situacijama jer može uvelike utjecati na pozitivan ili negativan ishod po život pacijenta (Tenali & Babu, 2023).

Analiza dosadašnjih istraživanja o vizualizaciji podataka i njenom utjecaju na donošenje odluka u zdravstvenim organizacijama pokazuje da postoji veliki potencijal za primjenu vizualizacije podataka u zdravstvenom sektoru. Učinkovita vizualizacija podataka može pomoći zdravstvenim profesionalcima da brzo i lako interpretiraju složene podatke i donose bolje informirane odluke.

Prema istraživanjima, vizualizacija podataka u zdravstvu je korisna u različitim područjima, kao što su praćenje epidemija, praćenje pacijentovog zdravstvenog stanja, praćenje korištenja resursa u bolnicama i sl. U ovim područjima, vizualizacija podataka može pomoći u razumijevanju i identificiranju trendova, odnosno promjena u podacima.

Jedno istraživanje provedeno na osoblju bolnice pokazalo je da vizualizacija podataka o pacijentovom zdravstvenom stanju može pomoći bržem i preciznijem donošenju odluka o dijagnozi i liječenju (Berros, Mendili, Filaly, & Idrissi, 2023). Drugo istraživanje je pokazalo da vizualizacija podataka može pomoći u identificiranju i praćenju epidemija, što može biti od velike važnosti u hitnim situacijama (Tenali & Babu, 2023).

Međutim, dosadašnja istraživanja također su ukazala na nekoliko izazova u primjeni vizualizacije podataka u zdravstvenom sektoru. Neki od tih izazova uključuju teškoće u prikupljanju i obradi velike količine podataka, nedostatak standardiziranih formata podataka, nedostatak standardiziranih metoda za vizualizaciju podataka i nedostatak obuke zdravstvenih profesionalaca za rad s vizualizacijom podataka (Berros, Mendili, Filaly, & Idrissi, 2023), (Khristich, 2023).

Unatoč ovim izazovima, vizualizacija podataka i dalje predstavlja važan alat za poboljšanje donošenja odluka u zdravstvenom sektoru. Potrebno je nastaviti sa istraživanjem u ovom području kako bi se bolje razumjelo kako vizualizacija podataka može biti korištena na najbolji način u zdravstvenim organizacijama.

### 2.1. Dosadašnja istraživanja o vizualizaciji podataka i njenom utjecaju na donošenje odluka u zdravstvenim organizacijama

Analiza podataka velikog obujma u zdravstvu slijedi strukturu koja je uglavnom ista kao tradicionalna zdravstvena informatika. Neovisni sustavi s instaliranom naprednom analitikom često su se koristili u zdravstvenim inicijativama.

Za dijagnozu bolesti koja se smatra najsmrtonosnijom bolešću na cijelom svijetu, bolesti srca, koristi se tehnologija koja se sastoji od dvije osnovne komponente. Prva je pohranjivanje podataka i vizualizacija, a druga je *streaming* obrada. Bolesti srca ozbiljan su zdravstveni problem i vodeći uzrok smrti u svijetu. Rana dijagnoza bolesti srca postala je vrlo važna u području medicinskih istraživanja (Tenali & Babu, 2023).

Istraživanja su također navela da bi analiza i vizualizacija podataka bila korisna za klasifikaciju bolesti bubrega, tumora na mozgu te otkrivanju dijabetesa (Tenali & Babu, 2023).

Analiza i vizualizacija podataka imala je veliki utjecaj za upravljanje pandemijskim uvjetima, kao što je predviđanje epidemije COVID-a i prepoznavanje slučajeva i širenje istog. Još uvijek se provode istraživanja o korištenju analitike podataka velikog obujma za predviđanje bolesti COVID-19. Međutim, još uvijek nedostaje precizna i rana identifikacija bolesti zbog količine medicinske dokumentacije poput različitih modaliteta medicinskog snimanja. U međuvremenu, digitalna slika sada je postala ključna za dijagnostiku COVID-a. Kako bi se epidemija učinkovito kontrolirala, važno je razumjeti njezine karakteristike i ponašanje. To znanje može se postići prikupljanjem i analizom relevantnih podataka velikog obujma. Uzimajući u obzir veliku količinu podataka o COVID-u koji su dostupni iz nekoliko izvora, važno je procijeniti ulogu analize podataka u sprječavanju širenja epidemije, kao i odrediti ključne izazove i potencijalne putove za buduće studije podataka o COVID-u (Tenali & Babu, 2023).

Jednim istraživanjem na NIH-u (National Institutes of Health) razvijen je općenamjenski, proširivi program za obradu slika i vizualizaciju kako bi se olakšalo slikovno istraživanje. Program naziva MIVAP (Medical Image Processing Analysis and Visualization) neovisan je o platformi i može obrađivati podatke koji potječu iz većine medicinskih modaliteta snimanja. Njime se korisnicima pružaju brojni osnovni i sofisticirani alati za analizu i vizualizaciju slike. Na istraživanju provedenom na starijim pacijentima s dijagnozom Alzheimerove bolesti te njihovoj rodbini prvog koljena, MIVAP se koristi za utvrđivanje suptilne promjene ukupnog volumena mozga i odabranih moždanih struktura u razdoblju od deset godina. Kako bi se zadovoljile šire potrebe zajednice korisnika, radi se na dodavanju raznih algoritama za registraciju te površinsko renderiranje u MIVAP (McAuliffe, i dr., 2002).

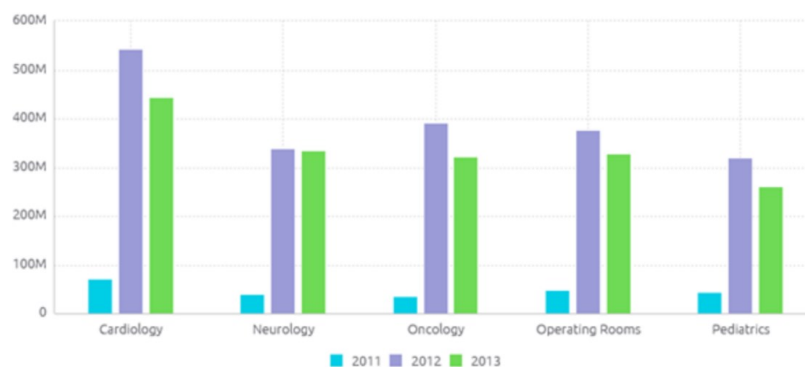
Drugo istraživanje istaknulo je važnosti tehnika vizualizacije podataka u zdravstvu sa glavnim prednostima koje su: poboljšanje pružene zdravstvene skrbi, brza dijagnoza bolesti, prepoznavanje obrazaca, pojednostavljenje prezentacije zdravstvenih podataka, ubrzanje učinkovitosti zdravstvene skrbi i poboljšanje otkrivanja grešaka. Zaključeno je da se prednosti tehnika vizualizacije jasno odražavaju u njihovoj učinkovitosti u procesu donošenja odluka, što rezultira poboljšanom sigurnošću pacijenata i kvalitetom skrbi. Također, povećava transparentnost medicinske naplate smanjenjem pogrešaka i slučajeva prijevare u zdravstvenim ustanovama (Abudiyab & Alanazi, 2022).



### 3. Metode i tehnike vizualizacije podataka u zdravstvenim organizacijama

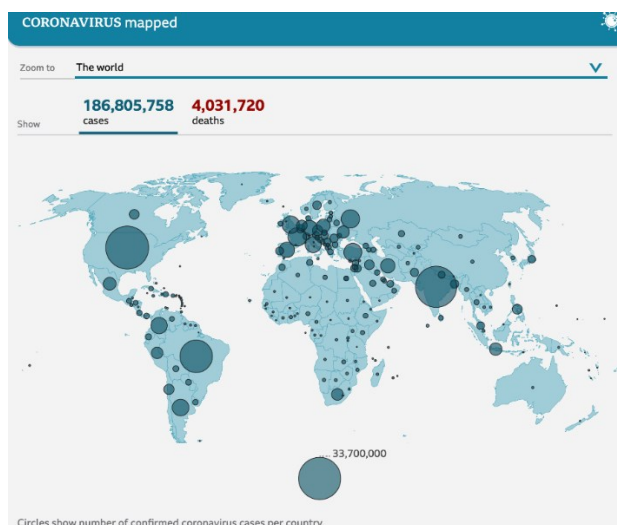
Postoji mnogo različitih metoda i tehnika za vizualizaciju podataka u zdravstvenim organizacijama. U nastavku su navedene neke od najčešćih metoda i tehnika.

1. **Grafovi:** Grafovi se uglavnom koriste za prikazivanje kvantitativnih podataka jer su jednostavni i intuitivni. Mogu se koristiti za prikaz promjena tijekom vremena, uspoređivanje skupina ili prikaz distribucije podataka. Grafovi koji se često koriste u zdravstvu su linijski grafikoni, stupčasti grafikoni, histogrami i kružni dijagrami. Na primjer, pomoću stupčastog grafikona mogu se prikazati troškovi prijema po odjelima kroz tri godine (slika 1).



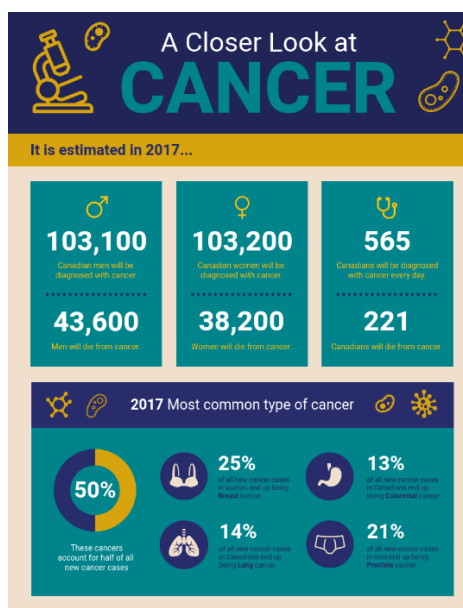
Slika 1: Troškovi prijema po odjelima (Sisense)

2. **Karte:** Karte se u zdravstvu koriste za prikazivanje prostorne distribucije podataka kao što su broj slučajeva bolesti u različitim područjima ili lokacija zdravstvenih ustanova. One mogu biti statične ili interaktivne, kako bi korisnici mogli interaktivno istraživati podatke. Primjer korištenja interaktivne karte sa prikazom broja slučajeva koronavirusa u svijetu po državi prikazuje slika 2. Korisnici mogu namjestiti određeni dio karte na koji žele zumirati te žele li prikaz ukupnog broja slučajeva ili ukupnog broja smrtnih slučajeva.



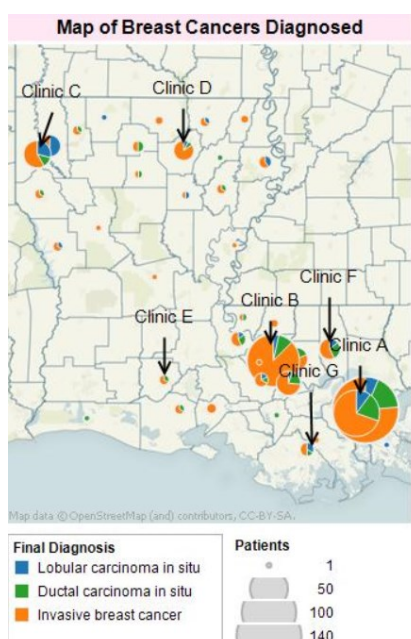
Slika 2: Prikaz broja slučajeva koronavirusa po državi (Dunskiy, 2023)

3. Infografički prikazi: Infografike su kombinacija grafova, slika i teksta. Mogu se koristiti za prikaz složenih podataka na jasan i privlačan način. Posebno su korisne za prikaz podataka široj publici koja nema stručno znanje u određenom području. Slika 3 prikazuje infogram sa brojem slučajeva raka u Kanadi u 2017. godini po kategorijama te postocima najčešćih tipova.



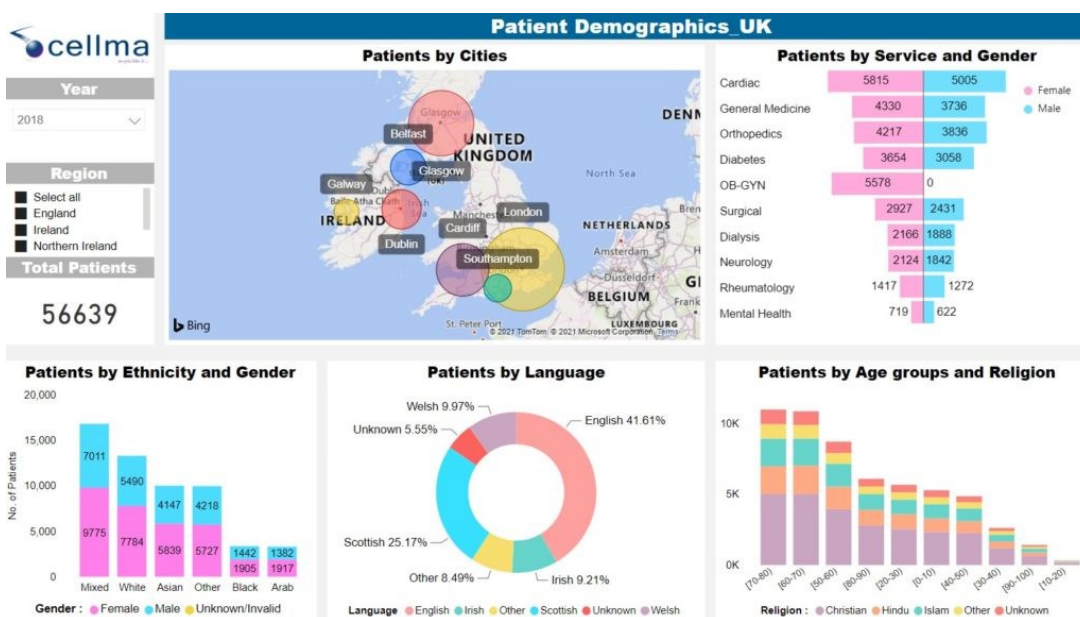
Slika 3: Prikaz podataka o raku u Kanadi 2017. godine (Abayowa, 2021)

4. Topografski prikazi: Topografski prikazi su vizualizacija u kojoj se prikazuje prostorna organizacija podataka. Na primjer, topografskim prikazom mogu se prikazati zdravstveni centri u određenom području u kojima se dijagnosticirao rak dojke (slika 4). Također, za svaki centar je pomoću kružnog grafikona prikazan udio pojedinog tipa raka dojke.



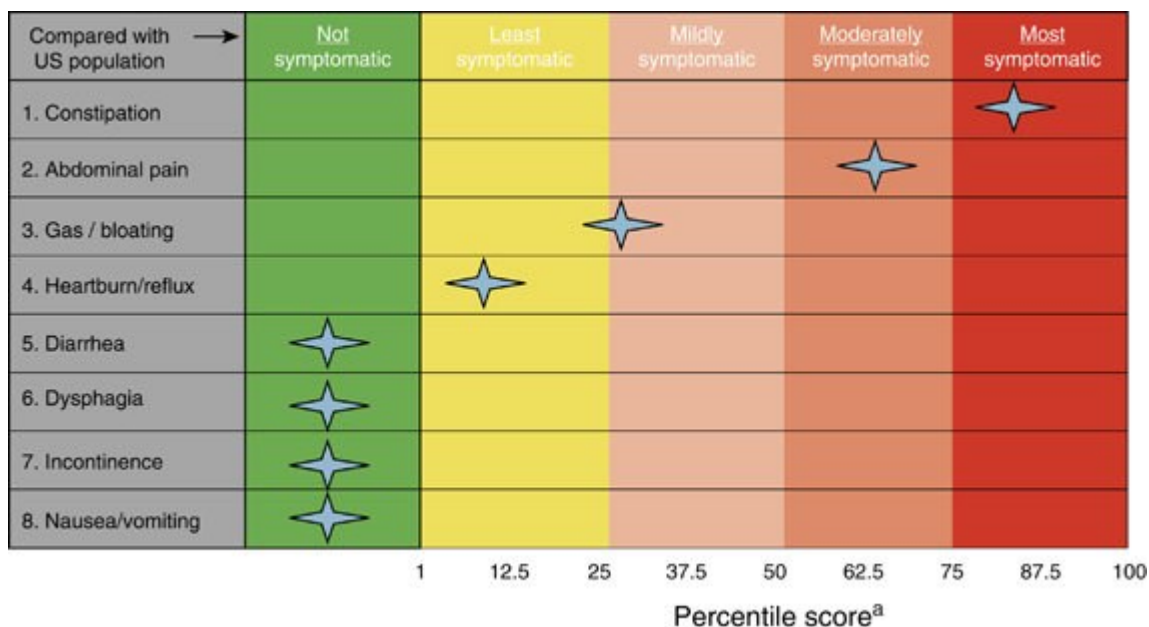
Slika 4: Prikaz klinika u kojima je dijagnosticiran rak dojke (Tableau)

5. Interaktivne vizualizacije: Interaktivne vizualizacije omogućuju korisnicima istraživanje i manipulaciju podacima na interaktivan način. Primjerice, korisnici mogu mijenjati vremenski raspon ili filtrirati podatke po određenim regijama (slika 5).



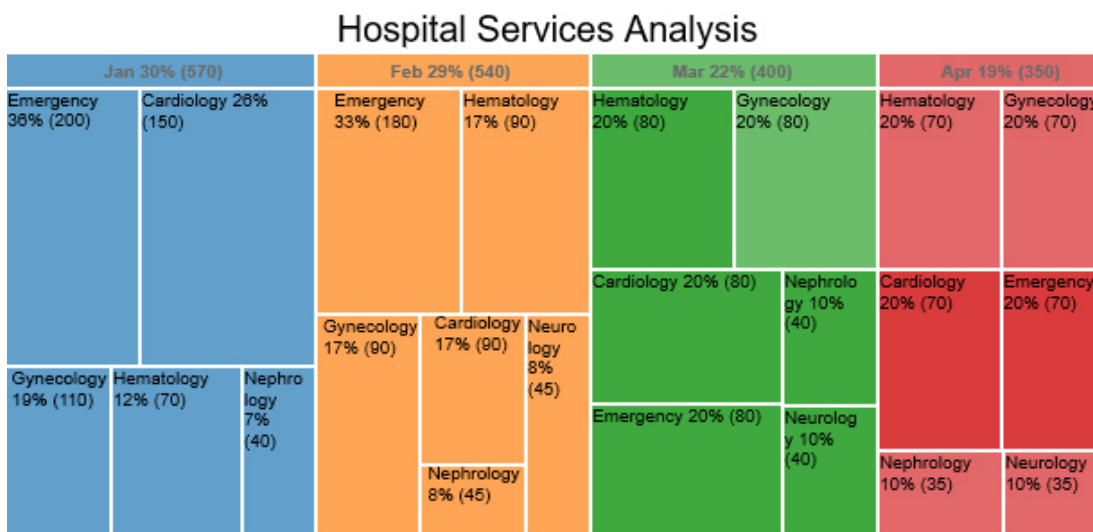
Slika 5: Prikaz podataka o pacijentima na temelju odabranih kriterija (RioMed, 2021)

6. Heat map: Heat map prikazuje razinu intenziteta nekog fenomena na karti ili u tablici koristeći boje. Na primjer, može se prikazati broj pacijenata po zdravstvenim ustanovama ili ukupna potrošnja resursa po bolnicama. Slika 6 prikazuje učestalost simptoma gastrointestinalnih bolesti na temelju prijave pacijenata.



Slika 6: Prikaz učestalosti pojedinih simptoma kod gastrointestinalnih bolesti (Chey, 2014)

7. Treemap: Treemap je prikaz hijerarhijske strukture podataka pomoću prostornih blokova. Njime se može prikazati hijerarhijska struktura zdravstvenih usluga, prikazujući usluge u blokovima različitih veličina i boja (slika 7).

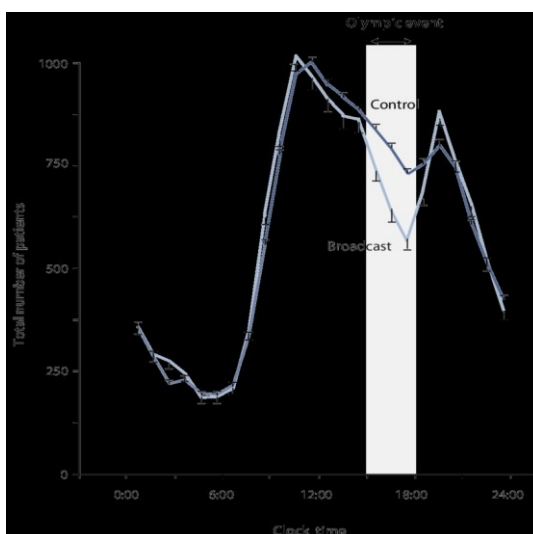


Slika 7: Prikaz hijerarhijske strukture zdravstvenih usluga (ChartExpo)

Osim prethodno navedenih metoda i tehnika, postoje i druge metode vizualizacije podataka koje se koriste u zdravstvenim organizacijama, kao što su mrežni grafikoni, *sunburst* grafikoni, *scatter* plotovi, tehnike crtanja, i druge. Odabir metode vizualizacije ovisi o ciljevima analize, vrsti podataka, publici koja će koristiti vizualizaciju i drugim čimbenicima.

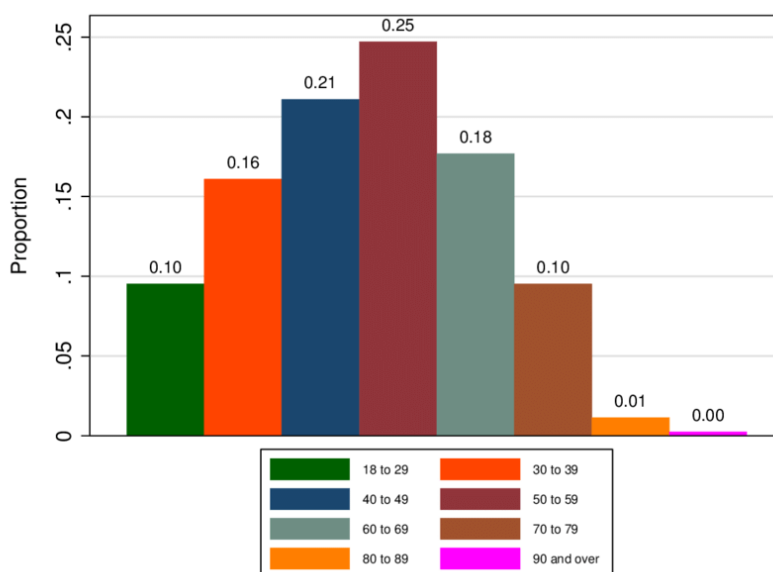
U nastavku se navodi nekoliko primjera vizualizacije podataka u analizi zdravstvenih podataka:

1. Linijski grafikon: Koristi se za vizualizaciju trendova u vremenskim serijama, poput prikaza promjene broja hospitalizacija u određenom razdoblju (slika 8).



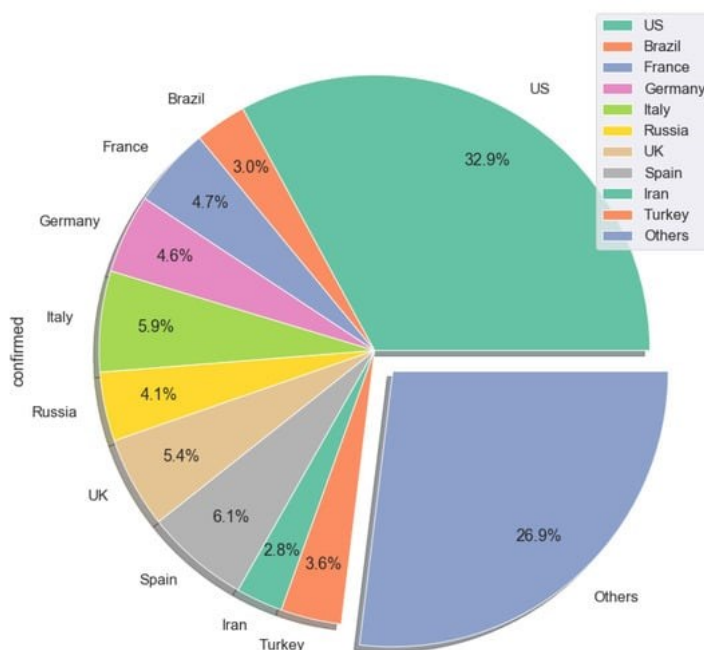
Slika 8: Prikaz promjene broja hospitalizacija tijekom 24 sata (Vermeulen, 2011)

2. Histogram: Koristi se za prikazivanje raspodjele vrijednosti u skupu podataka, poput prikaza godina pacijenata s određenom dijagnozom (slika 9).



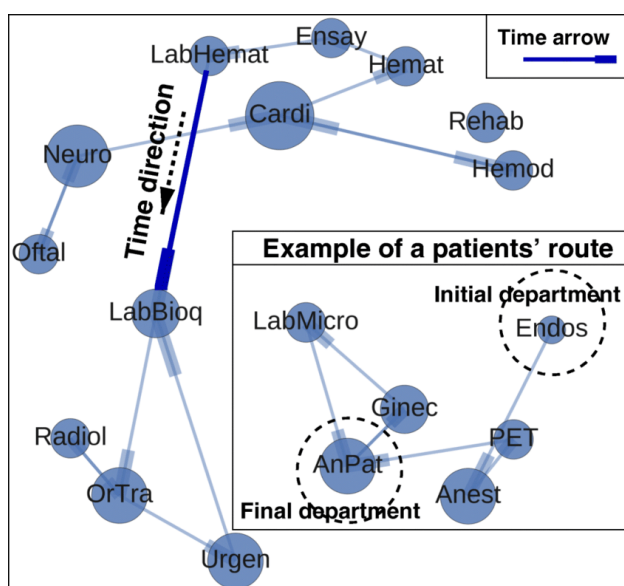
Slika 9: Prikaz godina pacijenata sa boli u donjem dijelu leđa (Froud, 2018)

3. Pie Chart: Koristi se za prikazivanje udjela različitih kategorija u skupu podataka, poput prikaza udjela broja slučajeva bolesti u određenoj državi (slika 10).



Slika 10: Prikaz distribucije ukupnog broja slučajeva koronavirusa po zemlji 2020. godine (Papastefanopoulos, Linardatos, & Kotstantis, 2020)

4. Network grafikon: Koristi se za prikazivanje veza između različitih entiteta u skupu podataka, poput prikaza ruta između različitih odjela (slika 11).



Slika 11: Prikaz rute pacijenata tokom tjedan dana (Salvatierra & Alvarez, 2020)

Ovo su samo neki od primjera vizualizacije podataka u analizi zdravstvenih podataka.

### 3.1. Uvjeti za vizualizaciju

Prije same vizualizacije, potrebno je zadovoljiti neke odredbe kako bi se ona mogla provesti. Kao prvo, važno je imati dovoljnu količinu medicinskih podataka. Vizualizacija podataka najbolje funkcionira kada se koriste velike količine dobro strukturiranih informacija. Na taj način se zdravstvenim djelatnicima omogućuje pregled različitih vrsta kombiniranih grafikona koji obuhvaćaju dulja razdoblja kako bi došli do sveobuhvatnih zaključaka. Zatim, bitno je imati dobro razrađene matematičke modele koji podržavaju vizualizaciju podataka o pacijentu. Softver na kojem se temelji vizualizacija zdravstvenih podataka mora sadržavati dobro utemeljenu matematičku logiku i algoritme. Taj mehanizam mora raditi besprijekorno kako bi medicinske podatke pretvorio u točnu, sveobuhvatnu grafiku. Važno je napomenuti da zdravstveni djelatnici ne moraju posjedovati napredne matematičke vještine te poznavati algoritme koji stoje iza same vizualizacije već trebaju donositi djelotvorne zaključke koji su primjenjivi na njihove ciljeve i aktivnosti. Posljednje, ali jednako bitno, je koristiti prikladan alatni okvir i sučelje za vizualizaciju podataka. Potrebno je da sučelje ima kontrole jednostavne za korištenje kako se zdravstveni djelatnici ne bi morali mučiti i gubiti vrijeme na učenje istih. Panel za upravljanje vizualizacijom se može implementirati u postojeća medicinska rješenja ili kreirati kao samostalna aplikacija (Khristich, 2023).

## 4. Prednosti i izazovi korištenja vizualizacije podataka u zdravstvu

### 4.1. Dobre strane vizualizacije podataka u zdravstvu

Sama vizualizacija podataka ne garantira bolje donošenje odluka u zdravstvenim organizacijama, već je potrebna kvalitetna analiza i interpretacija podataka. Međutim, dobro dizajnirana vizualizacija može pomoći u boljem razumijevanju podataka i donošenju informiranih odluka.

U nastavku se navodi 6 primjera kako vizualizacija podataka poboljšava zdravstvenu skrb (Khristich, 2023).

1. **Poboljšava njegu pacijenata:** Vizualizacija podataka omogućava lakše praćenje promjena zadovoljstva pacijenata. Mogu se jednostavno izraditi grafikoni koji prikazuju brojeve zadovoljnih i nezadovoljnih pacijenata tijekom određenog razdoblja te uspoređivati brojke iz mjesečne ili godišnje perspektive. Također se vizualizacijom može prikazati smanjenje ili povećanje zadovoljstva pacijenata po određenim skupinama. Nakon vizualizacije provodi se analiza razloga za navedene promjene.
2. **Vizualizacija za medicinske uređaje:** Medicinska oprema zahtijeva sveobuhvatno korisničko sučelje koje je lako čitljivo za zdravstvene djelatnike. Uz pomoć profesionalne vizualizacije medicinskih podataka liječnici i medicinske sestre mogu pratiti biosignale ili skenirati organsko tkivo pacijenata, tumačiti primljene signale u obliku dijagrama i animiranih vizualizacija kako bi identificirali zdravstveno stanje pacijenta, upravljati i podešavati medicinske uređaje za poboljšanje vizualnih detalja dok se ne postigne najbolja dijagnostička kvaliteta.
3. **Poboljšava pristup uvidu podacima za liječnike:** Medicinski stručnjaci mogu koristiti nadzorne ploče za vizualizaciju podataka povezane s medicinskim digitalnim zapisima. Na taj način mogu vizualizirati izvedbu i sate provedene s pojedinačnim pacijentima ili određenim medicinskim aktivnostima, vidjeti statistiku po pacijentu (stanje, planovi liječenja, rezultati), izraditi grafička izvješća o zdravstvenim podacima.
4. **Trendovi i prepoznavanje uzoraka u zdravstvenoj djelatnosti:** Mogu se otkriti padovi ili nagibi u određenim pojavama bolesti, utvrditi učestalost specifičnih simptoma u različitim skupinama pacijenata. Također, može se pronaći relevantnost za sezonske, rodne, društvene, ekonomske i druge čimbenike, definirati kako aktivnosti djelatnika, poput usvajanja novih protokola, utječu na stanje pacijenata i rezultate liječenja te crtati prognoze za buduća razdoblja na temelju trenutnih trendova i obrazaca podataka.
5. **Različite razine prikaza medicinskih podataka:** Uz pomoć prilagođenih alata za vizualizaciju kliničkih podataka, može se stvoriti širok raspon vrsta vizualnih prikaza podataka kako bi odgovarali ciljanoj publici. Mogu se izraditi grafikoni koji prikazuju metriku profesionalne medicinske usluge za investitore i dionike koji zahtijevaju izvješća o postignućima zdravstvene ustanove. Također se mogu raditi i infografike za sadašnje bolesnike i zdrave građane koje treba informirati o mogućim simptomima bolesti i metodama prevencije.
6. **Prediktivna analitika u zdravstvu koju pokreće umjetna inteligencija:** Uz pomoć umjetne inteligencije, vizualizacija također može pomoći u predviđanju budućih situacija u zdravstvu. Na primjer, može se predvidjeti nagli porast ili pad bolesti na temelju trendova, obrazaca i čimbenika, nadolazeći nedostatak radne snage u medicinskoj industriji te očekivana učinkovitost pojedinih planova i metoda liječenja u pojedinim skupinama bolesnika.

## 4.2. Izazovi vizualizacije podataka u zdravstvu

Analiza zdravstvenih podataka predstavlja proces prikupljanja, obrade i interpretacije podataka vezanih za zdravlje pojedinaca ili populacije. Mogućnosti koje pruža analiza zdravstvenih podataka su brojne, od identifikacije rizičnih čimbenika za bolesti, do praćenja trendova u zdravstvenim stanjima populacije i poboljšanja kvalitete zdravstvene skrbi.

Međutim, postoje i izazovi s kojima se susreću analitičari zdravstvenih podataka. Jedan od njih je pitanje zaštite privatnosti pacijenata i sigurnosti podataka, s obzirom na to da se radi o osjetljivim informacijama. Potrebno je osigurati adekvatnu zaštitu podataka i pridržavati se propisa i standarda vezanih za zaštitu privatnosti.

Drugi izazov je velika količina podataka koju treba obraditi i analizirati, što zahtijeva sofisticirane alate i tehnologije za prikupljanje, pohranu i obradu podataka. Potrebno je i uspostaviti učinkovit sustav praćenja i prikupljanja podataka te osigurati njihovu kvalitetu.

Treći izazov je usklađivanje različitih izvora podataka kako bi se osigurala kvalitetna analiza. Podaci o zdravstvenom stanju pacijenta mogu dolaziti iz različitih izvora, kao što su medicinske evidencije, laboratorijski nalazi, podaci o lijekovima te ankete i istraživanja. Potrebno je uskladiti i integrirati sve te izvore podataka kako bi se dobila potpuna i pouzdana slika zdravstvenog stanja pojedinca ili populacije.

Četvrti izazov je razumijevanje i interpretacija podataka. Analiza zdravstvenih podataka može biti vrlo složena i zahtjevna, te je potrebno imati stručnost za razumijevanje i interpretaciju podataka. Važno je imati i multidisciplinarni pristup analizi, uključujući stručnjake za statistiku, epidemiologiju, medicinu i druge relevantne discipline.

Uz adekvatnu pripremu i suradnju između stručnjaka različitih disciplina, analiza zdravstvenih podataka može biti vrlo korisna u poboljšanju kvalitete zdravstvene skrbi i donošenju informiranih odluka u zdravstvenoj politici (Berros, Mendili, Filaly, & Idrissi, 2023).



## 5. Analiza podataka, korištene metode i skupovi podataka

U ovom će se poglavlju prikazati analiza i vizualizacija podataka na skupu podataka o slučajevima koronavirusa u svijetu. Skup podataka 'COVID-19 Coronavirus' je preuzet s platforme Kaggle (Christy, 2022).

Kaggle<sup>1</sup> je platforma za natjecanje u znanosti o podacima i online zajednica znanstvenika koji se bave podacima i praktičara strojnog učenja pod Googleom. Kaggle omogućuje korisnicima pronalaženje i objavljivanje skupova podataka, istraživanje i izradu modela u mrežnom okruženju znanosti o podacima, rad s drugim znanstvenicima i inženjerima strojnog učenja te sudjelovanje u natjecanjima za rješavanje izazova znanosti podataka. Tvrtka je osnovana 2010. godine od strane Anthony Goldbloom i Bena Hammera s Nicholasom Gruenom kao predsjedateljem. Tvrtku je 8. ožujka 2017. preuzeo Google.

Prije same vizualizacije, izvršena je analiza podataka kako bi se znalo što se ima na raspolaganju te kako bi se pripremili podaci koji će se vizualizirati.

Odabrani skup podataka sadrži podatke o nazivu države te njezinom drugom nazivu ako ga ona ima, kratici države, populaciji države, kontinentu na kojem se država nalazi, ukupnom broju slučajeva korone, ukupnom broju smrtnih slučajeva zbog korone, ukupnom broju slučajeva korone na milijun stanovnika, ukupnom broju smrtnih slučajeva zbog korone na milijun stanovnika te postotku smrtnosti obzirom na broj ukupno prijavljenih slučajeva korone. Podaci se mogu svrstati u 3 kategorije: podaci o državi (naziv države, drugi naziv države, kratica države, populacija države, kontinent na kojem je država), podaci o broju slučajeva (ukupni broj slučajeva, ukupni broj slučajeva na milijun stanovnika) te podaci o broju smrtnih slučajeva (ukupni broj smrtnih slučajeva, ukupni broj smrtnih slučajeva na milijun stanovnika, postotak smrtnosti). Što se tiče tipa podataka, stupci naziv države, drugi naziv, kratica i kontinent su tipa object, populacija, ukupni broj slučajeva, ukupni broj slučajeva na milijun stanovnika, ukupni broj smrtnih slučajeva i ukupni broj smrtnih slučajeva na milijun stanovnika su tipa integer, a postotak smrtnosti je tipa float. Vremensko razdoblje prikupljanja podataka je od početka koronavirusa (studeni 2019.) do objavljivanja skupa podataka na Kaggle platformu (travanj 2022.).

Za analizu i vizualizaciju podataka korišten je alat Google Colab<sup>2</sup>.

Google Colab je proizvod Google Researcha koji omogućuje pisanje i izvršavanje proizvoljnog Python koda putem preglednika. Posebno je pogodan za strojno učenje, analizu podataka i obrazovanje.

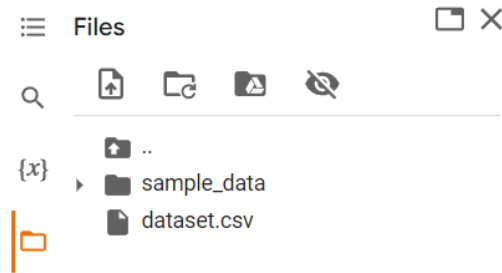
Unutar Google Colab-a napravljena je nova bilježnica u kojoj je pisan i izvršavan kod.

Prije samih naredbi, potrebno je učitati skup podataka u datoteke. (slika 12).

---

<sup>1</sup> Dostupna na web adresi: <https://www.kaggle.com/>

<sup>2</sup> Dostupan na web adresi <https://colab.google/>



Slika 12: Učitavanje skupa podataka u Google Colab

Pošto su podaci koji se koriste u CSV datoteci, za njihovo učitavanje koristi se funkcija `read_csv` iz `pandas` biblioteke koja je prethodno uvedena te omogućava da se koriste sve funkcionalnosti iz nje.

`Pandas` je softverska biblioteka napisana za programski jezik Python koja služi za manipulaciju i analizu podataka. Nudi korištenje struktura podataka i operacije za manipuliranje numeričkim tablicama i vremenskim serijama. Naziv je izveden iz pojma 'panel data', izraz za skupove podataka koji uključuju opažanja u više vremenskih razdoblja.

Funkcija `read_csv` učitava podatke i sprema ih u `pandas dataframe` objekt koji omogućava manipulaciju i analizu podataka.

Za analizu podataka koristile su se funkcije `head`, `dtypes` te `shape`.

Funkcija `head` služi za prikaz prvih nekoliko redaka učitanih podataka što je korisno za brzi uvid u strukturu i sadržaj podataka (slika 13).

```
import pandas as pd
data = pd.read_csv('dataset.csv')
data.head()
```

	Country	Other names	ISO 3166-1 alpha-3 CODE	Population	Continent	Total Cases	Total Deaths	Tot Cases//1M pop	Tot Deaths/1M pop	Death percentage
0	Afghanistan	Afghanistan	AFG	40462186	Asia	177827	7671	4395	190	4.313743
1	Albania	Albania	ALB	2872296	Europe	273870	3492	95349	1216	1.275058
2	Algeria	Algeria	DZA	45236699	Africa	265691	6874	5873	152	2.587216
3	Andorra	Andorra	AND	77481	Europe	40024	153	516565	1975	0.382271
4	Angola	Angola	AGO	34654212	Africa	99194	1900	2862	55	1.915438

Slika 13: Ispis prvih nekoliko redaka podataka

Za provjeru tipa podataka svakog stupca koristi se funkcija `dtypes` (slika 14).

```
[2] data.dtypes
```

```
Country                object
Other names            object
ISO 3166-1 alpha-3 CODE object
Population             int64
Continent              object
Total Cases            int64
Total Deaths           int64
Tot Cases//1M pop     int64
Tot Deaths/1M pop    int64
Death percentage      float64
dtype: object
```

Slika 14: Ispis tipova podataka za svaki stupac

Funkcija *shape* služi za prikaz dimenzije dataframea, odnosno broja redaka i stupaca. Po izlazu koda vidi se da skup podataka ima 225 redaka i 10 stupaca (slika 15).

```
▶ data.shape
(225, 10)
```

Slika 15: Ispis dimenzije dataframe-a

Za uklanjanje mogućih nedostajućih vrijednosti korištena je funkcija *dropna* koja uklanja redove koji sadrže barem jednu nedostajuću vrijednost (slika 16).

```
✓ [5] data = data.dropna()
```

Slika 16: Uklanjanje nedostajućih vrijednosti

Usljedili su daljnji koraci za implementaciju vizualizacija podataka.

Za početak su se uvele dvije biblioteke koje se koriste, a to su *matplotlib.pyplot* i *seaborn*.

*Matplotlib* je osnovna biblioteka za vizualizaciju podataka u programskom jeziku Python. Ona omogućava stvaranje različitih tipova grafova i dijagrama kako bi se bolje razumjeli i predstavili podaci. Temelji se na konceptu objekata koji predstavljaju različite elemente grafa, kao što su osi, naslovi, legende i sami podaci (Matplotlib).

*Seaborn* je nadogradnja *Matplotlib-a* koja olakšava stvaranje estetski privlačnih i informacijski bogatih statističkih grafova. Integrira se s *Pandasom* i pruža jednostavan način za vizualizaciju statističkih analiza podataka (Waskom, 2022).

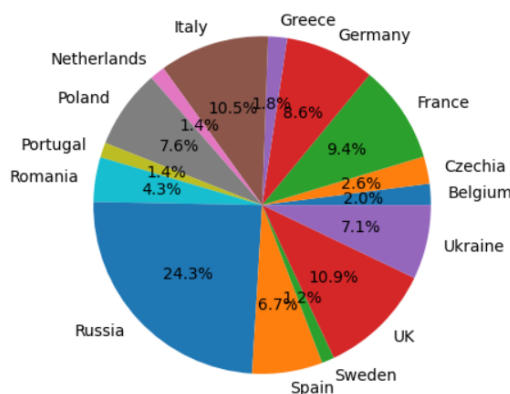
## 5.1. Kružni dijagram

Prva izrađena vizualizacija je prikaz broja smrtnih slučajeva u europskim zemljama koje imaju više od 10 milijuna stanovnika do travnja 2022. godine.

Za početak izvršila se filtracija podataka u dataframe-u kako bi se izdvojili samo oni retci koji zadovoljavaju uvjete da pripadaju kontinentu 'Europa' i da imaju populaciju veću od 10 milijuna stanovnika. Rezultat filtriranja sprema se u novi dataframe 'europske\_zemlje'. Zatim se stvara kružni grafikon koristeći podatke iz novog dataframe-a. Funkcija *plt.pie()* prima listu vrijednosti koje će biti prikazane u dijagramu (broj smrtnih slučajeva), listu oznaka za svaku vrijednost (naziv zemlje) te *autopct* koji dodaje postotak na svaki udio grafa. Pomoću *plt.title()* postavlja se naslov dijagrama, a *plt.show()* prikazuje stvoreni kružni dijagram (slika 17).

```
import matplotlib.pyplot as plt
import seaborn as sns
europske_zemlje=data[(data['Continent']=='Europe')& (data['Population']>10000000)]
plt.pie(europske_zemlje['Total Deaths'], labels=europske_zemlje['Country'],autopct='%1.1f%%')
plt.title('Broj smrtnih slučajeva u Europskim zemljama sa više od 10 milijuna stanovnika')
plt.show()
```

Broj smrtnih slučajeva u Europskim zemljama sa više od 10 milijuna stanovnika

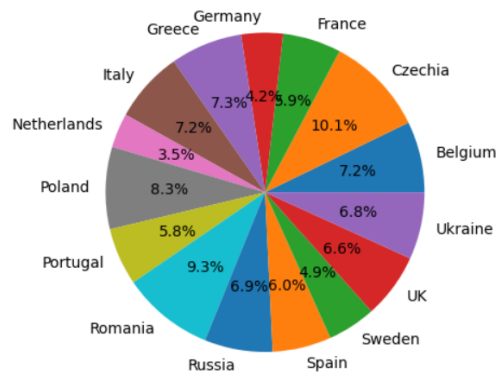


Slika 17: Stvaranje i prikaz kružnog dijagrama

Svaka zemlja ima prikazan udio drukčije boje na dijagramu kako bi se lakše uočavale razlike između njih, ali također za svaki udio prikazan je i postotak. Na temelju dobivenog dijagrama može se vidjeti da je zemlja s najvećim brojem smrtnih slučajeva Rusija, dok je Švedska s najmanje. Očekivano je da će Rusija imati najveći broj smrtnih slučajeva pošto ima i najveći broj stanovnika. Kako bi to izbjegli, potrebno je normirati ukupan broj smrtnih slučajeva. Korištenjem stupca ukupnog broja slučajeva na milijun stanovnika dobivamo precizniju i realniju sliku stanja u svijetu. Na temelju novostvorenog dijagrama može se vidjeti da je zemlja sa najvećom smrtnošću Češka, dok je Nizozemska s najmanjom (slika 18). Na ovom konkretnom primjeru može se zaključiti kolika je važnost normalizacije podataka te koliko ona utječe na krajnji rezultat.

```
europske_zemlje=data[(data['Continent']=='Europe')& (data['Population']>10000000)]
plt.pie(europske_zemlje['Tot\xa0Deaths/1M pop'], labels=europske_zemlje['Country'], autopct='%1.1f%%')
plt.title('Broj smrtnih slučajeva na milijun stanovnika u Europskim zemljama sa više od 10 milijuna stanovnika')
plt.show()
```

Broj smrtnih slučajeva na milijun stanovnika u Europskim zemljama sa više od 10 milijuna stanovnika



Slika 18: Stvaranje i prikaz poboljšanog kružnog dijagrama

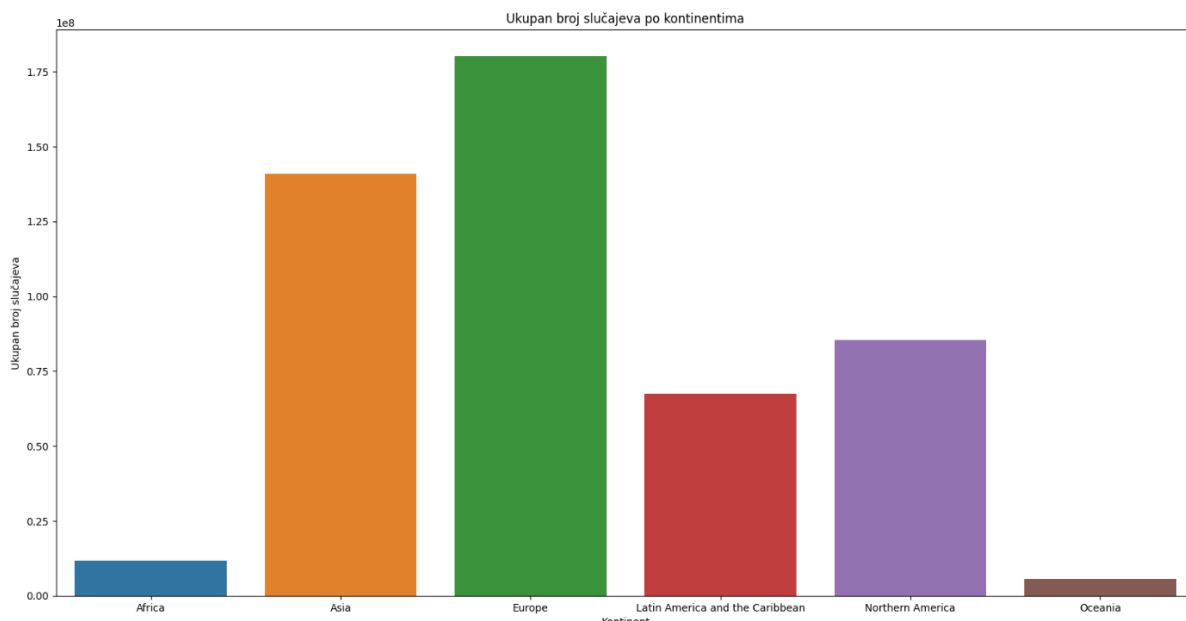
## 5.2. Stupčasti dijagram

Drugi primjer vizualizacije je prikaz ukupnog broja slučajeva korone u razdoblju od dvije godine po kontinentima koristeći stupčasti dijagram (slika 20). Za početak su se grupirali podaci po kontinentu funkcijom `data.groupby('Continent')` te se za svaki kontinent izračunao ukupan broj slučajeva zbrajanjem vrijednosti iz stupca 'Total Cases' funkcijom `['Total Cases'].sum()`. Pomoću `figsize` se namjestila veličina slike za grafički prikaz kako bi sav tekst stao u ispisu. Funkcija `barplot` iz `Seaborn` biblioteke služi za izradu stupčastog dijagrama. Preko parametara `x` i `y` se definiraju osi dijagrama, a preko `data` se definira dataframe iz kojeg se uzimaju podaci za dijagram. Naslov dijagrama se postavlja pomoću `plt.title`, a osi dijagrama imenujemo pomoću `plt.xlabel` i `plt.ylabel`. Slika 19 prikazuje kod za stvaranje stupčastog dijagrama.

```
slucajevi_po_kontinentu=data.groupby('Continent')['Total Cases'].sum().reset_index()
plt.figure(figsize=(20, 10))
sns.barplot(x='Continent', y='Total Cases', data=slucajevi_po_kontinentu)
plt.title('Ukupan broj slučajeva po kontinentima')
plt.xlabel('Kontinent')
plt.ylabel('Ukupan broj slučajeva')

plt.show()
```

Slika 19: Stvaranje stupčastog dijagrama



Slika 20: Prikaz ukupnog broja slučajeva po kontinentima stupčastim dijagramom

Iščitavanjem dobivenih rezultata iz stupčastog grafikona može se zaključiti da je najviše slučajeva koronavirusa bilo na kontinentu Europe, a najmanje na Oceaniji (Australija, Polinezija, Melanezija, Mikronezija).

### 5.3. Geografska mapa

Za treći grafikon korištena je biblioteka *GeoPandas* kako bi se izradila karta koja prikazuje broj slučajeva koronavirusa po državama (slika 22).

*GeoPandas* je projekt otvorenog koda koji olakšava rad s geoprostornim podacima u Pythonu. Proširuje tipove podataka koji koriste *Pandas* kako bi omogućio prostorne operacije na geometrijskim tipovima (*GeoPandas*).

Pomoću *gpd.read\_file* učitala se svjetska mapa iz *GeoPandas*-ovog skupa podataka *naturalearth\_lowres* koja sadrži osnovne informacije o svjetskim državama i geometriji njihovih granica. Zatim su se grupirali podaci po zemljama i sumirao broj slučajeva koronavirusa na milijun stanovnika za svaku zemlju. Nakon toga, spojila se svjetska mapa s prethodno stvorenim dataframeom 'slucajevi\_po\_drzavama' preko *world.merge* funkcije. Spaja se lijevo (*how='left'*) kako bi se zadržale sve zemlje iz svjetske mape, čak i ako nema podataka o slučajevima za neke od njih. Dakle, spajanjem lijevo uzimaju se ključevi dataframe-a *world* pa na dobivenom grafu vidimo kartu cijelog svijeta, a ne samo države za koje postoje podaci o broju slučajeva. Zatim se crta mapa tako što se prvo stvara nova figura, tj. graf i os za crtanje mape preko *plt.subplots*, a zatim se pomoću funkcije *plot* crta karta. Parametar *column* označava stupac po kojemu će se prikazivati boje na mapi, u ovom slučaju broj slučajeva na milijun stanovnika. Parametar *cmap* postavlja paletu boja, u ovom slučaju 'BuPu' je kombinacija plave i ljubičaste boje gdje boje prelaze od svijetloplave prema tamnoljubičastoj. Parametar *linewidth* postavlja debljinu granica država, a *edgecolor* postavlja boju granica. Kako bi se prikazala samo karta bez osi, koristi se funkcija *set\_axis\_off*. Na slici 21 prikazane su prethodno spomenute funkcije.

```

import geopandas as gpd
world = gpd.read_file(gpd.datasets.get_path('naturalearth_lowres'))

slucajevi_po_drzavama = data.groupby('Country')['Tot\xa0Cases//1M pop'].sum().reset_index()

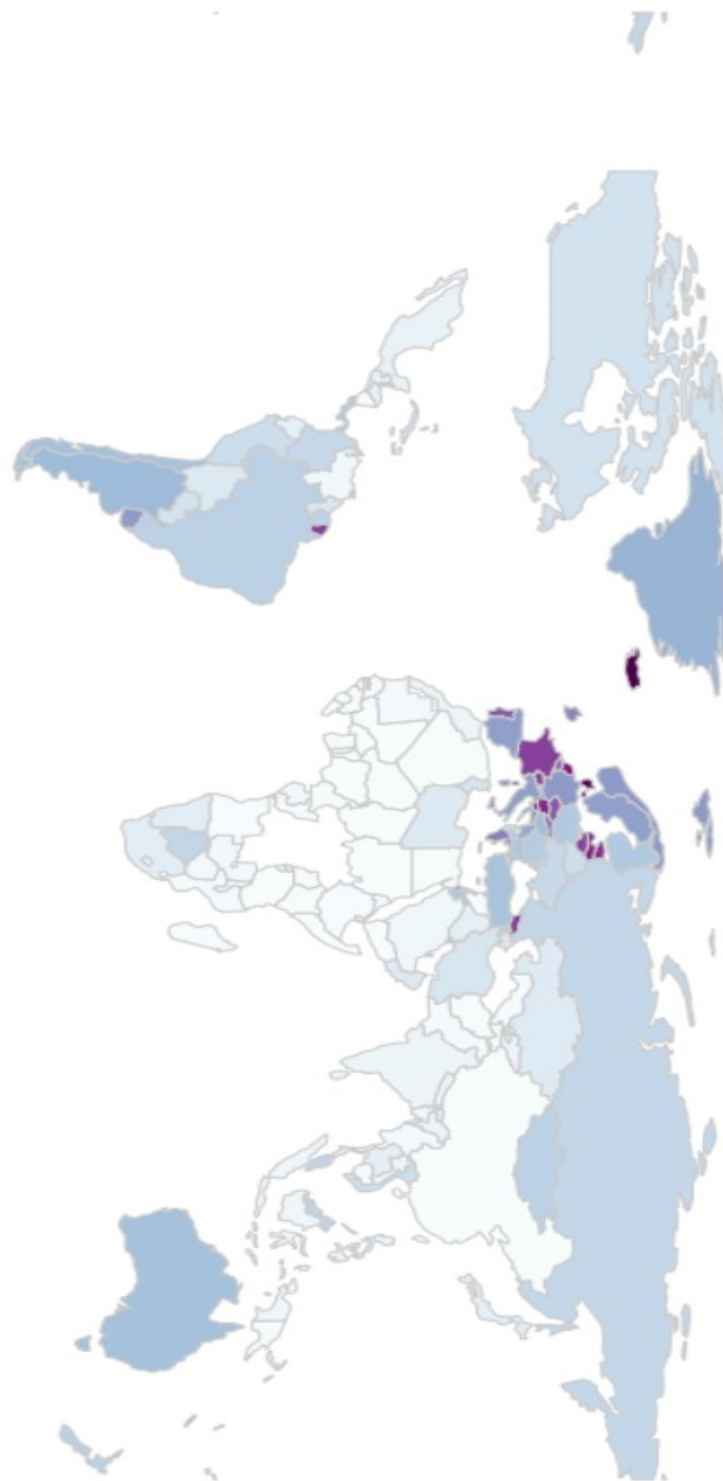
merged = world.merge(slucajevi_po_drzavama, how='left', left_on='name', right_on='Country')

fig, ax = plt.subplots(1, 1, figsize=(15, 10))
merged.plot(column='Tot\xa0Cases//1M pop', cmap='BuPu', linewidth=0.8, ax=ax, edgecolor='0.8', legend=True)
ax.set_title('Broj slu\u010dajeva COVID-19 po dr\u017eavama')
ax.set_axis_off()

plt.show()

```

*Slika 21: Stvaranje geografske mape*



Broj slučajeva COVID-19 po državama

Slika 22: Prikaz broja slučajeva koronavirusa po državama geografskom mapom



Ispisani dijagram prikazuje kartu svijeta te odvojene države. Na temelju boje može se vidjeti da su tamnije obojane države one koje imaju veći ukupni broj slučajeva koronavirusa na milijun stanovnika, dok su svjetlije obojane one sa manjim brojem slučajeva. Na prikazu je vidljivo da se zemlje sa većim brojem slučajeva koronavirusa uglavnom nalaze u Europi pošto na tom predjelu karte prevladava tamnija boja, dok je Afrika svijetlo obojana što znači da ima i manji broj slučajeva.

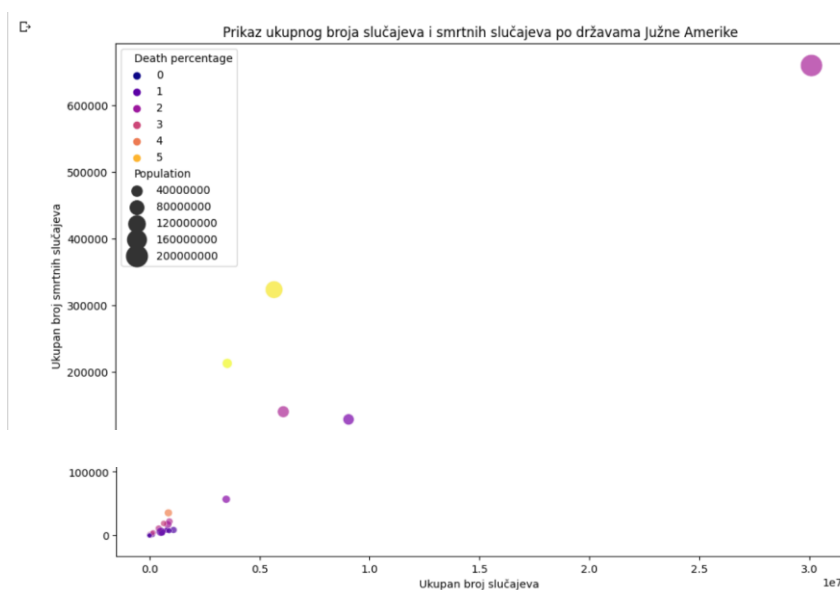
#### 5.4. Dijagram raspršenosti

Iduća vizualizacija bio je prikaz usporedbe ukupnog broja slučajeva i ukupnog broja smrtnih slučajeva zbog koronavirusa po državama Južne Amerike putem dijagrama raspršenosti (slika 24).

Za početak se izvršila filtracija podataka tako da se dobije podskup podataka samo za države Južne Amerike. Preko *figsize* se odredila dimenzija slike za crtanje grafa kako bi prikaz podataka bio čitljiviji. Za crtanje grafa koristi se *Seaborn*. Parametri *x* i *y* označavaju stupce koji će biti prikazani na x i y osi grafa, parametar *hue* označava stupac za bojenje točaka, u ovom slučaju postotak smrtnosti, *size* označava stupac koji određuje veličinu točaka, u ovom slučaju populacija. Parametar *data* specificira skup podataka koji će se koristiti za crtanje grafa, u ovom slučaju onaj koji se dobio filtracijom. Parametar *sizes* određuje raspon veličina točaka, *alpha* određuje transparentnost točaka, a *palette* paletu boja za bojenje točaka. Prethodno spomenuti postupak za izradu dijagrama prikazan je na slici 23.

```
[50] jamer=data[(data['Continent']=='Latin America and the Caribbean')]
plt.figure(figsize=(12, 8))
sns.scatterplot(x='Total Cases', y='Total Deaths', hue='Death percentage', size='Population', data=jamer, sizes=(20, 400), alpha=0.7,
palette='plasma')
plt.title('Prikaz ukupnog broja slučajeva i smrtnih slučajeva po državama Južne Amerike')
plt.xlabel('Ukupan broj slučajeva')
plt.ylabel('Ukupan broj smrtnih slučajeva')
plt.show()
```

Slika 23: Stvaranje dijagrama raspršenosti



Slika 24: Prikaz ukupnog broja slučajeva i smrtnih slučajeva u državama Južne Amerike dijagramom raspršenosti

Preko dobivenog grafa mogu se vizualno analizirati veze između ukupnog broja slučajeva, ukupnog broja smrtnih slučajeva, postotka smrtnosti i populacije za države Južne Amerike. Boja točaka označava postotak smrtnosti, veličina točaka označava veličinu populacije, a njihova pozicija na grafu odražava odnos između ukupnog broja slučajeva i smrtnih slučajeva. Može se primijetiti da uglavnom što je veći broj ukupnih slučajeva to je veći i ukupni broj smrtnih slučajeva u nekoj državi. Također, što je manja država, to ima i manji ukupni broj slučajeva što je očekivano pošto se ne gleda ukupan broj slučajeva na milijun stanovnika već cjelokupni.

## 6. Diskusija

Na temelju dosadašnjih istraživanja, analiza i vizualizacija podataka može imati glavnu ulogu u identificiranju novih načina liječenja neke bolesti, razvoju boljih dijagnostičkih testova i predviđanju budućih izbijanja bolesti. Također, može pomoći zdravstvenim organizacijama da učinkovitije upravljaju svojim resursima pružajući podatke o zdravstvenom stanju pacijenata u stvarnom vremenu i omogućujući im da brzo donose odluke na temelju najnovijih dostupnih informacija. Analiza i vizualizacija podataka u zdravstvu ne samo da pomaže zdravstvenim radnicima brže dijagnosticirati bolesti, nego također pomaže u točnom predviđanju budućih izbijanja kako bi se protiv njih moglo boriti učinkovitije nego prije.

### 6.1. Glavni zaključci izrađenih vlastitih vizualizacija

Na temelju vlastitih vizualizacija provedenih nad odabranim skupom podataka iz prethodnog poglavlja, može se izvesti nekoliko zaključaka i pretpostavki.

Rezultati su pokazali da je najmanja smrtnost od koronavirusa u europskim zemljama koje imaju više od 10 milijuna stanovnika bila u Nizozemskoj, Njemačkoj i Švedskoj, a najveća u Češkoj i Rumunjskoj. Mogući razlog tome je visoko razvijeno zdravstvo koje imaju Nizozemska, skandinavske zemlje i Njemačka te stoga imaju i manji broj smrtnih slučajeva. U odnosu na zemlje koje su obuhvaćene u skupu podataka, Češka i Rumunjska su među slabije razvijenima pa je za pretpostaviti da imaju i slabije razvijeni zdravstveni sustav i da je to razlog veće smrtnosti.

Što se tiče rezultata za ukupan broj slučajeva po kontinentima, pokazalo se da ih je najviše bilo u Europi, a najmanje u Australiji. Očienkavno je da će najmanje slučajeva biti u Australiji pošto je to kontinent s najmanje stanovnika pa time ima i najmanji broj slučajeva. Međutim, Afrika kao drugi najveći kontinent po broju stanovnika nalazi se odmah iza Australije. Može se pretpostaviti da je razlog taj što je Afrika slabo razvijena te se nisu provodila sustavna testiranja pa se ne može dobiti realna slika. Iako je Azija veća od Europe, ima manje slučajeva od nje. Mogući razlog je taj da ljudi u Europi više putuju iz jedne zemlje u drugu te se time brže prenosi virus i raste broj zaraženih.

Vizualizacijom veza između ukupnog broja slučajeva i smrtnih slučajeva u državama Južne Amerike dalo se zaključiti da što je veći ukupni broj slučajeva u državi, to je veći i smrtni broj slučajeva, što ima smisla jer jedno povlači drugo. Također, veći broj slučajeva imaju države s većom populacijom, a manji one koje imaju manju populaciju.

### 6.2. Prijedlozi za buduća istraživanja

Na temelju dosadašnjih istraživanja, uvođenje analitike i vizualizacije podataka u zdravstvo može imati ključnu ulogu za poboljšanje ukupne kvalitete života. Zdravstveni djelatnici te svi oni koji pružaju zdravstvene usluge mogu steći sveobuhvatnije razumijevanje zdravstvenih potreba zajednice iskorištavanjem velikih količina podataka koje generiraju različiti izvori poput prijenosnih uređaja, elektroničkih zdravstvenih kartona i platforma društvenih medija. Zahvaljujući tome, moglo bi doći do učinkovitijih i ciljanijih intervencija u rješavanju zdravstvenih problema, a ujedno i razvoja proaktivnih strategija zdravstvene zaštite kako bi se uopće izbjegle bolesti. Korištenje analitike i

vizualizacije podataka također može pomoći u optimizaciji raspodjele zdravstvenih resursa, smanjenju troškova i povećanju učinkovitosti. Na temelju najboljih praksi u području analitike i vizualizacije podataka u zdravstvu, pruža se opći okvir koji zdravstvene organizacije trebaju slijediti. Primjenom te jednostavne strategije, zdravstveni djelatnici mogu učinkovito iskoristiti potencijal analitike velikih podataka za poboljšanje ishoda pacijenata u medicinskim ustanovama (Berros, Mendili, Filaly, & Idrissi, 2023).

Smjernice za najbolji učinak korištenja analize i vizualizacije u zdravstvu (Berros, Mendili, Filaly, & Idrissi, 2023):

1. Definicija ciljeva i zadataka: Trebaju se jasno definirati ciljevi i ciljevi inicijative za vizualizaciju, kao što su poboljšanje ishoda pacijenata, smanjenje troškova zdravstvene skrbi ili poboljšanje kvalitete skrbi.
2. Razvoj sveobuhvatne podatkovne strategije: Potrebno je razviti sveobuhvatnu podatkovnu strategiju koja opisuje kako će se podaci za vizualizaciju prikupljati, pohranjivati, obrađivati i analizirati.
3. Identifikacija alata i aplikacija: Važno je uložiti u pravu tehnologiju i infrastrukturu koja će se koristiti.
4. Maksimiziranje korištenja postojećeg znanja: Prioritet je usvojiti perspektivu koja integrira i koristi postojeće znanje. Taj pristup će poboljšati razumijevanje podataka, olakšati sustavno stvaranje novih uvida i potaknuti kulturu vođenu podacima unutar medicinske ustanove.
5. Stvaranje medicinske mreže: Suradnja s pacijentima, pružateljima zdravstvenih usluga i istraživačima osigurati će usklađenost projekta s njihovim potrebama i ciljevima.
6. Uspostavljanje snažnog pravnog okvira za zaštitu osobnih podataka: Zaštita podataka ima ključnu ulogu u uspješnoj implementaciji. Posebnu pozornost treba posvetiti obradi osobnih podataka, pri čemu je važno voditi računa o pravnim okvirnim uvjetima i tehnološkim mogućnostima za njezinu provedbu.
7. Progresivni razvoj i kontinuirano praćenje: Progresivna integracija može pomoći u boljem praćenju i kontinuiranoj evaluaciji inicijative za analizu i vizualizaciju podataka kako bi se osiguralo da daje vrijednost i pozitivno utječe na skrb o pacijentima.

## 7. Zaključak

Uporabom analize i vizualizacije podataka, ne samo da se rješava mnoštvo problema, već se također otvaraju vrata za buduća ulaganja u poboljšanje zdravstvene skrbi. Njihovo korištenje omogućuje bolju organizaciju procesa i donošenje informiranih odluka. Gledajući u budućnost, jasno je da će se još više ulagati u razvoj ovih tehnologija kako bi se podigla kvaliteta i učinkovitost zdravstvene skrbi. Napredak tehnologije i sve veća upotreba digitalnih alata mijenjaju način na koji funkcionira zdravstveni sustav. Zdravstvene ustanove će se morati prilagoditi uporabi tehnologije kako bi pružile napredne usluge pacijentima. Time će se omogućiti personalizirana njega, šira dostupnost i bolja kvaliteta zdravstvenih usluga. Skup podataka koji sadrži informacije o pacijentima pomaže bolje razumjeti kako bolesti napreduju, kako djeluju terapije te potrebe pojedinaca. To znači da se može ciljano djelovati, spriječiti probleme i brže otkriti bolesti.

Unatoč velikim prednostima, zdravstveni sektor nosi i specifične izazove u primjeni analize i vizualizacije podataka. Njima je potrebno posvetiti posebnu pažnju kako bi se pravilno usmjerila istraživanja i razvoj. Sigurnost podataka, etička pitanja i integracija s postojećim zdravstvenim sustavima su samo od nekih izazova koji zahtijevaju multidisciplinarni pristup i kontinuirano unapređenje.

Ukupno gledajući, sve je više očigledno da je analiza i vizualizacija podataka kritična komponenta budućnosti zdravstvene skrbi. Kombinacija tehnologije, podataka i stručnog znanja omogućuje stvaranje naprednih zdravstvenih sustava koji koriste i pacijentima i zdravstvenim stručnjacima. Zbog ovog obećavajućeg potencijala, ulaganje u istraživanje i razvoj u ovom području postaje ključno kako bi se iskoristile sve mogućnosti za napredak i unapređenje zdravstvene skrbi.

## 8. Popis literature

- Abayowa, J. (6. April 2021). *Visualizing Healthcare Data With Infographics to Save Lives*. Preuzeto 9. September 2023 iz Venngage: <https://venngage.com/blog/healthcare-data-visualization/>
- Abudiyab, N. A., & Alanazi, A. T. (11. November 2022). *Visualization Techniques in Healthcare Applications: A Narrative Review*. Preuzeto 12. September 2023 iz National Library of Medicine: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9741729/>
- Berros, N., Mendili, F. E., Filaly, Y., & Idrissi, Y. E. (30. March 2023). *Enhancing Digital Health Services with Big Data Analytics*. Preuzeto 20. August 2023 iz MDPI Open Access Journals: <https://www.mdpi.com/2504-2289/7/2/64>
- ChartExpo. (n.d.). *How to Get Started with Healthcare Data Visualization?* Preuzeto 10. September 2023 iz ChartExpo: <https://chartexpo.com/blog/healthcare-data-visualization#>
- Chey, W. (December 2014). *Computer-Generated Vs. Physician-Documented History of Present Illness (HPI): Results of a Blinded Comparison*. Preuzeto 10. September 2023 iz ResearchGate: [https://www.researchgate.net/figure/Sample-heat-map-patient-report-of-GI-PROMIS-scores-Patients-complete-PROMIS-items-on\\_fig1\\_269169177](https://www.researchgate.net/figure/Sample-heat-map-patient-report-of-GI-PROMIS-scores-Patients-complete-PROMIS-items-on_fig1_269169177)
- Christy, R. (5. April 2022). *COVID-19 Coronavirus Pandemic*. Preuzeto 9. August 2023 iz Kaggle: <https://www.kaggle.com/datasets/rinichristy/covid19-coronavirus-pandemic>
- Dunskiy, I. (18. July 2023). *HEALTHCARE DATA VISUALIZATION: EXAMPLES & KEY BENEFITS*. Preuzeto 9. September 2023 iz Demigos: <https://demigos.com/blog-post/healthcare-data-visualization/>
- Froud, R. (January 2018). *Responsiveness, reliability, and minimally important and minimal detectable change of three electronic patient reported outcome measures for low back pain: a validation study (Preprint)*. Preuzeto 10. September 2023 iz ResearchGate: [https://www.researchgate.net/figure/Histogram-of-patient-age-The-figure-shows-a-histogram-of-patient-age-at-baseline\\_fig3\\_325820363](https://www.researchgate.net/figure/Histogram-of-patient-age-The-figure-shows-a-histogram-of-patient-age-at-baseline_fig3_325820363)
- Gattu, S. (16. September 2021). *How to Visualise data in Maps Using GeoPandas*. Preuzeto 18. August 2023 iz Analytics Vidhya: <https://www.analyticsvidhya.com/blog/2021/09/how-to-visualise-data-in-maps-using-geopandas/>
- GeoPandas. (n.d.). *GeoPandas 0.13.2*. Preuzeto 12. August 2023 iz GeoPandas: <https://geopandas.org/en/stable/>
- Google. (n.d.). *Colaboratory*. Preuzeto 16. August 2023 iz Google: <https://research.google.com/colaboratory/faq.html>
- Khristich, S. (12. January 2023). *Healthcare Data Visualization Software Development: Examples, Benefits & Challenges*. Preuzeto 15. June 2023 iz Tateeda:

<https://tateeda.com/blog/healthcare-data-visualization-examples-benefits-challenges#:~:text=Data%20visualization%20in%20the%20healthcare,even%20witho ut%20developed%20math%20skills>

Matplotlib. (n.d.). *Matplotlib: Visualization with Python*. Preuzeto 10. August 2023 iz Matplotlib: <https://matplotlib.org/>

McAuliffe, M. J., Lalonde, F. M., McGarry, D., Gandler, W., Csaky, K., & Trus, B. L. (7. August 2002). *Medical Image Processing, Analysis & Visualization*. Preuzeto 12. September 2023 iz IEEE: [https://www.researchgate.net/profile/Benes-Trus/publication/3908878\\_Medical\\_Image\\_Processing\\_Analysis\\_Visualization\\_in\\_Clinical\\_Research/links/56f2b65408aee25a6371bb78/Medical-Image-Processing-Analysis-Visualization-in-Clinical-Research.pdf](https://www.researchgate.net/profile/Benes-Trus/publication/3908878_Medical_Image_Processing_Analysis_Visualization_in_Clinical_Research/links/56f2b65408aee25a6371bb78/Medical-Image-Processing-Analysis-Visualization-in-Clinical-Research.pdf)

Papastefanopoulos, V., Linardatos, P., & Kotsiantis, S. (3. June 2020). *COVID-19: A Comparison of Time Series Methods to Forecast Percentage of Active Cases per Population*. Preuzeto 10. September 2023 iz MDPI Open Access Journals: <https://www.mdpi.com/2076-3417/10/11/3880>

Prskalo, L. N. (n.d.). *Uvod u analizu i vizualizaciju podataka*. Preuzeto 9. August 2023 iz Merlin.

RioMed. (6. September 2021). *Healthcare Reporting Dashboards: Data Visualisation that saves time and costs*. Preuzeto 10. September 2023 iz RioMed: <https://www.riomed.com/healthcare-reporting-dashboards-data-visualisation-that-saves-time-and-costs/>

Salvatierra, S., & Alvarez, M. (March 2020). *Characterization of the flow of patients in a hospital from complex networks*. Preuzeto 10. September 2023 iz ResearchGate: [https://www.researchgate.net/figure/Illustrative-example-of-a-network-graph-involving-the-routes-assistances-of-eight\\_fig3\\_330116221](https://www.researchgate.net/figure/Illustrative-example-of-a-network-graph-involving-the-routes-assistances-of-eight_fig3_330116221)

Seaborn. (n.d.). *seaborn.scatterplot*. Preuzeto 20. August 2023 iz Seaborn: <https://seaborn.pydata.org/generated/seaborn.scatterplot.html>

Sisense. (n.d.). *Healthcare Dashboards: Examples of Visualizing Key Metrics & KPIs*. Preuzeto 9. September 2023 iz Sisense: <https://www.sisense.com/blog/healthcare-dashboards-examples-visualizing-key-metrics/>

Tableau. (n.d.). *Health sciences center visualizes cancer research data*. Preuzeto 9. September 2023 iz Tableau: <https://www.tableau.com/louisiana-state-university-data-visualization>

Tenali, N., & Babu, G. R. (16. March 2023). *A Systematic Literature Review and Future Perspectives for Handling Big Data Analytics in COVID-19 Diagnosis*. Preuzeto 18. August 2023 iz Springer Link: <https://link.springer.com/article/10.1007/s00354-023-00211-8>

Vermeulen, M. J. (June 2011). *Emergency department visits during an Olympic gold medal television broadcast*. Preuzeto 10. September 2023 iz ResearchGate:

[https://www.researchgate.net/figure/Line-graph-showing-count-of-total-patients-visiting-an-emergency-department-based-on\\_fig1\\_51641468](https://www.researchgate.net/figure/Line-graph-showing-count-of-total-patients-visiting-an-emergency-department-based-on_fig1_51641468)

Waskom, M. (30. December 2022). *seaborn: statistical data visualization*. Dohvaćeno iz PyPI: <https://pypi.org/project/seaborn/>

Wikipedia. (22. July 2023). *Kaggle*. Preuzeto 22. August 2023 iz Wikipedia: <https://en.wikipedia.org/wiki/Kaggle>



## 9. Popis slika

Slika 1: Troškovi prijema po odjelima (Sisense) .....	7
Slika 2: Prikaz broja slučajeva koronavirusa po državi (Dunskiy, 2023).....	7
Slika 3: Prikaz podataka o raku u Kanadi 2017. godine (Abayowa, 2021).....	8
Slika 4: Prikaz klinika u kojima je dijagnosticiran rak dojke (Tableau) .....	8
Slika 5: Prikaz podataka o pacijentima na temelju odabranih kriterija (RioMed, 2021)..	9
Slika 6: Prikaz učestalosti pojedinih simptoma kod gastrointestinalnih bolesti (Chey, 2014).....	9
Slika 7: Prikaz hijerarhijske strukture zdravstvenih usluga (ChartExpo).....	10
Slika 8: Prikaz promjene broja hospitalizacija tijekom 24 sata (Vermeulen, 2011) .....	10
Slika 9: Prikaz godina pacijenata sa boli u donjem dijelu leđa (Froud, 2018).....	11
Slika 10: Prikaz distribucije ukupnog broja slučajeva koronavirusa po zemlji 2020. godine (Papastefanopoulos, Linardatos, & Kotsiantis, 2020) .....	11
Slika 11: Prikaz rute pacijenata tokom tjedan dana (Salvatierra & Alvarez, 2020).....	12
Slika 12: Učitavanje skupa podataka u Google Colab.....	16
Slika 13: Ispis prvih nekoliko redaka podataka.....	16
Slika 14: Ispis tipova podataka za svaki stupac.....	17
Slika 15: Ispis dimenzije dataframe-a.....	17
Slika 16: Uklanjanje nedostajućih vrijednosti.....	17
Slika 17: Stvaranje i prikaz kružnog dijagrama .....	18
Slika 18: Stvaranje i prikaz poboljšanog kružnog dijagrama.....	19
Slika 19: Stvaranje stupčastog dijagrama .....	19
Slika 20: Prikaz ukupnog broja slučajeva po kontinentima stupčastim dijagramom .....	20
Slika 21: Stvaranje geografske mape.....	21
Slika 22: Prikaz broja slučajeva koronavirusa po državama geografskom mapom .....	22
Slika 23: Stvaranje dijagrama raspršenosti .....	23

Slika 24: Prikaz ukupnog broja slučajeva i smrtnih slučajeva u državama Južne Amerike dijagramom raspršenosti .....	23
---	----

## 10. Prilozi

Prilog 1- programski kod izrađenih dijagrama